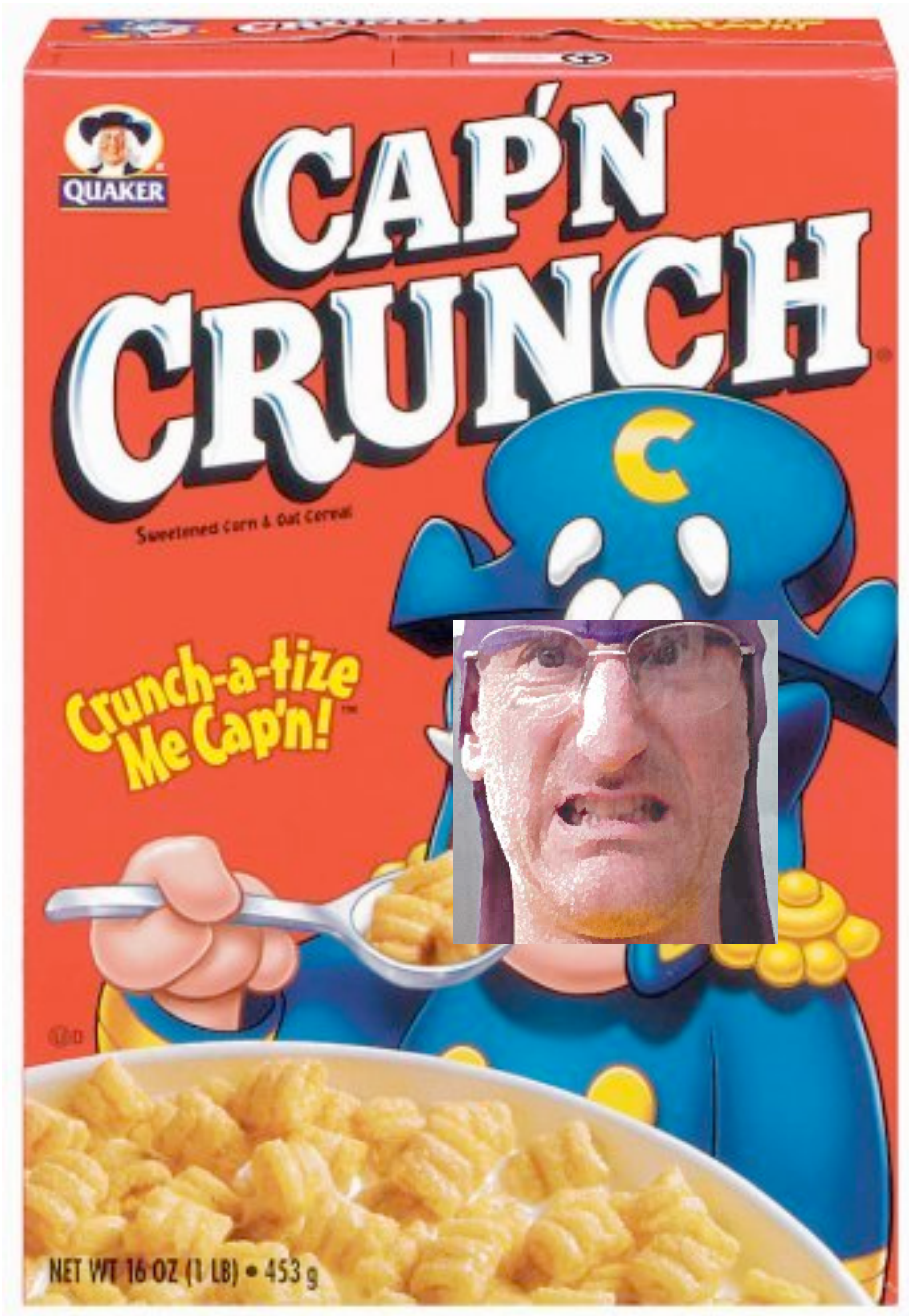
Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter

Tom Hamill and Jeff Whitaker

NOAA Earth System Research Lab

Physical Sciences Division



All due credit
to Jeff Whitaker,
chief number
cruncher,
EnKF innovator

(he's at ECMWF currently
implementing
the EnKF code there)

The ensemble Kalman Filter (EnKF)

- A method for the initialization of ensemble forecasts that is conceptually appealing for hurricanes
  - “Flow-dependent” background-error covariances may be useful to achieving quality analyses

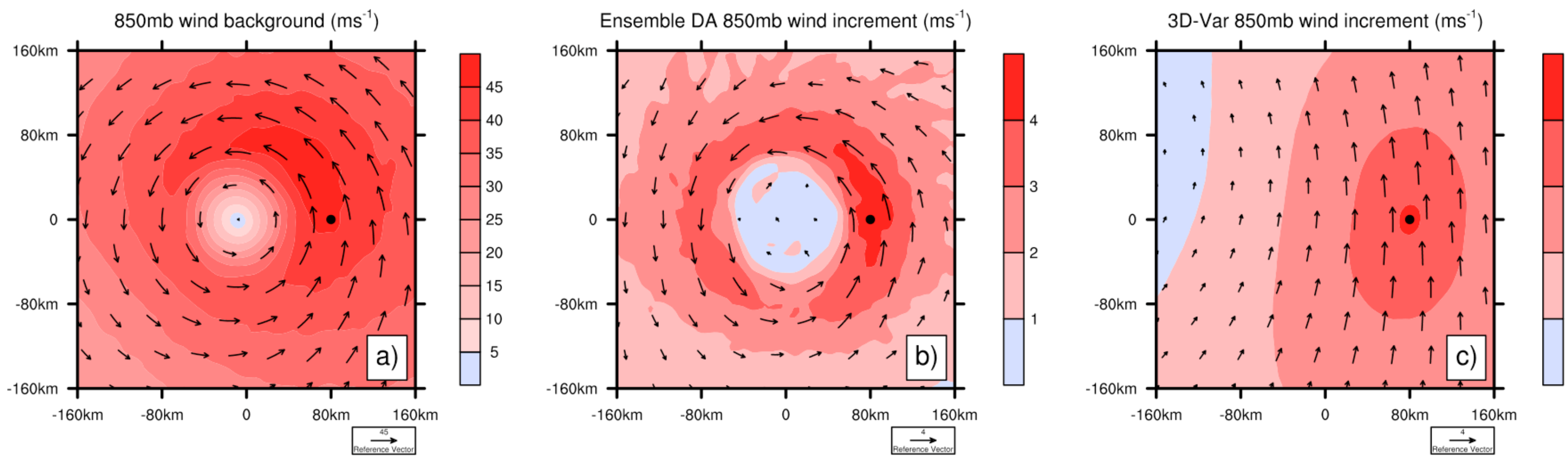


figure c/o Xuguang Wang, University of Oklahoma

(more)

Questions

- Can hurricane ensemble forecasts from a global model be improved substantially by:
  - using an EnKF for data assimilation & ensemble initialization?
  - using a higher-resolution version of global model?
  - using new "TCVitals" sea-level pressure observations in the data assimilation?
- How much of any improvement can be attributed to EnKF vs. higher resolution model vs. new obs?
- How do experimental forecasts compare with operational forecasts from worldwide centers?
- Multi-model ensembles provide improvement?

Testing performed

- Ran a global ensemble square-root filter ("EnKF") data assimilation
  - T382L64 (~35 km) version of NCEP GFS, 60 members
  - Full observational data stream + "TCVitals" (min central pressure)
  - 20-member ensemble forecasts to 7 days for most active days during hurricane season, late July to early October 2009.
- Other operational ensembles (next page)
- Also useful: deterministic forecasts from T382 GFS/EnKF, operational GFS/GSI and parallel GFS/GSI with TCVitals
- Compare against "best track" files compiled by NHC and Joint Typhoon Warning Center

Ensemble systems evaluated

- Run ourselves on NSF U Texas computer:
  - T382L64 "*GFS/EnKF*" (experimental)
  - 30-km NOAA "*FIM*" off GFS/EnKF IC's (experimental)
  - For diagnostics at the end of the talk, "*T126L28 GFS/EnKF*" initialized off T382L64 EnKF ICs
- Operational:
  - "*NCEP*" T126L28 GFS/GSI/ETR
  - "*CMC*" ensemble, 0.9-degree, L28, EnKF perts around 4D-Var control.
  - "*UKMO*" MOGREPS ensemble, 1.25*0.83-degree, L38, ETKF perts around 4D-Var control
  - "*ECMWF*" T399L62, v. 35r2 and 35r3 (with stochastic physics upgrade). Singular vector perts around 4D-Var control

What we don't have, and wish we did

- T382L64 GFS/EnKF and subsequent ensembles, *without* TCVitals observations
- T126L28 GFS/EnKF and subsequent ensembles (or T190L64) to cleanly examine effect of resolution
- A bigger sample (lackluster Atlantic season, only global-composite statistics likely to be worth interpreting).

Rules for including a particular storm in "homogeneous" comparisons of models A vs. B

- Storm must be tracked and at least tropical depression strength at initial time of forecast
- Ensemble scores com[...] forecasts computed.
- At least 8 members m[...] statistics; mean error [...] storms tracked.
- When performing "homogeneous" comparison of forecast model A to forecast model B, count a storm as a sample only when both models have forecast available.

This rule conveniently allows us to skirt the issue of whether the forecast model over-forecasts TC-genesis. In fact, both T382 GFS/EnKF and FIM/EnKF did over-forecast.

Definitions & metrics

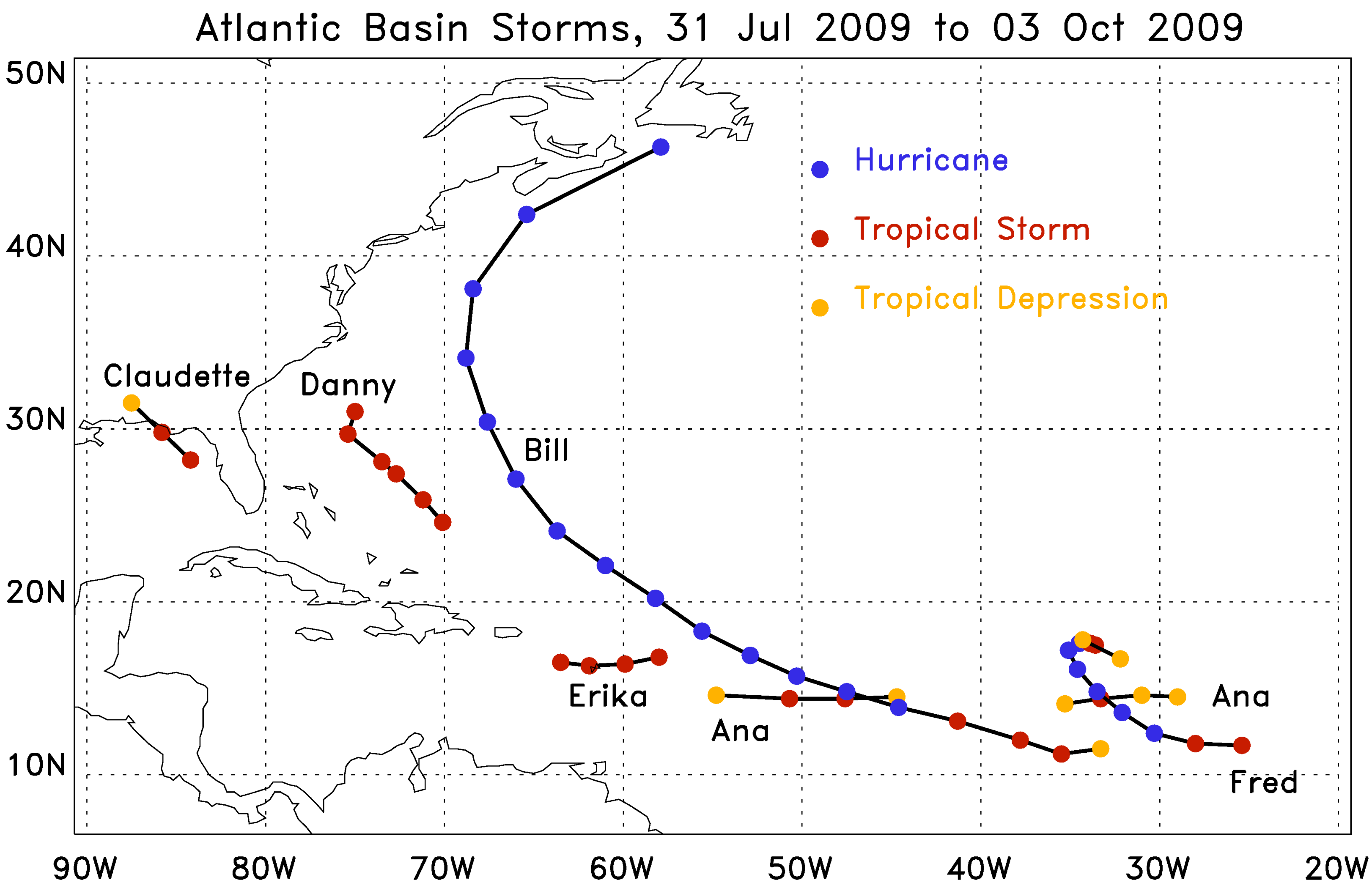
- Absolute error (km) of ens.-mean track forecast for the *i*th of *m* samples $E_i(t)$
- Abs. difference of the *j*th of *n* members from the ens mean: $D_{i,j}$
- Track average error: $\bar{E}(t) = \dfrac{\sum_{i=1}^{m} E_i(t)}{m}$
- Spread for ith sample ………………….. $S_i(t) = \dfrac{\sum_{j=1}^{n} D_{i,j}}{n}$
- Average spread : $\bar{S}(t) = \dfrac{\sum_{i=1}^{m} S_i(t)}{m}$

Data availability

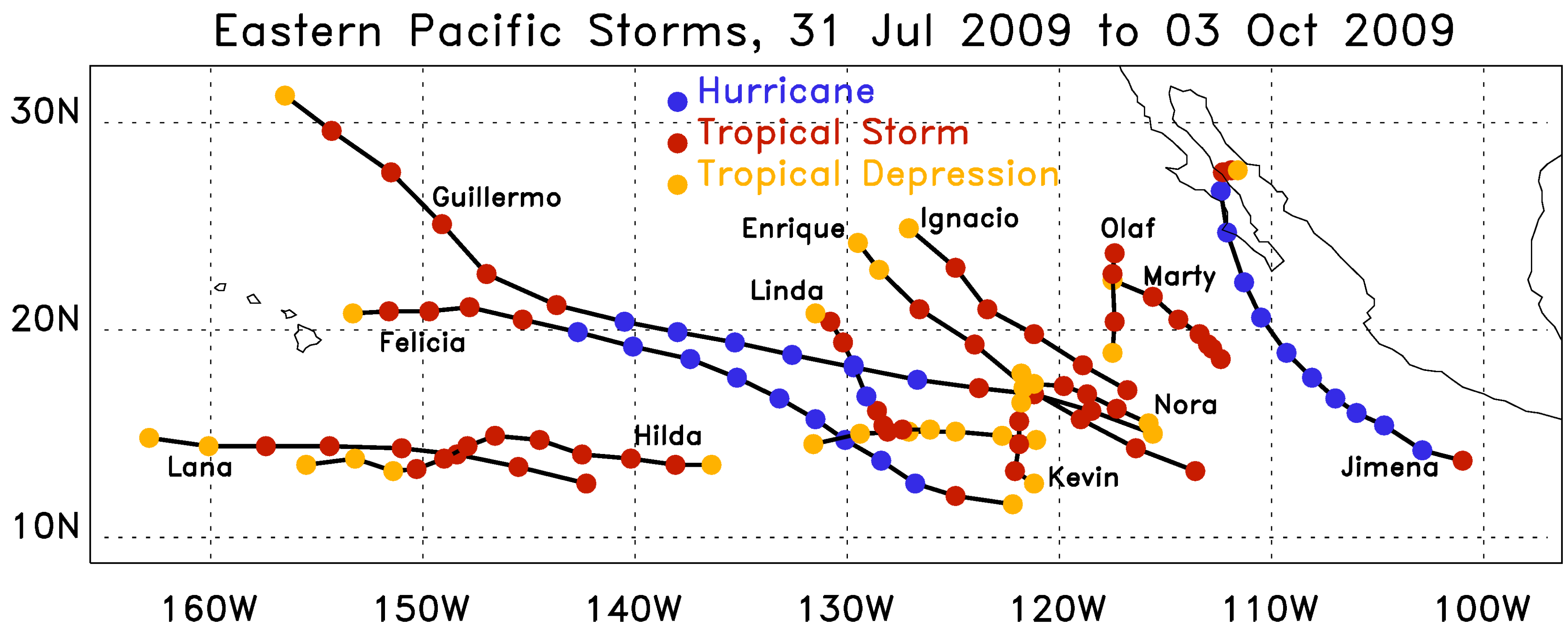
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
					Jul 31: F	Aug 1: F
2: F,E	3: F	4: F	5: F	6: F,U	7: F	8: F
9: F	10: F,G	11: F	12: F	13	14	15
16	17	18	19	20	21	22
23: U	24: F	25	26: G	27: F	28: F	29: F
30	31	Sep 1: G	2	3	4: F	5
6	7: F	8	9	10	11	12
13	14	15: C	16	17	18	19
20	21	22	23: F,E	24: F	25	26: E
27: E,U	28: U					

Table 1: Availability of 0000 UTC global ensemble forecast data between 31 July 2009 and 28 September 2009. For a particular date, "F" indicates that FIM ensemble data was unavailable for this initial time; E indicates that ECMWF ensemble was unavailable; U indicates UKMO; C indicates CMC; N indicates NCEP, and G indicates experimental GEFS/EnKF.

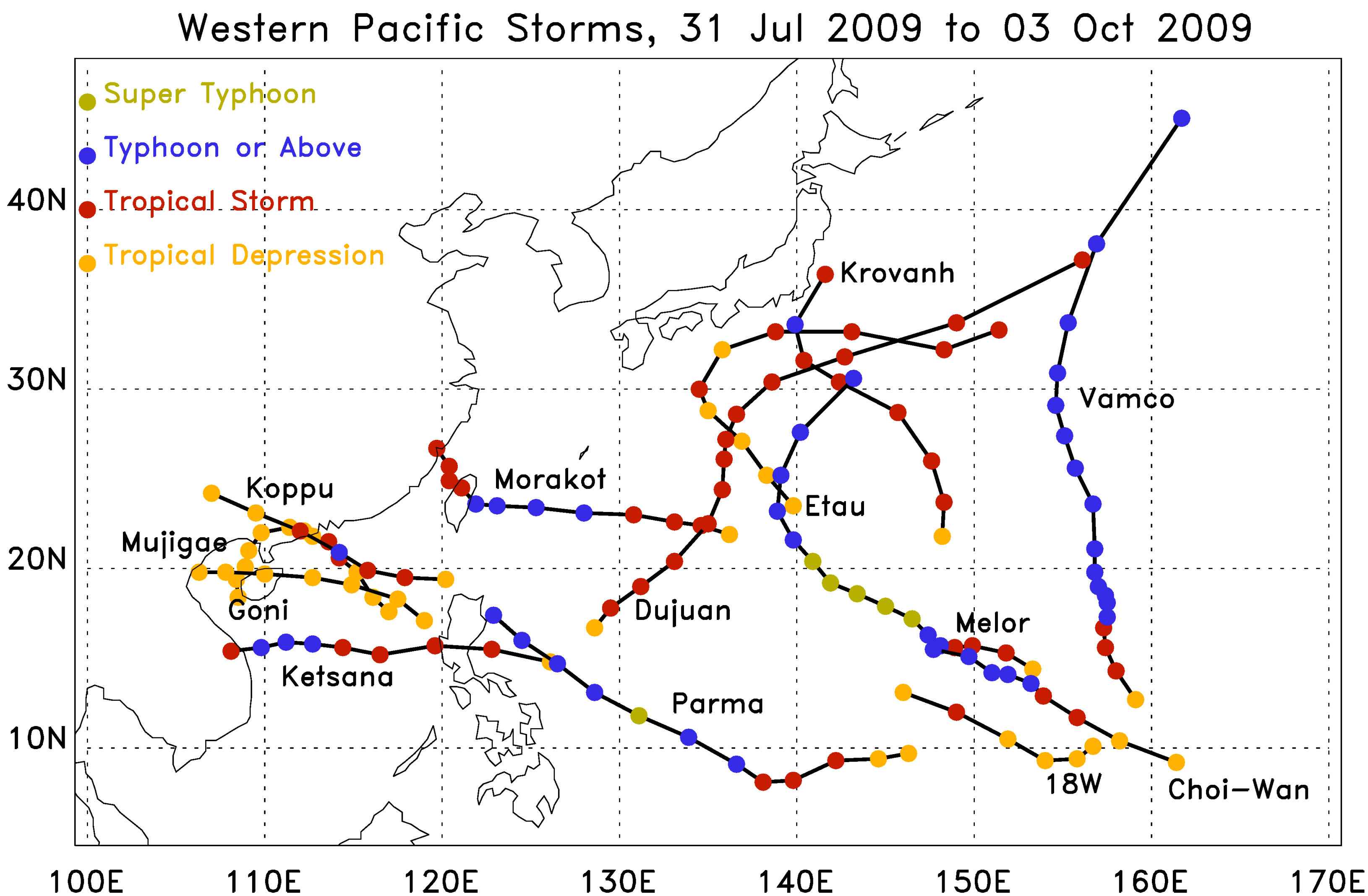
Review of Atlantic Basin activity



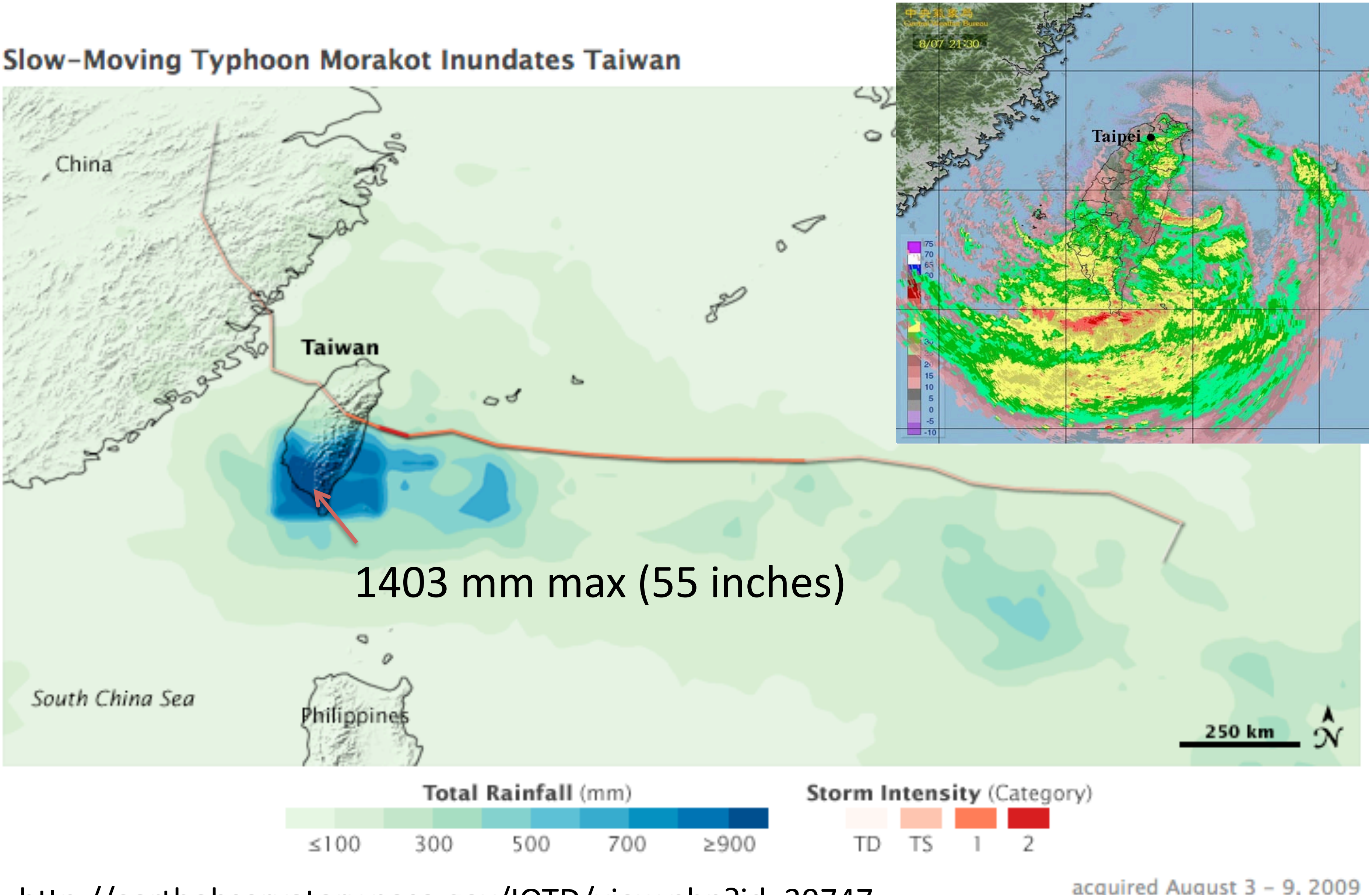
Review of Eastern-Pacific activity



Review of Western-Pacific activity



Typhoon Morakot (Taiwan floods)



http://earthobservatory.nasa.gov/IOTD/view.php?id=39747

http://www.boston.com/bigpicture/2009/08/typhoon_morakot.html

http://www.boston.com/bigpicture/2009/08/typhoon_morakot.html

http://www.boston.com/bigpicture/2009/08/typhoon_morakot.html

http://www.boston.com/bigpicture/2009/08/typhoon_morakot.html

Initialized 00 UTC
5 August 2009.

* indicates observed best-track position.

Bi-variate normal distribution fit to ensemble member positions; contour encloses 90% of fitted probability.

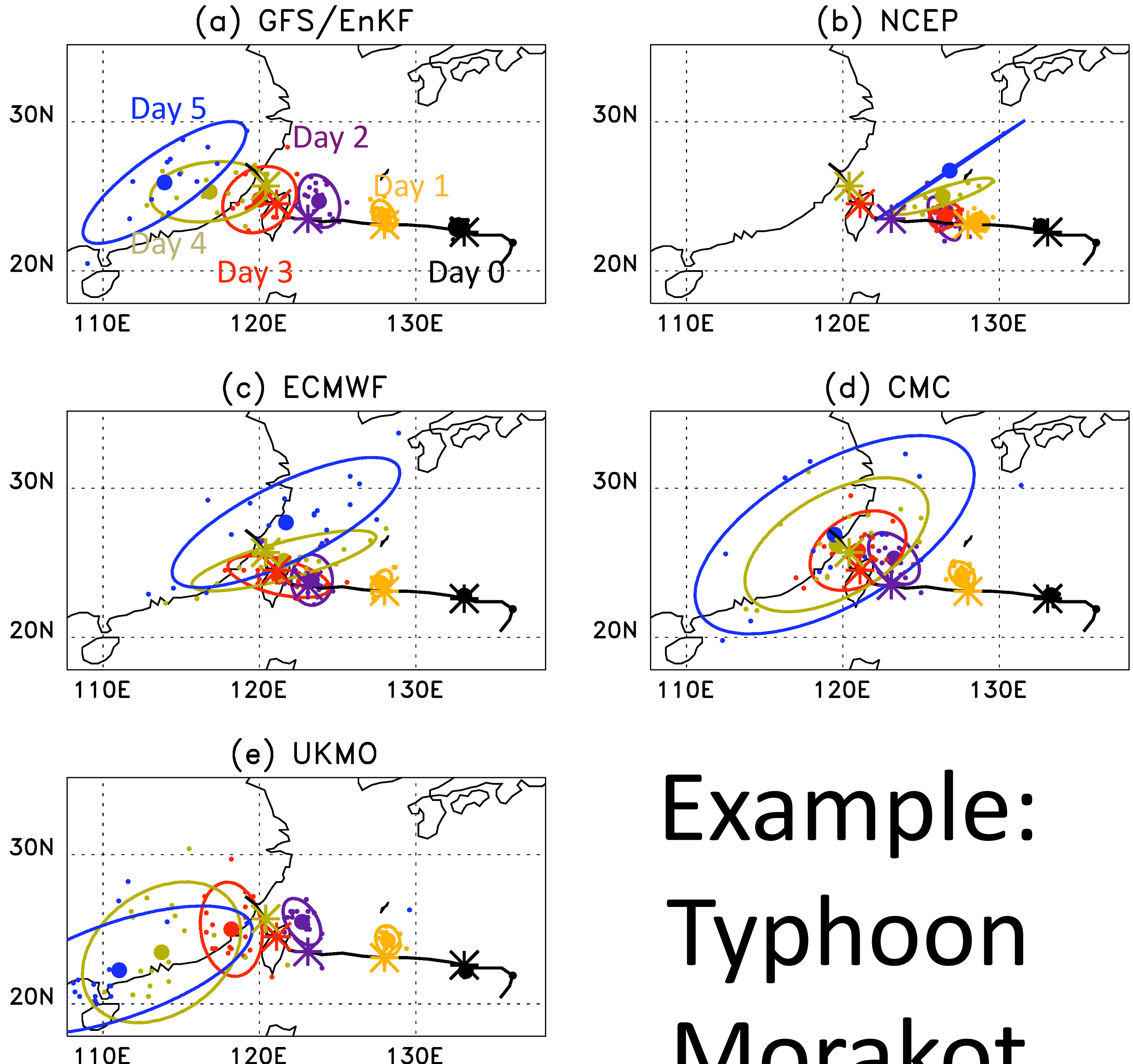
GEFS/EnKF a bit north and too fast.

NCEP has northward & westward bias, few members track.

ECMWF tracks decent up to Taiwan landfall

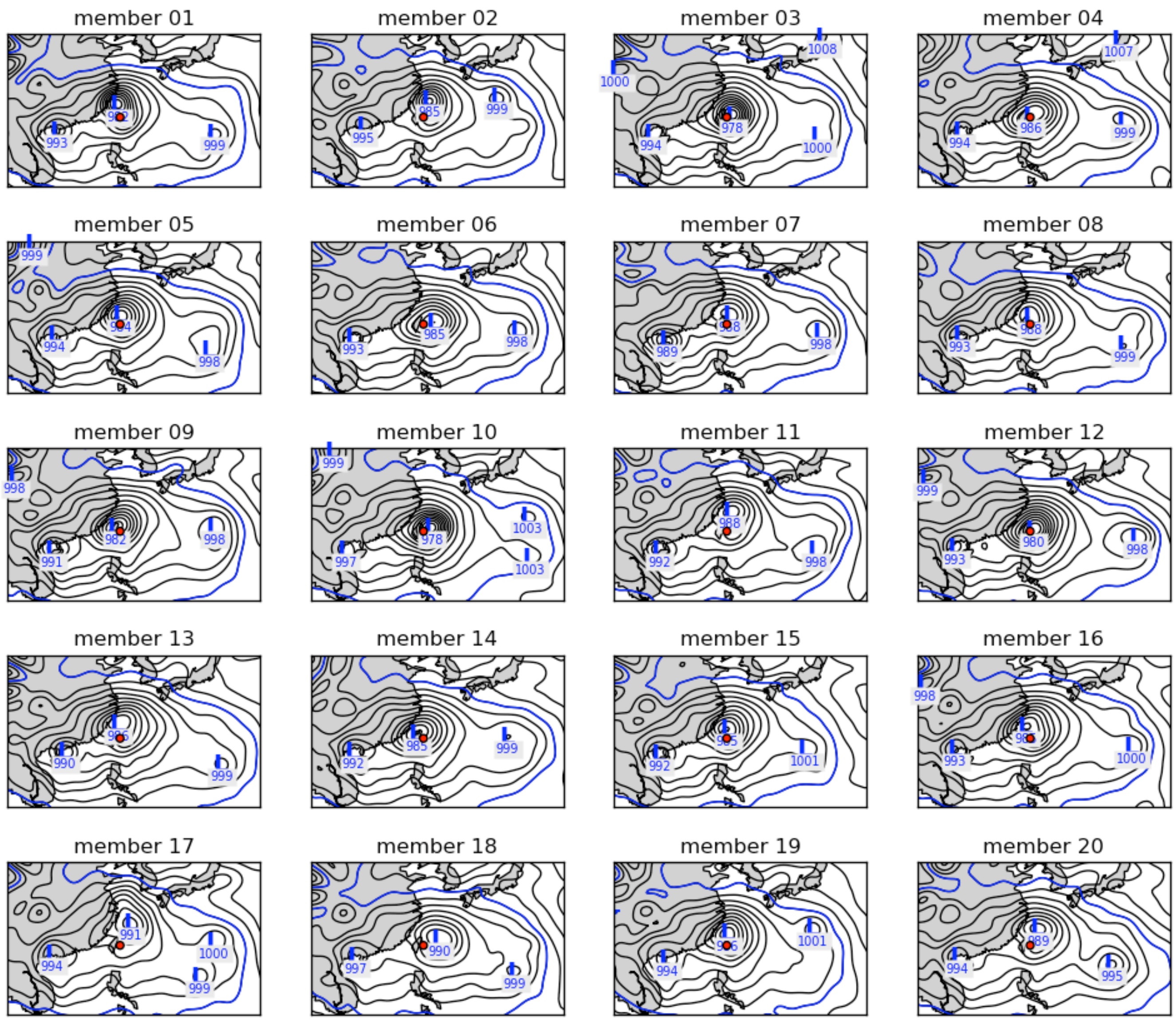
CMC has very large spread, esp. after landfall.

UKMO too north, too fast.



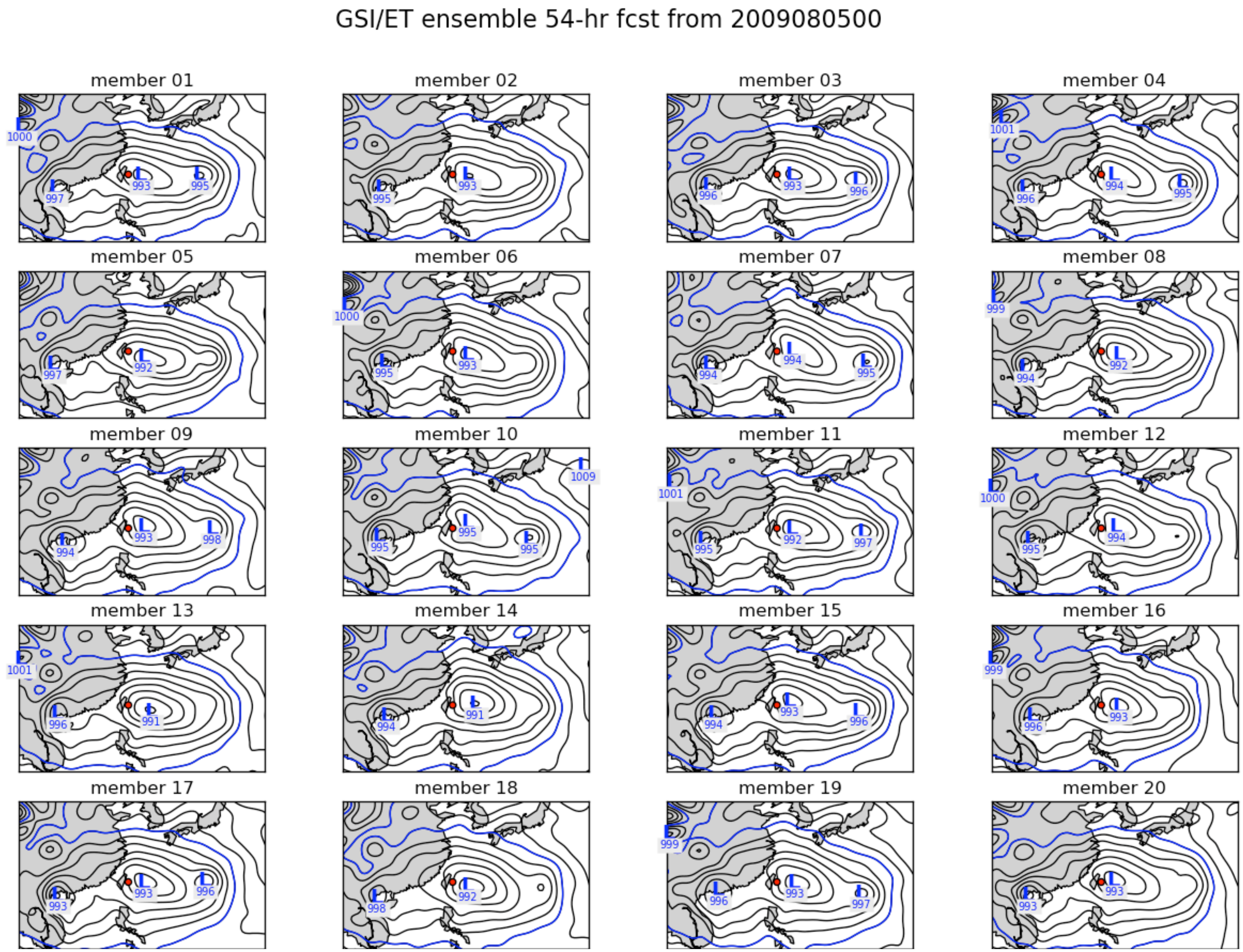
Example: Typhoon Morakot

54-h ensembles from T382 GFS & EnKF initial conditions.



Intense vortices in forecasts, with ensembles of forecast positions relatively close to the observed position (red dot).

54-h ensembles from experimental T382 GFS & GSI / operational ET perturbations



Note that GFS model resolution is much greater than current operational, T126

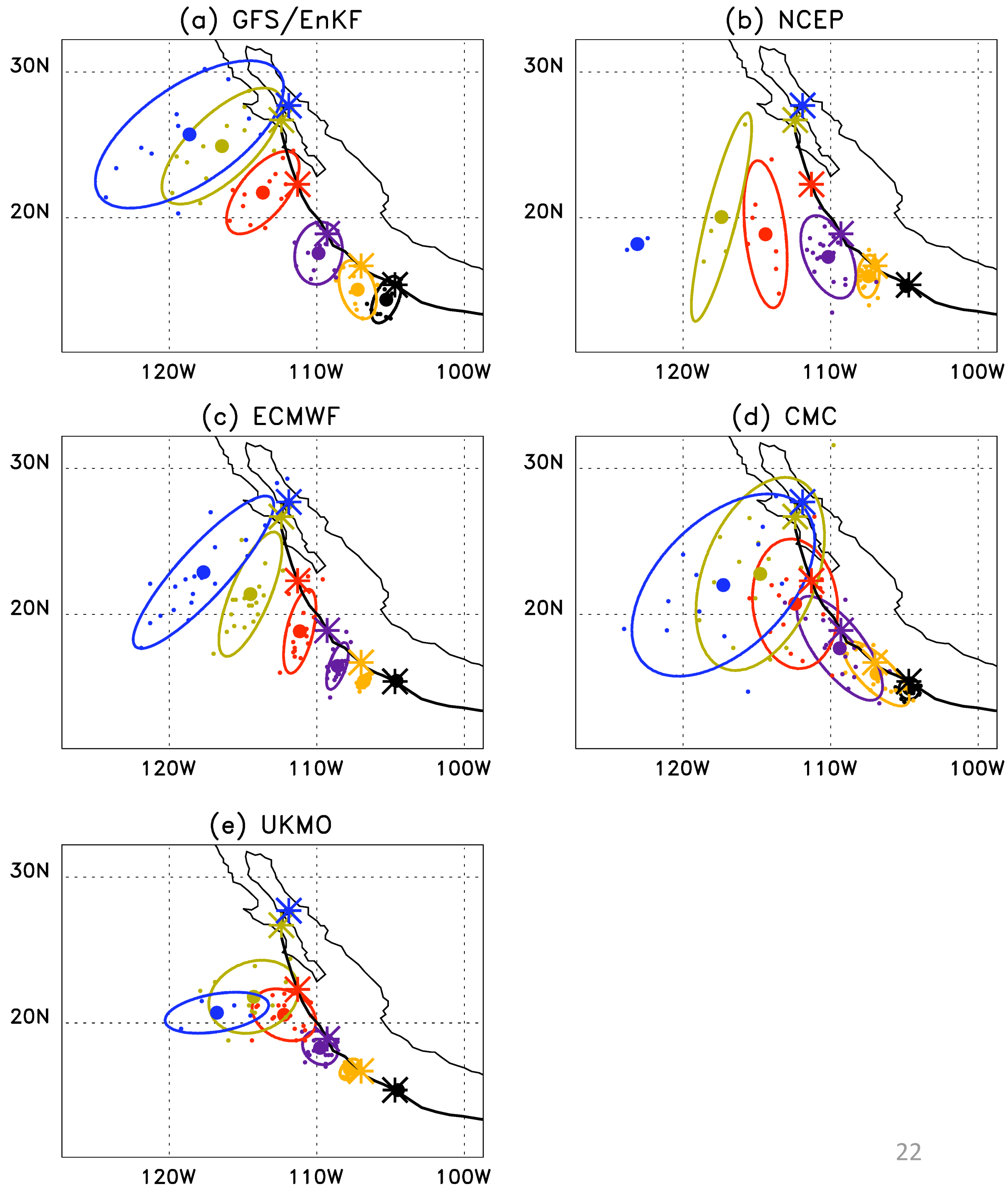
GSI-ET initialized ensemble produces less intense vortices, and forecasts are slow in moving typhoon west.

This operational version of GSI did not include TCVitals central pressure obs.

Example: hurricane Jimena

Initialized 00 UTC 30 Aug 2009

all models have westward bias; none of the forecasts particularly good.



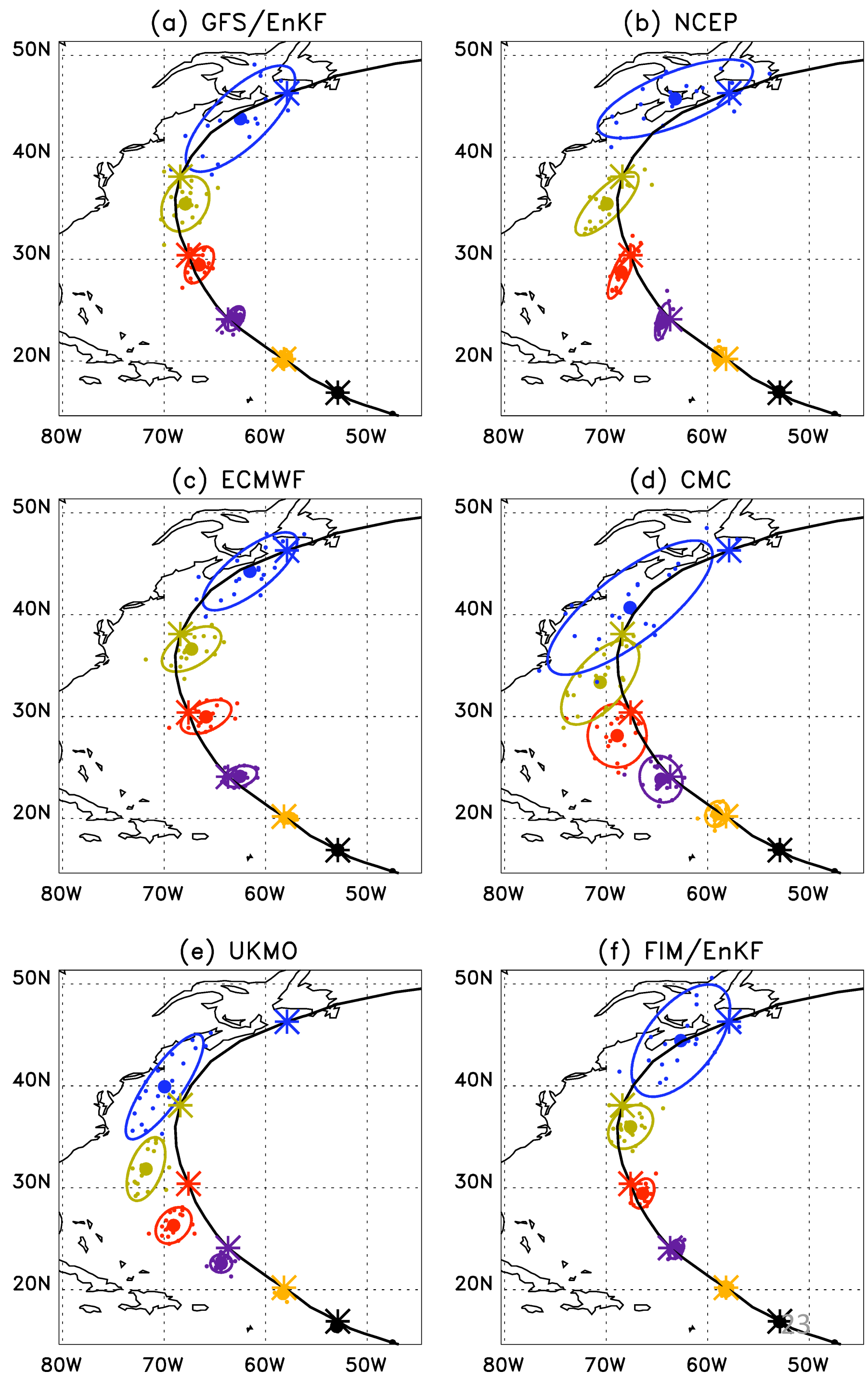
Example: Hurricane Bill

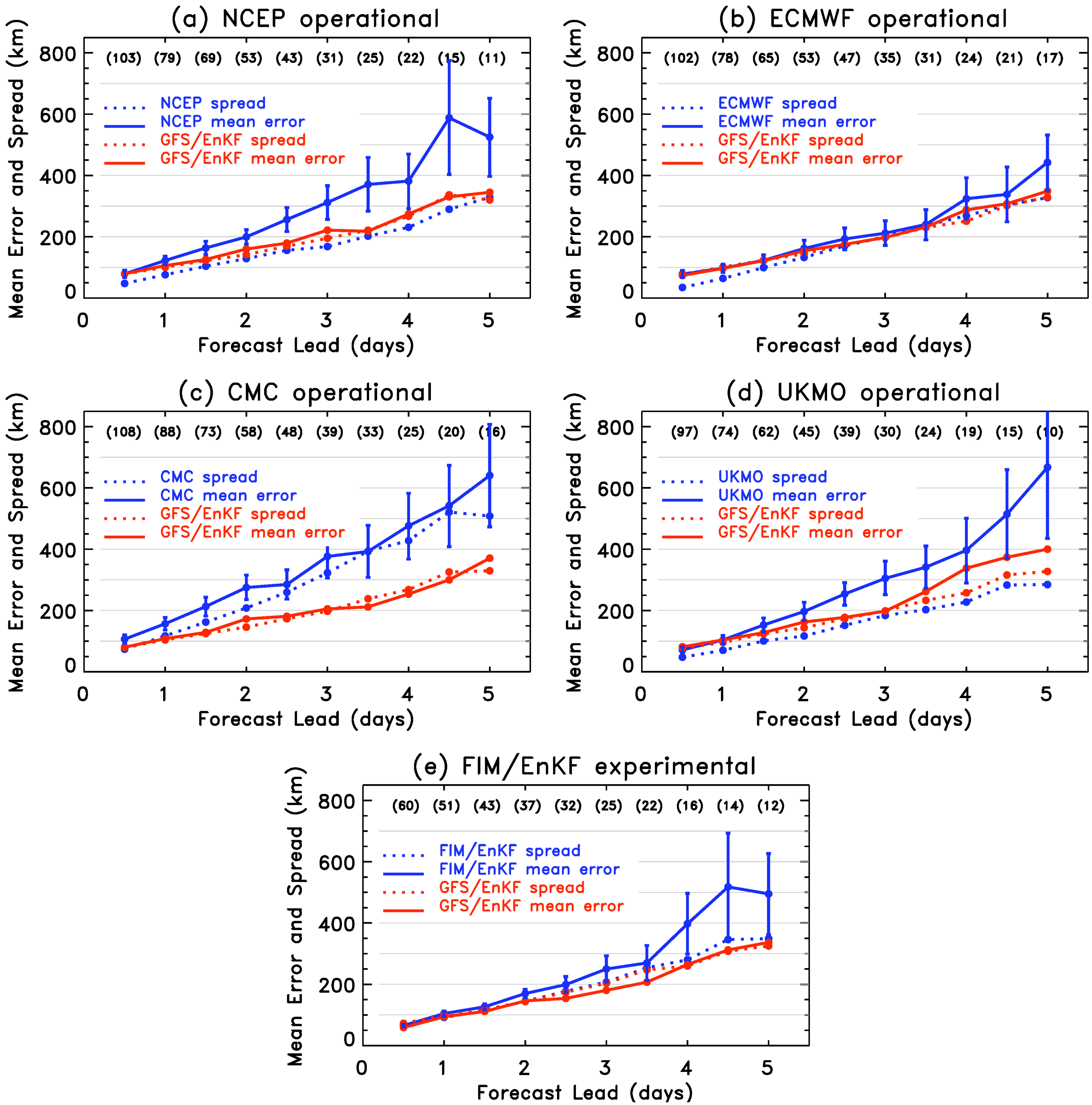
Initialized 00 UTC 19 August 2009.

All models slow, to varying extents.

GEFS/EnKF, ECMWF , NCEP, FIM tracks decent.

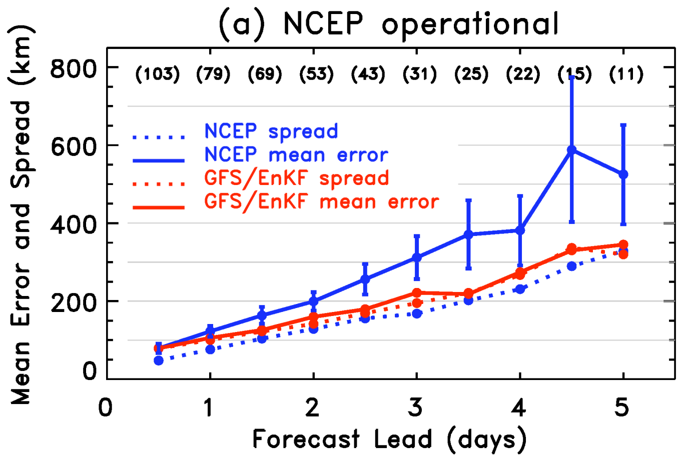
UKMO, CMC have westward bias.

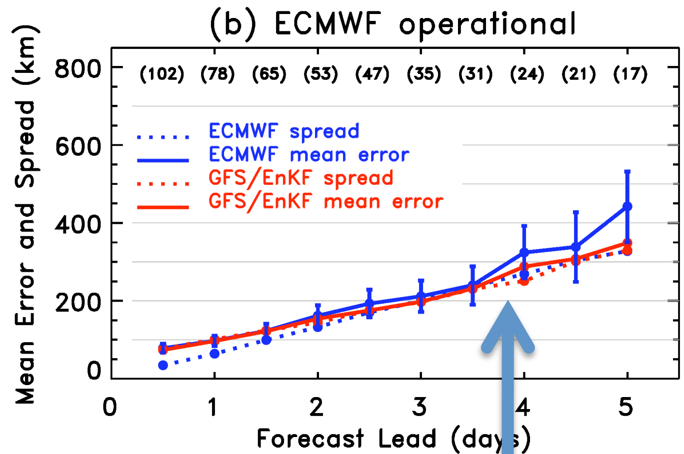




Track error major findings:

(1) Experimental T382 GFS/EnKF beats NCEP operational handily.
(2) Experimental T382 GFS/EnKF competitive with ECMWF
(3) Experimental T382 GFS/EnKF has overall spread-error calibration.
(4) FIM/EnKF not quite as skillful as GFS/EnKF.
(5) CMC not as skillful, but calibrated.
(6) UKMO not as skillful, under-spread.



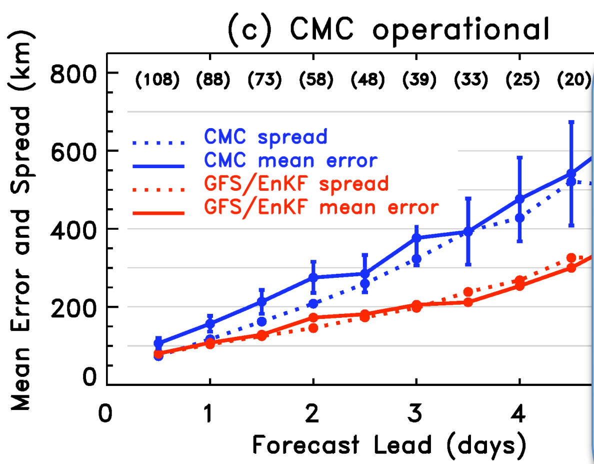


Track error major findings:

(1) Experimental T382 GFS/EnKF beats NCEP operational handily.

(2) Experimental T382

...e



A notable success for NCEP GFS and ESRL EnKF!

A word of caution: look at other norms, and ECMWF still has a substantial lead.

...2

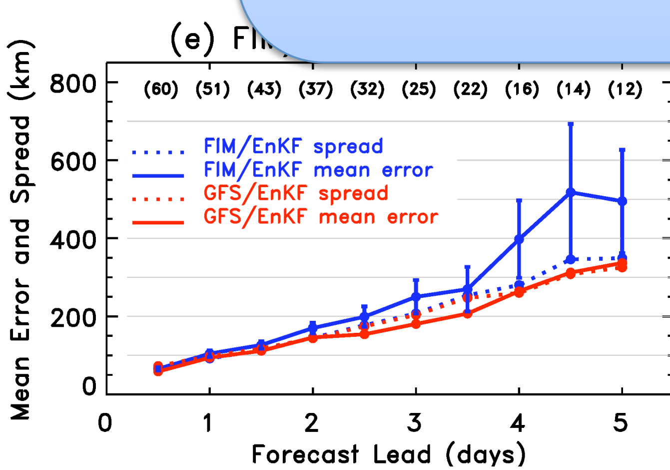
...on.

...te

...F.

...l,

but calibrated.



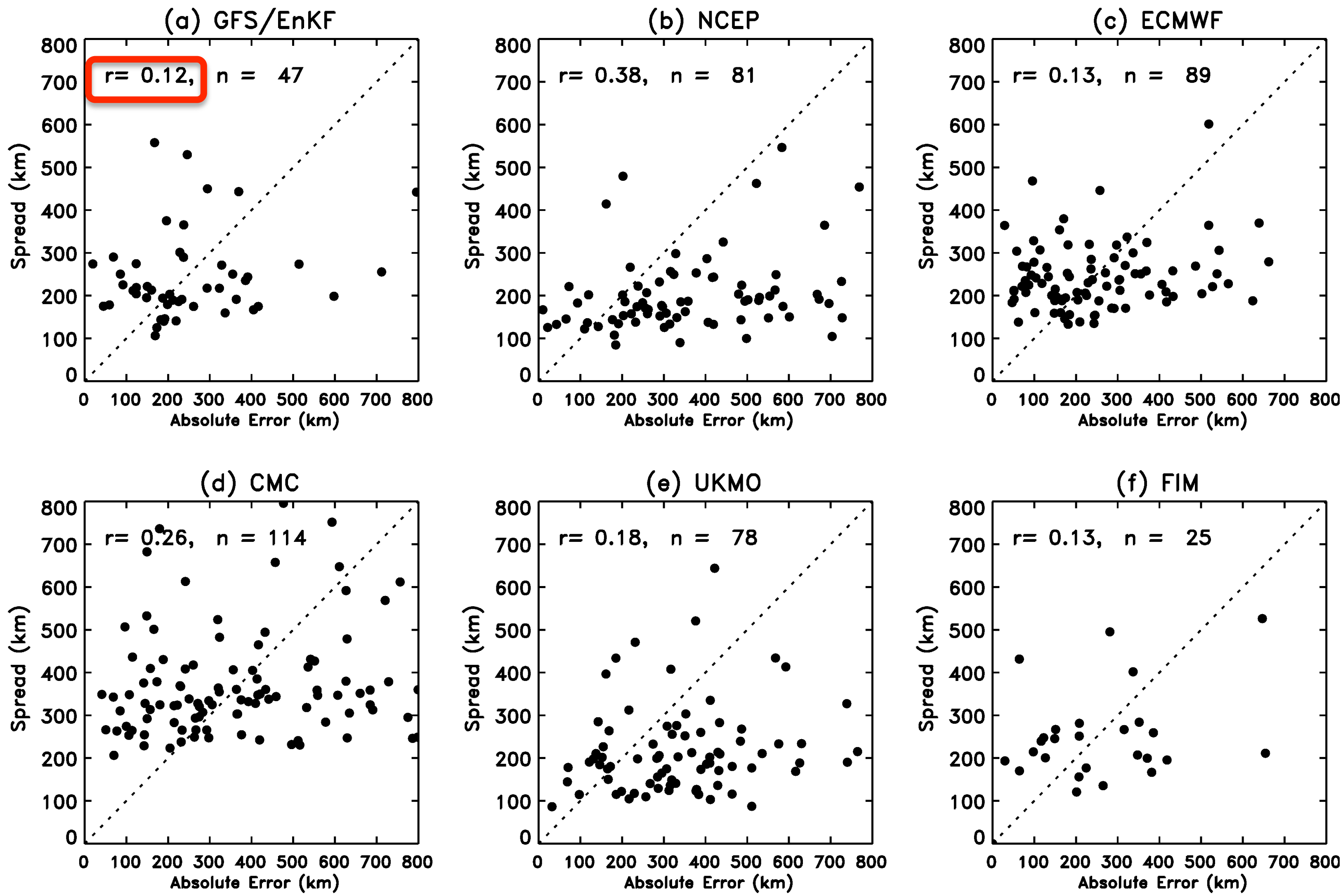
(6) UKMO not as skillful, under-spread.

Example: RMS error and AC, Z500

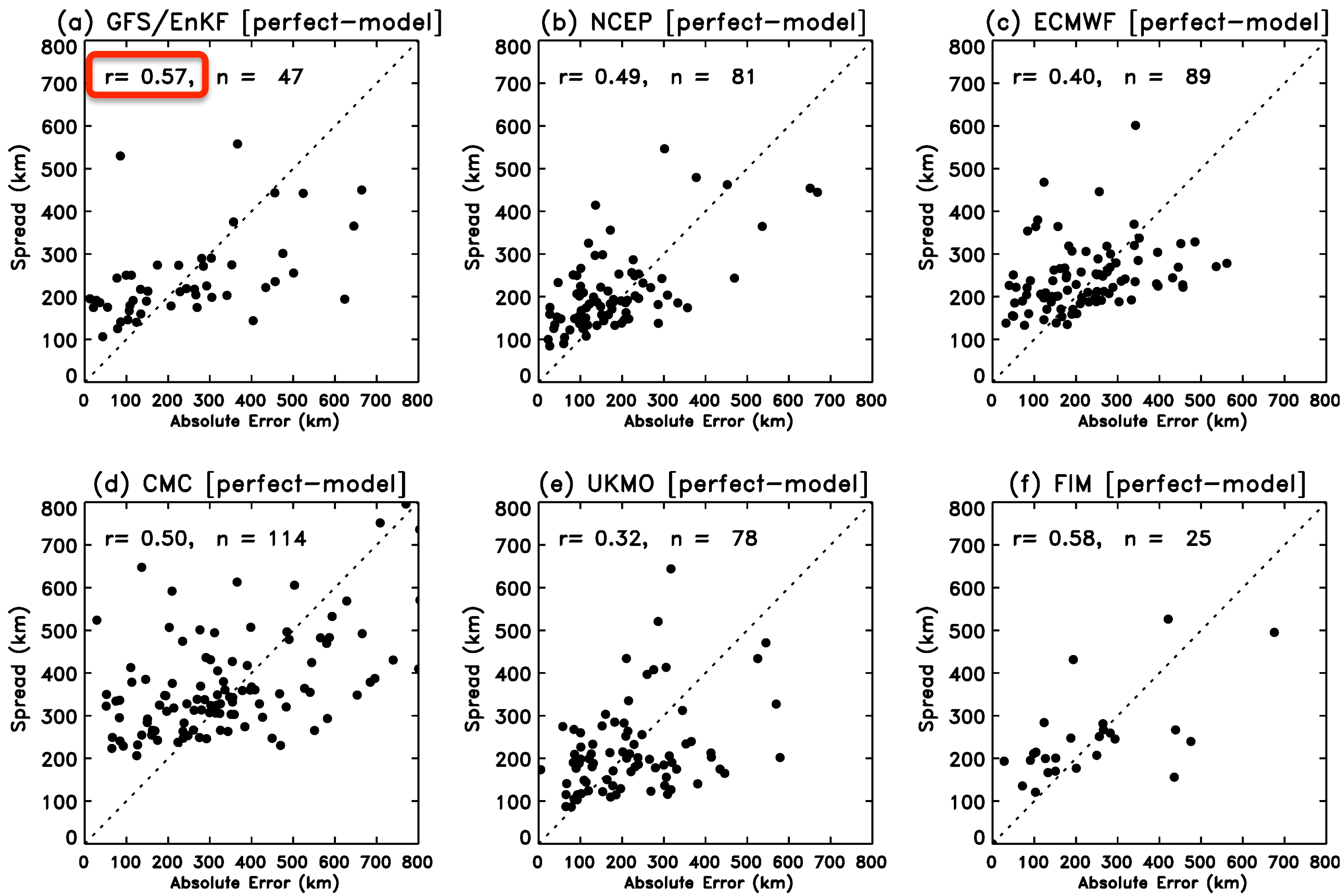
	ECMWF T399	GFS/GSI-ET T126	GFS/EnKF T382
RMSE 500-mb height, N. Hem.	32.35	39.98	36.22
RMSE 500-mb height, N. Hem.	51.14	63.12	56.72
AC 500-mb height, N. Hem.	0.888	0.832	0.854
AC 500-mb height, N. Hem.	0.891	0.829	0.856

Table 1: Errors and anomaly correlations of forecasts from the 2009 operational ECMWF T399 ensemble-mean forecasts, the operational GFS-based ensemble at NCEP (GSI initial condition, T126 forecast model), and the experimental T382 GFS ensemble initialized with the EnKF. All errors are measured with respect to the own products's analysis, and all verifications are performed on a 2.5-degree lat-lon grid. RMSE indicated the root-mean square error, AC the anomaly correlation.

Spread-error relations? (Day-3)



Spread-error relations, perfect-model assumption



Ellipse eccentricity analysis

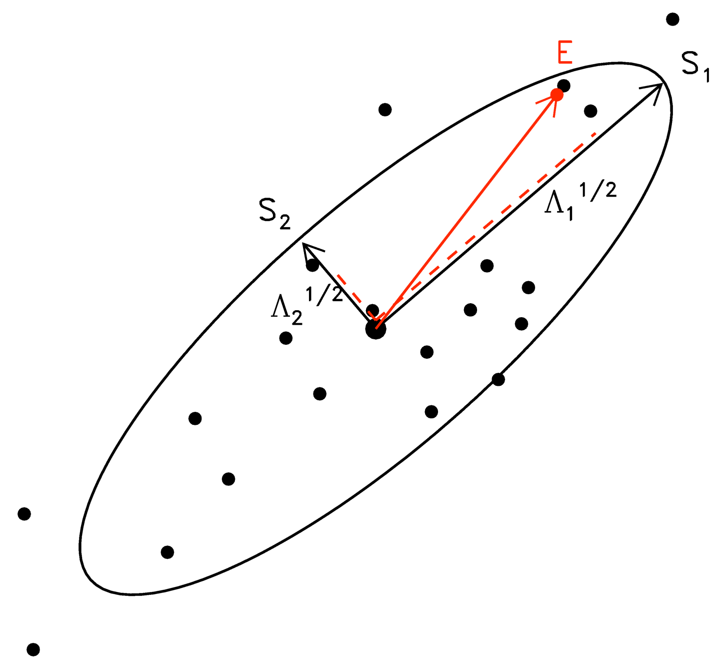
Question: are errors projections larger along the direction where the ellipse is stretched out?

$$\mathbf{x}'_\lambda = \left(x_{\lambda(1)} - \overline{x}_\lambda,\ \ldots\ , x_{\lambda(nt)} - \overline{x}_\lambda\right) \Big/ \left(nt-1\right)^{1/2}$$

$$\mathbf{x}'_\phi = \left(x_{\phi(1)} - \overline{x}_\phi,\ \ldots\ , x_{\phi(nt)} - \overline{x}_\phi\right) \Big/ \left(nt-1\right)^{1/2}$$

$$\lambda = longitude, \quad \phi = latitude, \quad nt = \#\, tracked$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_\lambda \\ \mathbf{x}'_\phi \end{bmatrix}$$



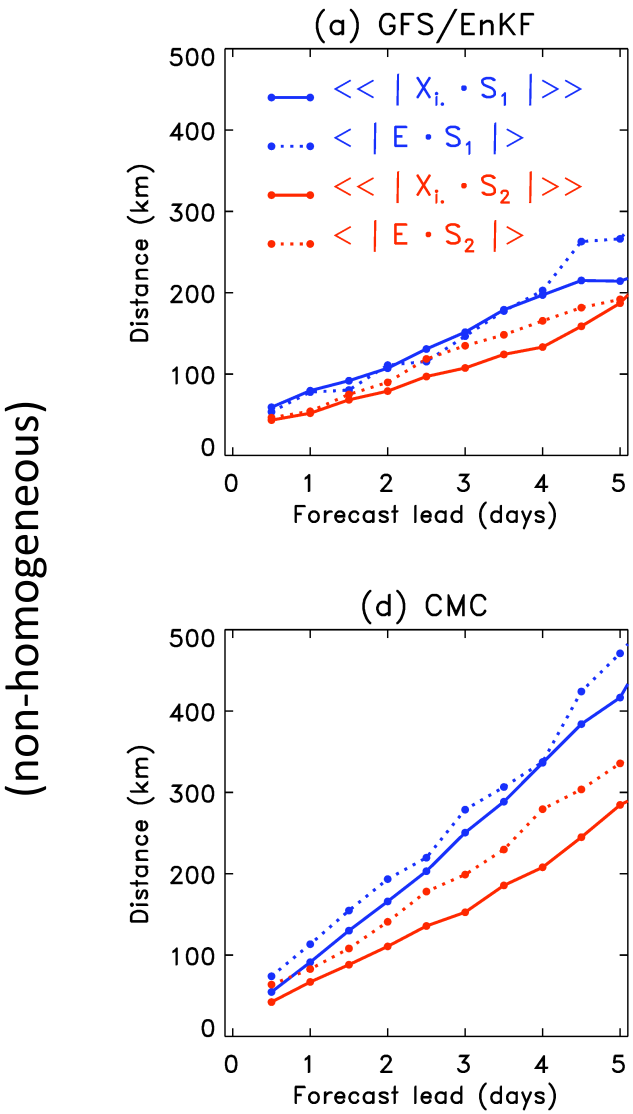
$$\mathbf{F} = \mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{S}\Lambda\mathbf{S}^{-1} = \mathbf{S}\Lambda\mathbf{S}^{\mathrm{T}} = \left(\mathbf{S}\Lambda^{1/2}\right)\left(\mathbf{S}\Lambda^{1/2}\right)^{\mathrm{T}}$$

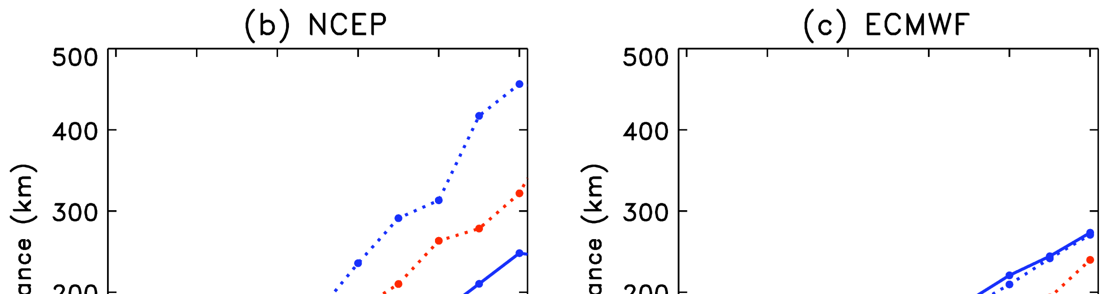
$\left\langle \left| \mathbf{E} \bullet \mathbf{S}_1 \right| \right\rangle$ should be consistent with $\left\langle \left\langle \left| \mathbf{X}_{i\bullet} \bullet \mathbf{S}_1 \right| \right\rangle \right\rangle$

$\left\langle \left| \mathbf{E} \bullet \mathbf{S}_2 \right| \right\rangle$ should be consistent with $\left\langle \left\langle \left| \mathbf{X}_{i\bullet} \bullet \mathbf{S}_2 \right| \right\rangle \right\rangle$

$\langle \bullet \rangle$ = average over cases; $\langle\langle \bullet \rangle\rangle$ = average over cases, members

Ellipse eccentricity analysis





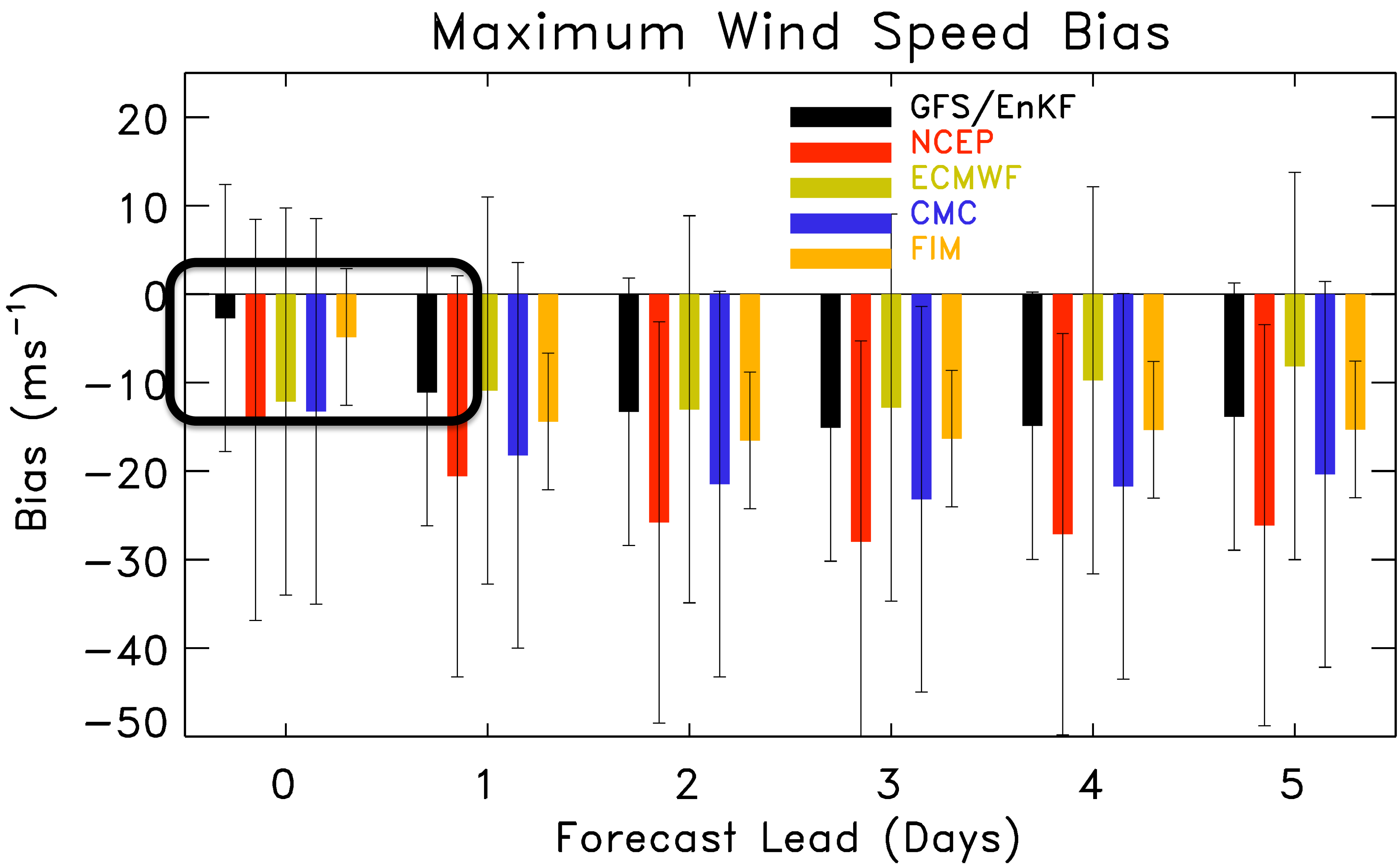
Notes for GFS/EnKF:

(1) Along major axis of ellipse, consistent average projection error of errors and projection of members; spread well estimated.

(2) Along minor axis of ellipse, slightly larger projection of errors than projection of members. Too little spread.

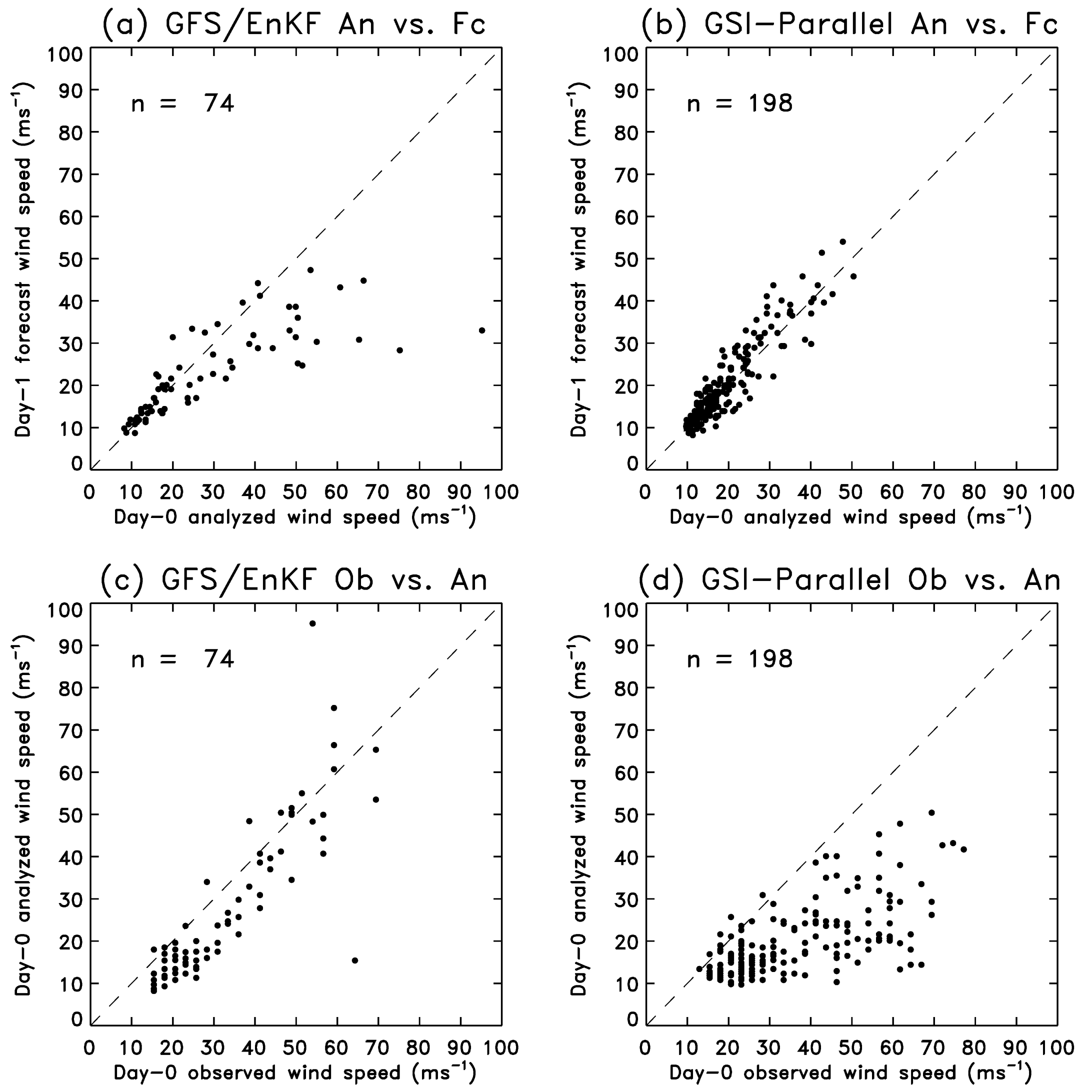
(3) Together, imply more isotropy needed.

(4) Still (dashed lines) some separation of projection of error onto ellipses indicates there is some skill in forecasting ellipticity.



non-homogeneous; error bars are 5th, 95th percentiles of normal distribution fit to data.

Source of rapid decrease of GFS/EnKF wind speeds between day 0 and day 1?

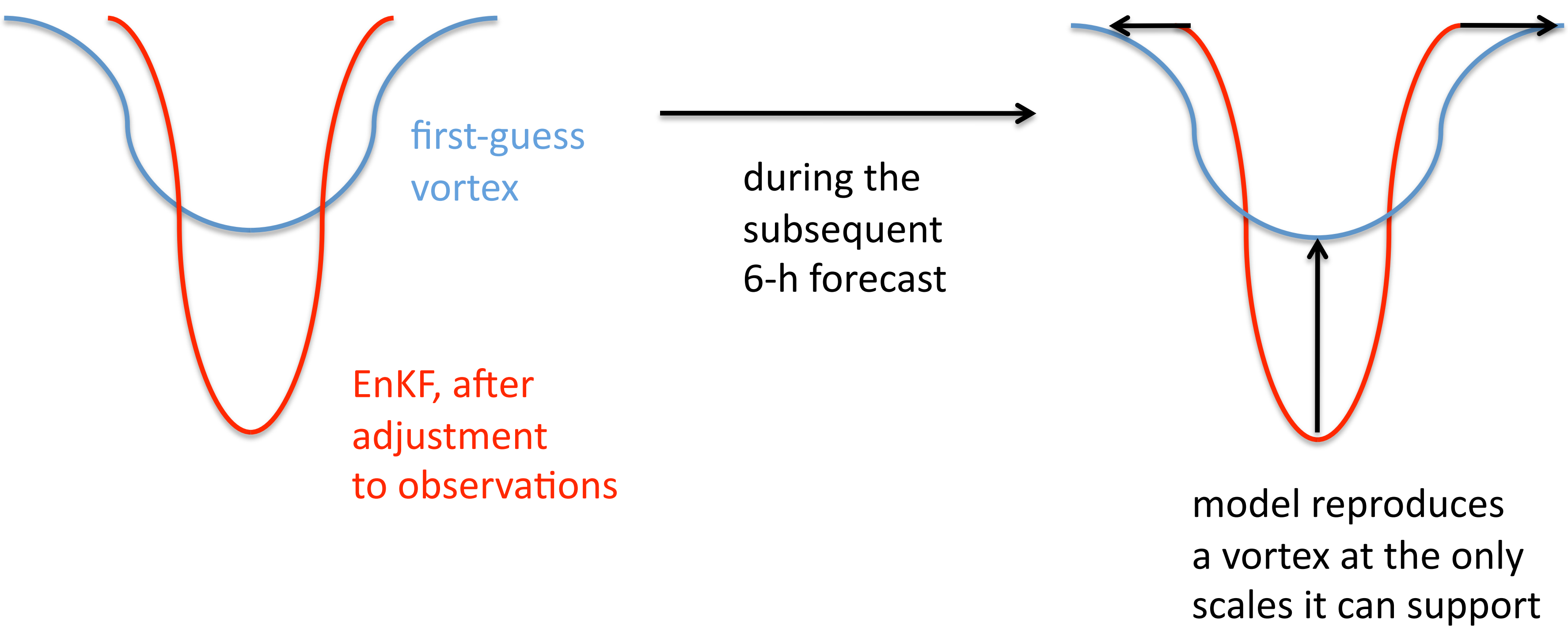


Rapid decrease in speed of GFS/EnKF forecasts; not so for GSI initialized

But looking at analyzed wind speeds, GFS/EnKF produces appropriately strong vortex, GSI does not.

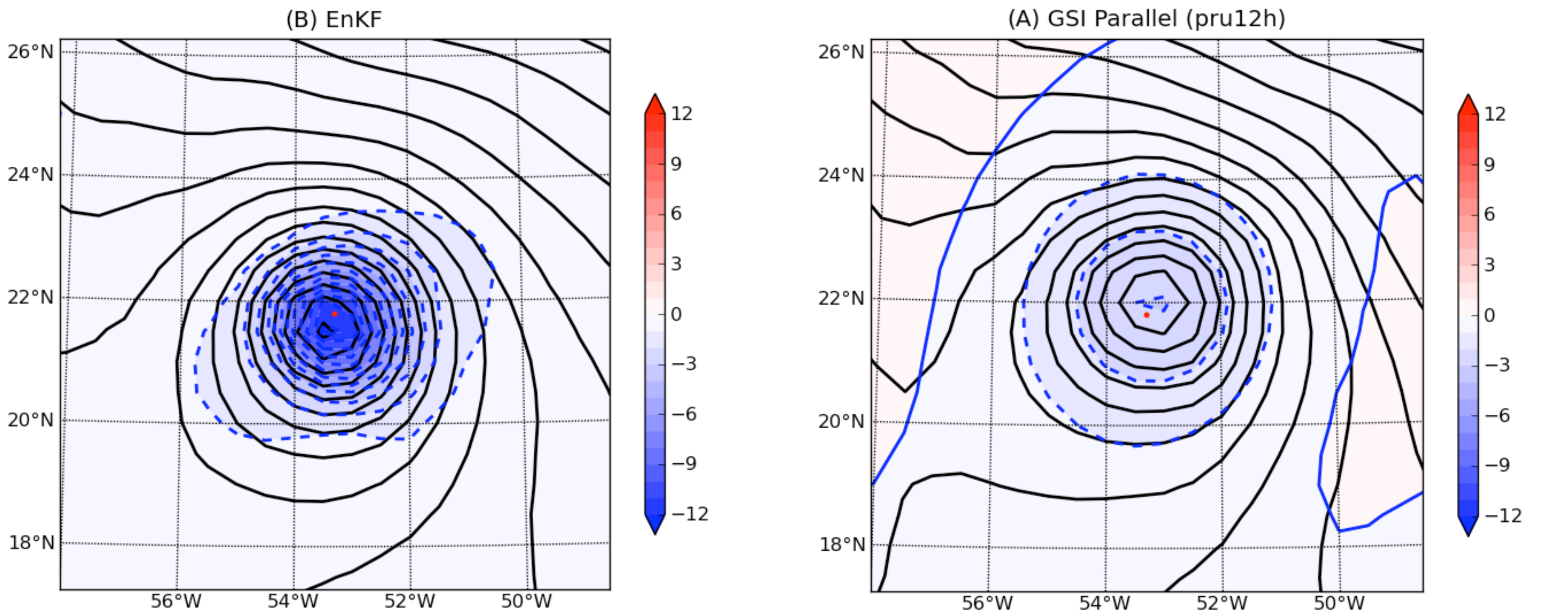
Why the persistent under-forecast of the strength of vortices?

at coarse resolution the model simply cannot support intense vortices, even if properly analyzed

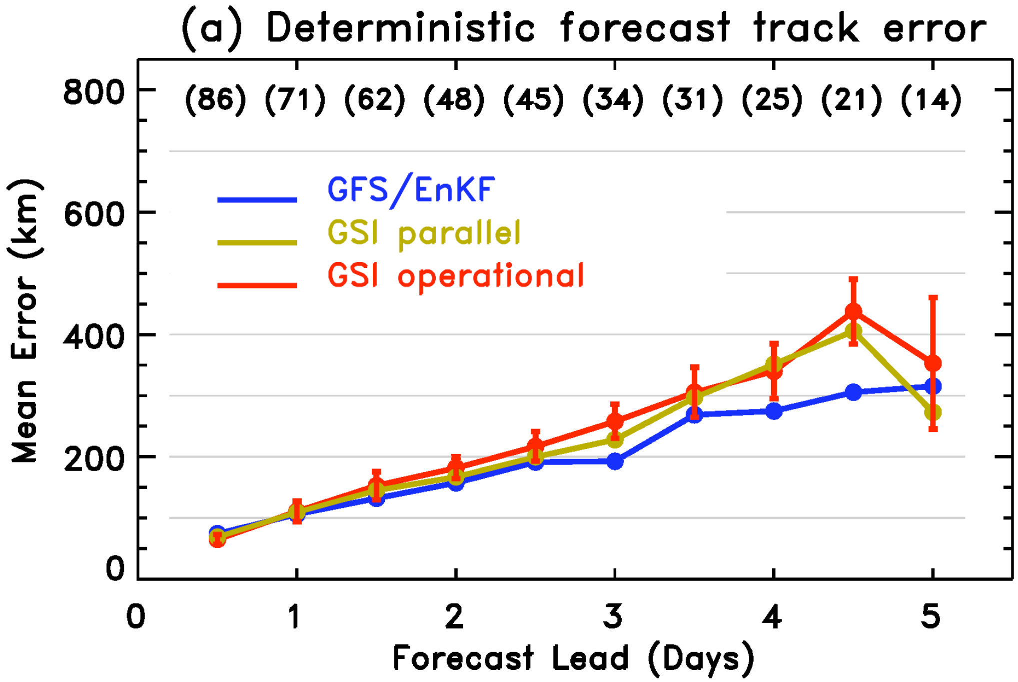


Increments in GFS/EnKF and GSI-parallel to TCVitals SLP

Hurricane Ike, 00 UTC 4 September 2008

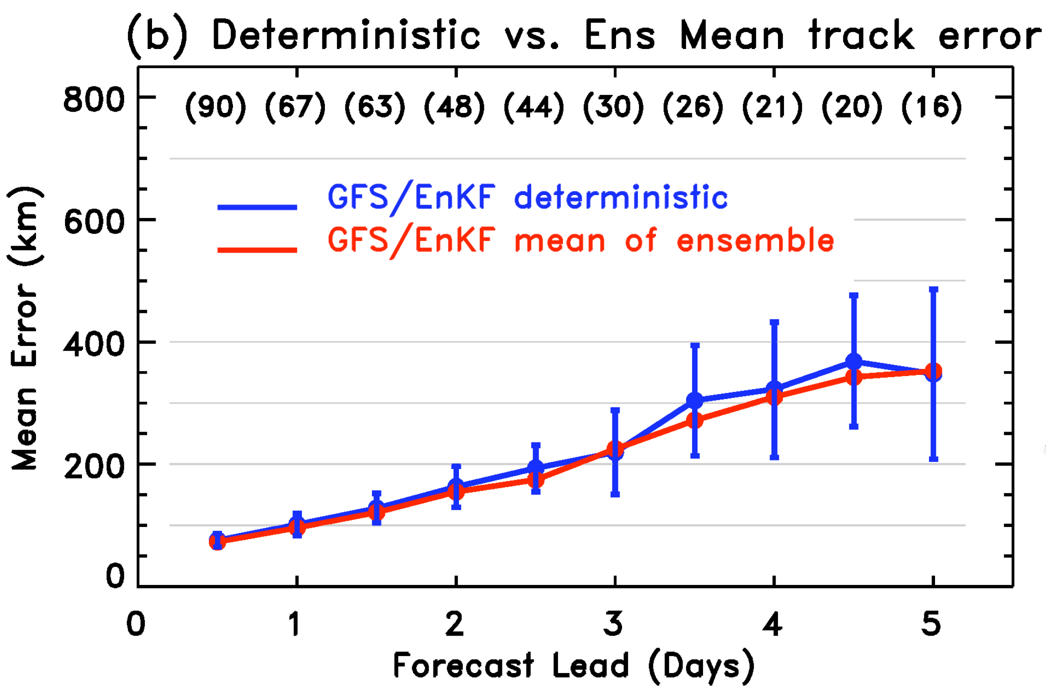


Change to EnKF initializes much deeper, tighter vortex; contours every 1 hPa.



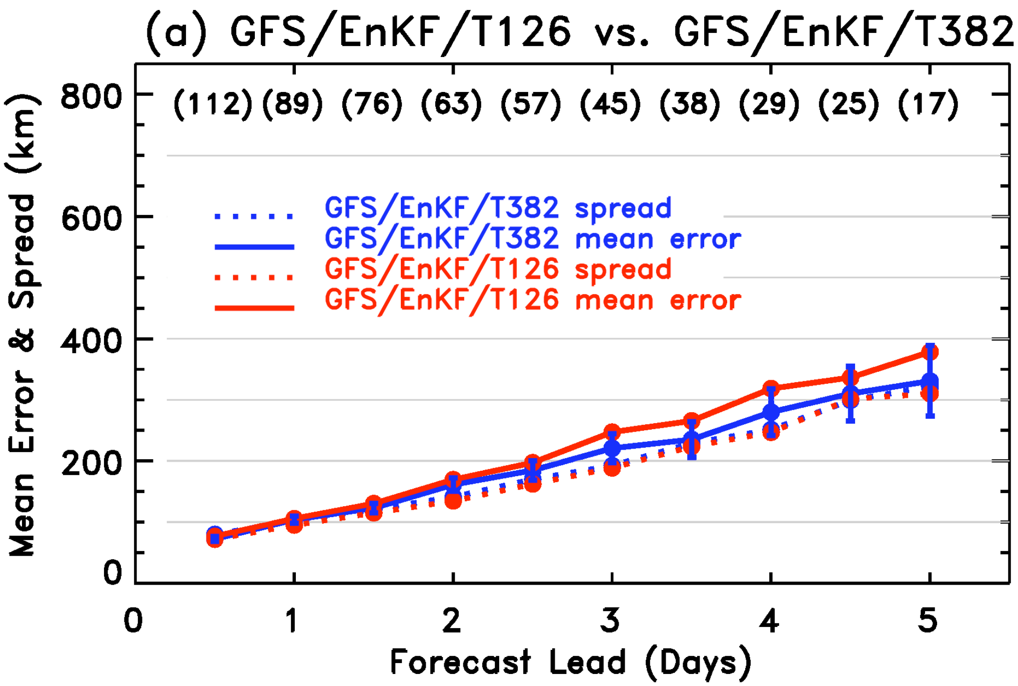
Trying to understand effects of TCVitals, EnKF, resolution, ensemble averaging

Surprisingly, GFS parallel with TCVitals didn't have appreciably lower errors than GSI operational. Maybe different effect with EnKF? Also, note Daryl Kleist, NCEP/EMC saw bigger impact with 2008 season.

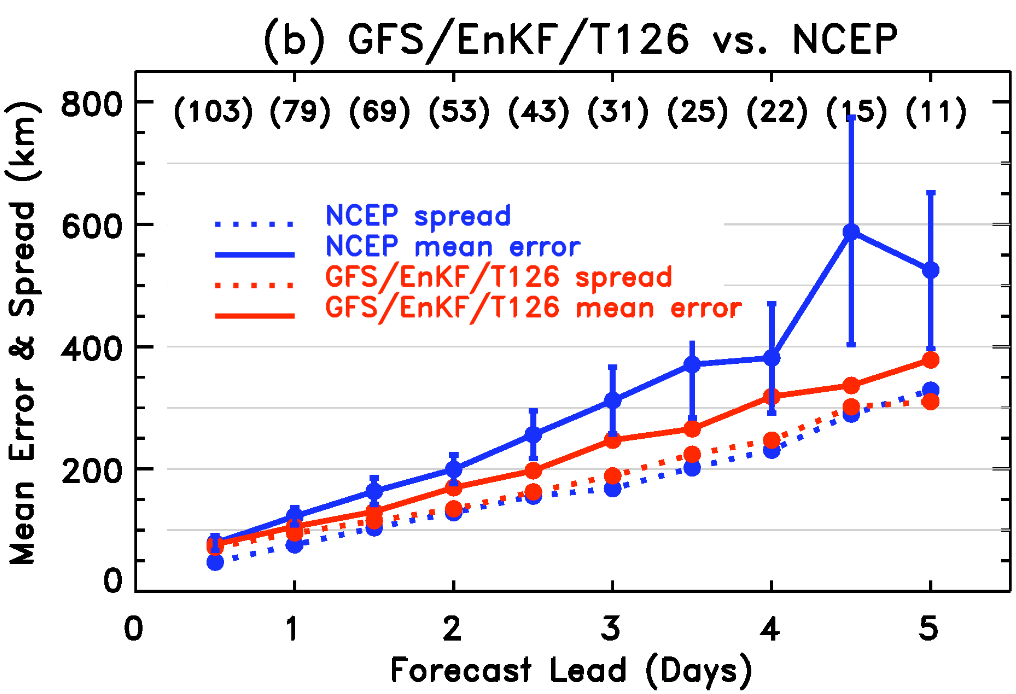


Ensemble averaging did not appear to have a large effect on track forecast performance.

Trying to understand effects of TCVitals, EnKF, resolution, ensemble averaging

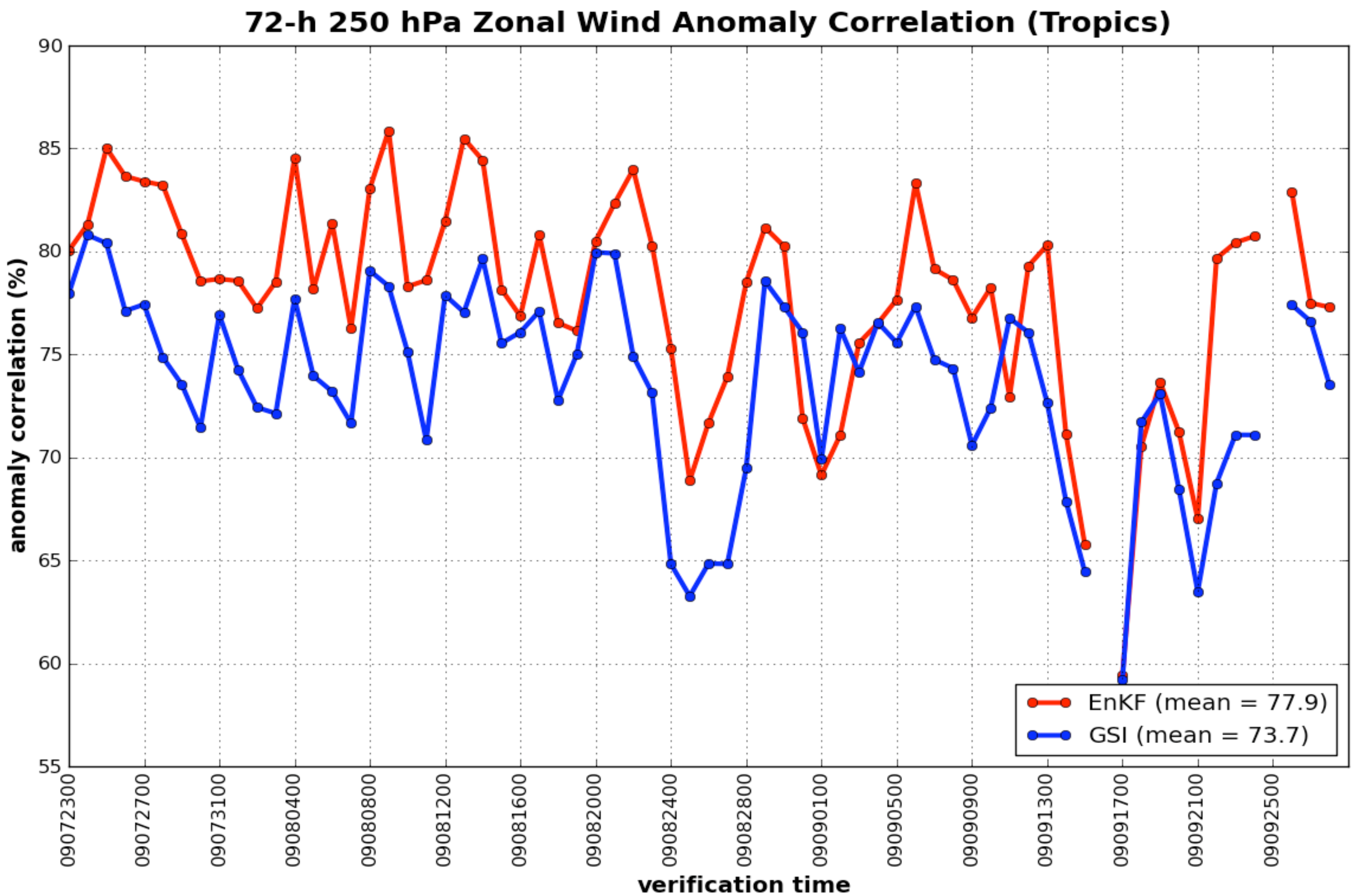


modest impact on track forecast from degrading the resolution of the forecast (still T382 during the data assimilation).



larger impact of GFS/EnKF at T126 vs. operational. However, 2009 operational version had more diffusion, so that complicates analysis.

Tropical winds from parallel tests of deterministic T382 GFS/GSI & GFS/EnKF



EnKF provides consistently better wind forecasts; better steering of tropical cyclones?

Improvements next summer

- Higher resolution, if possible.
- Vortex relocation to ens. mean prior?
- TCVitals obs error estimates incorporated into data assimilation?

	Dvorak	vs.	in-situ	error	(mb)	
TD	TS	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
3.3	6.1	9.5	11.7	13.9	16.1	19.1

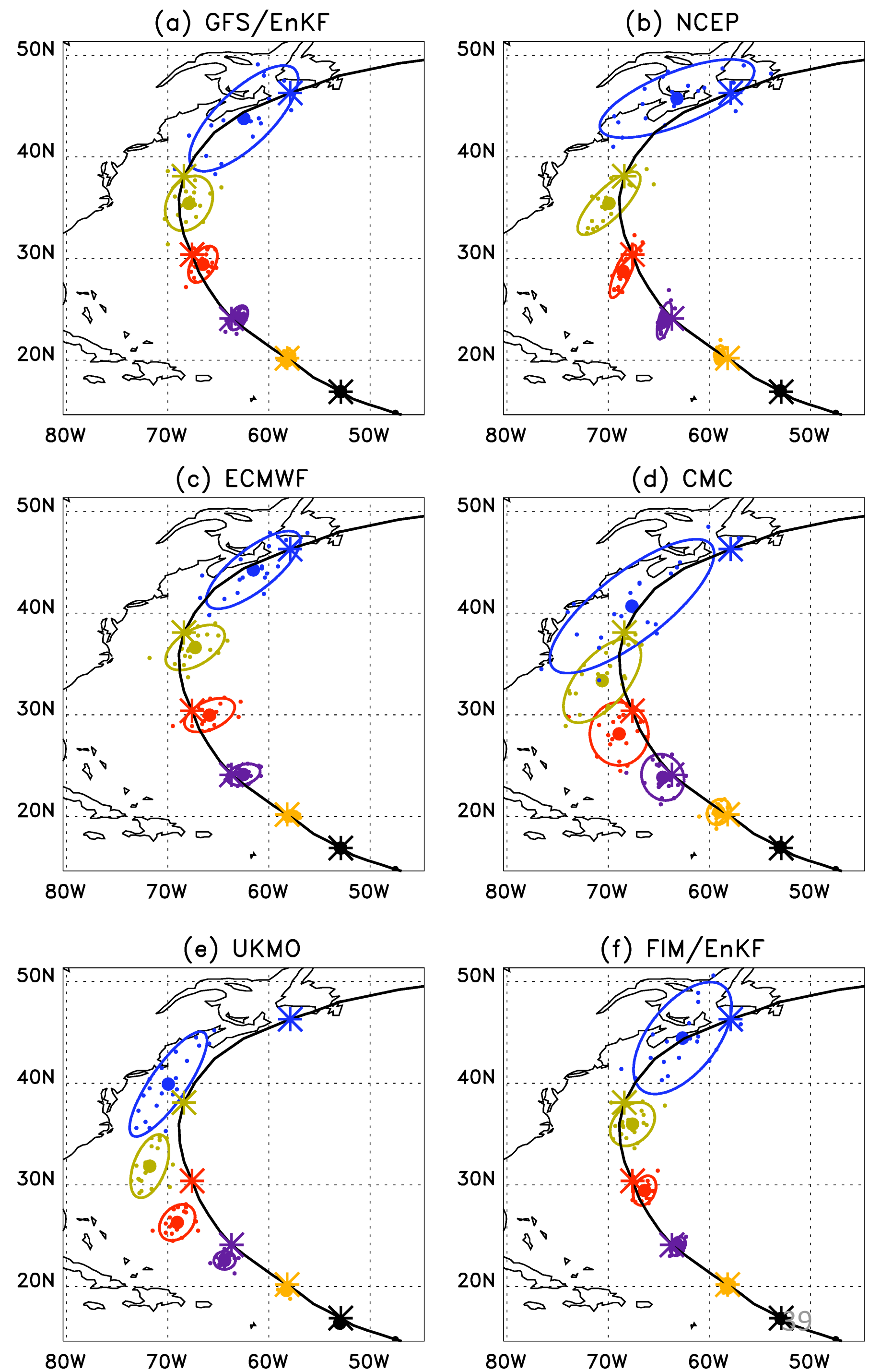
- Also: estimate representativeness errors of TCVitals
- Better treatment of model-error covariances.

* table c/o Ryan Torn, U Albany

Multi-model forecast? Hurricane Bill

Initialized 00 UTC 19 August 2009.

What if we combine the forecasts in some fashion, using their error statistics?



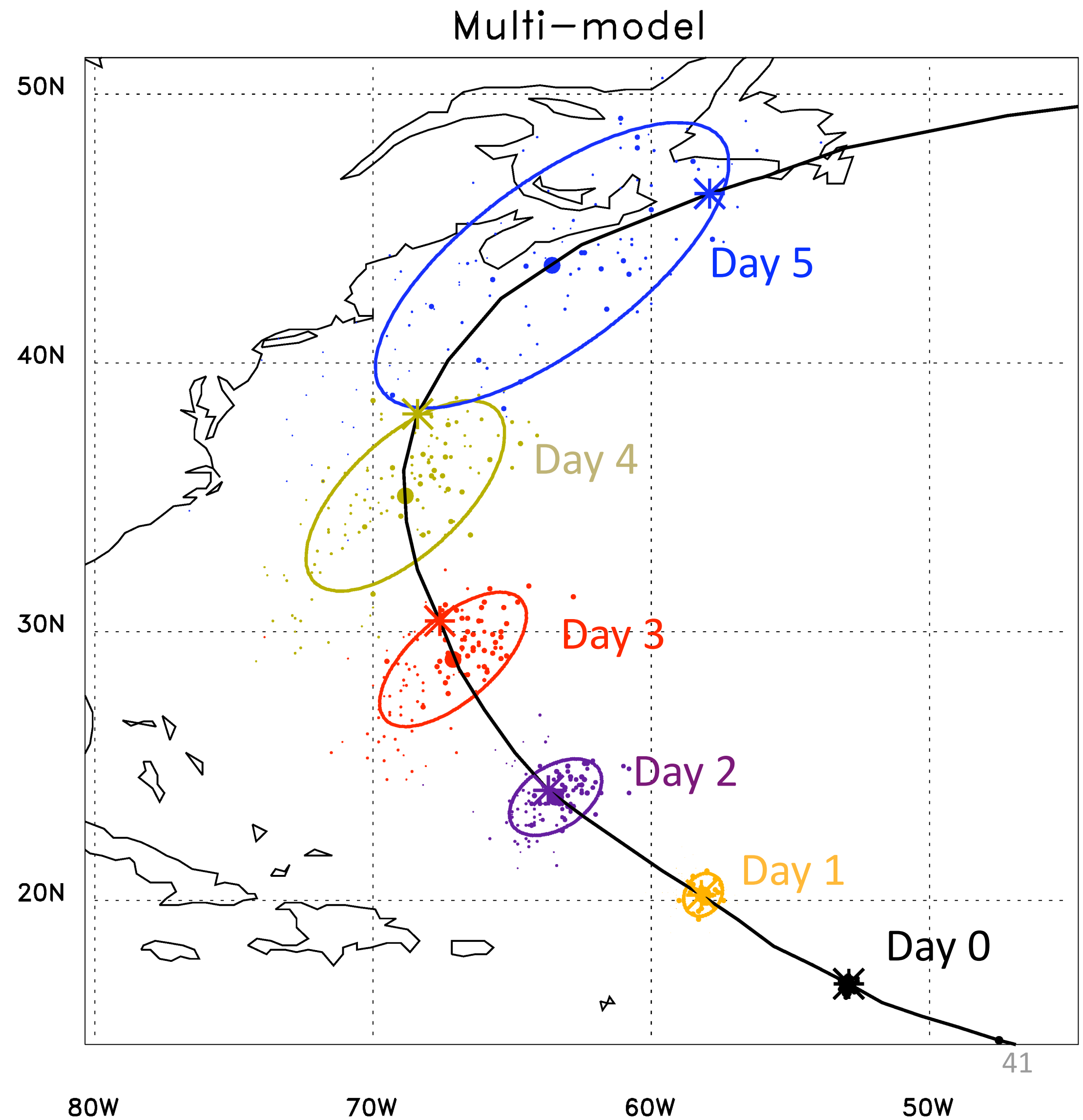
Proposed multi-model technique

- Estimate average absolute error $\sigma_{m,t}$ of the ensemble-mean forecast for a given lead time *t* and forecast model *m*, quasi-cross validated (e.g., when estimating error for Bill, don't use Bill data, but ok to use every other storm).
- Set weights for every available member forecast to be $1/\sigma^2_{m,t}$
- Estimate weighted ensemble mean and weighted ensemble covariance matrix.

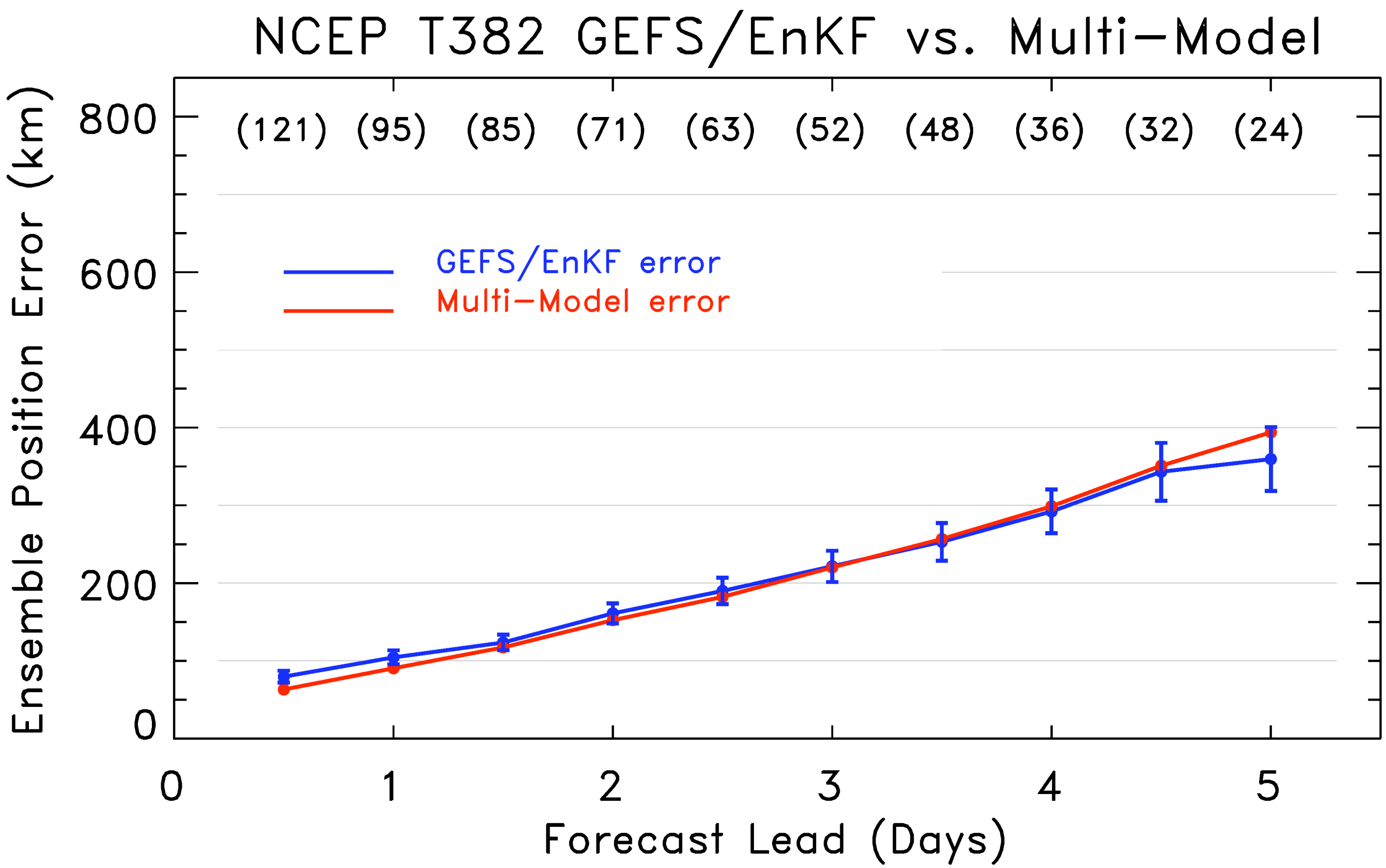
An experimental multi-model product

Dot area is proportional to the weighting applied to that member

• = ens. mean position
* = observed position



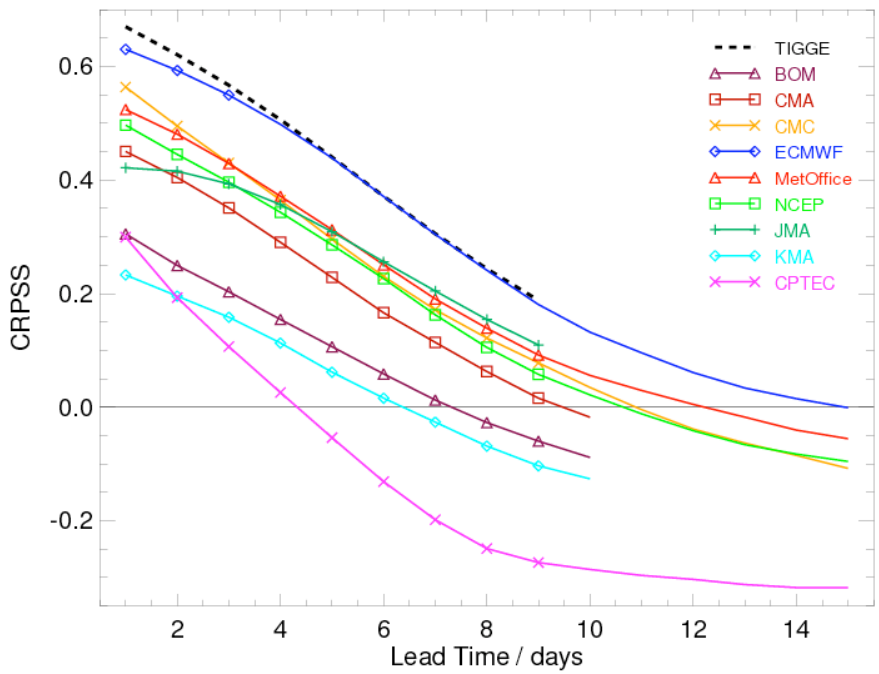
Multi-model error (GEFS/EnKF, ECMWF, FIM, UKMO, CMC, NCEP)



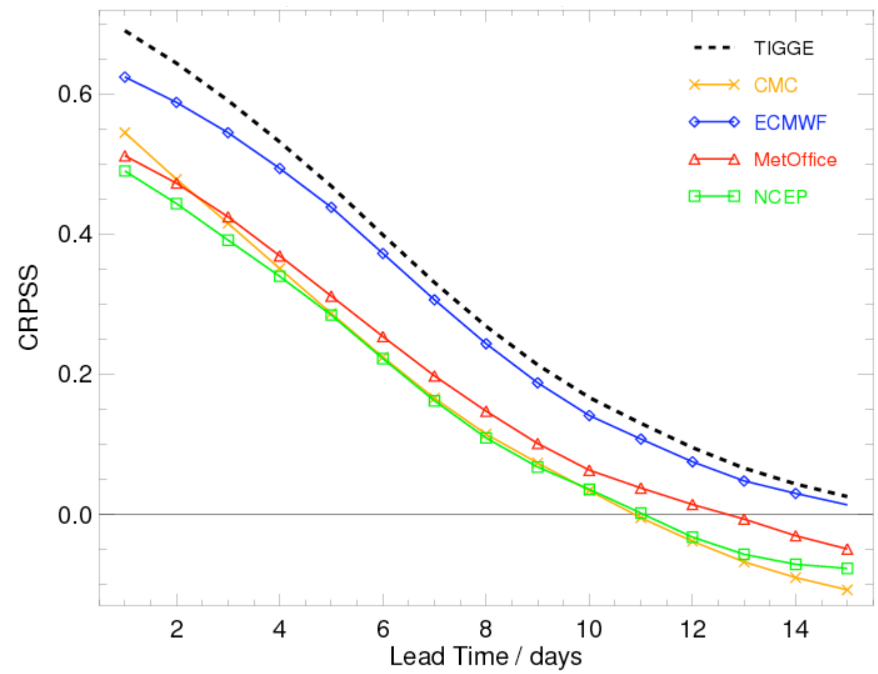
Not much improvement from multi-model. Why?

Multi-model 2-m temperature forecasts

All models (with 30-day bias correction) & multi-model "TIGGE"



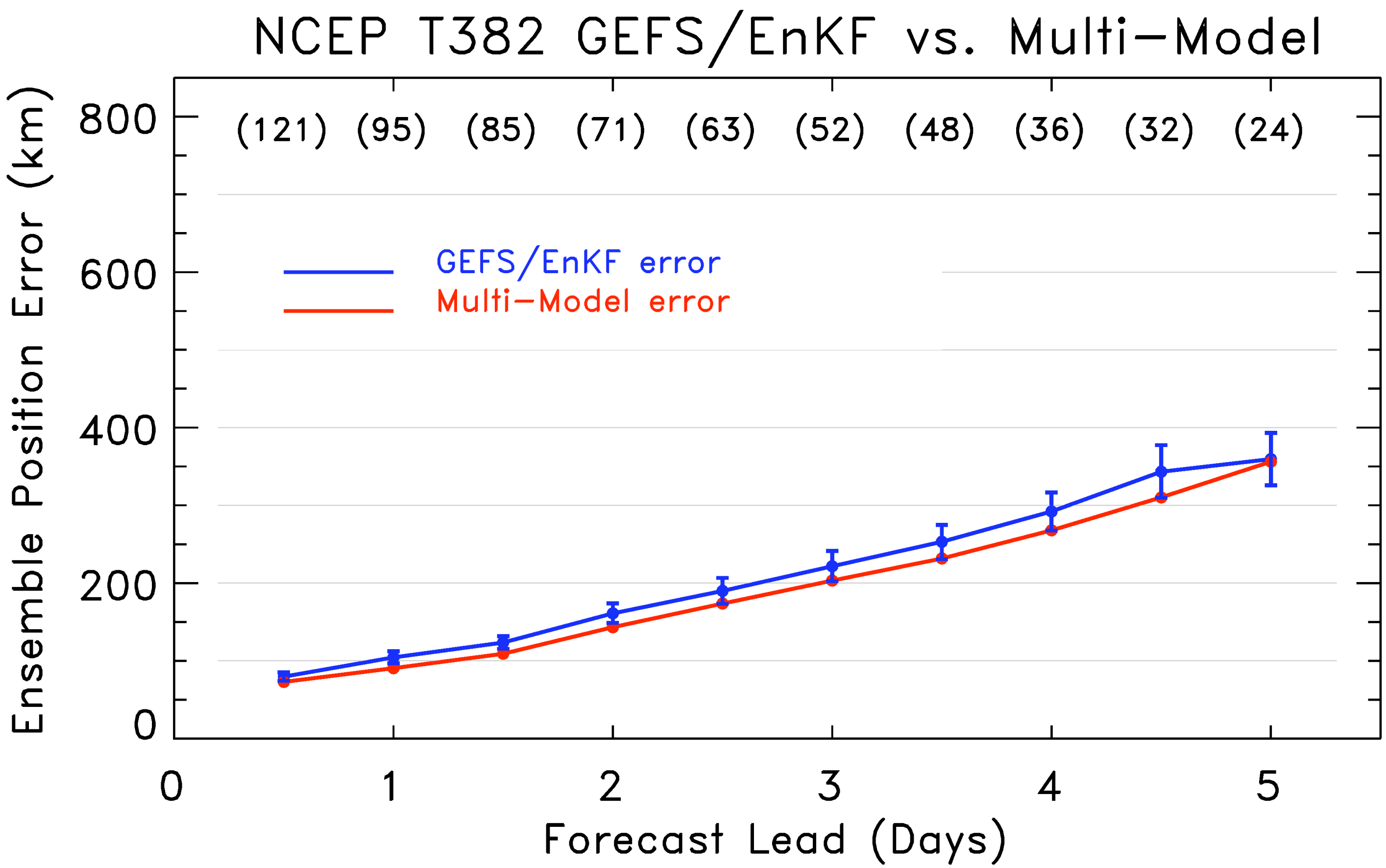
Best 4 models (with 30-day bias correction) & multi-model "TIGGE"



collaborative work led by Renate Hagedorn, ECMWF; conditionally accepted, MWR

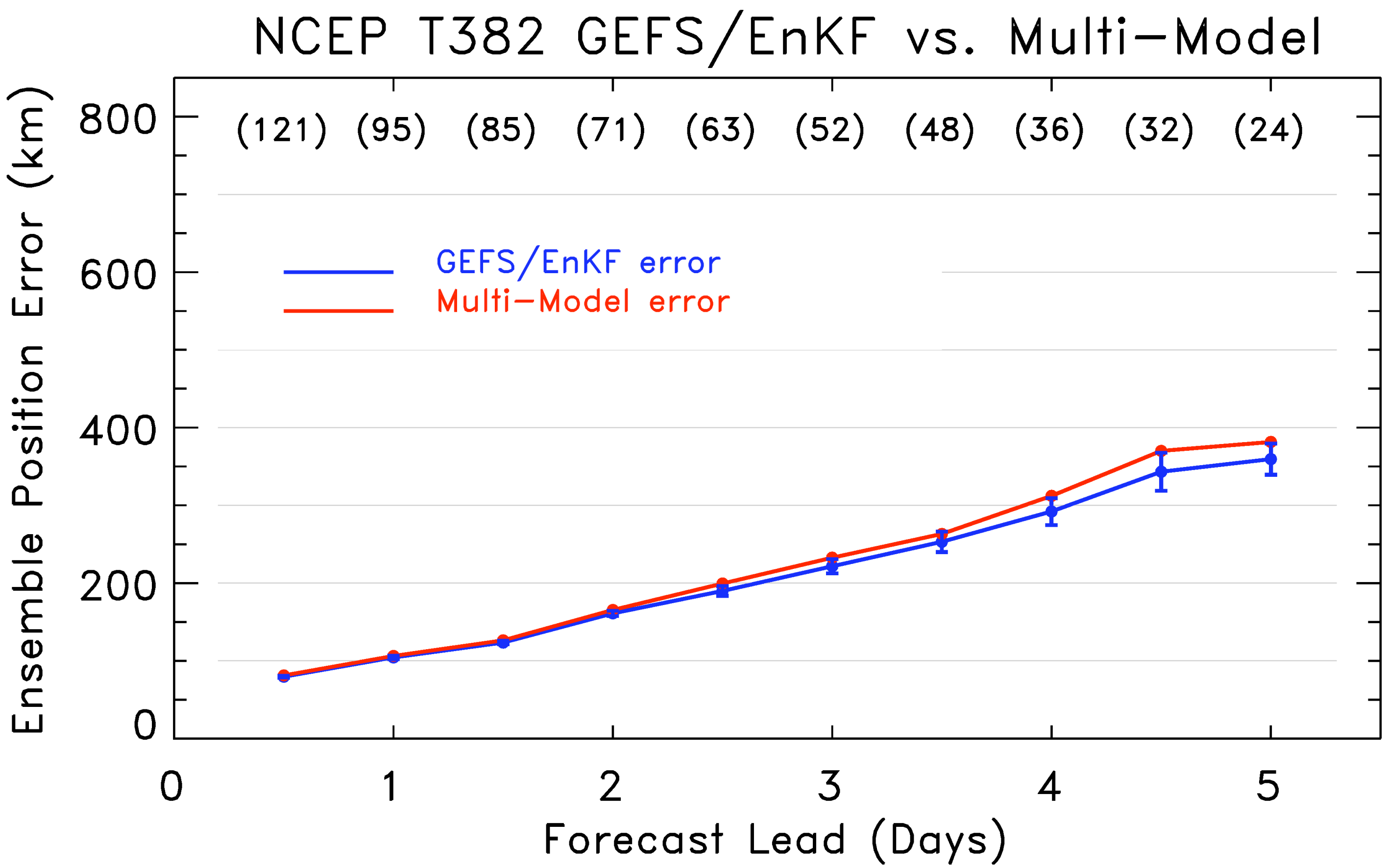
Lesson: discard less skillful models?

Multi-model error (GEFS/EnKF, ECMWF only)



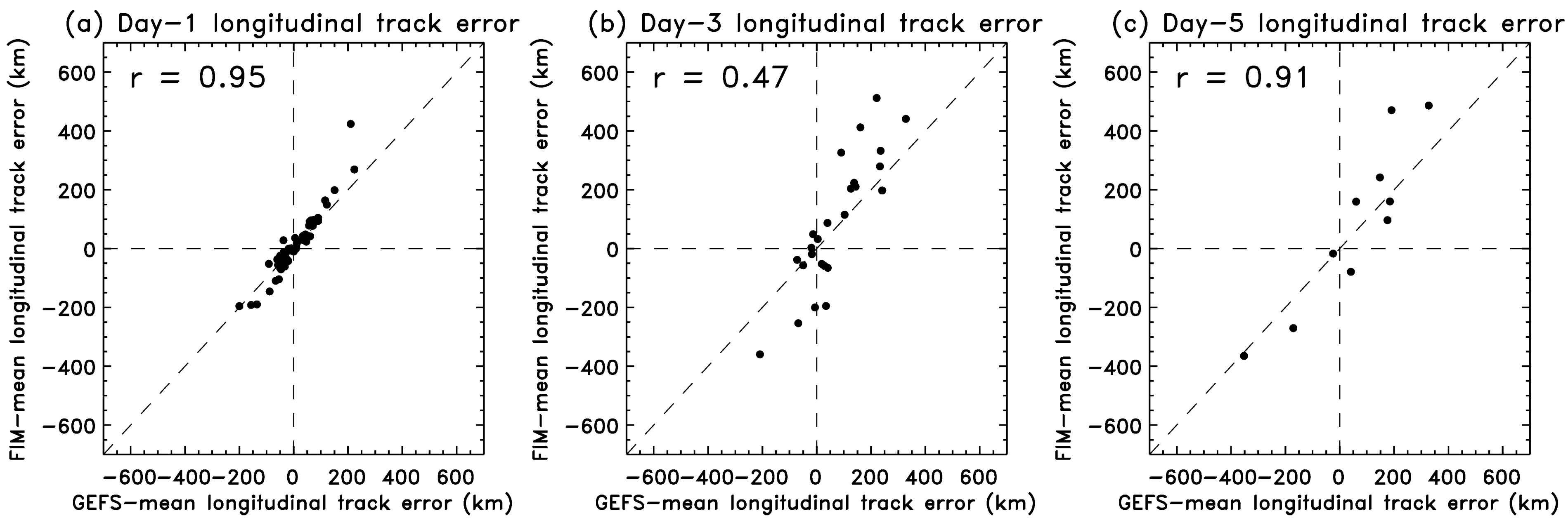
Now some improvement, ~ 6 - 9 hours lead.

Multi-model error (GEFS/EnKF, FIM only)



Degradation relative to GFS/EnKF alone. Why?

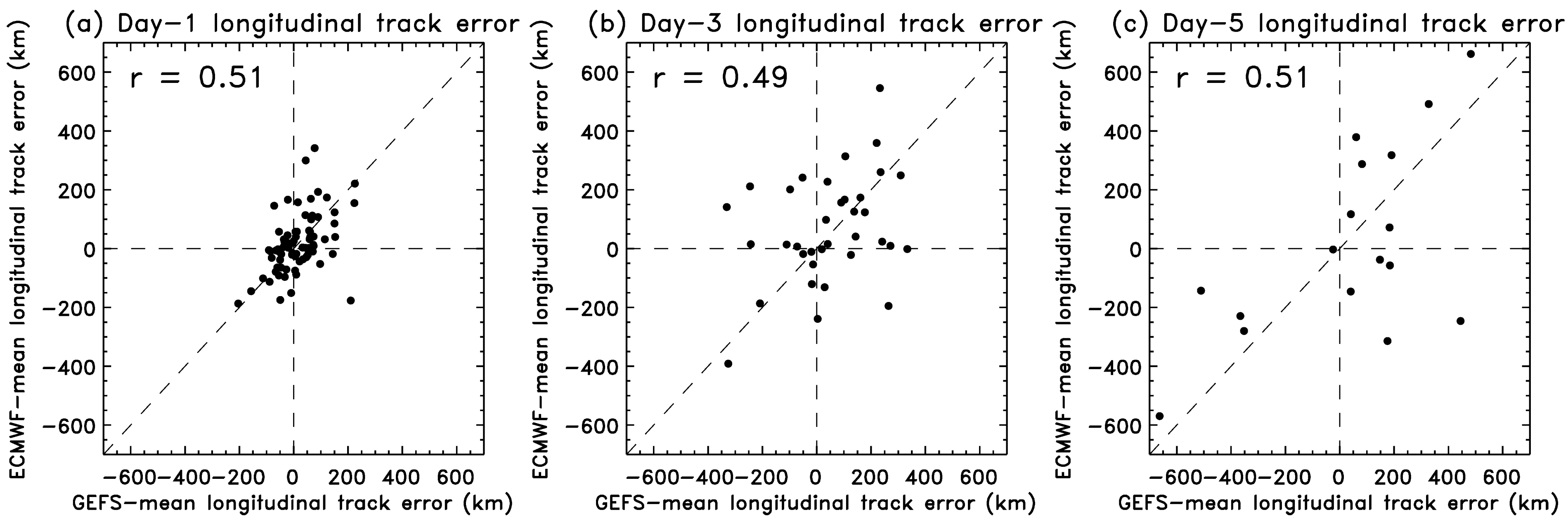
Correlation of errors, GFS/EnKF & FIM



r is the Spearman rank correlation.

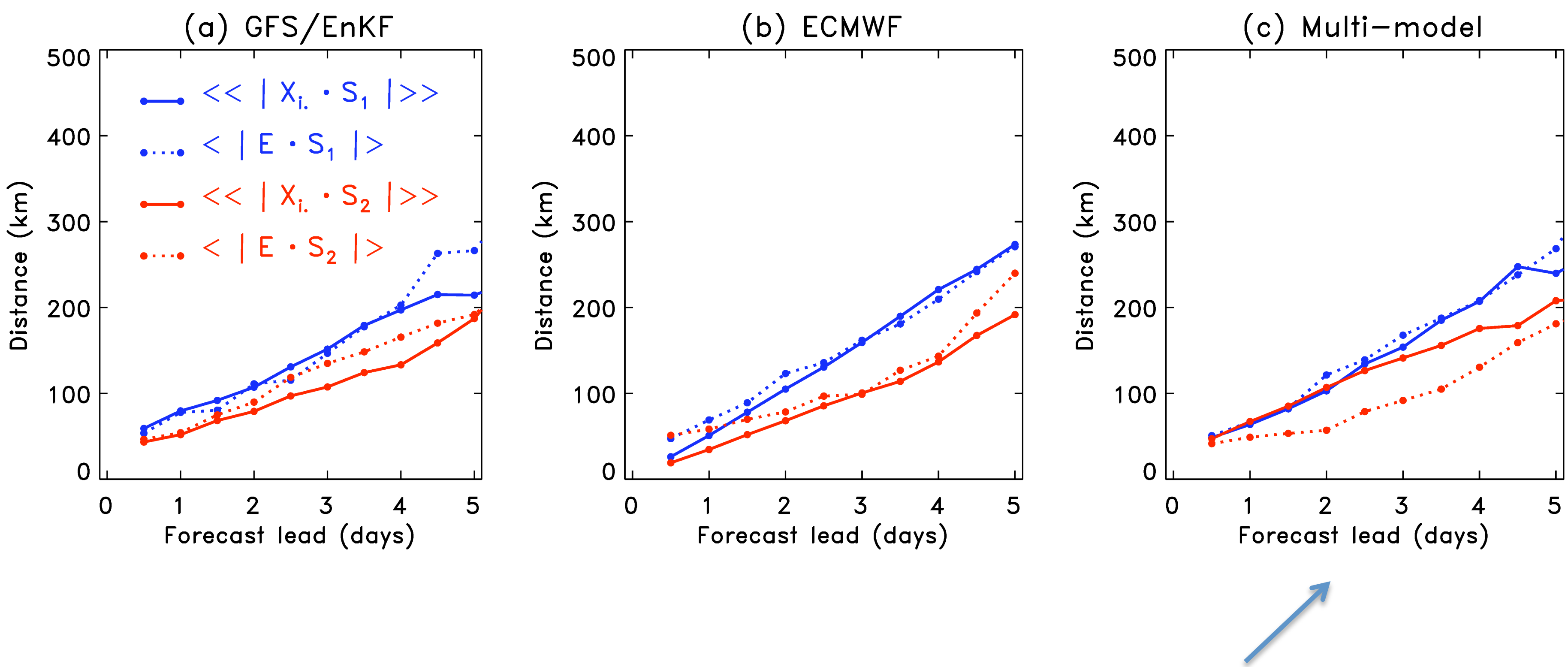
Multi-model forecasts generally predicated on the assumption that models provide independent information. In this case, FIM errors are highly co-linear with GFS/EnKF errors.

Correlation of errors, GFS/EnKF & ECMWF



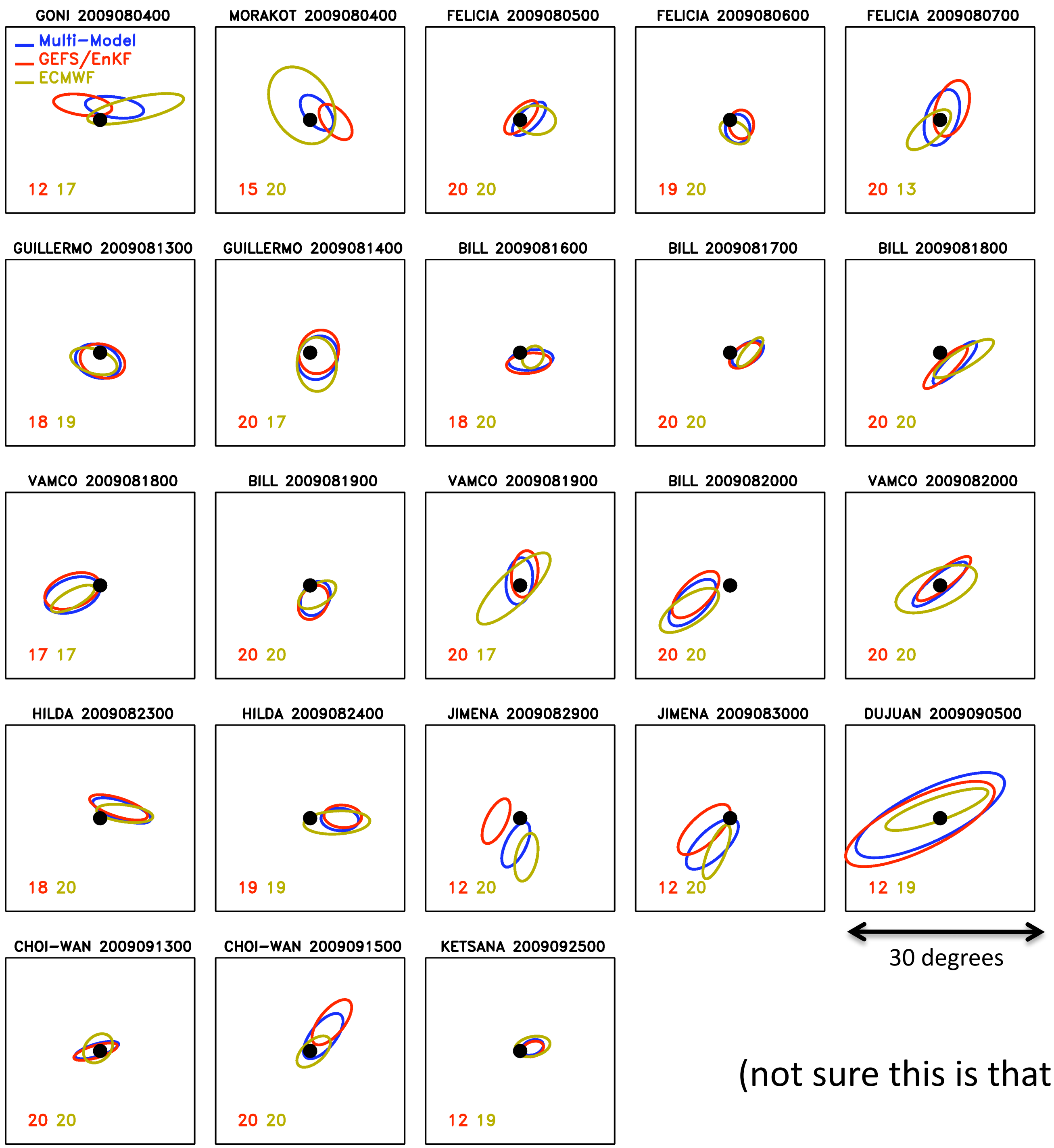
Less co-linearity of forecast errors between GFS/EnKF & ECMWF systems; the greater independence of their forecast errors permits a multi-model improvement.

Multi-model ellipse eccentricity



suggests that multi-model forecast now has a bit too much spread in directions of trailing eigenvectors. Also, the projection of error onto the trailing eigenvector has decreased.

Day-4 Ellipses



Dot: observed position (always at center)

Red oval: GEFS/EnKF

Green oval: ECMWF

Blue oval: Multi-model

Numbers: # members tracked

(not sure this is that illuminating)

Conclusions, part 1

- EnKF + high-resolution global model showed remarkable success in 2009 season
  - track forecasts competitive with state-of-the art ECMWF forecast ensemble.
  - track forecasts clearly better than NCEP, CMC operational, FIM.
  - good consistency between ensemble spread and error.
  - generally better tropical wind analyses.
  - information on ellipticity of track positions useful
- Improvement in TC forecasts likely due to increased model resolution, EnKF, and TCVitals.
  - however, forecast resolution had smaller impact when data assimilation with hi-res EnKF
  - TCVitals had small impact in GSI; parallel tests with/without in EnKF not conducted. Presumed effect larger.

Conclusions, part 2

- Multi-model forecasts provided some benefit to tracks
  - No benefit when using all ensemble systems, including poorer ones
  - Some benefit of just GFS/EnKF + ECMWF
- Some questions
  - improve methods for vortex initialization in EnKF. Incorporate relocation?
  - methods for treating hurricane-related model error?
  - resolution impacts of global model in EnKF?
  - effect of assimilating position and intensity of TCVitals separately?
  - will nesting of high-resolution regional EnKF and SREF forecasts provide even better results?

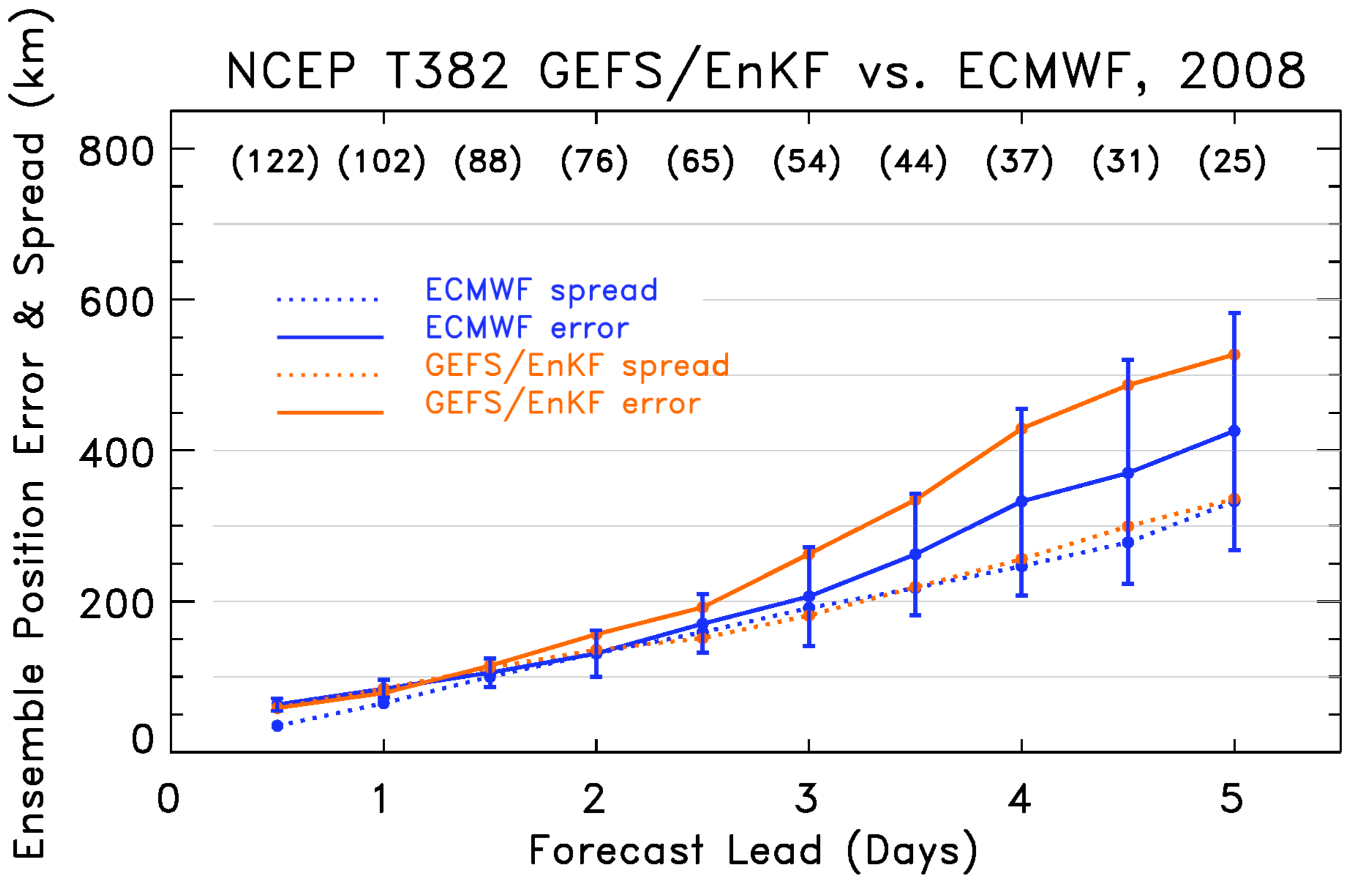
The future

- ESRL and EMC have a signed agreement to work together toward the development of a hybrid XD-Var/EnKF, operational 3 years hence.
- We will continue testing the EnKF with as high a resolution global model as possible through HFIP.
- We partner with AOML scientists to develop a *consistent* nested regional WRF/EnKF system.

Acknowledgments

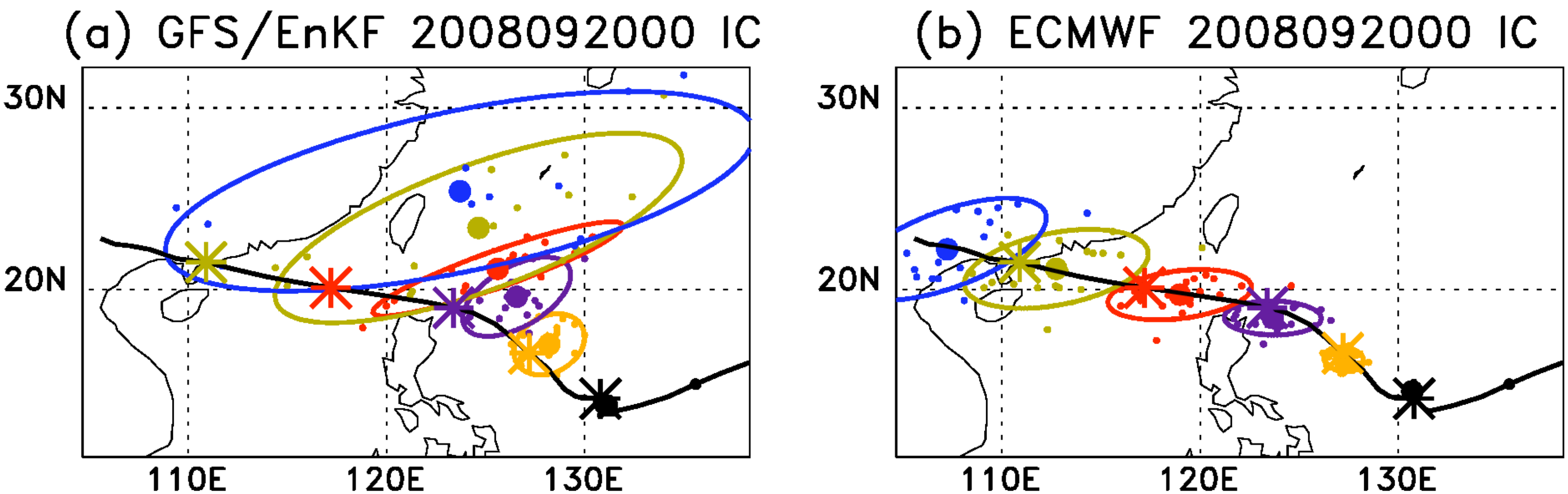
– NOAA HFIP program for 2009 summer test support
– NOAA THORPEX, ESRL base for EnKF development
– U. Texas NSF supercomputer for CPU cycles
– Phil Pegion for programming support, analysis.
– Mike Fiorino (tracker files, software sanity check)
– Other ESRLites, e.g., Chris Harrop for scripting, Stan Benjamin for FIM development and management, Paula McCaslin for display software
– Tim Marchok (tracker files)
– Daryl Kleist (GFS parallel runs)
– Chris Snyder, Jeff Anderson for ensemble bull sessions and algorithmic ideas.
– and you, for your constructive feedback

2008 season, GFS/EnKF vs. ECMWF



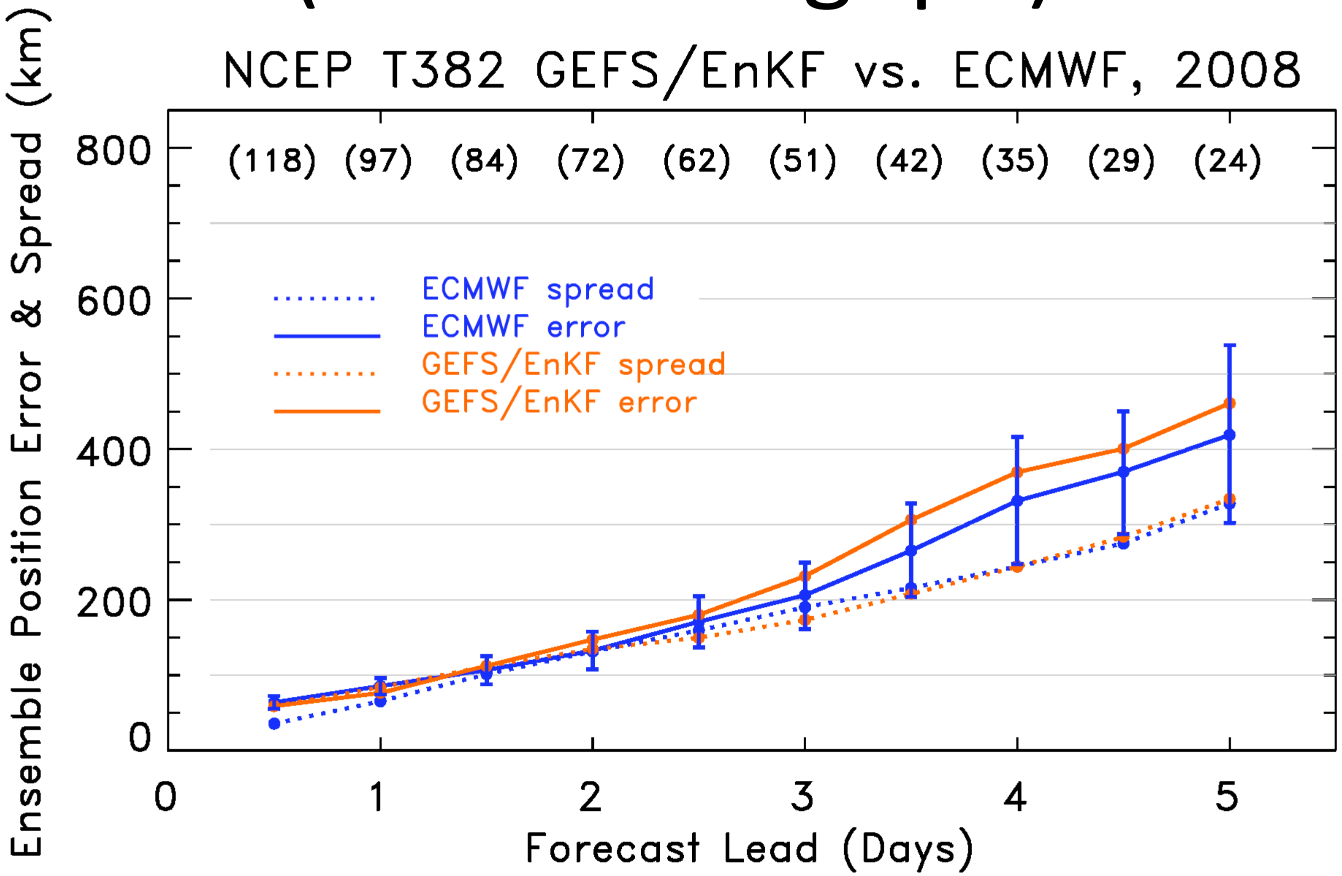
For 2008, EnKF-based system does not compare favorably to ECMWF. Why?

Typhoon Hagupit, 2008 GFS/EnKF bust



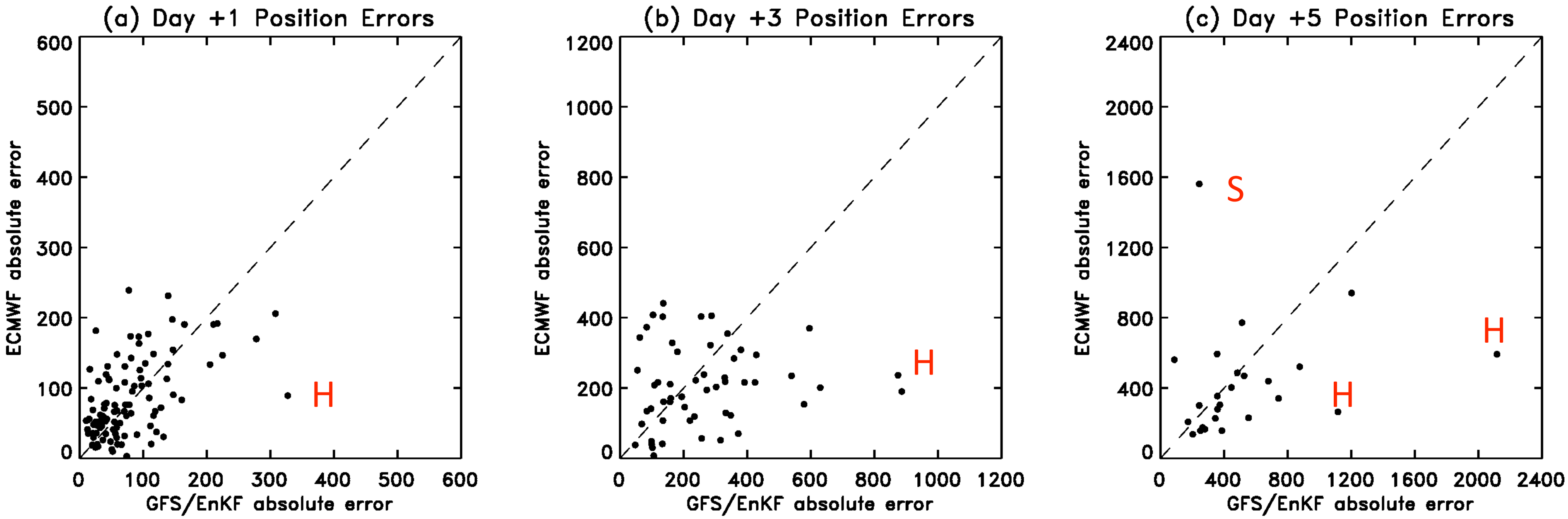
GFS statistics are strongly affected by two samples from Typhoon Hagupit. Here's one.

2008 season, GFS/EnKF vs. ECMWF (without Hagupit)



Now GFS/EnKF is more similar in performance, though in 2008 both systems exhibit a lack of spread.

GFS/EnKF vs. ECMWF track errors



H = Hagupit, S = Sinlaku

Canonical EnKF update equations (for time *t*)

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}\left(\mathbf{y}_i - H\mathbf{x}_i^b\right)$$

$$\mathbf{K} = \mathbf{P}^b H^{\mathrm{T}}\left(H\mathbf{P}^b H^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

$$\mathbf{P}^b = \mathbf{X}\mathbf{X}^{\mathrm{T}}$$

$$\mathbf{X} = \left(\mathbf{x}_1^b - \overline{\mathbf{x}^b}\ , \ldots , \mathbf{x}_n^b - \overline{\mathbf{x}^b}\right)$$

$$\mathbf{y}_i = \mathbf{y} + \mathbf{y}_i^{'}$$

$$\mathbf{y}_i^{'} \sim N(0, \mathbf{R})$$

Notes: (1) An ensemble of *n* parallel data assimilation cycles is conducted, assimilating *perturbed observations* .

(2) Background-error covariances are estimated using the ensemble.

(back)

Propagation of state and error covariances in EnKF

$$\mathbf{P}^a(t) = \left\langle \left[\mathbf{x}_i^a(t) - \overline{\mathbf{x}}_i^a(t)\right]\left[\mathbf{x}_i^a(t) - \overline{\mathbf{x}}_i^a(t)\right]^T \right\rangle$$

(**P**^{a} never explicitly formed)

$$\mathbf{x}_i^b(t+1) = M\mathbf{x}_i^a(t)$$

if forecast model is "perfect"; *M* is forward model operator

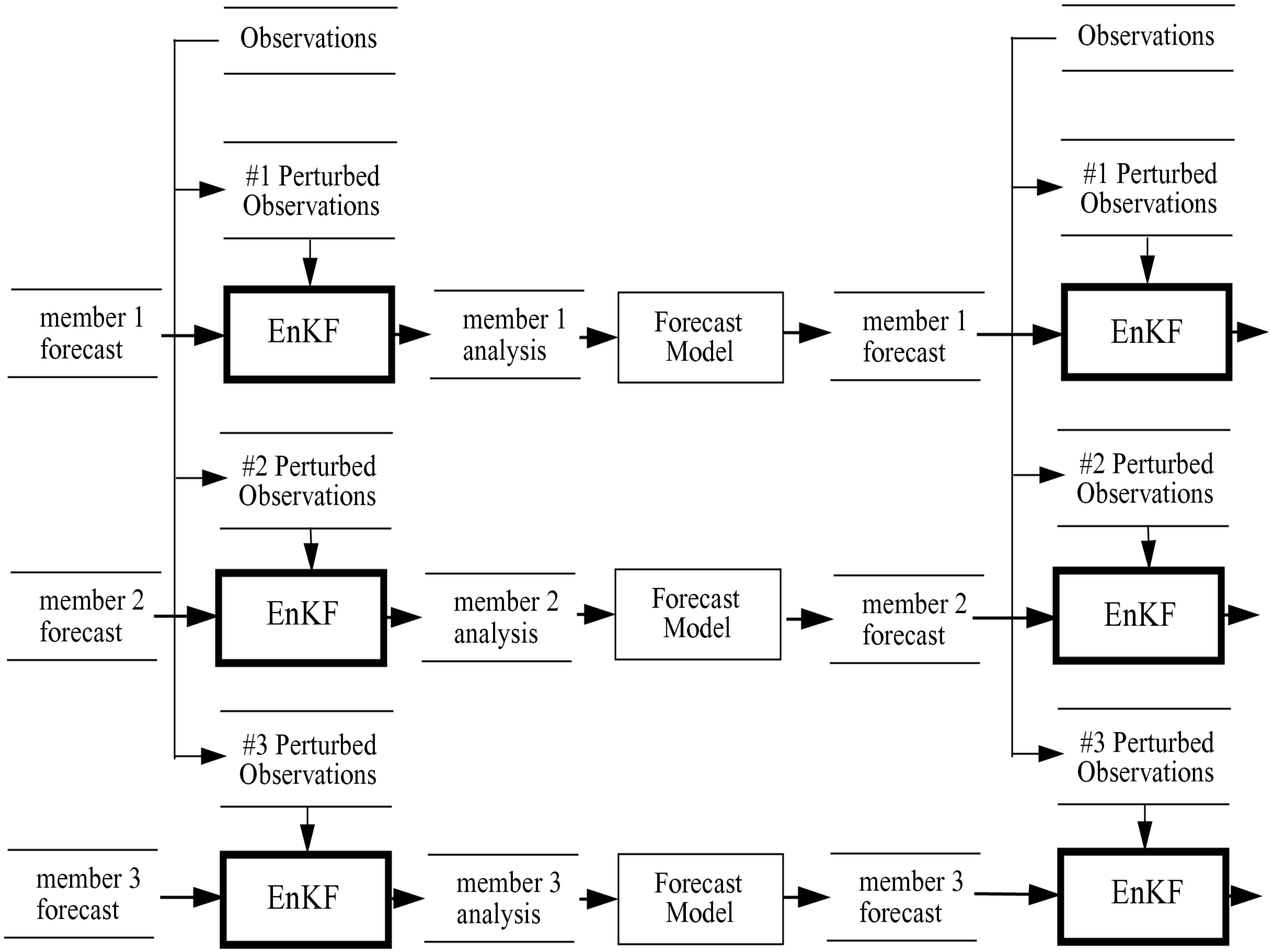
- or -

$$\mathbf{x}_i^b(t+1) = M\mathbf{x}_i^a(t) + \eta_i$$

$$\left\langle \eta_i \eta_i^T \right\rangle = \mathbf{Q}$$

…or something similar,
if forecast model imperfect.

Perfect-model EnKF schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)