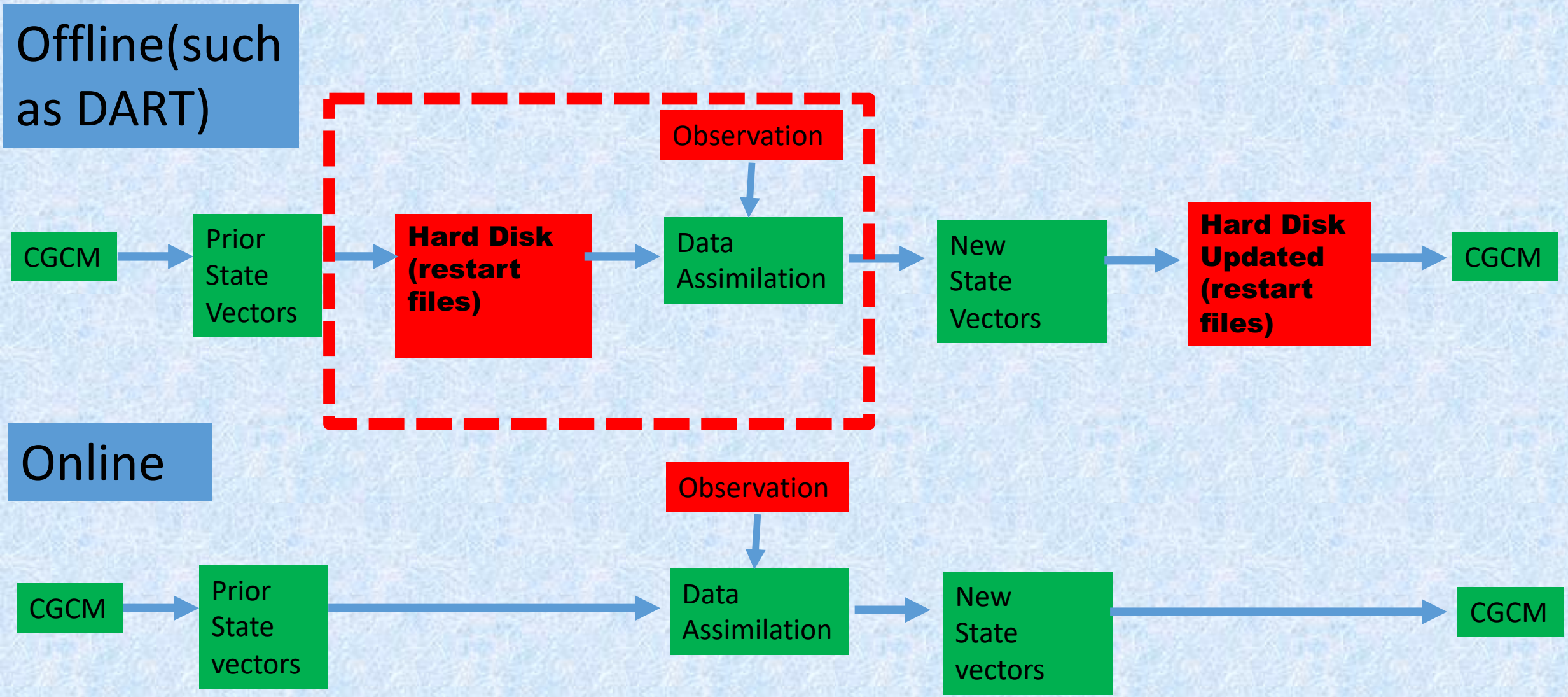
System Design and Evaluation of a Real Time Online Hybrid Data Assimilation System based on E3SM

S. Wu (UW-Madison)
Z. Liu (OSU)
L.V. Roekel (LANL)
S. Balwinder (PNNL)

Outline

- 1. Online vs Offline data assimilation system
- 2. Implementation of an online DA system into E3SM
- 3. Results
- 4. Efficiency comparisons: Online vs Offline DA system
- 5. Summary and Perspective

1. Online vs Offline Data Assimilation System



Main differences: Hard Disk I/O

Key implementation steps for an online DA system

- **1. Set up an online parallel run for ensemble simulation.**
- **2. Build the communicators across processors(P2P)**
- **3. Implement DA package**
- **4. Carry out experiments**
- **5. Verification**
- **……**

Hardest

Easiest

Tested compset and resolution:

- **Compset=A_WCYCL1850S_CMIP6**
- **Resolution=ne30_oECv3_ICG**
- **Machine=Cori-KNL**

Code modifications(an example).

- **mpas-ocean(seaice)** /buildnml
- my $NINST_OCN = 4;
- my $inst_string;
- my $inst_counter = 1;
- while ($inst_counter <= $NINST_OCN) {
- …….
- $sysmod = "cp $RUNDIR**/$STREAM_NAME** $RUNDIR**/$STREAM_NAME${inst_string}**";
- **}**

Courtesy of Coupler development team.

2. Implementation of an online DA system into E3SM

Consistency of ensemble run and single run

	Cori-KNL single run		Cori-KNL online parallel run(ninst=4)	
	temperatureMin	temperatureMax	temperatureMin	temperatureMin
Day1	-1.803327042822610	31.714405225595101	-1.803327042822610	31.714405225595101
Day2	-1.803310266115530	31.743831529932802	-1.803310266115530	31.743831529932802
Day3	-1.803302579358390	31.828462783695400	-1.803302579358390	31.828462783695400
Day4	-1.803293675869160	31.716819040285198	-1.803293675869160	31.716819040285198
Day5	-1.803472311151500	31.351666918635999	-1.803472311151500	31.351666918635999

E3SM HYBRID DA system settings

- Online ensemble size=8 (running simultaneously)
- Stationary ensemble size=100
- B matrix

  B = alpha*B_EAKF + (1-alpha)*B_EnOI

  EAKF system alpha=1

  HYBRID system 0<alpha<1

  EnOI system alpha=0

2. Implementation of an online DA system into E3SM

Initial conditions and stationary vectors

- 8 online members

  Member 1: 1980.1.1 Deck mem 1
  Member 2: 1980.1.1 Deck mem 2
  Member 3: 1980.1.1 Deck mem 3
  Member 4: 1980.1.1 Deck mem 4
  Member 5: 1980.1.1 Deck mem 5
  Member 6: 1990.1.1 Deck mem 1
  Member 7: 1990.1.1 Deck mem 2
  Member 8: 1990.1.1 Deck mem 3

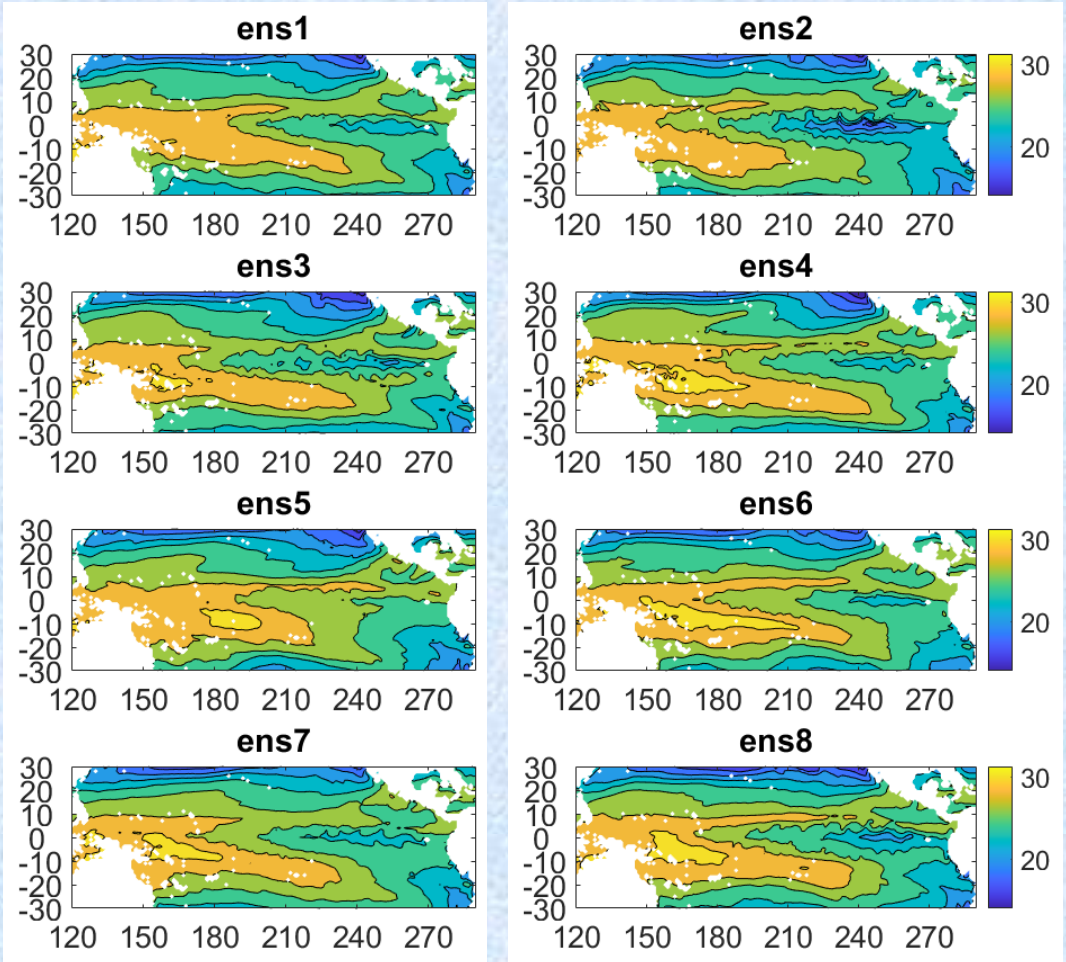
- 100 Stationary members

  Member 1: 1951-2000 Deck mem1
  Member 2: 1951-2000 Deck mem2

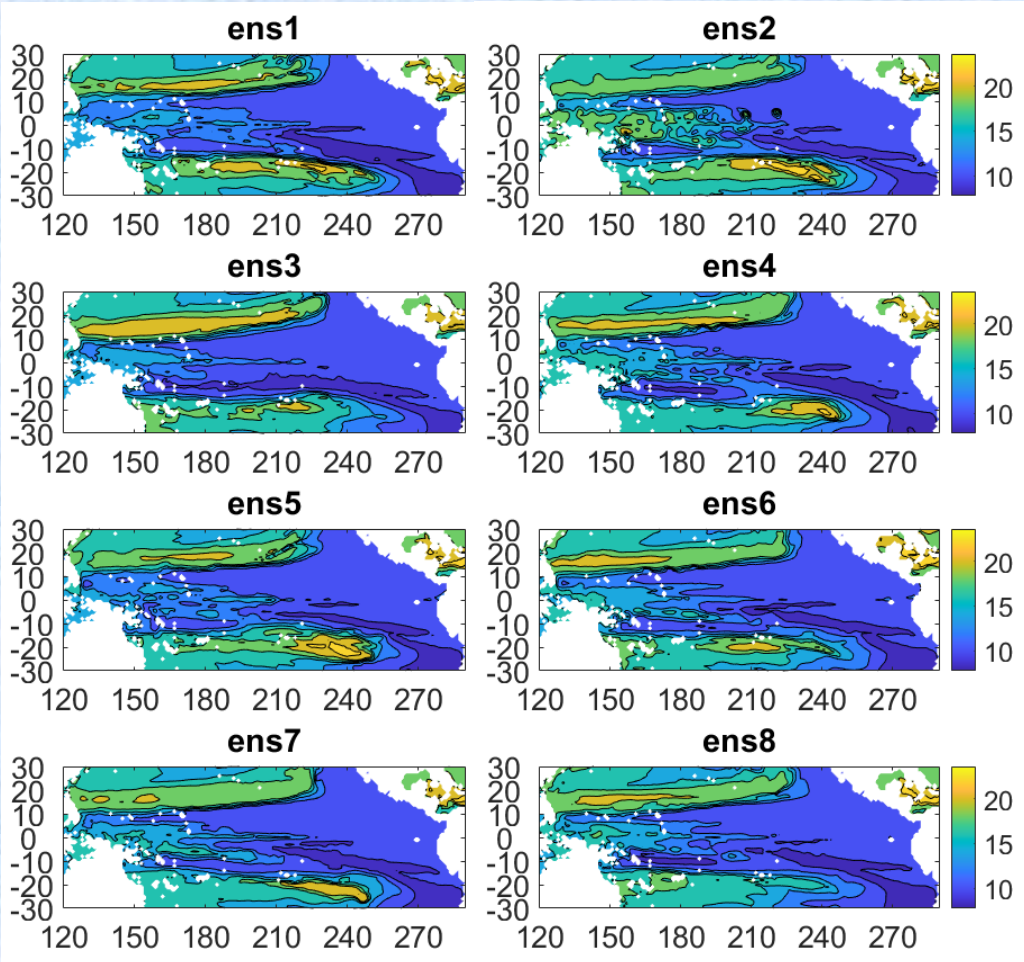
2. Implementation of an online DA system into E3SM

Initial conditions of E3SM DA system (Pacific SST and T200 as an example)

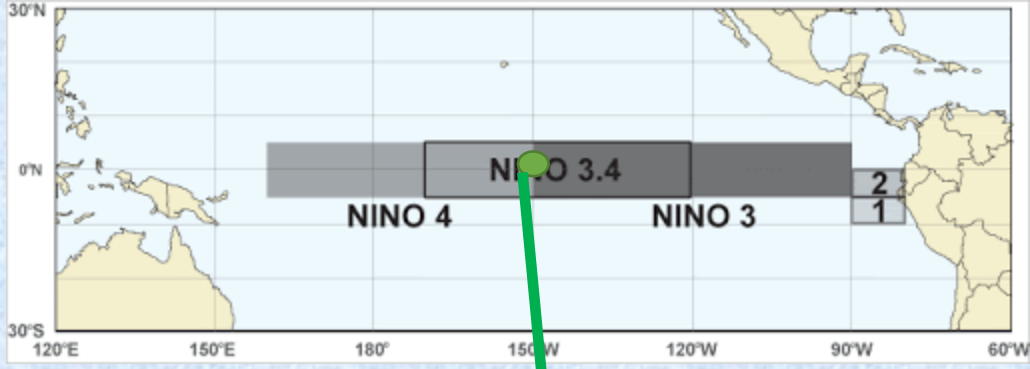
SST

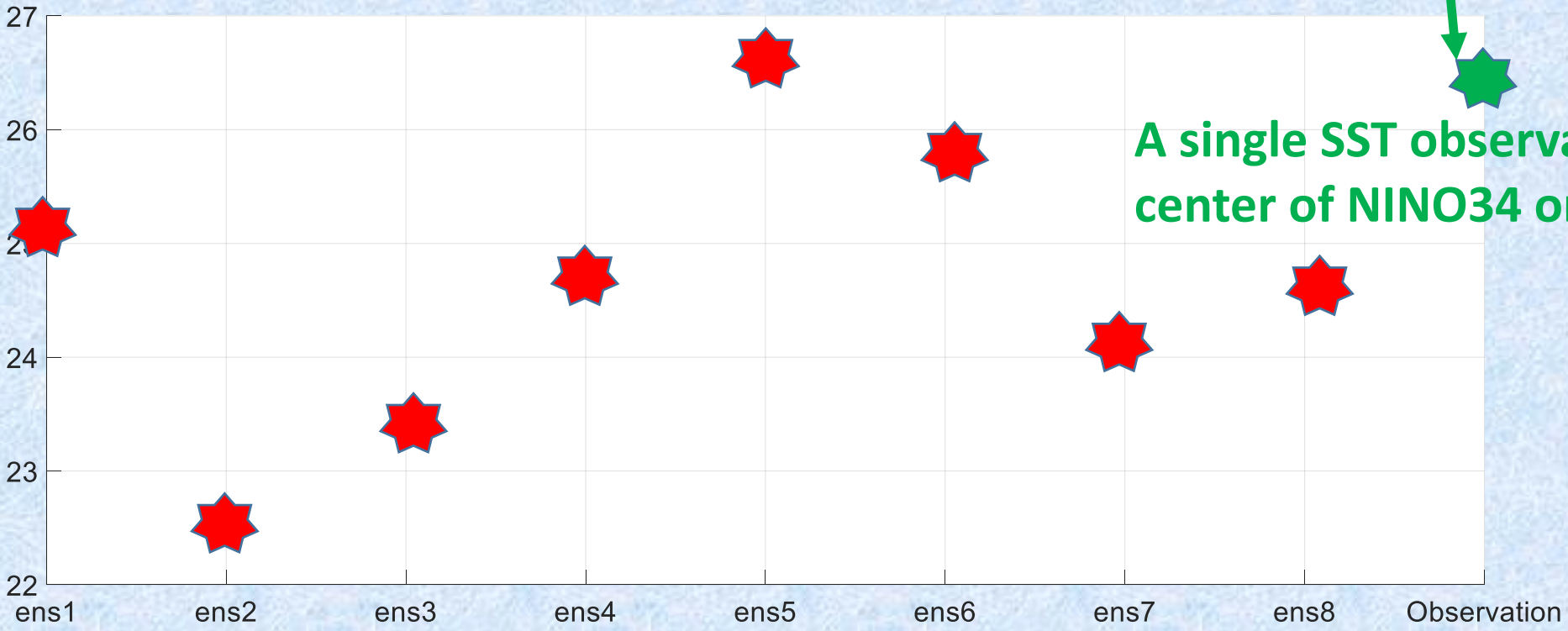


T200



3. Results:
Test 1: Single Observation

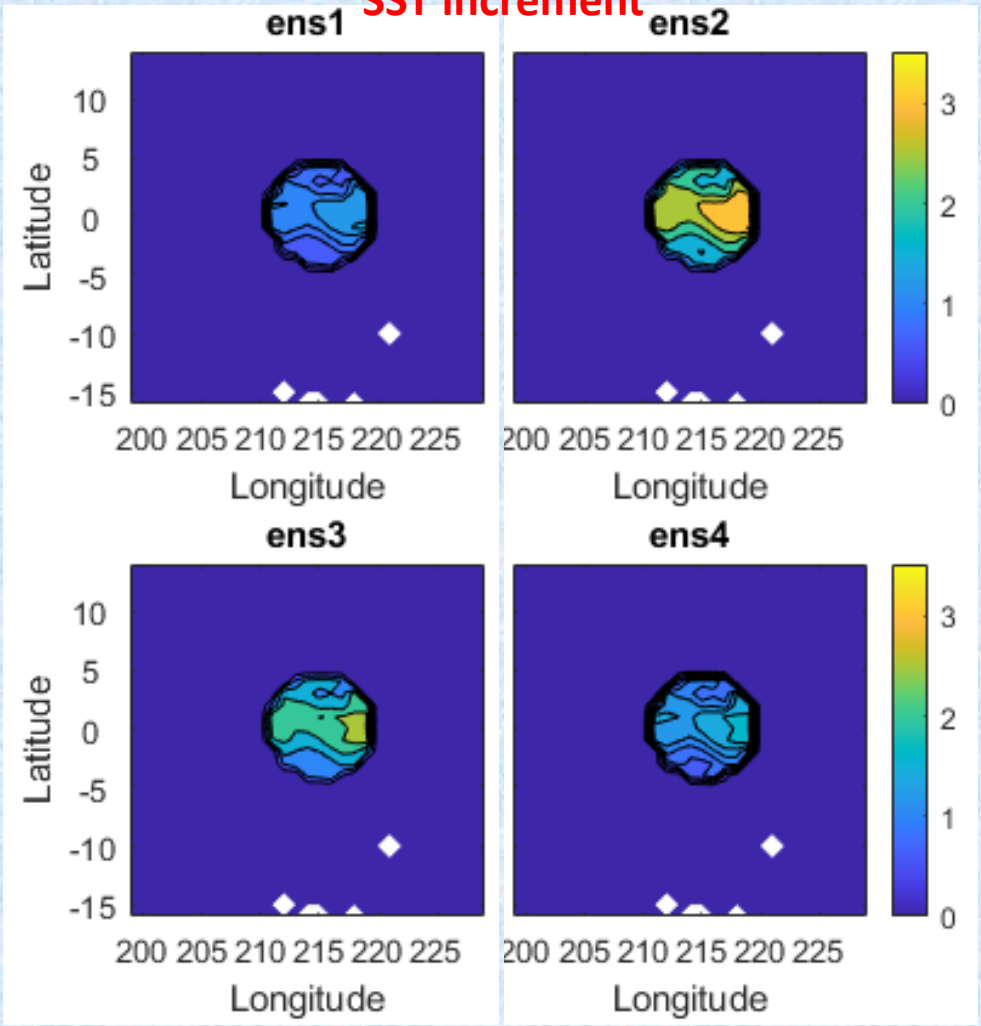




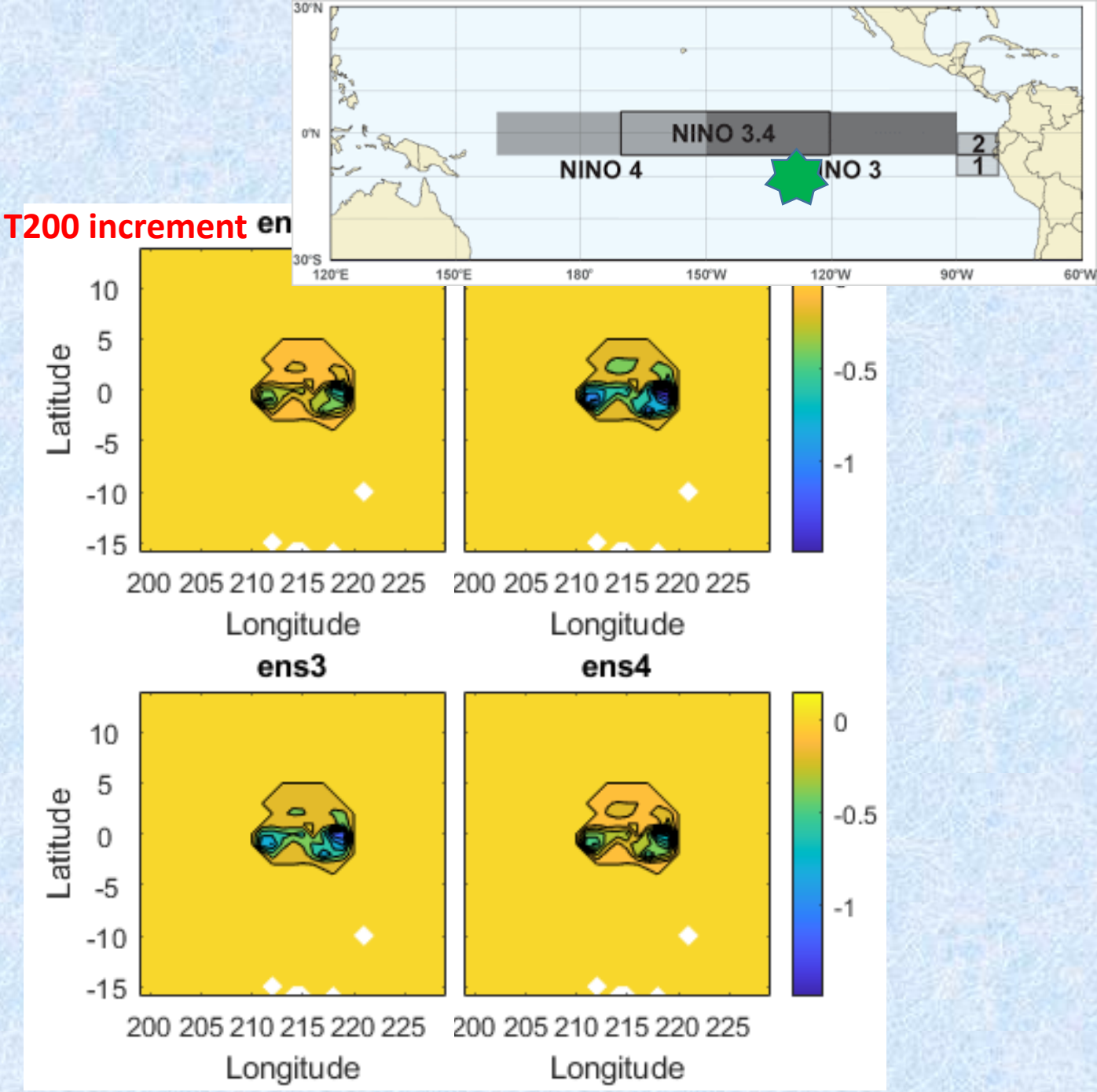
A single SST observation at the center of NINO34 on 1/1/1982

3. Results:
Test 1: Single Observation

SST increment



T200 increment



Impact of a single SST observation is broadened to a region both horizontally and vertically through flow dependent covariance.

3. Results:
Test 2: Global SST assimilation

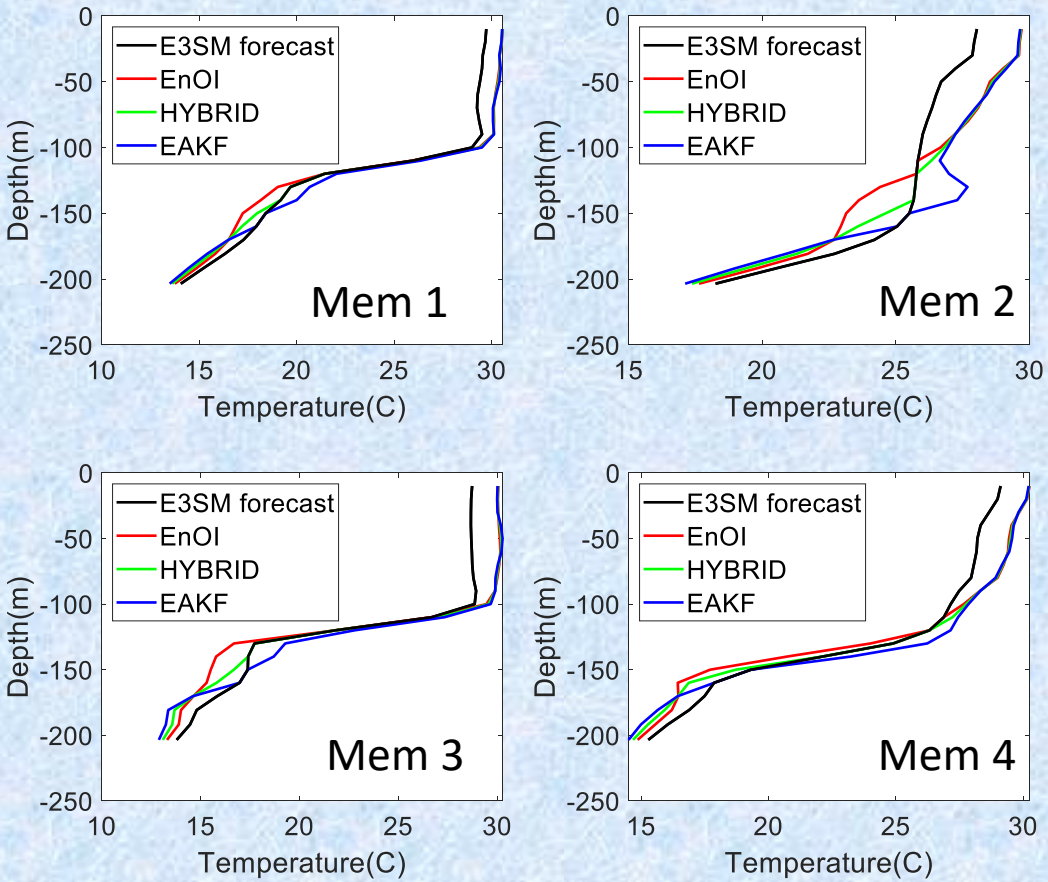
Global SST assimilation experiment settings

- Global points: 235160(MPASO) vs 691150(OISST)
- OISST is interpolated to MPASO grid
- Assimilation region(0-360, 60S-60N)
- R_SST = 0.5
- Horizontally point to point. No localization yet!
- vertically, z=1-20, roughly 10m-200m
- Alpha=1;0.5;0

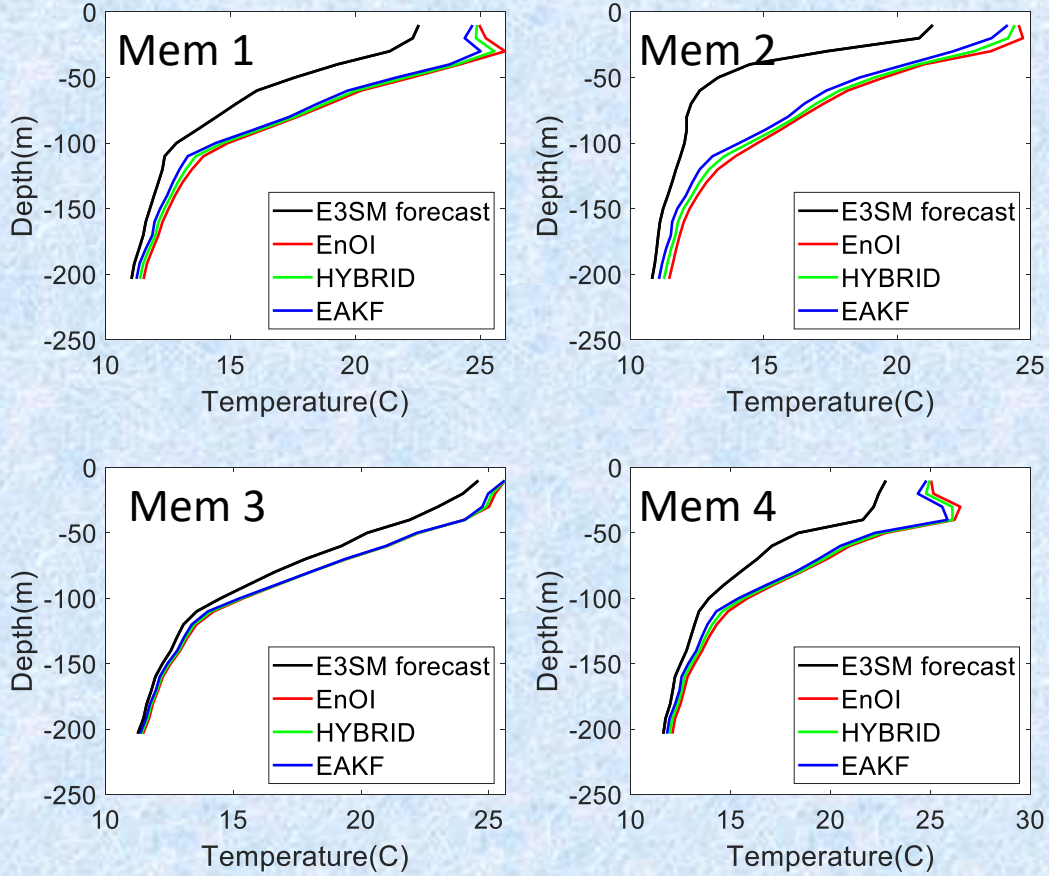
3. Results:
Test 2: Global SST assimilation

EnOI vs EAKF vs HYBRID

Western equatorial Pacific

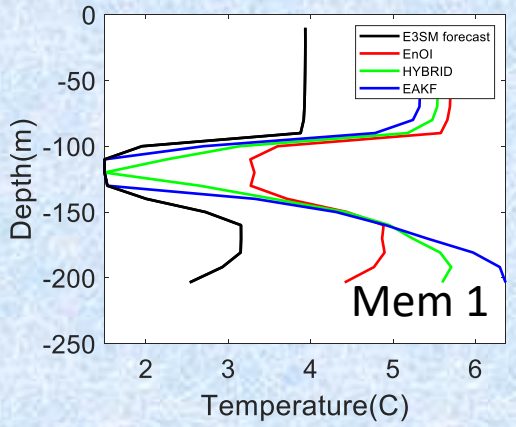


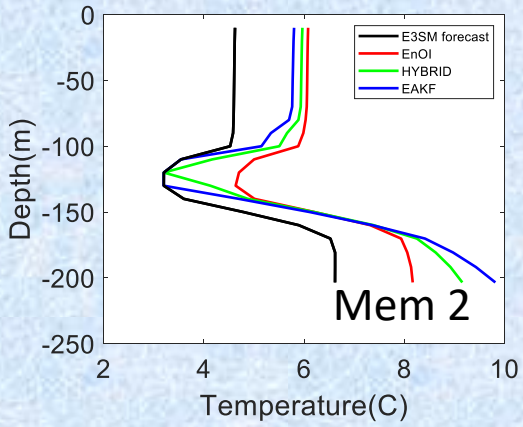
Eastern equatorial Pacific.

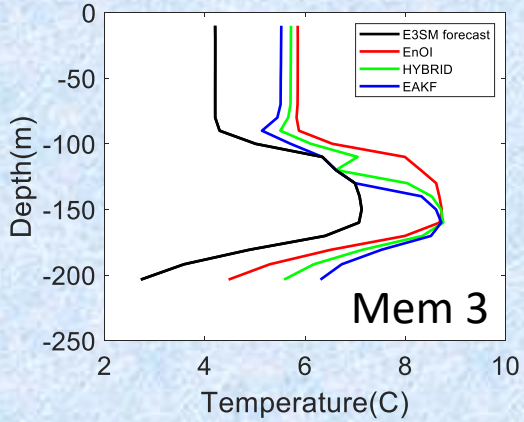


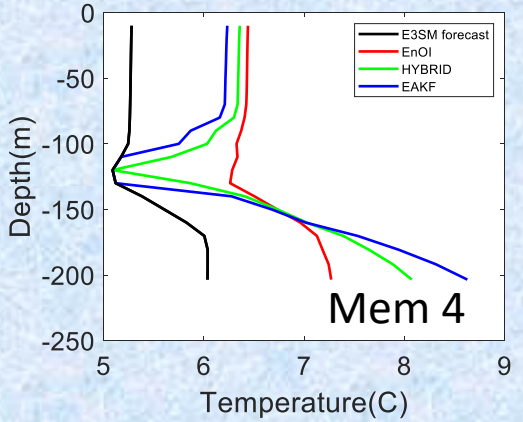
3. Results:
Test 2: Global SST assimilation

North Pacific

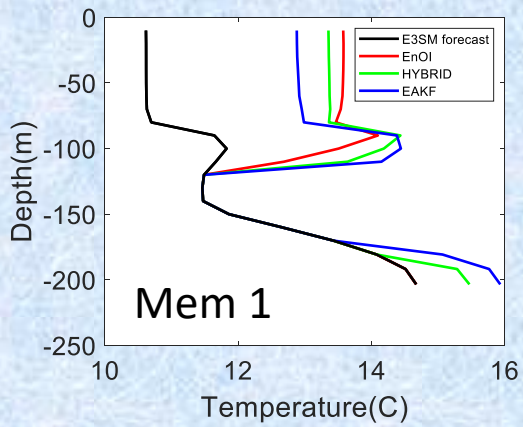


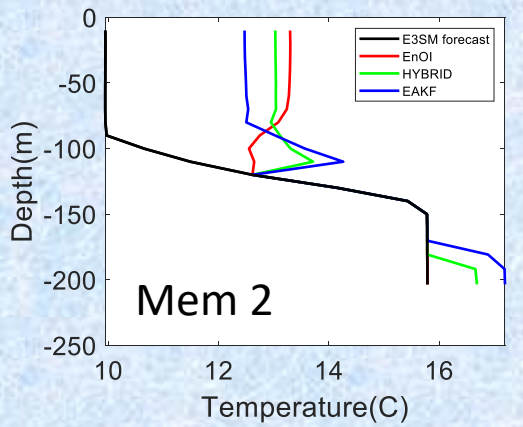


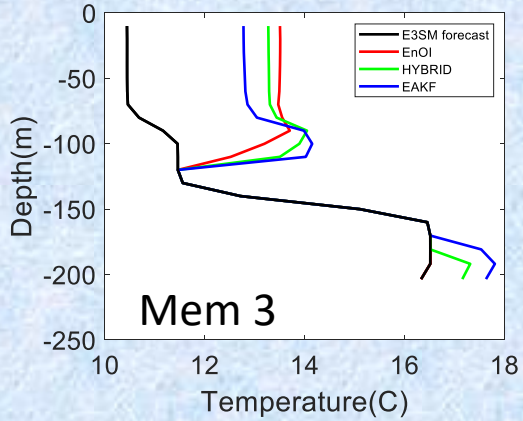


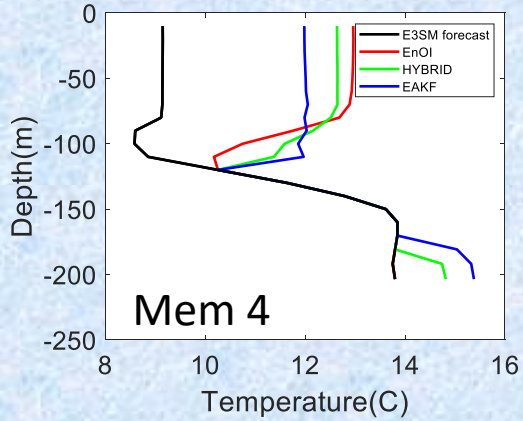


North Atlantic









Different schemes produce similar or quite different results depending on the regions.

Our HYBRID DA system is so flexible that we can make use of it to explore the proper schemes for model initialization or bias reduction.

4. Efficiency Comparison: Online vs Offline DA system

No DA Run

Simulation length = 30 days, 8 members, daily output	Online	Offline	Offline/Online
Physical time*	~200 minutes	~420 minutes	~2 times
Number of Nodes needed	125 nodes	31 nodes * 8 members = 248 nodes	~2 times
charged	33253	134582	~4 times

DA Run**
Offline/Online ~4 times

*Assuming 8 offline members can run simultaneously and there is no waiting time.
** This number depends on the amount of observations assimilated and parallel settings of both online and offline data package.

Summary

- A Real Time Online Hybrid Data Assimilation System has been developed for E3SM model, with reasonable testing results.

- Our results convinced the significant improvement (4x) of the efficiency of online over offline data assimilation system.

Perspective

- Coupled DA system is critical for the improvement of forecasting capability --- A primary future objective of E3SM.

- Online Ensemble DA system is a high-workload task

  -- efficient in the future GPU architecture.

Thanks!

2. Implementation of an online DA system into E3SM

DA Cost function:

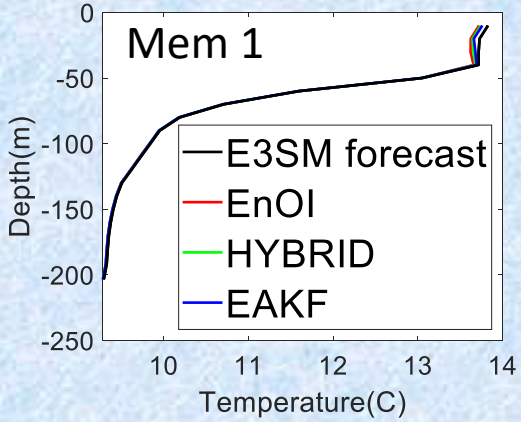
$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}]),$$

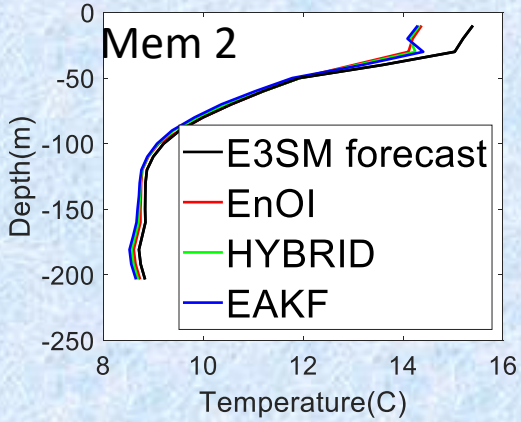
B matrix is the key

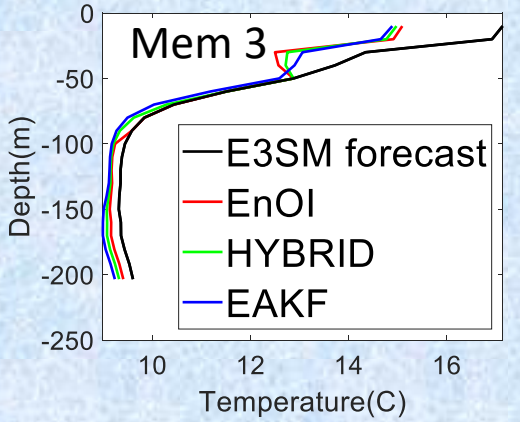
For EnOI B_EnOI is stationary/unchanged. A modified version has seasonal cycle.
For EnKF B_EnKF is estimated by ensemble vectors and changes with time
For HYBRID B_HYBRID = alpha*B_EnKF+(1-alpha)*B_EnOI e.g. (Hamill 2000)

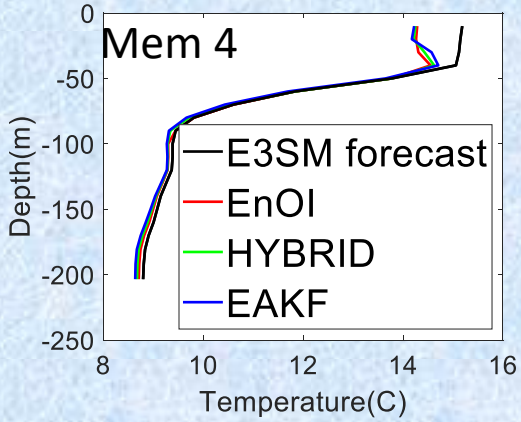
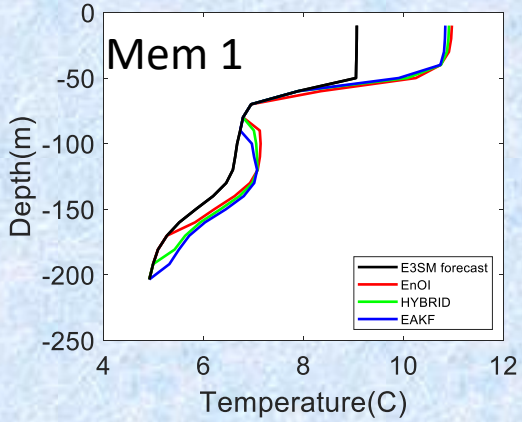
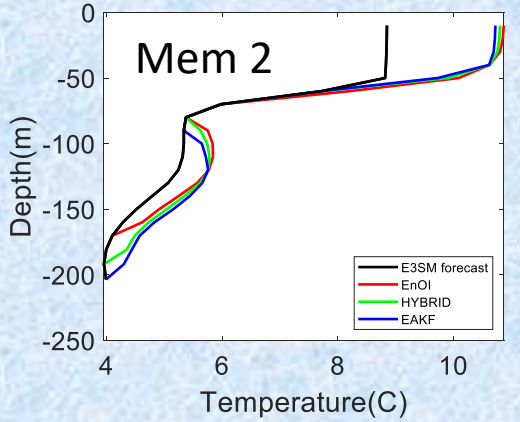
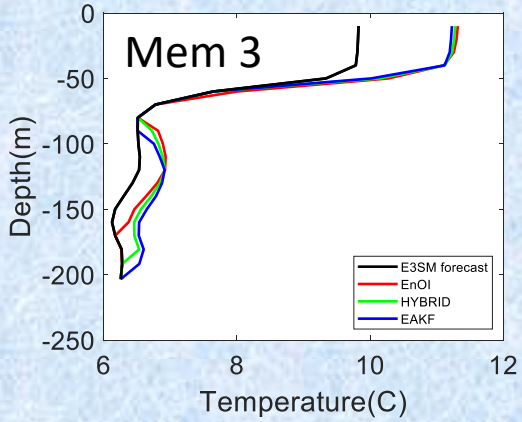
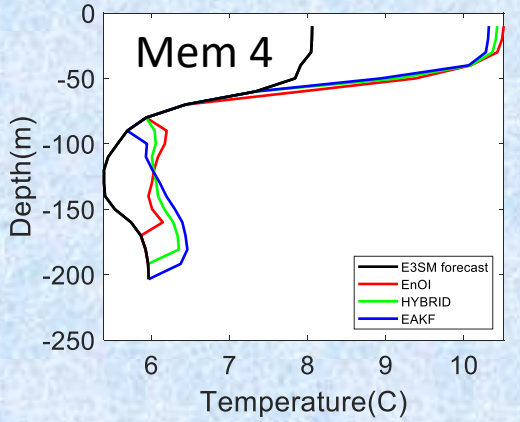
3. Results:
Test 2: Global SST assimilation

South Pacific.









South Atlantic.