An Oceanographic Perspective on Climate State Estimation

Boulder, June 2010

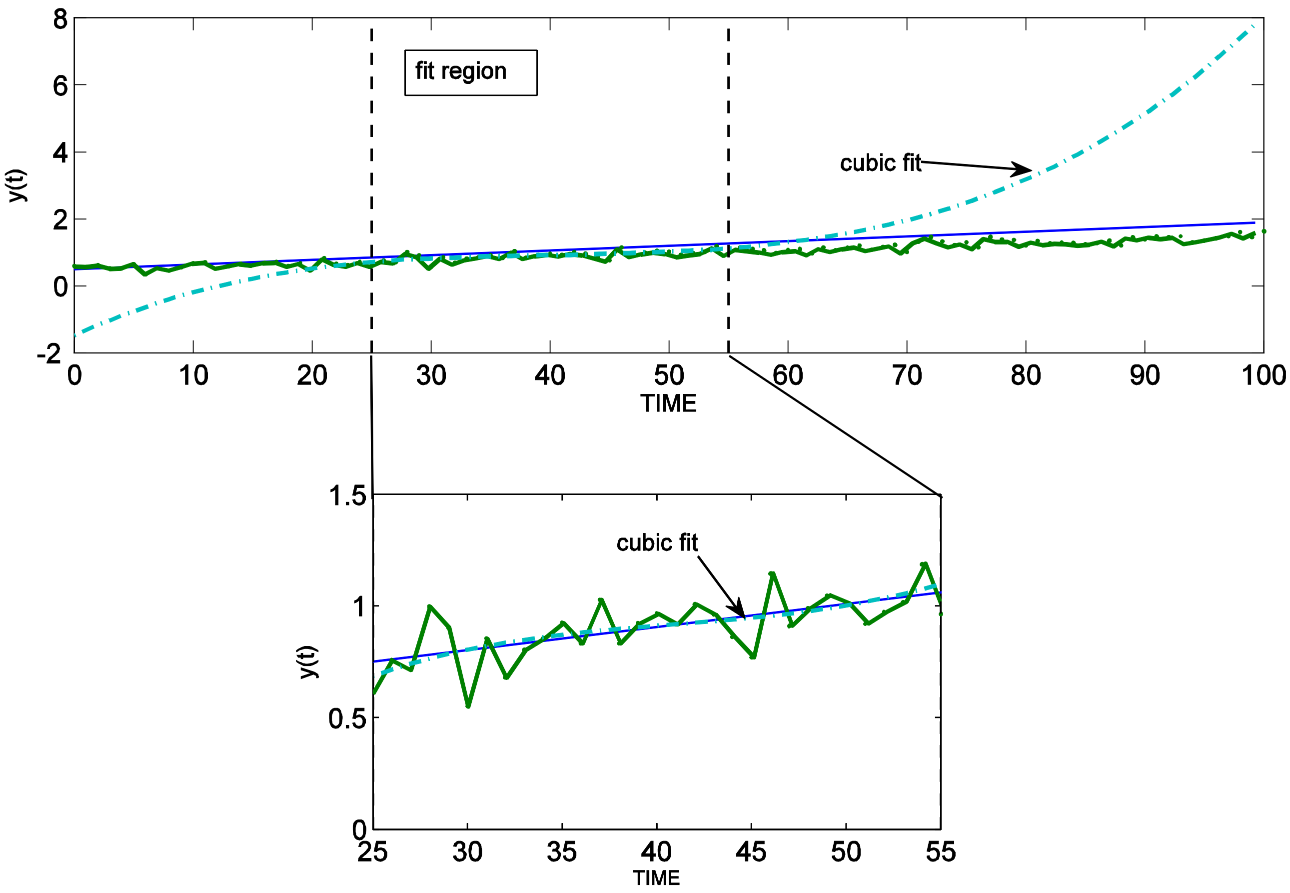
Carl Wunsch, MIT

This work dates back to about 1992, when it became clear that physical oceanographers would, for the first time, have data sets near-global in scope and continuous in time (the World Ocean Circulation Experiment, WOCE). The question was how to use these new data?

“Data assimilation” is a very sophisticated collection of numerical techniques developed primarily for *prediction* in the context of numerical weather prediction (NWP).

Prediction and state estimation (or *smoothing* or *interpolation)* are very different problems and many methods useful for weather forecasting are not suitable for state estimation.

Recall the elementary problem of interpolating noisy data with a polynomial:



This is too simple to be more than a metaphor---but worth remembering.

Let us examine the meteorological “reanalyses” for a bit of context.

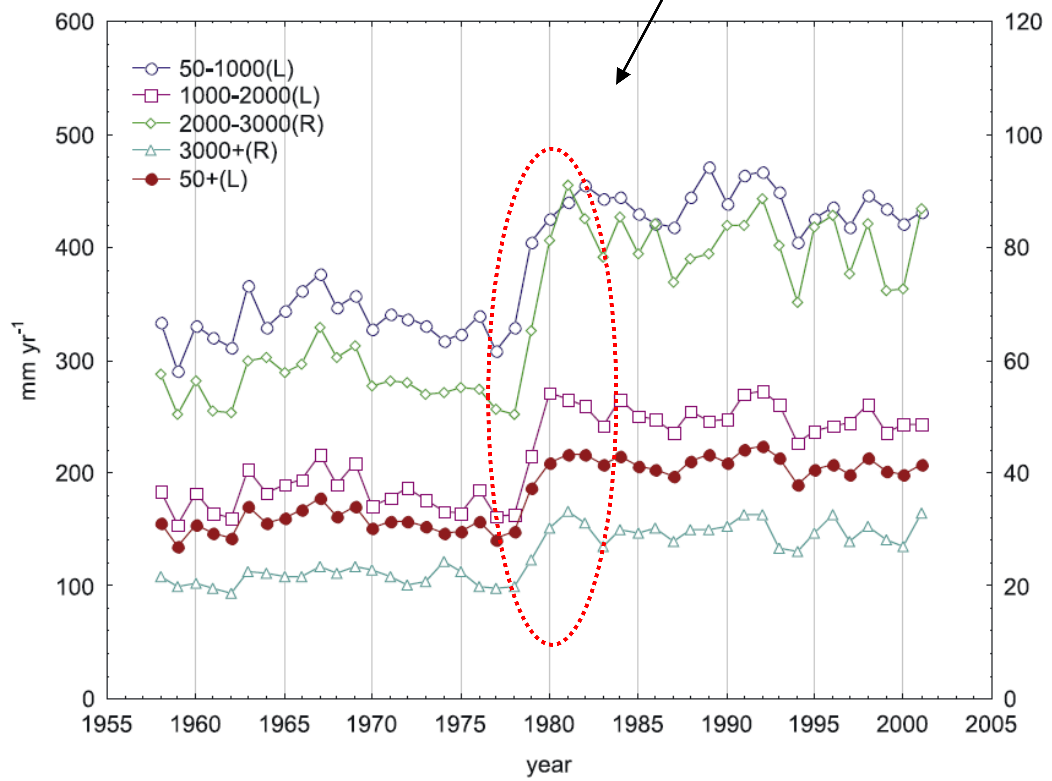
Generally speaking, these are based upon the weather forecast systems, but with the model and assimilation methodology held fixed over multi-decadal intervals. They have been widely used to study the climate system for various trends.

(I am indebted to David Bromwich, Ohio State, for the comparisons that follow.)

arrival of polar orbiters

Mean annual Antarctic net precipitation (P-E) from ERA-40 reanalysis for various elevation areas.

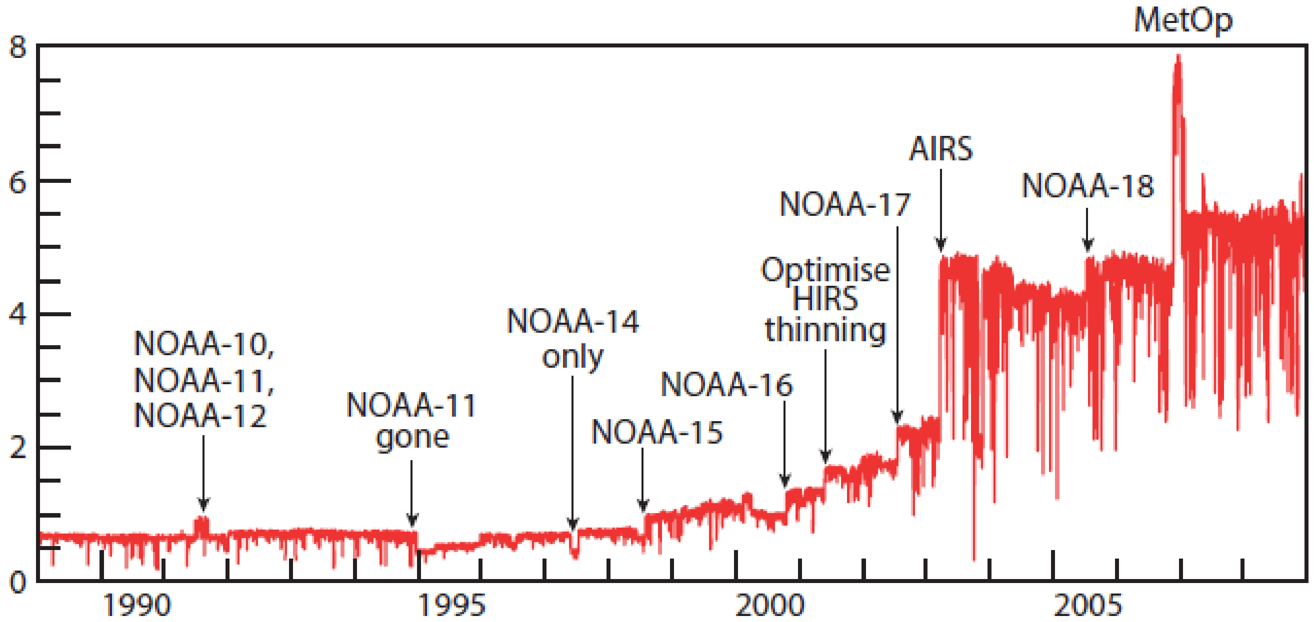
[Bromwich et al. 2007, adapted from Van de Berg et al. 2005]



- Spurious trends in the high latitudes resulting from changes in the observing system, especially the assimilation of satellite observations in the late 1970s.
- Jump in Antarctic P-E in 1978-79, particularly marked at high elevations.

- A related scenario in the 1990s-2000s?
- Dramatic increase in the amount and quality of satellite observations assimilated into the reanalyses (or available for assimilation).

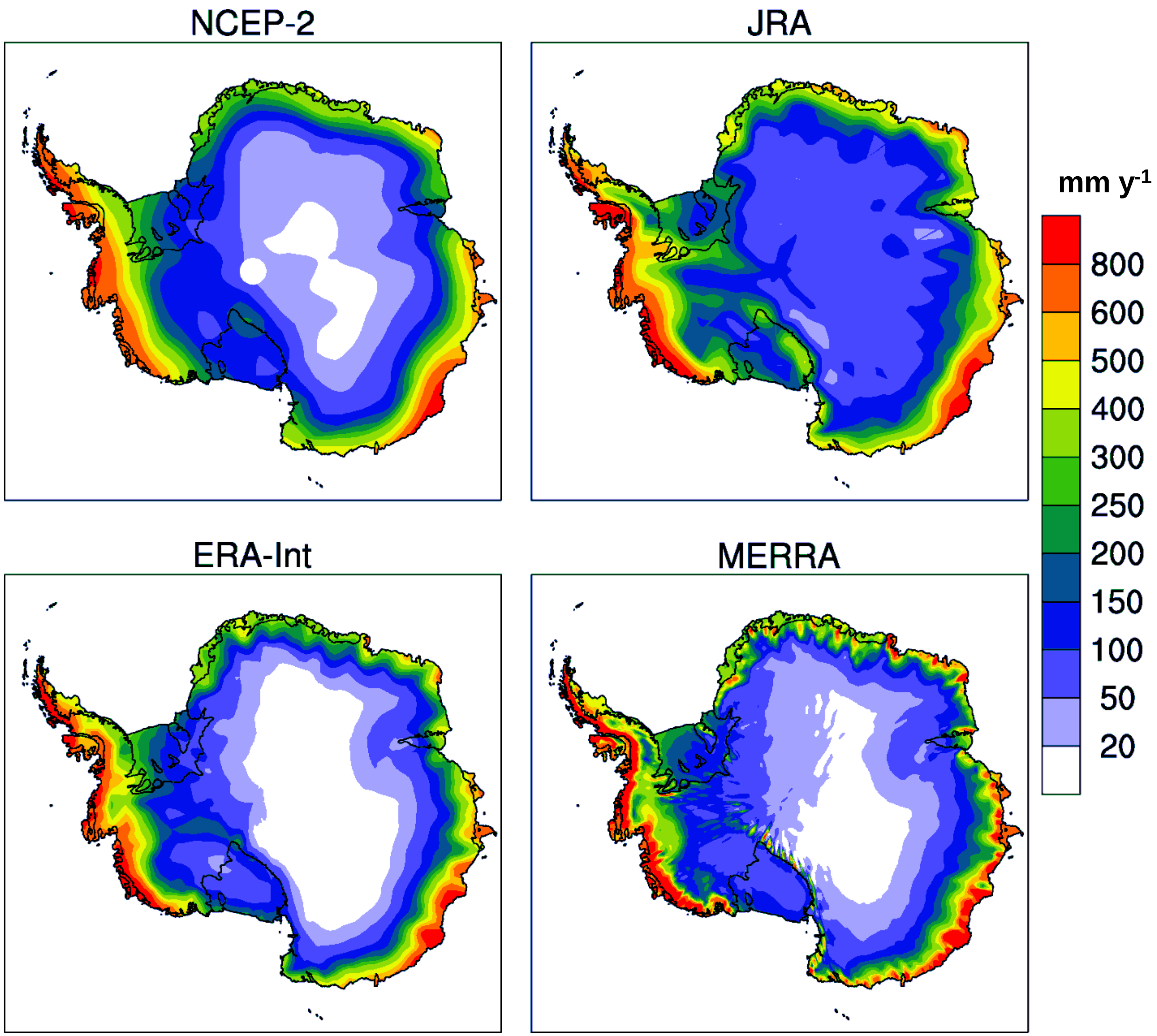
Number of observations assimilated in ERA-Interim



[Dee et al., 2009, *ECMWF Newsletter* (119)]

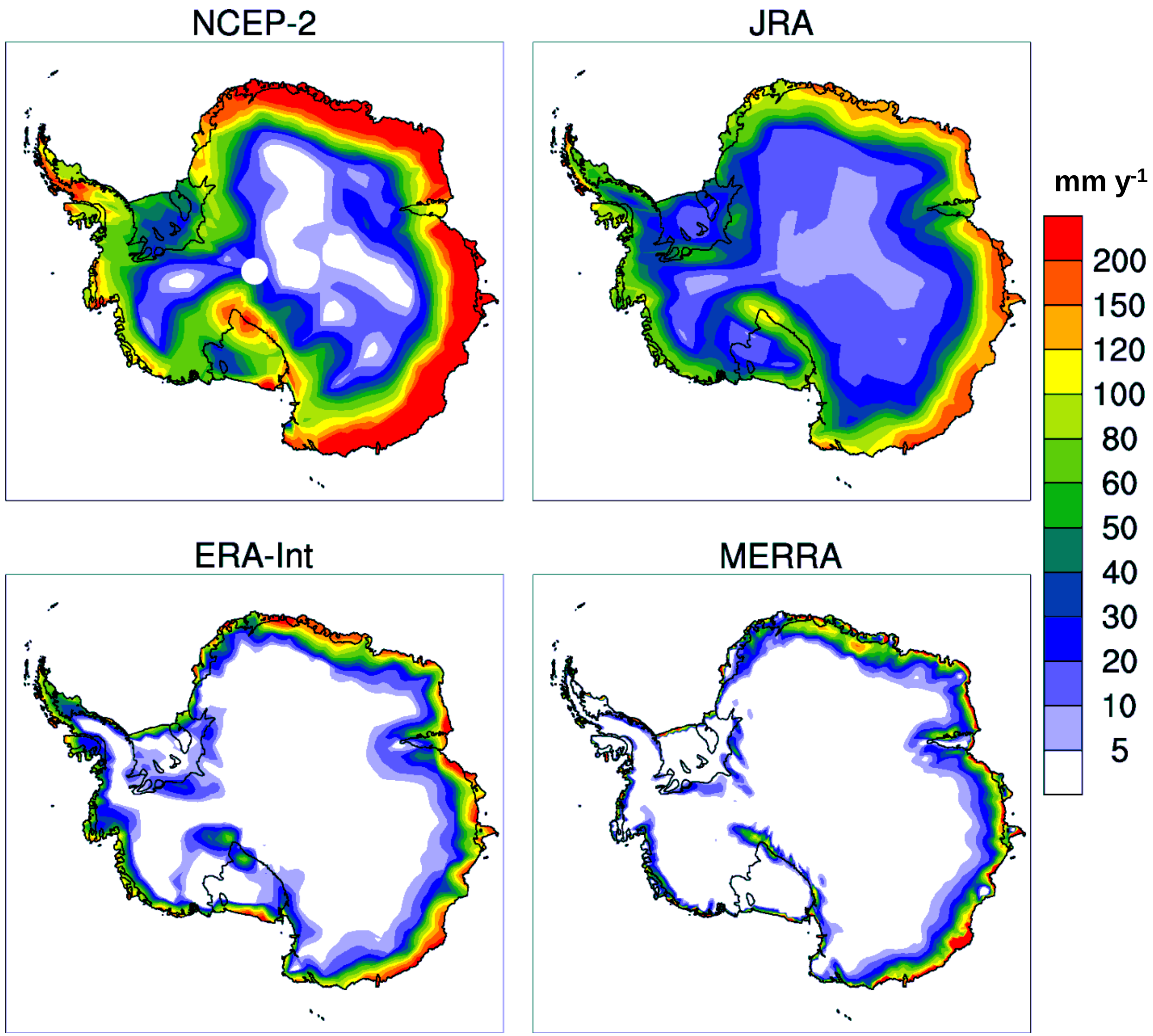
Mean annual precipitation (P)

1989-2008



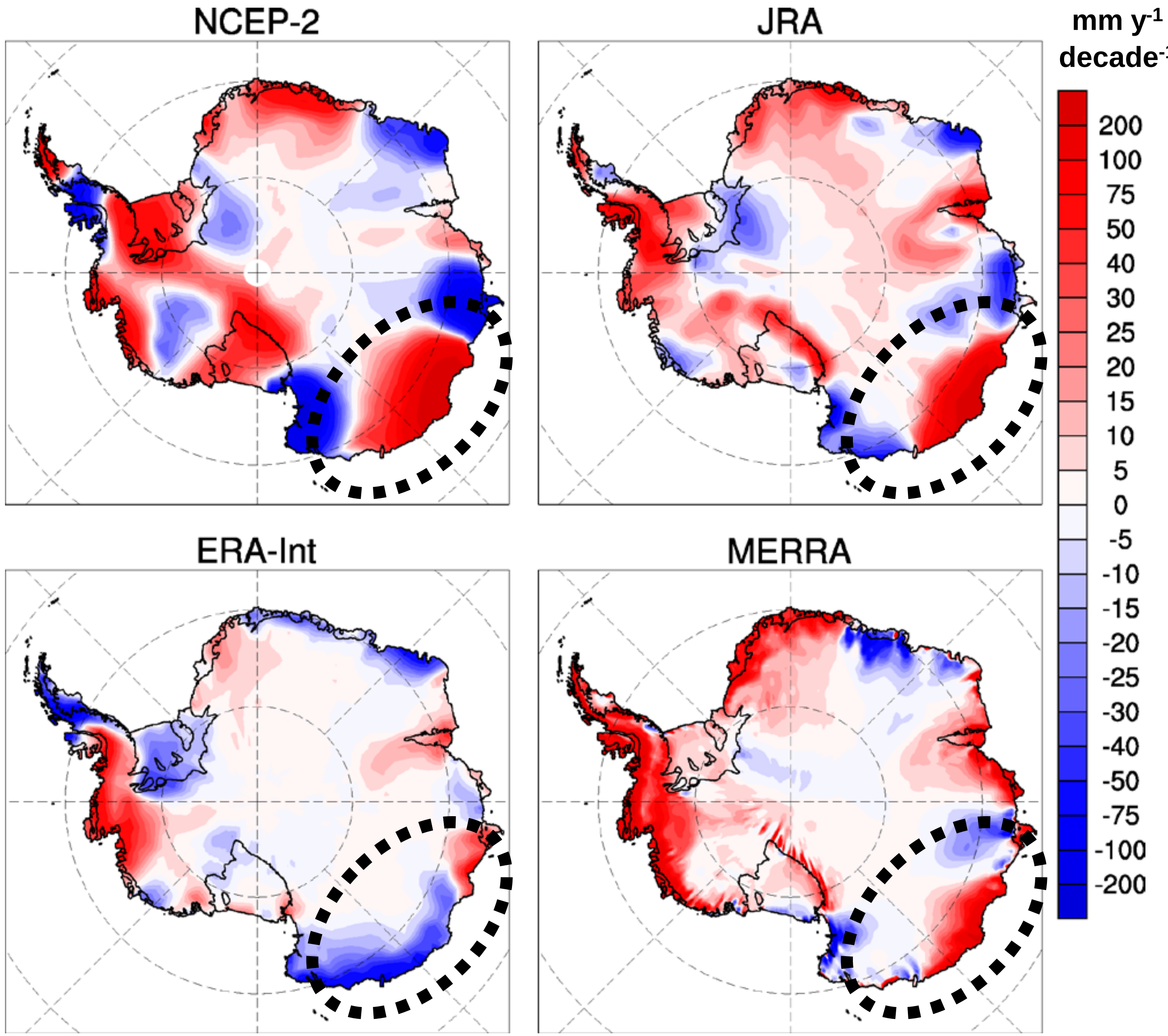
Mean annual evaporation (E)

1989-2008

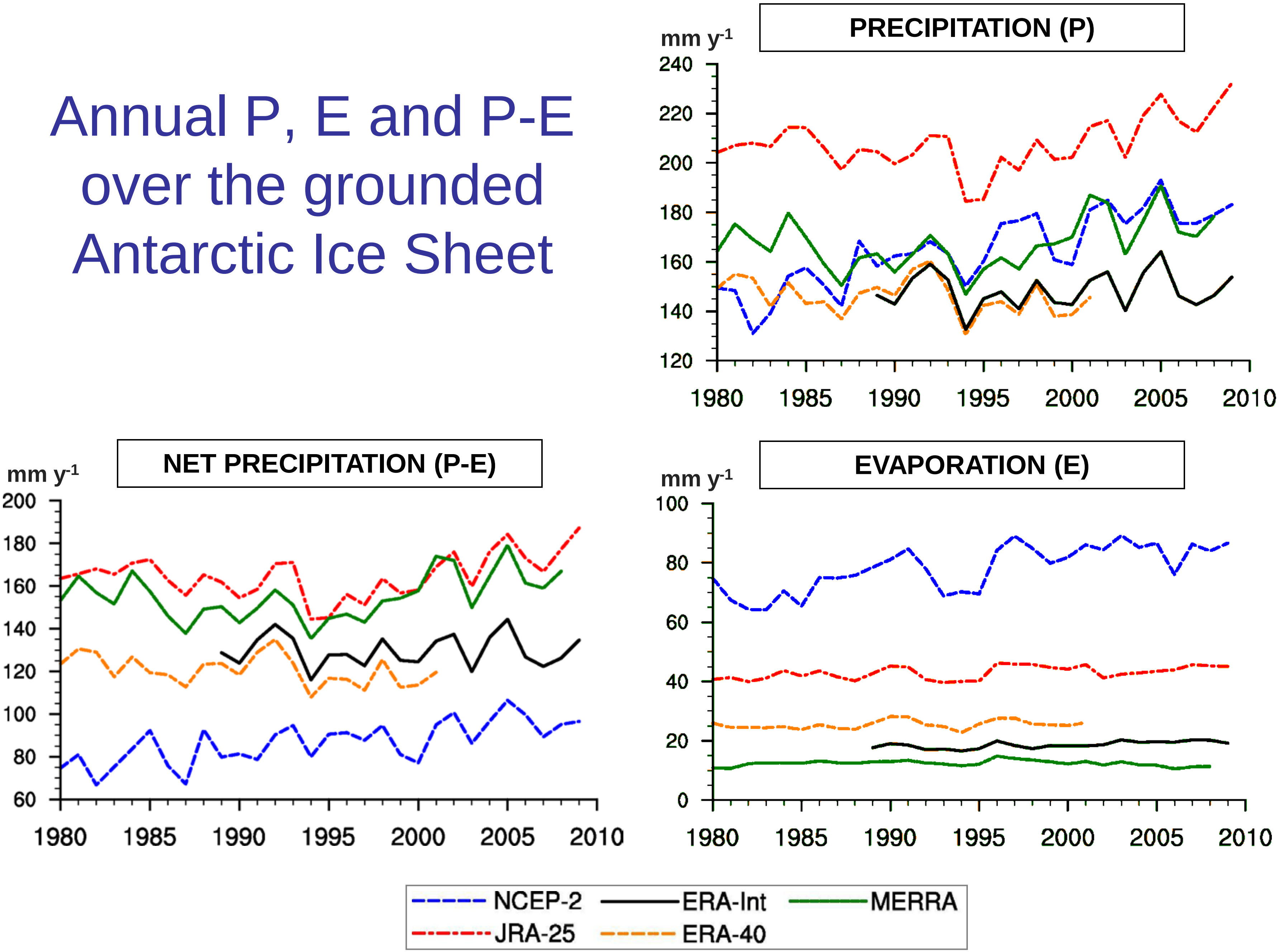


1989-2008 linear trends in annual P-E

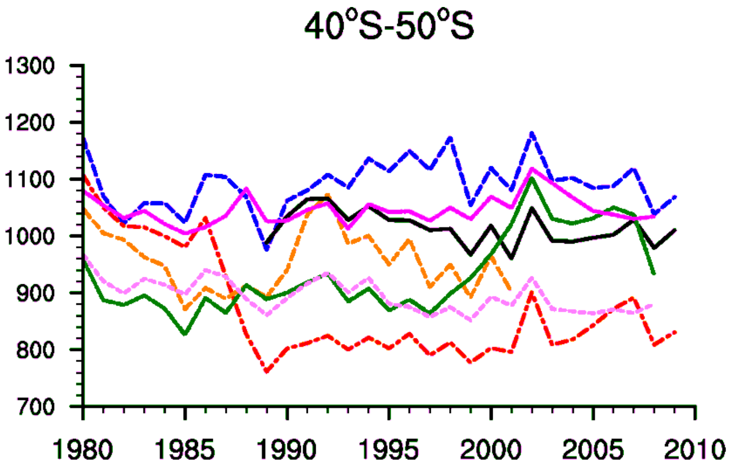
(D. Bromwich, Byrd Polar Research Center, Ohio)

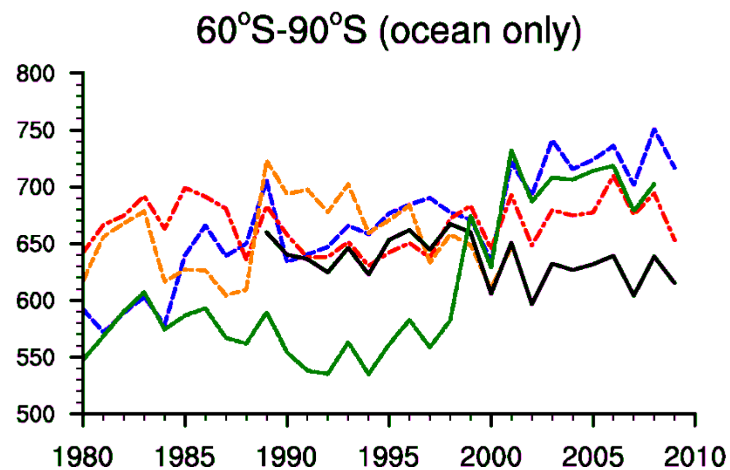


Annual P, E and P-E over the grounded Antarctic Ice Sheet

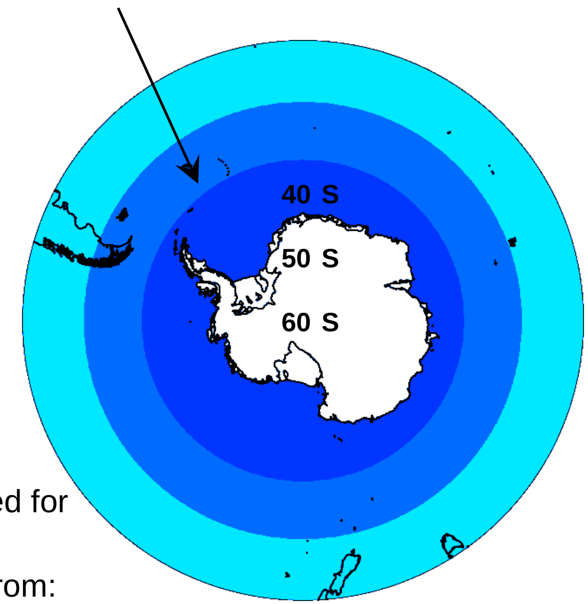


Precipitation and PW changes over the Southern Ocean





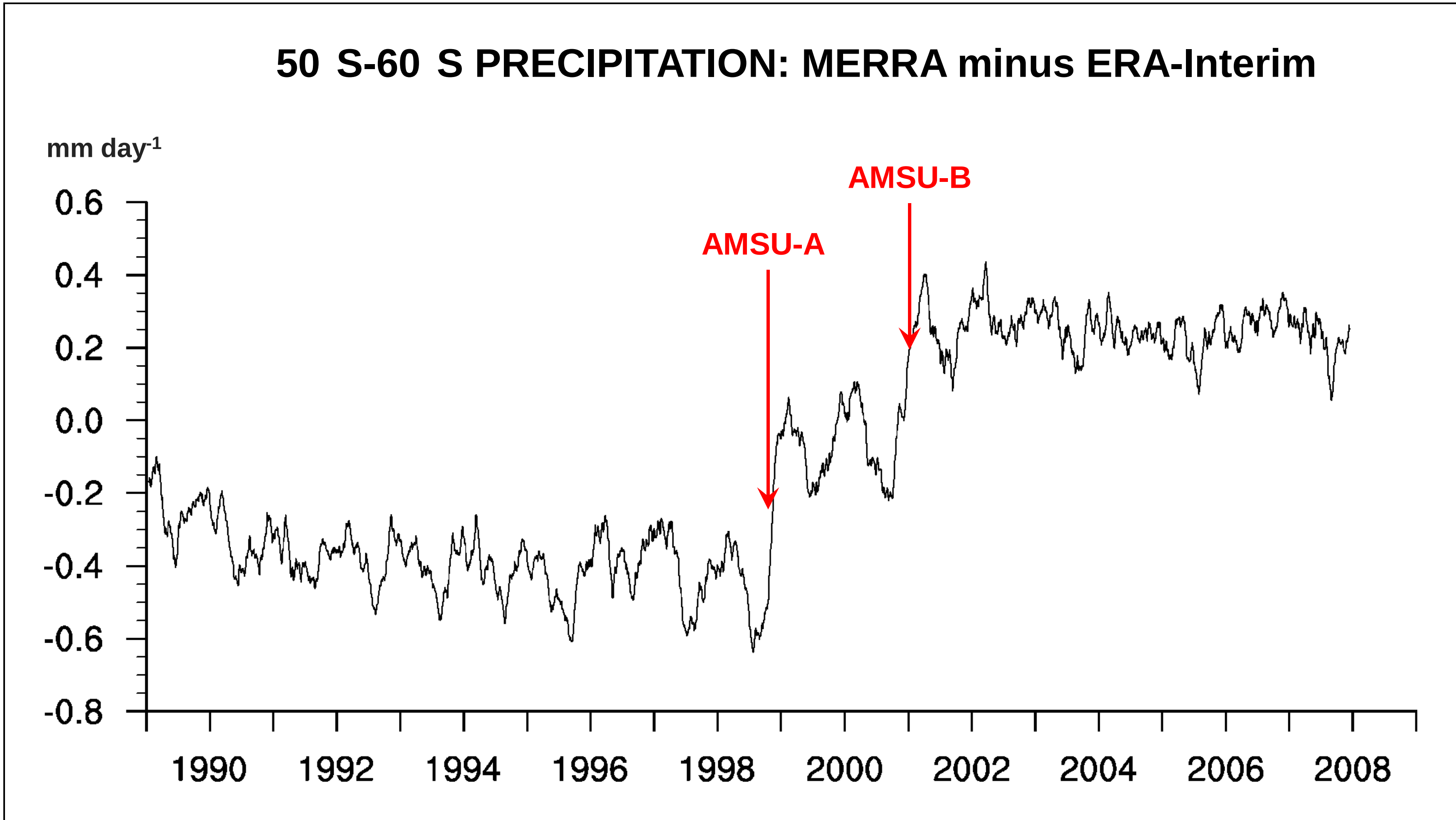
- Zonal means of precipitation and total precipitable water (PW) are examined for different <u>latitude bands</u>.



Additional datasets are included for latitudes **40 S-60 S**:

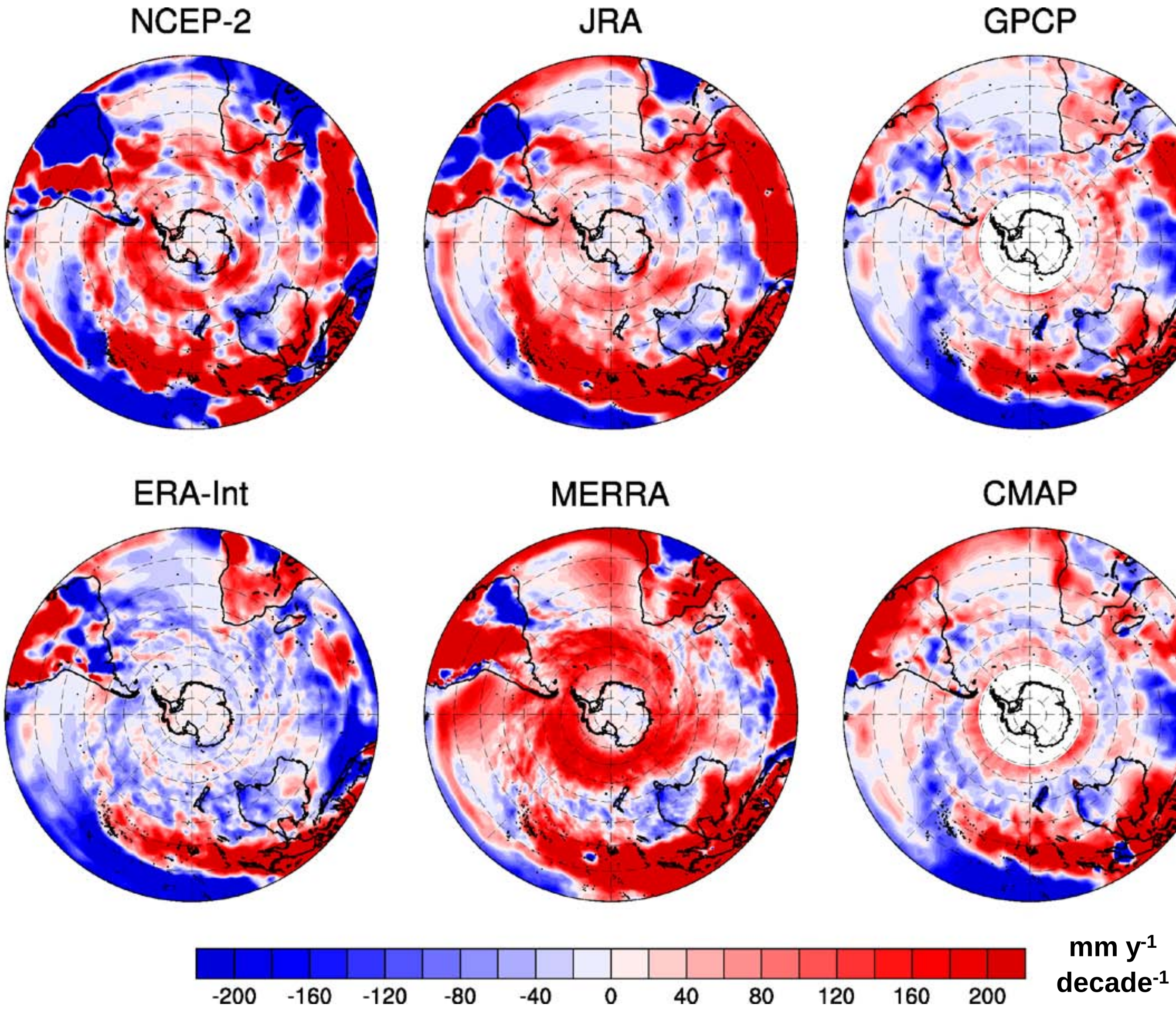
- Precipitation estimates from:
  - Global Precipitation Climatology Project (**GPCP**)
  - Climate Prediction Center Merged Analysis of Precipitation (**CMAP**)
- PW estimates from **SSM/I** (over ice-free ocean only)

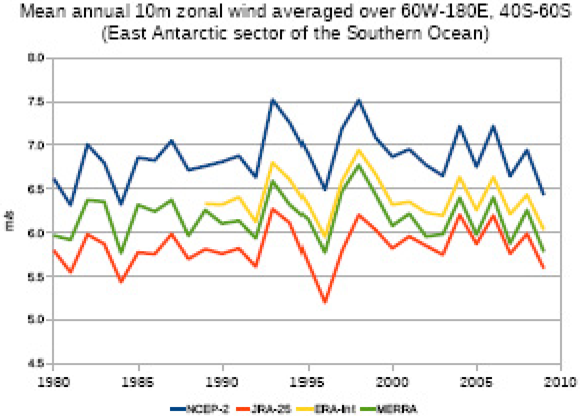
Spurious trends in MERRA precipitation

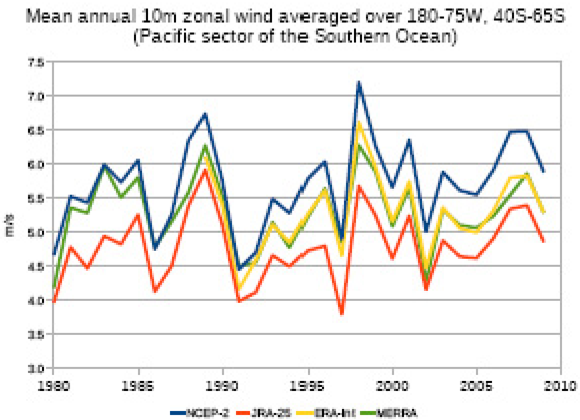


The figure shows the 2-month running average difference between forecast <u>daily</u> precipitation from MERRA and from ERA-Int, spatially averaged over the 50°S-60°S latitude band.

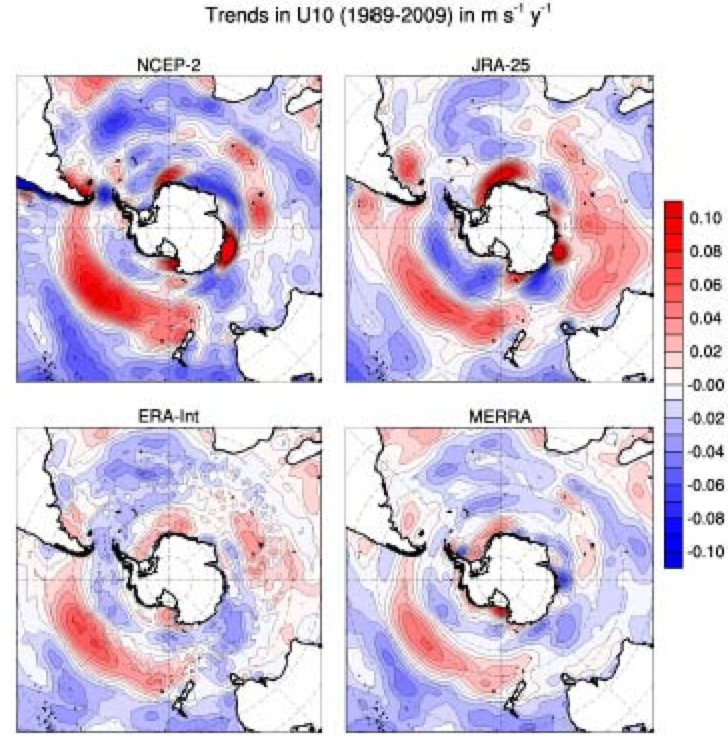
1989-2008 linear trends in annual P
(Southern Hemisphere)







NOTE: We included ERA-Interim for completeness but there are known issues with the zonal and meridional wind fields in this reanalysis dataset (as well as in ERA-40). Caution is required here.



Zonal Winds

D. Bromwich, private comm. 2010

Air-sea flux imbalances in analyses & reanalyses are a major problem when used as boundary conditions on the ocean circulation. Annual mean imbalances:

	mean [cm/year]
NCEP/NCAR-I ocean $E-P$	15.1
NCEP/NCAR-I ocean $E-P-R$	6.2
ECCO-GODAE ocean $E-P-R$	3.9
NCEP/NCAR-I global $E-P$	6.1
NCEP/DOE-II global $E-P$	-73.9

Similar heat budget issues

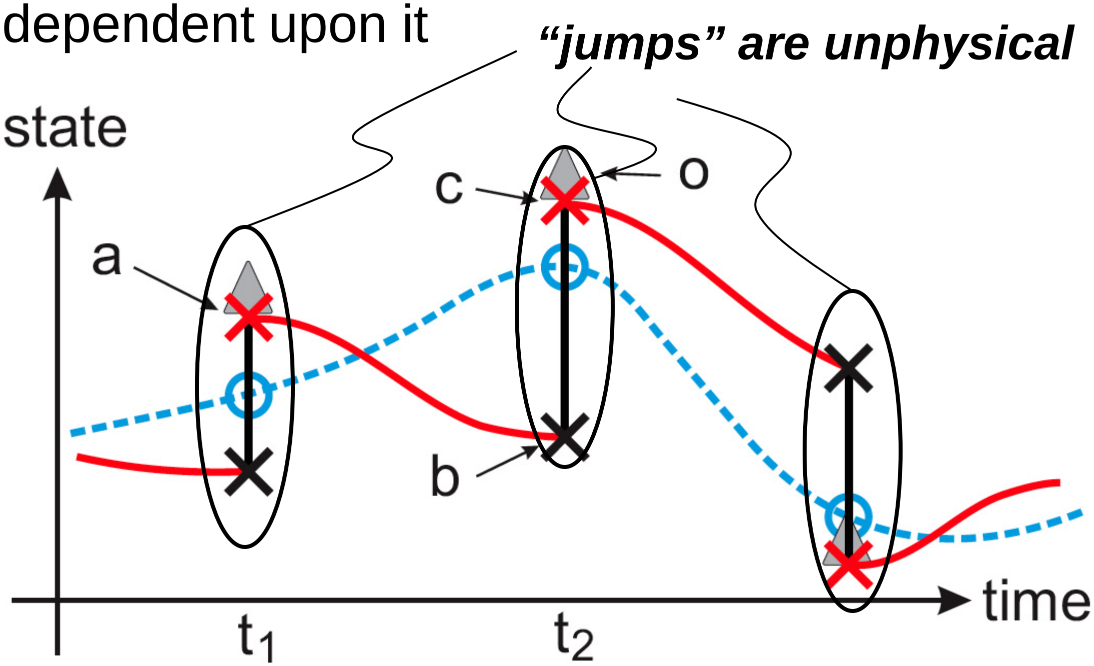
The reanalyses are derived from *weather forecast* models in which global water/heat balances are of no concern.

What is the runoff rate? How much does it vary? How much is climatological ice melt? Difficult to mode the ocean state.

None of this would be of serious zero order concern if the estimates were accompanied by error bars.

Estimation (interpolation) vs. forecasting (extrapolation)

- Atmosphere
  - Relatively abundant data sampling of the 3-dim. atmosphere
  - Most DA applications target the problem of optimal forecasting
  - ➔ find initial conditions which produce best possible forecast;
  - ➔ *dynamical consistency or property conservation *NOT* required*
- Ocean
  - Very sparse data sampling of the 3-dim. ocean
  - Understanding past & present state of the ocean is a major issue all by itself, the forecasting dependent upon it
  - ➔ use observations in an optimal way to extract max. information about oceanic changes
  - ➔ *dynamic consistency & property conservation *ESSENTIAL** over climate time scales

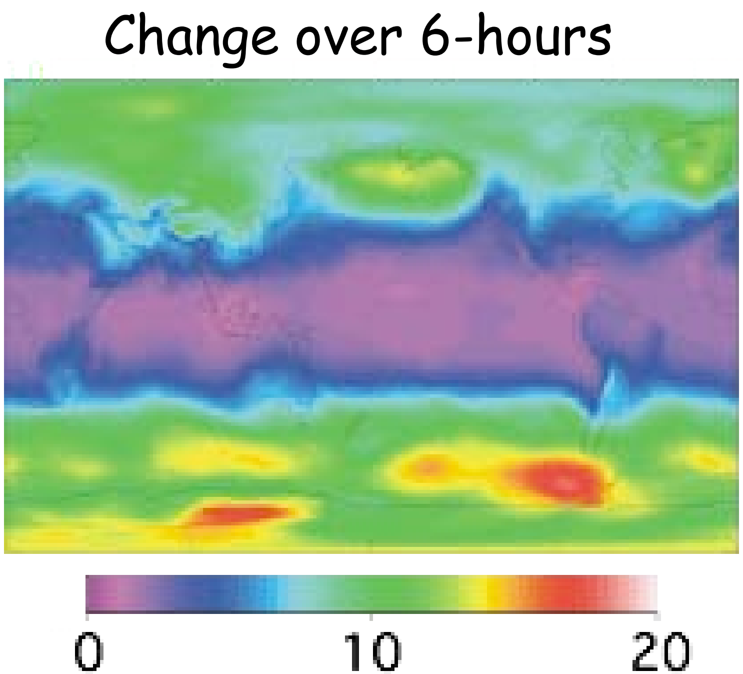


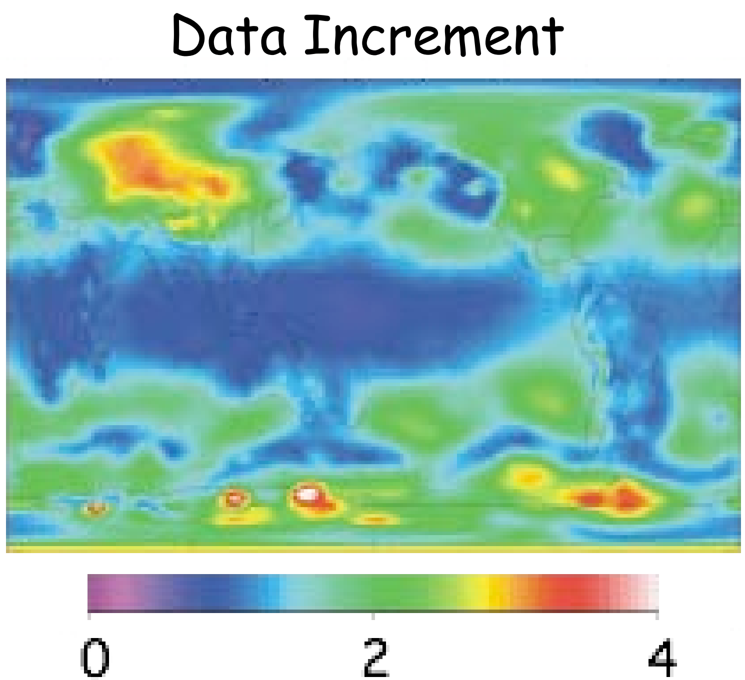
Importance of a physically consistent solution

Atmospheric reanalyses contain large air-sea flux imbalances. For example, the NCEP/NCAR reanalysis has an ocean freshwater flux imbalance of **6.2 cm/yr**, about 20 times larger than the observed **3 mm/yr** sea level rise.

They also contain discontinuities during “assimilation” updates. For example, standard deviation of NCEP **surface pressure** analysis shows that **24%** of the atmosphere’s mass change is physically **unaccounted for** (I. Fukumori, JPL).

mass conservation?



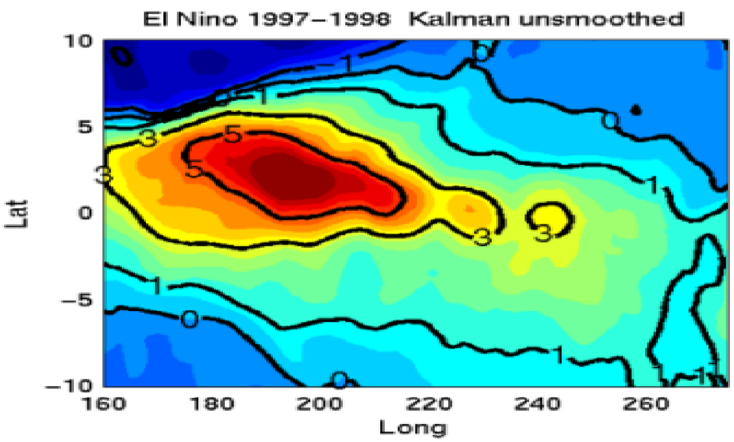


mbar

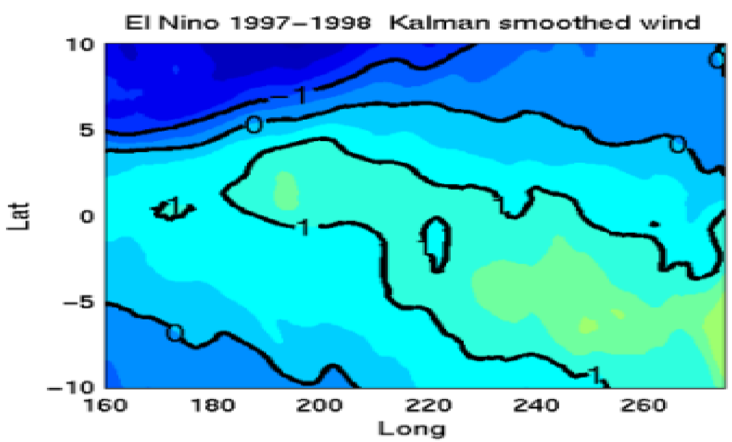
Contrary to atmospheric data assimilation, whose primary objective is NWP, need climate solutions satisfy model equations exactly, for example, conserving tracer properties.

Example tracer application: CO_2 Sea Air Flux

Estimate of CO_2 flux during 97-98 El Niño (mol/m2/yr) based on Kalman filter solution

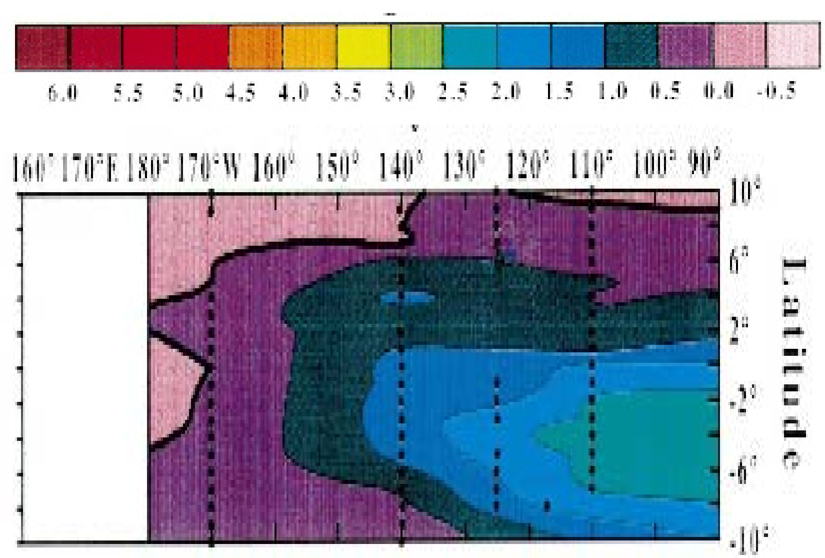


Estimate based on smoothed solution



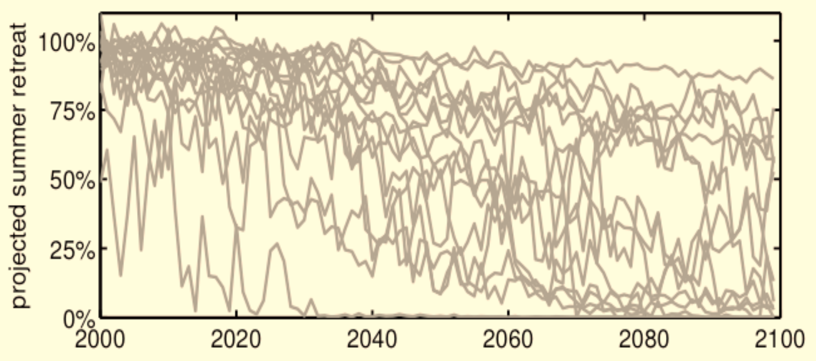
McKinley, 2002

Observed estimate of CO_2 flux during 92-93 El Niño (mol/m^2/yr)



Feely et al., 1999

Can one predict from these solutions?



IPCC AR4 scenarios for Arctic September ice cover 100 years into the future.
(I. Eisenman, J. Wettlaufer, 2007) Models were all tuned to recent conditions.

Climate models now contain nearly 1 million lines of computer
code and have been assembled over 50 years by hundreds of individuals.

ECCO Origins and Goals (1992+)

- Origin in WOCE: first global data sets. Exploit, particularly, the satellite data (altimetry, scatterometry) plus *all* of the in situ observations.
- Directed at *global* climate scales (*decadal*+)---contrast to short-term, non-globally balanced estimates
- Initially, intended as a demonstration---no one had ever done anything like this in the ocean before. Computers of 1992 were not really adequate.
- *Remains unique*

Most so-called data assimilation efforts are based upon the experience with numerical weather forecasting. ECCO *deliberately* did not follow that route for several reasons:

State estimation and prediction are very different goals. (Think of prediction as extrapolation, state estimation as interpolation.)

NWP (as used in the "reanalyses") does *not* conserve global water, energy, momentum, etc. These are essential for understanding climate change.

NWP methods introduce non-physical state jumps at the 6-hourly analysis times. These render impossible the calculation of budgets (freshwater, heat) essential to understanding climate.

NWP methods, more generally, discard information about prior times available from formal future observations.

Compared to data assimilation, the ECCO choices lead to a much higher computational load as well as the necessity for a major effort to assign error estimates to the huge data sets.

Our conclusion was that this price had to be paid in order to understand the climate system.

Ocean State Estimation

- **Synthesize …**

  - … all (diverse/disparate) types of observations,
  - … taking optimal advantage of sparse observations,
  - … with best-known dynamics/physics,
  - … into dynamically consistent, time-evolving estimate,
  - … obeying known conservation laws,
  - … to enable time-varying budget calculations,
  - … and diagnostic of un-observable quantities (e.g., MOC)
  - … with quantification of posterior uncertainties

ECCO-GODAE estimates are from ordinary least-squares solutions obtained by “adjoining” the model to a model-data misfit function using an ancient mathematical trick: Lagrange multipliers:

$$J = [\mathbf{x}(0) - \mathbf{x}_0]^T \mathbf{P}(0)^{-1} [\mathbf{x}(0) - \tilde{\mathbf{x}}_0]$$

misfit to Initial conditions

$$+ \sum_{t=1}^{t_f} [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)]^T \mathbf{R}(t)^{-1} [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)]$$

misfit to the observations

$$+ \sum_{t=0}^{t_f-1} \mathbf{u}(t)^T \mathbf{Q}(t)^{-1} \mathbf{u}(t)$$

adjustable parameters (controls) includes model error

$$- 2 \sum_{t=1}^{t_f} \mu(t)^T [\mathbf{x}(t) - \mathbf{L}[\mathbf{x}(t-1), \mathbf{B}\mathbf{q}(t-1), \Gamma \mathbf{u}(t-1)]].$$

the model

and seek the stationary point.

vectors of Lagrange multipliers, AKA, the adjoint or dual solution

In electrical engineering, called the Pontryagin Minimum Principle, in meteorology 4DVAR, in oceanography the adjoint method, …. Just least-constrained squares.

Very flexible: can cope with *any* constraint (e.g. averages over any time and/or space span) that can be written down in ths form. **R,Q** can have arbitrarily complex spatial structures (if known).

Two major difficulties: the size of the problem, and the need to understand errors in everything.

Finding the stationary point(s) of J is a (numerical) engineering problem. Not easy, but the principle is not in doubt. The *Kalman filter is not* a method for finding that point (was invented by Kalman to solve a prediction problem and fails to use any data in the formal future)

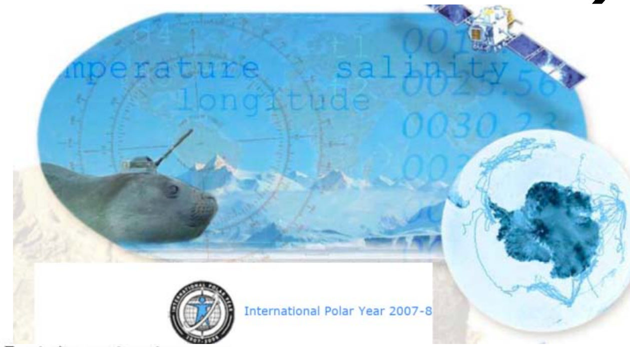
Our particular approach has been to use automatic/algorithmic differentiation (AD) software tools. Solved by iteration relying upon knowledge of the partial derivatives of J with respect to x(t), u(t), using. Will skip all that here.

After (near) optimization, the solution analyzed is from the freely running gcm---thus it satisfies the known model equations and can be used e.g., to calculate energy or heat or mass budgets.

DATA TYPE	Source	Spatial Extent	Variable(s)	Duration	Number of values
Altimetry: TOPEX/POSEIDON	PODAAC	Global, equatorward of 66.5 degrees	height anomaly, temporal average	1993-2002 2003-2005	(4500/day) 3.0×10^{7}
Altimetry: Jason	PODAAC	Global equatorward of 66.5 degrees	height anomaly, temporal average	2002-2007	included above
Altimetry: Geosat-followon	US Navy, NOAA	Global, equatorward of 72 degrees	height anomaly	2001-2007	(4300/day) 2.6×10^{7}
Altimetry: ERS-1/2, ENVISAT	AVISO	Global, equatorward of 81.5 degrees	height anomaly	1992-2007	(3800/day) 2.2×10^{7}
Hydrographic climatology	Gouretski and Koltermann (2004)	global, 300m to seafloor	temperature, salinity	inhomogeneous average	(monthly) 1.7×10^{7}
Hydrographic climatology	World Ocean Atlas (2001), Conkright et al. (2002)	global to 300m	temperature, salinity	average seasonal cycle	included above
CTD synoptic section data	Various, including WOCE Hydro. Prog.	global, all seasons, to 3000m	temperature, salinity	1992-2005	(17000 profilx10s) $2.\times10^{6}$
XBTs	D. Behringer (NCEP)	global, but little So. Ocean	temperature	1992-2006	(470000 profiles) 1.2×10^{7}
ARGO Float profiles	IFREMER	global, above 2500m	temperature, salinity	1992-2007	(416000 profiles) 7×10^{6}
Sea Surface Temperature	Reynolds and Smith (1999)	global	temperature	1992-2007	(monthly) 8.0×10^{6}
Sea Surface Salinity	Etudes Climatiques de l'Oce'an Pacifique (ECOP)	tropical Pacific	salinity	1992-1999	(monthly) 5.5×10^{6}
TMI AMSRE	NASA/NOAA discoverearth.org	global	temperature	1998-2007	(daily) 2.1×10^{8}
Geoid (GRACE mission)	GRACE SM004-GRACE3 CLS/GFZ (H. M. Rio)	global	mean dynamic topography	NA	(1 deg) 5.8×10^{4}
Bottom Topography	Smith&Sandwell(1997)+ETOP05	Smith/Sandwell to 72.006, ETOP05 to 79.5	water depth	NA	(1 deg) 5.8×10^{4}
Toga-TAO, Pirata array	PMEL, NOAA	tropical Pacific	temperature, salinity	1992-2006	(daily) 2.2×10^{6}
SeaOS	Sea Mammal Research U. St. Andrews, Scotland	Southern ocean	temperature, salinity	2004-2007	(24590 profiles) 5.6×10^{5}
Rapid	BODC	Atlantic 26N	temperature, salinity	2004-2005	8.4×10^{5}
Florida Current transport	NOAA/AOML	Florida Straits	Transport	1992-2007	(1value/day) 5.8×10^{3}
FORCING:					
Windstress-scatterometer	PODAAC	global	stress	1992-2006	9×10^{8}
Windstress	NCEP/NCAR reanalysis Kalnay et al. (1996)	global	stress	1992-2007	(192x94-6hr) 4.5×10^{8}
Heat Flux	NCEP/NCAR reanalysis	global	lw+sensible+latent heat	1992-2007	(192x94-6hr) 2.3×10^{8}
Freshwater Flux	NCEP/NCAR reanalysis	global	evap-precip	1992-2007	(192x94-6hr) $2..3\times10^{8}$
Short/long Wave Radiation (experimental)	NCEP/NCAR reanalysis	global	Sw	1992-2007	(192x94-6hr) 2.3×10^{8}
					Total variables 3.1×10^{9}
WITHHELD (as of April 2006):					
tide gauges		global, sparse			
Tomographic integrals		N. Pacific	velocity/temperature		
Float and Drifter Velocities		Global	heat content		

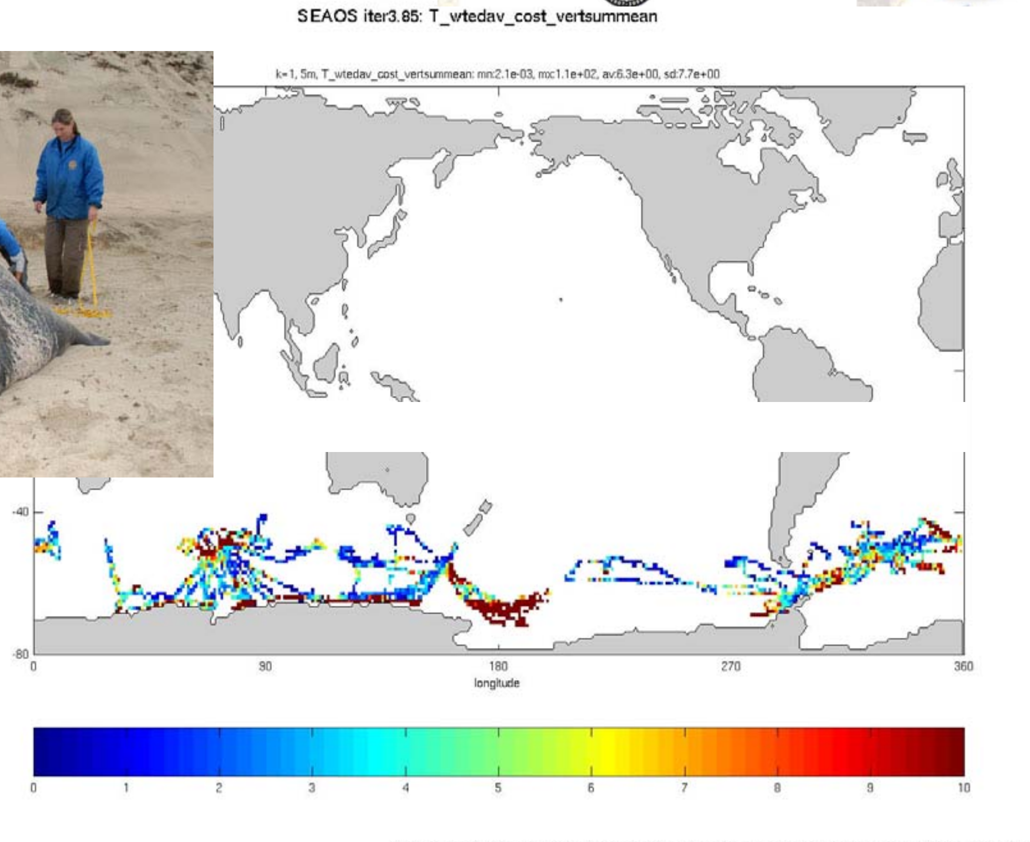
Southern Elephant Seals as Oceanographic Samplers (SEaOS)

- CTD-type observations from seals in SO



Sea Mammal Research Unit,
University St. Andrews, UK,
British Antarctic Survey

Courtesy M. Meredith

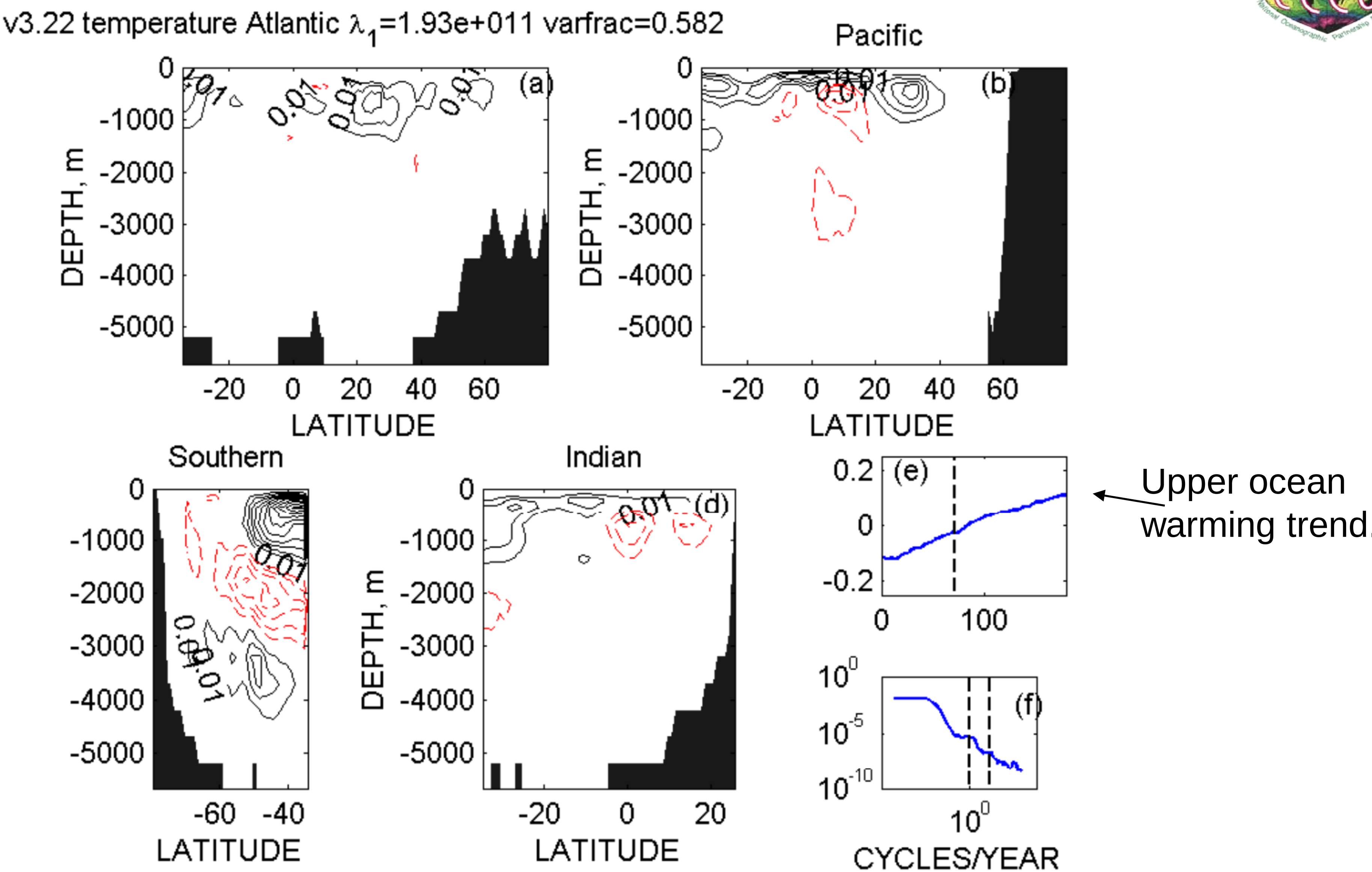


No animals were harmed in the making of this Powerpoint presentation

In round numbers, the ECCO-GODAE effort at MIT is using 2 billion observations over 16 years (including meteorological estimates), with 10^{12} unknowns (both state vector and control vector values).

It does work (see many published papers http://www.ecco-group.org). Computationally painful.

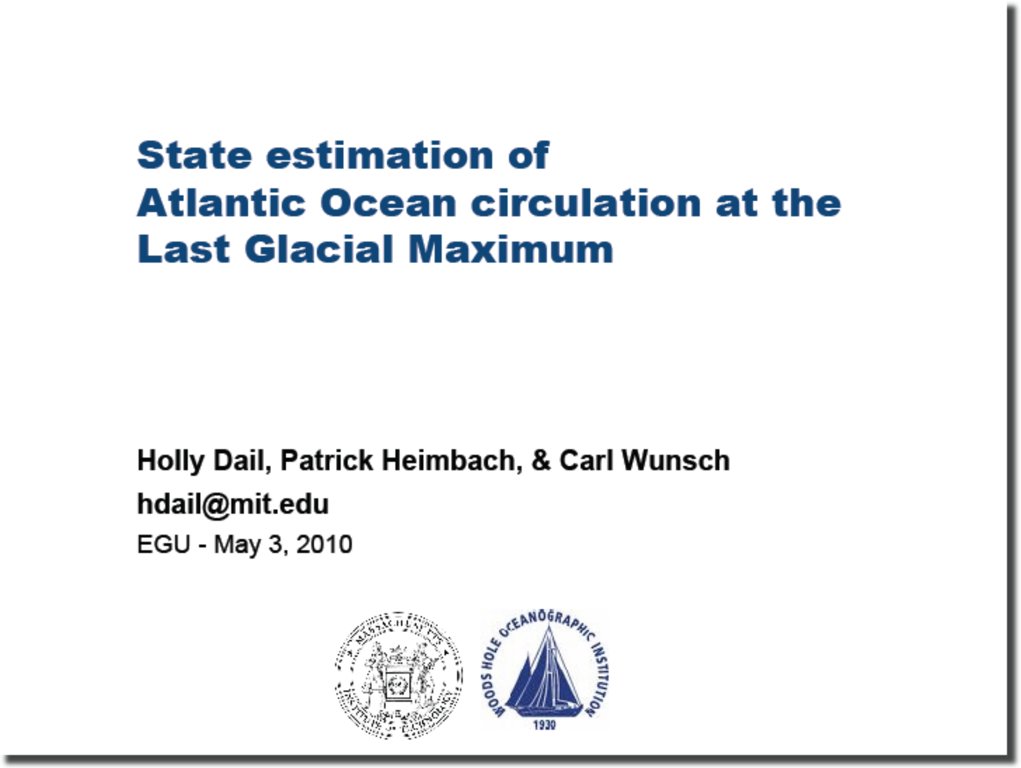
EOF 1 Temperature 58% of var.



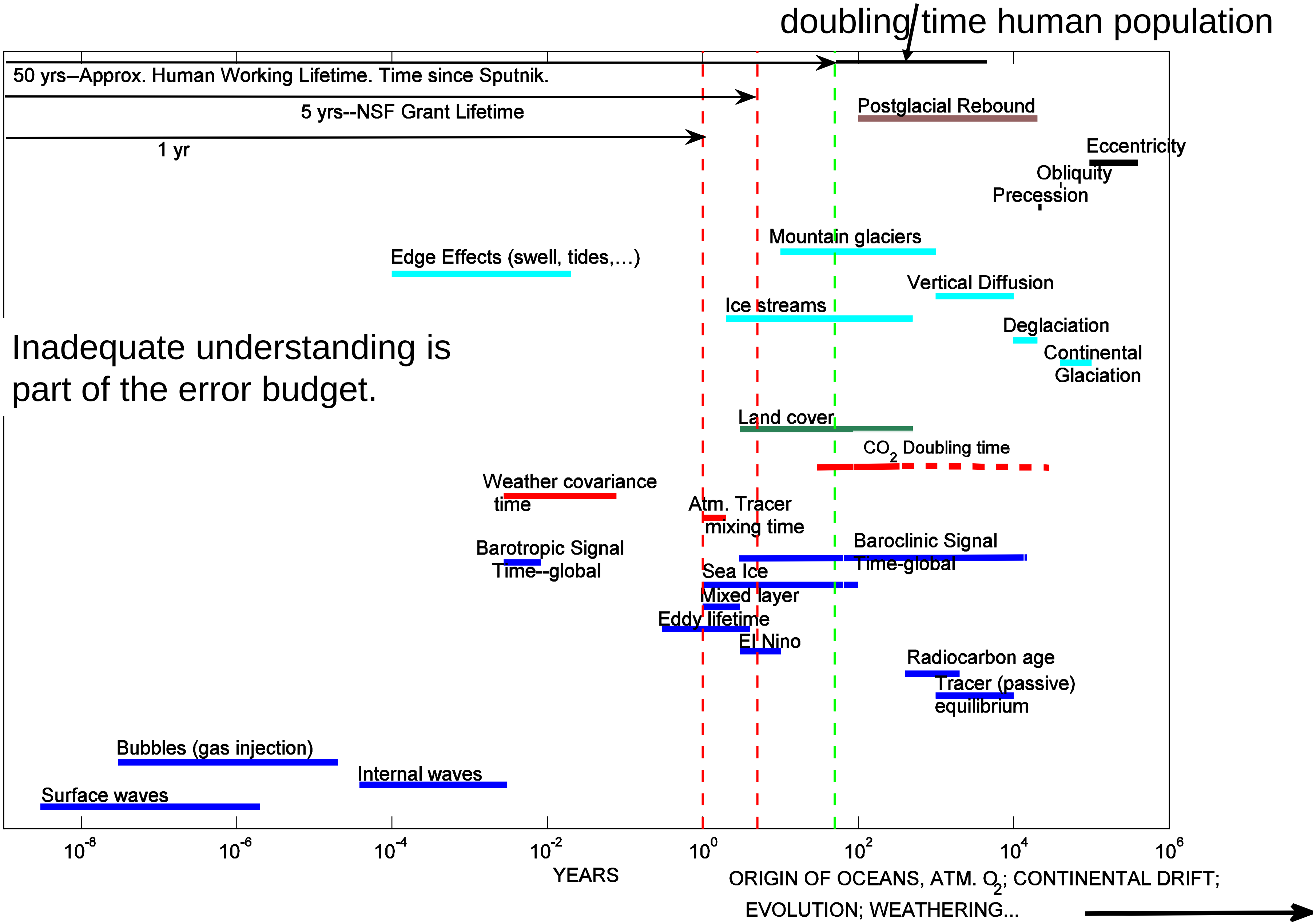
Warming is an artifact of the NCEP/NCAR reanalysis. Upper ocean observations are inadequate to preclude it---despite all the measurements!

Works also for the paleo problem: Example (preliminary) results from PhD thesis of Holly Dail: the ocean circulation during the last glacial maximum (LGM)

Dail presentation:



Some Time Scales of the Climate System



How to characterize model error? Resolution issues loom largest.

1/9deg

mixed layer depth

1deg.

1/54 deg.

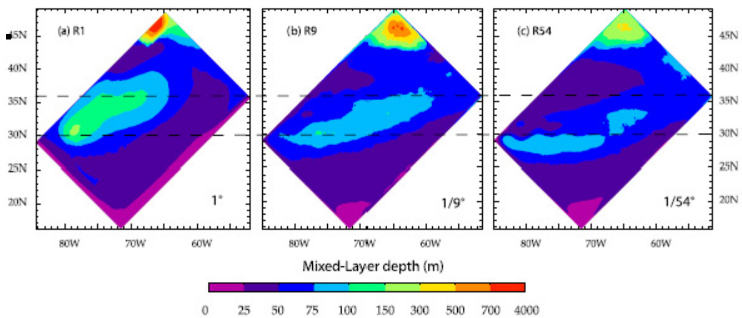


Fig. 9. Ten-year-mean mixed-layer depth (MLD) in experiments R1, R9 and R54. The MLD is computed as the interface of the surface layer whose density does not exceed the surface density by more than 0.01.

Lévy et al., Ocean Mod., 2010

mean density along 72W

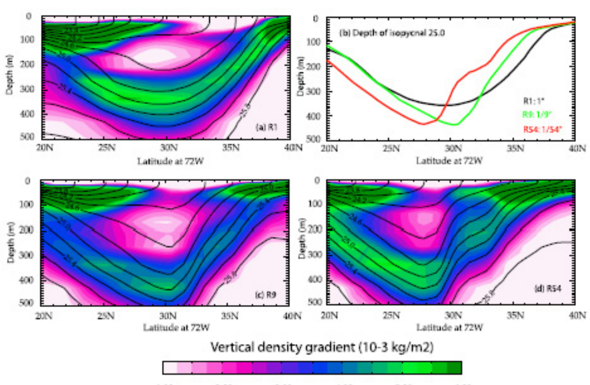


Fig. 11. Ten-year-mean density (black contours) along a section at 72°W in experiments R1, R9 and R54. The colors show the intensity of the vertical density gradient. To facilitate the comparison, the depth of the 25.0 isopycnal is reported in panel (d) for the three experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

meridional heat transport

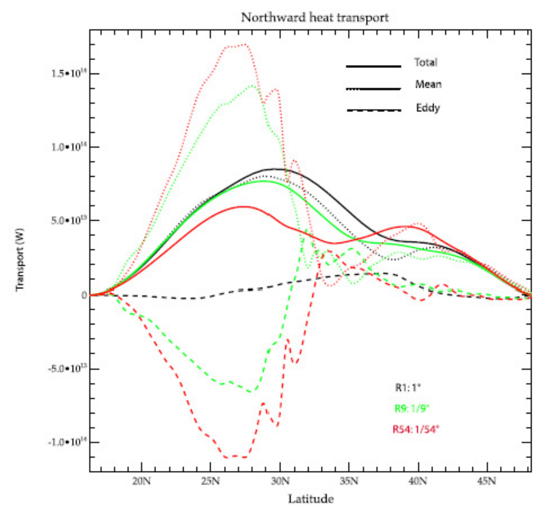
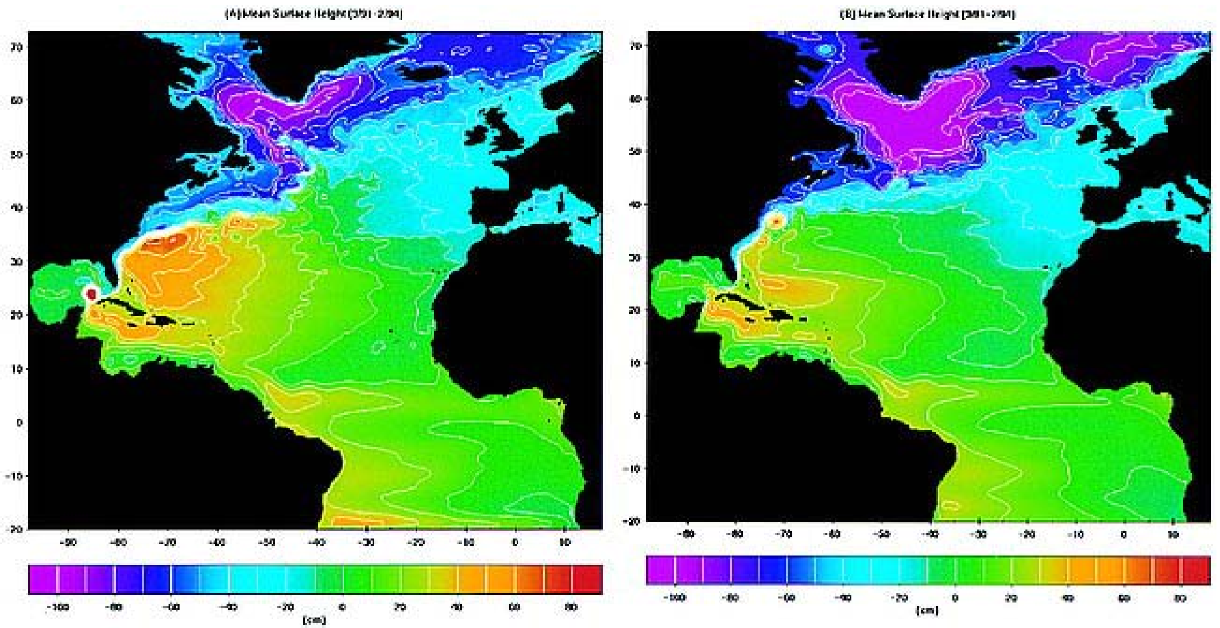


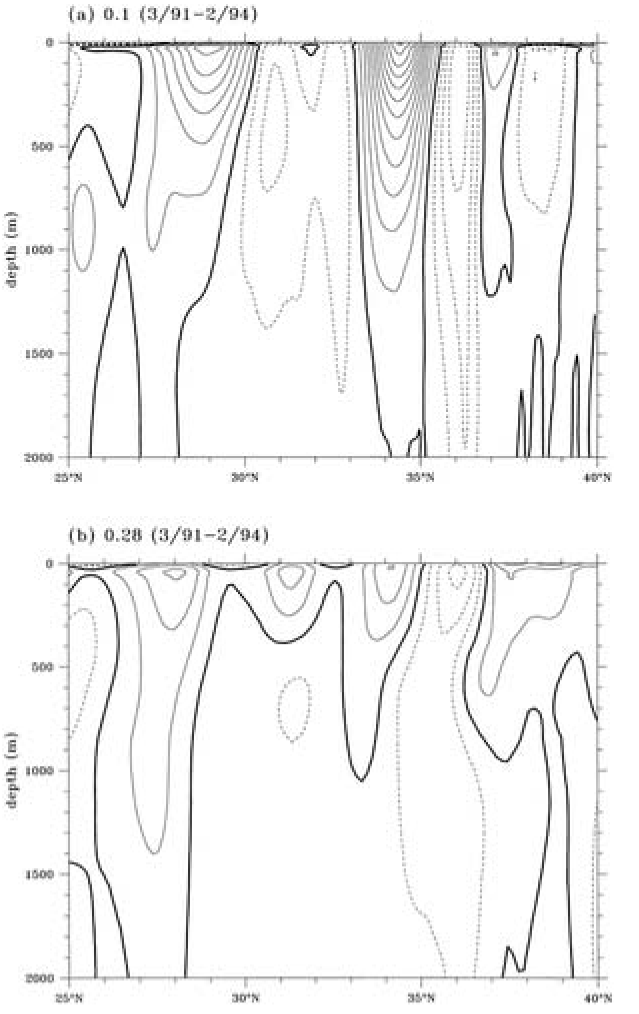
Fig. 12. One-year-mean northward heat transport (in W) in experiments R1 (black), R9 (green) and R54 (red). The plain line shows the "total" heat transport, computed from the integration of 1 year-mean meridional heat fluxes. The dotted line shows the "mean" heat transport, computed from the 1 year-mean flow and 1 year-mean temperature distribution. The dashed line shows the "eddy" contribution, computed as the difference between the "total" and "mean" contributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)



Time mean sea surface height,0.1deg,0.28deg models

Smith, Maltrud, Bryan, Hecht, 2000, JPO

An old-fashioned question: is numerical convergence an issue? What equations are being solved?



EKE Meridional sections of the time-mean zonal velocity averaged between 35°W and 25°W for (a) the 0.1° run and (b) the 0.28° run.

The error estimate problem dominates the subject. Two major divisions:

state estimates
forecasts

Classes of errors:
Resolution failure coupled with inadequate subgrid-scale parameterizations
Missing physics (e.g., sea-ice ocean interactions)
Mis-specified data errors, including inadequate spatial/temporal covariances
Erroneous initial conditions
Erroneous boundary conditions (mainly in sub-system modeling, e.g., the ocean driven by an overlying atmosphere).

A mixture of systematic and stochastic elements.

How to estimate uncertainties/errors?
Several approaches:

$$J = A + \mathbf{x}^T \mathbf{H}^{-1} \mathbf{x}$$ ‡

(1) Inverse Hessian

where
$\mathbf{x} = [\mathbf{x}(0), \mathbf{x}(1), \ldots \mathbf{x}(t_f), \mathbf{u}(0), \mathbf{u}(1), \ldots, \mathbf{u}(t_f - 1)]$
and $\mathbf{H}$ is the Hessian. Find the inverse Hessian and then, using the RMS misfits, have the linearized uncertainty. Inverse Hessian is square of the dimension of the combined state and control vectors. A linear estimate.

(2)Fokker-Planck equation

$$\frac{\partial P(\mathbf{x}(t))}{\partial t} + \frac{\partial}{\partial \mathbf{x}(t)} \left\{ \mathbf{L}(\mathbf{x}(t), \ldots) P(\mathbf{x}) - \theta(t) \frac{\partial P(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \right\} = 0$$

(Fokker-Planck equation) where $P(\mathbf{x}(t))$ is the probability density of the state vector, $\mathbf{L}$ is the model time-stepping operator and $\theta(t)$ is a stochastic noise process.

(3) Singular vectors at initial times
(4) Eigenvectors at initial times
(5) Monte-Carlo (Ensemble methods)
….

See e.g., T. Palmer, 2000, Rep. Prog. Phys.

Dimensionality is a problem with all of them. (1) and (3) are linearizations. Ensemble sizes are minute compared to the state and control vector dimensions (thus, singular covariances and unexplored uncertainty spaces); colored (space-time) structures unaccounted for.

No one wants the uncertainty of the individual elements of $\mathbf{x}(t)$, $\mathbf{u}(t)$, but normally some consequences of them. But which consequences should be the focus?

Stored heat content, uptake rate of carbon, sea surface temperatures, snow cover, sea ice extent in September, North Atlantic precipitation trends,….

How to sort these into some priorities? Want methods that permit a fairly general application.

The same issues pertain to forecasts of climate elements---but no data are involved at all (simpler?)

Problem Priorities:

(1) Error estimates. Estimates are nearly useless without them. Forecasts are impossible to interpret.

(2) Full climate system estimates (ocean; atmosphere; cryosphere---particularly sea ice, but including glaciers; ultimately biological activity)

(3) Computational efficiency improvements so resolution issues can be addressed.

(4) Much better understanding of trends in key climate measurements (radiation, water vapor, etc.)

(5) Efficient estimates for non-Gaussian statistics

(6) Model errors from lack of resolution, mis-constructed parameterizations

(7) Construction of spatial covariances of data error

A major problem:

It is clear that the atmospheric reanalyses fail as climate state products as they do not have closed budgets and exhibit discontinuities leading to unphysical trends.

A self-consistent climate state estimate should be carried out using a fully coupled atmosphere-ocean-cryosphere (land and sea ice) system, and estimation methods using realistic error estimates for all data types, and producing dynamically self-consistent results.

The principle is clear. The doing of it is ill-matched to a small academic group. How will this ultimately be carried out? (Probably not by the existing reanalysis groups.)

Thank you.