POP+ DART

Parallel Ocean Program +

Data Assimilation Research Testbed

Accelerated Scientific Discovery Presentation NCAR HPC User Group May 2022 Meeting



Ben Johnson & Moha Gharamti DAReS · TDD · CISL · NCAR Anna-Lena Deppenmeier OS · CGD · NCAR Ian Grooms *CU-Boulder* Scientific Background Emergent Phenomena







Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS



Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE

SMALL-SCALE THERMODYNAMICS



Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE

SMALL-SCALE THERMODYNAMICS



Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE

SMALL-SCALE THERMODYNAMICS



Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS

Increasing resolution \rightarrow



Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS

Increasing resolution \rightarrow





Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS

Increasing resolution \rightarrow





Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS

Increasing resolution \rightarrow





Dynamical or Physical Regime

LAMINAR / BAROTROPIC

BAROCLINIC RESOLVING CYCLOSTROPHIC / SUBMESOSCALE SMALL-SCALE THERMODYNAMICS

Increasing resolution \rightarrow







Horizontal resolution

Dynamical or Physical Regime

LAMINAR /

BAROTROPIC

Loop current

Horizontal resolution

IN THE OCEAN

SMALL-SCALE **THERMODYNAMICS**

CYCLOSTROPHIC / **SUBMESOSCALE**

BAROCLINIC

RESOLVING



Software Components Community Earth System Model + Data Assimilation Research Testbed

Model Hierarchy Parallel Ocean Program (POP) in Eddy-Resolving and Eddy-Parameterizing Configurations





Horizontal resolution

Dynamical or LAMINAR / CYCLOSTROPHIC / SMALL-SCALE BAROCLINIC Physical BAROTROPIC SUBMESOSCALE RESOLVING **THERMODYNAMICS** Regime 1° latitude/longitude 0.1° latitude/longitude Horizontal resolution 1 km – 100 m > 10 m IN THE **OCEAN** Loop current



> 10 m

Horizontal resolution

Dynamical or LAMINAR / CYCLOSTROPHIC / SMALL-SCALE BAROCLINIC Physical BAROTROPIC **SUBMESOSCALE** RESOLVING **THERMODYNAMICS** Regime 1° latitude/longitude 0.1° latitude/longitude Horizontal resolution 1 km – 100 m Low Resolution Ocean Configuration (g17 grid) Loop current

High Resolution Ocean Configuration (t13 grid)

Loop current Horizontal resolution 1 km – 100 m 1° latitude/longitude 0.1° latitude/longitude > 10 m Dynamical or LAMINAR / CYCLOSTROPHIC / SMALL-SCALE BAROCLINIC Physical BAROTROPIC **SUBMESOSCALE** RESOLVING **THERMODYNAMICS** Regime 1° latitude/longitude 0.1° latitude/longitude Horizontal resolution 1 km – 100 m > 10 m Loop current

Low Resolution Ocean Configuration (g17 grid)



Detecting Model Error Using Data Assimilation to Diagnose where the Model Physics Produce Improbable Outcomes



High Resolution Ocean Configuration (t13 grid)

PARAMETERIZED Loop current 1° latitude/longitude 0.1° latitude/longitude Horizontal resolution 1 km – 100 m Dynamical or LAMINAR / CYCLOSTROPHIC / BAROCLINIC Physical BAROTROPIC **SUBMESOSCALE** RESOLVING Regime 1° latitude/longitude 0.1° latitude/longitude Horizontal resolution 1 km – 100 m Low Resolution Ocean PARAMETERIZED PARAMETERIZED Configuration Loop current (g17 grid)

> 10 m **SMALL-SCALE THERMODYNAMICS**

PARAMETERIZED

> 10 m

PARAMETERIZED

25

Scientific Workflow Data Assimilation Cycling



















DART Overview



General-purpose codebase Interfaces with atmospheric, oceanographic, cryospheric, land surface & space weather models.



Works with any type of observation Satellites, weather balloons, undersea gliders, radar stations, GPS, etc. Each type can pose unique challenges.



Range of users Can be compiled without MPI for use on laptops; can be compiled with MPI for running on thousands of nodes

Written in FORTRAN

Implements algorithms currently used in weather forecasting and experimental techniques

Accommodates huge state vectors; one-way MPI communications

https://dart.ucar.edu

Computational Background
Experimental Setup



Model Configuration





Model Configuration	Low-resolution (g17)	High-resolution (t13)
Model Grid	Offset Greenland	Poseidon Tripole
Grid points in 1 instance	5,160,960	535,680,000
Derecho CPU hours	135,924	44,142,416
Long term archiving	3.7 TB	87.9 TB
Forcing	CAM6 Reanalysis (f09)	CAM6 Reanalysis (f09)

Team Experience
Precursor Projects

Isohaline Salinity Budget of the North Atlantic Salinity Maximum

FRANK BRYAN Climate and Global Dynamics Division, National Center for Atmospheric Research,* Boulder, Colorado

SCOTT BACHMAN

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge United Kingdon

OCTOBER 2016

(Manuscript received 21 August 2014, in final form 4 November 2014)

In this study, the salinity budget of volumes bounded by isohaline surface a high-resolution numerical simulation of hydrography and surface fluxes. W instantaneous volume-integrated salt uated from time-averaged data. In this variability of the salinity maximum ar instantaneous and time-averaged moc water mass is determined. This study fi by the mesoscale eddies is approxima climatological-mean conditions. The i rates associated with surface forcing as the surface net evaporation in the No 7 Sverdrups (Sv: $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) of the simulation, whereas the estimate

In each of the subtropical eyres of the g there exists a distinct surface salinity max

closed isohaline contours. They are an ex the coupling of the ocean and atmosphere

hydrologic cycle. The surface salinity may

cated in the vicinity of regional extrema i

These surface features connect to equate

westward-extending subsurface salinity max

core depths of 100-150 m) sometimes is

* The National Center for Atmospheric Research

Corresponding author address: Frank Bryan, NC

by the National Science Foundation.

3000, Boulder, CO 80307-3000

DOI: 10.1175/JPO-D-14-0172.1

E-mail: bryan@ucar.edu

oration but are not exactly collocated

1. Introduction

724

IOHNSON ET AL

VOLUME 45

³Climatological Annual Cycle of the Salinity Budgets of the Subtropical Maxima

2981

BENJAMIN K. JOHNSON Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

> FRANK O. BRYAN National Center for Atmospheric Research," Boulder, Colorado

SEMYON A. GRODSKY AND JAMES A. CARTON Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

(Manuscript received 19 October 2015, in final form 17 June 2016)

ABSTRACT

Six subtropical salinity maxima (Smax) exist: two each in the Pacific, Atlantic, and Indian Ocean basins. The north Indian (NI) Smax lies in the Arabian Sea while the remaining five lie in the open ocean. The annual cycl of evaporation minus precipitation (E - P) flux over the S_{max} is asymmetric about the equator. Over the Northern Hemisphere Smax, the semiannual harmonic is dominant (peaking in local summer and winter), while over the Southern Hemisphere S_{max} , the annual harmonic is dominant (peaking in local winter). Re gardless, the surface layer salinity for all six Smax reaches a maximum in local fall and minimum in local spring his study uses a multidecade integration of an eddy-resolving ocean circulation model to compute salinit budgets for each of the six Smar The NI Smar budget is dominated by eddy advection related to the evolution of the seasonal monsoon. The five open-ocean S_{max} budgets reveal a common annual cycle of vertical diffusive fluxes that peak in winter. These Smax have regions on their eastward and poleward edges in which the vertical salinity gradient is destabilizing. These destabilizing gradients, in conjunction with wintertime surface cooling generate a gradually deepening wintertime mixed layer. The vertical salinity gradient sharpens at the base of he mixed layer, making the water column susceptible to salt finger convection and enhancing vertical diffusive salinity fluxes out of the S_{max} into the ocean interior. This process is also observed in Argo float profiles and is related to the formation regions of subtropical mode waters.

© 2015 American Meteorological Society 1. Introduction

Benotes Open Access content.

by the National Science Foundation.

DOI: 10.1175/JPO-D-15-0202.1 © 2016 American Meteorological Society

Park, MD 20742 E-mail: bjohnson@atmos.umd.edu

basins, causing evaporation to exceed precipitation, elevating SLS and creating six subtropical salinity maxima The broad patterns of surface layer salinity (SLS) are (hereinafter S_{max}). Five of the S_{max} lie primarily in the set by the interaction of surface waters with the atmoopen ocean: the Pacific and Atlantic S_{max} are located sphere (Wüst 1935). The descending branch of the around 25°N and 20°S; the south Indian Smax is around Hadley cell drives anticyclonic flow over the ocean 30°S (Gordon et al. 2015). The north Indian (NI) Smax lies in the Arabian Sea and Gulf of Oman around 18°N. It is bounded by coasts to the north and west. For brevity, we will refer to each S_{max} by the first initial of its hemisphere and ocean basin, for example, NA Smax denotes the North Atlantic subtropical salinity maxima. ^a The National Center for Atmospheric Research is sponsored The Smax are bounded by different isohaline values given in Table 1. We selected these values as those that define the maximum volume that is confined to the subtropical Corresponding author address: Benjamin Johnson, Department region of each basins' corresponding hemisphere. of Atmospheric and Oceanic Science, University of Maryland, The annual cycle of evaporation minus precipitation College Park, 3424 Computer and Space Science Bldg., College (E - P) flux over the subtropical regions is asymmetric about the equator (Large and Yeager 2009). Over the

Unauthenticated | Downloaded 05/03/22 04:56 AM UTC

Yellowstone ASD t12 Runs

- Two long-term integrations of POP in its high-resolution, eddy-resolving configuration were completed using the Yellowstone ASD period and a University Large Allocation using roughly 5M core hours on Yellowstone.
- These experiments are described in Bryan and Bachman (2015) and Johnson et al. (2016).
- The output has been heavily studied and the experiments are colloquially known as the "ASD Runs" in CGD's Oceanography Section.

scientific reports

(R) Check for updates

OPEN A new CAM6 + DART reanalysis with surface forcing from CAM6 to other CESM models

Kevin Raeder^{1,4}^{SC}, Timothy J. Hoar^{1,4}, Mohamad El Gharamti^{1,4}, Benjamin K. Johnson^{1,4}, Nancy Collins^{1,4}, Jeffrey L. Anderson^{1,5}, Jeff Steward^{2,4} & Mick Coady^{3,5}

An ensemble Kalman fitter reanalysis has been archived in the Research Data Archive at the National Center for Atmospheric Research. It used a CAM6 configuration of the Community Earth System Model (CESM), several million observations per day, and the Data Assimilation Research Testbed (DAR7). The data saved from this global, ~ 1° resolution, 80 member ensemble span 2011–2019. They include ensembles of: sub-daily, real world, atmospheric forcing for use by all of the nonatmospheric models of CESM, weekly, CAM6, restart file sets; 6 hourly, prior hindcast estimates of the assimilated observations; 6 hourly, land model, plant growth variables, and 6 hourly, ensemble mean, gridded, atmospheric analyses. This data can be used for hindcast studies and data assimilation using component models of CESM; CAM6, CHMS, CICES, POP2. MOM6, MOSART, and CISM; and non-CESM Earth system models. This large dataset (- 120 Tb) has a unique combination of a large ensemble, high frequency, and multivear time span, which provides opportunities for robust statistical analysis and use as a machine learning training dataset.

"Data assimilation" ("DA") is the term used in many geophysical sciences for the merging of observations with a model state created by a (usually) numerical model of the physical system. The model state used in this dataset is a "hindcast" because it represents a past state'. The model state is referred to as "prior" to the assimilation. The result of assimilating observations into a hindcast is a "reanalysis", which is a better description of the system than either the observations or the model state individually. Observations and model hindcasts have both valuable information and errors. Successful DA keeps the information from both and reduces the errors. It also reduces the prior uncertainty^{-2,0}.

The reanalyses created by the Data Assimilation Research Testbed⁴⁵ (DART) use an 80 member ensemble of similar hindcasts, which leverages the power of statistics to give a more comprehensive estimate than a single hindcast can provide. The ensemble of reanalyses is a sample of the probability distribution of the physical system after the observations are assimilated (the "posterior")⁶.

The mean of this ensemble can be viewed as the most likely weather. The spread of the ensemble tells us how much confidence to have in it (smaller spread implies more certainty) or how much variability we should expect.

DA was developed to improve the initial conditions used in numerical weather forecasts¹, but its use is spreading into studies of other Earth system components, the oceans, land, cryosphere, etc. These components are strongly forced by the atmosphere. In order to successfully model them, the atmospheric forcing must be specified correctly, both its mean and variability. The first motivation for the creation of this dataset was to provide that forcing in the context of an Earth system modeling framework, in which the atmospheric forcing can be applied to the nonatmospheric components consistently and conveniently. This is accomplished by running an atmospheric reanalysis for as long as possible and archiving all of the fluxes and other variables as frequently as required by the nonatmospheric forcing. The nonatmospheric models can be run repeatedly without the cost of regenerating the atmospheric forcing. The nonatmospheric model runs could be single or ensemble hindcasts to study the model behavior, or they could be the hindcasts used in generating reanalyses of the nonatmospheric components (see Fig. 1 for an illustration).

¹National Center for Atmospheric Research, CISL/DAReS, Boulder, CO 80305, USA. ²Spire Global Inc., 1690 38th St., Boulder, CO 80301, USA. ³National Center for Atmospheric Research, CISL/CSG, Boulder, CO 80305, USA. ³These authors contributed equally: Kevin Raeder, Timothy J. Hoar, Mohamad El Gharamti, Benjamin K. Johnson, Nancy Collins and Jeff Steward. ³These authors jointly supervised this work: Jeffrey L. Anderson and Mick Coady. ⁵⁵email: raeder@ucar.edu

https://doi.org/10.1038/s41598-021-92927-0

Scientific Reports (2021) 11:16384

nature portfolio

CAM6 + DART Reanalysis

- Raeder et al., (2021) completed a 10year atmospheric reanalysis using DART and the atmospheric component of CESM, the Community Atmosphere Model 6
- Reanalysis spans 2011-2019 with 2020
 currently underway
- Required roughly 17M CPU hours on Cheyenne
- Revealed inefficiencies in model build scripts due to directory locking



KiloCAM Experiment

- Johnson and Gharamti are conducting a thousand-member CAM6 ensemble experiment to test DART algorithms on Shaheen II, a Cray XC40 at King Abdullah University of Science and Technology.
- Shaheen II has 6174 nodes with Intel
 Haswell processors and can
 theoretically compute 7.2 Pflop/s. The
 thousand-member experiment uses up
 to 3000 nodes at one time, nearly 50%
 of the system's nodes.

Closing Remarks
Desired Outcomes

Desired Outcomes

- Produce a hierarchical reanalysis spanning four complete years, 2012-2015, that encompasses the termination of the 2012 La Niña event, the transition of PDO from its cool to warm phase in 2014 and the onset of the subsequent 2015 El Niño event.
- Allow for enhanced understanding of mode water formation, continuing the work of Johnson et al. (2016)
- Sensitivity testing of adaptive inflation algorithms, continuing the work of Gharamti (2018)
- Studying interannual variability of the Equatorial Pacific Cold Tongue, continuing the work of Deppenmeier et al. (2021)
- Improved parameterization of mesoscale eddies, continuing the work of Grooms and Kleiber (2019).

Deppenmeier, A.-L., F. O. Bryan, W. S. Kessler, and L. Thompson, 2021: Modulation of Cross-Isothermal Velocities with ENSO in the Tropical Pacific Cold Tongue. J. Phys. Oceanogr., 51, 1559–1574, <u>https://doi.org/10.1175/JPO-D-20-0217.1</u>.

Johnson, B. K., F. O. Bryan, S. A. Grodsky, and J. A. Carton, 2016: Climatological Annual Cycle of the Salinity Budgets of the Subtropical Maxima. *J. Phys. Oceanogr.*, **46**, 2981–2994, <u>https://doi.org/10.1175/JPO-D-15-0202.1</u>.

- Gharamti, M., 2018: Enhanced Adaptive Inflation Algorithm for Ensemble Filters. *Mon. Wea. Rev.*, **146**, 623–640, <u>https://doi.org/10.1175/MWR-D-17-0187.1</u>.
- Grooms, I., and W. Kleiber, 2019: Diagnosing, modeling, and testing a multiplicative stochastic Gent-McWilliams parameterization. *Ocean Modelling*, **133**, 1–10, https://doi.org/10.1016/j.ocemod.2018.10.009.

Photo Credits

- Trade wind and loop current schematics produced by NOAA. Cropped and resized to fit within the slides.
- Nor'Easter, tornado and cloud photos produced by NOAA. Cropped and resized to fit within the slides.
- Mesoscale eddy figure produced by NASA-GSFC. Public domain.
- Langmuir circulation photo from The Eastern Bering Sea Shelf : Oceanography and Resources / edited by Donald W. Hood and John A. Calder. No known copyright issue.
- Double diffusion figure produced by Sandia National Lab. © 2007-2011 Sandia Corporation.
- Photos of Shaheen II and Campus Library © King Abdullah University of Science and Technology.
- Offset Greenland and Poseidon Tripole Grid schematics © UCAR.
- All remaining DART figures and schematics © UCAR.

Thanks For Your Attention
Any Questions?