Ensembles for long-term earth system modelling

J.C. Hargreaves, J.D. Annan,

Frontier Research System for Global Change
Yokohama
Japan

and

Proudman Oceanographic Laboratory
Birkenhead
UK

July 2003

Introduction

- In collaboration with the GENIE project, we are building an ensemble modelling system based on an ensemble Kalman filtering approach.

- The model is the low resolution coupled ocean-atmosphere model C-GOLDSTEIN, which forms the dynamical core of the GENIE paleo-climate Earth System Model now under development.

- We have implemented an ensemble Kalman filter to reinitialise the ensemble. A novel development is the use of differing (variable) parameters for the ensemble members, which are updated along with the state variables.

GENIE – Grid ENabled Integrated Earth System Modelling

The GENIE model is a Grid-based, modular, distributed and scaleable Earth System Model.

The scientific focus is on long-term and paleo-climate change, especially through the last glacial maximum (21kyr BP) to the present interglacial, and the future long-term response of the Earth system to human activities.

A realistic ESM for this purpose must include models of the atmosphere, ocean, sea-ice, marine sediments, land surface, vegetation and soil, ice sheets and the energy, biogeochemical and hydrological cycling within and between components.

The GENIE Team

Principal Investigator

Paul Valdes	University of Reading	p.j.valdes@reading.ac.uk

Project Coordinator

Tim Lenton	CEH Edinburgh	tlent@ceh.ac.uk

Management Team

Melvin Cannell	Edinburgh CEH Edinburgh
Simon Cox	Southampton e-Science Centre
John Darlington	London e-Science Centre
Richard Harding	CEH Wallingford
Steven Newhouse	London e-Science Centre
Tony Payne	University of Bristol
John Shepherd	University of Southampton
Andy Watson	University of East Anglia

Research team and Collaborators

James Annan	Frontier Research System for Global Change, Japan
Chris Brockwell	University of East Anglia
Peter Cox	Hadley Centre
Neil Edwards	University of Bern, Switzerland
Murtaza Gulamali	London e-Science Centre
Julia Hargreaves	Frontier Research System for Global Change, Japan
Phil Harris	CEH Wallingford
Bob Marsh	Southampton Oceanography Centre
Andrew Price	Southampton e-Science Centre
Andy Ridgwell	University of East Anglia SOMEWHERE IN THE USA
Andrew Yool	Southampton Oceanography Centre

C-Goldstein AO-EMIC – the backbone of GENIE

- Ocean — Frictional Geostrophic (Edwards and Shepherd 2002)
- Atmosphere — Energy and Moisture balance (simlilar to Weaver et al. 2001)
- Zero thickness dynamical and thermodynamic sea ice
- No dynamical land surface (yet): $T_{surf} = T_{atmos}$ over land, no evaporation, precipitation is added to coastal grid cells according to a runoff map

Parameter Estimation in Climate Models

Many model parameters are uncertain, and the parameter estimation problem has recently become prominent in climate modelling.

Typical approaches

(eg. Knutti et al.; Edwards; Schneider von Deimling and Held; Stainforth et al. - EGS April 2003)

- 2-10 parameters estimated simultaneously
- 1,000-2,000,000 model runs — cost may be exponential in number of parameters
- naive comparison with data

The Ensemble Kalman Filter

Most operational Data Assimilation is aimed at reducing initialisation error, while largely ignoring the model error.

Ensemble Kalman Filter is used in forecasting - produces an ensemble that spans the PDF of the model given the data and any prior information.

There is no reason why the model parameters cannot be incorporated into the ENKF, through augmenting the model state to include parameter values and augmenting the model with auxiliary ‘parameter equations’. It then becomes a tool for parameter estimation.

Our implementation - Ensemble Kalman Filter

Based on parallel domain decomposition algorithm of Kepenne (2000).

- 4-12 parameters estimated simultaneously
- 54 model runs, independent of number of parameters
- theoretically optimal use of data

An Identical Twin Test

We illustrate how the ENKF works for parameter estimation by an Identical Twin test.

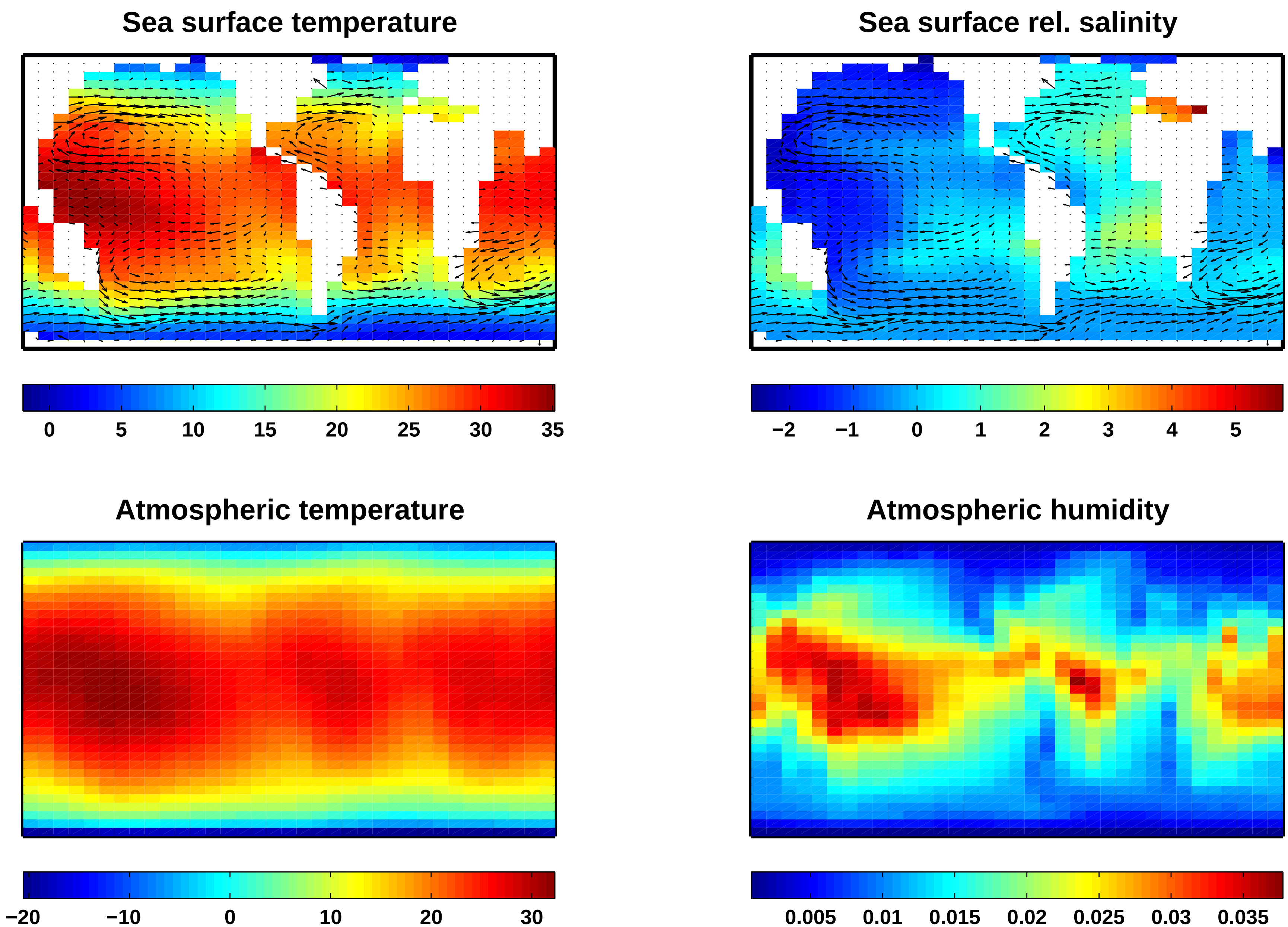
1. A 'truth' run, with a particular set of parameters is run to steady state.

2. A 54 member ensemble with incorrect (but identical) parameters is run to steady state from random initial conditions.

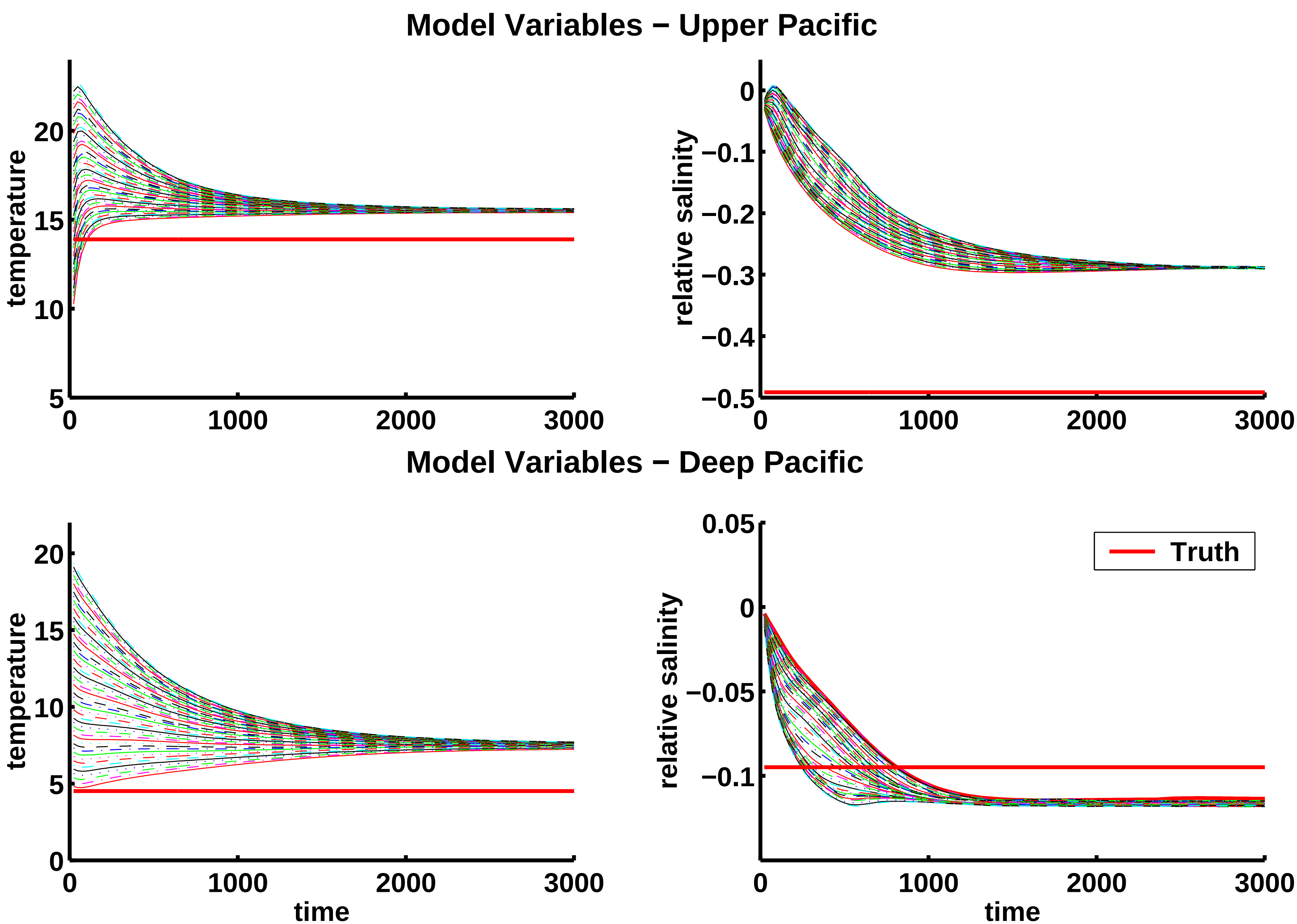
3. The standard ENKF is implemented in the model with incorrect parameters.

4. The parameter varying ENKF is implemented for the same model.

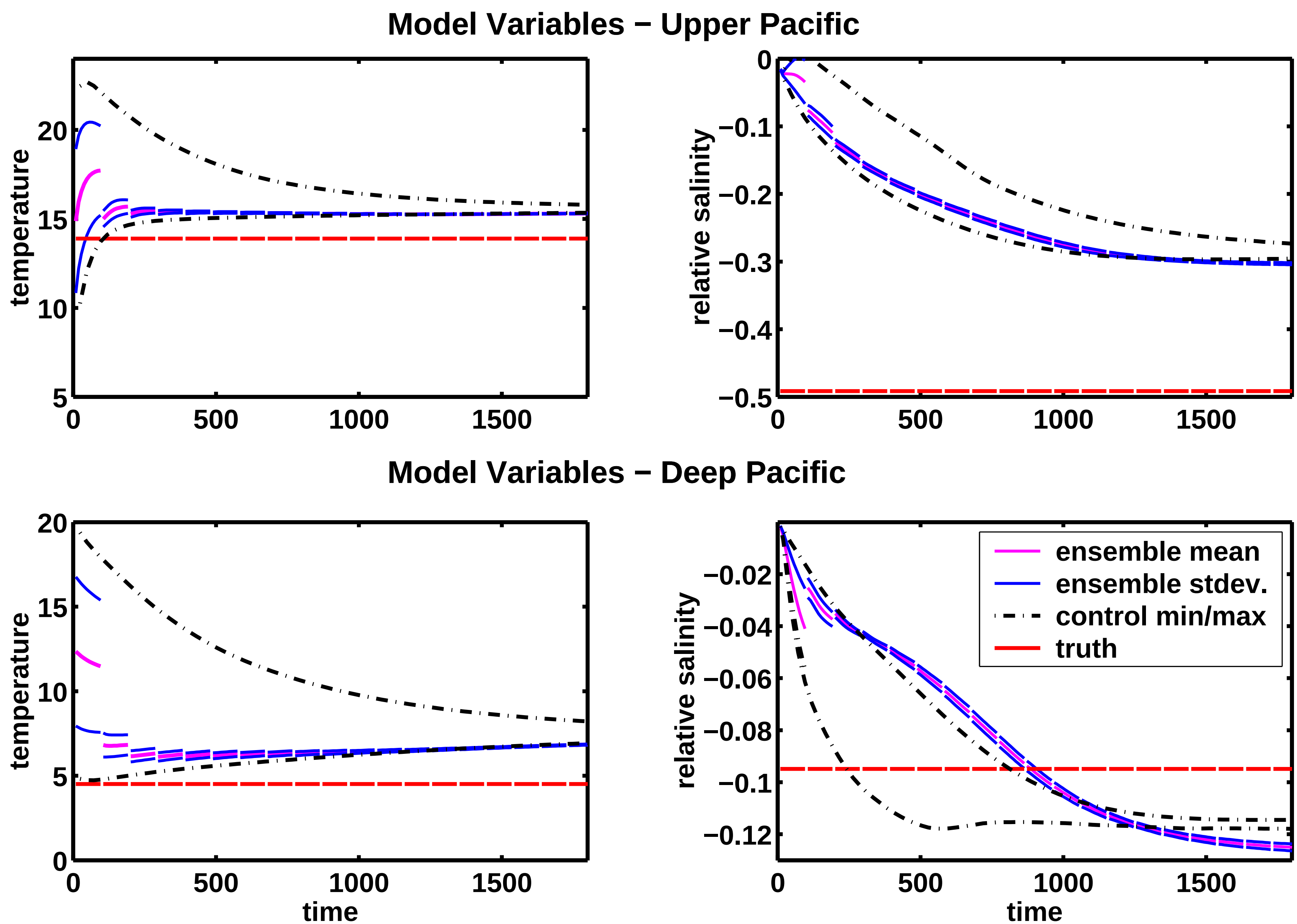
Typical model fields (used as 'truth' run)
AO-EMIC: C-Goldstein (Edwards and Marsh, 2003)



Model climatologies with wrong parameters

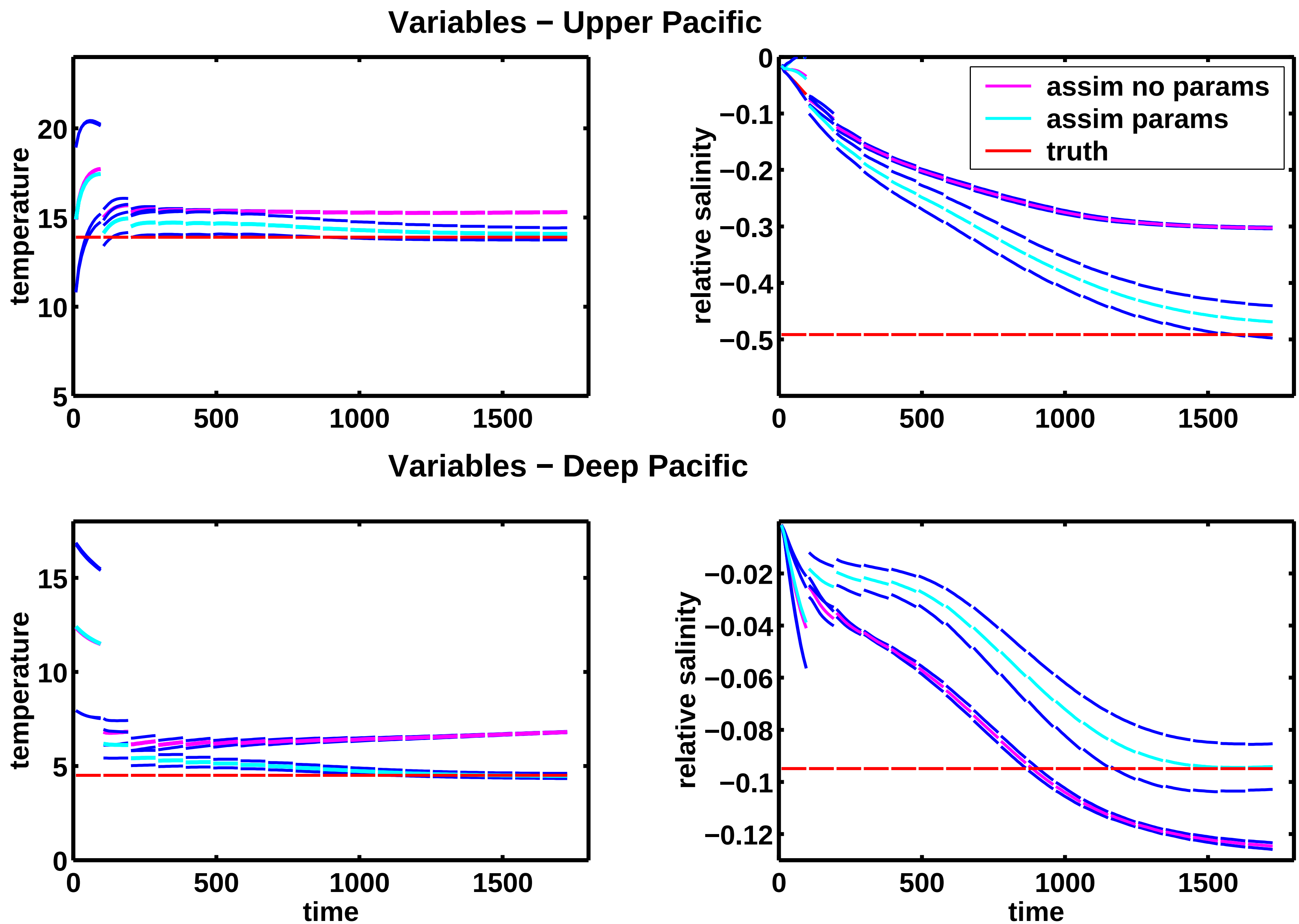


Assimilation of temperature with fixed (wrong) parameters



This illustrates the problem of false confidence mentioned previously

Comparison of fixed and varying parameters



A side note on climatological assimilation

In theory a steady-state climatology requires only a single analysis cycle:

Initialise model ensemble $\rightarrow$ Run to steady state $\rightarrow$ Assimilate data $\rightarrow$ We are done!

In practice:

- Few if any prior members are reasonably close to the posterior distribution.

- Many potential prior members are numerically unstable.

- The posterior ensemble members are not in balance due to the various nonlinearities in the model and numerical approximations in this implementation of the EnKF algorithm.

An alternative (novel?) method for climatological assimilation

Initialise with an arbitrary ensemble, assimilate the data and prior at regular intervals while also applying a modest ensemble inflation factor. Correction to the error estimate of the observations and prior is required to compensate.

The prior, observations and analysis covariances are related by:

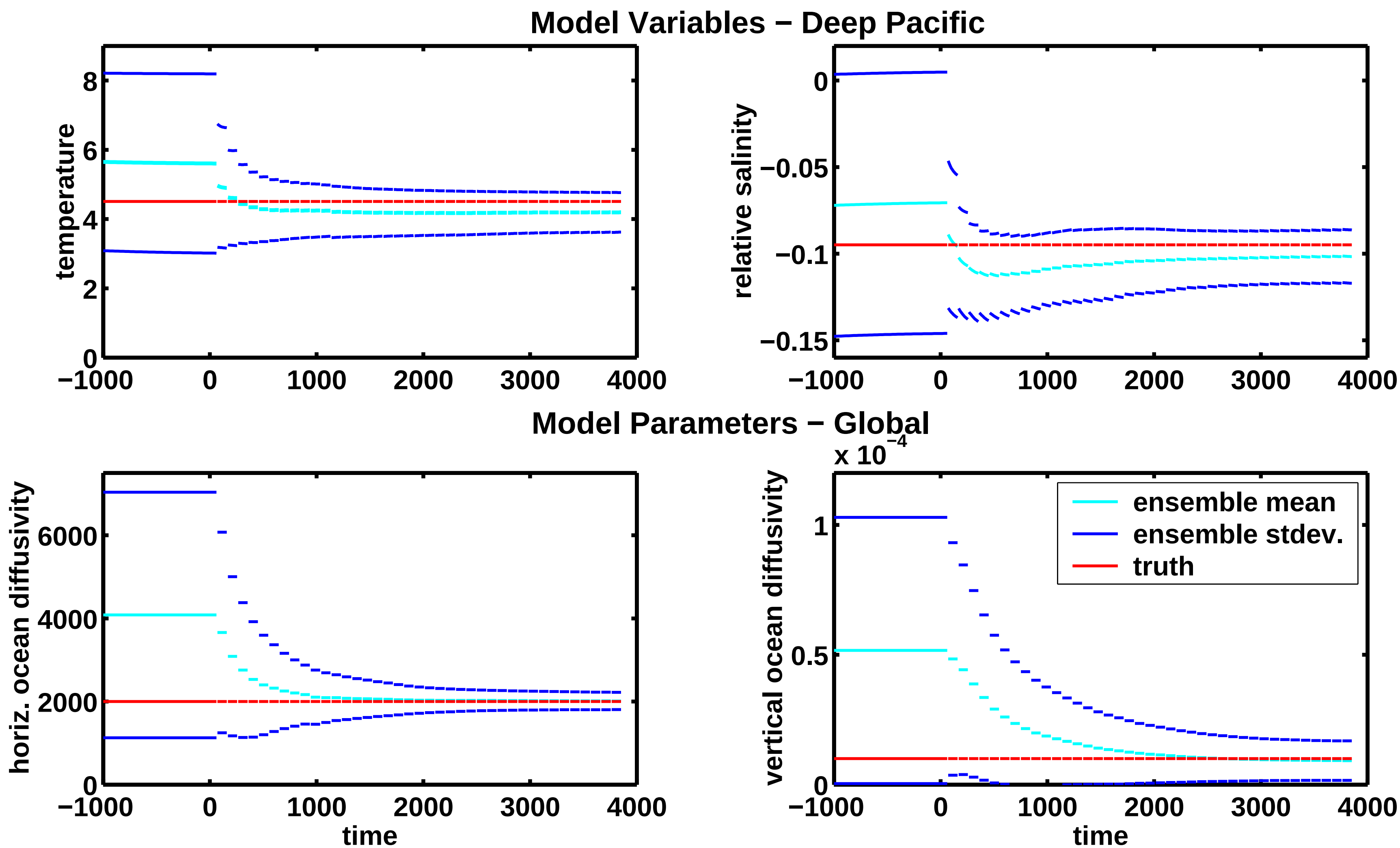
$$\frac{1}{\sigma_p^2} + \frac{1}{\sigma_o^2} = \frac{1}{\sigma_a^2}$$

If we repeatedly inflate the ensemble by a factor E, and reinsert the same prior and data with an error estimate artificially inflated by a factor C we should reach the same solution so long as we choose:

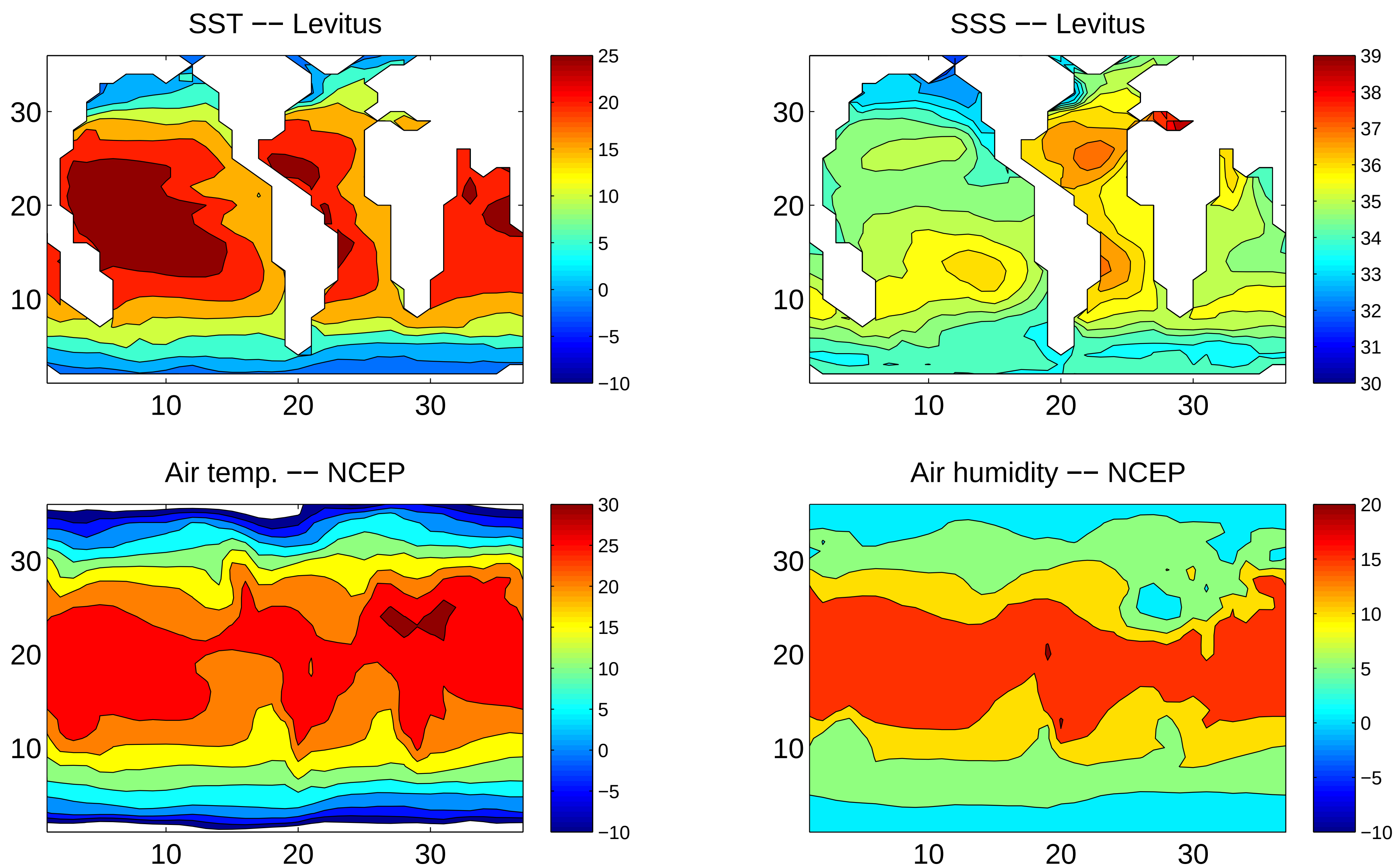
$$C^2 = \frac{E^2}{E^2 - 1}$$

The shocks due to innovations are much smaller, the ensemble members remain numerically stable and the system converges (pretty well)

Identical twin experiment with expansion factor (4 params varied)

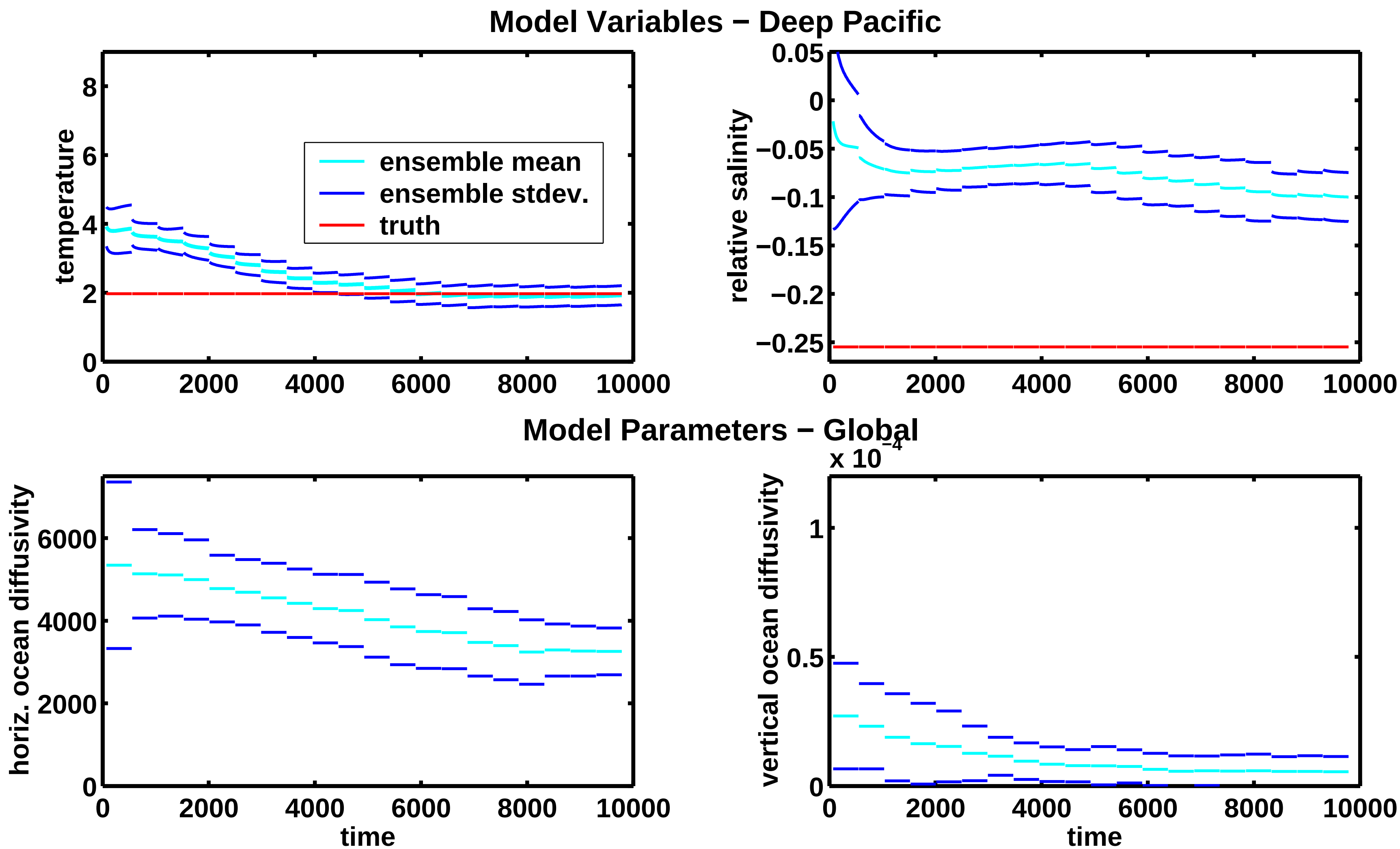


Assimilation with real data – The data

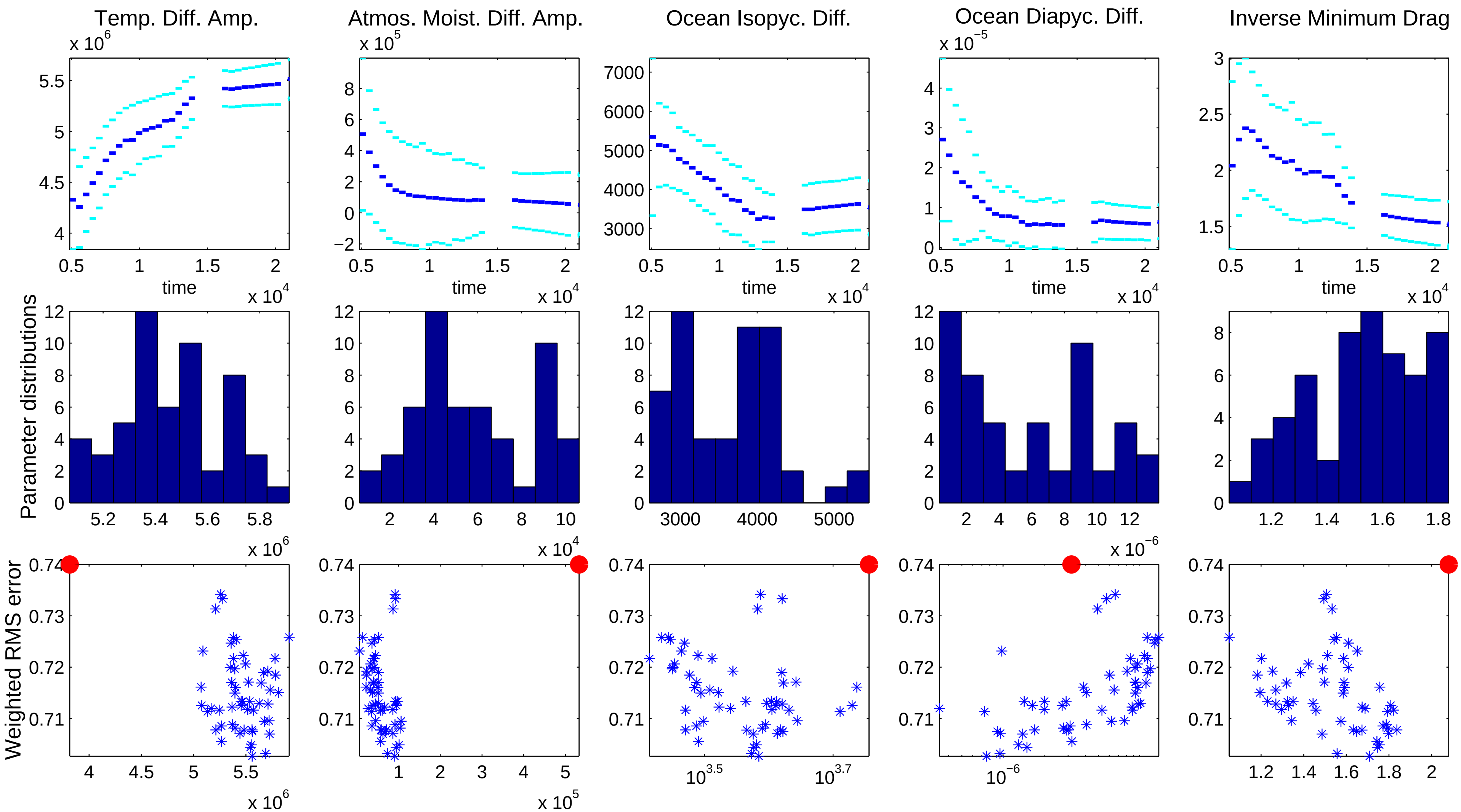


Levitus (1998) ocean temperature and salinity (3D); NCEP surafce (1000 mb) stmospheric temperature and humidity.

Assimilation with real data, varying 11 parameters



Assimilation with real data, varying 11 parameters



● = best fit from 1000 random parameter runs (next best fit, RMS difference =0.8)[Edwards, 2003]

Conclusions

We are using an EnKF for parameter estimation in an intermediate complexity climate model. Identical Twin tests work well and preliminary results with real data are promising.

Compared to ad-hoc parameter sampling techniques:-

- There is a smaller computational overhead
- The whole ensemble occupies a sparsely populated low error regime
- Meaningful parameter distributions can be produced.