Application of spatial verification methods to ensembles

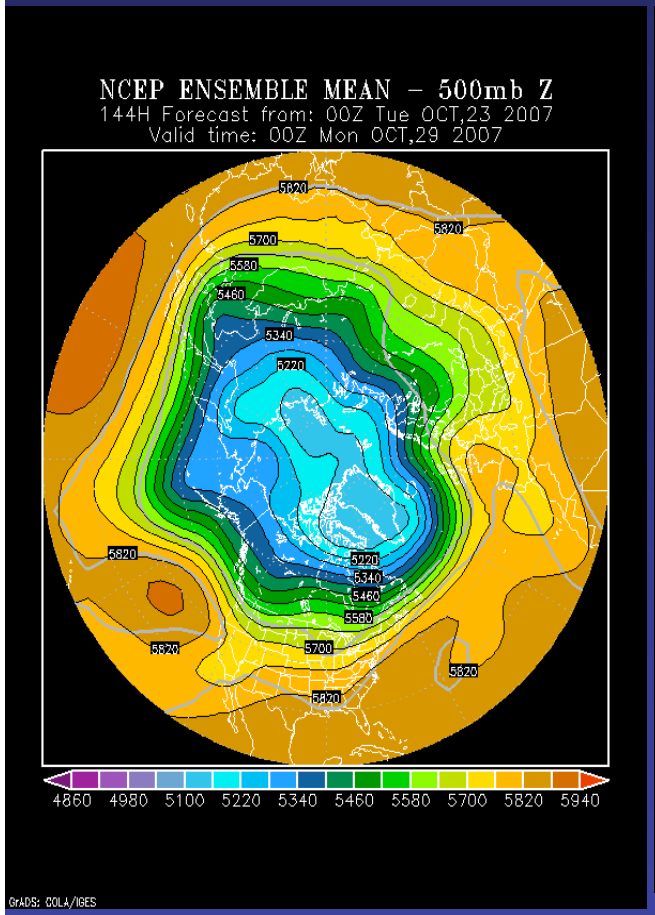
William A. Gallus, Jr.[1] and Beth Ebert[2] (with assistance from many others)

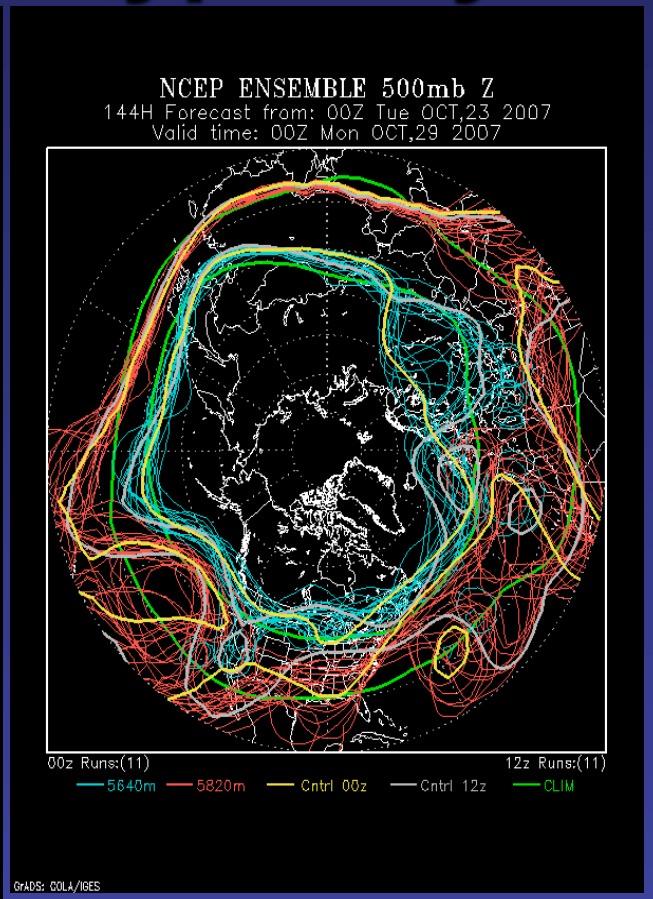
[1]Iowa State University

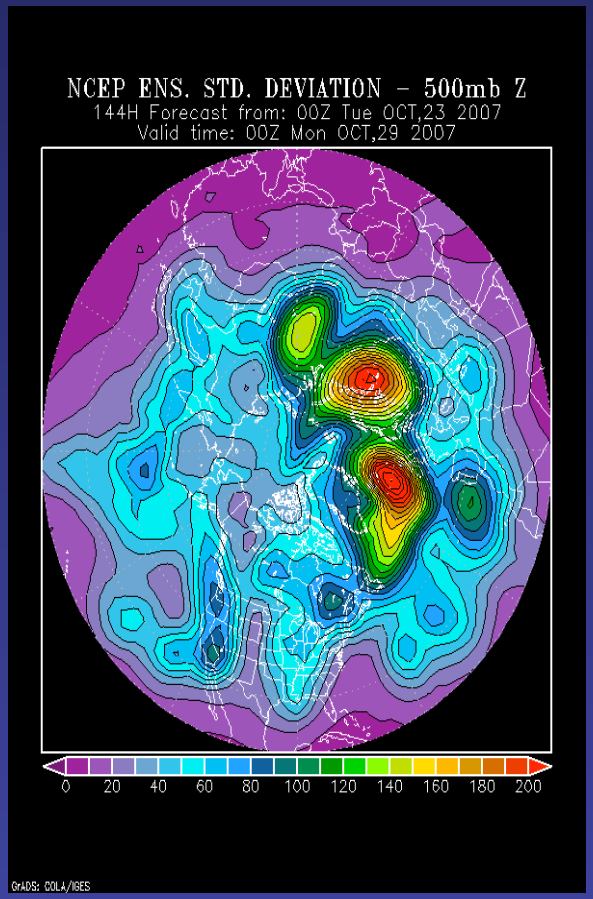
[2]CAWCR Weather & Environmental Prediction Group

August 27, 2009

Ensemble Forecasts - how are these typically used?







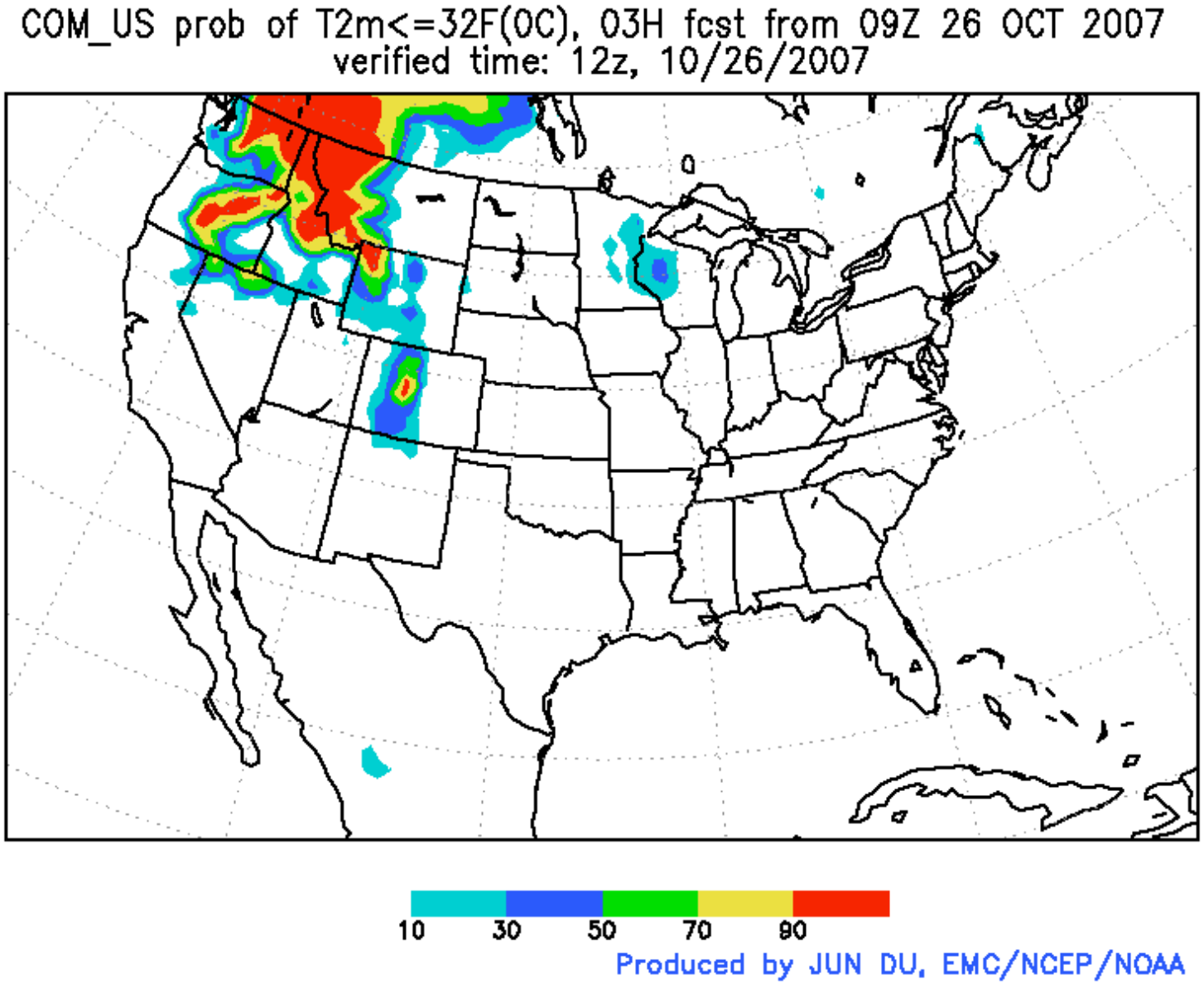
Mean

Spaghetti Plots – could be thought of as "objects"

(uncertainty)

Standard Deviation (uncertainty)

Or…. Probability information

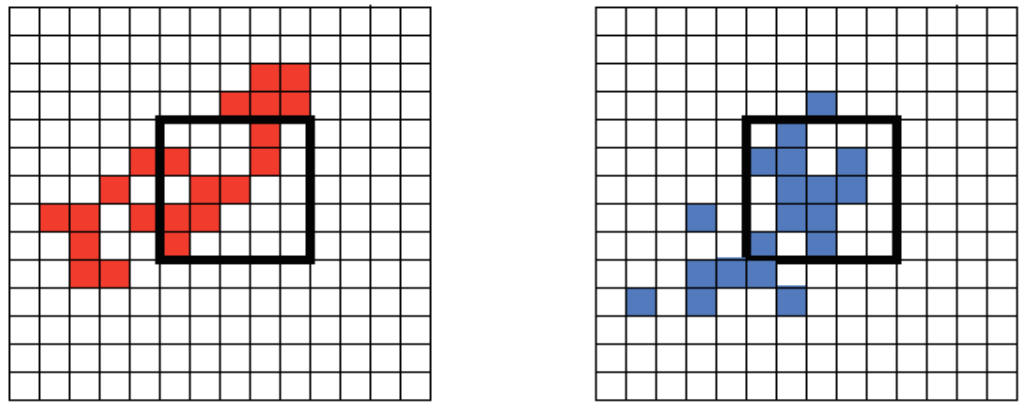


How can/should we apply spatial verification techniques to ensembles?

- Obviously any technique can be applied to the mean (which can be determined in a variety of ways)
- Likewise, probabilities can be used as the forecast field (e.g., a threshold probability can be used to define a system/object)

Example already in literature: **Fractions Skill Score**

- Compares fractional coverage in forecast with fractional coverage in observations (Roberts and Lean 2005)
- Has been used in examination of time-lagged 4 km ensemble by UKMET (Mittermaier 2007)

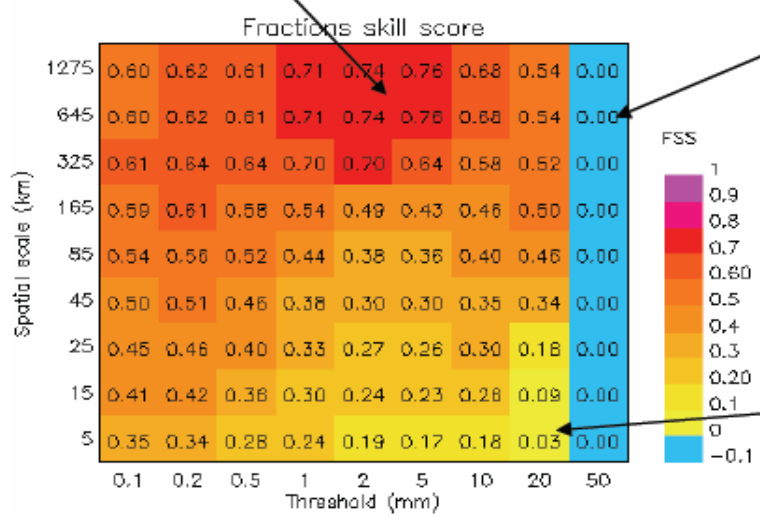


$$\mathrm{FSS} = 1 - \frac{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(P_{fcst} - P_{obs})^2}{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}P_{fcst}{}^2 + \frac{1}{\mathrm{N}}\sum_{i=1}^{N}P_{obs}{}^2}$$

Decision model – Useful forecast has similar frequency of forecast and observed events within a spatial window

Greatest skill for light rain, large scale

No skill at heaviest rain rate (none fcst)



Little skill for heavy rain, small scales

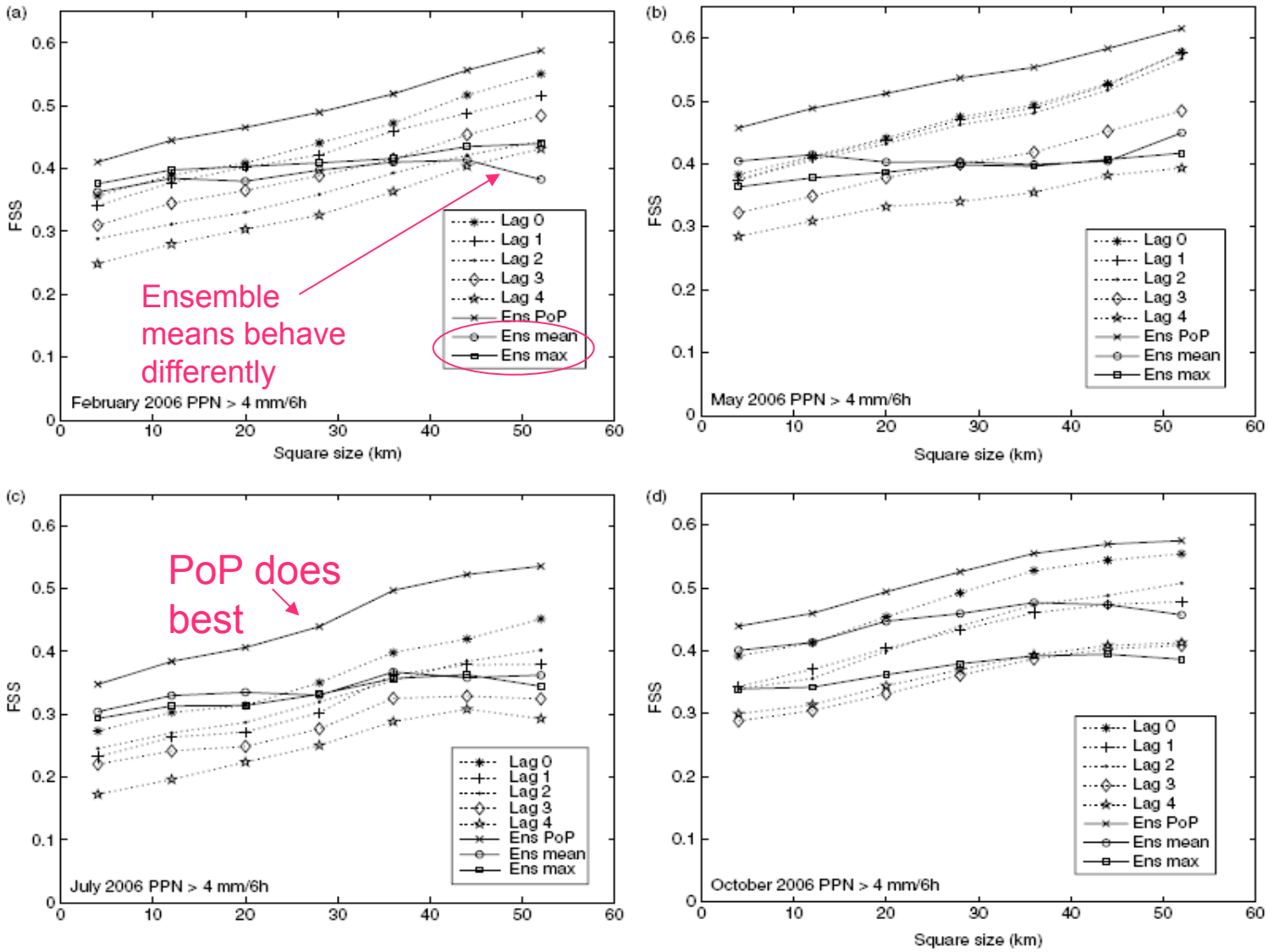
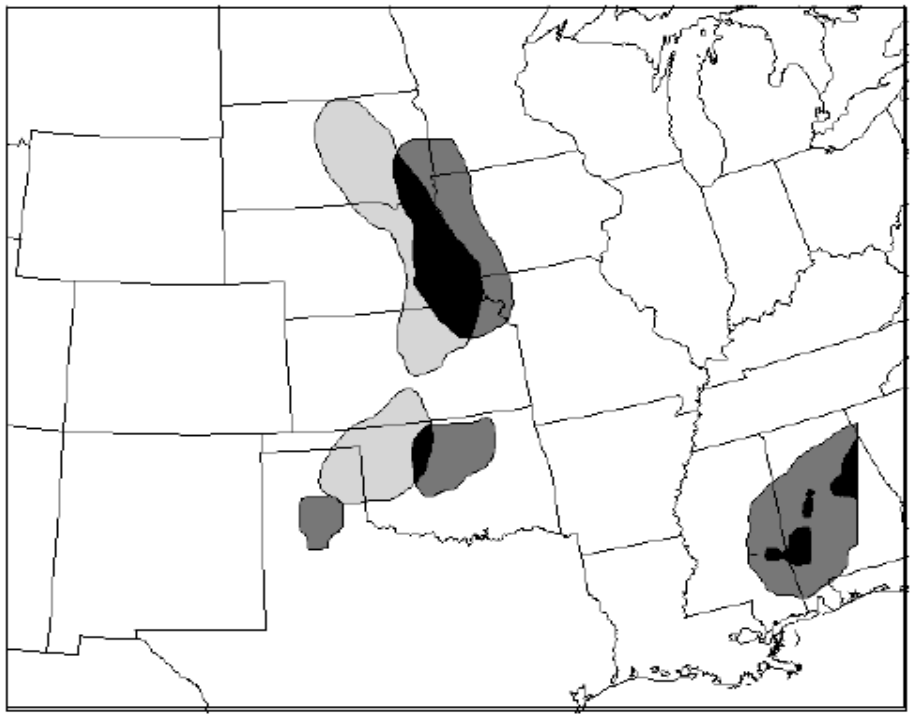


Figure 4. Mean monthly FSS at 4 mm/6 h, as a function of averaging lengths up to 52 km, for: (a) February 2006; (b) May 2006; (c) July 2006; (d) October 2006. Solid lines represent ensemble products; dotted lines represent individual deterministic forecasts (ensemble members).

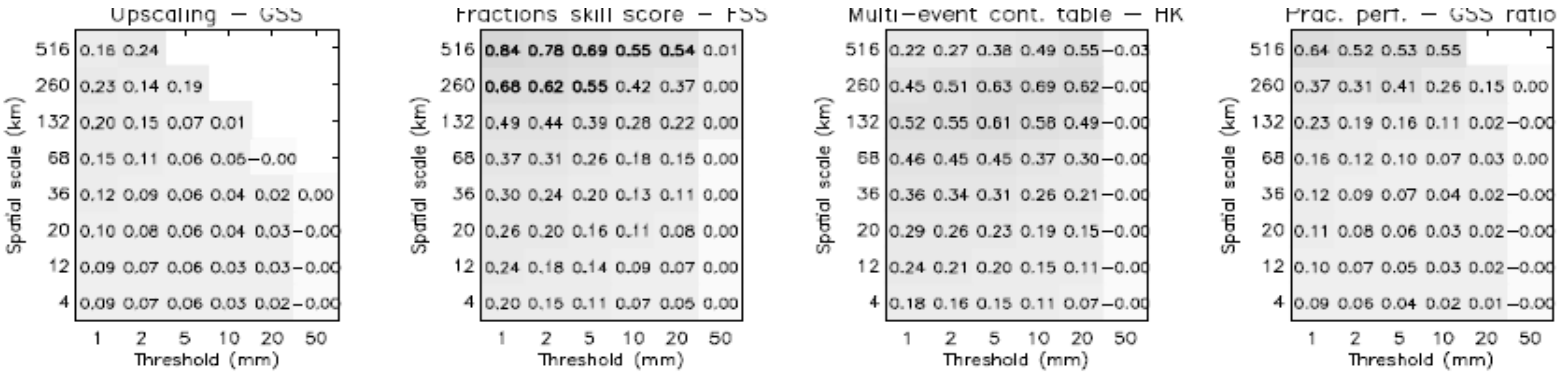
From Mittermaier 2007 QJRMS

What about other new spatial techniques?

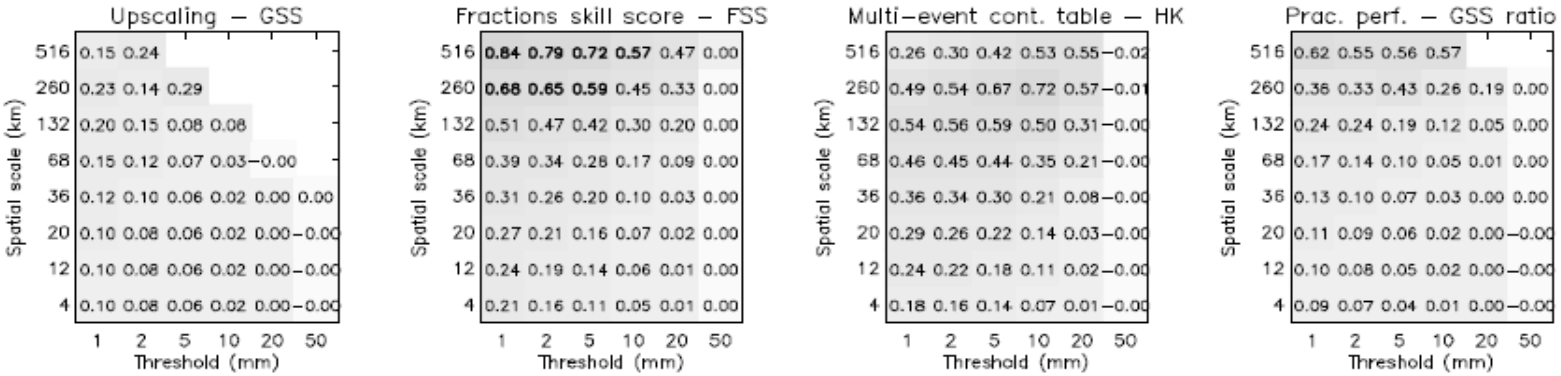
Fuzzy Verification (from Ebert 2009)



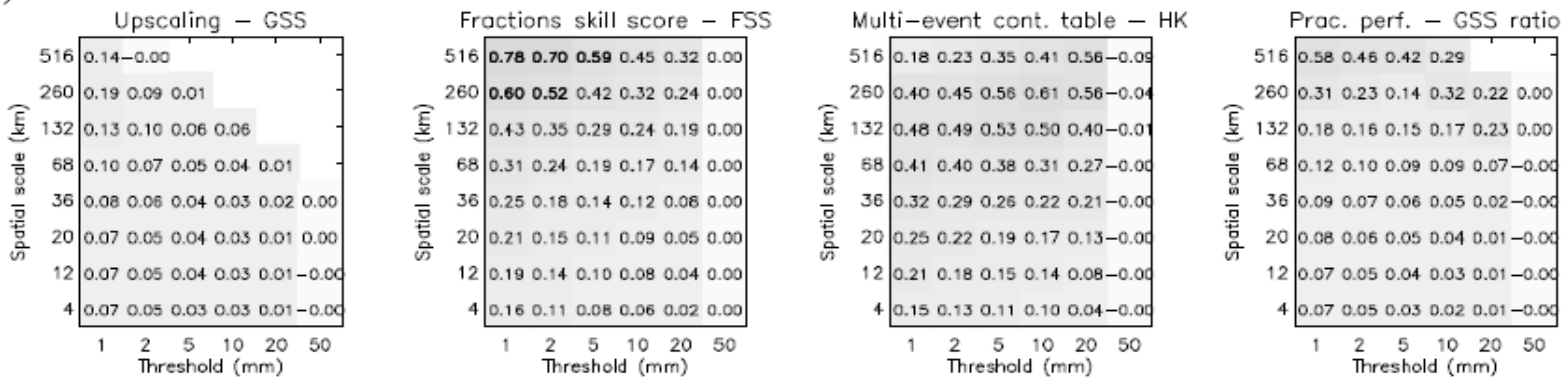
Fuzzified forecast in light gray, hindcast (obs) in dark gray – this could be done to an ensemble of forecasts.


Upscaling – GSS
Fractions skill score – FSS
Multi–event cont. table – HK
Prac. perf. – GSS ratio
Spatial scale (km)
Threshold (mm)
516 260 132 68 36 20 12 4
1 2 5 10 20 50

b)


Upscaling – GSS
Fractions skill score – FSS
Multi–event cont. table – HK
Prac. perf. – GSS ratio
Spatial scale (km)
Threshold (mm)
516 260 132 68 36 20 12 4
1 2 5 10 20 50

c)


Upscaling – GSS
Fractions skill score – FSS
Multi–event cont. table – HK
Prac. perf. – GSS ratio
Spatial scale (km)
Threshold (mm)
516 260 132 68 36 20 12 4
1 2 5 10 20 50

SAL (Structure, Amplitude, Location) – Wernli et al. 2009

- Could be applied to all members, generating cloud of SAL values. If a few were near (0,0,0), this would imply some members captured the event realistically.
- Not much done yet with application of this to ensembles (the case with many of the techniques)

Example from Wernli et al. (2009) paper – could represent an ensemble

date	model	domain	S	A	L	ME	RMSE	nRMSE	HSS	FBI
0601	2CAPS	large	0.46	0.21	0.06	0.07	2.75	0.03	0.28	0.92
0601	4NCAR	large	-0.22	0.13	0.07	0.04	2.79	0.03	0.30	0.86
0601	4NCEP	large	0.68	0.34	0.29	0.12	2.89	0.03	0.19	1.15
0601	2CAPS	south	-1.33	0.42	0.08	0.21	2.34	0.06	0.34	0.77
0601	4NCAR	south	-1.69	0.40	0.12	0.20	3.01	0.07	0.38	0.67
0601	4NCEP	south	-1.57	-0.12	0.26	-0.05	1.34	0.05	0.24	0.67
0603	2CAPS	south	-0.79	-0.02	0.26	0.00	0.70	0.12	0.19	0.99
0603	4NCAR	south	-0.92	-0.25	0.31	-0.02	0.71	0.12	0.13	0.69

Intensity-scale verification (Casati 2004)

- Used in Casati and Wilson (2007, MWR) to verify lightning probability forecasts, and thus could be used with any ensemble probability forecast.

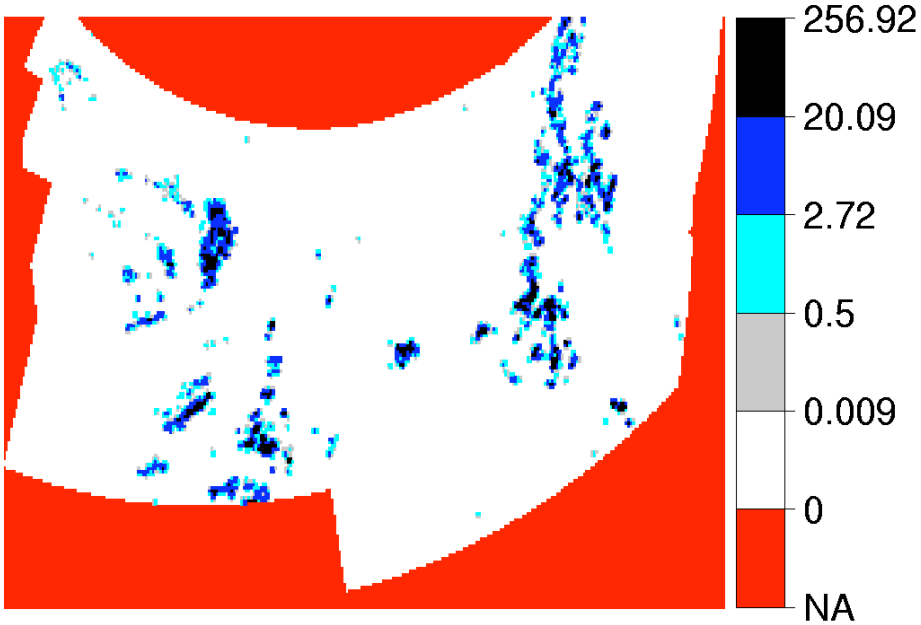
Case study

Forecast valid on 20/07/2004 in the 18:00 to 21:00 time window, 21h lead time. Thresholds:

T = 1/2 ANY lightning

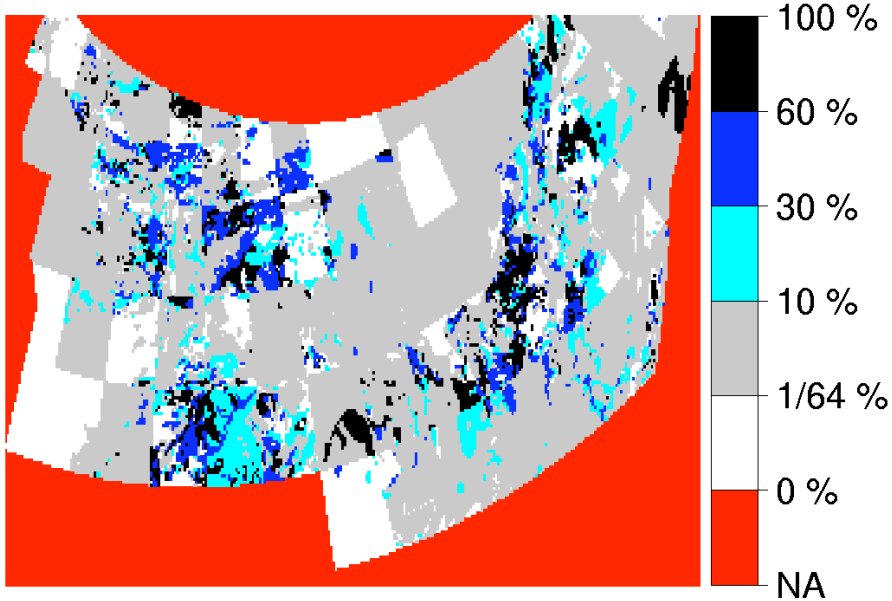
T = e^3 INTENSE lightning

lightning OBS (strikes/3h)

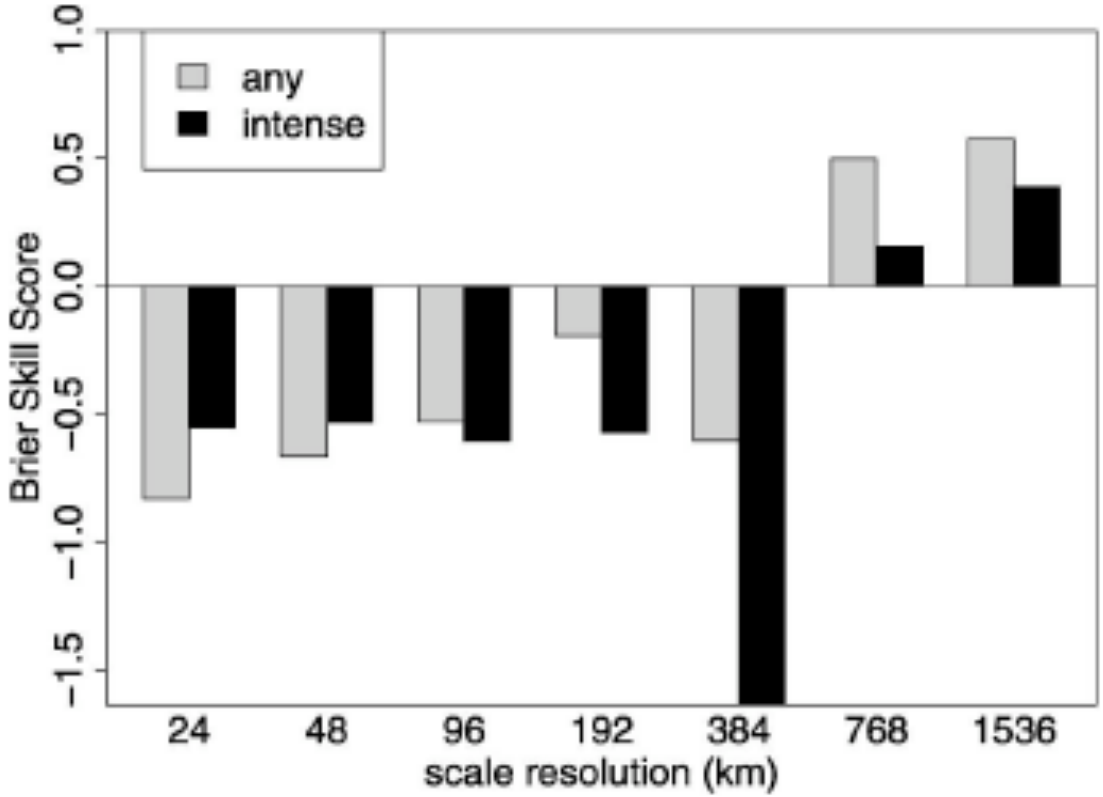


OBS frequency of ANY lightning

FOR prob of ANY lightning



Brier skill score on different scales



• Positive skill on large scales (> 700 km); negative skill on small scales (< 350 km); transition: 5th to 6th scale ~ 500 km (size of 5°x5° long-lat sectors)

• over-forecast of 400 km features for frequent lightning = poor skill

Other Scale-dependent verification

- Different type of scale dependent verification was performed on ECMWF EP
  S by Jung and Leutbecher (2008, QJRMS)

<u>Aims</u>

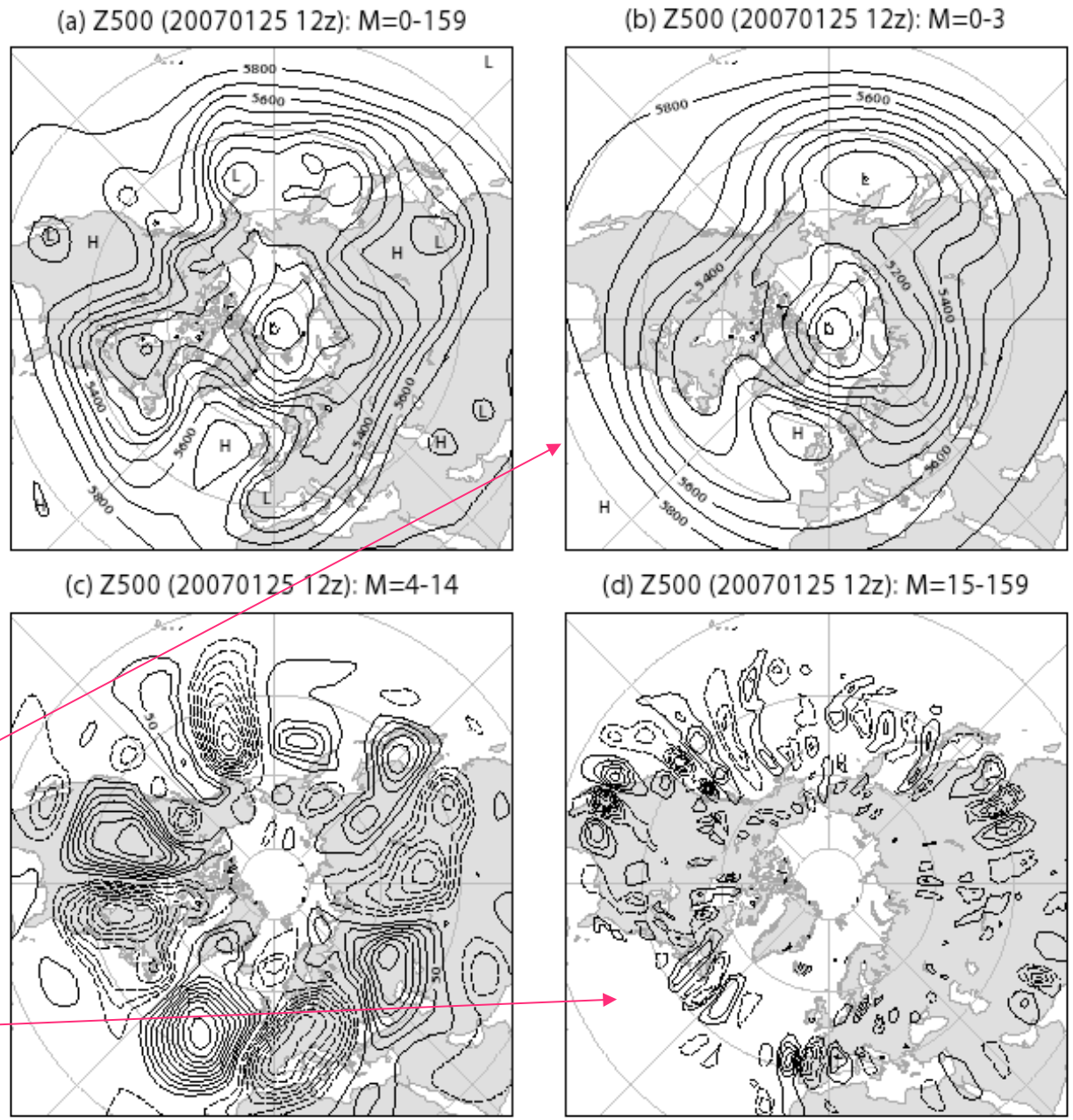
- Analyse the **scale dep**
  e
  ndency of the spread-skill relationship
- Eval
  u
  a
  te **skill for different scales** (BSS, RPSS)

Scale decomposition of Z500 fields

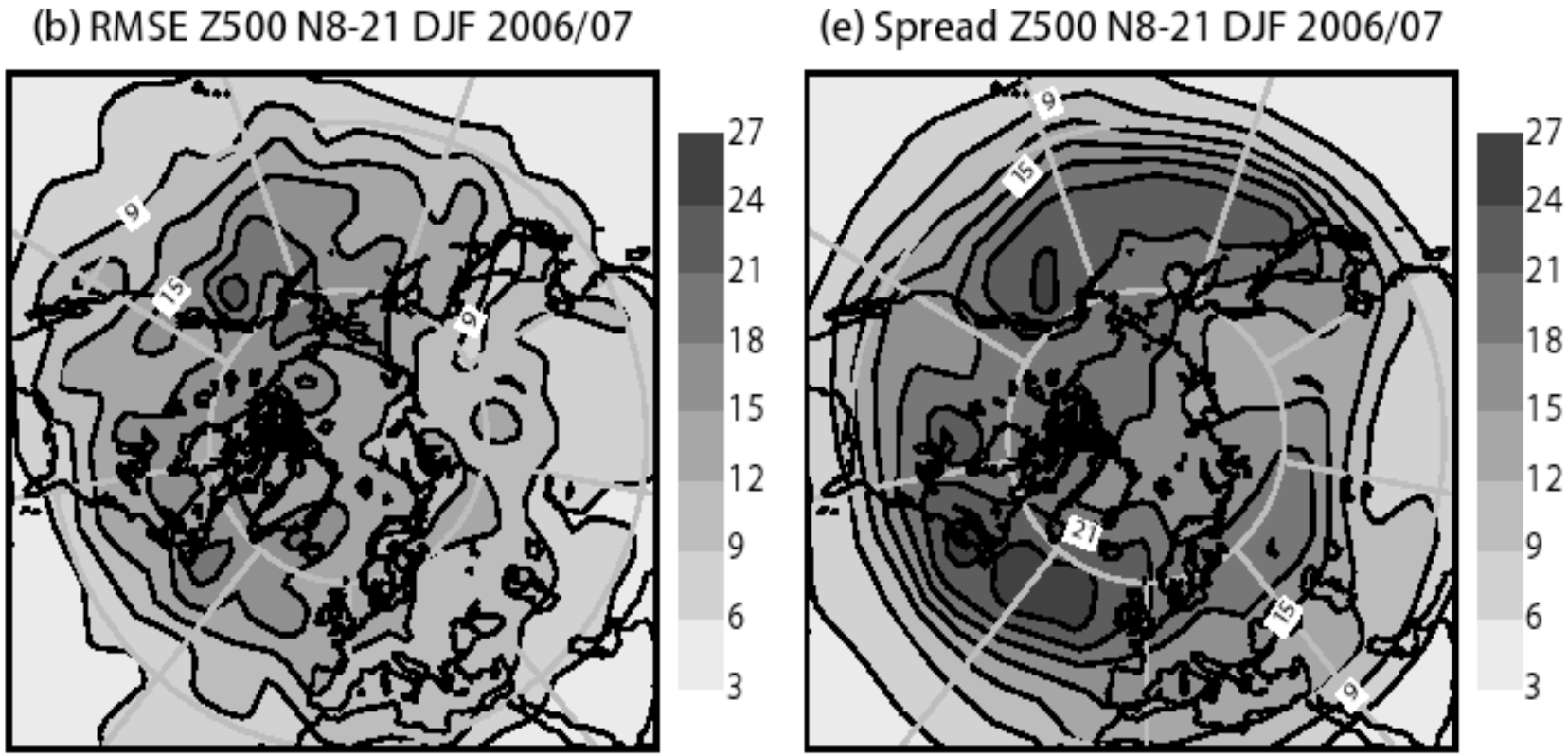
Filter: spherical harmonics

Two filters: total (N) and zonal (M) wavenumber filter

Separate:
planetary (b)
synoptic (c)
sub-synoptic (d)
scales

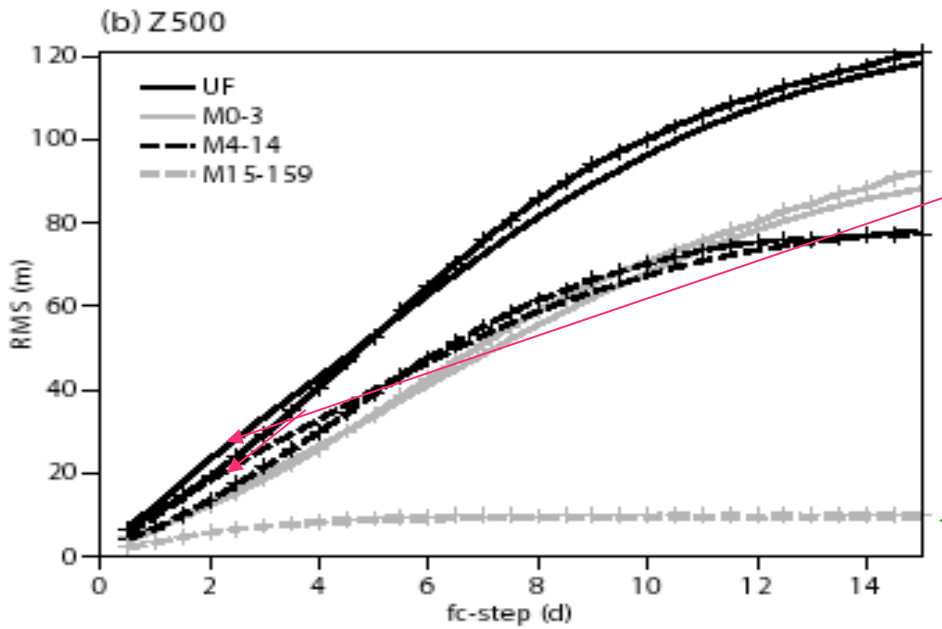


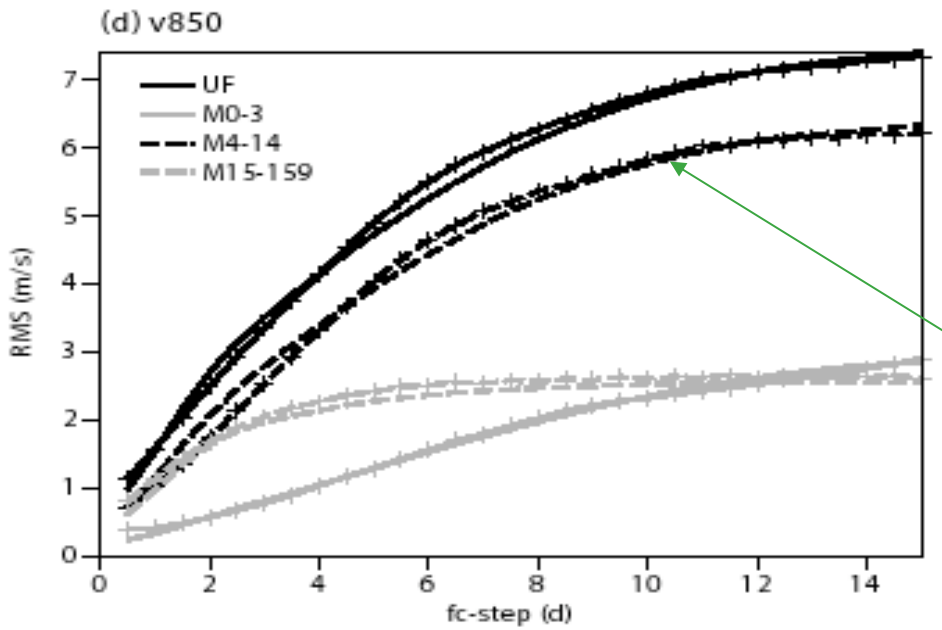
Spread-Skill Maps



Synoptic scales, D+2: over-dispersiveness of the ensemble

NOTE: max spread and error correspond the North Atlantic and North Pacific storm track regions





STDEV = no symbols, RMSE = symbols

Z500

Up to D+4: over-dispersiveness
Major contribution: synoptic

Beyond D+4: under-dispersiveness
Major contribution: synoptic and planetary

Synoptic and planetary scales equally contribute total error; sub-synoptic does not contribute much

V850

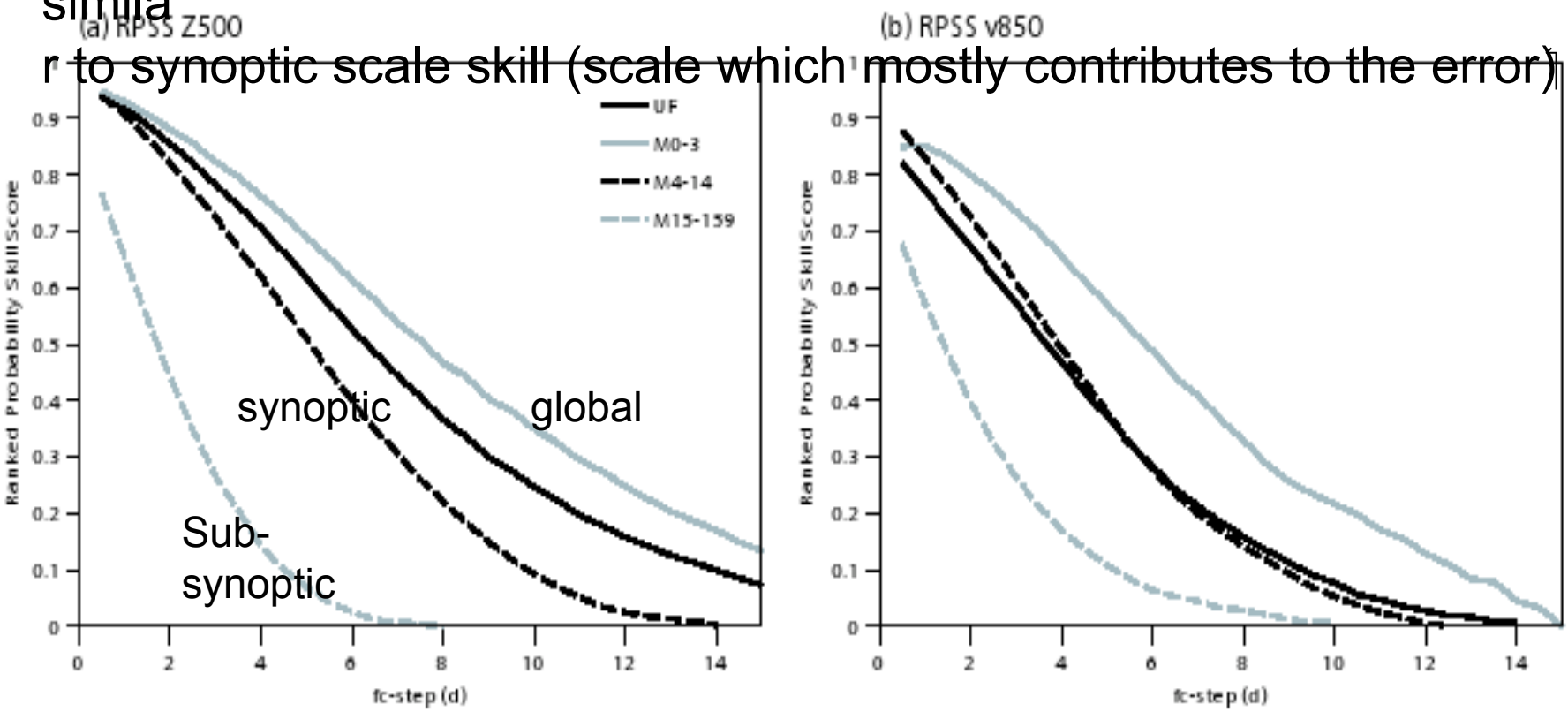
Up to D+4: over-dispersiveness
Major contribution: synoptic

Scales contribute in different proportion to total error, major contribution is from synoptic

RPSS on different scales: The larger the scale, the better the skill

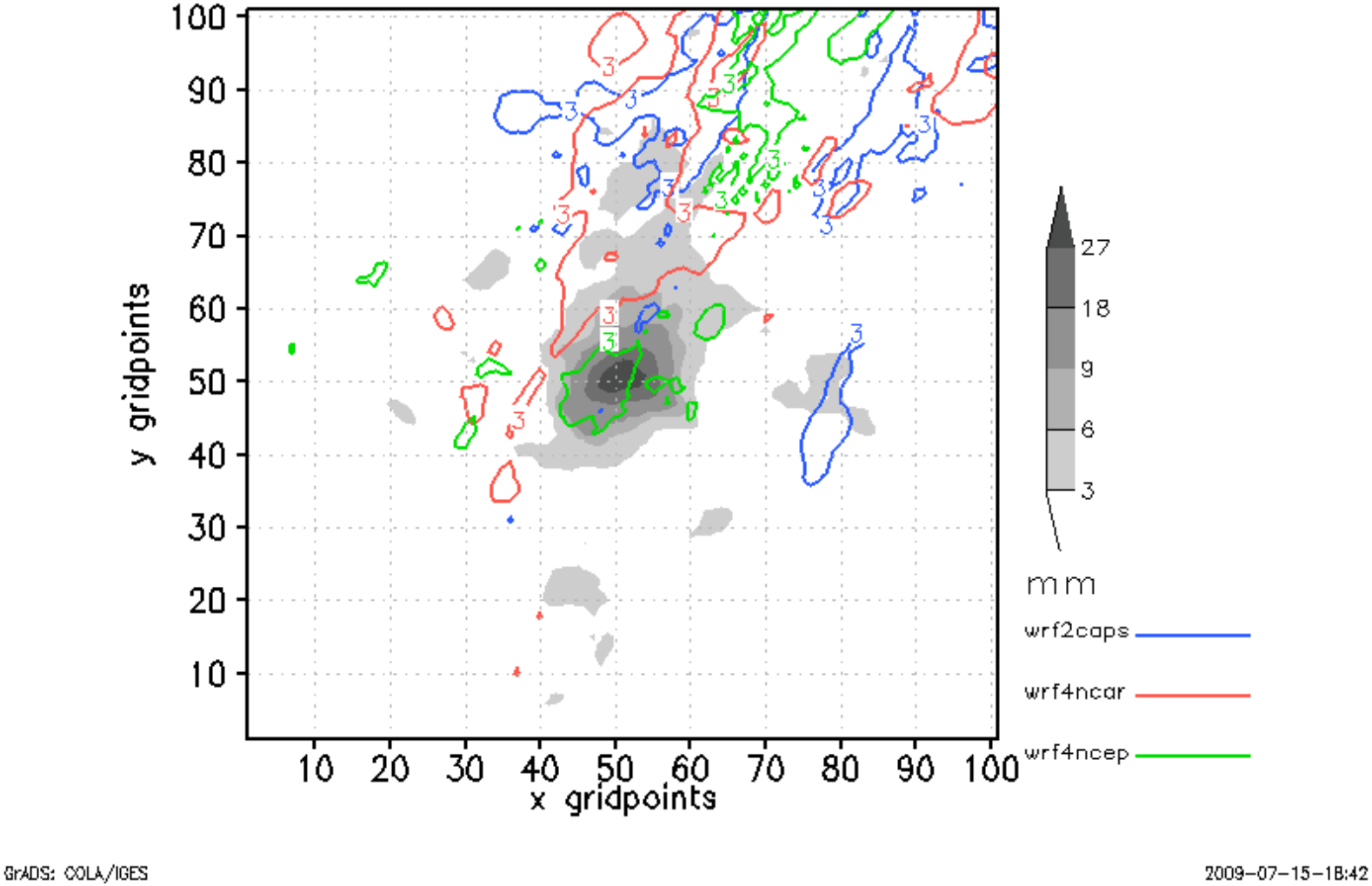
Z500: total skill between planetary and synoptic skill (scales which e
q
ually contributed to the error); sub-synoptic scales have worst skill

V850: total skill
simila
r to synoptic scale skill (scale which mostly contributes to the error)



Nachamkin Compositing Technique

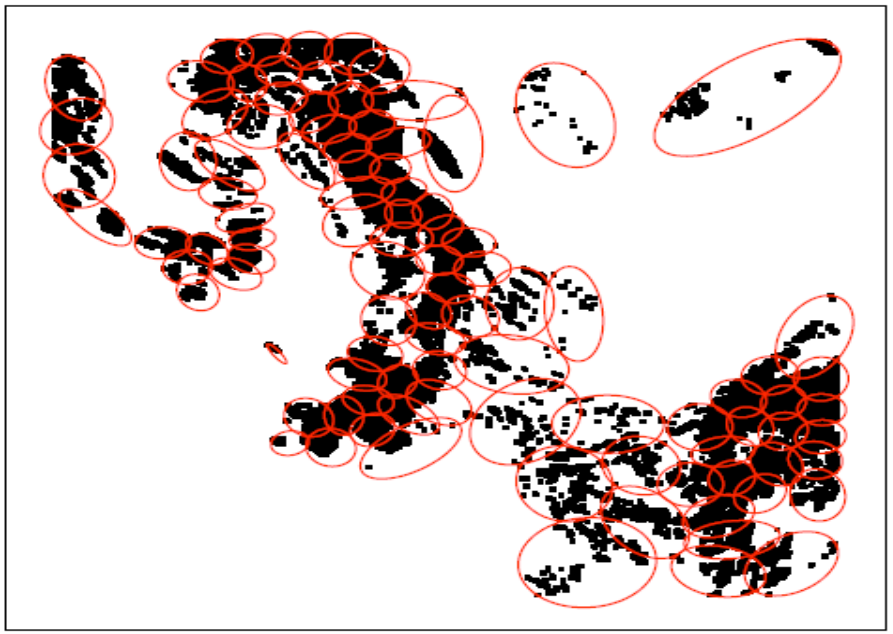
- Could be easily applied to ensembles to get a feel for systematic biases among members
- Better used with probability information



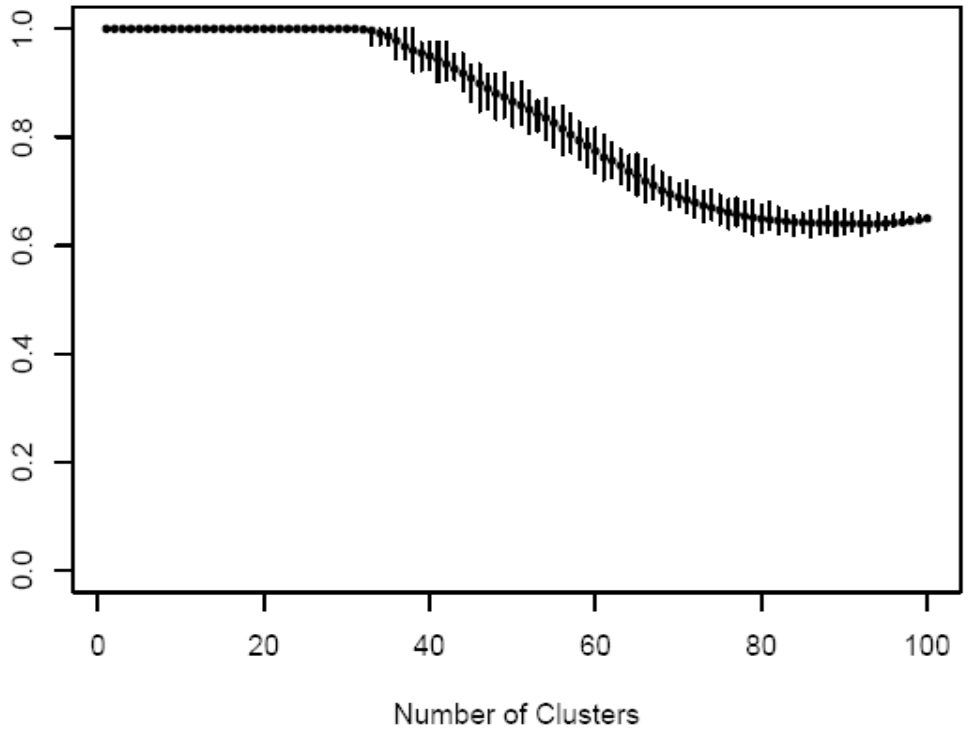
From Nachamkin – can think of this as 3 member ensemble

Cluster Analysis (Marzban et al. 2009)

- Could be applied to ensembles by clustering on field consisting of forecasts and observations from all members. -- This would be like superimposing fields from ensemble members and then identifying the “objects” in the composite fields



CSI



Variogram Analysis (Marzban et al. 2009)

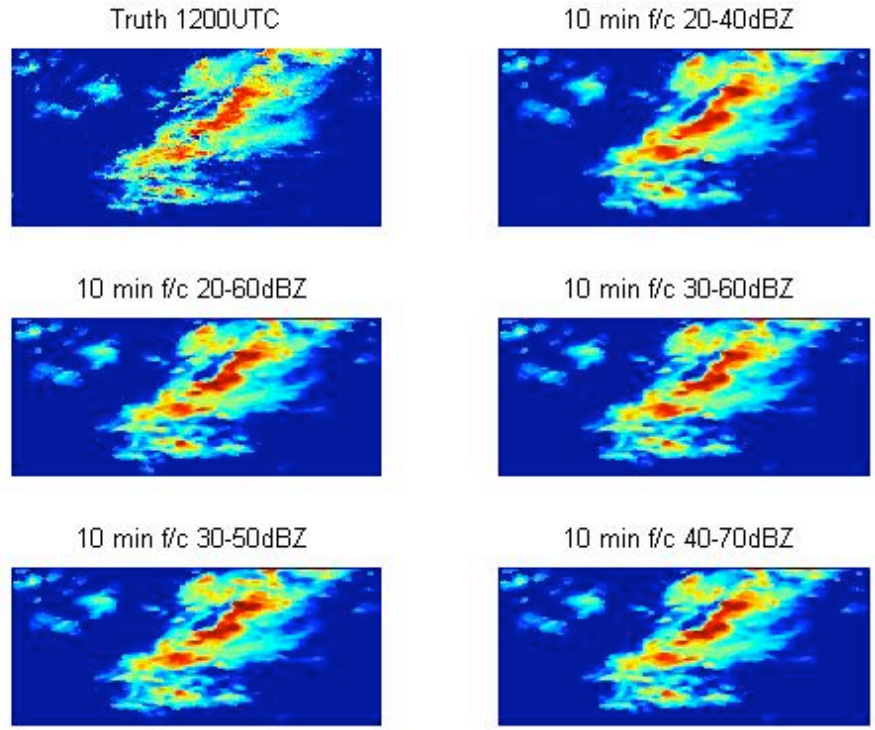
- Not as straightforward to apply since it is comparing textures of forecast field to observed, and if different ensemble members have different textures, combined field would not make much sense. Could do it with each member, though.

Optical flow (Marzban et al. 2009)

- Somewhat like variograms in the problems to face. If verifying all members of ensemble over a set of forecasts, and compositing the OF field, a good measure of reliability may be obtained (composite OF field should go to zero with a perfectly reliable ensemble over long series of forecasts)

Procrustes (Lack et al. 2009)

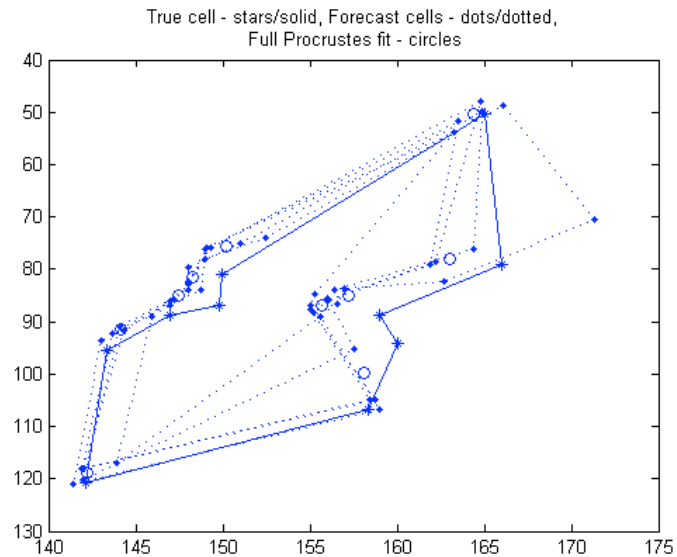
Example Ensemble forecast



Five 10-minute lead-time forecasts compared to the actual radar reflectivity at 1200 UTC on 13 January 2007.

Example fit for one cell in previous forecast

- The solid line shows the actual, the dotted line connecting dotted landmarks show the cells from each forecast realization, and the circles show the common full Procrustes fit.



Ensemble displacement error

	10 minute forecast	
f/c	Cell 1	Cell 2
x translation	0.300	-0.679
95% **credible set**	[0.092, 0.529]	[-0.923, -0.450]
y translation	0.534	-1.176
95% **credible set**	[0.349, 0.756]	[-1.423, -0.880]

Individual forecast translation errors

	10 minute forecast – cell 1	
f/c	x translation	y translation
1	-1.336	-2.723
2	0.024	1.098
3	1.895	3.634
4	-0.404	-0.576
5	-0.679	-1.176

	10 minute forecast – cell 2	
f/c	x translation	y translation
1	-0.294	-0.466
2	0.932	0.522
3	-0.698	-0.340
4	-0.197	-0.510
5	0.300	0.534

DAS (Displacement and Amplitude Score) – Keil and Craig 2009

- Can be used to identify best ensemble member

Case	Description of perturbation	DAS	DIS/D_{max}	AMP/I_0	Rank
1	Shift 3 pts right -5 pts up	0.18	0.07	0.12	1
2	Shift 6 pts right -10 pts up	0.24	0.13	0.11	2
3	Shift 12 pts right -20 pts up	0.38	0.26	0.12	3
4	Shift 24 pts right -40 pts up	0.69	0.50	0.19	5
5	Shift 48 pts right -80 pts up	1.20	0.30	0.90	7
6	Shift 12 pts right -20 pts up and intensity times 1.5	0.69	0.25	0.44	5
7	Shift 12 pts right -20 pts up and intensity minus 0.01	0.46	0.27	0.19	4

Object-based techniques…

Operational Forecaster Uncertainty Needs and Future Roles

DAVID R. NOVAK

NOAA/National Weather Service, Eastern Region Headquarters, Scientific Services Division, Bohemia, New York

DAVID R. BRIGHT

NOAA/National Weather Service, Storm Prediction Center, Norman, Oklahoma

MICHAEL J. BRENNAN

NOAA/National Weather Service, Hydrometeorological Prediction Center, Camp Springs, Maryland

(Manuscript received 20 March 2008, in final form 6 June 2008)

ABSTRACT

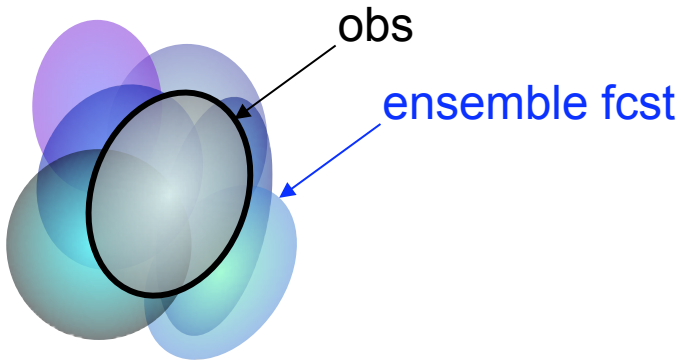
Key results of a comprehensive survey of U.S. National Weather Service operational forecast managers concerning the assessment and communication of forecast uncertainty are presented and discussed. The survey results revealed that forecasters are using uncertainty guidance to assess uncertainty, but that limited data access and ensemble underdispersion and biases are barriers to more effective use. Some respondents expressed skepticism as to the added value of formal ensemble guidance relative to simpler approaches of estimating uncertainty, and related the desire for feature-specific ensemble verification to address this skepticism. Respondents reported receiving requests for uncertainty information primarily from sophisticated users such as emergency managers, and most often during high-impact events. The largest request for additional training material called for simulator-based case studies that demonstrate how uncertainty information should be interpreted and communicated.

Respondents were in consensus that forecasters should be significantly involved in the communication of uncertainty forecasts; however, there was disagreement regarding if and how forecasters should adjust objective ensemble guidance. It is contended that whether forecasters directly modify objective ensemble guidance will ultimately depend on how the weather enterprise views ensemble output (as the final forecast or as a guidance supporting conceptual understanding), the enterprise's commitment to provide the nec-

December 2008 *WAF*

Forecasters have indicated a desire for **feature-specific verification approaches** (e.g., Ebert and McBride 2000, Davis et al. 2006) to be **applied to ensemble forecasts**.

Verifying "features" in ensembles



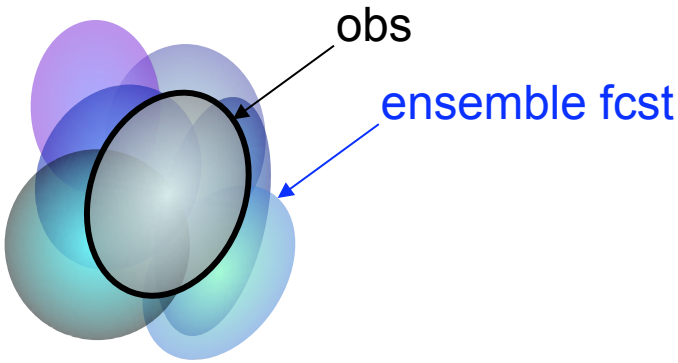
What might the ensemble forecast look like?

- spatial probability contour maps
- ensemble mean
- distributions of object properties (attributes)
  - location, size, intensity, etc.

Verifying "features" in ensembles

Strategies for verifying ensemble predictions of objects

1. Verify objects in probability maps
2. Verify "ensemble mean"
   - spatially averaged forecast objects
   - generated from average object properties
3. Verify distributions of object properties
   - many samples – use probabilistic measures
   - individual cases

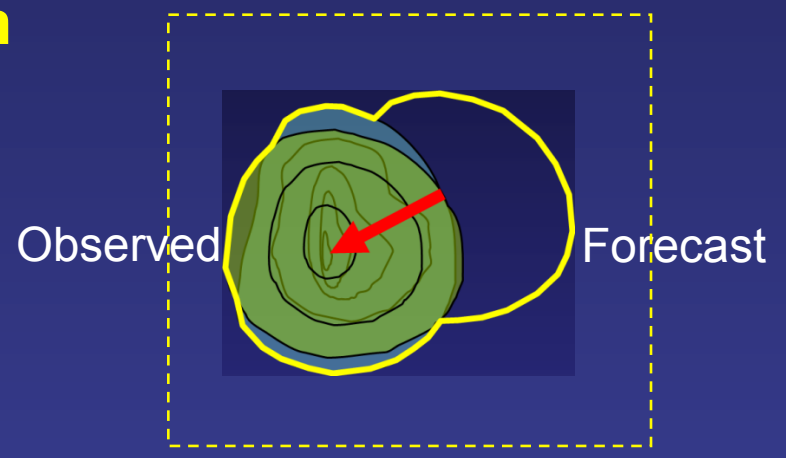


Object-based Methods

- Contiguous Rain Area (CRA – Ebert and McBride 2000)
- Method for Object-based Diagnostic Evaluation (MODE – Davis et al. 2006a,b)

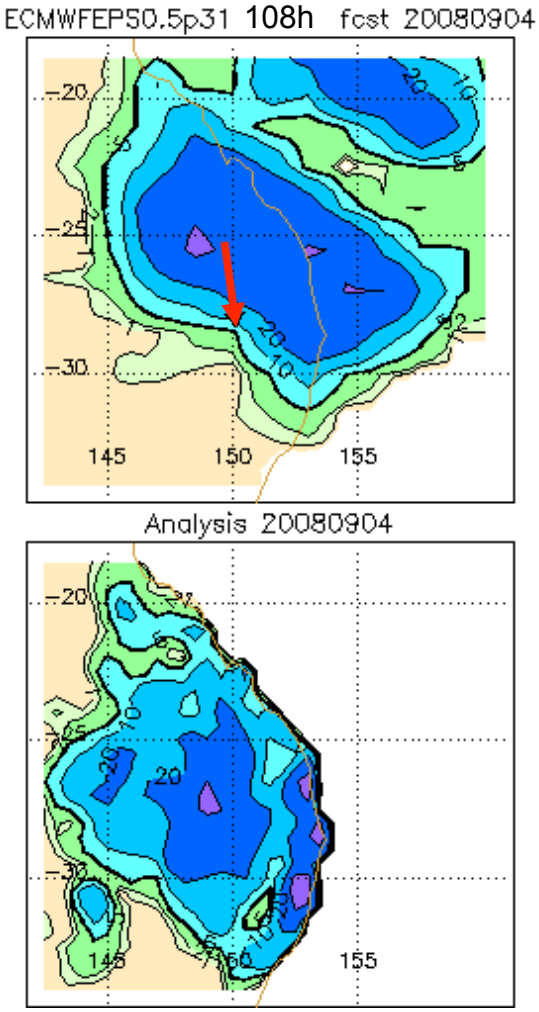
How does CRA work?

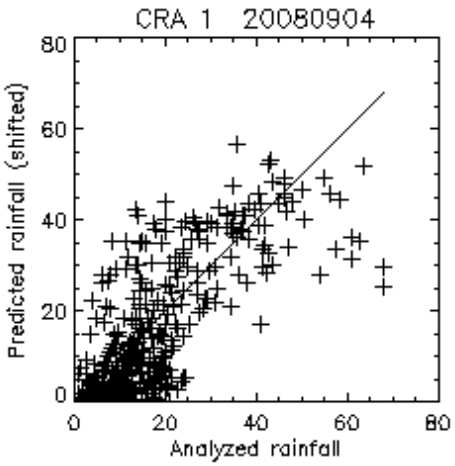
- Find Contiguous Rain Areas in the fields to be verified



- Define a rectangular search box around CRA to look for best match between forecast and observations
- Displacement determined by shifting forecast within the box until MSE is minimized or correlation coefficient is maximized

CRA verification for an ensemble member





ECMWFEPS0.5p31 84-** fcst 20080904 n=387
(-33.50°,143.00°) to (-19.00°,160.00°)
Verif. grid=0.500° CRA threshold=5.0 mm/d

	Analysed	Forecast
# gridpoints ≥5 mm/d	358	233
Average rainrate (mm/d)	17.88	15.46
Maximum rain (mm/d)	68.07	56.66
Rain volume (km³)	19.12	16.53

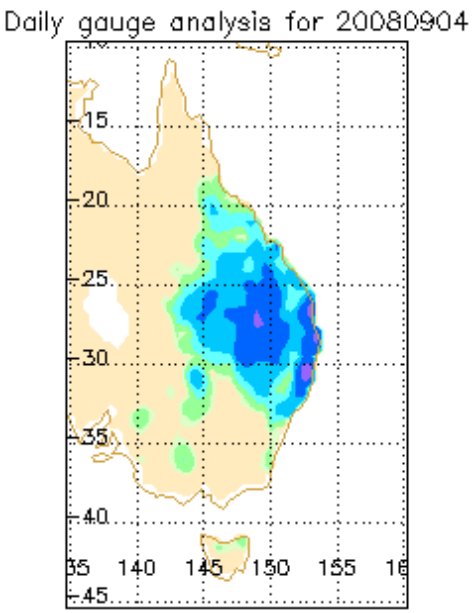
Displacement (E,N) = [-0.50°,3.00°]

	Original	Shifted
RMS error (mm/d)	17.84	10.29
Correlation coefficient	0.243	0.750

Error Decomposition:

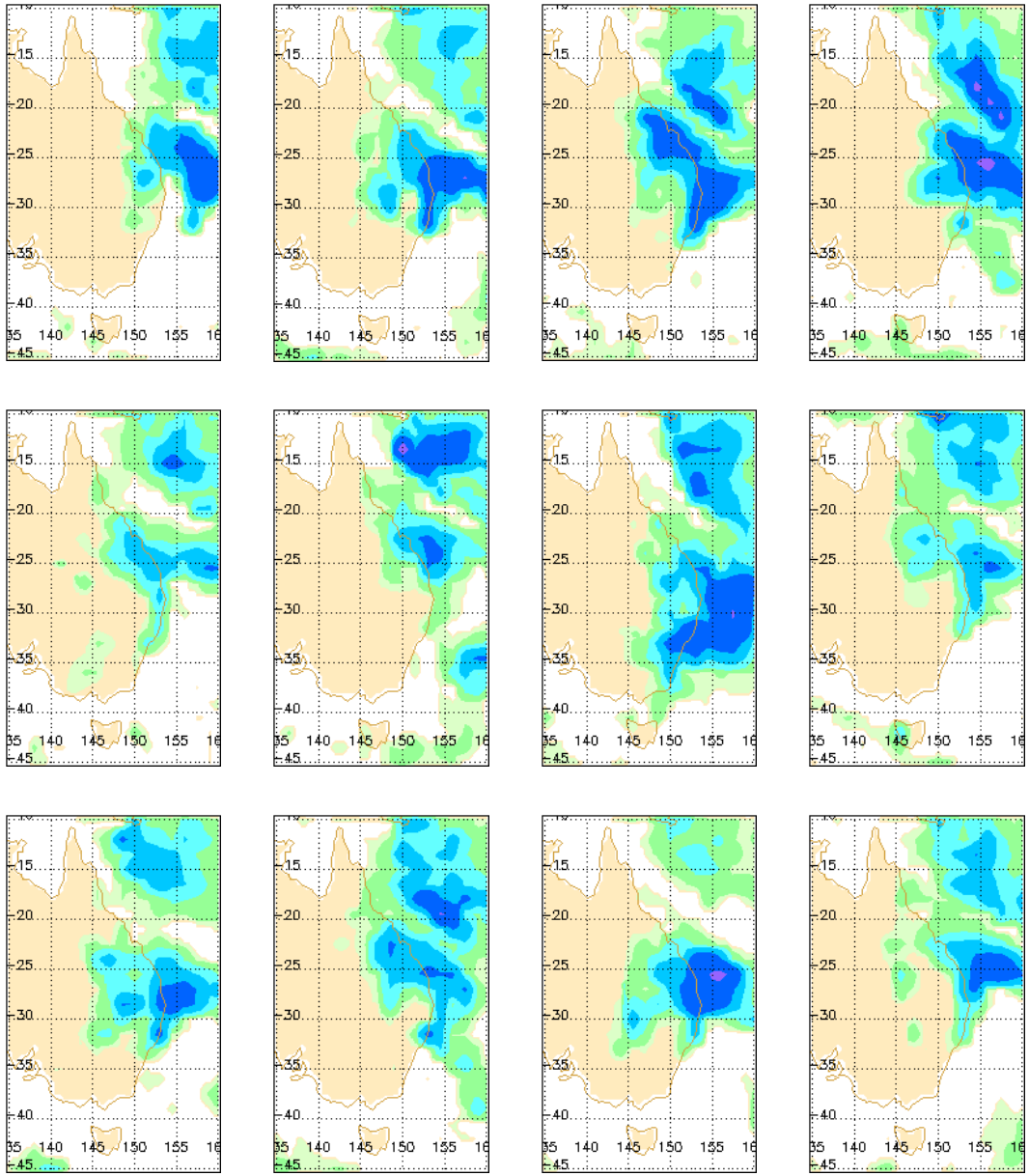
Displacement error	66.7%
Volume error	1.8%
Pattern error	31.4%

Example: East coast low, September 2008



Observed
00 UTC 5 Sept 2008

ECMWF EPS
4.5 day forecast
(first 12 of 51 members)



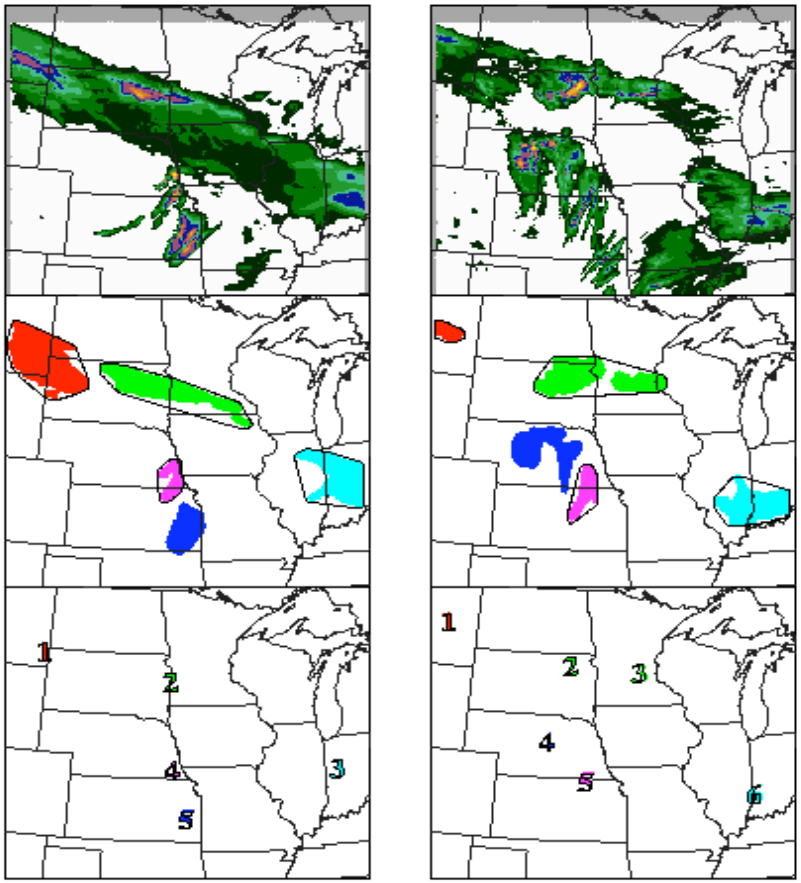
How does MODE work?

- Identifies objects using convolution-thresholding
- Merging and matching are performed via fuzzy logic approaches (systems don’t have to be contiguous)

Example of MODE output

MODE: APCP_06 at SFC

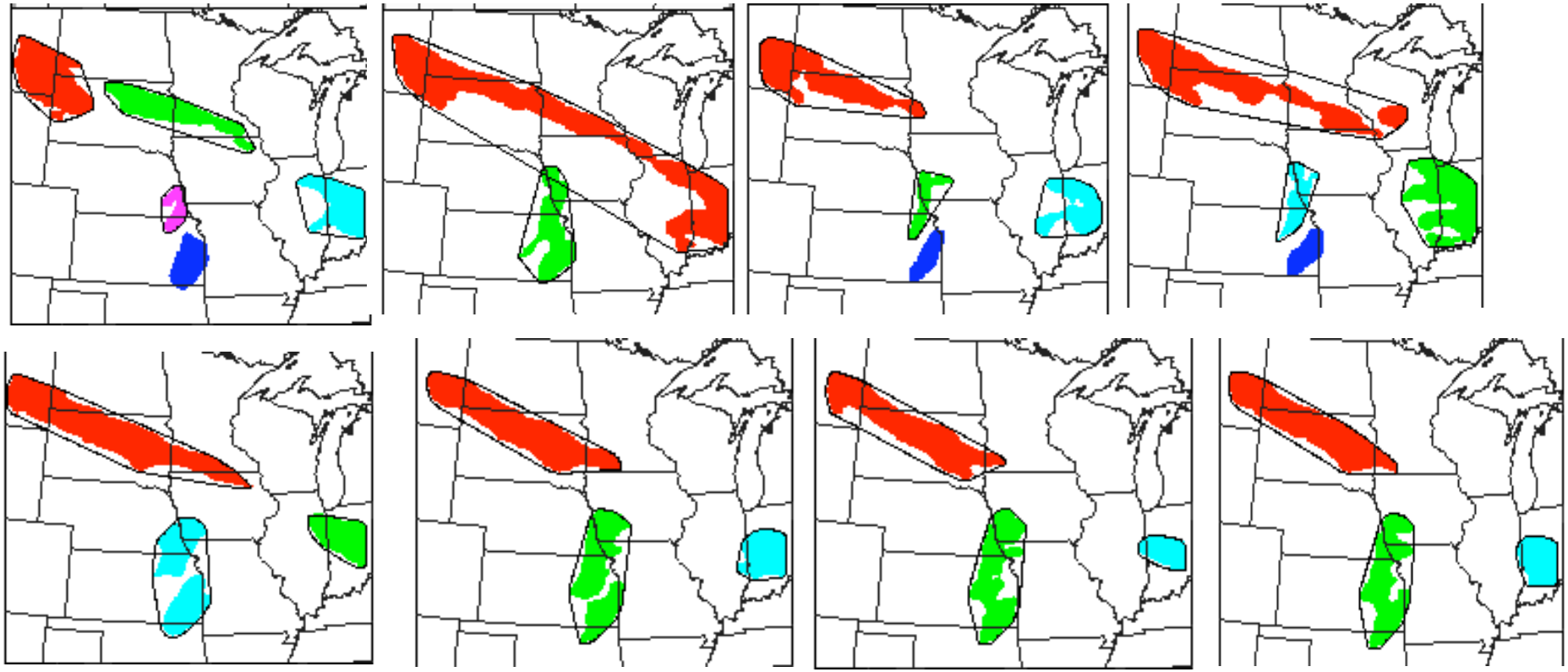
Forecast | Observation



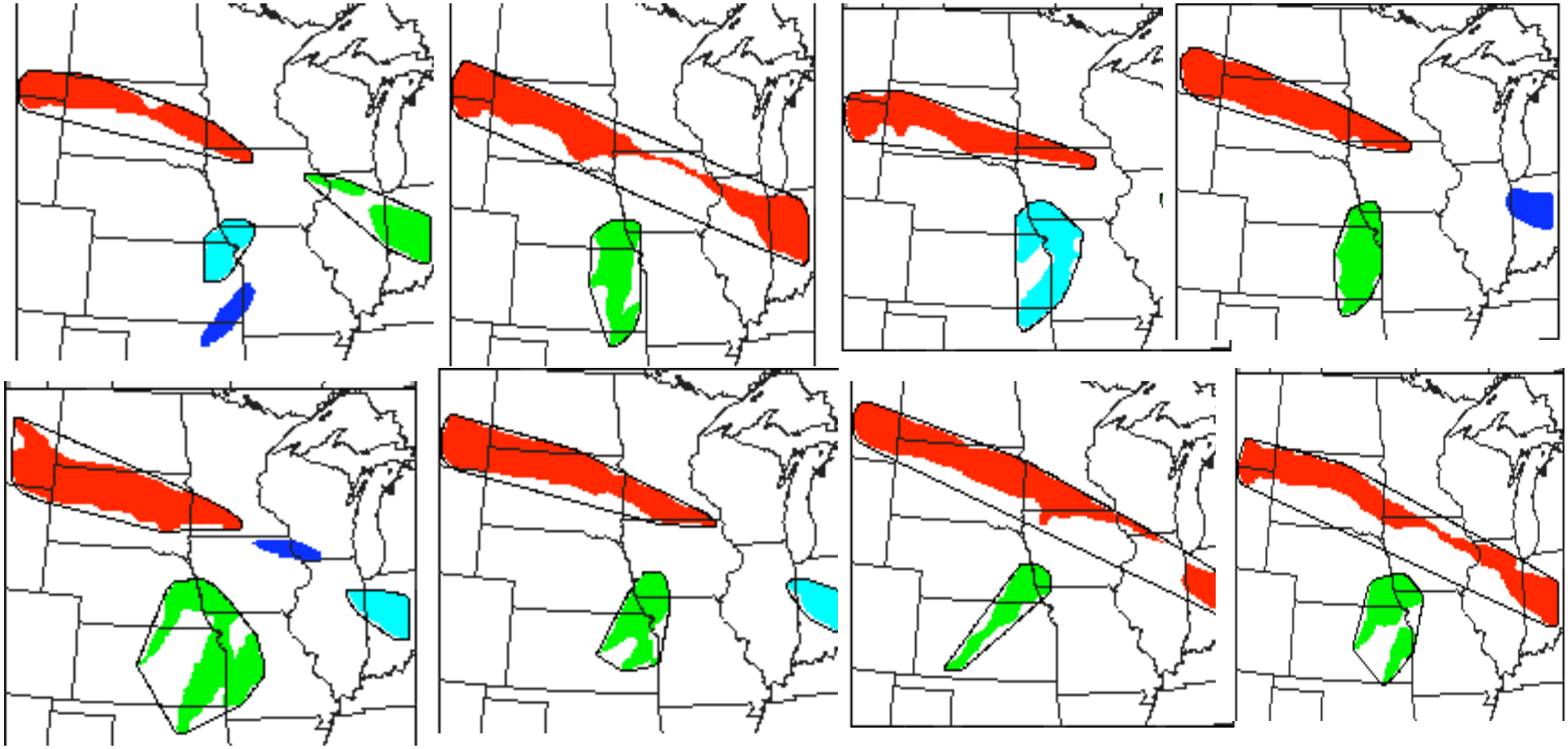
Fcst	Obs	Interest
4	5	0.9736
2	2	0.9254
3	6	0.9233
1	1	0.8798
2	3	0.7256
5	5	0.6837
4	4	0.6244
2	4	0.6110
5	4	0.5884
1	2	0.5652
1	3	N/A
4	1	N/A
3	3	N/A
4	3	N/A
5	3	N/A
1	4	N/A
5	1	N/A
3	4	N/A
2	1	N/A
3	1	N/A
1	5	N/A
2	5	N/A
3	5	N/A
3	2	N/A
4	2	N/A
1	6	N/A
2	6	N/A
5	2	N/A
4	6	N/A
5	6	N/A

Init Time:	Apr 6, 2006 00:00:00
Valid Time:	Apr 7, 2006 00:00:00
Lead Time:	24 hours
Accum Time:	6 hours
Fuzzy Engine Weights	
Centroid Distance:	2.00
Boundary Distance:	4.00
Convex Hull Distance:	0.00
Angle Difference:	1.00
Area Ratio:	1.00
Intersection/Area:	2.00
Complexity Ratio:	0.00
Intensity Ratio:	0.00
Total Interest Thresh:	0.70

	Forecast	Observation
Mask Missing:	on	on
Mask Gr/Ply:	off/off	off/off
Raw Thresh:	<= 1000.00 mm	>= 0.00 mm
Conv Radius:	3 gs	3 gs
Conv Thresh:	>= 6.25 mm	>= 6.25 mm
Area Thresh:	>= 10 gs	>= 10 gs
Inten Perc:	102 th	102 th
Inten Thresh:	>= 1000.00 mm	>= 1000.00 mm
Merge Thresh:	>= 5.00 mm	>= 5.00 mm
Merging:	thresh	thresh
Matching:	match/fcst merge	
Simple(M/U):	5 (4/1)	6 (5/1)
Composites:	4	4



April 6 18-24 hour forecast from Mixed Physics/Dynamics ensemble



Same forecast but from mixed initial/boundary condition ensemble

Tests to be discussed…

- Examination of rain area, volume, rate, and displacement forecasts using MODE and CRA for two sets of WRF ensembles in central U.S. (Gallus 2009, WAF) - EMPHASIS ON TRENDS IN SPREAD AND AGREEMENT WITH TRADITIONAL APPROACHES
- Examination of ECMWF EPS forecasts for heavy rain events in Australia – EMPHASIS ON FORECASTING

Methodology for WRF ensembles

- Precipitation forecasts from two sets of ensembles were used as input into MODE and CRA:

1) Two 8 member 15 km grid spacing WRF ensembles (Clark et al. 2008), one with mixed physics alone (Phys), the other with perturbed IC/LBCs alone (IC/LBC) -- initialized at 00 UTC

Two 8-member ensembles

- 6h precipitation forecasts from both ensembles over 60 h integration periods for 72 cases were evaluated
- Clark et al. (2008) found that spread & skill initially were better in Phys than in IC/LBC but after roughly 30h, the lack of perturbed LBCs "hurt" the Phys ensemble, and IC/LBC had faster growth of spread, and better skill later
- Also, a diurnal signal was noted in traditional skill/spread measures

Do these same trends show up in the object parameters?

Methodology (cont.)

2) Second set included 5 members of a 10 member 4 km ensemble (ENS4) run by CAPS with mixed LBCs/IC/physics and 5 similar members of a 15 member 20 km ensemble (ENS20), studied in Clark et al. (2009)

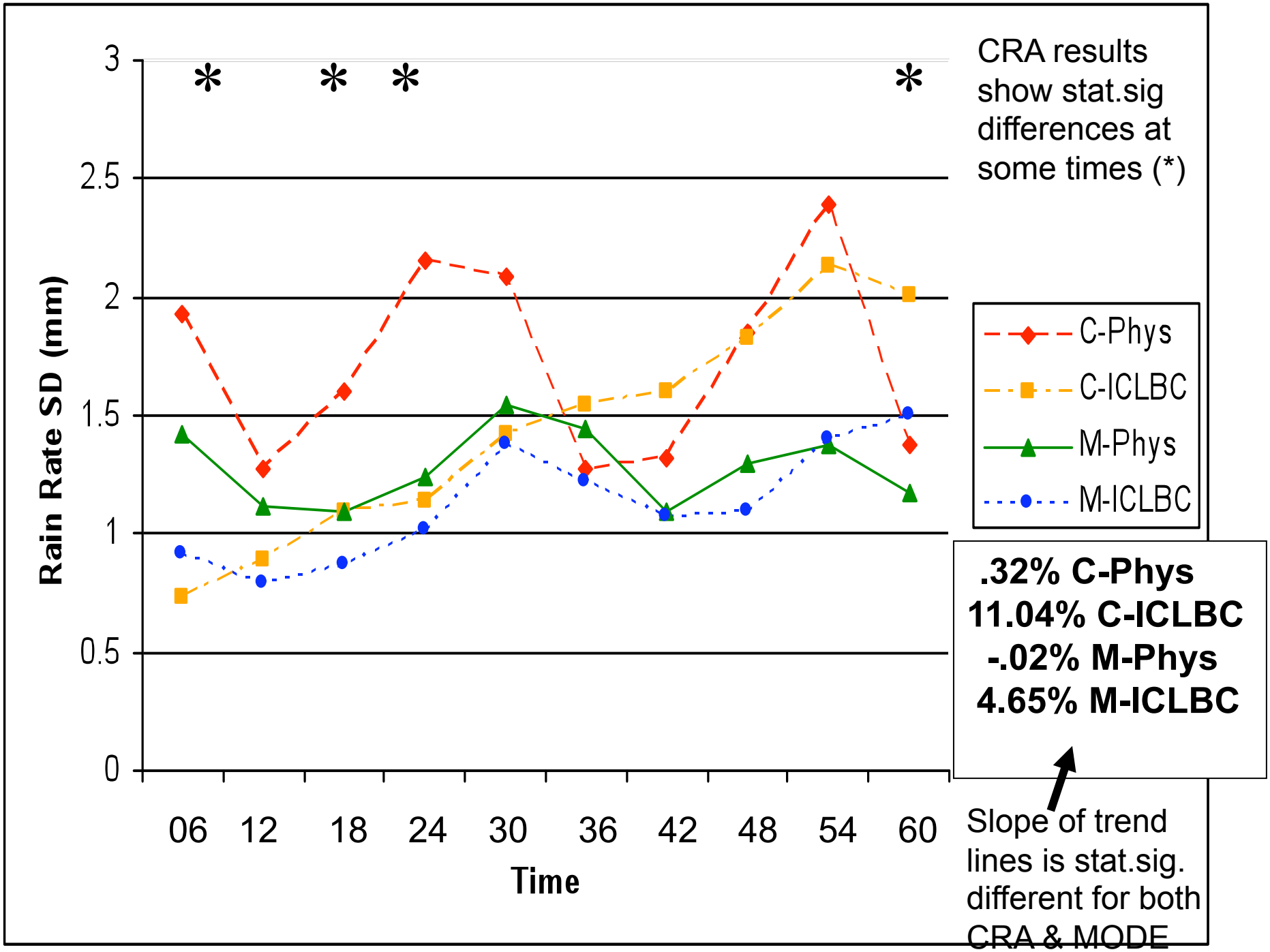
But results skipped in interest of time…

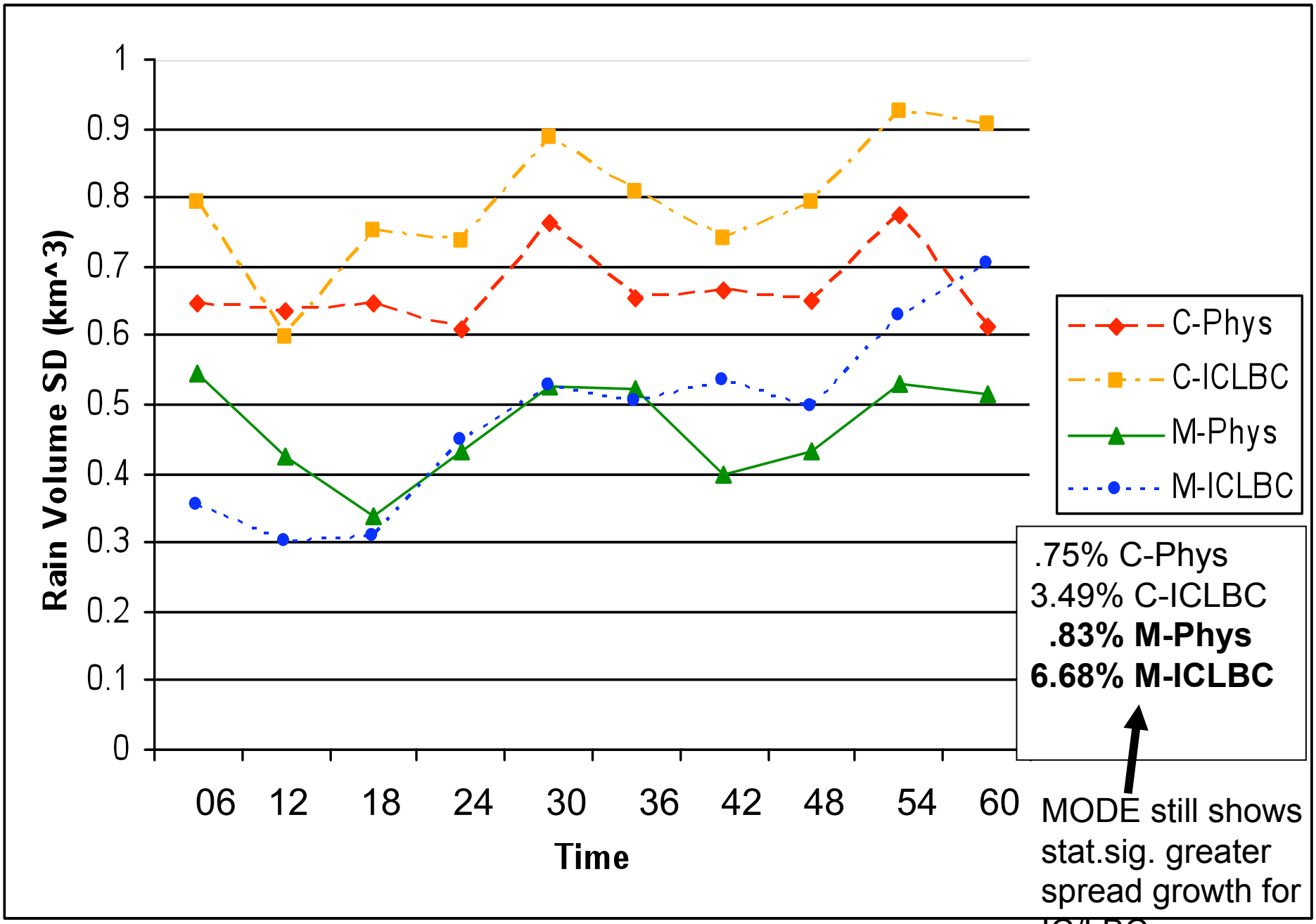
Other forecasting questions:

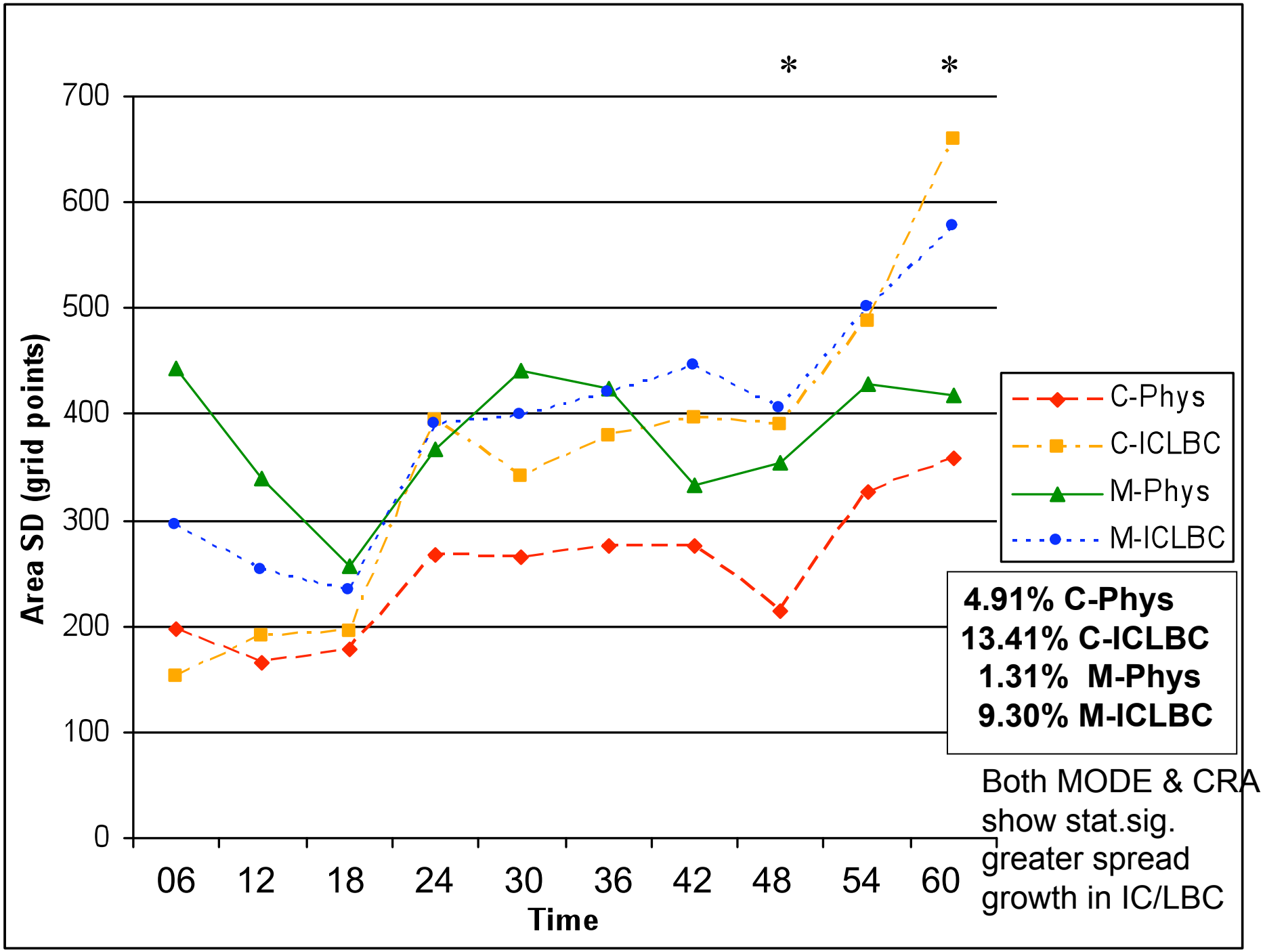
- Is the mean of the ensemble's distribution of object-based parameters (run MODE/CRA on each member) a better forecast than one from an ensemble mean (run MODE/CRA once)?
- Does an increase in spread imply less predictability?

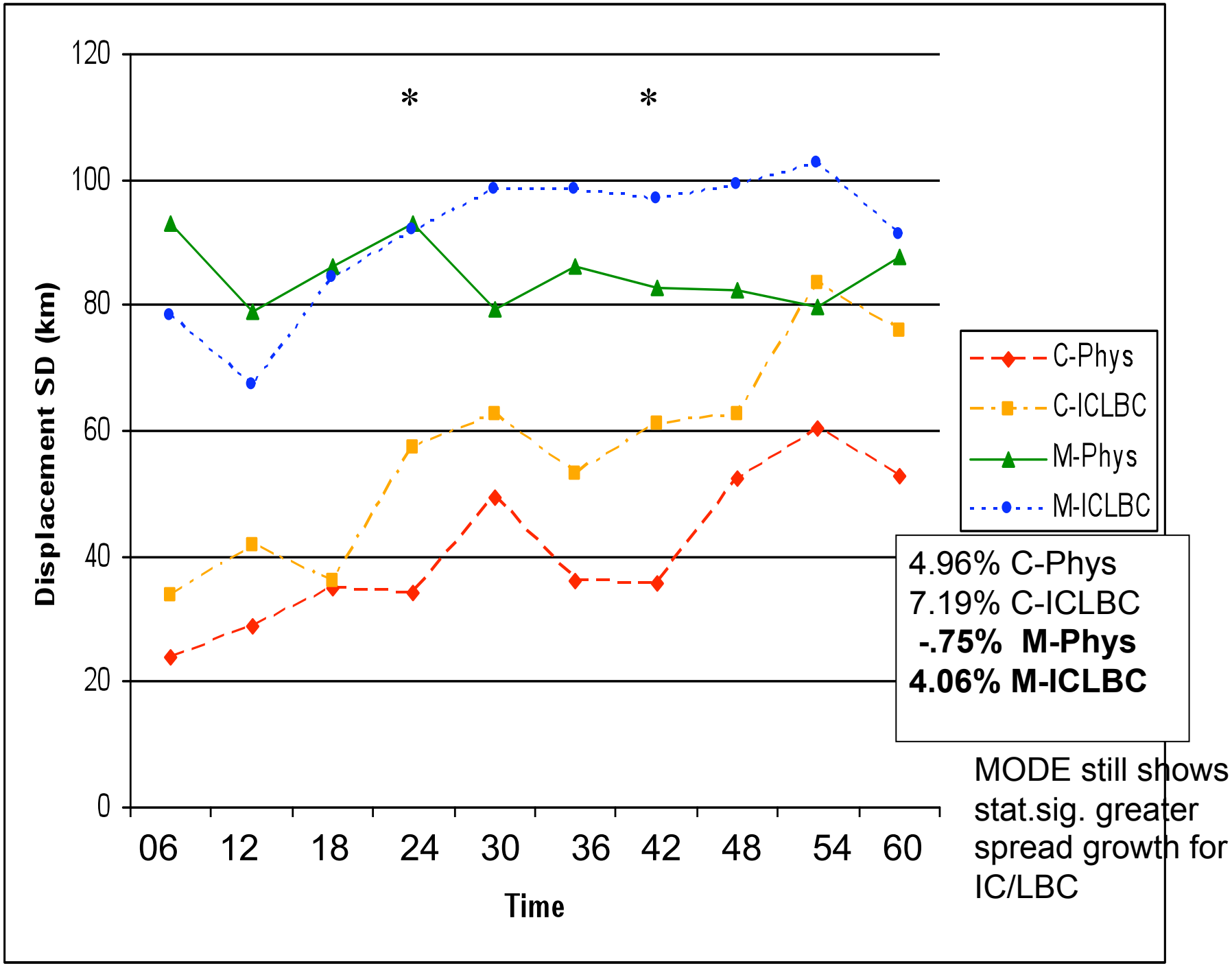
Object-based technique parameters examined:

- Areal coverage
- Mean Rain Rate
- Rain Volume of system
- Displacement error


CRA results show stat.sig differences at some times (*)
Rain Rate SD (mm)
3
2.5
2
1.5
1
0.5
0
*
*
*
*
06
12
18
24
30
36
42
48
54
60
Time
C-Phys
C-ICLBC
M-Phys
M-ICLBC
.32% C-Phys
11.04% C-ICLBC
-.02% M-Phys
4.65% M-ICLBC
Slope of trend lines is stat.sig. different for both CRA & MODE

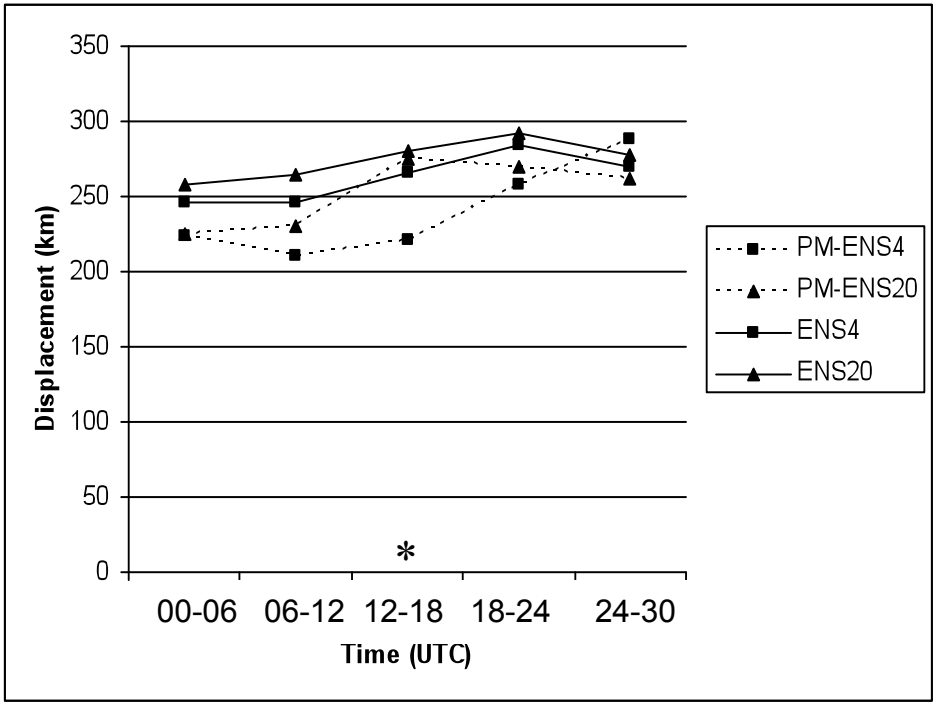

Rain Volume SD (km^3)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
06
12
18
24
30
36
42
48
54
60
Time
C-Phys
C-ICLBC
M-Phys
M-ICLBC
.75% C-Phys
3.49% C-ICLBC
.83% M-Phys
6.68% M-ICLBC
MODE still shows stat.sig. greater spread growth for IC/LBC

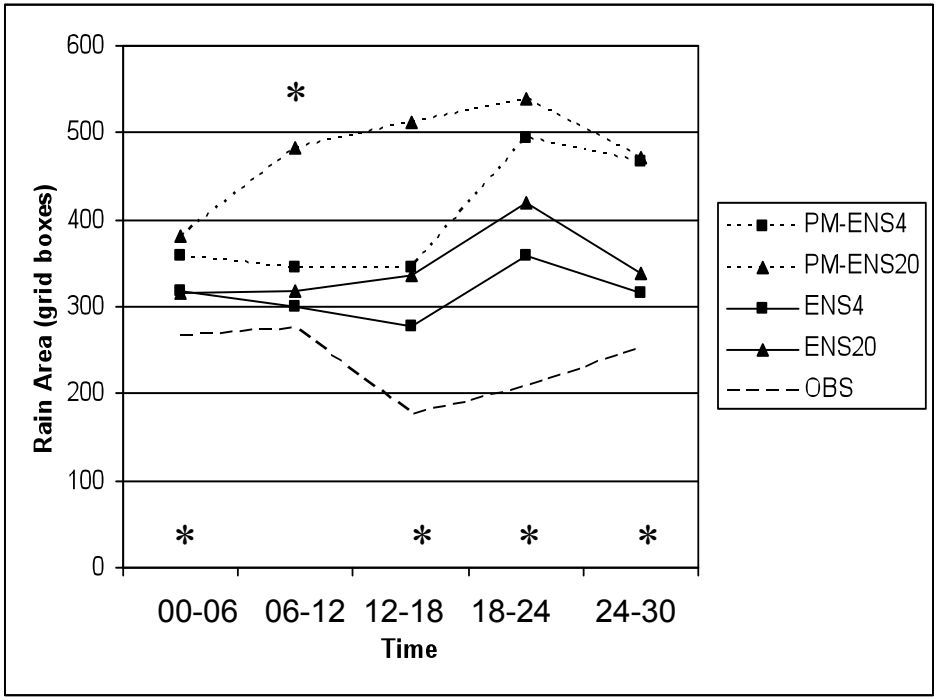

*
*
700
600
500
400
300
200
100
0
Area SD (grid points)
06
12
18
24
30
36
42
48
54
60
Time
C-Phys
C-ICLBC
M-Phys
M-ICLBC
4.91% C-Phys
13.41% C-ICLBC
1.31% M-Phys
9.30% M-ICLBC
Both MODE & CRA show stat.sig. greater spread growth in IC/LBC


120
100
80
60
40
20
0
Displacement SD (km)
*
*
06
12
18
24
30
36
42
48
54
60
Time
C-Phys
C-ICLBC
M-Phys
M-ICLBC
4.96% C-Phys
7.19% C-ICLBC
-.75% M-Phys
4.06% M-ICLBC
MODE still shows stat.sig. greater spread growth for IC/LBC

Conclusions (8 members)

- Increased spread growth in IC/LBC compared to Phys does show up in the 4 object parameters, especially for Areal Coverage, and moreso in MODE results than CRA
- Diurnal signal (more precip at night) does show up some in Rate, Volume, and Areal coverage parameters



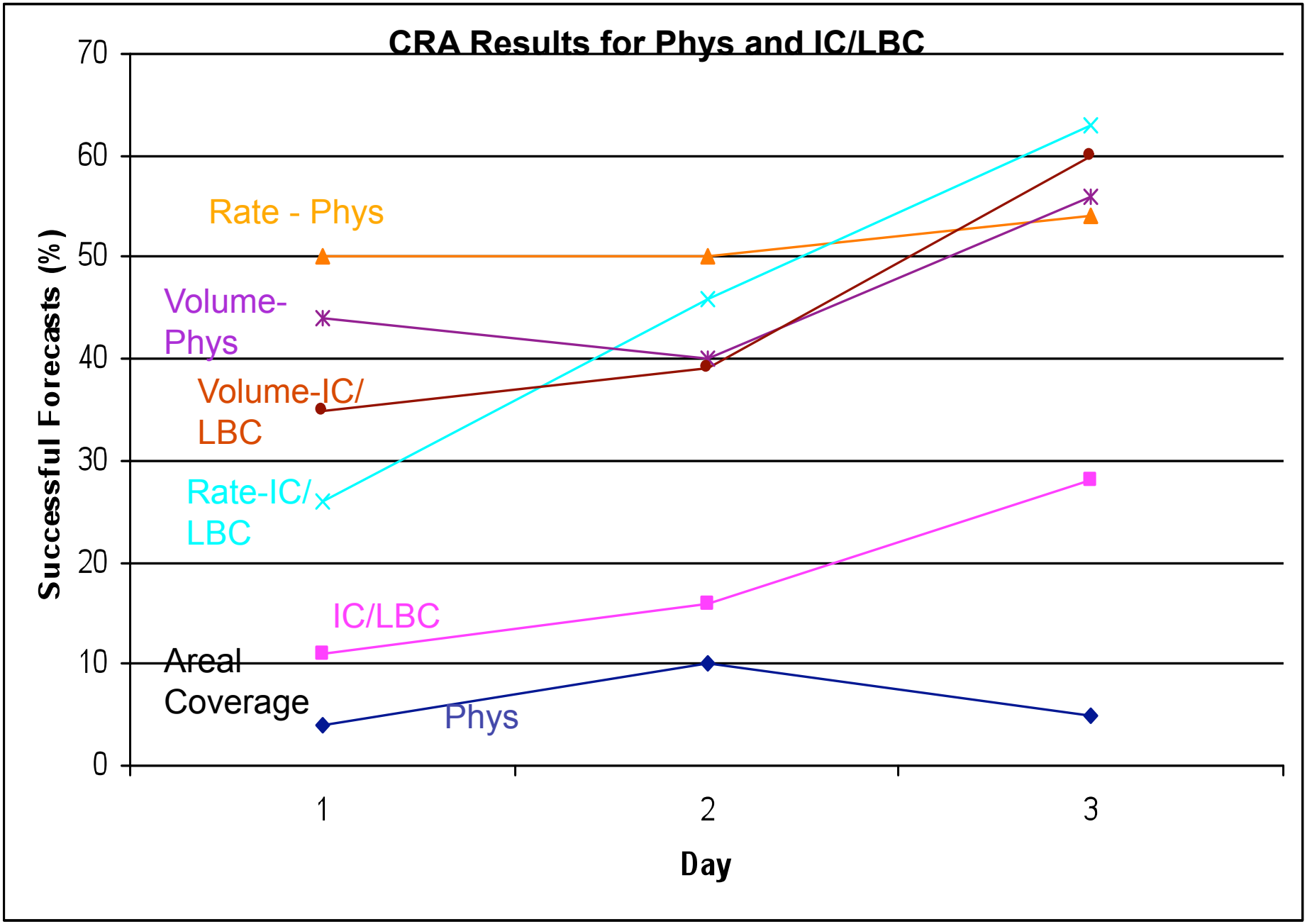


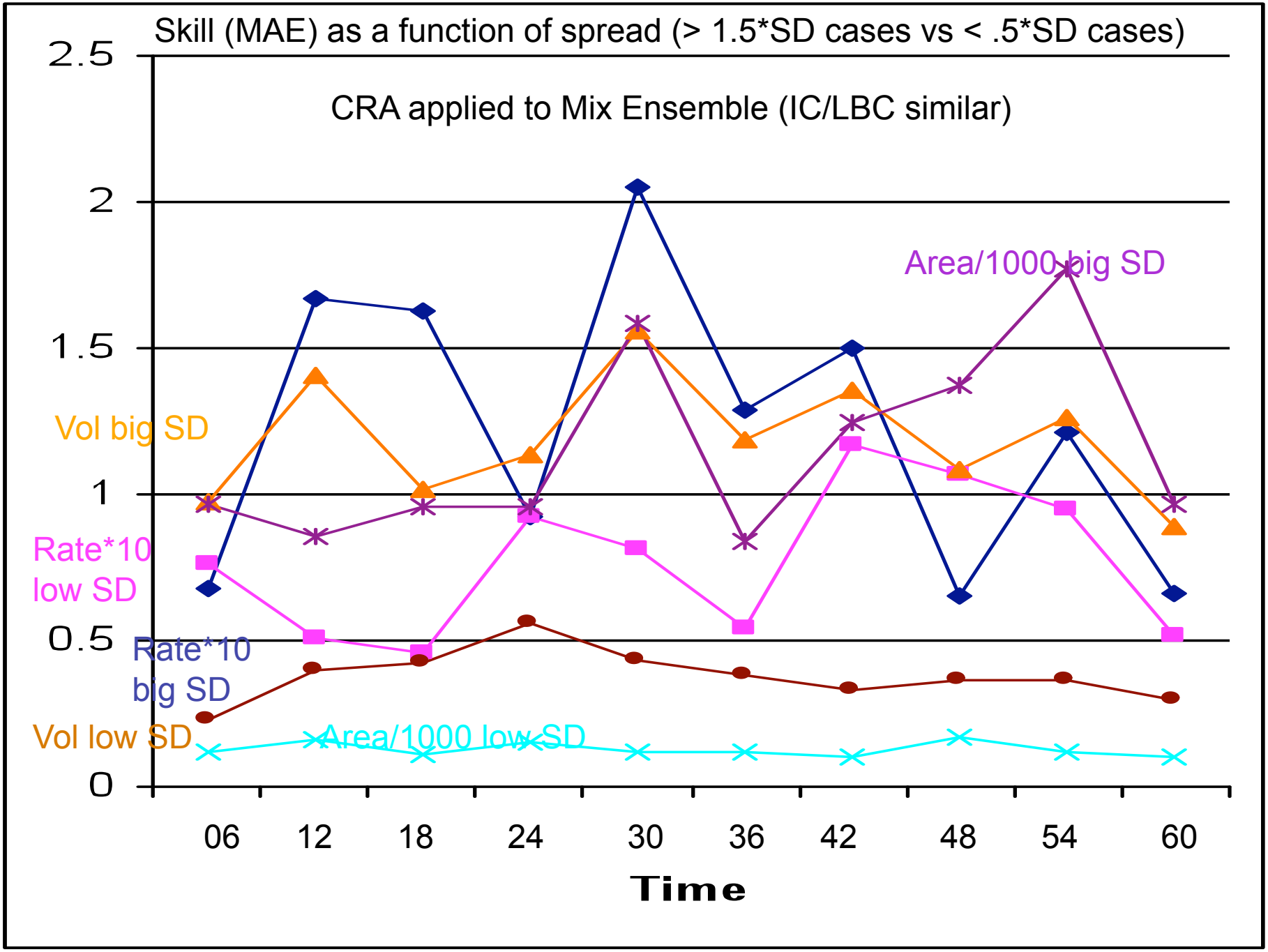
Asterisks at top (bottom) indicate stat.sig. difference for ENS20 (ENS4)

Comparison of Probability Matched Ensemble Mean (Ebert 2001) values to an average of the MODE output from each ensemble member

Note: PM may result in better forecast for location (smaller displacements) but a much worse forecast for area (also not as good for volume and rate – not shown)

Percentage of times the observed value fell within the min/max of the ensemble




Skill (MAE) as a function of spread (> 1.5*SD cases vs < .5*SD cases)
CRA applied to Mix Ensemble (IC/LBC similar)
2.5
2
1.5
1
0.5
0
Area/1000 big SD
Vol big SD
Rate*10
low SD
Rate*10
big SD
Vol low SD
Area/1000 low SD
06
12
18
24
30
36
42
48
54
60
Time

Conclusions (forecasting approaches)

- **For some parameters, application of MODE or CRA to PM-mean might be fine; for others it is better to use MODE/CRA on each member**
- **System rain volume and areal coverage show a clear signal for better skill when spread is smaller, not so true for rate**
- ***Average rain rate for systems is not as big a problem in the forecasts as areal coverage (and thus volume), which might explain lack of spread-skill relationship***
- ***AREAL COVERAGE IS ESPECIALLY POORLY FORECASTED***

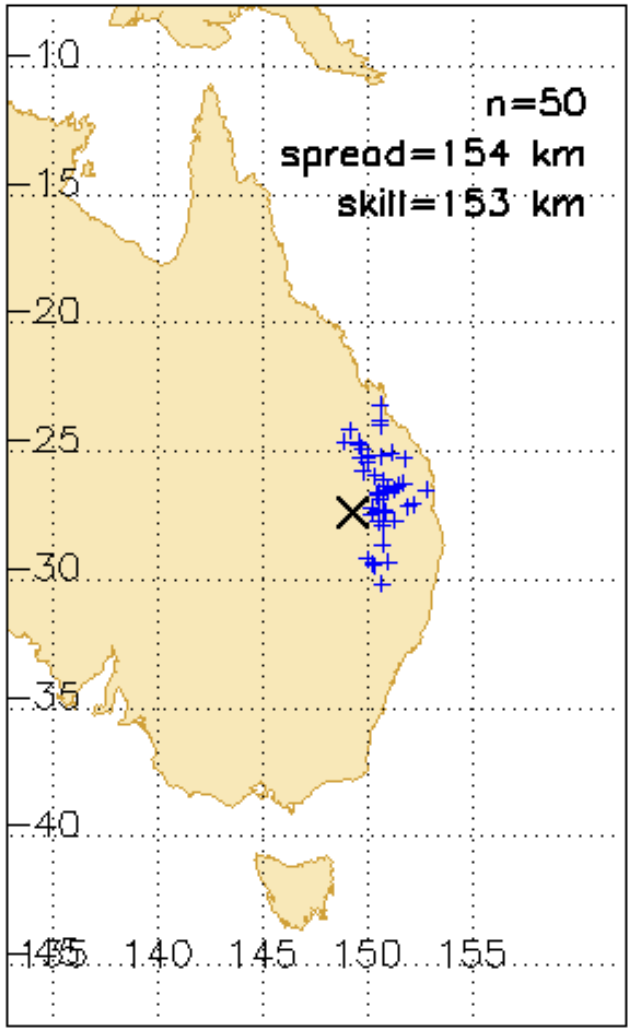
Difference in behavior of MSE and SD/Variance

- Clark et al. (2008) found both MSE and Variances had maxima at time of diurnal max in precipitation (with MSE having higher amplitude signal), but ETS was best at these times – reflecting bias or small displacements
- Object parameters did not have MSE or SDs (or Variances) cleanly reflecting the diurnal precipitation signal.

Some other questions to address using object approach

- **What does the number of members predicting an event say about its likelihood?**
- **Was the location of the center of the observed system found within an envelope (convex hull) of ensemble members' centers?**

Distribution of object locations



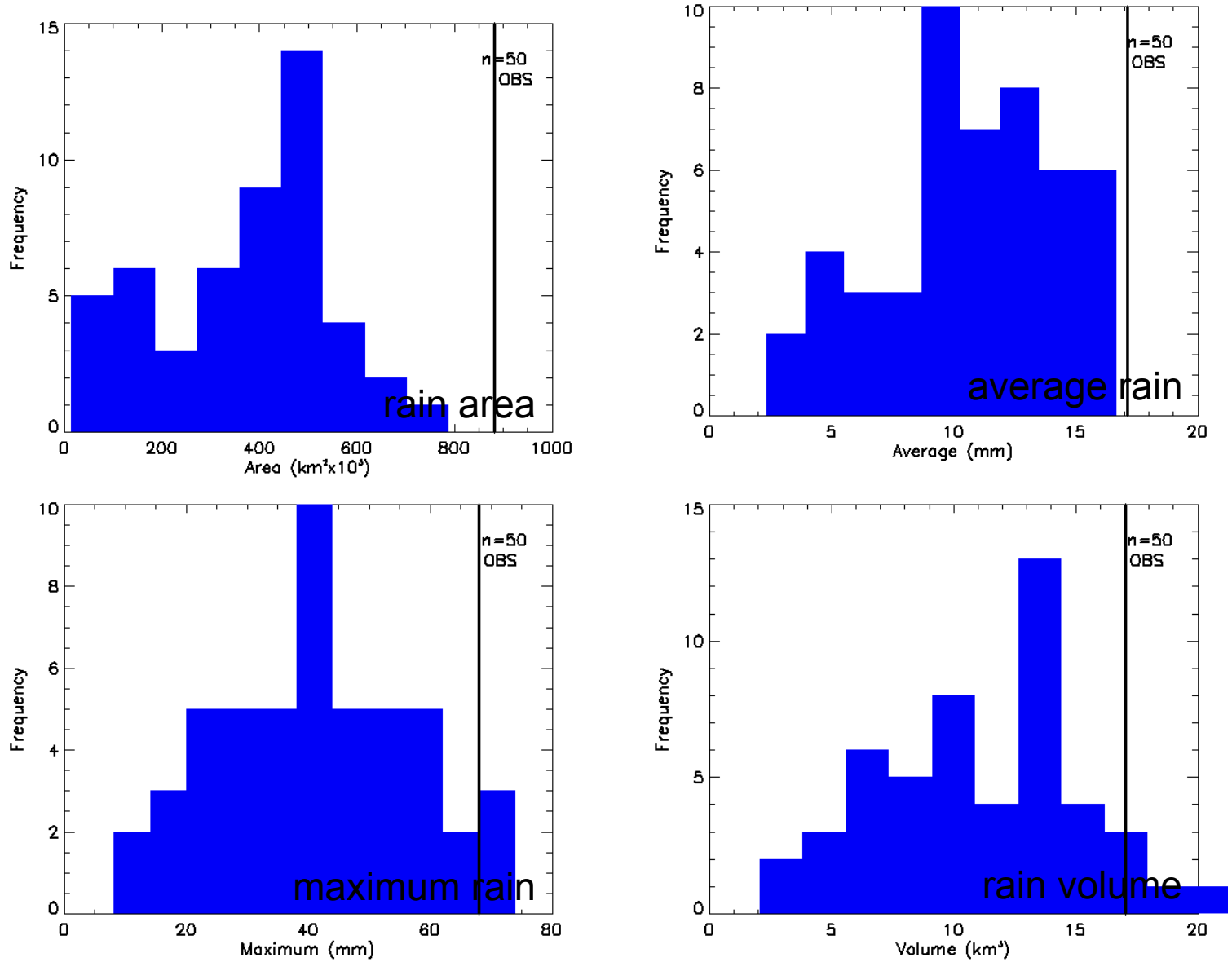
ECMWF EPS 4.5 day forecast
valid 00 UTC 5 Sept 08

location = mass weighted center of event

spread = average distance to ensemble median location

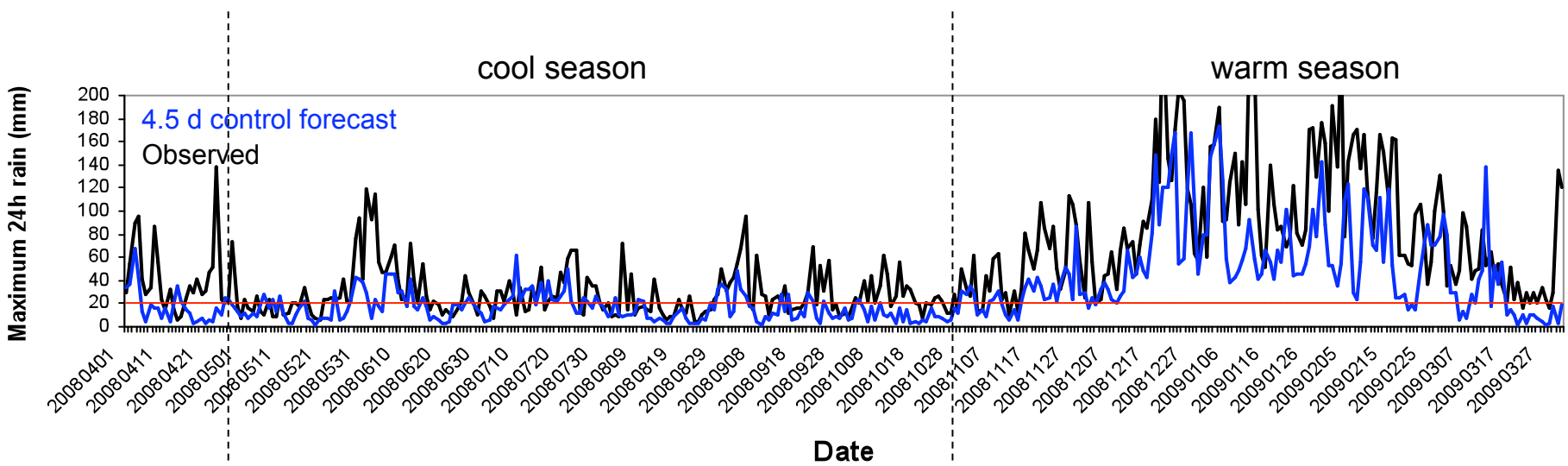
skill = distance between ensemble median and observed location

Distributions of object attributes

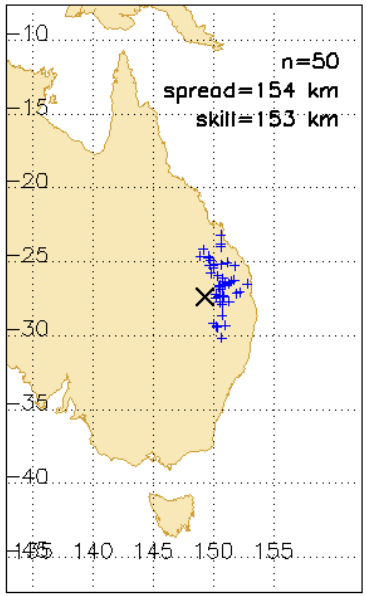


ECMWF EPS heavy rain verification
1 April 2008 – 31 March 2009

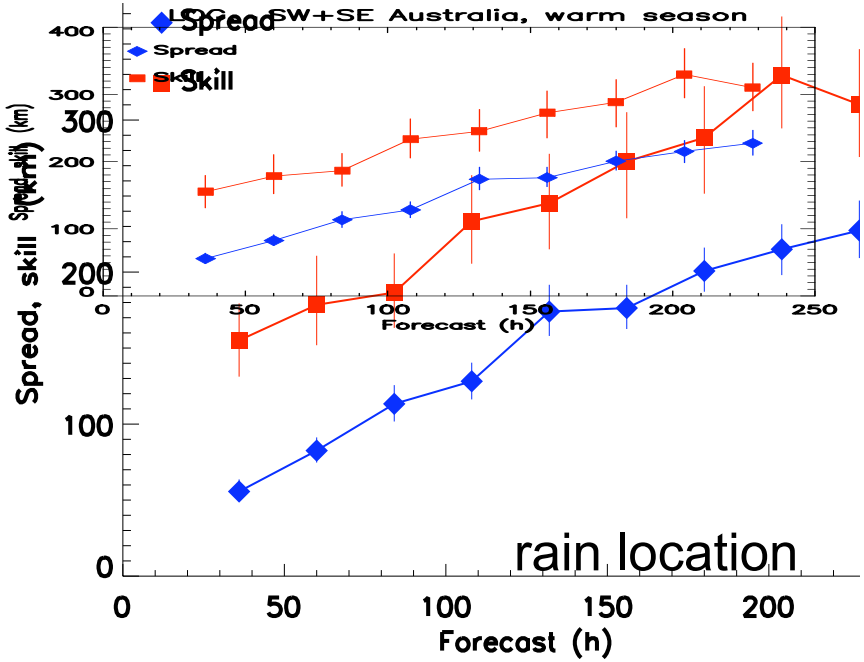
- ECMWF EPS @ T_L399L91 resolution (~0.5º)
- Forecasts of 24h rain accumulation to 228 h
- Verified against BOM operational daily rain gauge analysis (0.25º)
- 0.5º verification grid
- Verify individual ensemble members' CRAs defined by 5 mm d^{-1} threshold
- Focus on events with observed or forecast maxima of ≥ 20 mm d^{-1}
- Stratify results by season and region



Spread and skill for location forecasts



CRAs with max rain ≥ 20 mm d^{-1}
April 2008 – March 2009
Warm season, southern Australia
368 observed events
(plus 37 false alarms)



Spread = average distance to ensemble median location

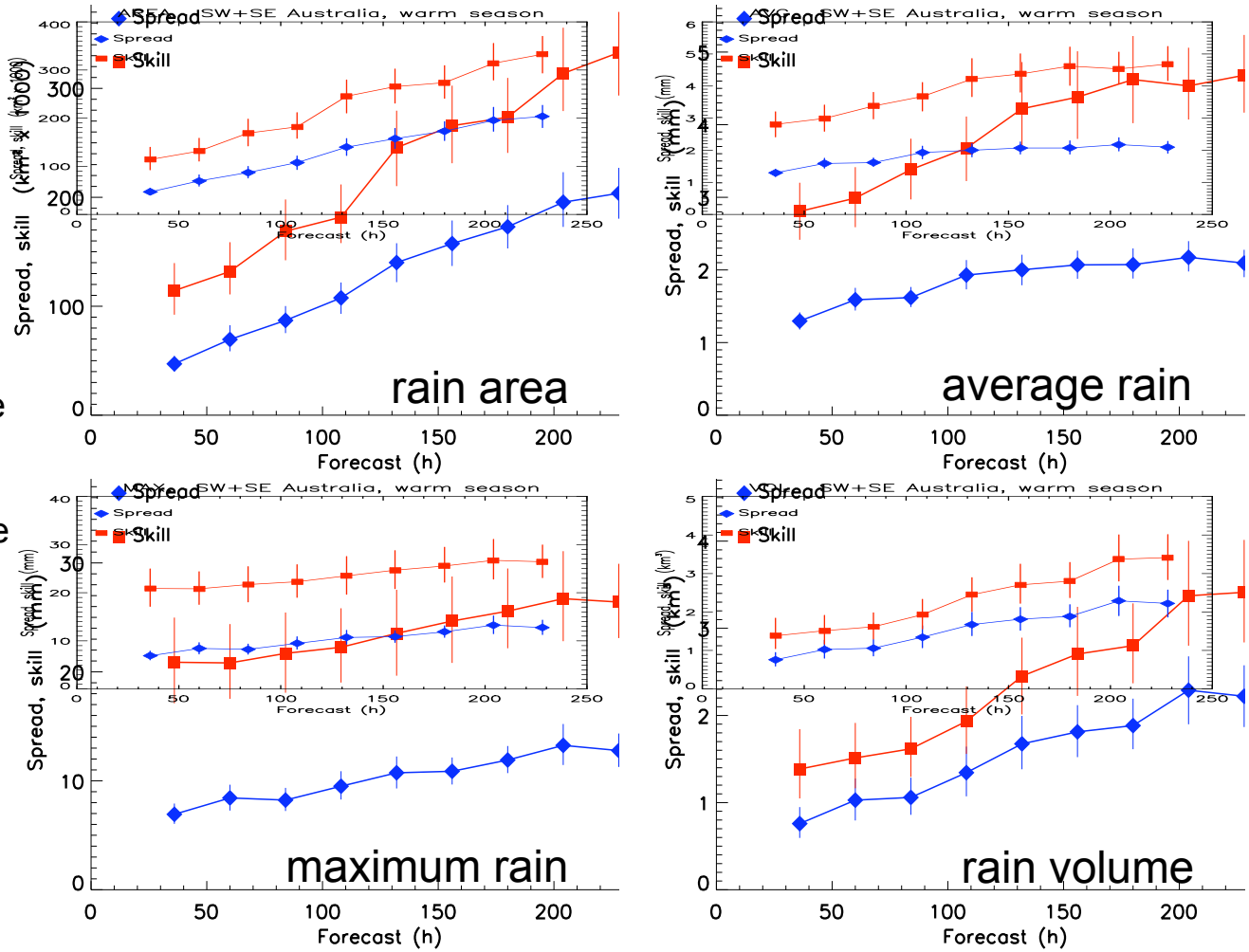
Skill = distance between ensemble median and observed location

Spread and skill

CRAs with max rain ≥ 20 mm d^{-1} April 2008 – March 2009
Warm season, southern Australia

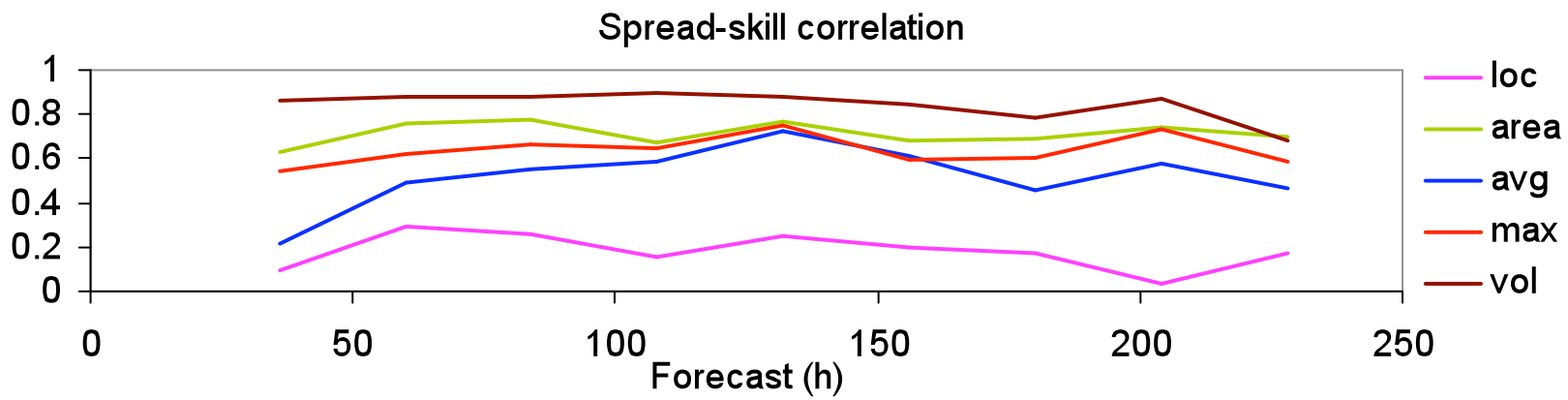
Spread = SD of ensemble attributes

Skill = RMSE of ensemble mean attribute



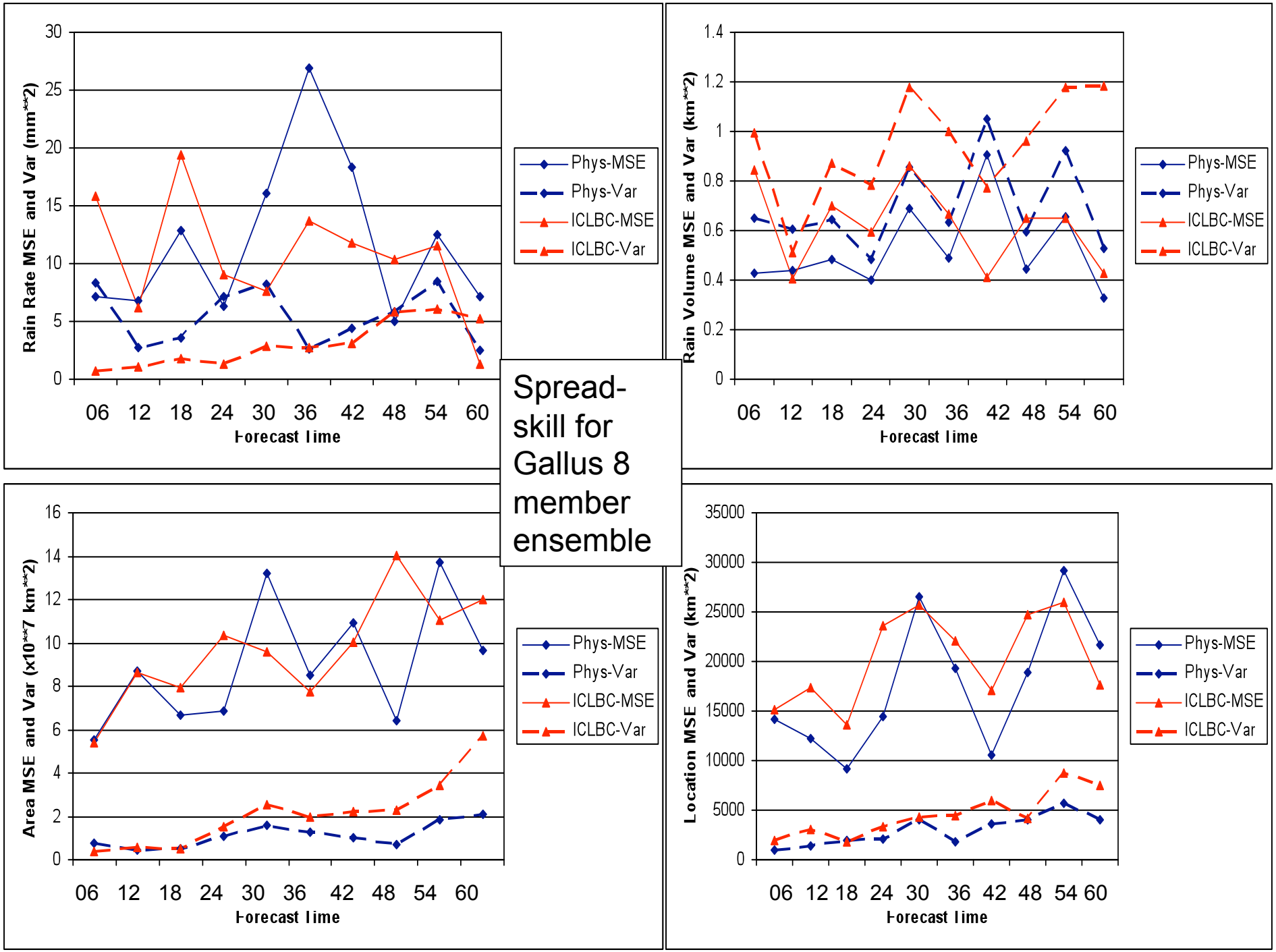
Object spread-skill correlations

CRAs with max rain ≥ 20 mm d^{-1} April 2008 – March 2009
Warm season, southern Australia



Attribute	object location	object area	object mean rainfall	object maximum rainfall	object rain volume
Mean spread-skill correlation	0.18	0.71	0.52	0.64	0.84

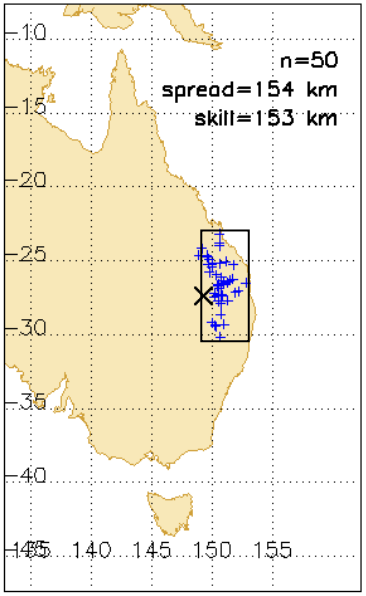
→ Uncertainty in heavy rain system properties can be predicted reasonably well except for location

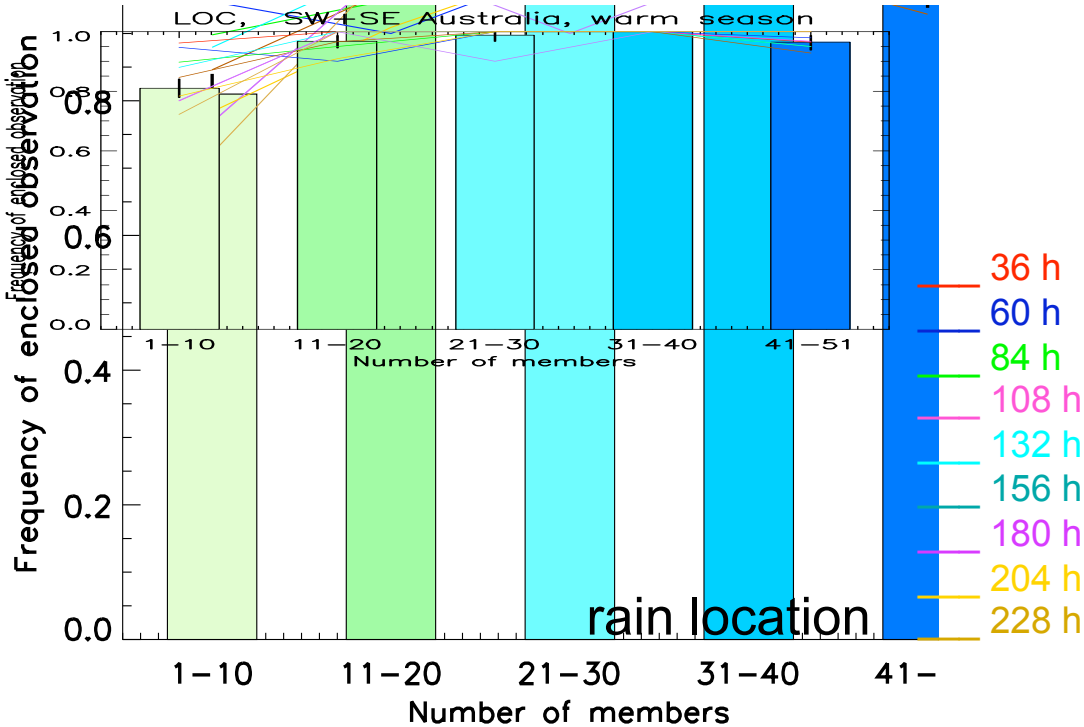


Spread-skill for Gallus 8 member ensemble

How often is the observed value within the ensemble distribution?

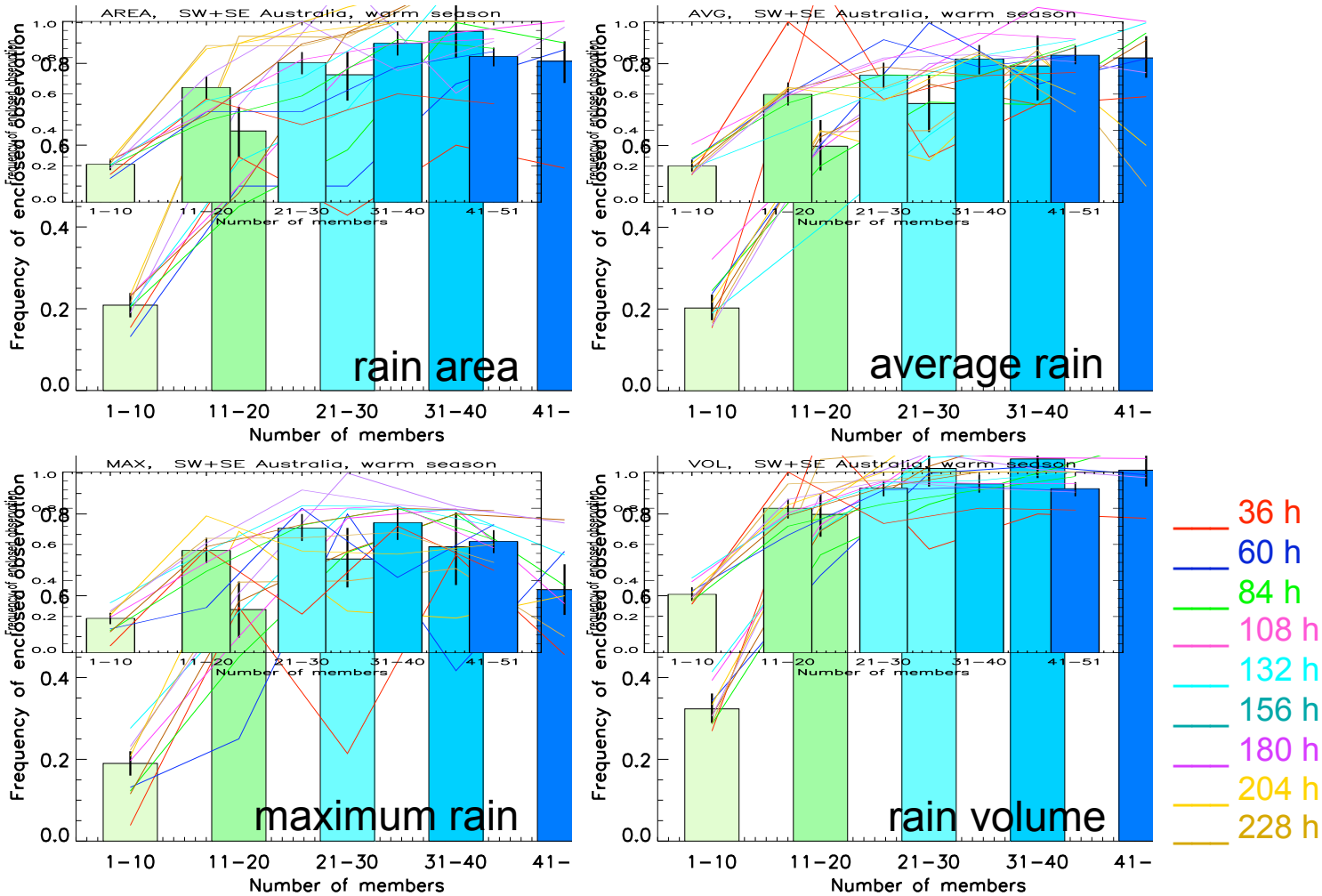
CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia





How often is the observed value within the ensemble distribution?

CRAs with max rain ≥ 20 mm d^{-1} Warm season, southern Australia



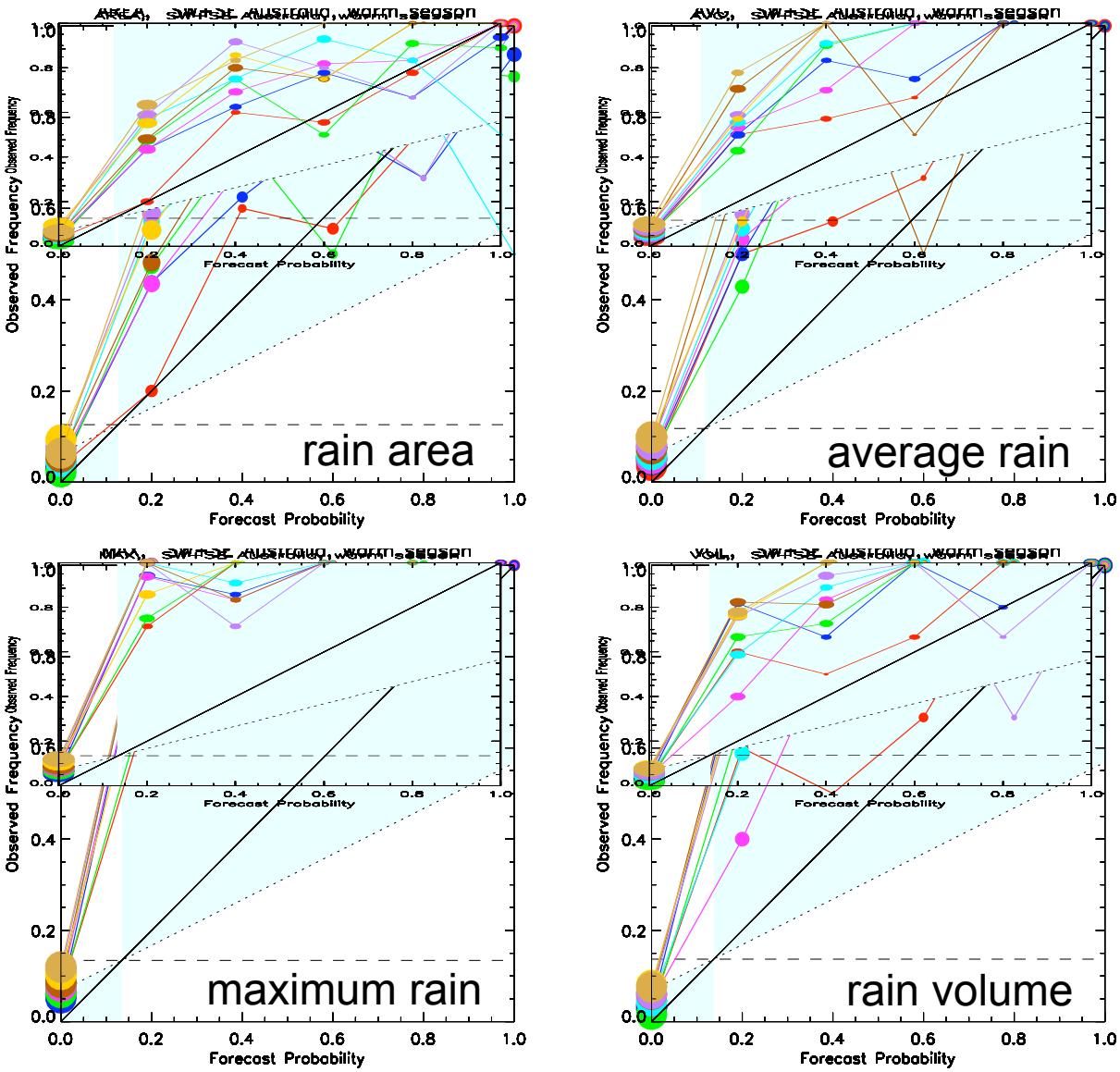
Probabilistic verification

- How accurate is the **distribution of features attributes** for the ensemble members?
- What thresholds to use? (these are for heavy rain)

Attribute	Median value
object area	$2x10^5$ km^2
object mean rainfall	10 mm d^{-1}
object maximum rainfall	40 mm d^{-1}
object rain volume	2.5 km^3

Does the number of ensemble members predicting an event have meaning?

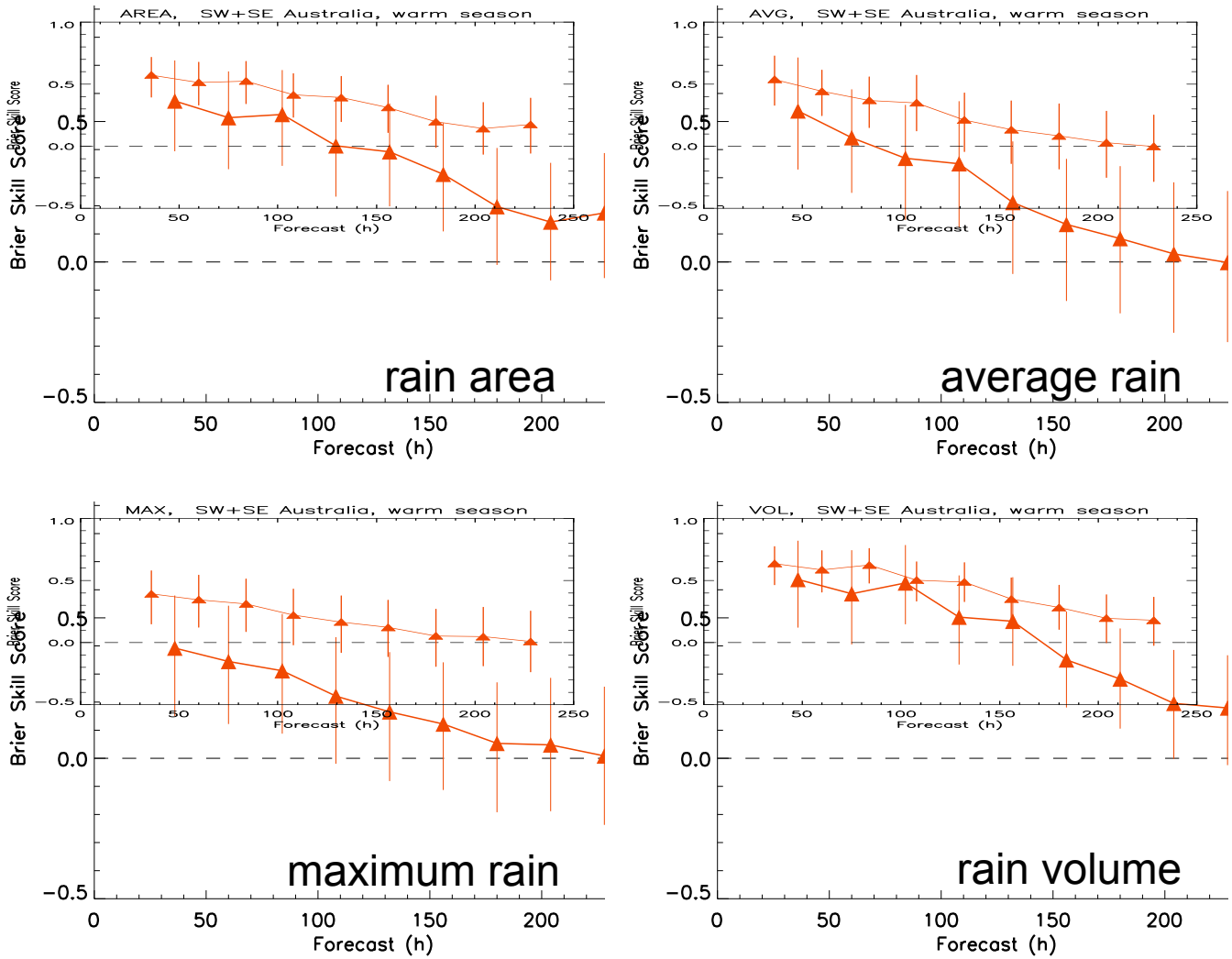
Forecast probability = fraction of ensemble members ≥ threshold



36 h
60 h
84 h
108 h
132 h
156 h
180 h
204 h
228 h

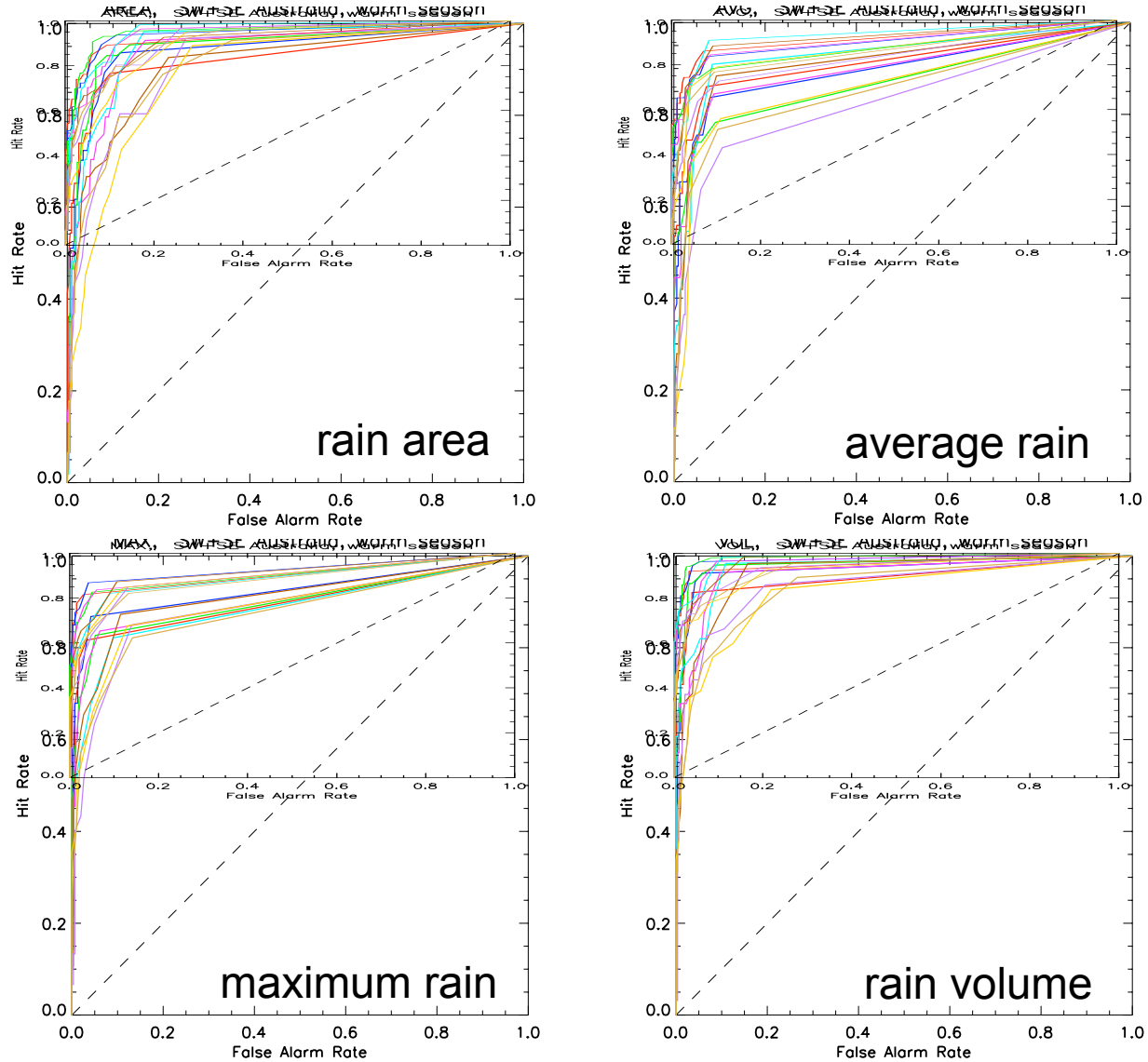
Does the number of ensemble members predicting an event have meaning?

Brier skill score with respect to sample climatology of observed events



Does the number of ensemble members predicting an event have meaning?

ROC varies the number of ensemble members exceeding the threshold required for a "yes event" forecast



36 h
60 h
84 h
108 h
132 h
156 h
180 h
204 h
228 h

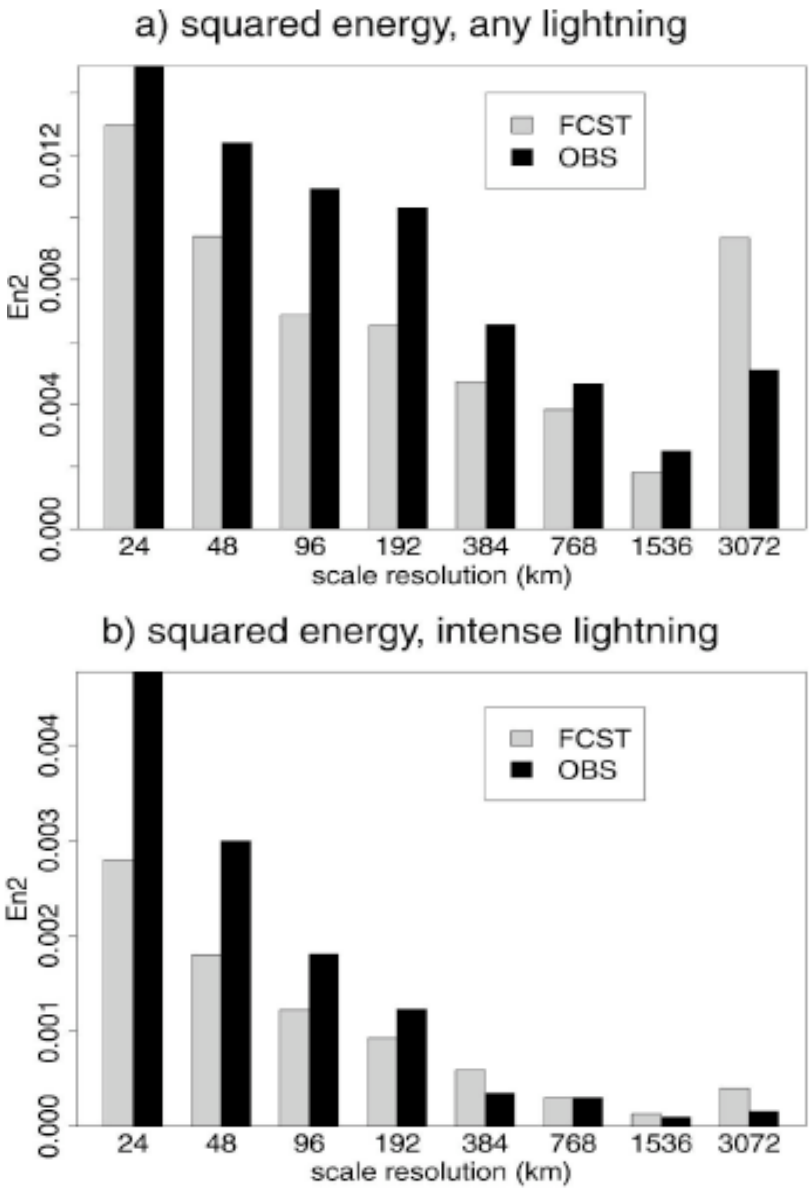
Early results

- **Object-based verification looks promising for evaluation of heavy rain events and other weather features**
- The ensemble spread is less than the RMSE of the ensemble mean for all object attributes (location, rain area, mean & max rain, rain volume)
- Spread-skill correlation is high (except for location)
  → **spread can be used to predict uncertainty**
- The more members predicting an event, the greater the chance of its observed properties being enveloped by the ensemble
- Probability forecasts for events were biased low
- Brier skill score consistently positive
- ROC shows high discrimination ability → calibration should help!

Acknowledgments

- Thanks are given to the many researchers who contributed information about their spatial techniques, including Steven Lack, Neil Fox, Caren Marzban, Barb Casati, Jason Nachamkin, Marion Mittermaier, and Heini Wernli
- Thanks to the DTC for funding the workshops and my own DTC visit
- Partial support for my own work comes from NSF grants ATM-0537043 and ATM-0848200

Energy on different scales

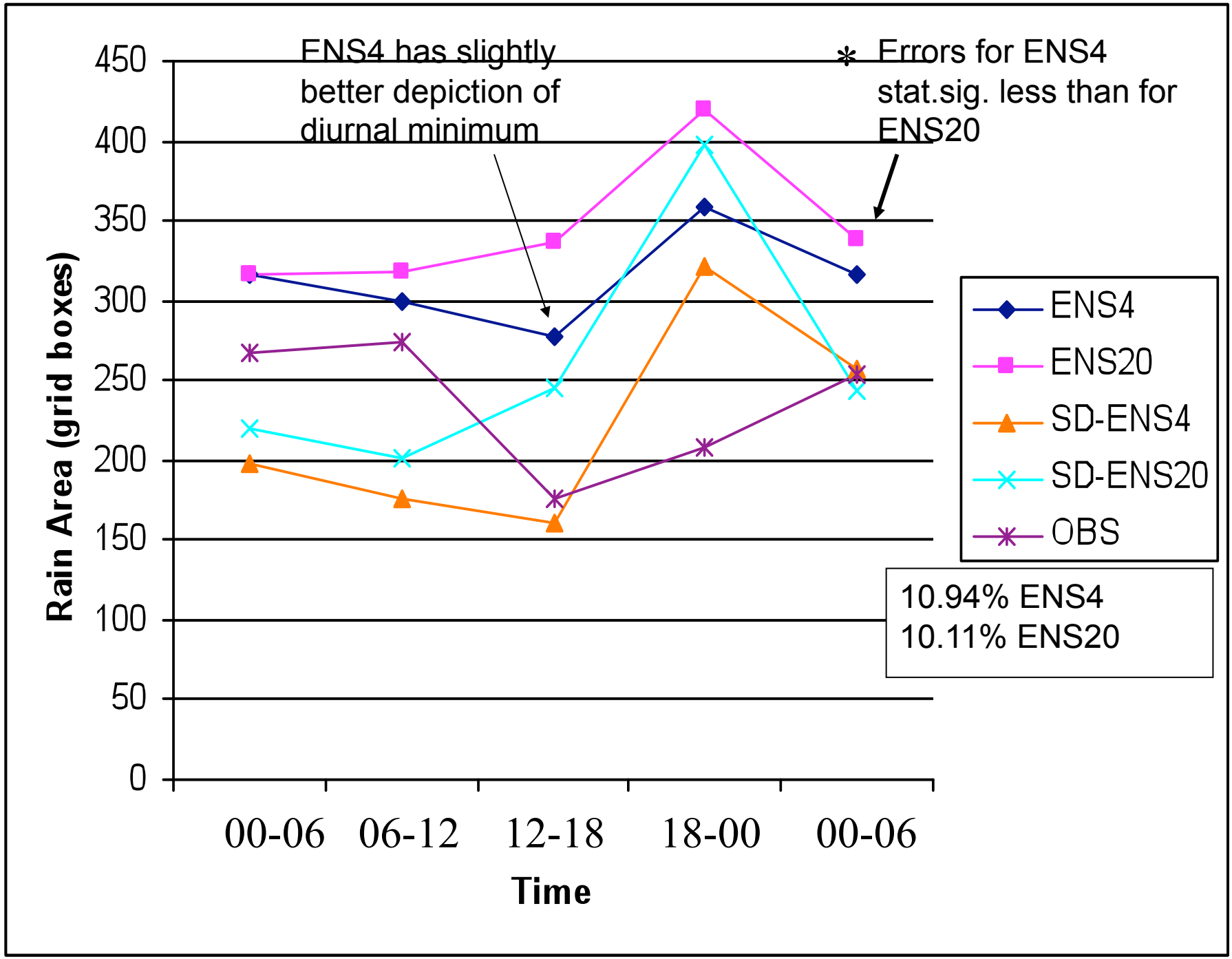


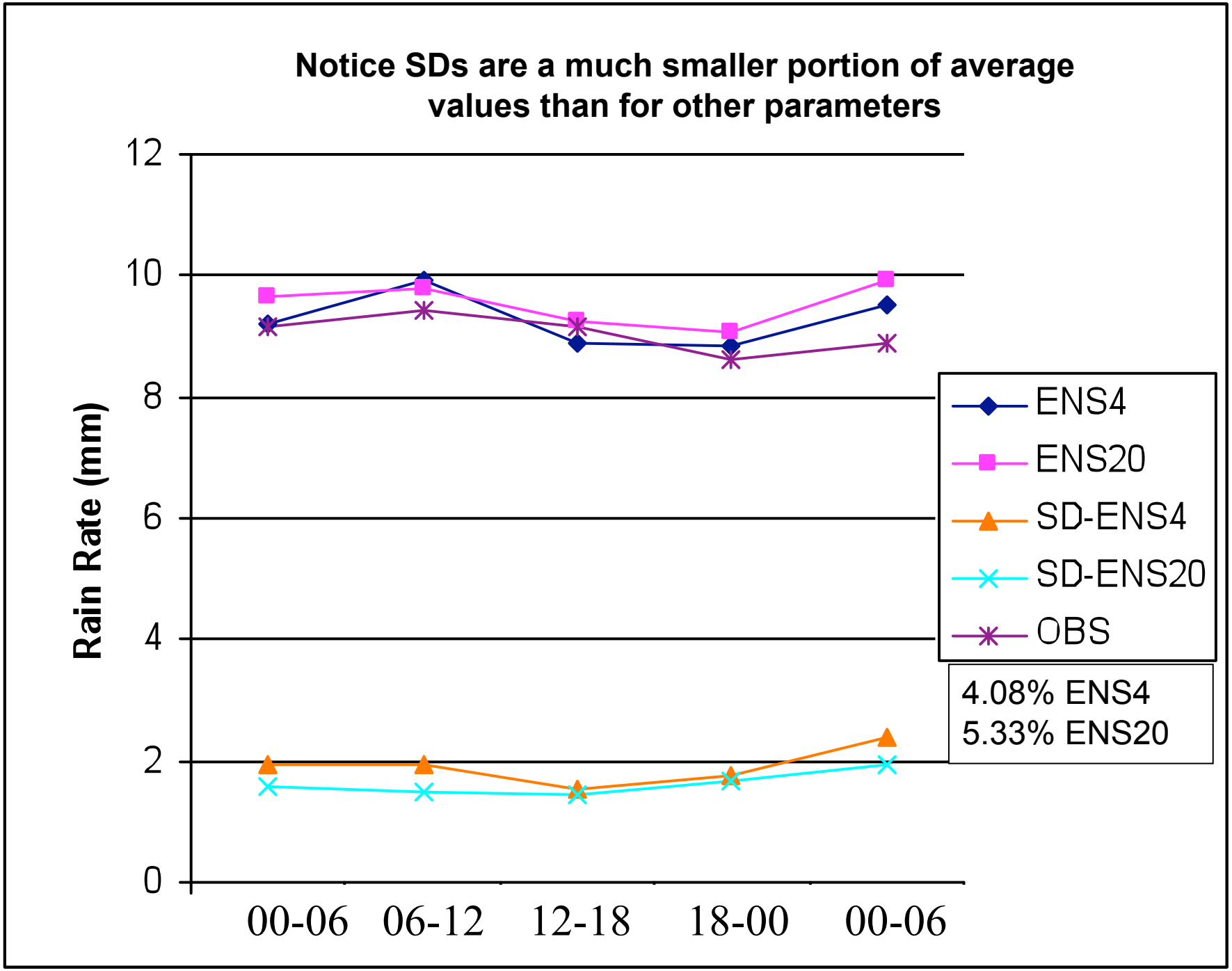
- Squared energy (En2) = average of squared values
- En2 provides feedback on the number of events on each scales (& magnitude): more events for low thresholds and less for intense lightning
- **Small scales have the largest number of events, the events decrease as the scale increases**
- Comparison of forecast and obs En2 provide a measure of the bias on different scales: **under-forecast on small scales, over-forecast on largest scale** and large scales for intense events

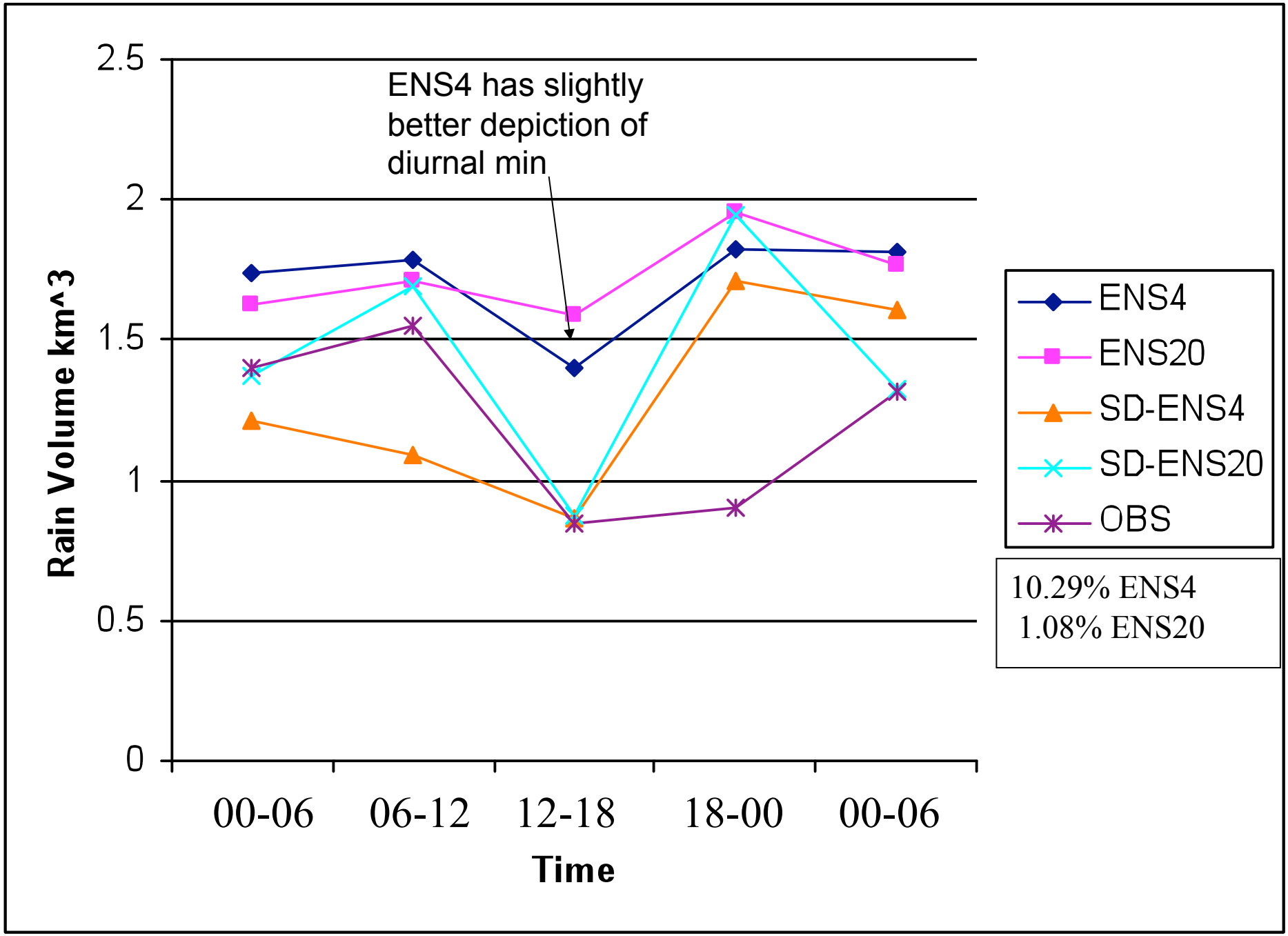
ENS4 vs ENS20

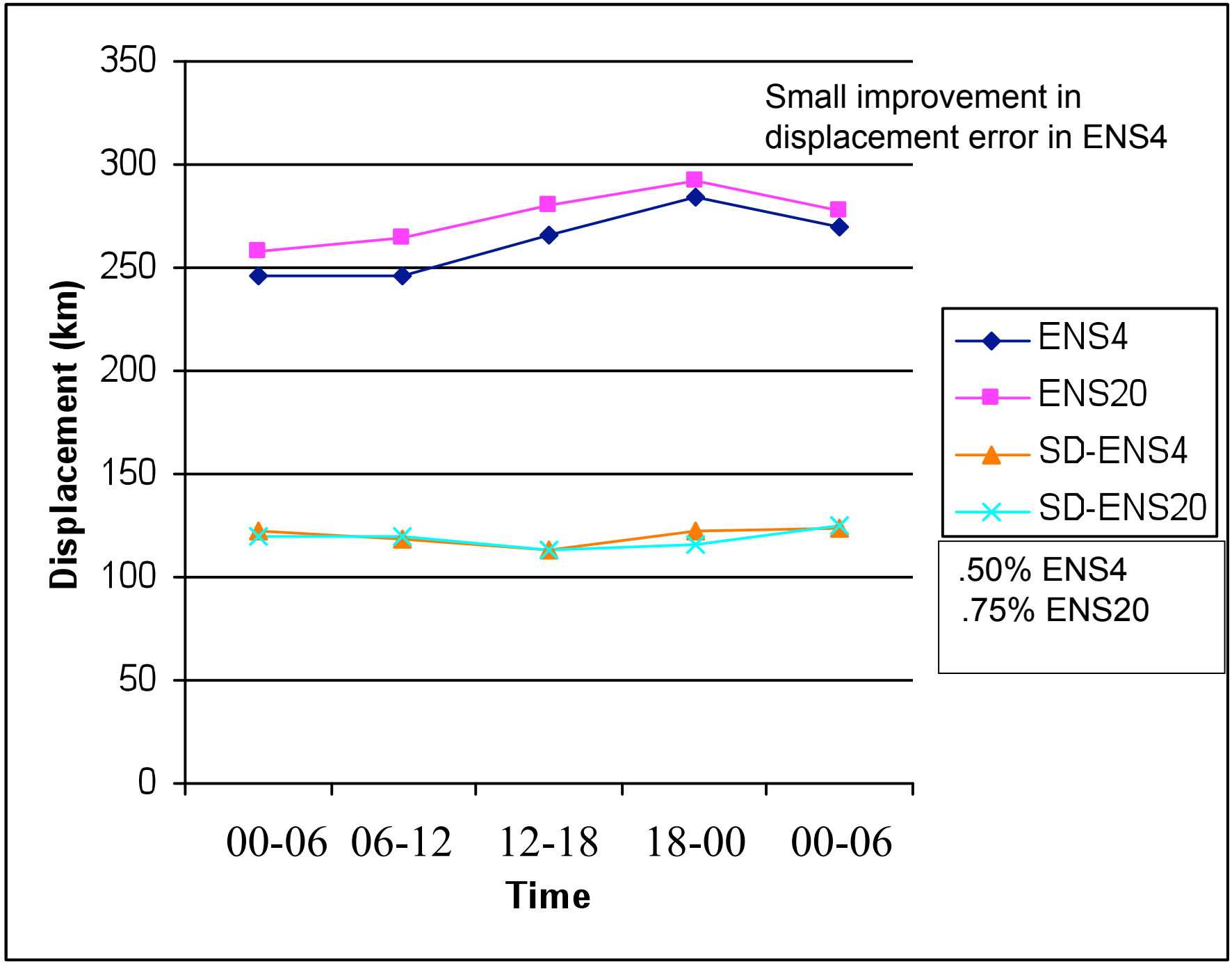
- 6 hour precipitation evaluated over 30 h integrations from 23 cases, with all precipitation remapped to a 20 km grid
- Clark et al. (2009) found <u>spread growth was faster in ENS4 than in ENS20</u>
- <u>ENS4 had a much better depiction of the diurnal cycle</u> in precipitation

Question: Do these two results also show up in the object parameters


ENS4 has slightly better depiction of diurnal minimum
* Errors for ENS4 stat.sig. less than for ENS20
Rain Area (grid boxes)
450
400
350
300
250
200
150
100
50
0
00-06
06-12
12-18
18-00
00-06
Time
ENS4
ENS20
SD-ENS4
SD-ENS20
OBS
10.94% ENS4
10.11% ENS20

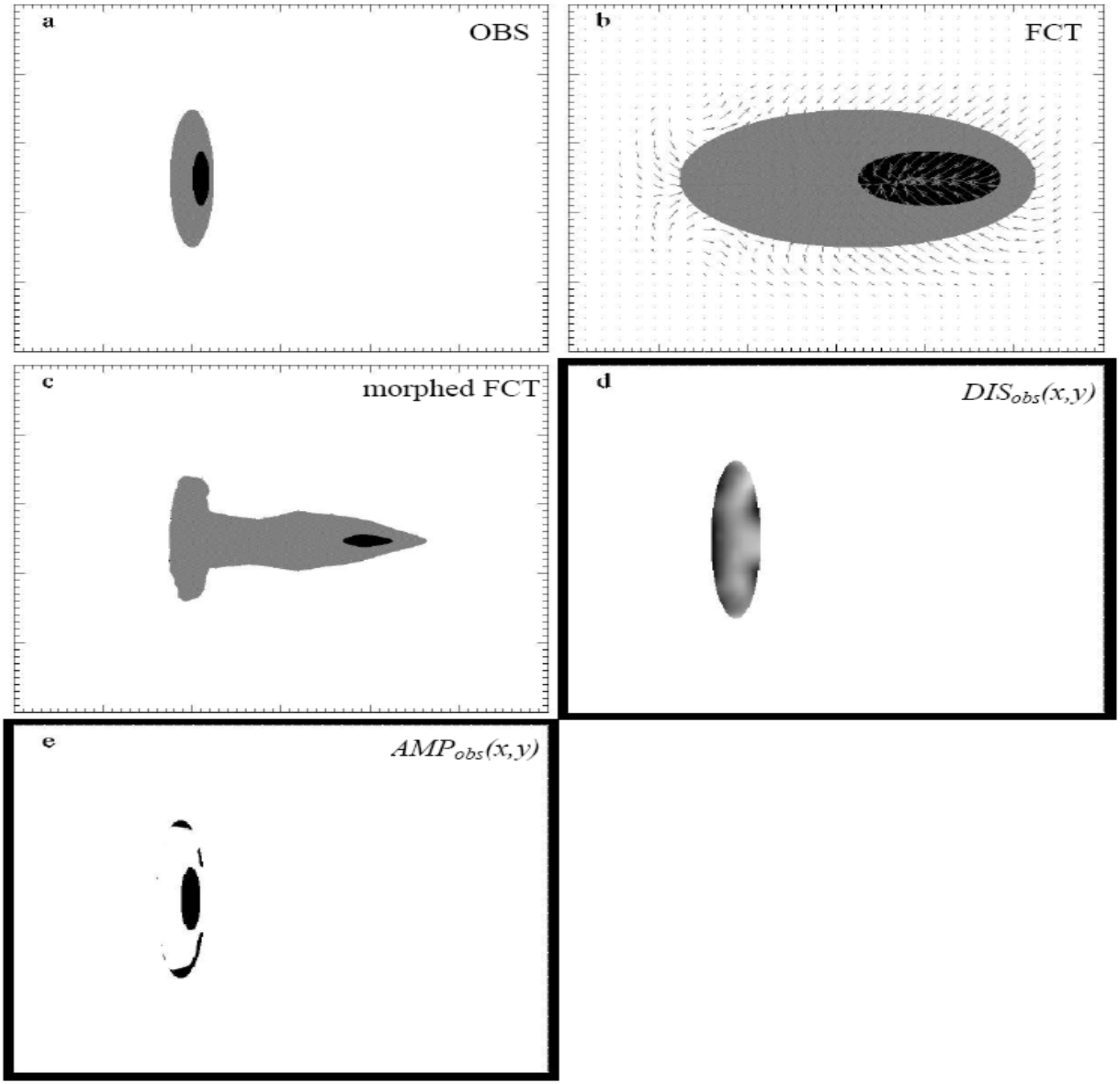

Notice SDs are a much smaller portion of average values than for other parameters
Rain Rate (mm)
12
10
8
6
4
2
0
00-06
06-12
12-18
18-00
00-06
ENS4
ENS20
SD-ENS4
SD-ENS20
OBS
4.08% ENS4
5.33% ENS20


ENS4 has slightly better depiction of diurnal min
Rain Volume km^3
2.5
2
1.5
1
0.5
0
00-06
06-12
12-18
18-00
00-06
Time
ENS4
ENS20
SD-ENS4
SD-ENS20
OBS
10.29% ENS4
1.08% ENS20


Small improvement in displacement error in ENS4
Displacement (km)
350
300
250
200
150
100
50
0
00-06
06-12
12-18
18-00
00-06
Time
ENS4
ENS20
SD-ENS4
SD-ENS20
.50% ENS4
.75% ENS20

Conclusions (ENS4 VS ENS20)

- Hint of better diurnal signal in ENS4 (Area and Volume)
- ENS4 seems more skillful (but not usually statistically significant)
- Volume best shows faster spread growth in ENS4 compared to ENS20, but results not as significant as in 8 member ensembles
- Rain rate has less variability among members, and is better forecasted


a
OBS
b
FCT
c
morphed FCT
d
$DIS_{obs}(x,y)$
e
$AMP_{obs}(x,y)$