Wunsch and Heimbach 2013:
Dynamically and Kinematically Consistent Global Ocean Circulation and Ice State Estimates

SIO 219 winter 2014

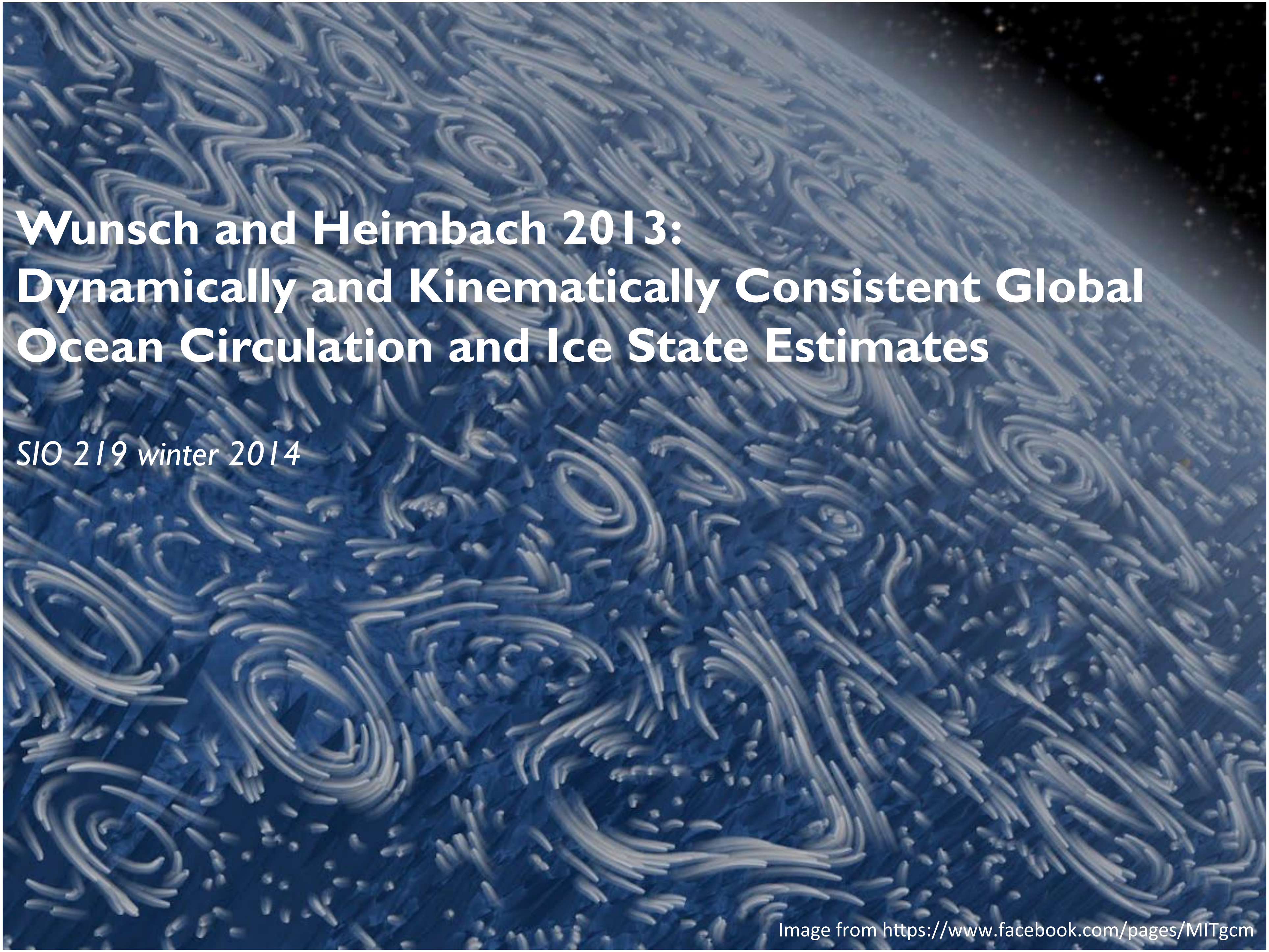
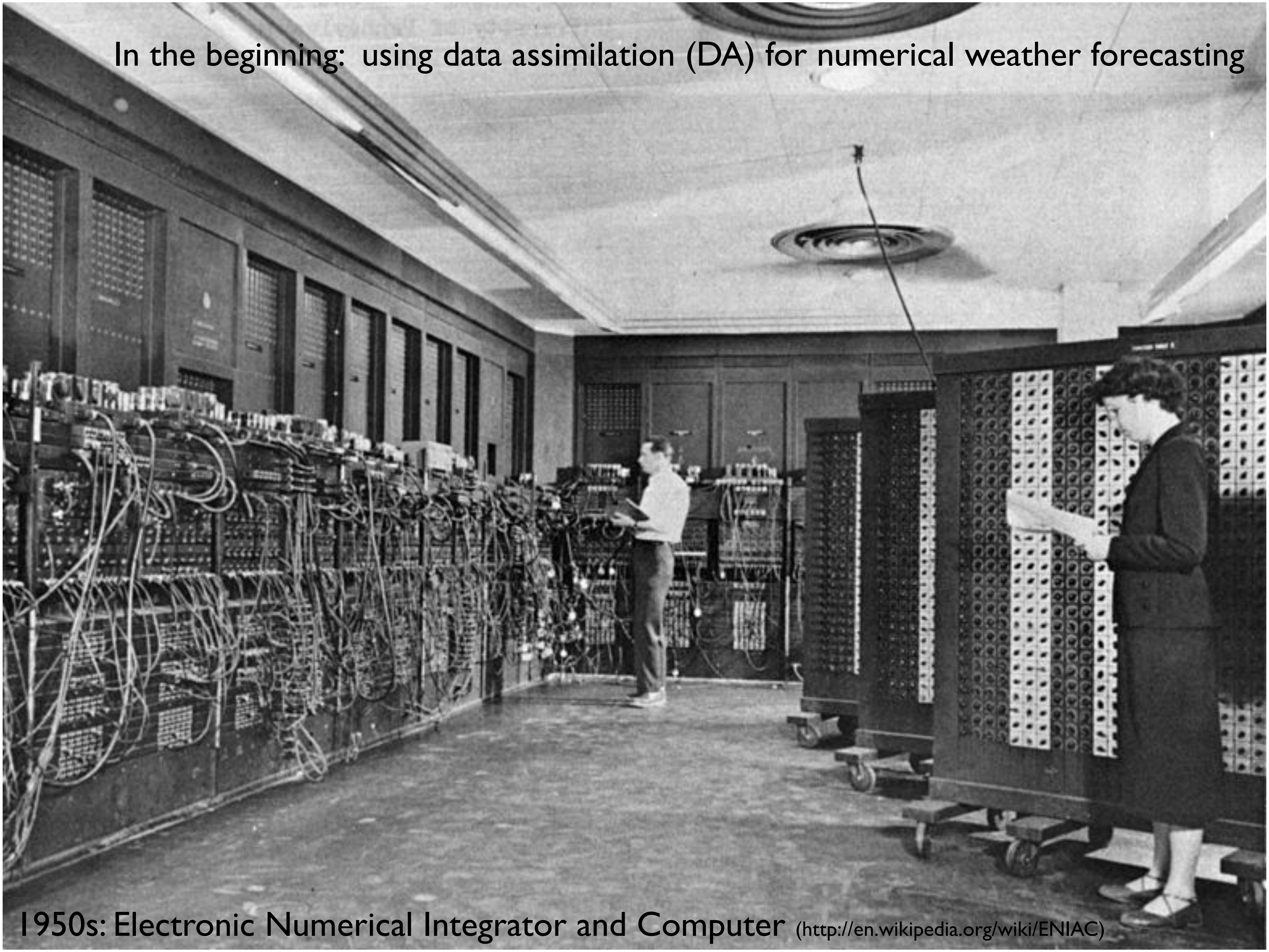


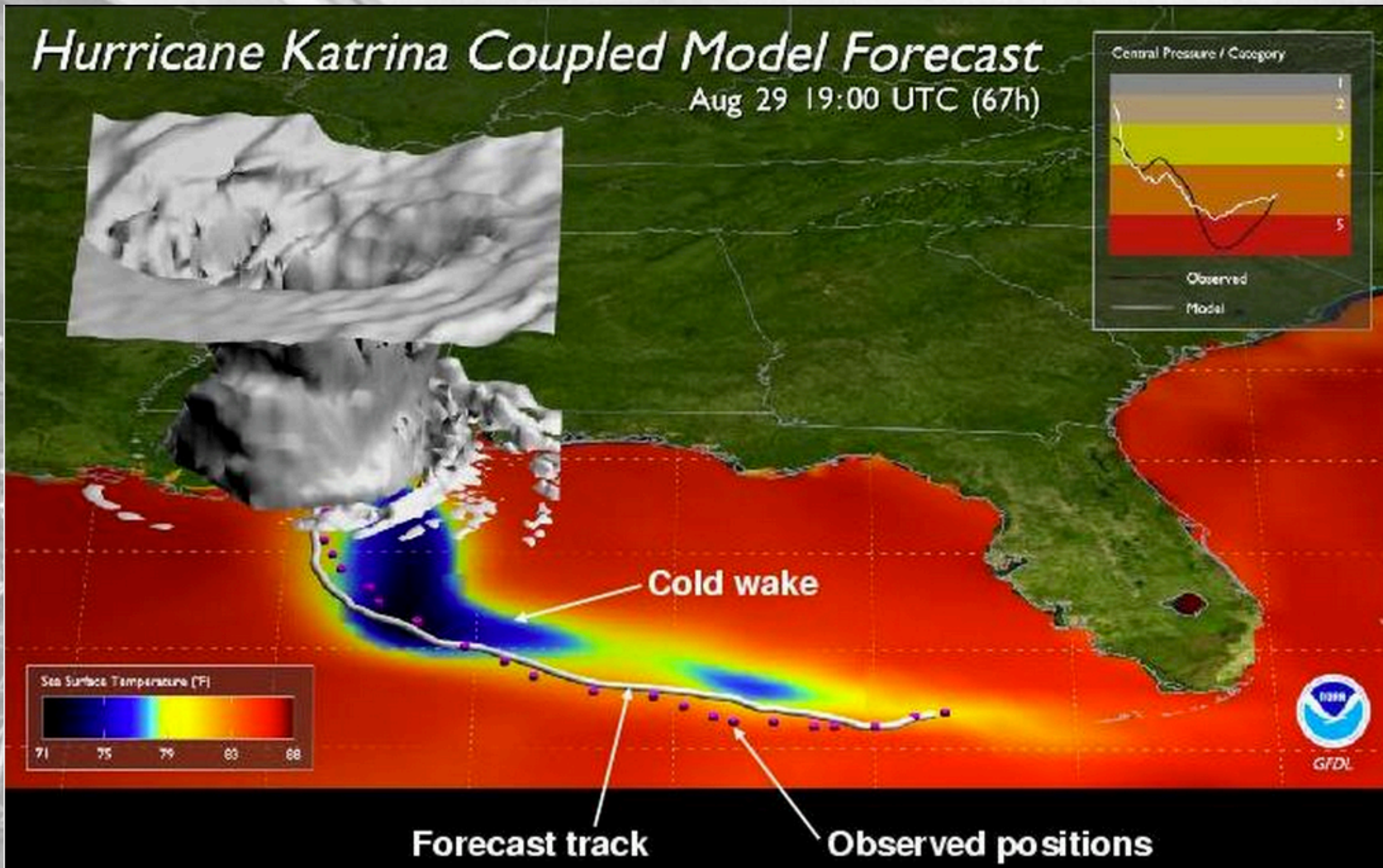
Image from https://www.facebook.com/pages/MITgcm

In the beginning: using data assimilation (DA) for numerical weather forecasting



1950s: Electronic Numerical Integrator and Computer (http://en.wikipedia.org/wiki/ENIAC)

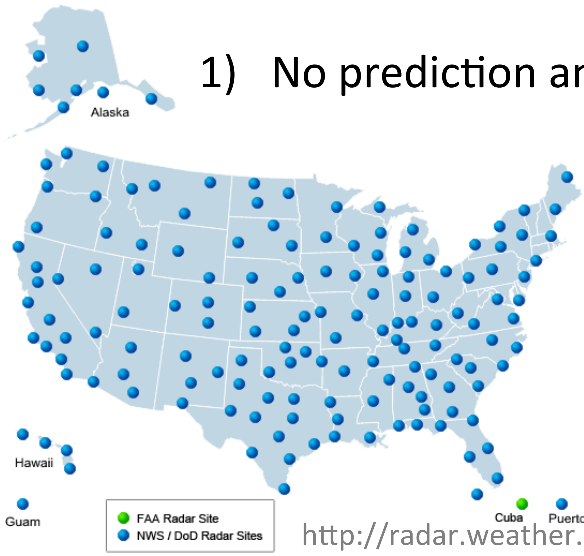
In the beginning: using data assimilation (DA) for numerical weather forecasting



Today: NOAA GFDL Hurricane Model (http://www.gfdl.noaa.gov/operational-hurricane-forecasting)

1950s: Electronic Numerical Integrator and Computer (http://en.wikipedia.org/wiki/ENIAC)

Why can't we do this for the ocean?



1) No prediction and observation infrastructure like the National Weather Service.

http://radar.weather.gov/index_li

2) Oceanography goals that cannot be met via short-term prediction methods:

- *Long time-scale predictions*
- *Understanding governing physics*
- *Closing heat, energy, water budgets.*

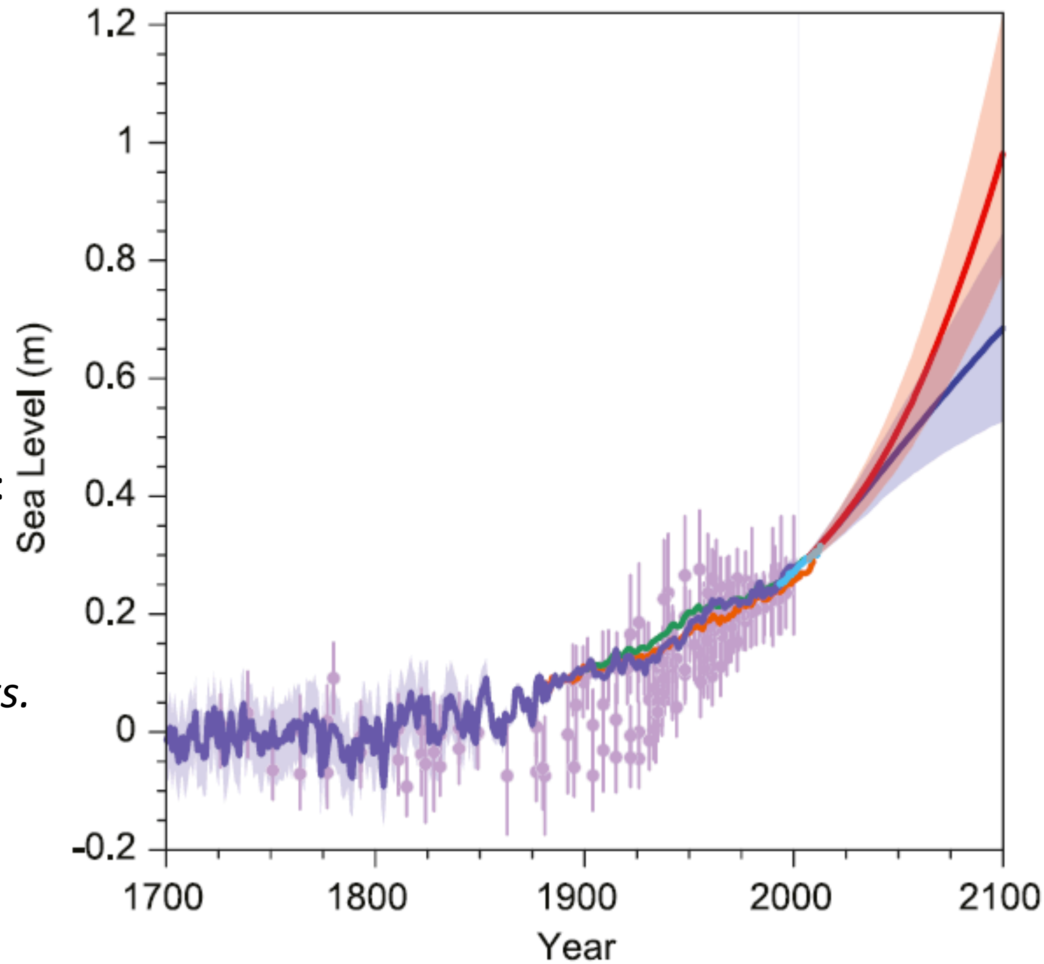
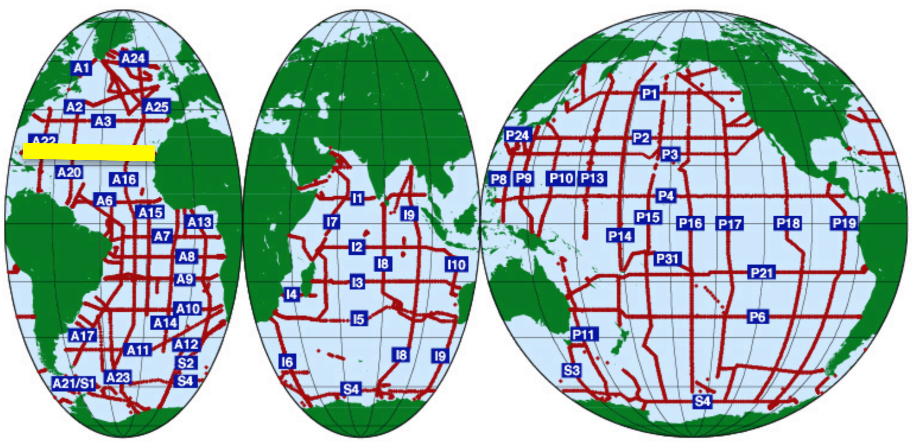
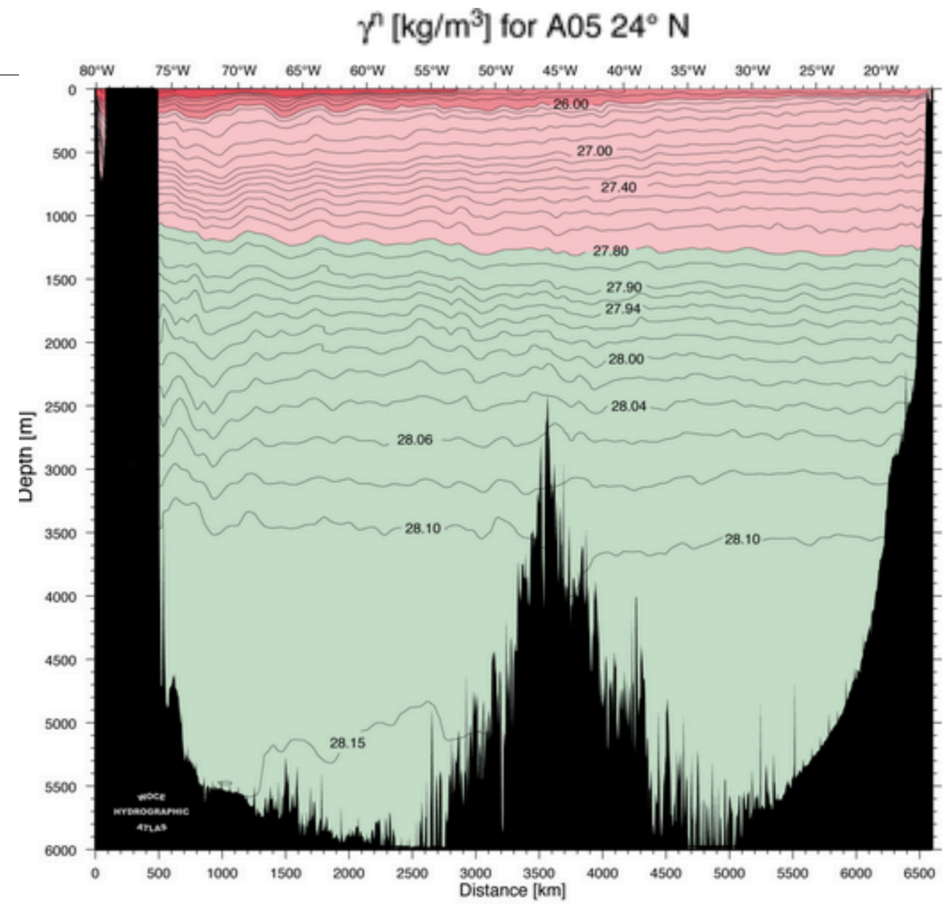


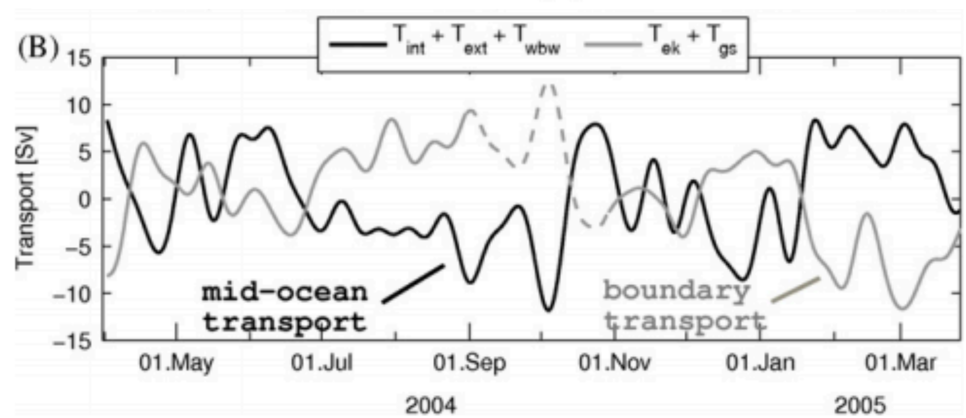
Fig. 1. Past and future sea-level rise. For the past, proxy data are shown in light purple and tide gauge data in blue. For the future, the IPCC projections for very high emissions (red, RCP8.5 scenario) and very low emissions (blue, RCP2.6 scenario) are shown. Source: IPCC AR5 Fig. 13.27.

World Ocean Circulation Experiment

Largest sources of error in ‘synoptic’ transects: temporal variability.
Thus ocean must be treated as a *fundamentally time-varying system*.







15-day low-pass filtered transport fluctuations across A05 (Rayner et al., 2011)

Generic model of a physical system: model state

$$\mathbf{x}(t) = L(\mathbf{x}(t-\Delta t), \mathbf{q}(t-\Delta t), \mathbf{u}(t-\Delta t))$$

where:

t is time, discrete at intervals *Δt*

x(t) is the model state at time *t*

x(t-Δt) is the model state at the previous time interval

q(t-Δt) are the known forcings, sources, sinks, boundary and initial conditions, and internal model parameters, at the previous time interval

u(t-Δt) are control variables: any such elements that are regarded as only partly or wholly unknown, hence subject to adjustment; and model errors

L is an operator and can involve any mathematically defined function. In practice, a computer code working on arrays of numbers.

Generic model of a physical system: model 'observation'

$$\mathbf{y}(t) = \mathbf{E}(t)\mathbf{x}(t) + \mathbf{n}(t)$$

where:

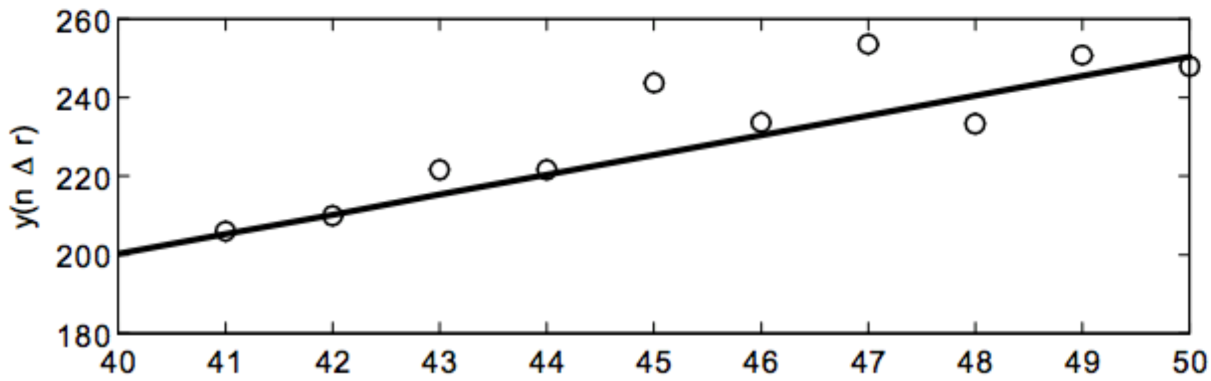
t is time, discrete at intervals *Δt*

y(t) is an observation at time *t*

x(t) is the model state at time *t*

n(t) is the noise in the observations at time *t*

This direction leads us to consider ordinary least-squares as one practices it in beginning science courses. Many of the conceptual issues can be understood from the most elementary problem of fitting a straight line to data:



We can write the problem as above:

$$\frac{d^2y}{dt^2} = 0,\ y\left(41\Delta r\right) = y_{41} \pm \varepsilon_{41},\ y_{42} \pm \varepsilon_{42}, \ldots,$$

($r = 41, 42, \ldots$,is totally arbitrary). Could discretize the equation as above,

$$y\left(r + \Delta r\right) - 2y\left(r\right) + y\left(r - \Delta r\right) = 0 \tag{4}$$

and is a set of simultaneous equations, although with errors in some of them (the ones involving the observations, and maybe the model isn't perfect either).

Generic model of a physical system: straight line example (2)

Alternatively, we can reformulate it as

$$y = a + bt$$

which reduces the number of unknowns to 2 instead of all of $y(41)$, $y(42)$, etc.
(Am setting $\Delta r = 1$.)
Thus

$$\begin{aligned} a + 41b &= y(41) \\ a + 42b &= y(42) \\ &.. \\ a + (40+N)b &= y(40+N) \end{aligned}$$

or

$$\left\{\begin{array}{cc} 1 & 41 \\ 1 & 42 \\ . & . \\ . & . \\ 1 & 40+N \end{array}\right\} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y(41) \\ y(42) \\ . \\ . \\ y(40+N) \end{bmatrix}$$

or

$$\mathbf{Ex} = \mathbf{y}$$

Generic model of a physical system: straight line example (3)

but which from the graph we know is contradictory. No straight line will produce a solution in the mathematical sense unless $N = 2$. Should really write it,

$$\mathbf{Ex} \approx \mathbf{y} \tag{5}$$

Generic model of a physical system: straight line example (4)

But equations are much easier to deal with than constructs like, (5), so convert it,

$$\mathbf{Ex}+\varepsilon=\mathbf{y}$$

where ε represents the noise. It's still just a set of simultaneous equations except now we *could* write it as,

$$\left\{\begin{array}{cccccc} 1 & 41 & 1 & 0 & . & 0 \\ 1 & 42 & 0 & 1 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 40+N & 0 & 0 & 0 & 1 \end{array}\right\}\left[\begin{array}{c} a \\ b \\ \varepsilon(41)) \\ \varepsilon(42) \\ . \\ \varepsilon(40+N) \end{array}\right]=\left[\begin{array}{c} y(41) \\ y(42) \\ . \\ . \\ y(40+N) \end{array}\right]$$

which is once again

$$\mathbf{E}_1\mathbf{x}=\mathbf{y}$$

except now there are still N equations, but $N+2$ unknowns. Or use the finite form to the same end.
One is taught in school to solve this problem by minimizing $\sum \varepsilon_i^2 = \varepsilon^T\varepsilon$ (**why?**).

Constrained optimization problem

For Gaussian data and controls, we seek to minimize the scalar term:

$$J = \sum_{m=0}^{M} (\mathbf{y}(t) - \mathbf{E}(t)\widetilde{\mathbf{x}}(t))^{\mathrm{T}} \mathbf{R}^{-1}(t)(\mathbf{y}(t) - \mathbf{E}(t)\widetilde{\mathbf{x}}(t)) + \sum_{m=0}^{M-1} \widetilde{\mathbf{u}}(t)^{\mathrm{T}} \mathbf{Q}^{-1}(t)\widetilde{\mathbf{u}}(t), \quad t = m\Delta t,$$

where $\widetilde{\mathbf{x}}(t)$ and $\widetilde{\mathbf{u}}(t)$ exactly satisfy model equations

$$\mathbf{x}(t) = L(\mathbf{x}(t-\Delta t), \mathbf{q}(t-\Delta t), \mathbf{u}(t-\Delta t)) \quad \textit{and} \quad \mathbf{y}(t) = \mathbf{E}(t)\mathbf{x}(t) + \mathbf{n}(t)$$

and ***R(t)*** and ***Q(t)*** are weights given by the variances of observational noise and controls, respectively:

1st and 2nd moment controls

$$< \mathbf{u}(t) > = 0,$$
$$\left\langle \mathbf{u}(t)\mathbf{u}(t')^T \right\rangle = \mathbf{Q}(t)\delta_{tt'}$$

1st and 2nd moment observational noise

$$< \mathbf{n}(t) > = 0,$$
$$\left\langle \mathbf{n}(t)\mathbf{n}(t')^T \right\rangle = \mathbf{R}(t)\delta_{tt'}$$

DA and reanalysis products

Data assimilation seeks to minimize the variance of the state about the true value sometime in the future:

$$\mathrm{diag}\left\langle \left(\widetilde{\mathbf{x}}(t_{\mathrm{now}}+\tau)-\mathbf{x}(t_{\mathrm{now}}+\tau)\right)\left(\widetilde{\mathbf{x}}(t_{\mathrm{now}}+\tau)-\mathbf{x}(t_{\mathrm{now}}+\tau)\right)^{\mathrm{T}}\right\rangle$$

Thus optimization problem seeks to minimize scalar

$$\begin{aligned} J_1 = &\left(\widetilde{\mathbf{x}}(t_{\mathrm{now}})-\widetilde{\mathbf{x}}(t_{\mathrm{now}},-)\right)^{\mathrm{T}}\mathbf{P}(t_{\mathrm{now}},-)^{-1}\left(\widetilde{\mathbf{x}}(t_{\mathrm{now}})-\widetilde{\mathbf{x}}(t_{\mathrm{now}},-)\right) \\ &+\left(\mathbf{y}(t_{\mathrm{now}})-\mathbf{E}(t_{\mathrm{now}})\mathbf{x}(t_{\mathrm{now}})\right)^{\mathrm{T}}\mathbf{R}(t_{\mathrm{now}})^{-1}\left(\mathbf{y}(t_{\mathrm{now}})-\mathbf{E}(t_{\mathrm{now}})\mathbf{x}(t_{\mathrm{now}})\right), \end{aligned}$$

Where

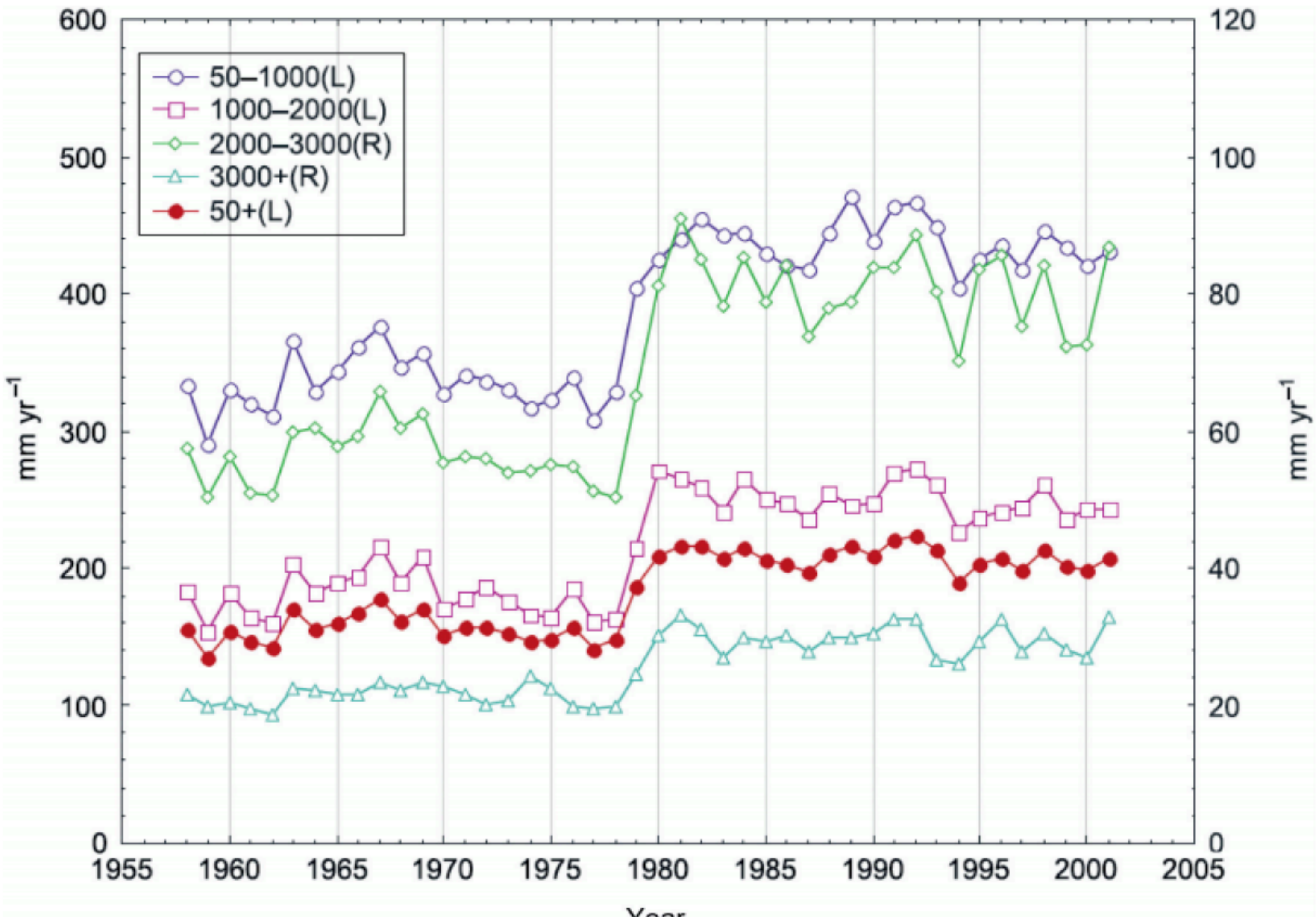
$\widetilde{\mathbf{x}}(t_{\mathrm{now}},-)$ is a future prediction

$\mathbf{P}(t_{\mathrm{now}},-)$ is the data variance; and

$\mathbf{R}(t_{\mathrm{now}})$ is the observational noise variance

Issues with reanalysis products

1) Spurious trends due to changes in the observational systems



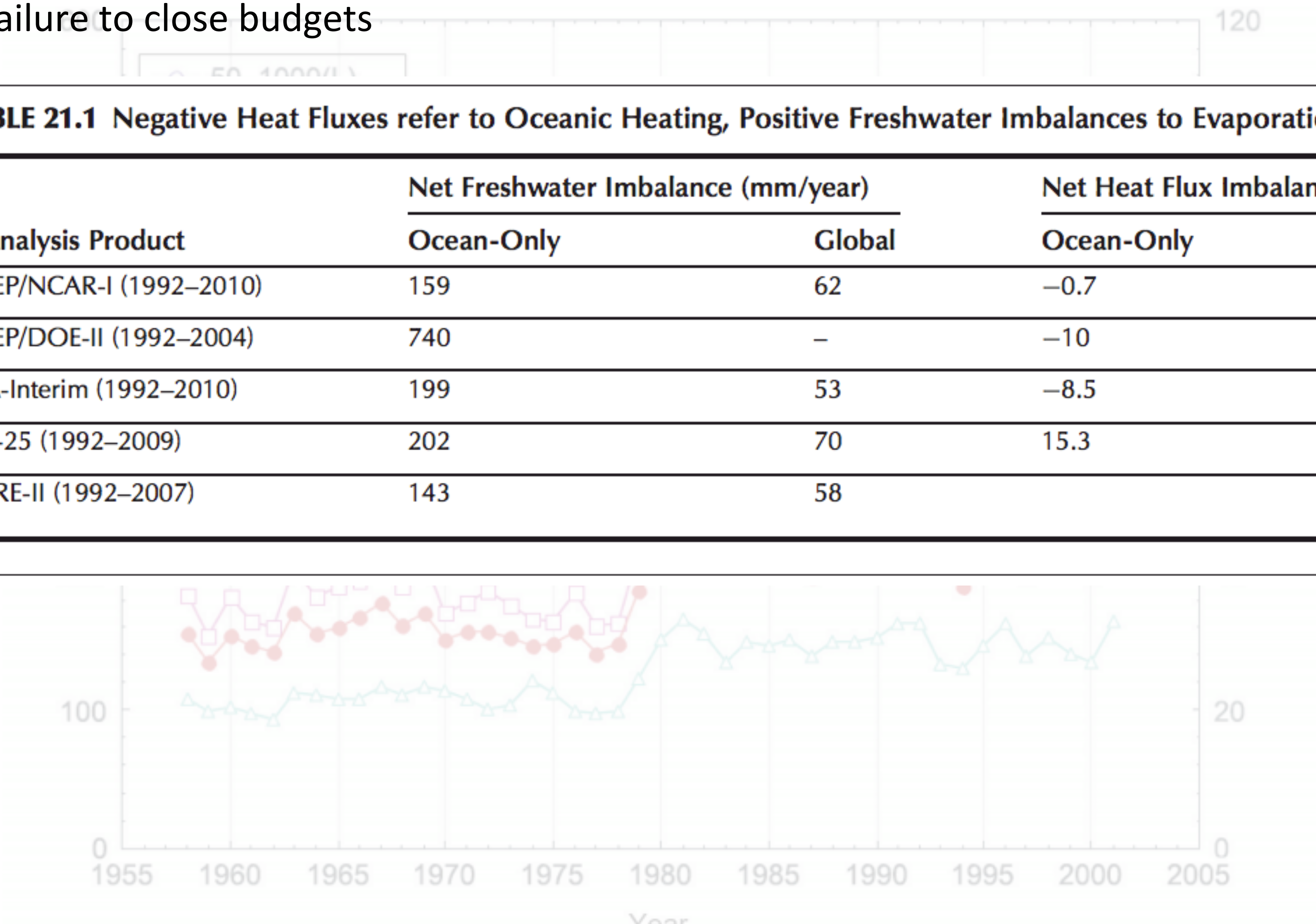
Mean annual precipitation minus evaporation over the Antarctic as a function of time in the ECMWF reanalysis ERA-40

Issues with reanalysis products

1) Spurious trends due to changes in the observational systems
2) Failure to close budgets

TABLE 21.1 Negative Heat Fluxes refer to Oceanic Heating, Positive Freshwater Imbalances to Evaporation

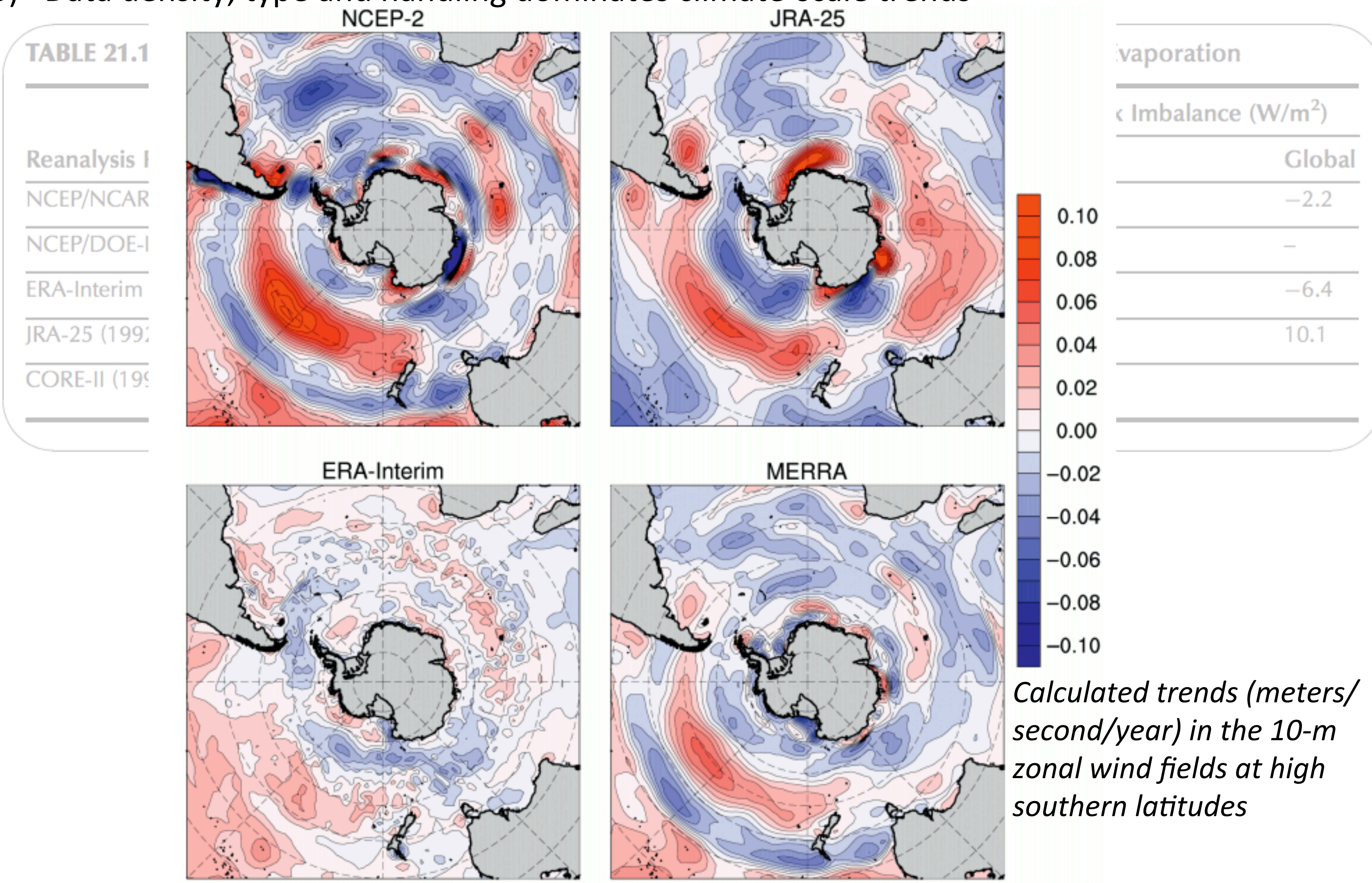
	Net Freshwater Imbalance (mm/year)		Net Heat Flux Imbalance (W/m^2)	
Reanalysis Product	**Ocean-Only**	**Global**	**Ocean-Only**	**Global**
NCEP/NCAR-I (1992–2010)	159	62	−0.7	−2.2
NCEP/DOE-II (1992–2004)	740	–	−10	–
ERA-Interim (1992–2010)	199	53	−8.5	−6.4
JRA-25 (1992–2009)	202	70	15.3	10.1
CORE-II (1992–2007)	143	58		



Mean annual precipitation minus evaporation over the Antarctic as a function of time in the ECMWF reanalysis ERA-40

Issues with reanalysis products

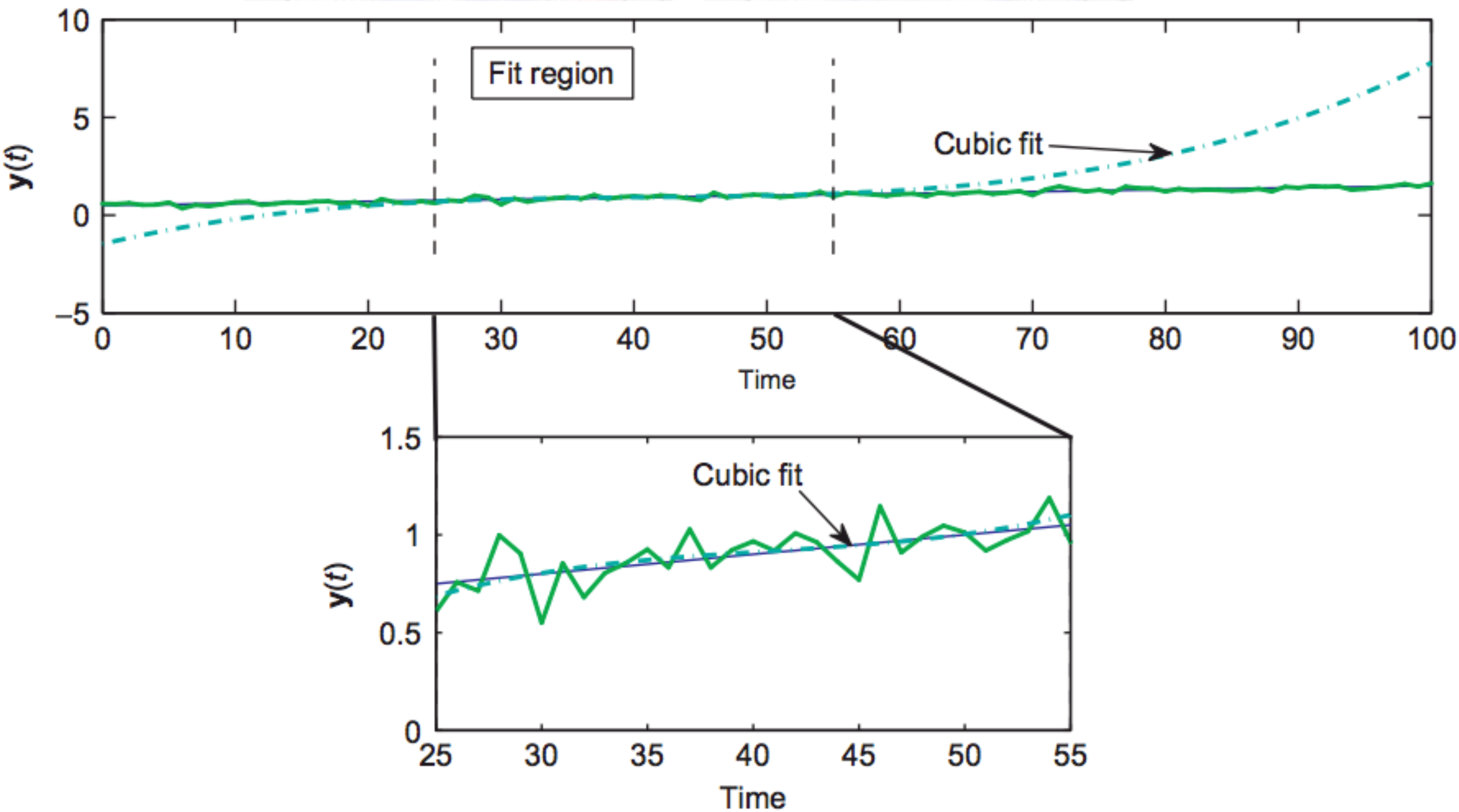
1) Spurious trends due to changes in the observational systems
2) Failure to close budgets
3) Data density, type and handling dominates climate-scale trends



Calculated trends (meters/second/year) in the 10-m zonal wind fields at high southern latitudes

Issues with reanalysis products

1) Spurious trends due to changes in the observational systems
2) Failure to close budgets
3) Data density, type and handling dominates climate-scale trends
4) Cannot be used for prediction: the Parable of the Cubic Polynomial



ECCO

TABLE 21.3 Published ECCO Family State Estimates, Divided Roughly into Categories

Label and Version	Hor./Ver. Grid	Domain	Duration	Scope	References
ECCO-Production Sustained production of decadal climate state estimates (former ECCO-GODAE)					
ver.0 (ECCO-MIT)	2°/22	80°N/S	1992–1997	First ECCO product—proof of feasibility	Stammer et al. (2002, 2004)
ver.1 (ECCO-SIO)	1°/23	80°N/S	1992–2002	Begin of 1° sustained production	Köhl et al. (2007)
ver.2 (ECCO-GODAE)	1°/23	80°N/S	1992–2004	Air–sea flux constraints for sea level studies	Wunsch and Heimbach (2006, 2007)
(OCCA)	1°/50	80°N/S	2004/2005/2006/2007	Atlas from 1-year "synoptic snapshots"	Forget (2010)
(GECCO)	1°/23	80°N/S	1951–2000	50-year solution covering NCEP/NCAR period	Köhl and Stammer (2008a,b)
ver.3 (ECCO-GODAE)	1°/23	80°N/S	1992–2007	Switch to atmospheric state controls and sea ice	Wunsch and Heimbach (2009)
ver. 4 (ECCO-Production)	1°/50	Global	1992–2010	First full-global estimate including Arctic	Forget et al. (in preparation, 2013)
ECCO-ICES Ocean–ice interactions in Earth system models (former ECCO2)					
ver.1 (CS510 GF)	18 km/50	Global	1992–2002/2010	Green's function optimization, of eddying model	Menemenlis et al. (2005a,b)
ECCO-JPL near real-time filter and reduced-space smoother					
ver.1 (KF)	1°/46	80°N/S	1992–present	Near-real-time Kalman Filter (KF) assimilation	Fukumori et al. (1999)
ver.2 (RTS)	1°/46	80°N/S	1992–present	Smoother update of KF solution	Fukumori (2002)
Regional efforts					
Southern Ocean (SOSE)[a]	1/6°/42	25°–80°S	2005–2009	Eddy-permitting SO state estimate	Mazloff et al. (2010)
ECCO2 Arctic and ASTE[a]	18 and 4 km/50	Arctic and SPG	1992–2009	Arctic/subpolar gyre ocean–sea ice estimate	Nguyen et al. (2011, 2012)
North Atlantic	1°/23	25°–80°N	1993	Experimental 2° versus 1° nesting	Ayoub (2006)
Subtropical Atlantic	1/6°/42	–	1992/1993	Experimental 1° versus 1/6° nesting	Gebbie et al. (2006)
Tropical Pacific	–	–		Experimental 1° versus 1/3° nesting	Hoteit et al. (2006, 2010)
Labrador Sea and Baffin Bay	–	–	1996/1997	First full coupled ocean–sea ice estimate	Fenty and Heimbach (2013a,b)

The global decade+estimates are labeled as "ECCO-Production," while others are either regional or experimental.
[a]*Denotes ongoing efforts.*

...and SOCOM

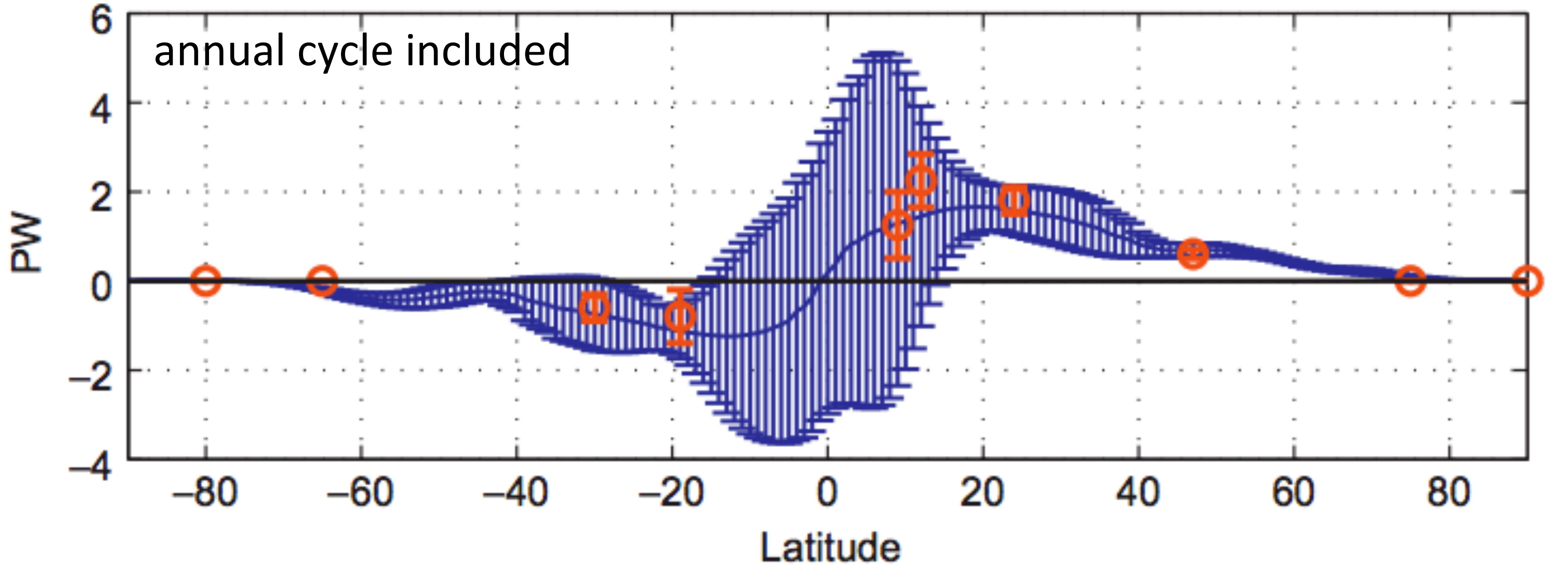
ECCO assimilated data

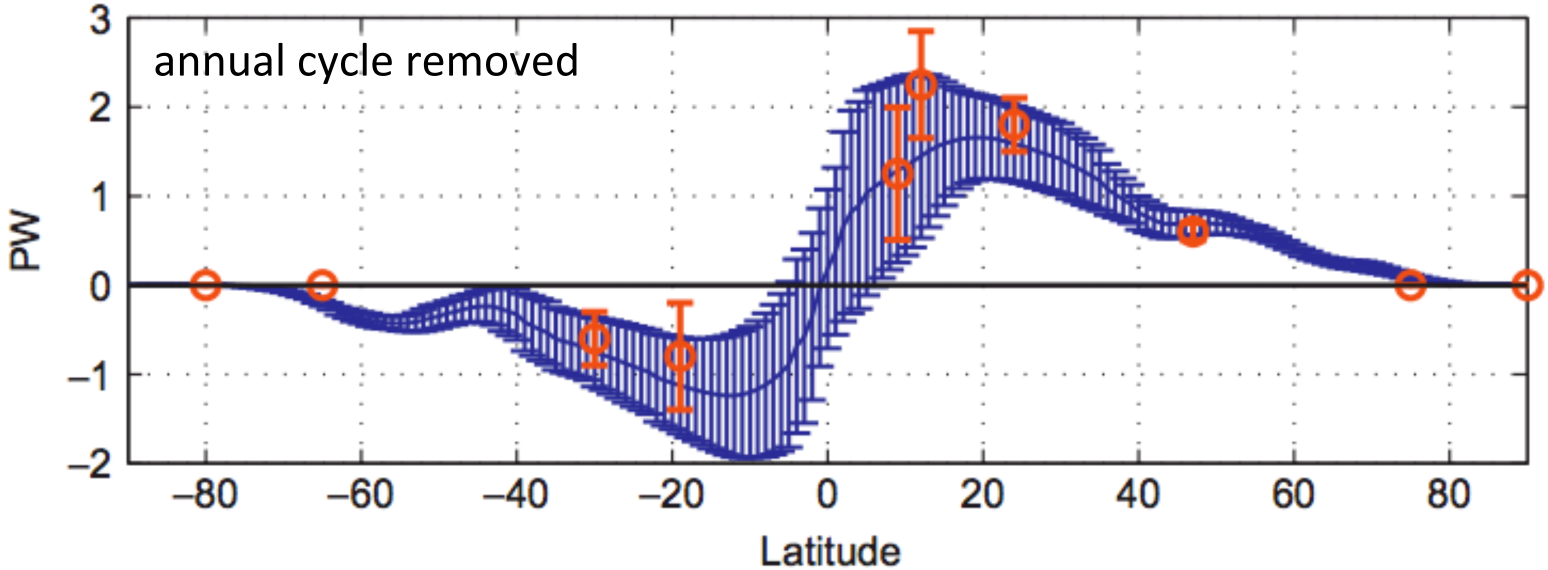
TABLE 21.2 Data Used in the ECCO Global 1° Resolution State Estimates Until About 2011

Observation	Instrument	Product/Source	Area	Period	dT
Mean dynamic topography (MDT)	• GRACE SM004-GRACE3	CLS/GFZ (A.M. Rio)	Global	Time-mean	Mean
	• EGM2008/DNSC07	N. Pavlis/Andersen & Knudsen	Global		
Sea level anomaly (SLA)	• TOPEX/POSEIDON	NOAA/RADS & PO.DAAC	65°N/S	1993–2005	Daily
	• Jason	NOAA/RADS & PO.DAAC	82°N/S	2001–2011	Daily
	• ERS, ENVISAT	NOAA/RADS & PO.DAAC	65°N/S	1992–2011	Daily
	• GFO	NOAA/RADS & PO.DAAC	65°N/S	2001–2008	Daily
SST	• Blended, AVHRR (O/I)	Reynolds & Smith	Global	1992–2011	Monthly
	• TRMM/TMI	GHRSST	40°N/S	1998–2004	Daily
	• AMSR-E (MODIS/Aqua)	GHRSST	Global	2001–2011	Daily
SSS	Various *in situ*	WOA09 surface	Global	Climatology	Monthly
In situ T, S	• Argo, P-Alace	Ifremer	"Global"	1992–2011	Daily
	• XBT	D. Behringer (NCEP)	"Global"	1992–2011	Daily
	• CTD	Various	Sections	1992–2011	Daily
	• SEaOS	SMRU & BAS (UK)	SO	2004–2010	Daily
	• TOGA/TAO, Pirata	PMEL/NOAA	Tropics	1992–2011	Daily
Mooring velocities	• TOGA/TAO, Pirata	PMEL/NOAA	Tropics	1992–2006	Daily
	• Florida Straits	NOAA/AOML	North Atlantic	1992–2011	Daily
Average T, S	• WOA09	WOA09	"Global"	1950–2000	Mean
	• OCCA	Forget (2010)	"Global"	2004–2006	Mean
Sea ice cover	• Satellite passive microwave radiometry	NSIDC (bootstrap)	Arctic, SO	1992–2011	Daily
Wind stress	QuickScat	• NASA (Bourassa)	Global	1999–2009	Daily
		• SCOW (Risien & Chelton)		Climatology	Monthly
Tide gauge SSH	Tide gauges	NBDC/NOAA	Sparse	1992–2006	Monthly
Flux constraints	From ERA-Interim, JRA-25, NCEP, CORE-2 variances	Various	Global	1992–2011	2-day to 14-day
Balance constraints			Global	1992–2011	Mean
Bathymetry		Smith & Sandwell, ETOPO5	Global	–	–

An estimated 22×10^9 individual values have been used in Equation (21.6), of which about 4×10^8 are assigned to the control terms.

ECCO-Production version 4: global heat transport

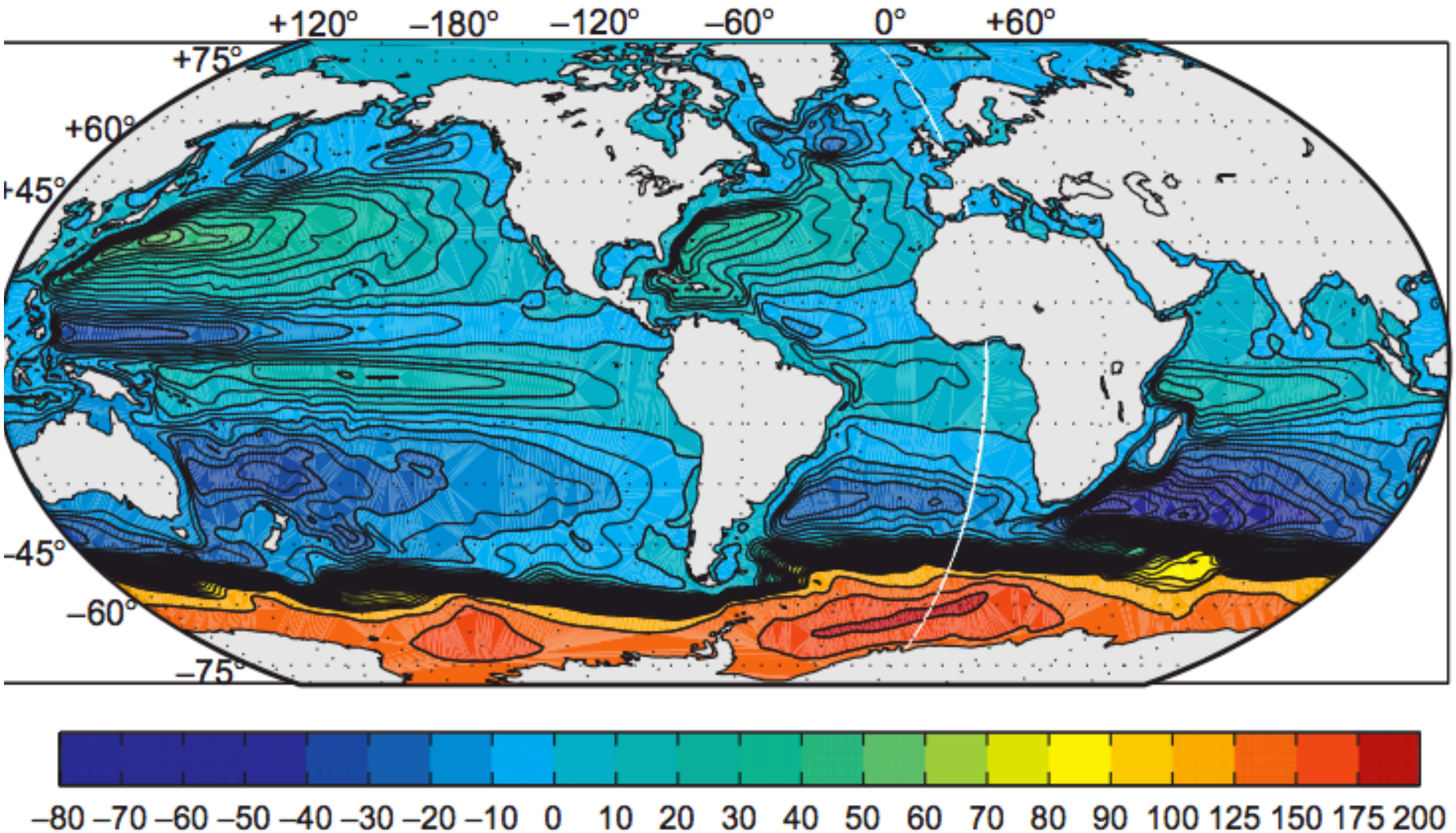




error bars are standard error about the mean;

red dots show estimates from observations (Ganachaud and Wunsch, 2002).

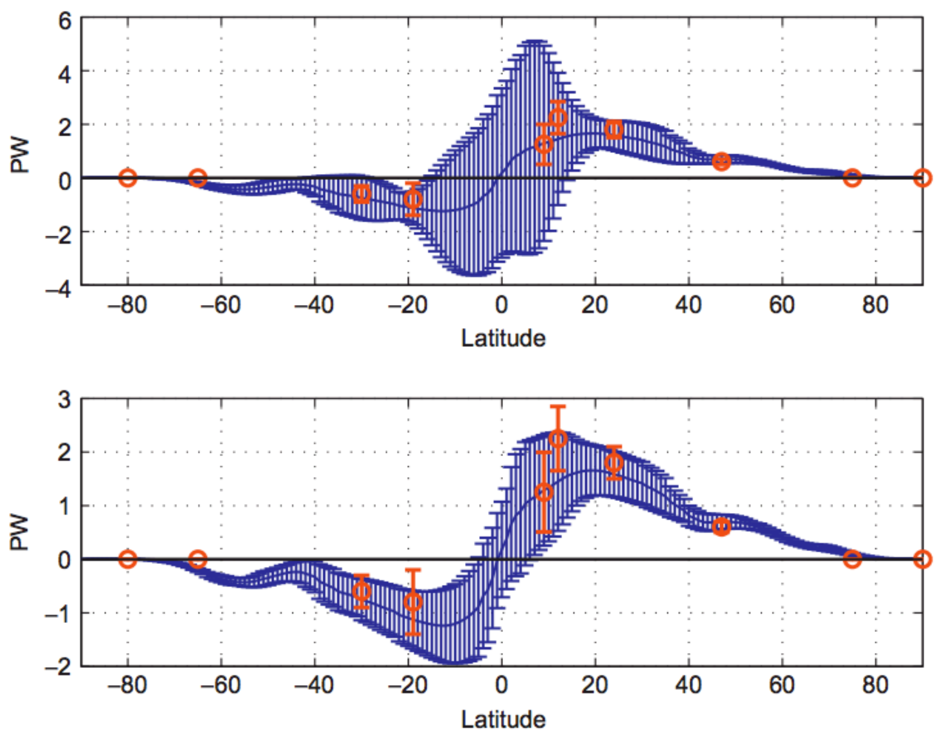
ECCO v3.73: Top-to-bottom transport streamfunction



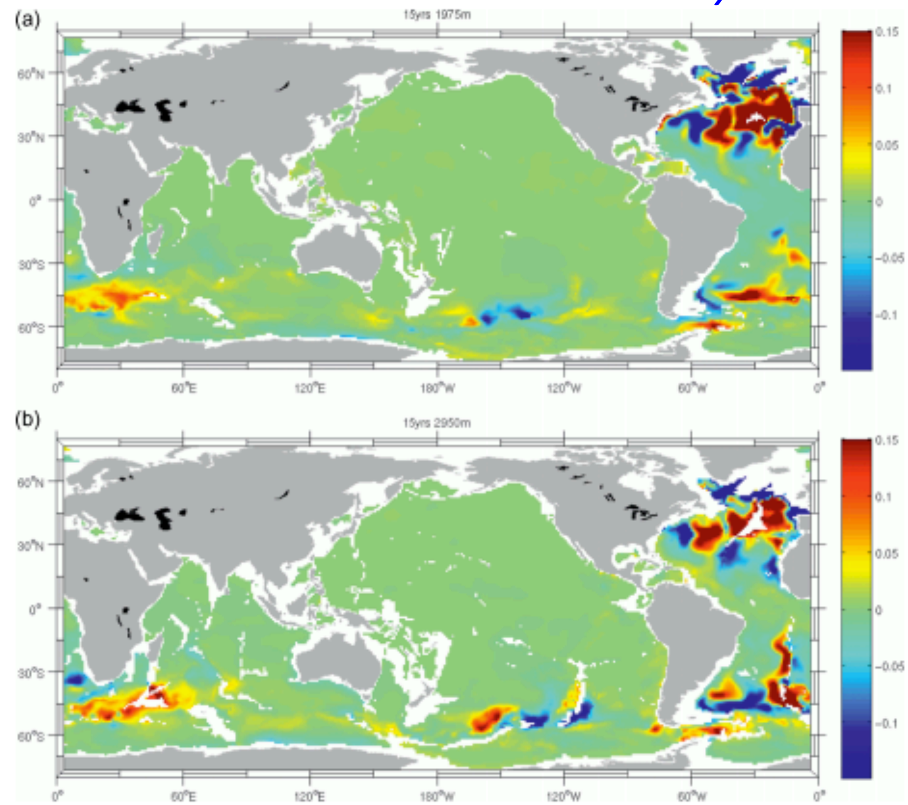
Estimating uncertainties

Determining the true uncertainty in a state estimate remains a difficult, unsolved problem. What can we use in the meantime?

Standard error from temporal variability



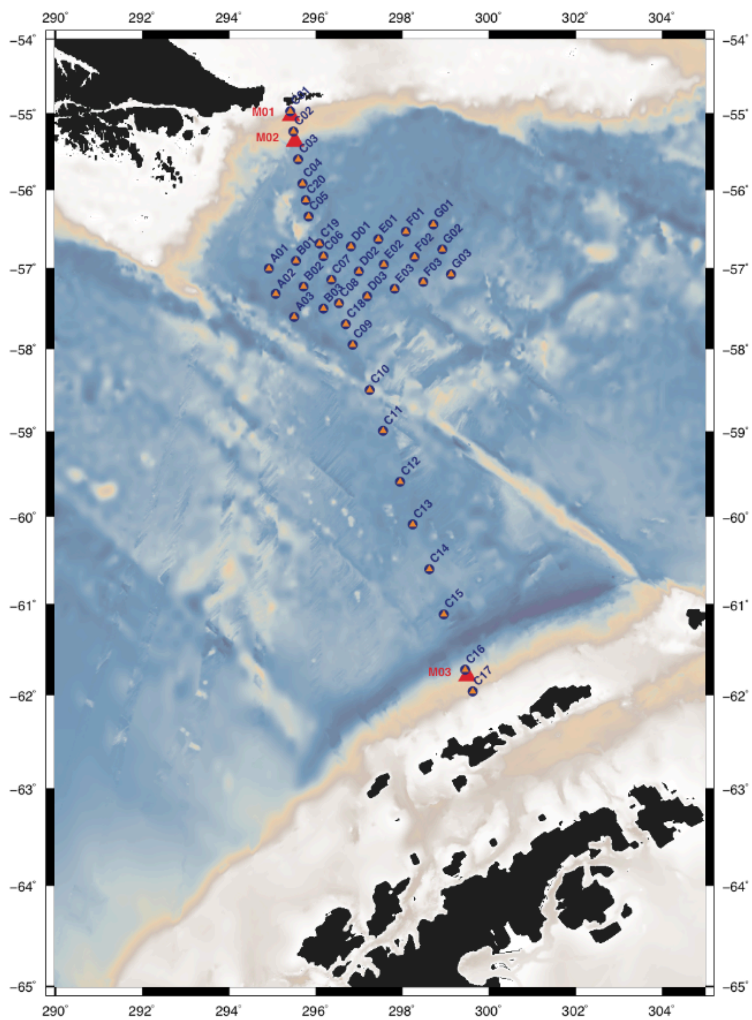
Sensitivity studies

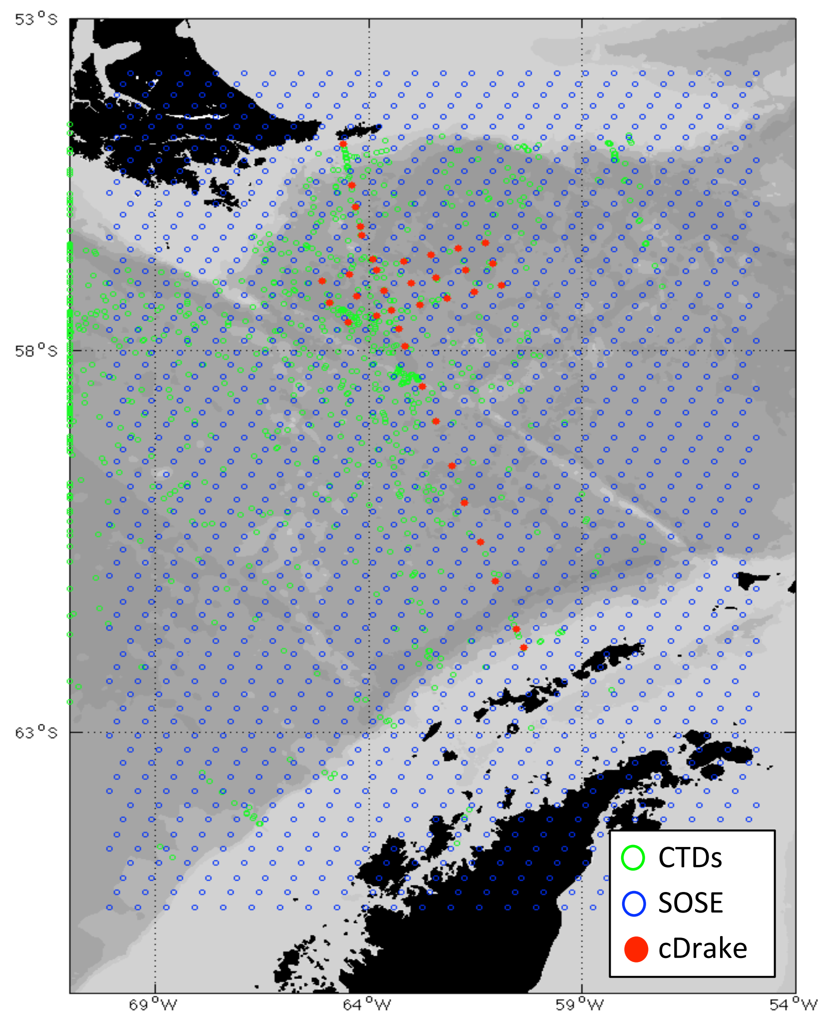


Estimating uncertainties

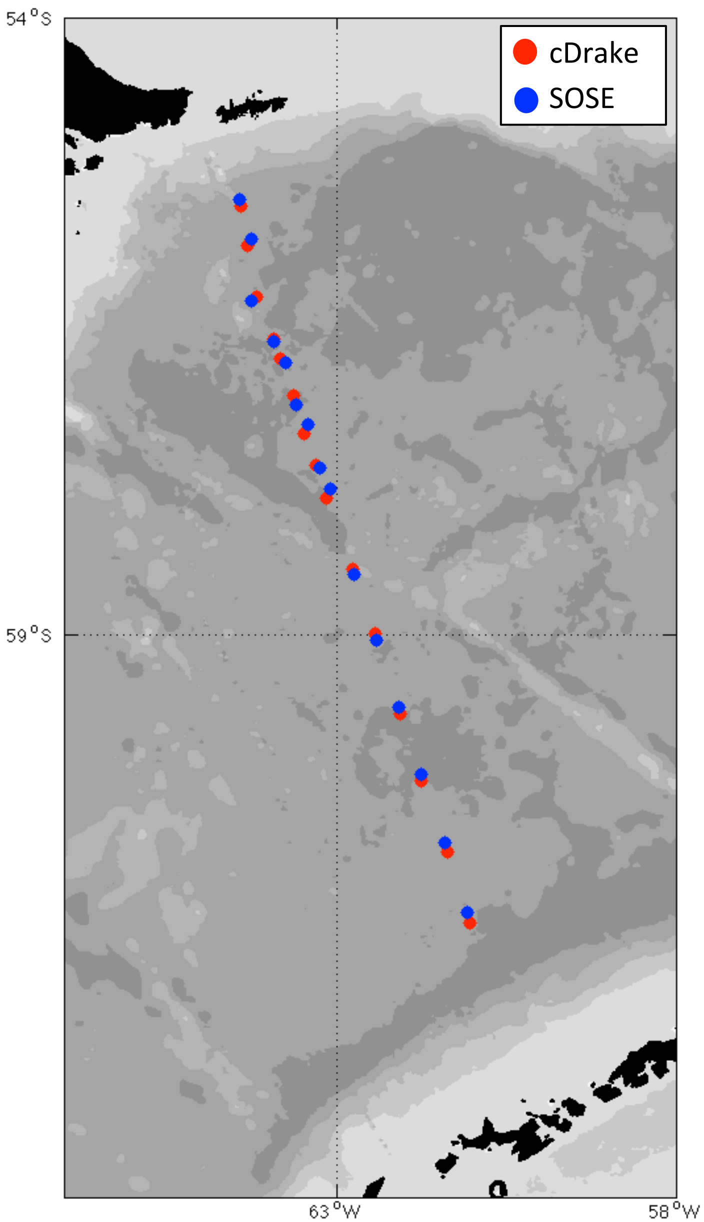
Determining the true uncertainty in a state estimate remains a difficult, unsolved problem. What can we use in the meantime?

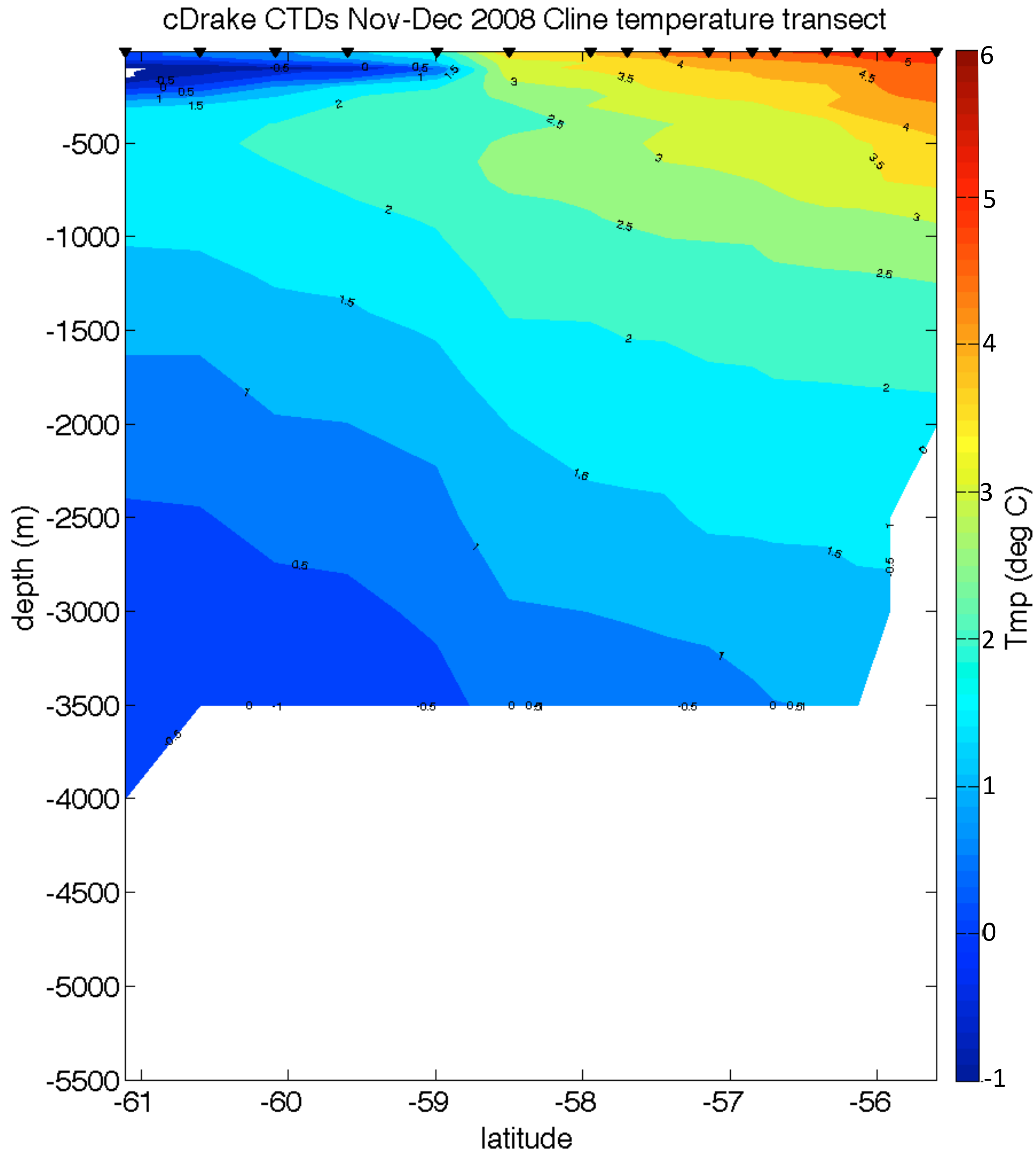
Direct comparisons between observations and model output



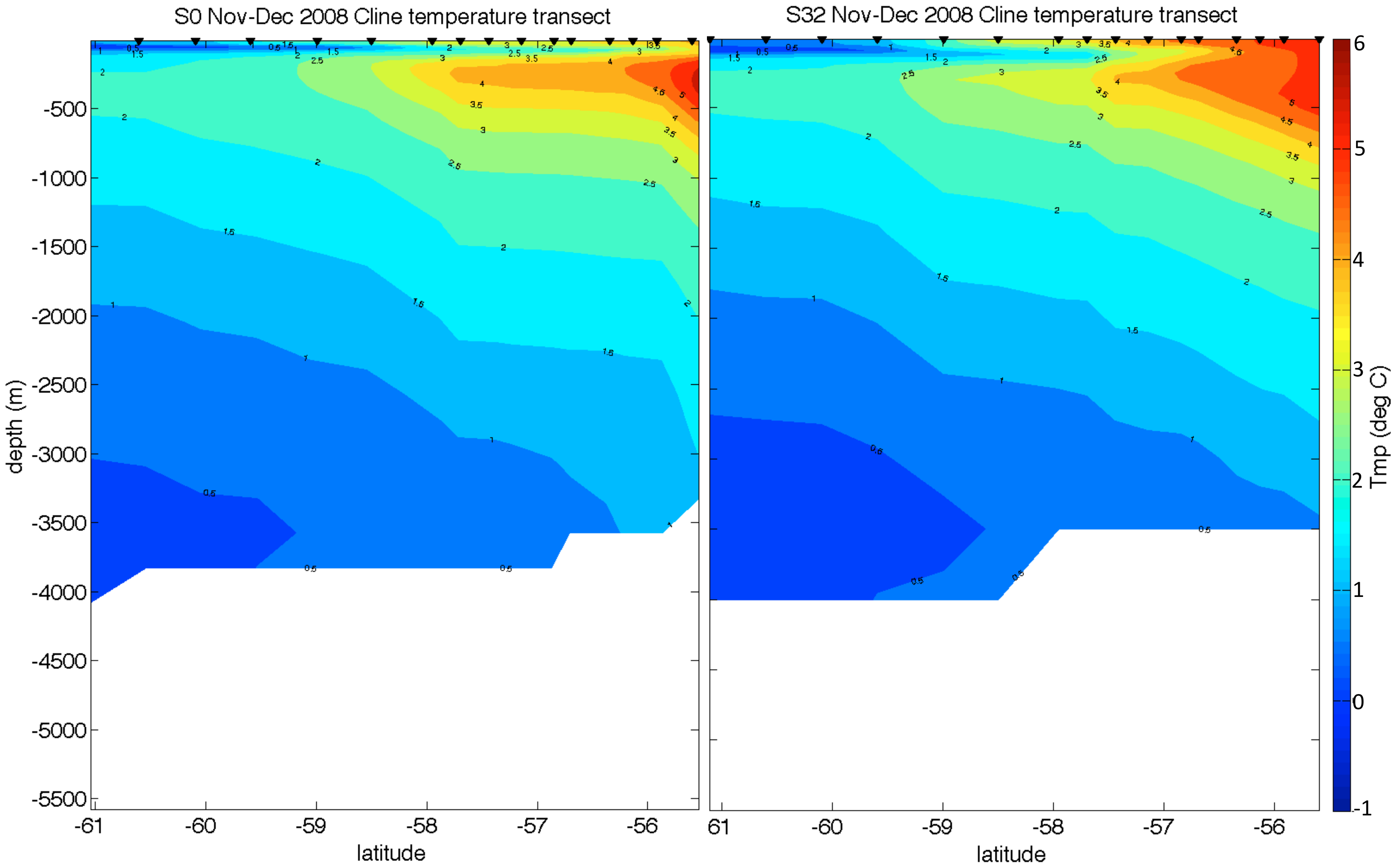


cDrake C-line temperature transect, Nov. and Dec. 2008

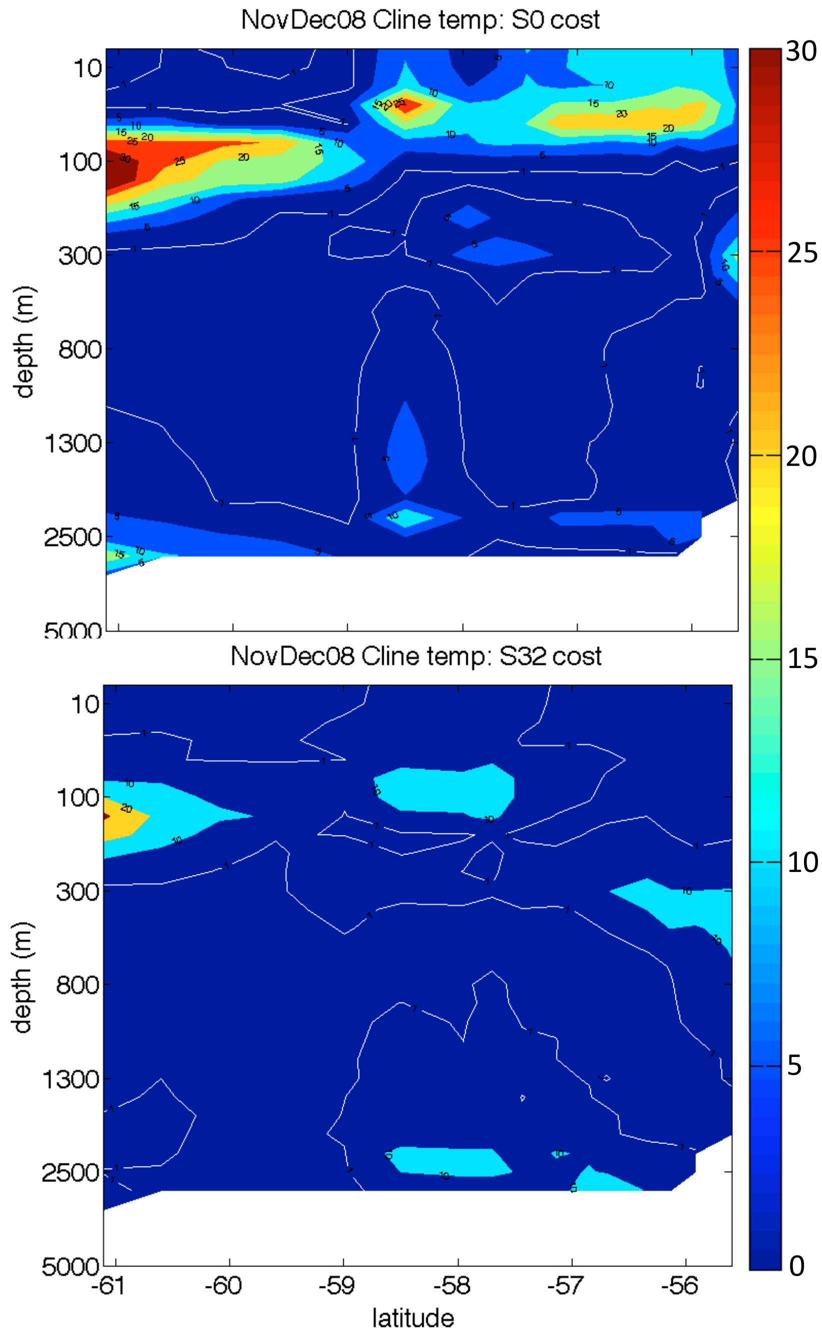




C-line temperature transect: SOSE 0 and SOSE 32

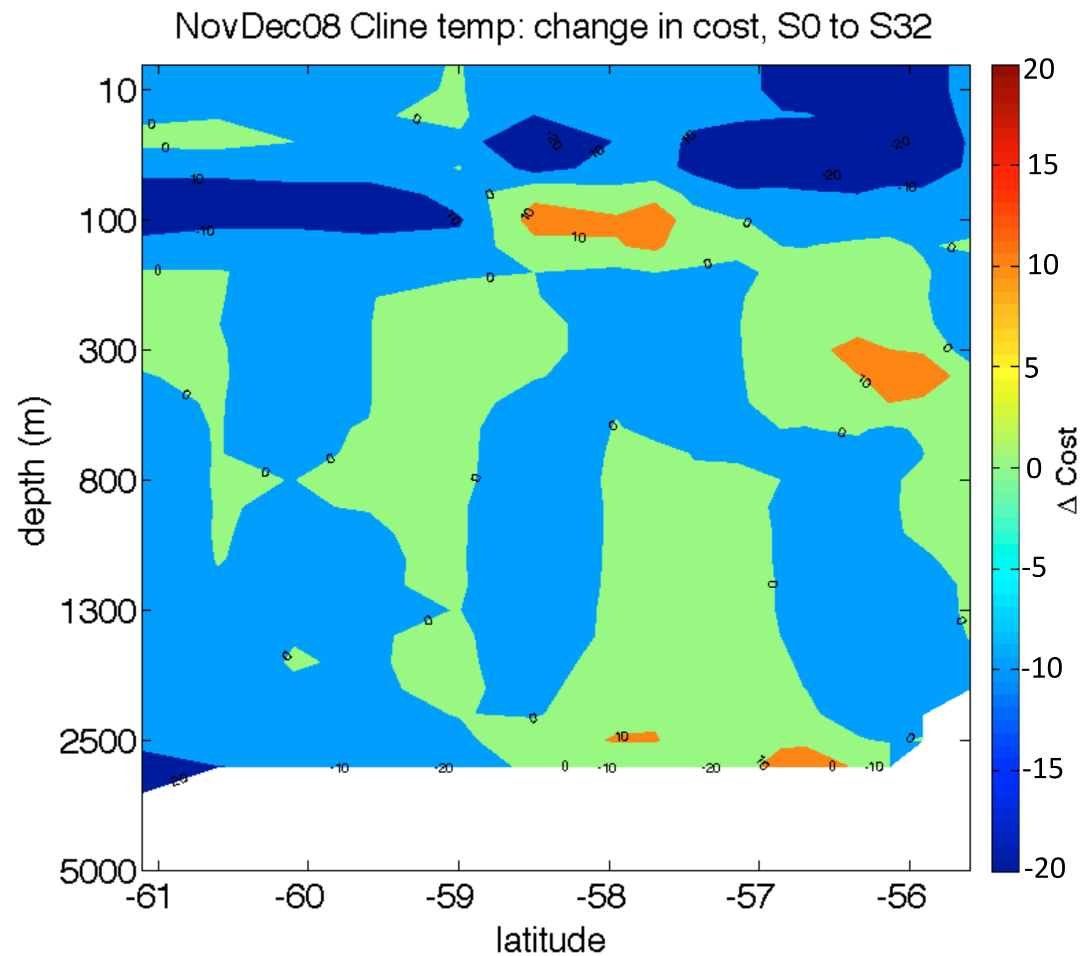


C-line temperature transect: Improvement, SOSE 0 to SOSE 32



Percent improvement, SOSE 0 to SOSE 32:

$$100 * \left(1 - \frac{\sum (SOSE32 - observations)^2}{\sum (SOSE0 - observations)^2}\right) = 53.18\%$$



SOSE/cDrake τ anomaly comparison: RMS error for three SOSE iterations

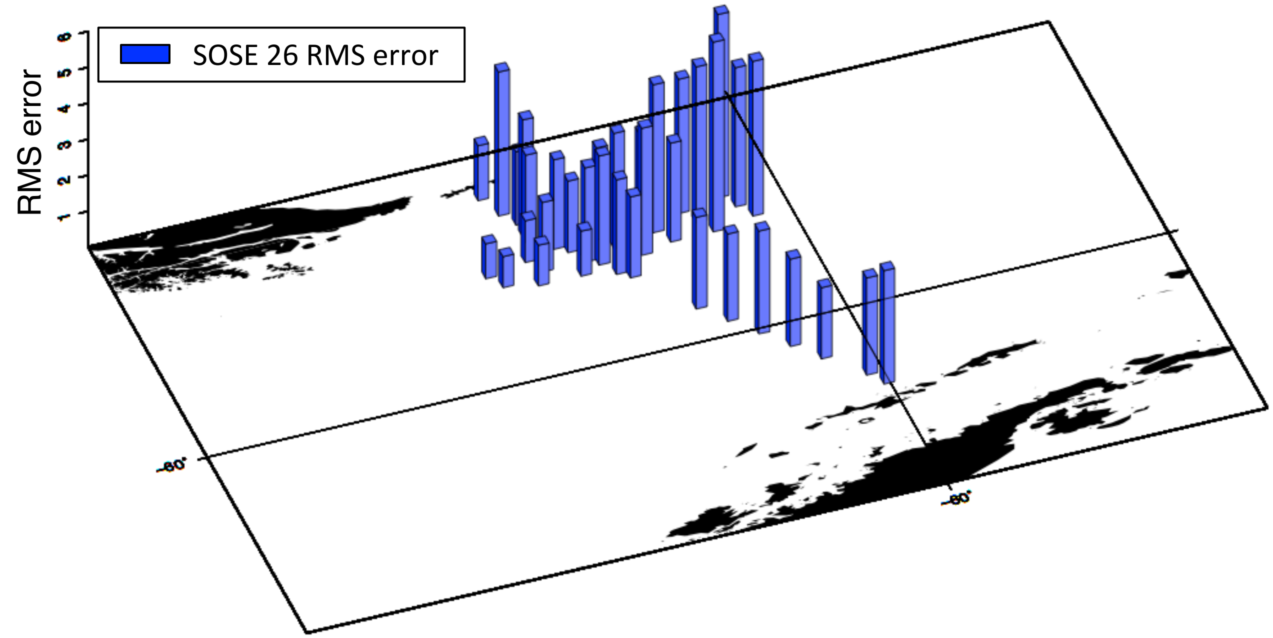
Full-depth τ anomaly at cDrake array locations, 2008-2009, RMS error:

$$\sqrt{\sum\left(\tau_{SOSE}-\tau_{OBS}\right)^2}$$

for SOSE 0 (ms) = 3.7

for SOSE 13 (ms) = 3.3

for SOSE 26 (ms) = 2.9



Normalized RMS error:

$$\frac{\sqrt{\sum\left(\tau_{SOSE}-\tau_{OBS}\right)^2}}{\sigma_{OBS}}$$

for SOSE 0 = 1.7396

for SOSE 13 = 1.6357

for SOSE 26 = 1.4074