Practical applications of hybrid variational-ensemble data assimilation approaches

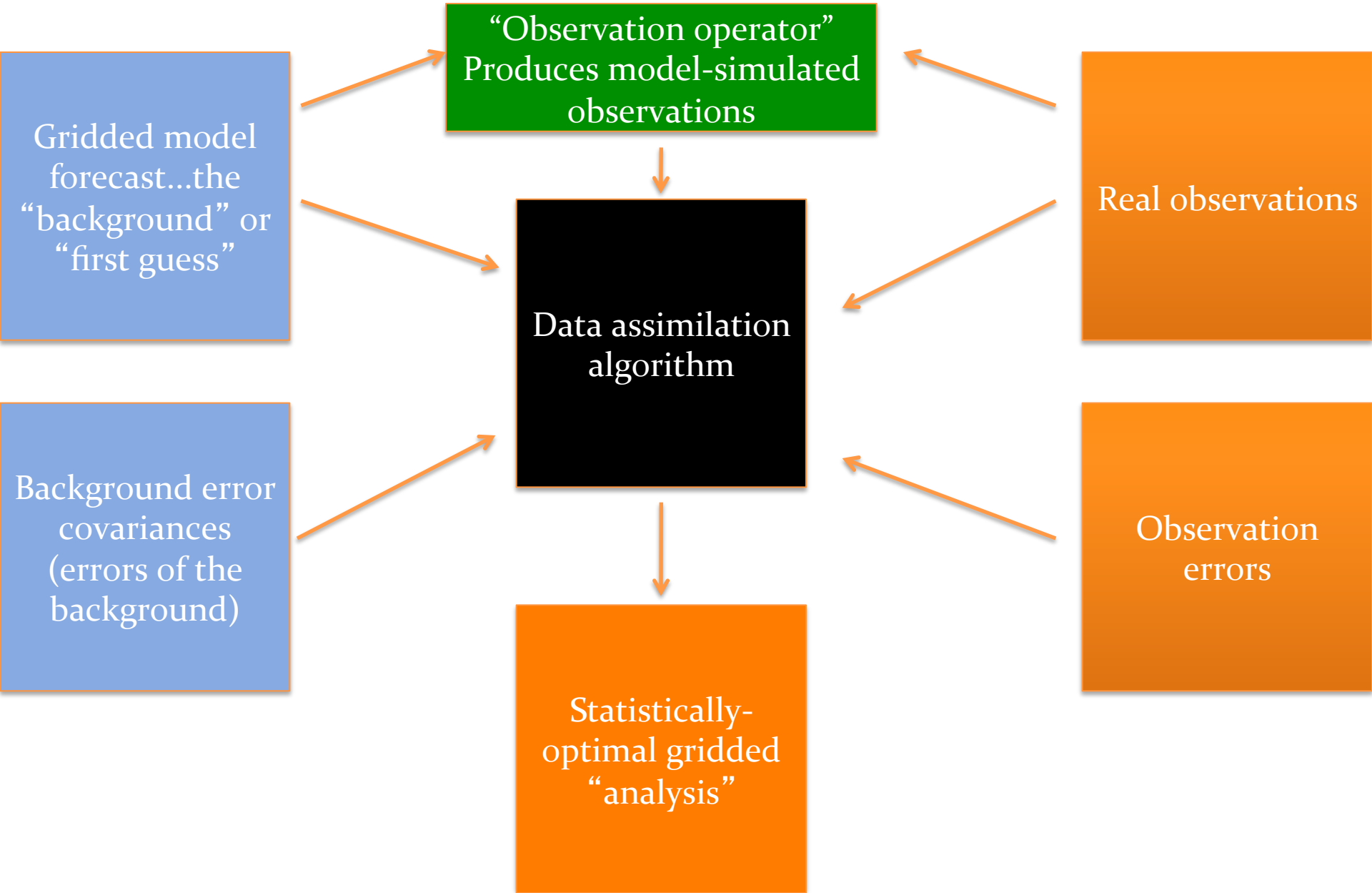
Craig Schwartz

(schwartz@ucar.edu)

NCAR/MMM

August 4, 2015

What is data assimilation?



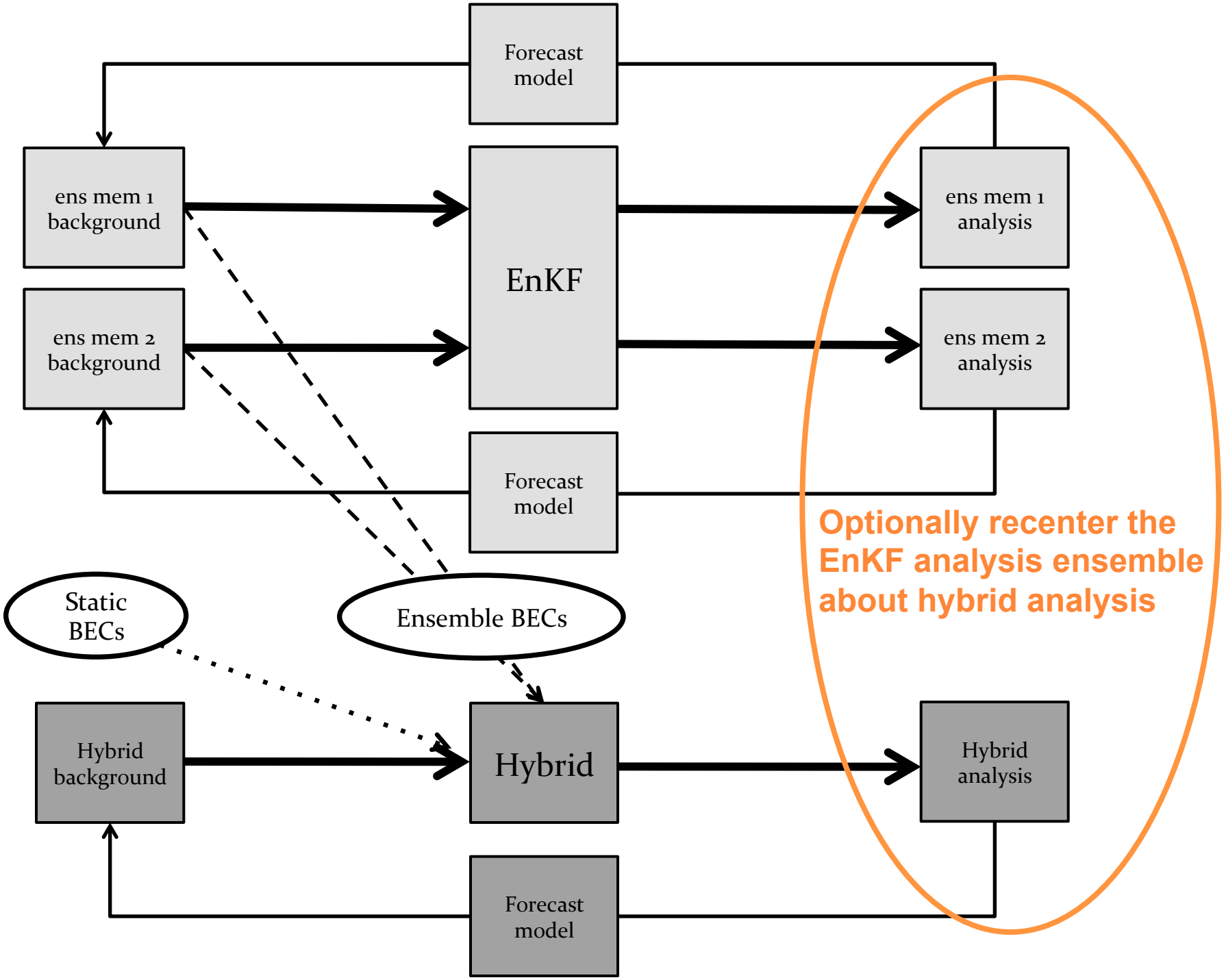
A few data assimilation approaches

- Three-dimensional variational (3DVAR)
  - Background error covariances (BECs) typically fixed/ time-invariant
  - May yield poor results when actual flow differs from that encapsulated within the fixed "climatology"
  - Produces a deterministic analysis

- Ensemble Kalman filter (EnKF)
  - Time-evolving, "flow-dependent" BECs estimated from short-term ensemble forecasts
  - Produces an ensemble analysis

A few data assimilation approaches

- "Hybrid" variational/ensemble
  - Incorporates ensemble background errors within a variational (e.g., 3DVAR) framework
  - Combination of fixed and time-evolving BECs
  - Traditionally produces a deterministic analysis like (3DVAR)


Forecast model
ens mem 1 background
EnKF
ens mem 1 analysis
ens mem 2 background
ens mem 2 analysis
Forecast model
Optionally recenter the EnKF analysis ensemble about hybrid analysis
Static BECs
Ensemble BECs
Hybrid background
Hybrid
Hybrid analysis
Forecast model

Dual-resolution hybrid DA with a focus on limited-area applications

Dual-resolution hybrid DA

- The hybrid method has repeatedly been shown to initialize better forecasts than 3DVAR
  - Also usually comparable to or better than pure EnKFs

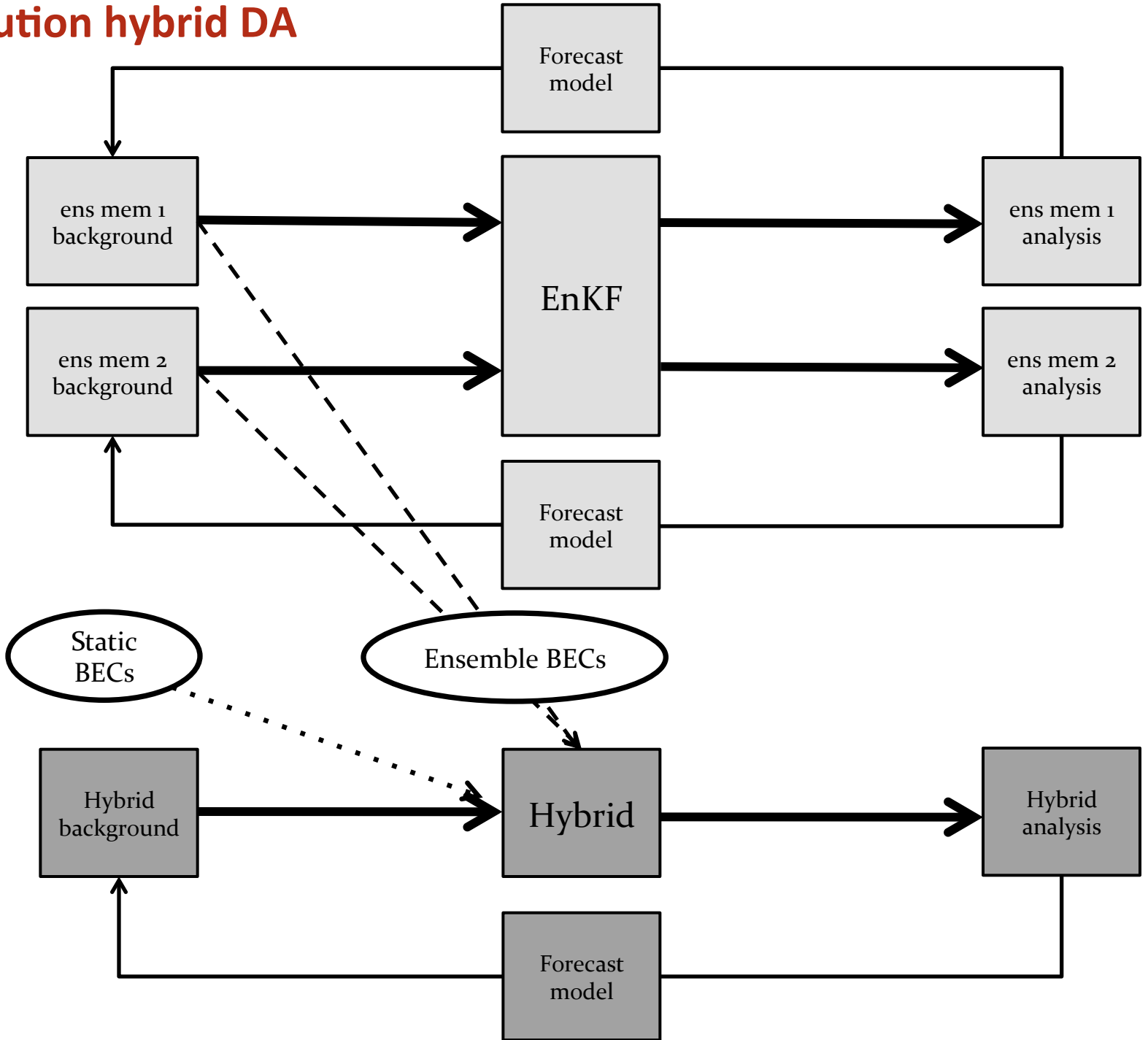
- Compared to 3DVAR, the primary expense of the hybrid is the ensemble of forecasts
  - Similar expense as pure EnKF

Goal of dual-resolution hybrid DA

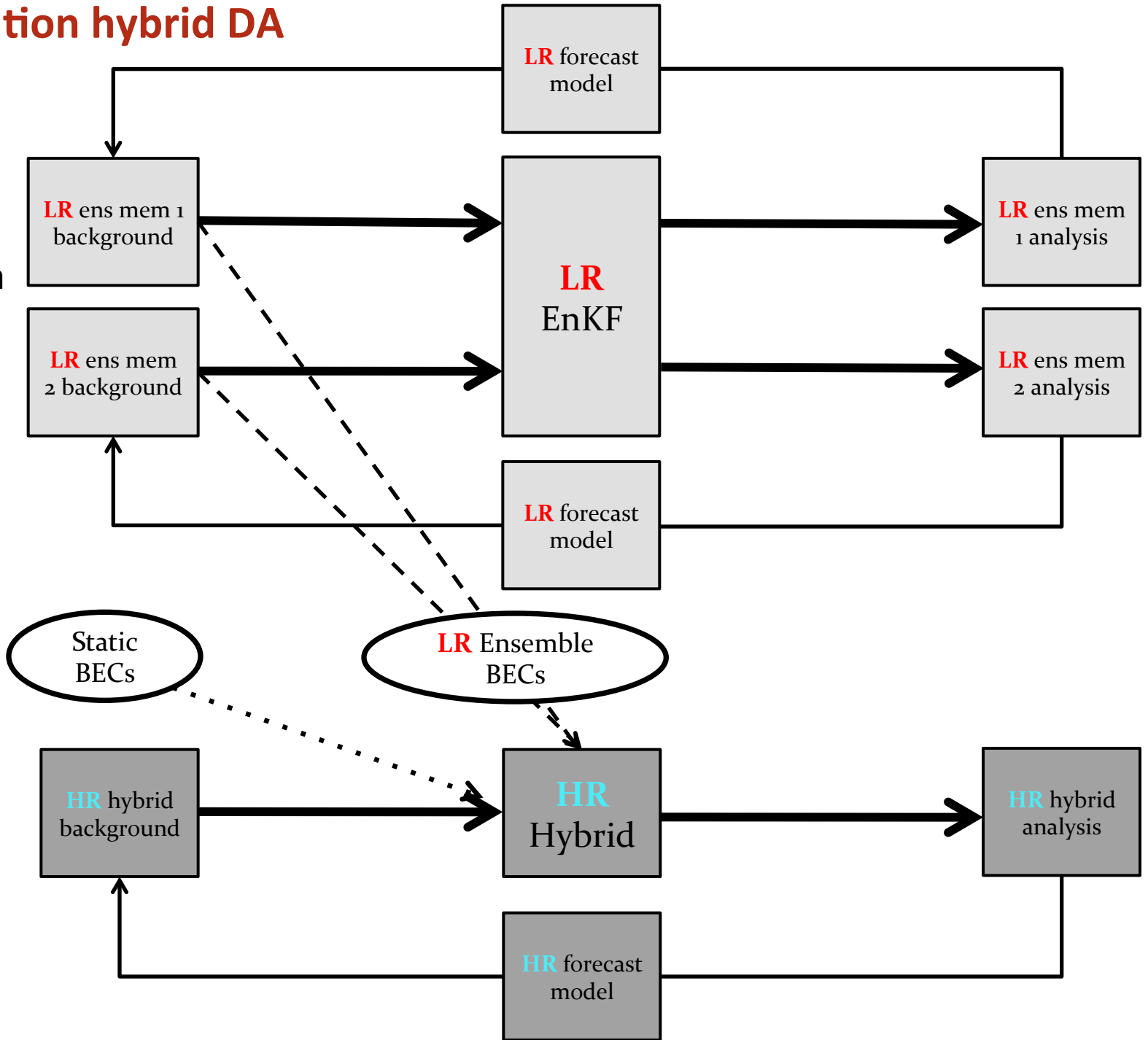
- We wish to perform high-resolution (HR) hybrid analyses using ensemble perturbations from a low-resolution (LR) ensemble
  - The ensemble *perturbations* provide flow-dependent BECs
- This approach considerably reduces computational expense since producing HR ensemble forecasts is unnecessary
  - The analyses are also faster
- Called "dual-resolution" (DR) hybrid DA

Single-resolution hybrid DA

Both ensemble and deterministic "circuits" have identical resolutions



Dual-resolution hybrid DA



Dual-resolution hybrid DA

- Practical aspect: Saving computational resources
- Scientific question: Does resolution of the ensemble perturbations matter?
- To examine these questions, several DA experiments were performed (Schwartz et al. 2015; MWR)

Experiments assessing utility of DR hybrid

- Produced 45- and 15-km analyses, but focus is on 15-km

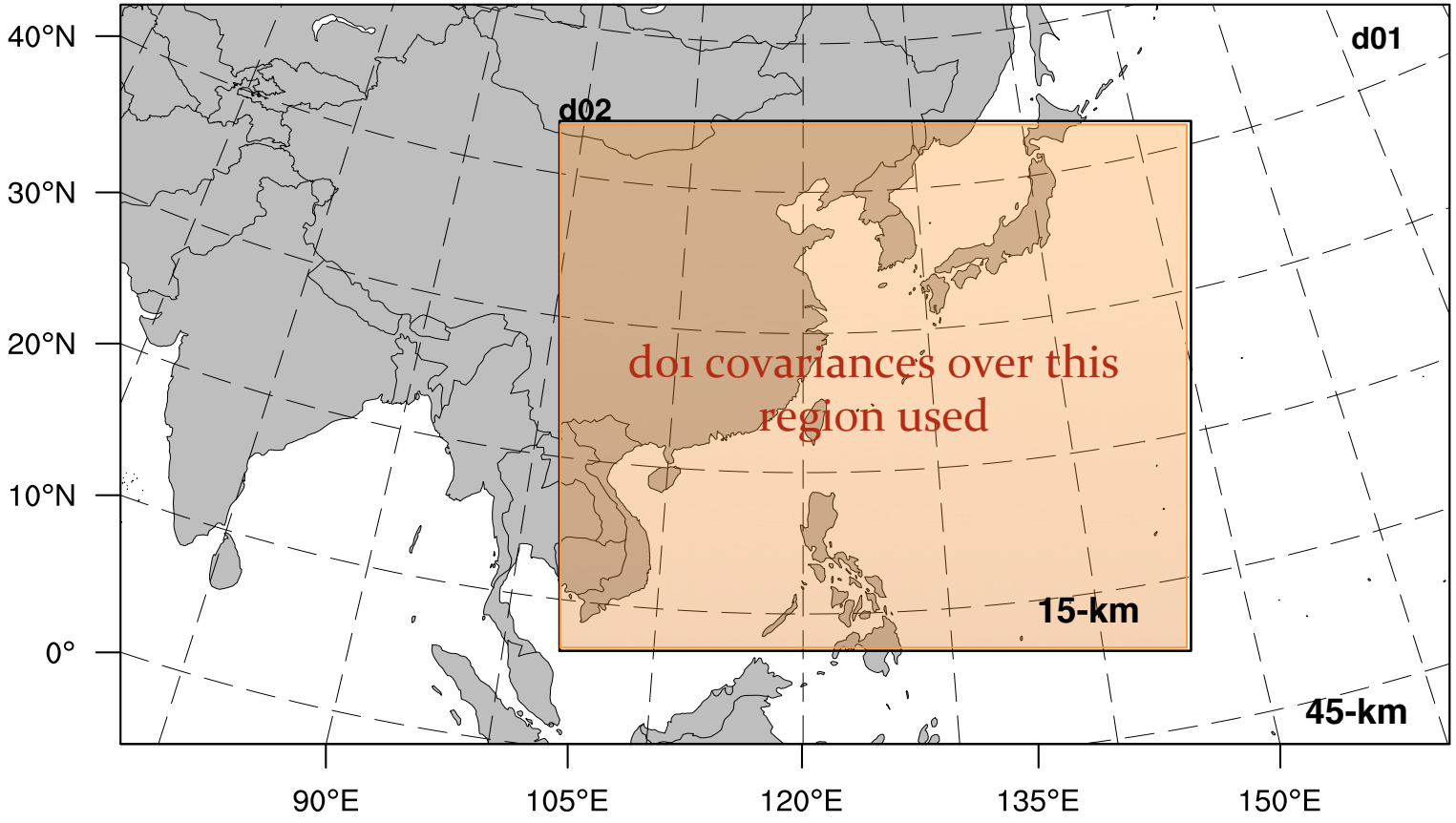
- Four continuously-cycling experiments:
  1) Single-resolution (SR) hybrid analyses at 15-km grid length:
     a *15-km* ensemble was required
  2) Dual-resolution (DR) hybrid analyses at 15-km grid length:
     a *45-km* ensemble was required
  3) Same as #2, but with EnKF re-centering
  4) Pure 3DVAR analyses at 15-km grid length (control run)

Some model configurations

- Version 3.3.1 of the WRF-ARW model
- 45 vertical levels, 30 hPa top
- Lateral boundary conditions from GFS forecasts
- Full suite of physical parameterizations

Nested WRF model configuration

- For DR analyses, produce an analysis on d02 (15-km) using ensemble BECs from d01 (45-km)



More settings

- An EnKF updated a 32-member WRF ensemble
  - Ran both 45- and 15-km cycling EnKF systems
  - 15-km ensembles provided input to 15-km *single-resolution* hybrid analyses
  - 45-km ensembles provided input to 15-km *dual-resolution* hybrid analyses
- Cyclic data assimilation (6-hr period) from 4-28 September 2008
- 72-hr nested WRF model forecasts initialized from hybrid and 3DVAR analyses beginning 1800 UTC 8 September—*focus on the 15-km forecasts*

EnKF assimilation parameters

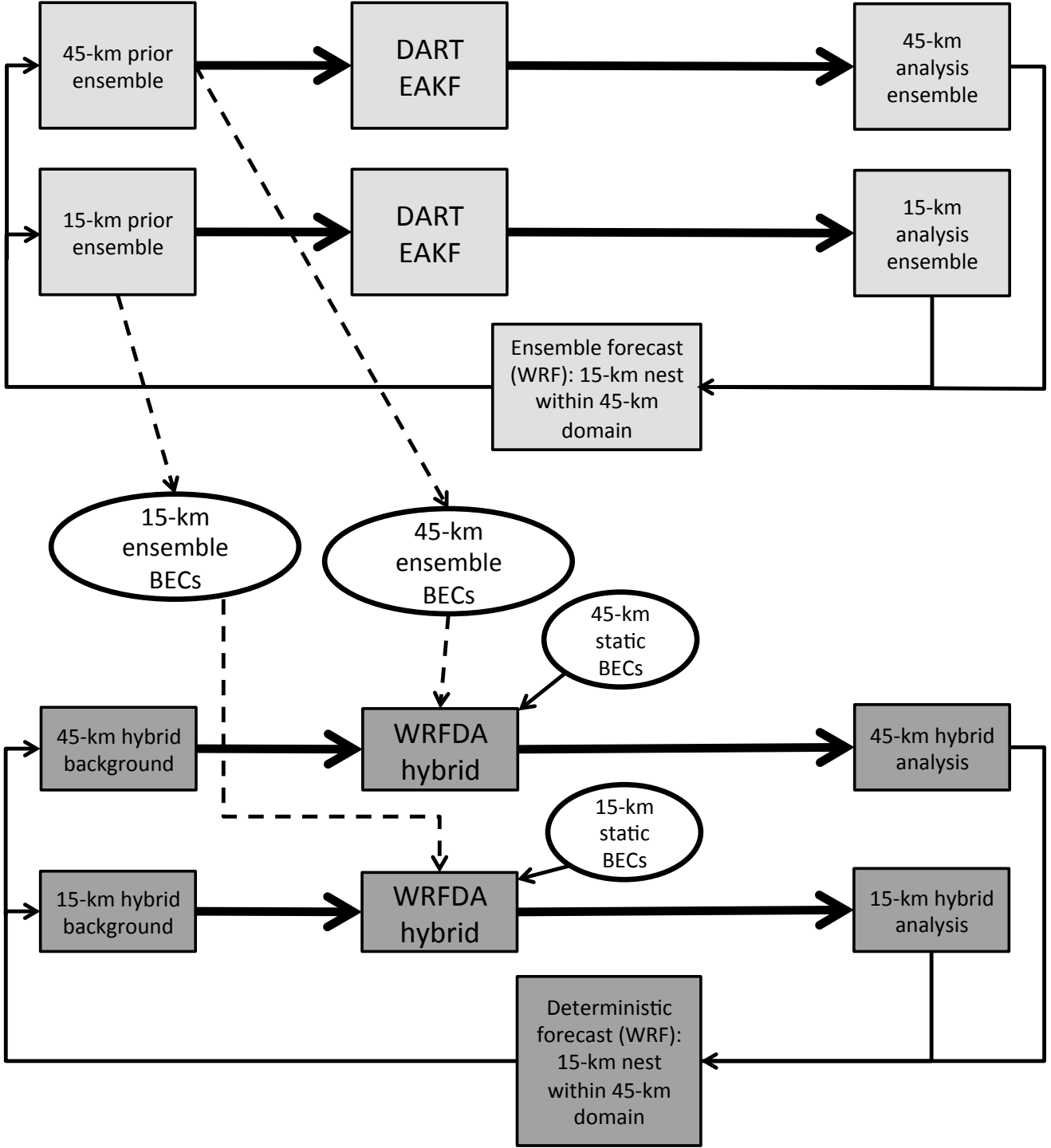
- The Data Assimilation Research Testbed (DART)
  - Ensemble Adjustment Kalman Filter (EAKF)
  - 32 ensemble members
  - Horizontal localization cutoff: ~1280-km from observations
  - Vertical localization cutoff: ~10-km from observations
  - Prior adaptive inflation

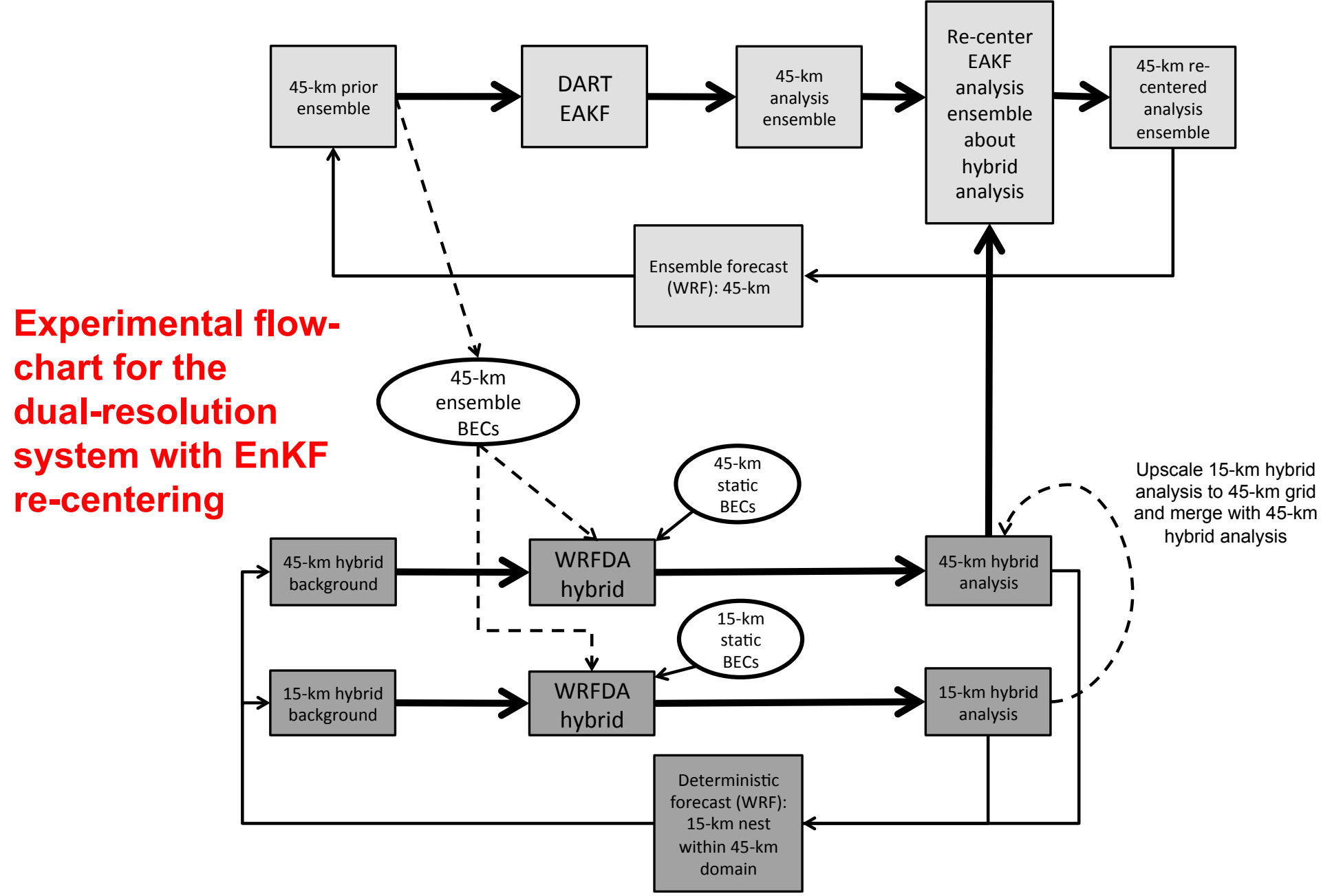
- Stochastic kinetic-energy backscatter scheme within WRF was used during model advances
  - Helped with ensemble spread
- Perturbed lateral boundary conditions for ensemble of model advances

Hybrid assimilation parameters

- Used the WRFDA hybrid system
- 75% ensemble, 25% climatological contributions to total BECs
- Horizontal localization: Similar as EnKF
- Vertical localization: Length-scale increased with height
- Same inflation as in DART applied to the prior perturbations before they were ingested into the hybrid

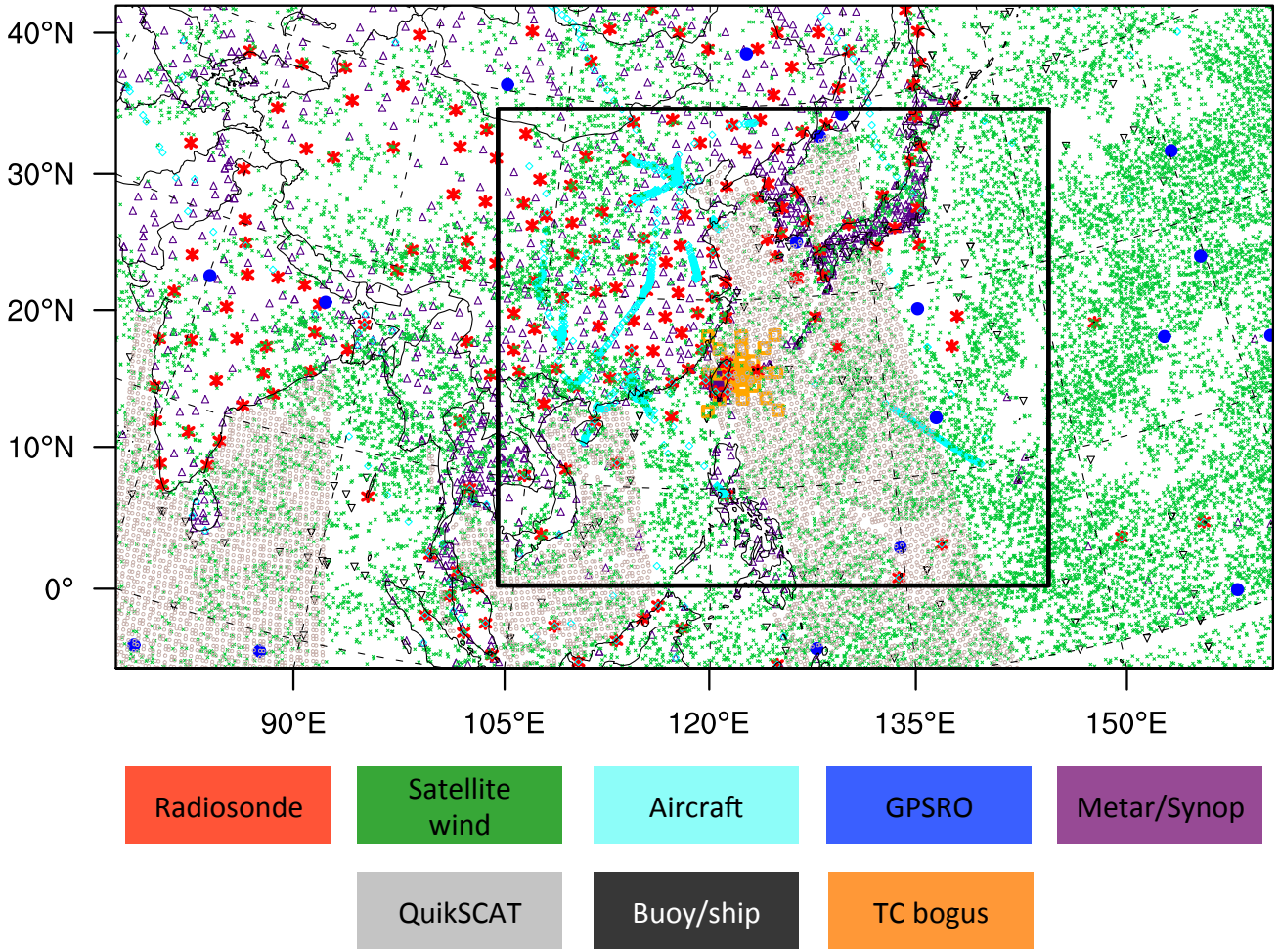
Experimental flow-chart for the single-resolution system




Experimental flow-chart for the dual-resolution system with EnKF re-centering
45-km prior ensemble
DART EAKF
45-km analysis ensemble
Re-center EAKF analysis ensemble about hybrid analysis
45-km re-centered analysis ensemble
Ensemble forecast (WRF): 45-km
45-km ensemble BECs
45-km static BECs
45-km hybrid background
WRFDA hybrid
45-km hybrid analysis
Upscale 15-km hybrid analysis to 45-km grid and merge with 45-km hybrid analysis
15-km static BECs
15-km hybrid background
WRFDA hybrid
15-km hybrid analysis
Deterministic forecast (WRF): 15-km nest within 45-km domain

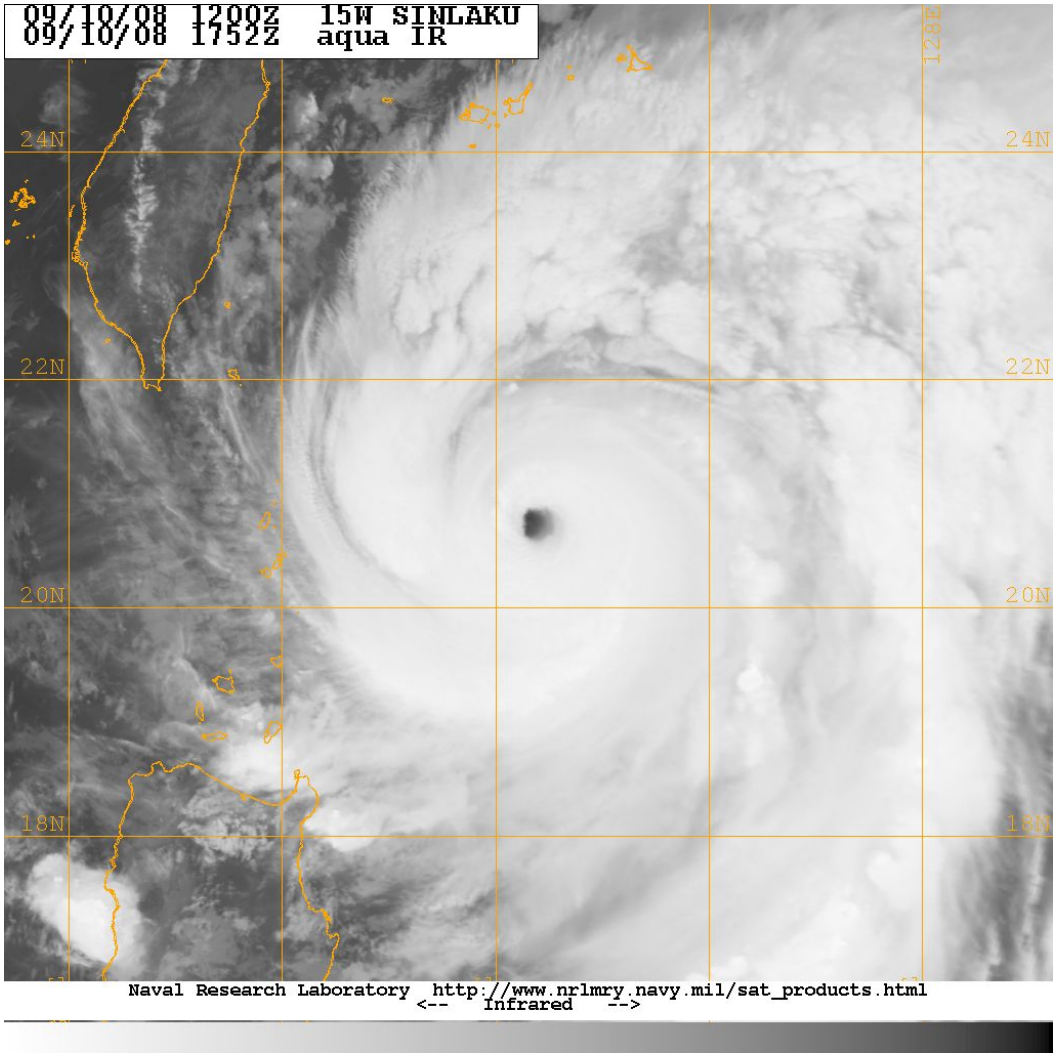
Observation coverage

- Snapshot of available observations



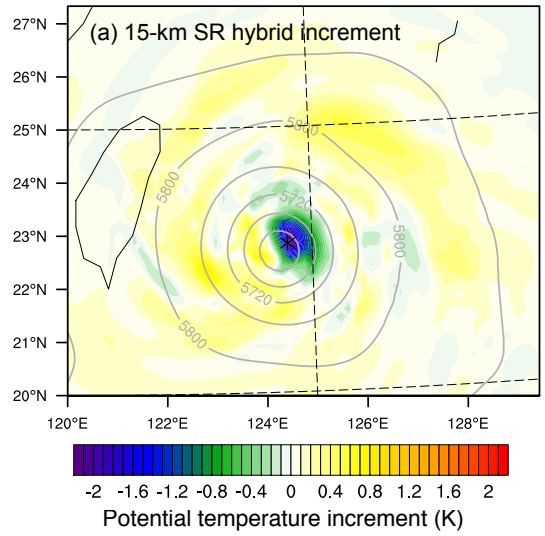
Single observation experiment

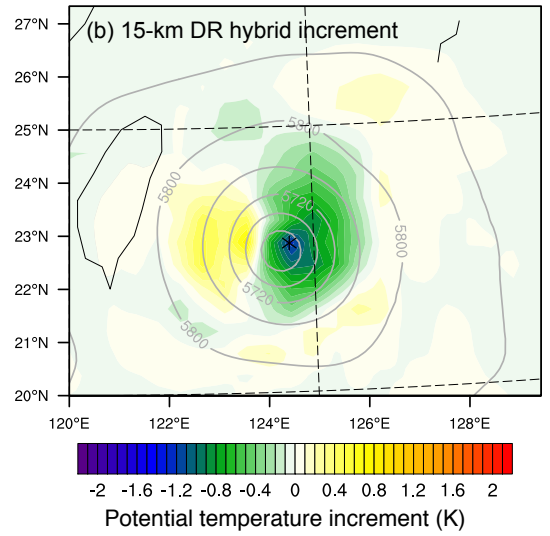
- Temperature observation (innovation -2K, ob. error 1K) placed in center of typhoon Sinlaku

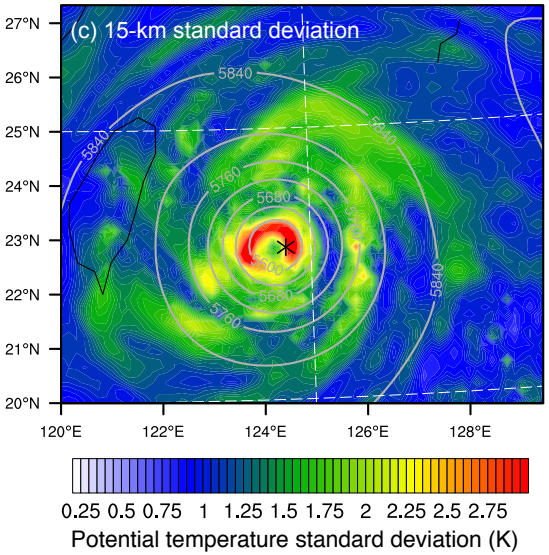


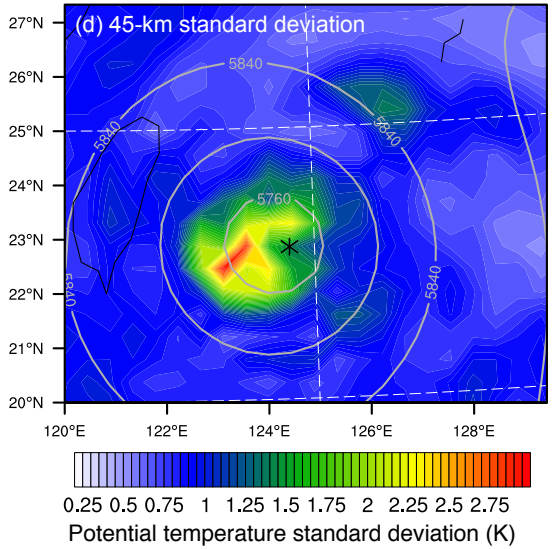
Single observation experiment

- Temperature observation (innovation -2K, ob. error 1K) placed in center of typhoon
- **Potential temperature** increments and spread





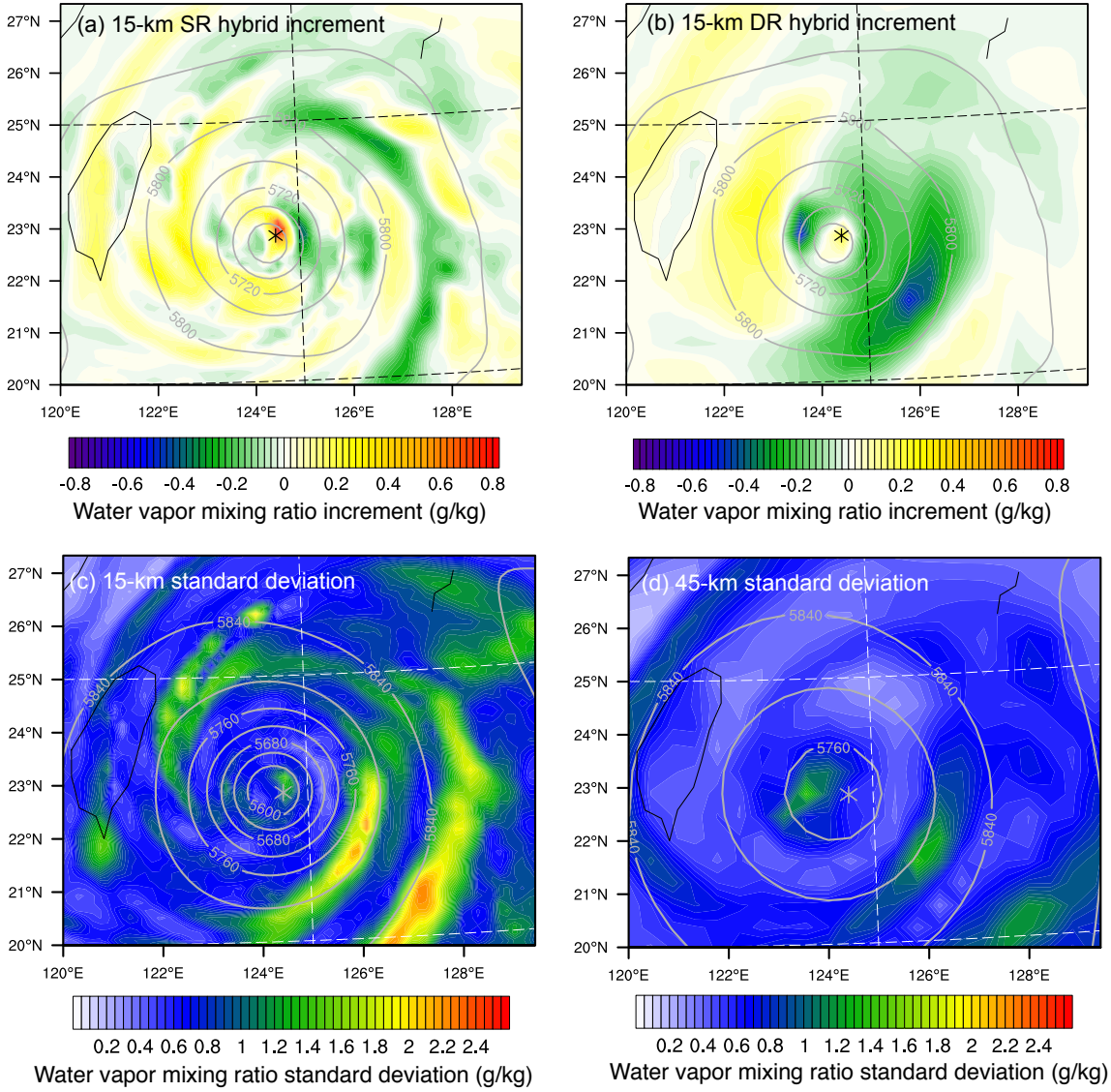




Single observation experiment

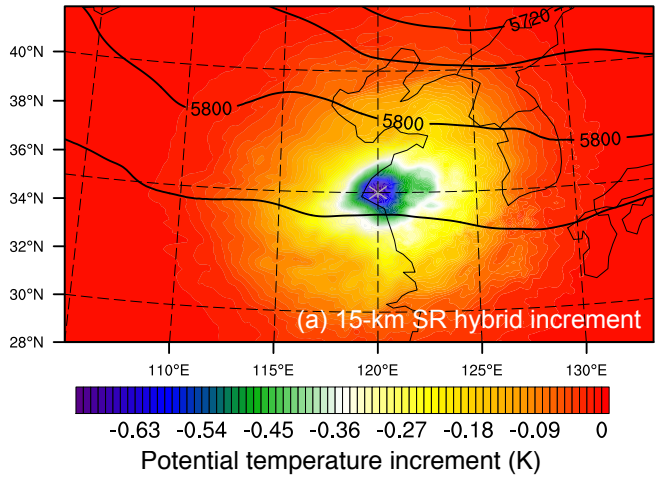
- Temperature observation (innovation -2K, ob. error 1K) placed in center of typhoon

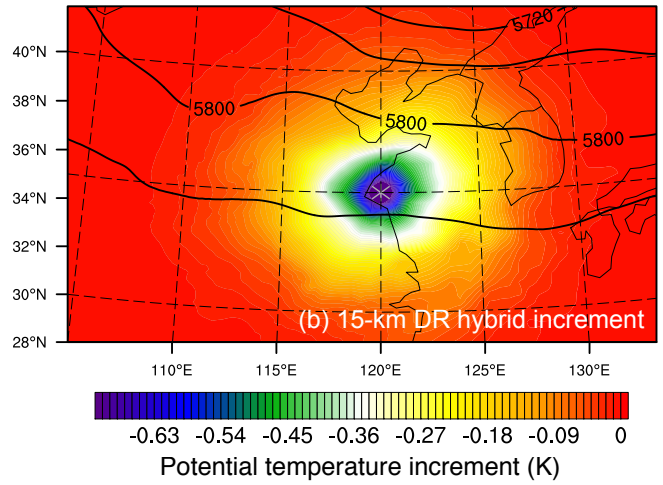
- **Mixing ratio** increments and spread

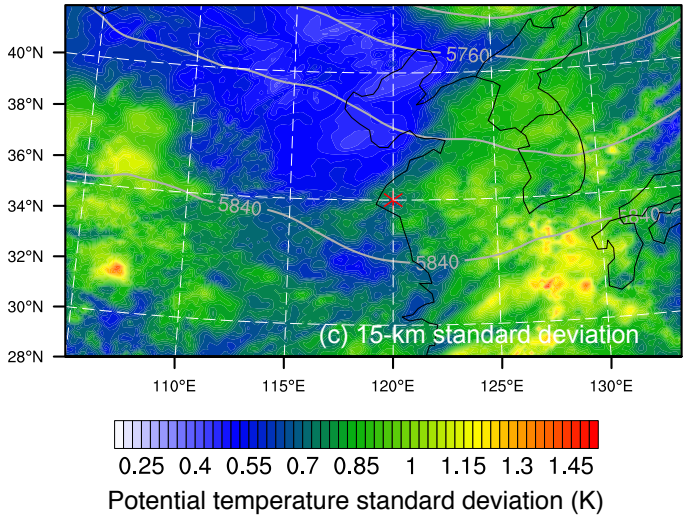


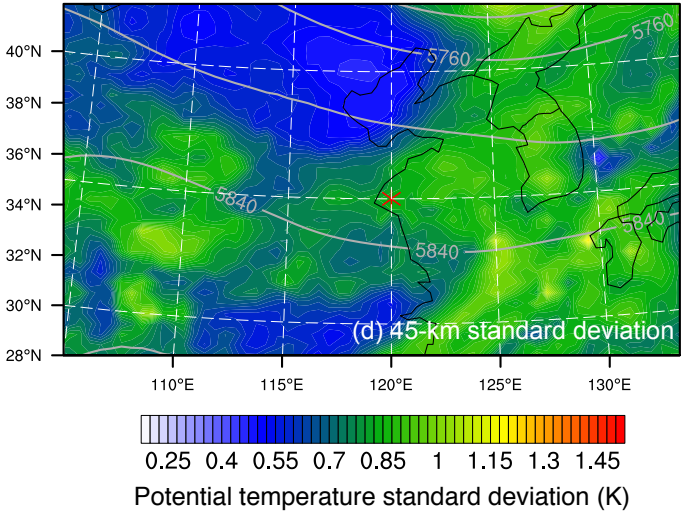
Single observation experiment

- Temperature observation (innovation -2K, ob. error 1K) placed in westerly flow
- Potential temperature increments and spread



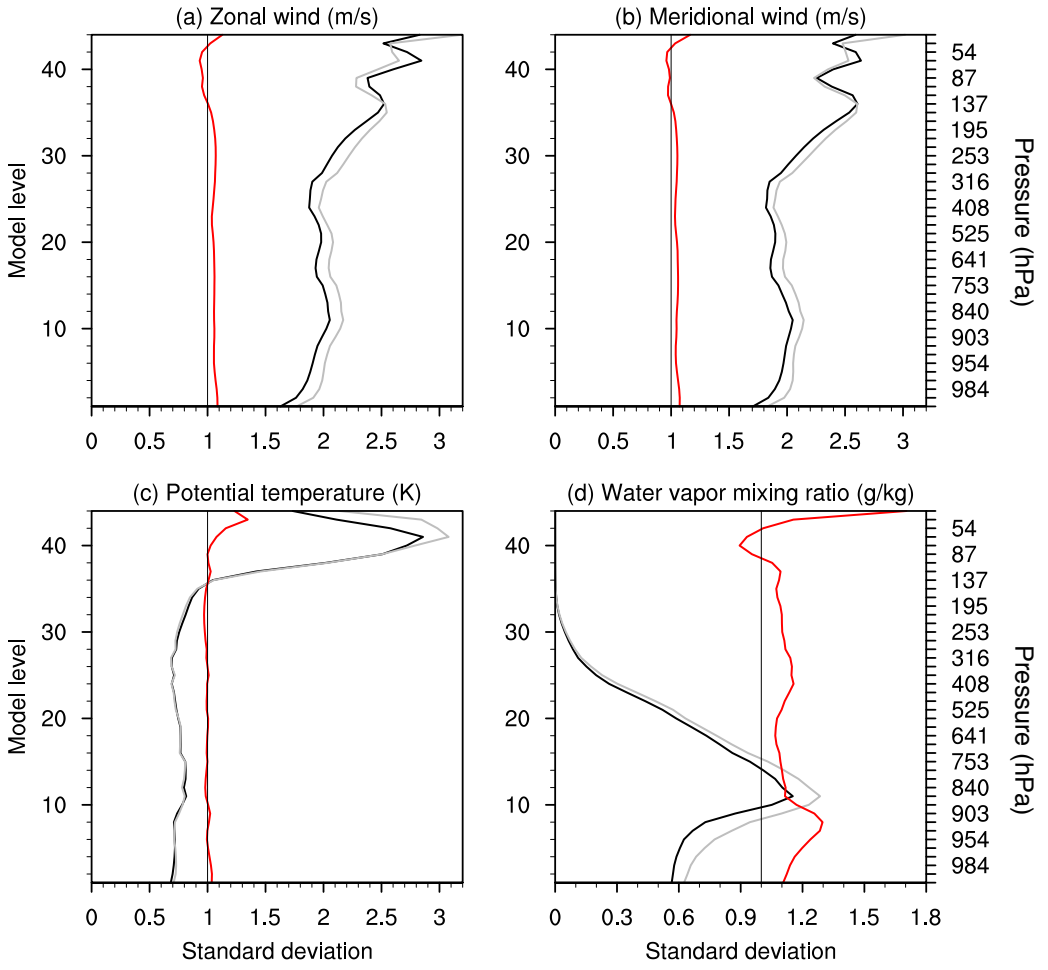






Ensemble spreads

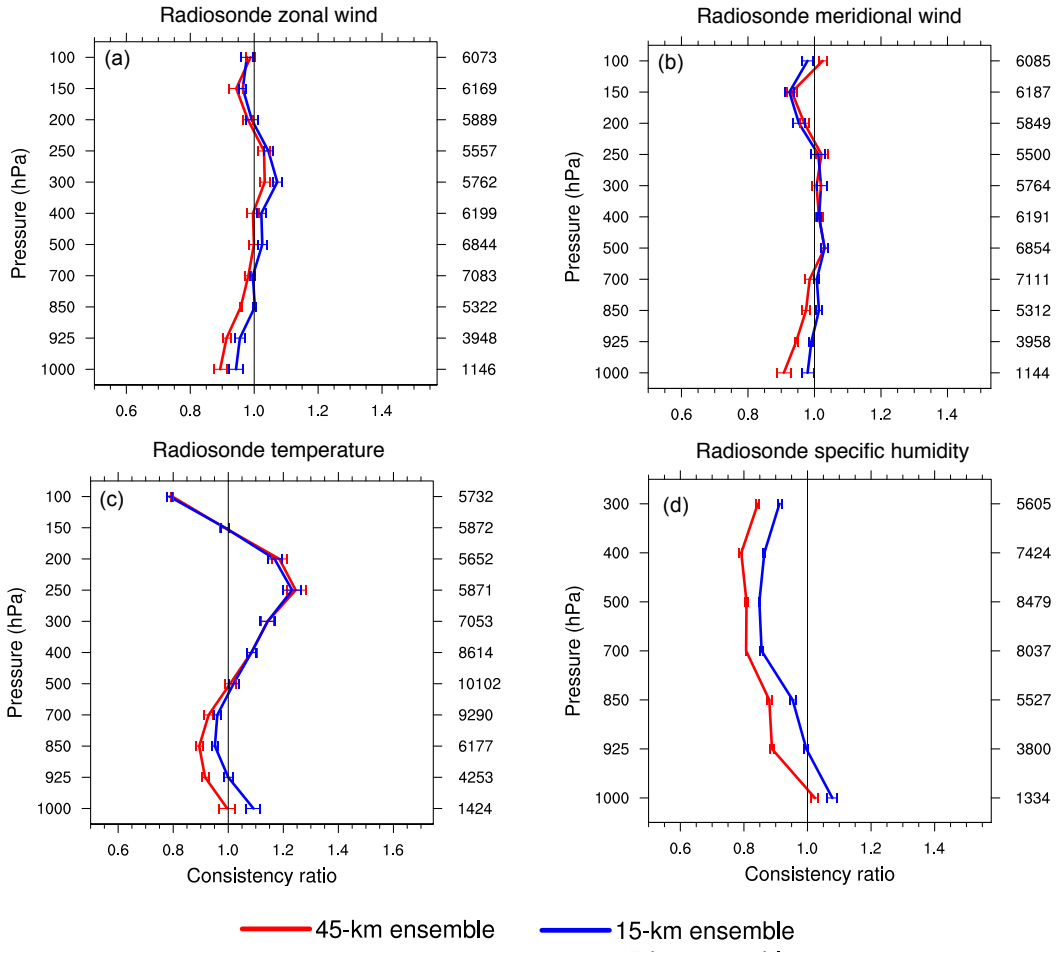
- Average prior ensemble spreads (over 15-km domain)
- Averaged between 1800 UTC 8 September and 0000 UTC 28 September
- After prior inflation



Ensemble spread/skill

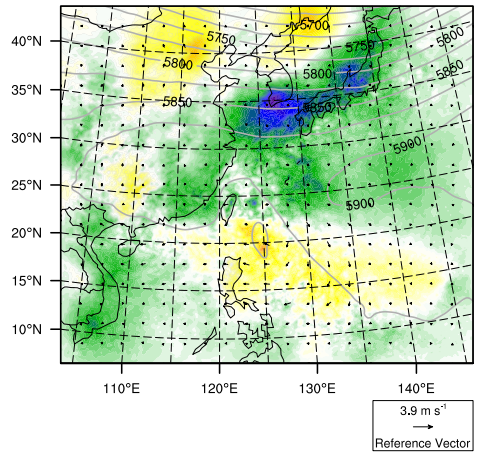
- How did the EAKF perform?
- Consistency ratios (CRs) averaged between 1800 UTC 8 September and 0000 UTC 28 September
- After prior inflation

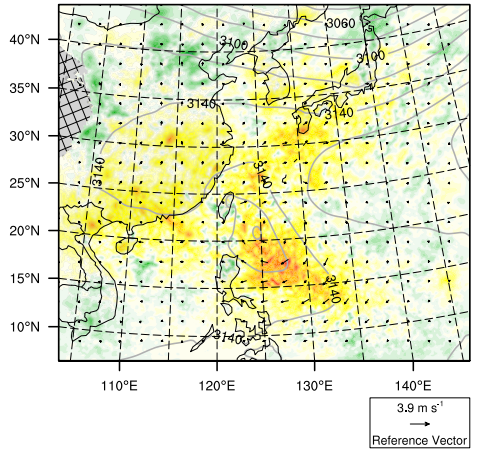
$$CR = \frac{Total\ spread}{RMSE}$$

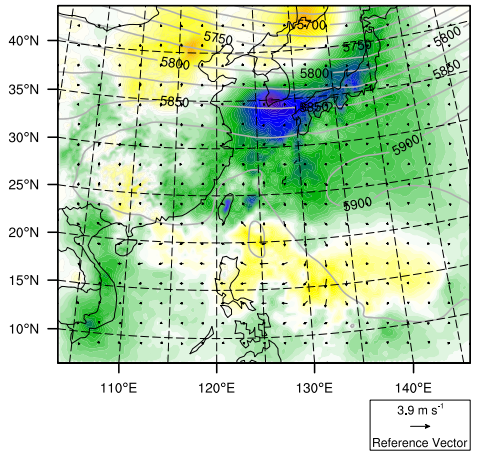


Mean analysis increments

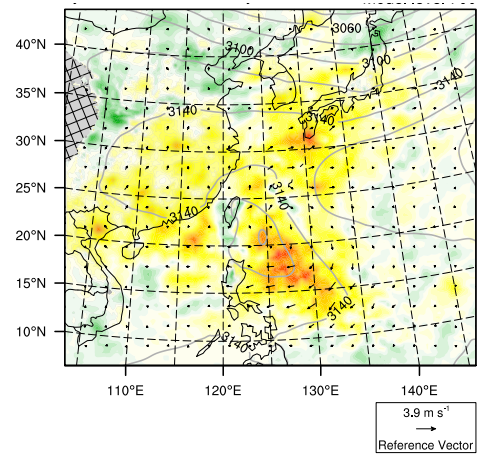
- Increments averaged between 1800 UTC 8 and 0000 UTC 28 September
- Which row is DR and which one is SR?
  - Can you tell the difference?







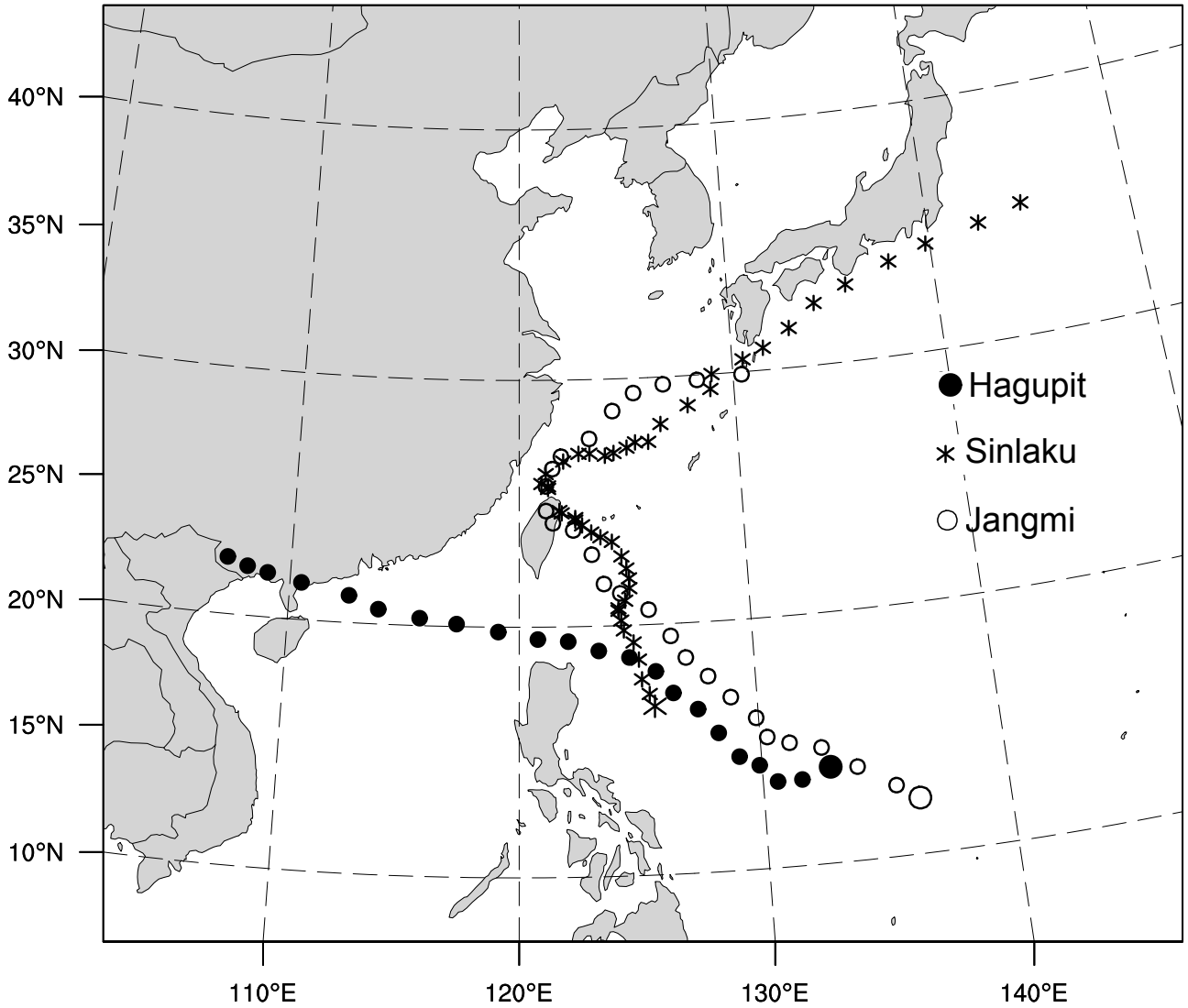
500 hPa potential temperature increment (K)



700 hPa water vapor mixing ratio increment (g/kg)

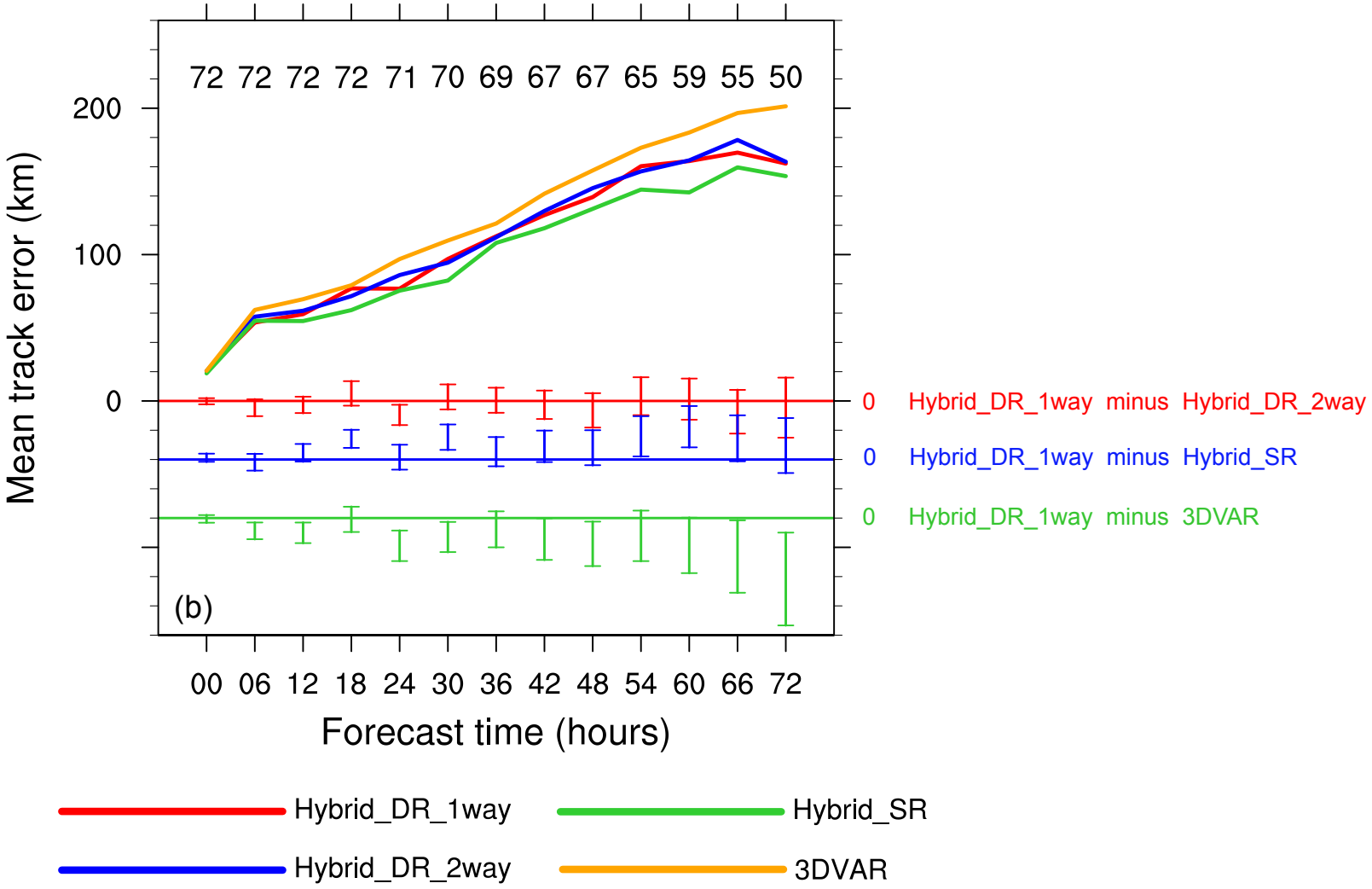
Typhoon track forecasts

- Verified 15-km forecasts against tracks of 3 typhoons



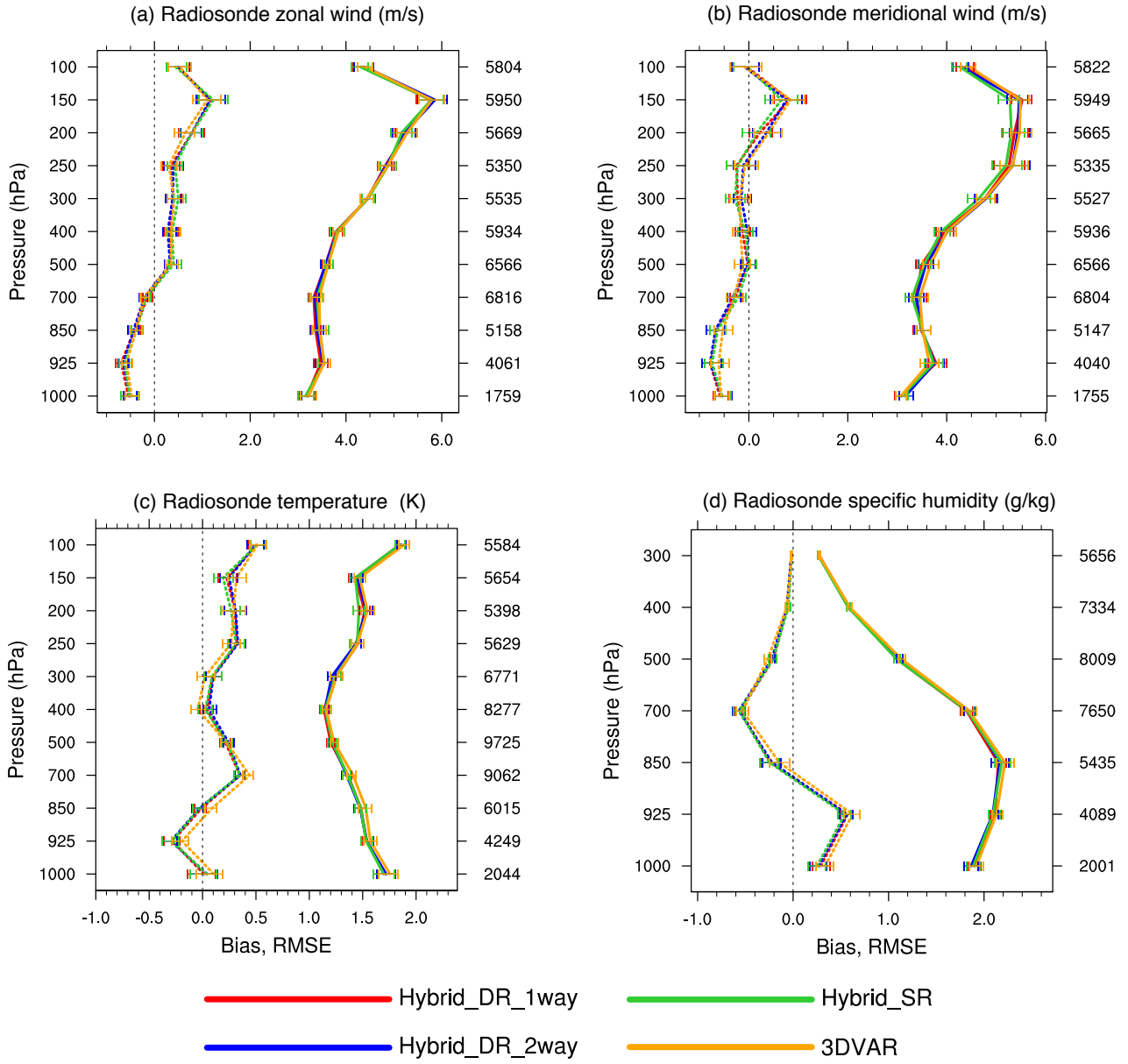
Typhoon track forecasts

- Verified 15-km forecasts against tracks of 3 typhoons



Verification against radiosondes

- 24-hr forecasts
- Aggregated over 78 forecasts initialized between 1800 UTC 8 and 0000 UTC 28 September



Toward higher-resolution analyses

Background

- Convection-allowing forecasts have traditionally been initialized from convection-parameterizing analyses
  - These forecasts have been good, but there are spin-up issues

- This work produces convection-allowing *analyses* over a large (e.g., 3000 km^2) domain
  - Many case studies have examined convection-allowing data assimilation over small domains

The challenge

- Ultimately, we want to produce analyses at convection-permitting resolution using flow-dependent BECs derived from ensembles

- But, for DA purposes, computational constraints currently do not permit high-resolution *ensembles* over domains large enough to resolve synoptic scale features
  - This will change fairly soon

Practical approach

- However, we can combine high-resolution *deterministic* backgrounds with low-resolution ensembles in a DR hybrid approach
  - Does not solve the issue of high-resolution ensembles
  - Permits high-resolution within ensemble DA systems

- Not clear how mixing a convection-allowing background with a convection-parameterizing ensemble will work

DA experiments

- Full-cycling (6-hr period) between May 4 – June 30, 2013

- Five DA experiments (analyses every 6-hrs):
  - Pure EnKF ("EnKF") : 20-km cycling
  - Pure 3DVAR ("3DVAR 20-km") : 20-km cycling
  - Pure 3DVAR ("3DVAR 4-km") : *4-km* cycling
  - SR Hybrid ("Hybrid 20-km") : 20-km cycling
  - DR Hybrid ("Hybrid 4-km") : *4-km* cycling

- Hybrid runs coupled to a 20-km, 50-member EnKF

- All assimilated identical conventional observations

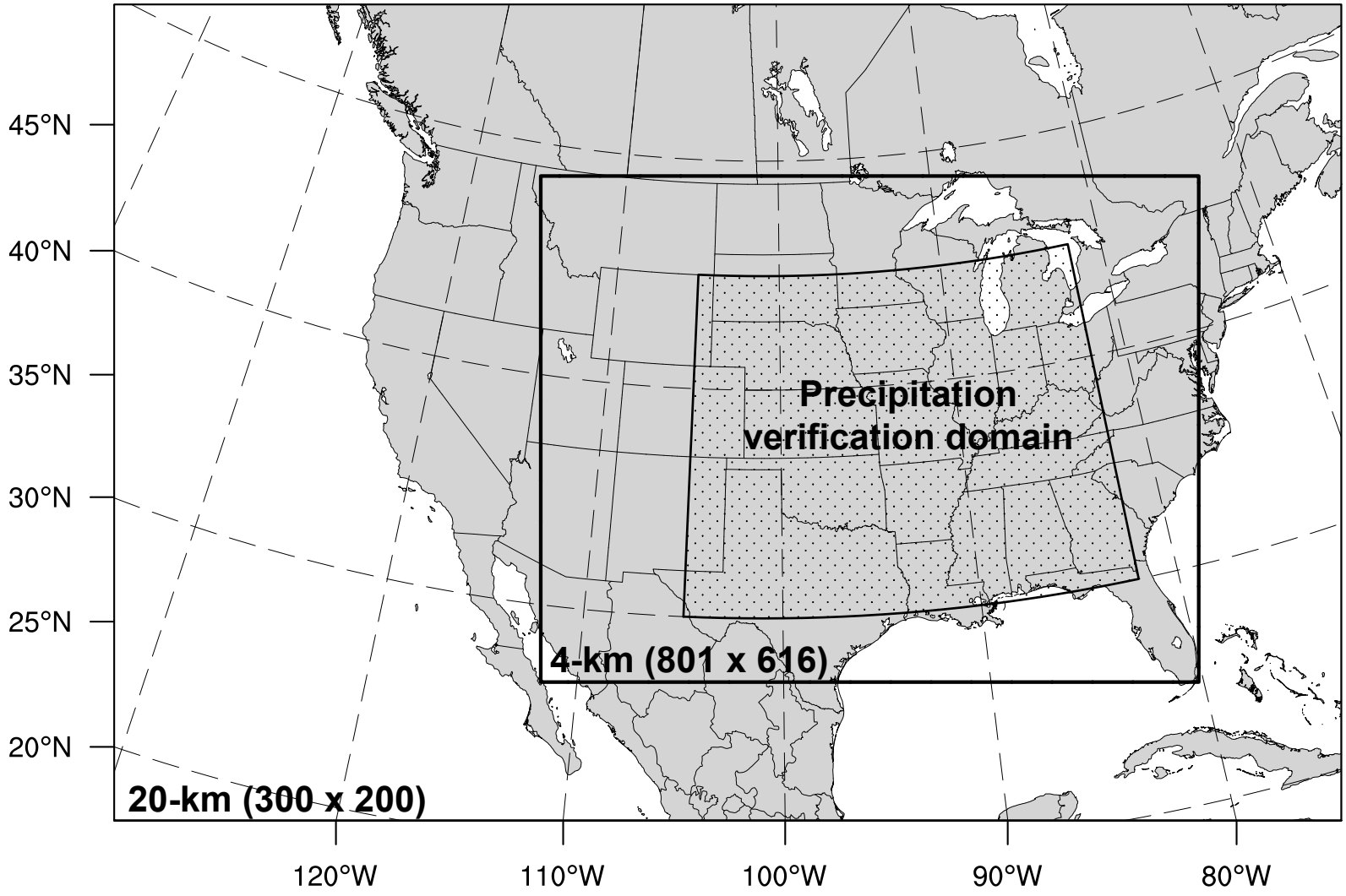
Selected data assimilation settings

- 50 ensemble members in hybrid/EnKF

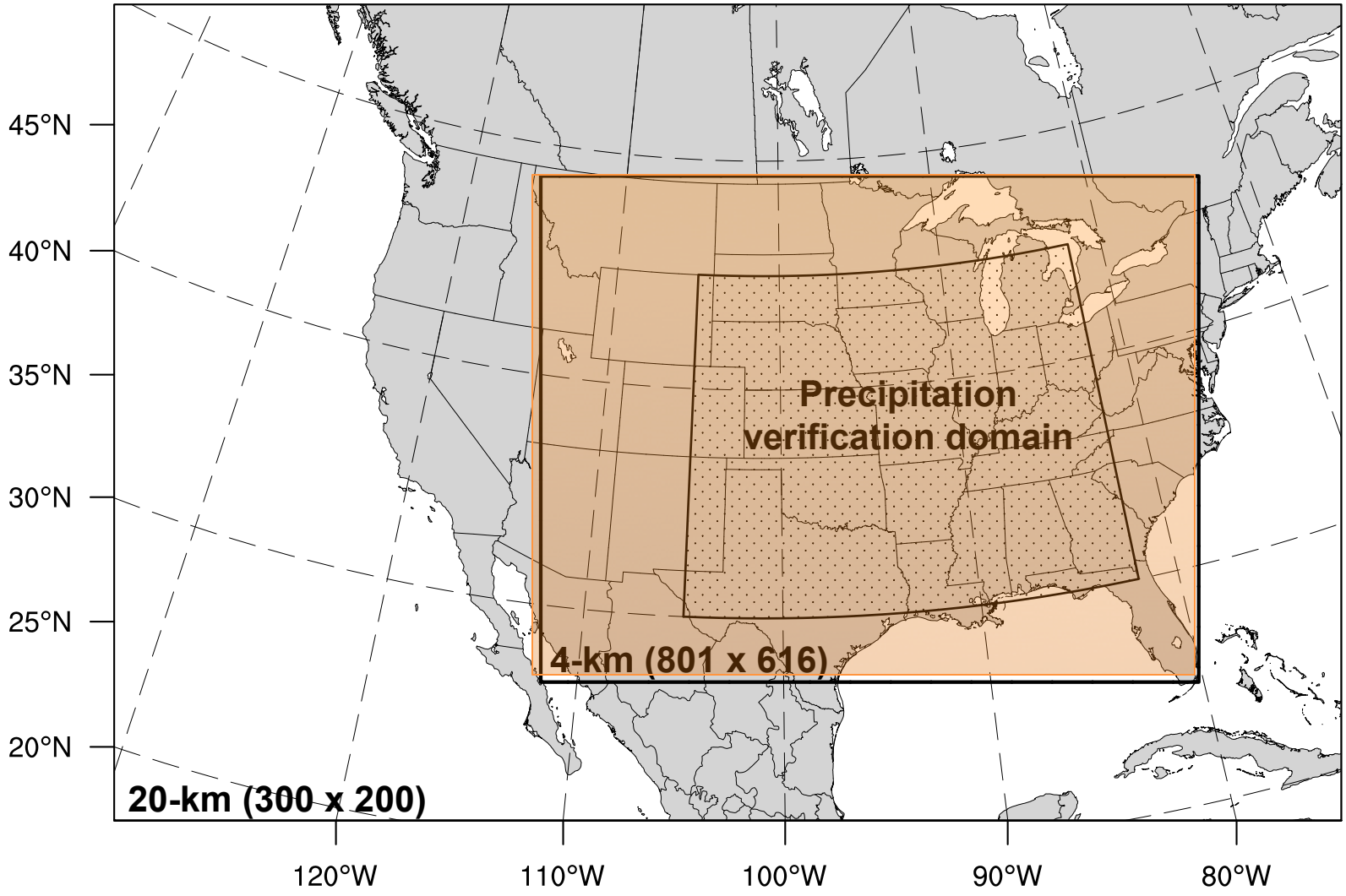
- Hybrid: 75% of background errors from ensemble, 25% from the static contribution

- Used posterior inflation for EnKF and localization in EnKF and hybrid

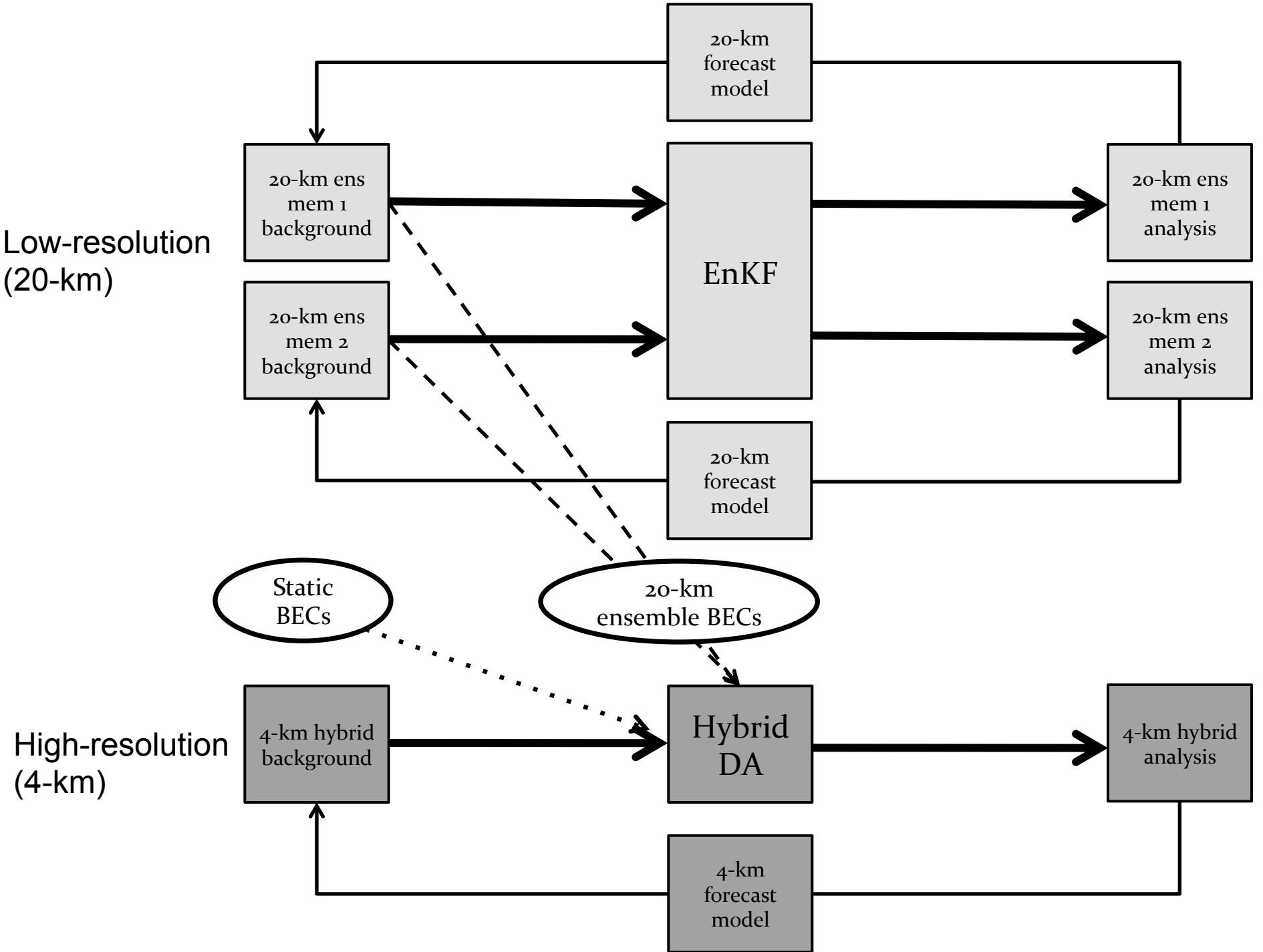
Computational domain



Computational domain



20-km ensemble covariances over orange region used for 4-km DR analyses


Low-resolution
(20-km)
20-km
forecast
model
20-km ens
mem 1
background
20-km ens
mem 2
background
EnKF
20-km ens
mem 1
analysis
20-km ens
mem 2
analysis
20-km
forecast
model
Static
BECs
20-km
ensemble BECs
High-resolution
(4-km)
4-km hybrid
background
Hybrid
DA
4-km hybrid
analysis
4-km
forecast
model

Forecast initialization

- 0000 UTC analyses initialized 36-hr 4-km WRF forecasts

- 4-km initial conditions were *downscaled 20-km analyses* in the 20-km 3DVAR, EnKF, and SR hybrid experiments

- *True 4-km analyses* initialized 4-km forecasts in the DR hybrid and 4-km 3DVAR experiments

- Forecast differences between SR and DR hybrid experiments due to analysis resolution

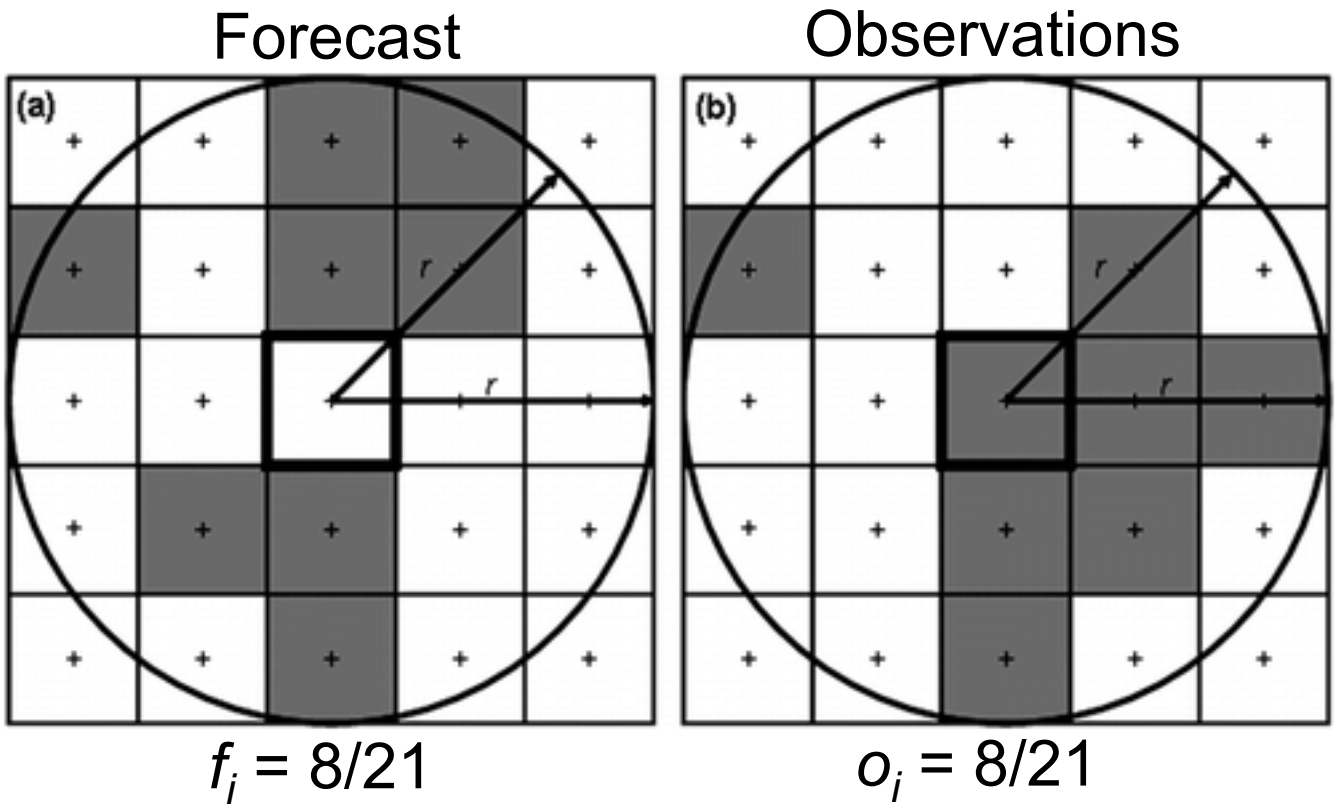
- Control: Interpolate 0000 UTC GFS analyses directly onto the domain and run forecasts

Precipitation verification

- Focus on 4-km hourly precipitation forecasts
- NCEP Stage IV observations as "truth"
- All precipitation statistics aggregated over 55 4-km forecasts
- Fractions skill score (FSS) quantifies displacement errors

Fractions skill score (FSS)

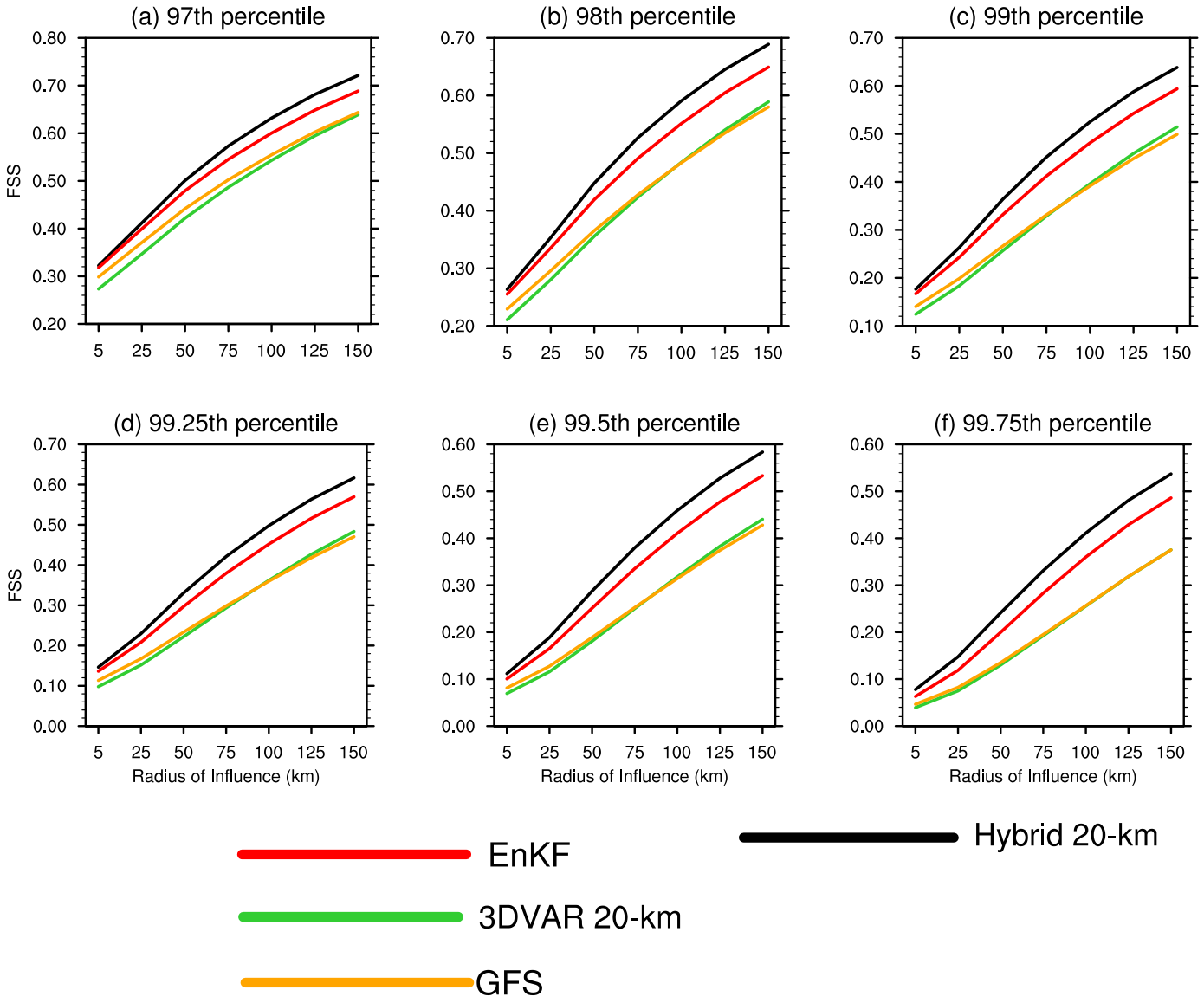
- A neighborhood approach to verification, which is needed to verify high-resolution forecasts



- Directly compare observed and forecast fractions to compute the FSS

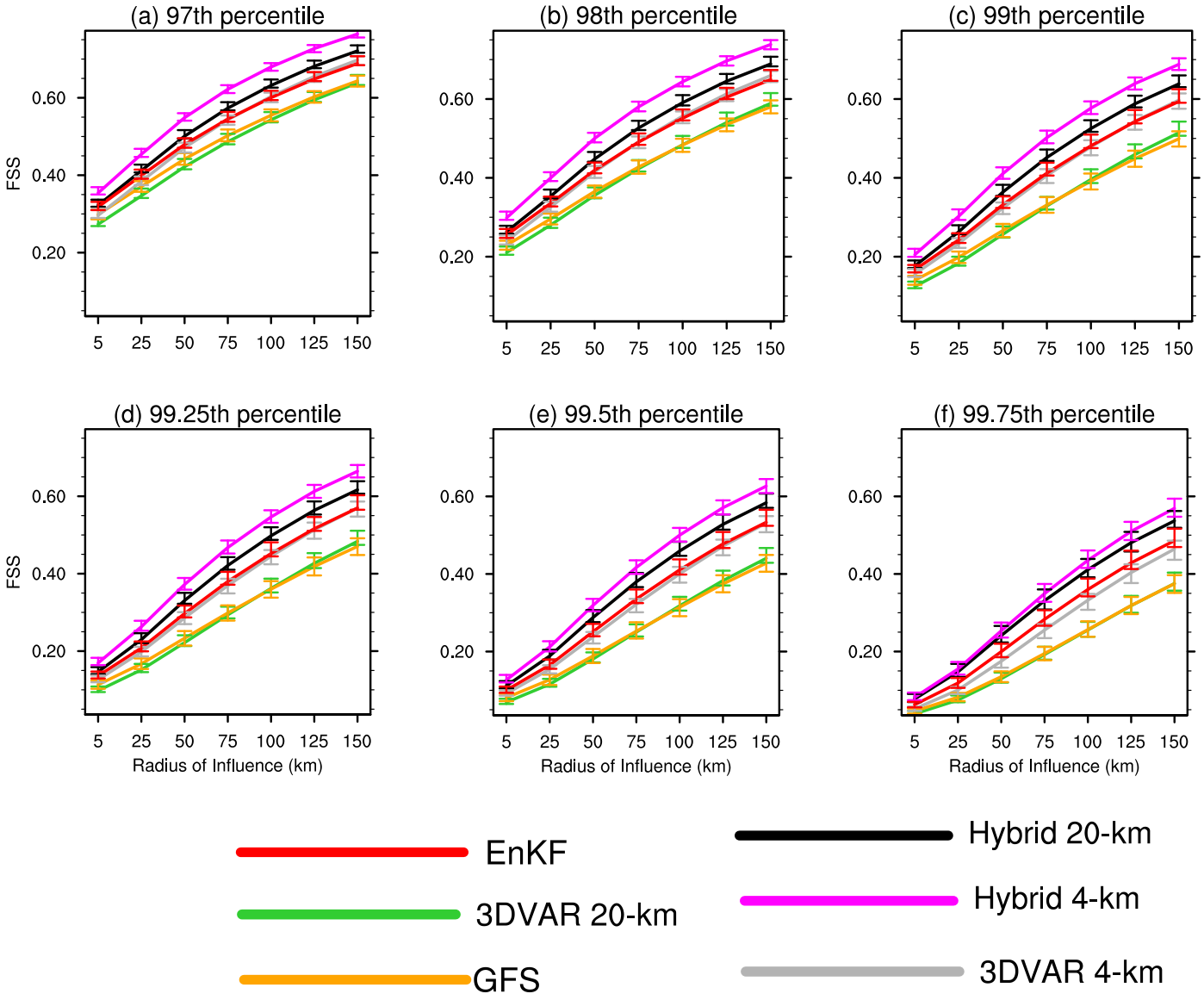
Precipitation verification: the first 12-hrs

- Fractions skill score (FSS) aggregated over the **first 12 forecast hours** and 55 4-km forecasts



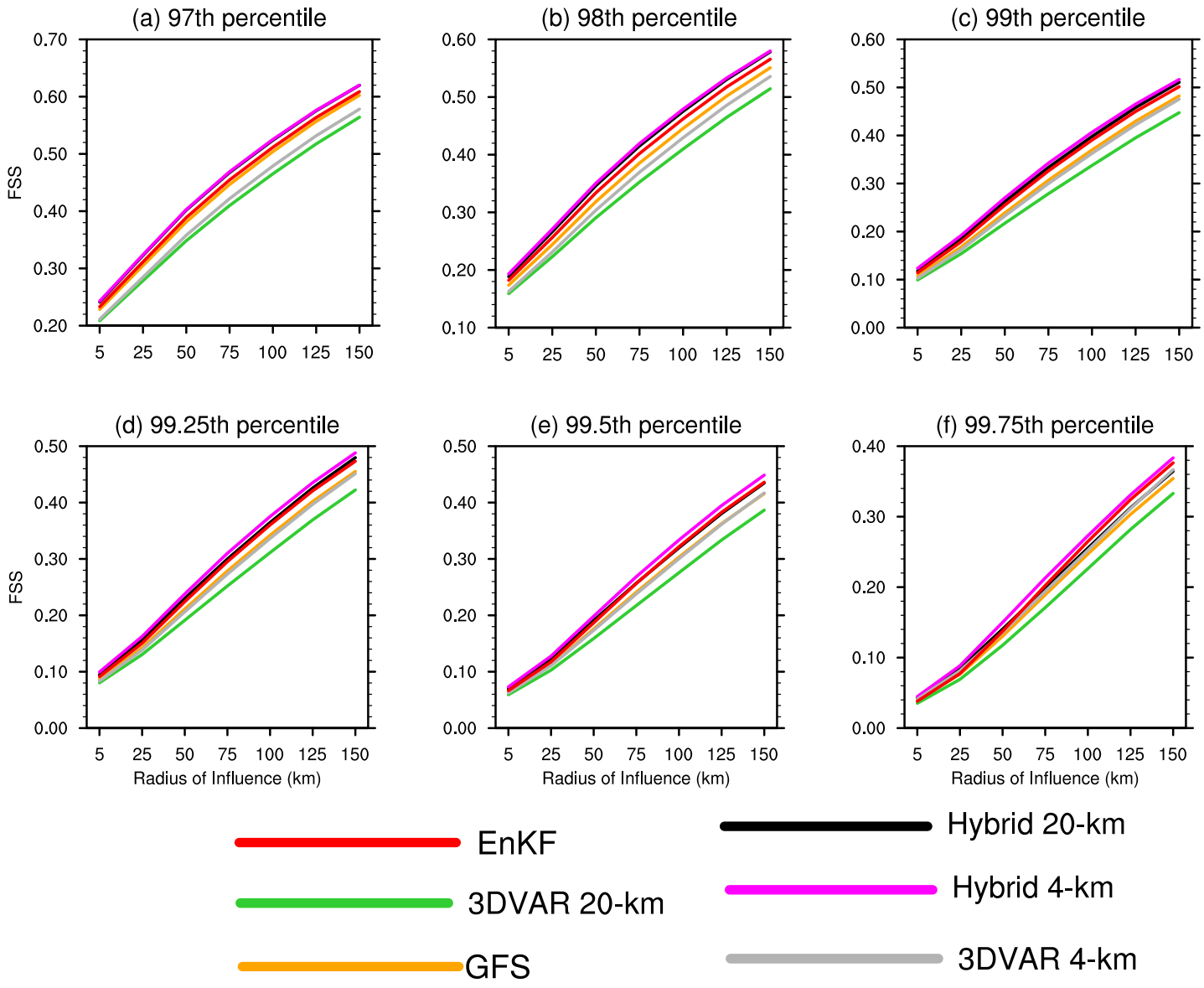
Precipitation verification: the first 12-hrs

- Fractions skill score (FSS) aggregated over the **first 12 forecast hours** and 55 4-km forecasts



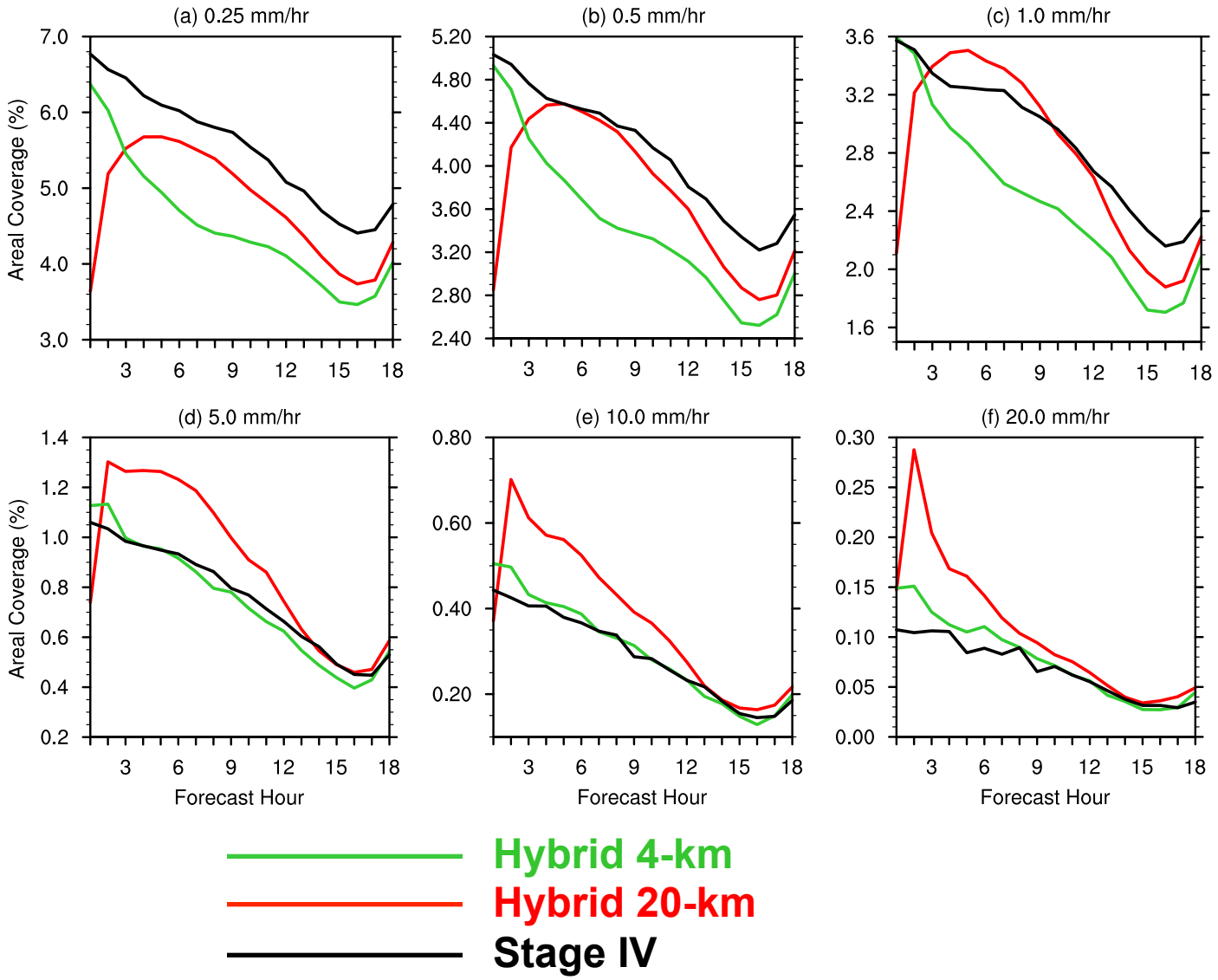
Precipitation verification: 18-36-hrs

- Fractions skill score (FSS) aggregated over **forecast hours 18-36** and 55 4-km forecasts



Areal coverages of precipitation

- Aggregate fractional coverage of precipitation exceeding certain thresholds aggregated over all 55 forecasts

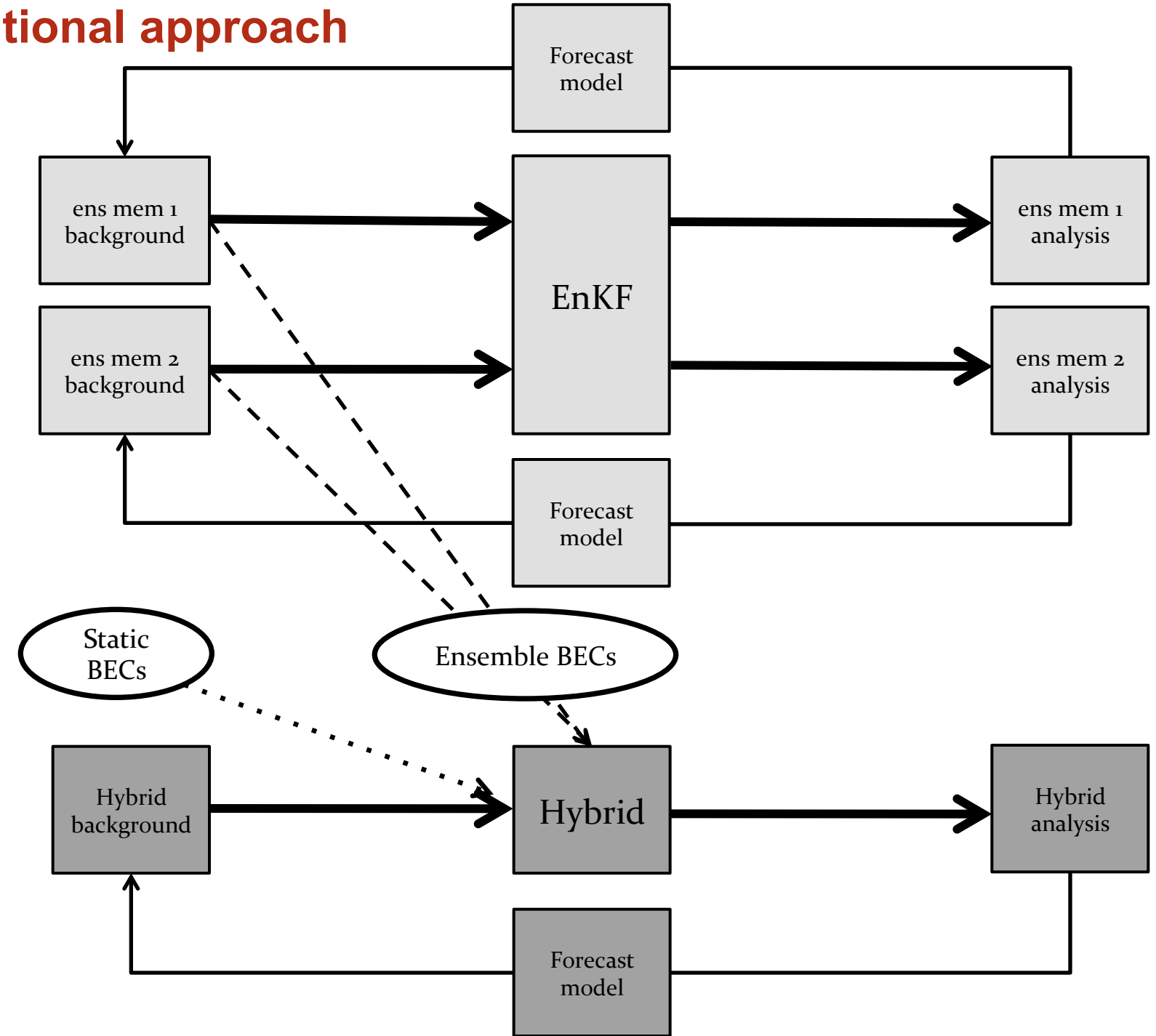


Introduction to integrated variational/ ensemble DA approaches

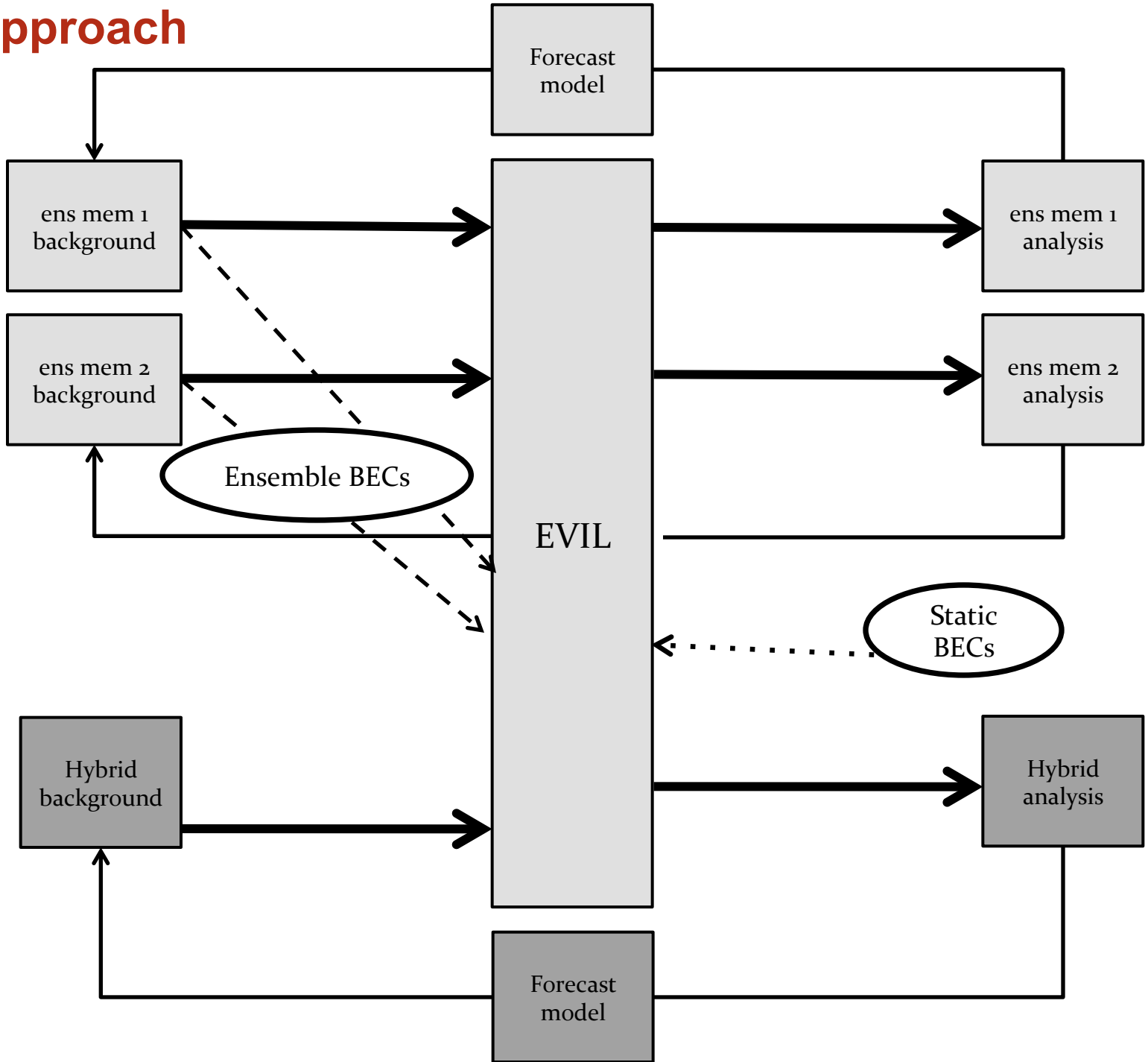
EVIL DA

- EVIL stands for "Ensemble variational integrated localized"
- Developed by Tom Auligne (NCAR)
- Simultaneously updates both an ensemble and deterministic background
- Uses properties of the variational minimization to produce the analysis ensemble

Traditional approach



EVIL approach



Summary

- The hybrid method is a practical way of incorporating ensemble BECs into DA systems

- Dual-resolution method is also a highly practical tool
  - Still not completely clear how much is lost through use of coarse-resolution ensemble

- Integrated variational/ensemble methods (e.g., EVIL) are “hot” research topics