Strongly Coupled Data Assimilation with a Linear Inverse Model

[using emulators to rapidly prototype coupled-DA strategies for S2S forecasting]

Gregory Hakim[1], Lindsey Taylor[1], Chris Snyder[2], Matthew Newman[3], and Steven Penny[4]

[1]University of Washington, [2]NCAR, [3]NOAA ESRL – NOAA/ESRL, [4]Sofar Ocean Technologies

Research sponsored by NOAA NWS/STI Award NA20NWS4680053

Strongly Coupled Data Assimilation (SCDA)

- Observation assimilation updates all modeled components of Earth system
- **SCDA**: Extremely computationally demanding (e.g. Liu et al. 2013; Penny & Hamill 2017)
    - Coupled forecast and “strongly coupled” DA: cross-medium updates
- “Weakly coupled” approximations (**WCDA**) (e.g. Saha et al. 2006; Zhang et al. 2007; Penny et al. 2019)
    - Separate DA in atmosphere & ocean with a <u>coupled forecast step</u>
    - No “cross covariance” influence from observations
    - Potentially “incompatible” states
    - Still very computationally demanding

Is SCDA Worth the Expense?

Need an efficient way to **estimate the benefits**

We propose a framework for this estimate using a **linear emulator**

- Low dimensionality promotes experimentation/prototyping
- Unlike “toy” models, the emulator skillfully forecasts observed fields
- Allows unapproximated Kalman filter, and WCDA-SCDA evaluation

Linear Inverse Models (LIMs)

e.g. Penland (1989); Newman et al. (2003); Breeden et al. (2020)

$$\frac{d\mathbf{x}}{dt} = \mathcal{N}(\mathbf{x}) \approx \mathbf{L}\mathbf{x} + \xi \quad \rightarrow \quad \mathbf{x}(\tau) = \mathbf{G}_\tau \mathbf{x}_0 + \epsilon \qquad \mathbf{G}_\tau = e^{\mathbf{L}\tau}$$

Solve for G empirically from sample data:

$$\mathbf{G}_\tau = cov(\mathbf{x}_\tau, \mathbf{x}_0)(cov(\mathbf{x}_0, \mathbf{x}_0))^{-1}$$

and the noise error-covariance matrix

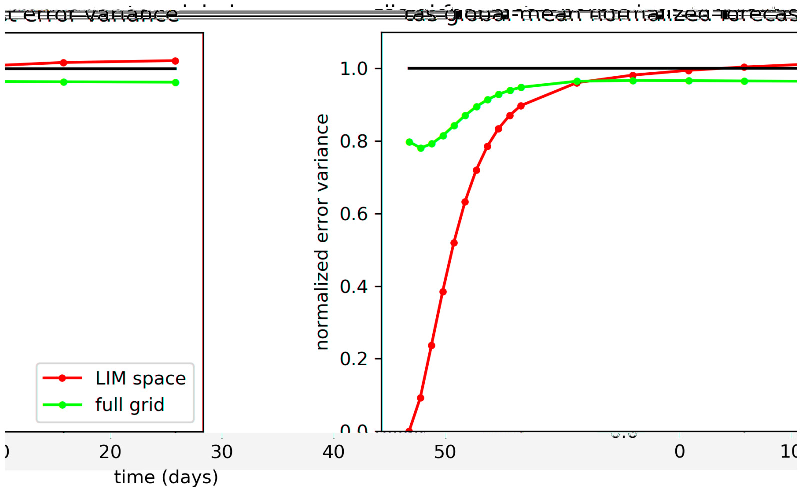
$$cov(\epsilon, \epsilon) = \mathbf{N}_\tau = \mathbf{C} - \mathbf{G}_\tau \mathbf{C} \mathbf{G}_\tau^{\mathrm{T}} \qquad \mathbf{C} = cov(\mathbf{x}_0, \mathbf{x}_0)$$

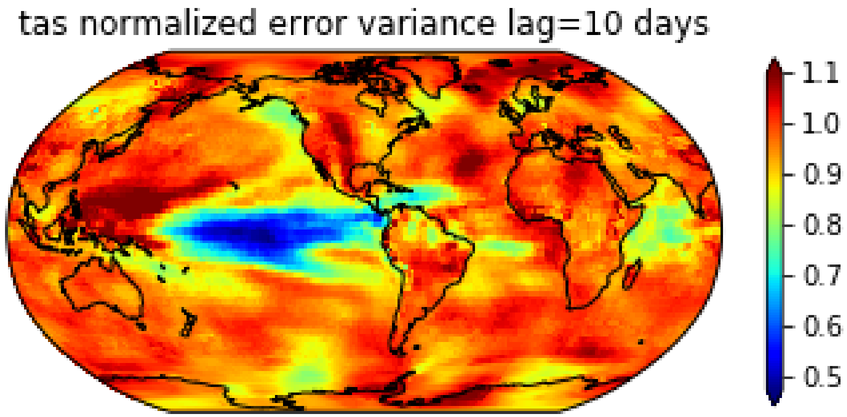
LIM Training

- Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010)
- Global gridded fields of **2m air temperature** (T_{2m}), **SST**, 850 hPa **u** & **v**, and **OLR**
  - diurnal & 5-day running mean; seasonal cycle removed
- Truncate to leading 30 EOFs for each variable (150 degrees of freedom in total)
- LIM **training** period: 1979-2003 (9130 days)
- LIM **validation** period: 2004-2010 (2556 days)

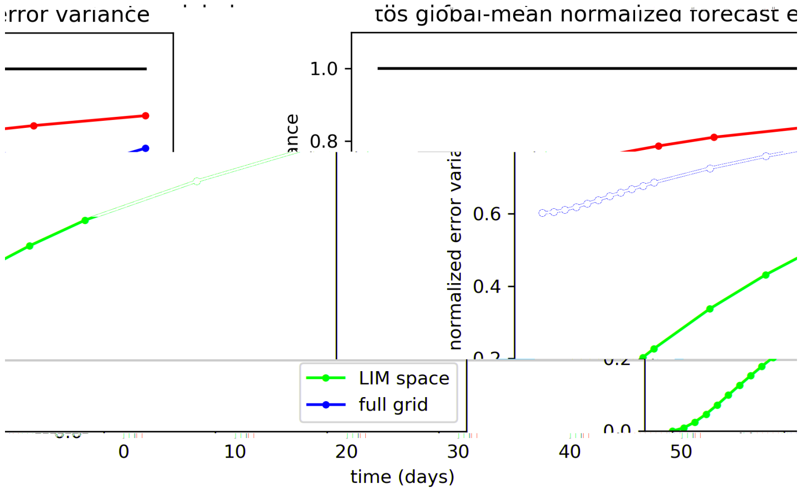
LIM forecast verification 2004-2010

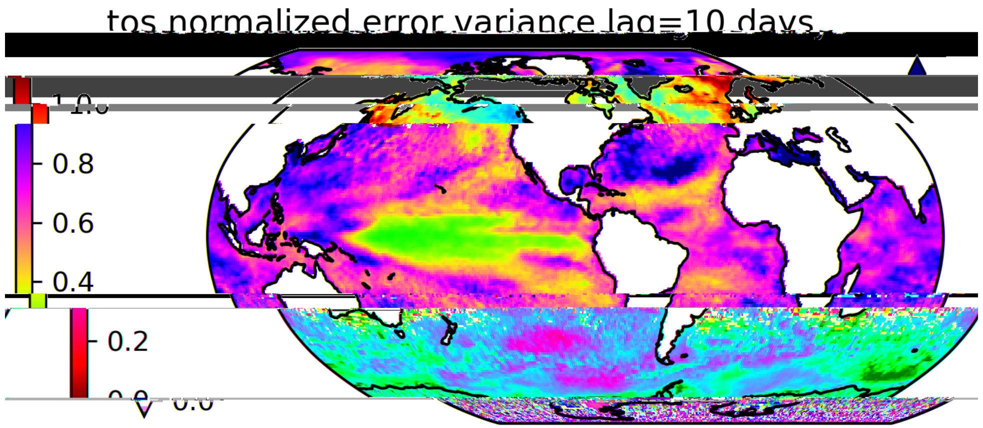
T_{2m}





SST





Kalman Filter using the LIM

analysis

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_f)$$

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_f$$

$$\mathbf{K} = \mathbf{P}_f\mathbf{H}^{\mathrm{T}}\left[\mathbf{H}\mathbf{P}_f\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right]^{-1}$$

forecast

full matrices!

$$\mathbf{x}_f = \mathbf{G}_t\mathbf{x}_a$$

$$\mathbf{P}_f = \mathbf{G}_t\mathbf{P}_a\mathbf{G}_t^{\mathrm{T}} + \mathbf{N}_t$$

Cycling time t = 1 day

Pointwise observations from CFSR

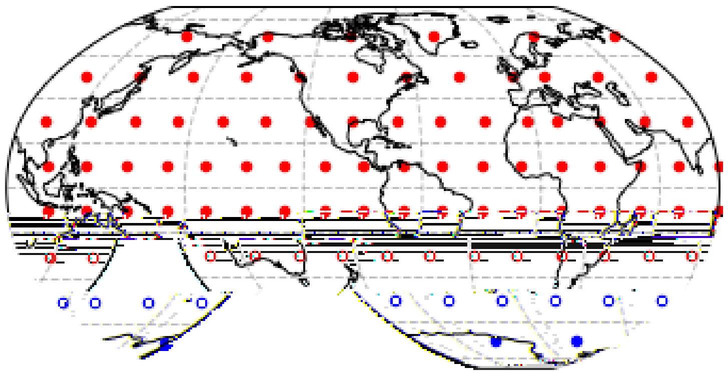
$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} = \underbrace{\mathbf{H}^*\mathbf{U}\mathbf{x}}_{\text{obs from LIM basis}} + \underbrace{\mathbf{H}^*\eta}_{\text{truncation error}}$$

$\mathbf{H}^*$: obs operator on CFSR grid

R derived from truncation error

T_{2m}

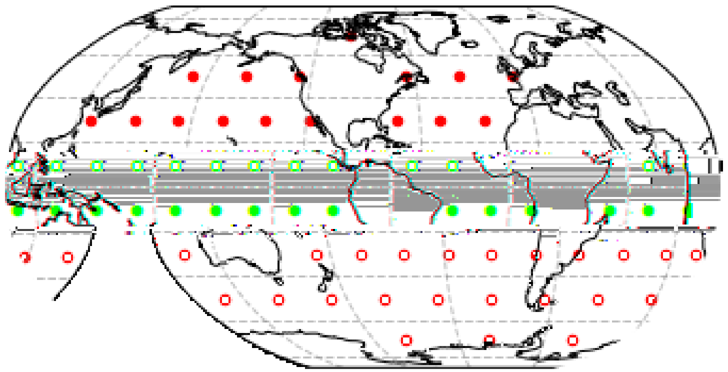
tas observation locations



104 locations

SST

tos observation locations

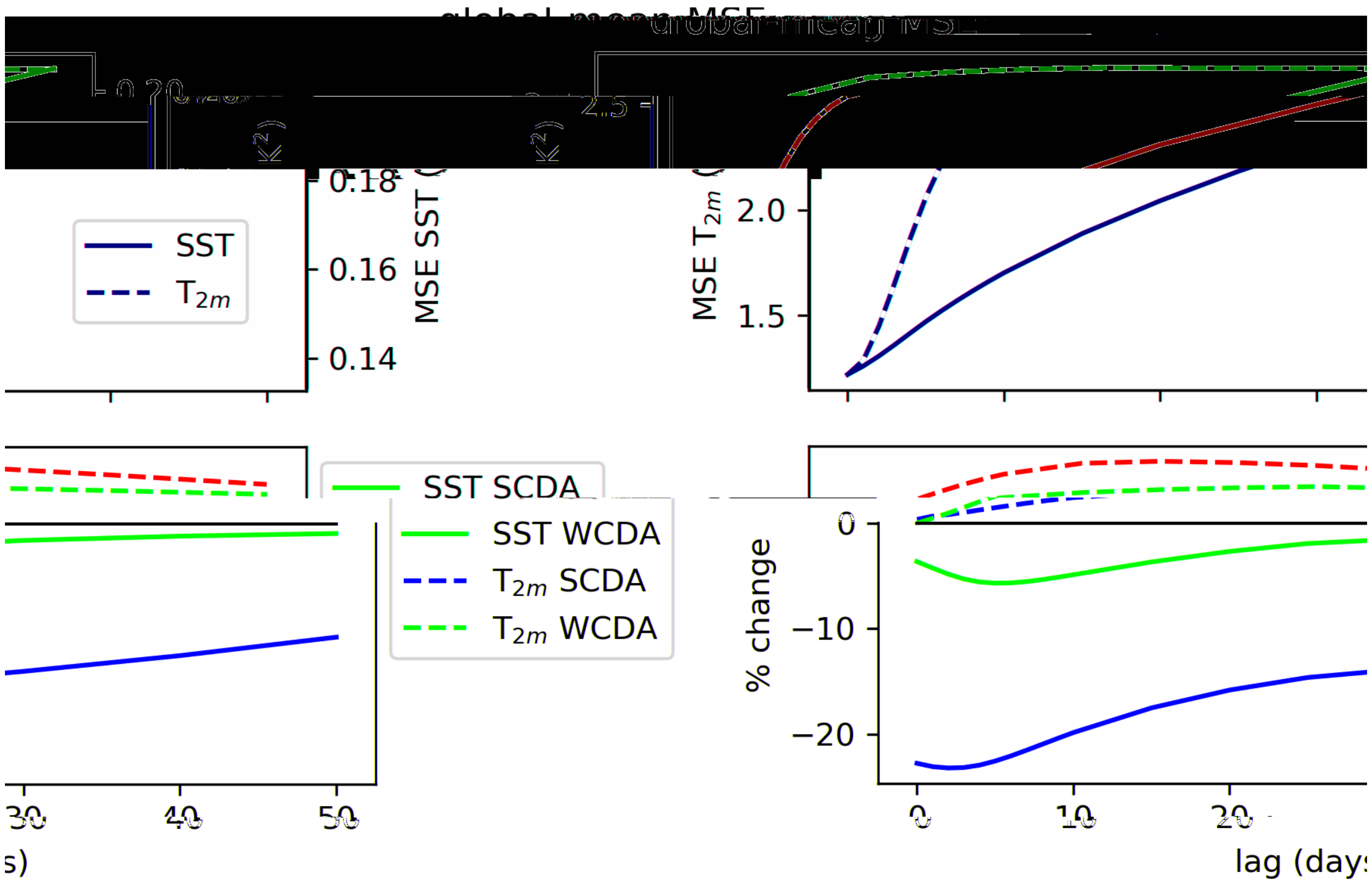


73 locations

DA Experiments

1. Uncoupled: **atmosphere**
    - Atmosphere components of the LIM & atmosphere observations
2. Uncoupled: **ocean**
    - Ocean components of the LIM & ocean observations
3. Weakly coupled (**WCDA**)
    - Full LIM; separate DA in atmosphere & ocean; no cross covariances
4. Strongly coupled (**SCDA**)
    - Full LIM; fully coupled DA

Global-mean Mean Squared Error



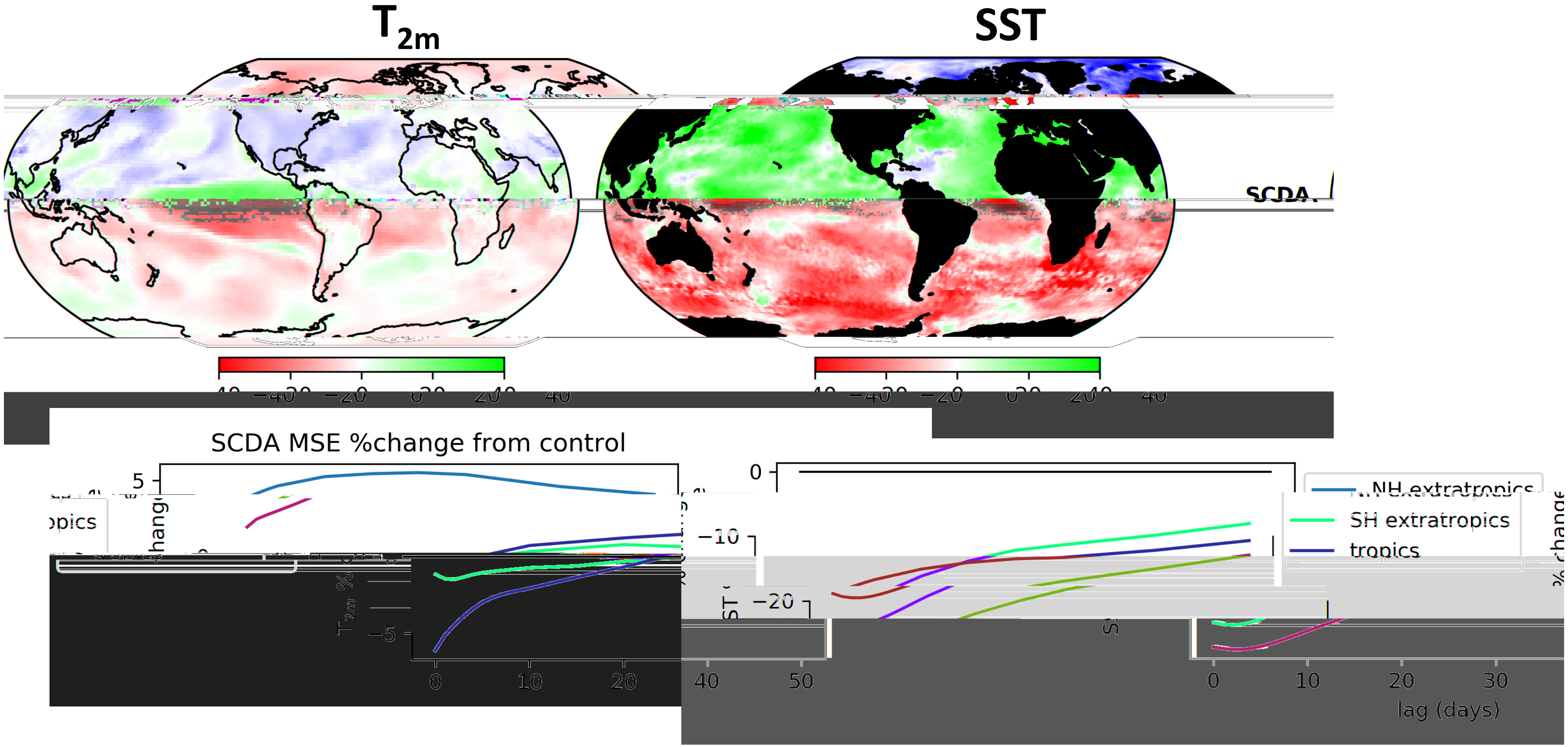
uncoupled (**control**) experiments

WCDA & **SCDA**

change from control experiments

- <u>large positive impact on SST</u>
- weak adverse impact on T_{2m}

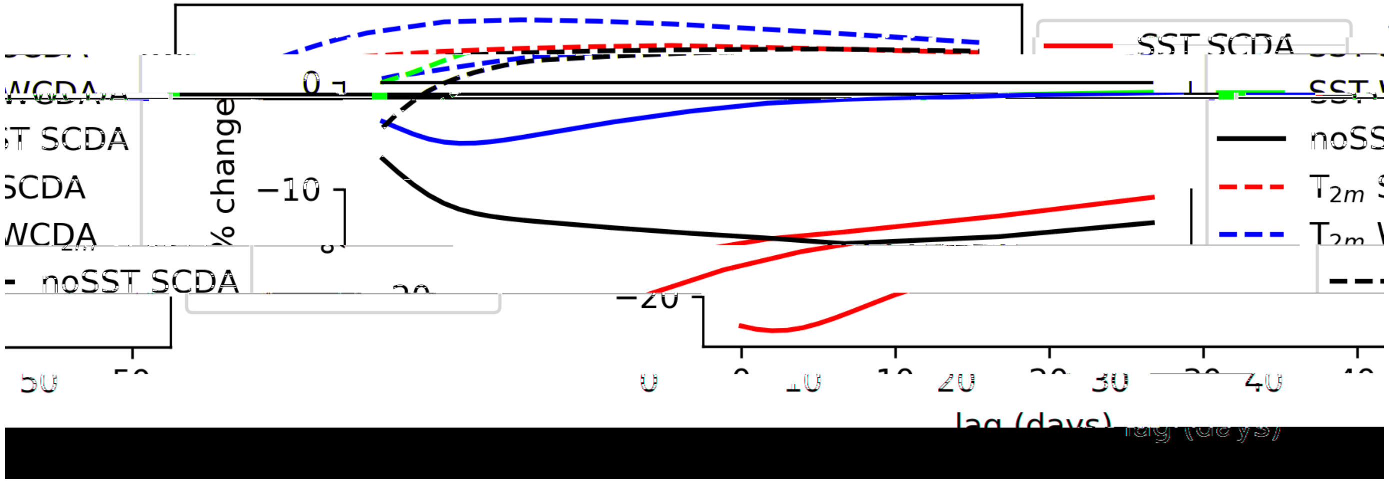
SCDA Change from Uncoupled 10-day Forecasts



- improved tropics
- degraded NH extratropics

- global improvement

SCDA without SST observations

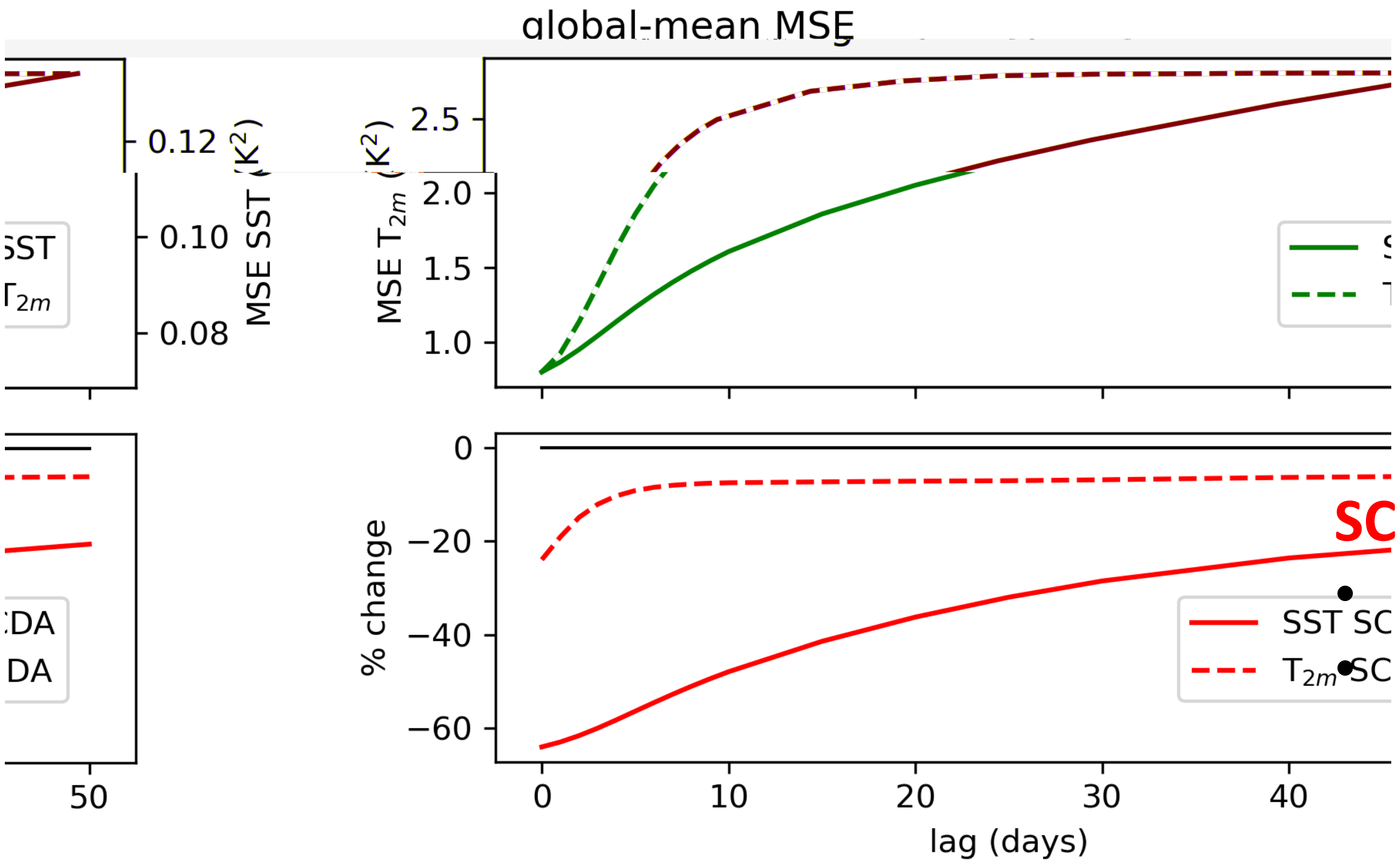


WCDA & SCDA

change from control experiments

- better T2m analyses
- SST analyses still better than control
- long-lead forecasts similar

SCDA Results During the Training Period



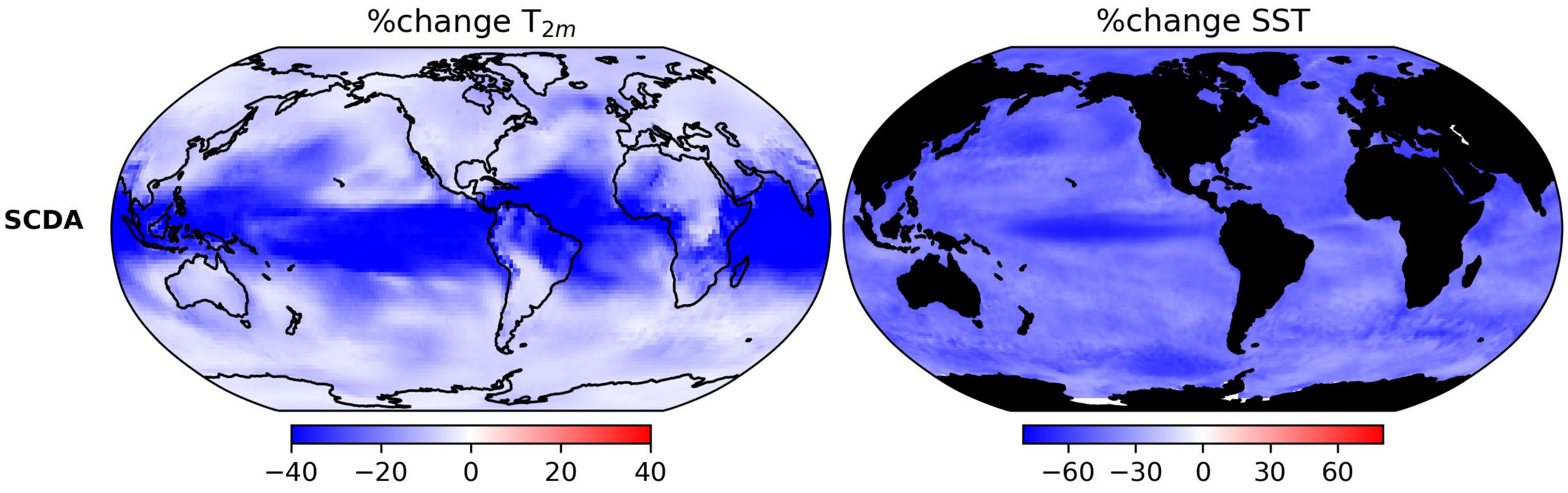
uncoupled (**control**) experiments

- smaller errors than validation

SCDA change from control experiment

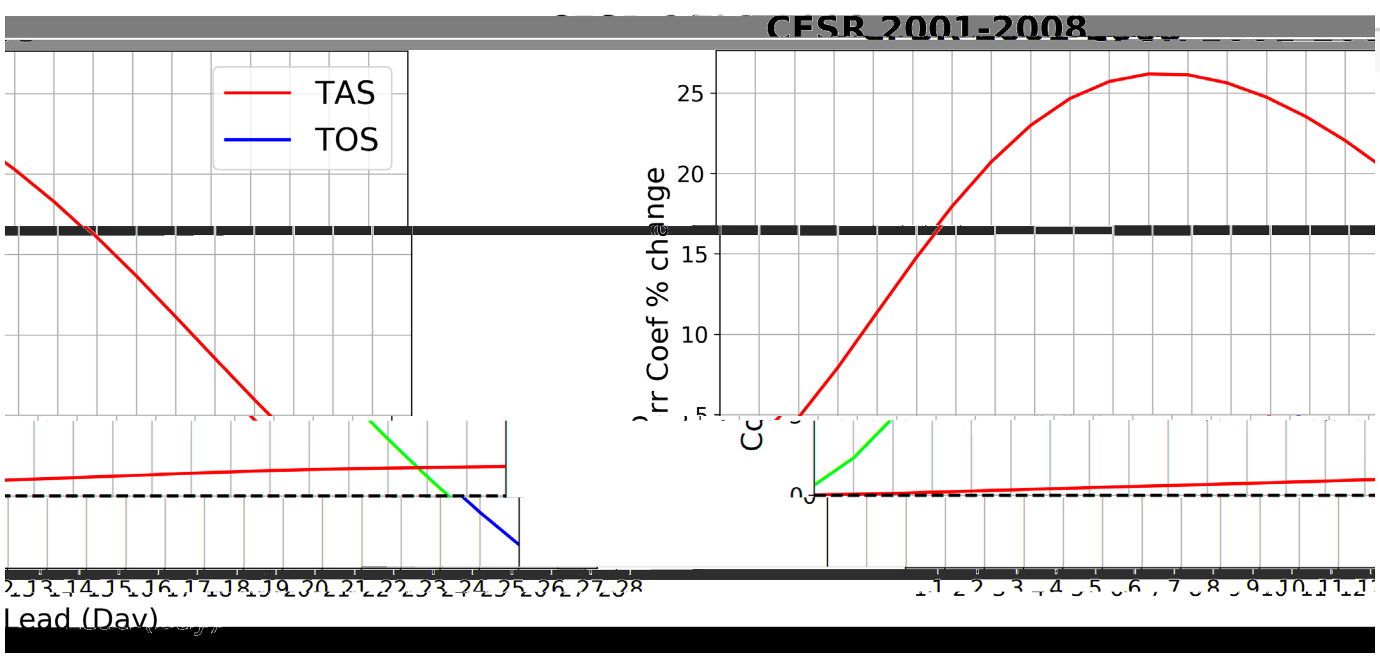
- large improvements
- non-stationary climate

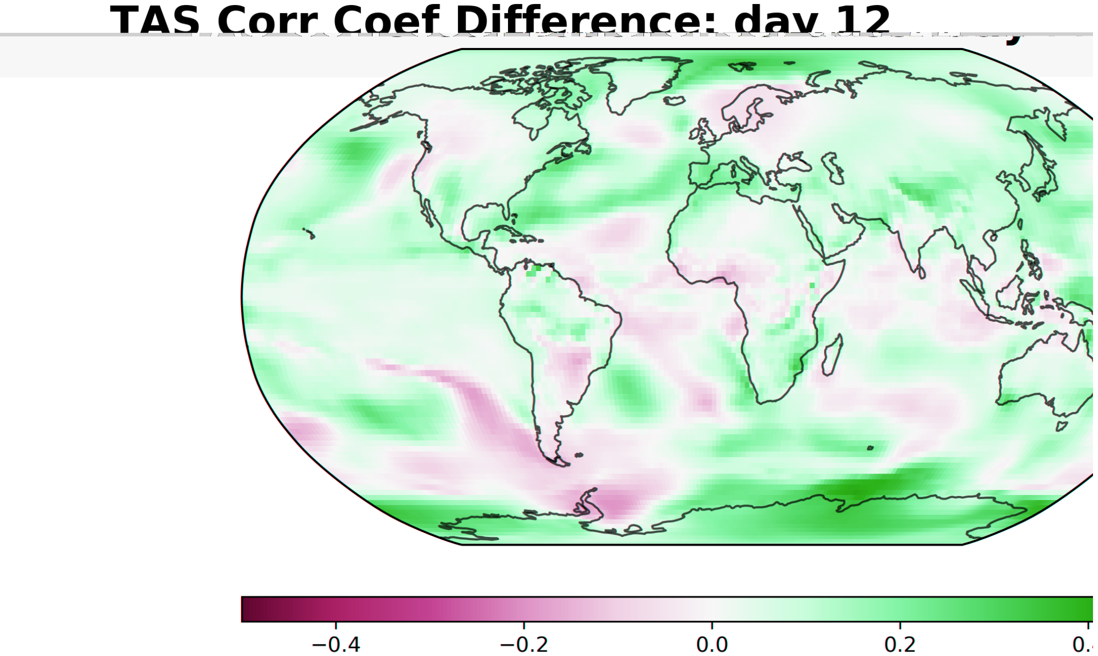
SCDA Training Period 10-day Forecast Change



Spatial pattern of % change in MSE relative to the control case.

Experiments with larger training data: JRA55





- JRA LIM trained 1958-2000
- CFSR LIM trained 1980-2000
- Validation: CFSR 2001-2008

Lindsey Taylor (UW)

Summary: LIM DA **Proof of Concept**

- **LIMs provide a flexible tool for coupled-DA hypothesis testing**
  - Kalman filtering, 4DVAR, ensemble strategies, etc.
- **SCDA performs better than WCDA in all experiments**
- **Relative to control, single-domain, experiments, SCDA:**
  - improves SST forecasts: ~10–25% better in global mean; 40+% locally
  - improves tropical T_{2m} forecasts; degrades extratropical NH T_{2m} forecasts
  - large T_{2m} improvement during training does not hold up in validation
- Currently exploring other LIM bases not truncated on EOFs; ensembles; seasonality

Hakim, G. J., C. Snyder, S. G. Penny, and M. Newman, 2022: Subseasonal forecast skill improvement from strongly coupled data assimilation with a linear inverse model. *Geophysical Research Letters*, **49,** e2022GL097996.
DOI: 10.1029/2022GL097996

Thank you!