

World Ocean Database and World Ocean Atlas 2009



*An Ocean of Data
on Your Desktop*

Hernan Garcia, Tim Boyer, Sydney Levitus, Ricardo Locarnini,
John Antonov, Olga Baranova, Dan Seidov, Igor Smolyar,
Daphne Johnson, Alexey Mishonov, and Melissa Zweng

National Oceanographic data Center, USA
DATA DISCOVERY, ACCESS AND ARCHIVE



www.nodc.noaa.gov

PICES 2009 Annual Meeting (Jeju, Korea)

OUTLINE

- 🌐 **World Ocean Database (WOD) & World Ocean Atlas (WOA) 2009.**
- 🌐 **Value of historical, integrated databases to examine basin-scale variability.**
- 🌐 **Value of data sharing**





World Ocean Database 2009

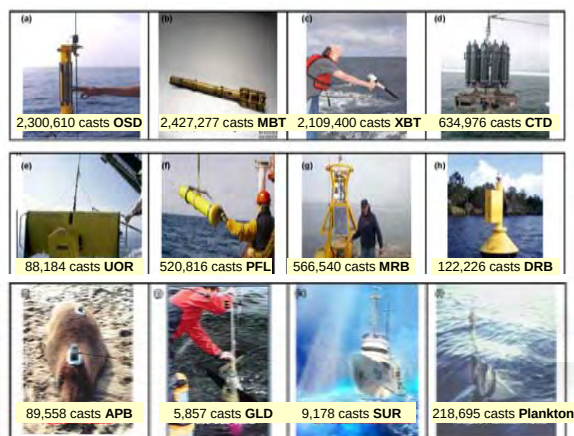
A global, comprehensive, integrated, scientifically quality-controlled oceanographic database in one well documented digital format + unrestricted access to original data & metadata (1772-2009)

Available online: Spatial, time-sorted subsets, and "user search" online (WODSelect) with **monthly updates** and **new data posted quarterly** (www.nodc.noaa.gov).

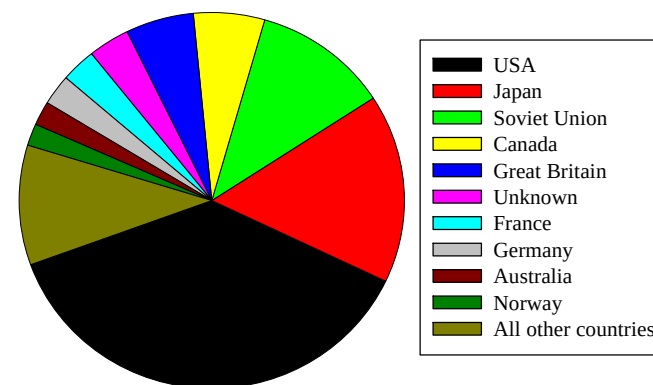


Parameter	Number of profiles (preliminary count)
Temperature	8,896,052 (1772-2009)
Salinity	3,533,325 (1874-2009)
Dissolved oxygen	848,519 (1898-2008)
Phosphate	454,933 (1922-2008)
Silicate	335,140 (1921-2008)
Nitrate	268,388 (1925-2008)
pH	196,844 (1910-2007)
Total Chlorophyll	212,825 (1933-2008)
Alkalinity	52,313 (1921-2008)
pCO ₂	3,382 (1967-2008)
Dissolved inorganic carbon	13,146 (1958-2008)
Tritium	1,618 (1984-2003)
Helium	2,116 (1984-2003)
Delta Helium-3	2,086 (1985-2003)
Delta Carbon-14	956 (1990-2003)
Delta Carbon-13	928 (1991-2003)
Argon	75 (1993-1993)
Neon	1,308 (1987-2002)
CFC-11	11,272 (1985-2008)
CFC-12	11,279 (1985-2008)
CFC-113	2,799 (1990-2006)
Delta Oxygen-18	94 (1993-1996)
Plankton	218,695 (1905-2008)
Transmissivity (BAC)	13,190 (1963-2007)

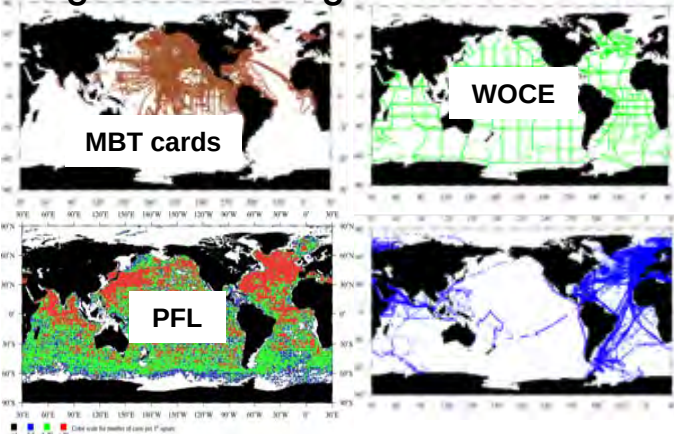
Multiple Sampling Instruments



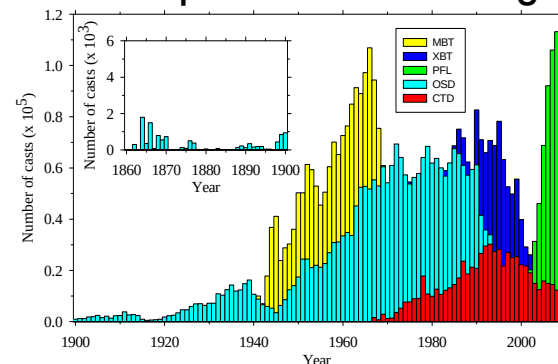
International Data Sources



Digital & Non-digital Data Sources



Temporal Data Coverage

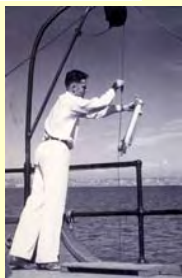


NODC Data Producers:

World Ocean database (WOD) and World Ocean Atlas (WOA) are made possible because of the data that scientists and institutions worldwide provide to national, regional, and World Data Centers. **NODC is dedicated to providing all available oceanographic data to as wide an audience as possible without restriction**

WOD also relies on voluntary data contributions from

1. IOC Member States through the IOC/IODE system and the International Council for Science (ICSU) WDC system
2. NODC bilateral agreements and international exchanges
3. Data from national and international projects and databases: GODAR, GTSP, GLOBEC, MEDAR/MEDATLAS, WOCE, JGOFS, CLIVAR, Argo, CCHDO, ICES, PISCO, and many others...

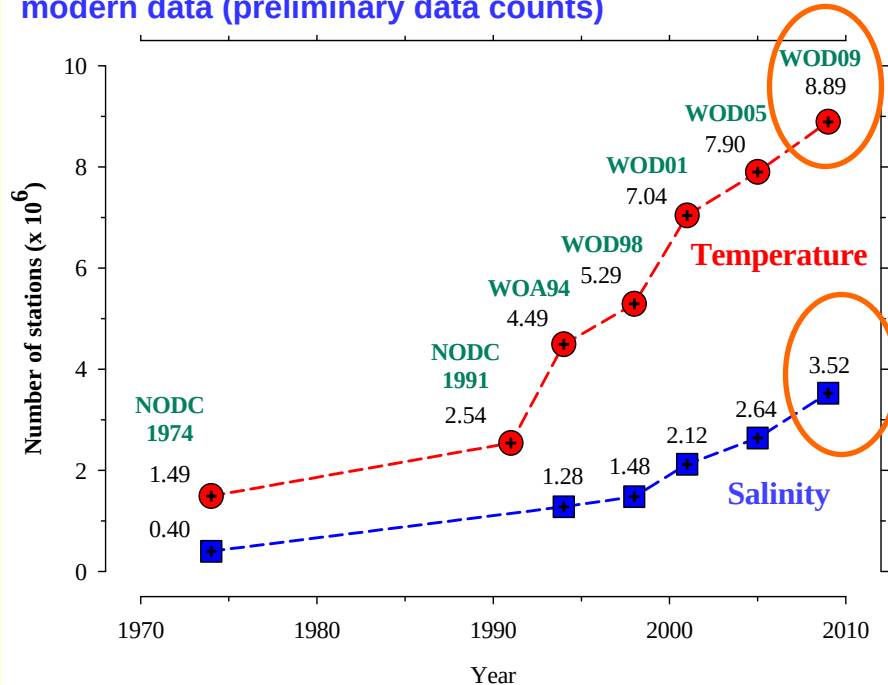


NODC, IODE, ICSU, and many others are recognized for their achievements.

How successful are they based on IMPACT and OUTPUT?

Output: Growth

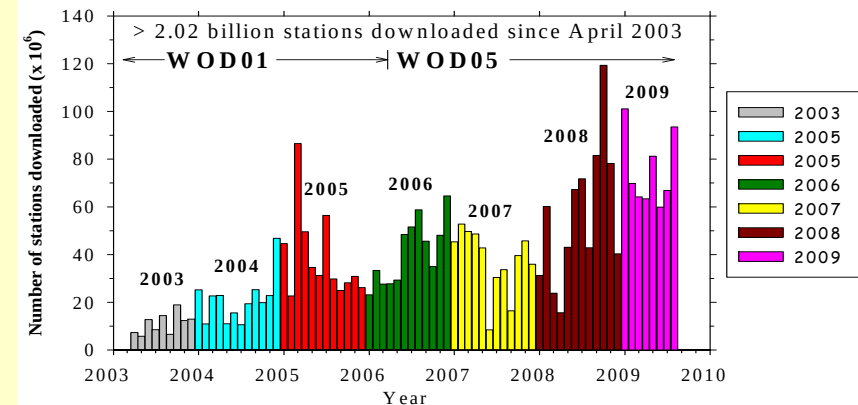
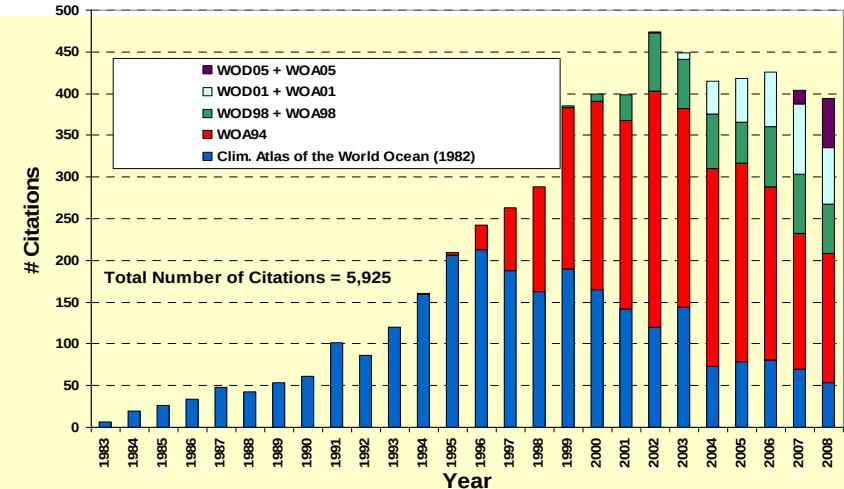
The Global Oceanographic Data Archaeology and Rescue (**GODAR**) and WOD projects have recovered and made available online without restrictions much historical and modern data (preliminary data counts)



Since April 2003; >2.02 billion stations (1,201 gigabytes) downloaded. About half the data have been served to countries other than the USA (>80% of OSD data in WOD09 are from non-USA data sources).

Impact: Science Citations and usage

WOD and objective analyses are used by the research and operational communities as evidenced by number of citations in the scientific literature and data downloads



World Ocean Database 2009 (WOD09).

	INSTRUMENT TYPE	WOD05	ADDED	WOD09	% increase
1	Bottle (OSD)	2,258,437	272,431	2,530,868	12.1
2	High Resolution Conductivity/Temperature/Depth (HCTD)	443,953	191,023	634,976	43.0
3	Mechanical Bathythermograph (MBT)	2,421,940	5,337	2,427,277	0.2
4	Expendable Bathythermograph (XBT)	1,930,413	178,987	2,109,400	9.3
5	Moored Buoys (e.g., TAO, TRITON, PIRATA)	445,371	121,169	566,540	27.2
6	Drifting Buoys	108,564	13,662	122,226	12.6
7	Profile Floats (P-ALACE, SOLO, APEX, PROVOR)	168,988	351,818	520,816	208.2
8	Undulating Ocean Recorder (e.g., Towed CTD)	46,699	41,485	88,184	88.8
9	Glider	338	5,519	5,857	1632.8
10	Autonomous Pinniped Bathythermograph	75,665	13,933	89,558	18.4

7



Other data in WOD09: Meteorological and sea-state

DATASET	SOURCE
OSD	Bottle, low-resolution Conductivity-Temperature-Depth (CTD), low-resolution XCTD data, and plankton data
CTD	High-resolution Conductivity-Temperature-Depth (CTD) data and high-resolution XCTD data
MBT	Mechanical Bathythermograph (MBT) data, DBT, micro-BT
XBT	Expendable (XBT) data
SUR	Surface only data (bucket, thermosalinograph)
APB	Autonomous Pinniped Bathythermograph - Time-Temperature-Depth recorders attached to elephant seals
MRB	Moored buoy data from TAO (Tropical Atmosphere-Ocean), PIRATA (moored array in the tropical Atlantic), MARNET, and TRITON (Japan-JAMSTEC)
PFL	Profiling float data
DRB	Drifting buoy data from surface drifting buoys with thermistor chains
UOR	Undulating Oceanographic Recorder data from a Conductivity/Temperature/Depth probe mounted on a towed undulating vehicle
GLD	Glider data

Variables	DATASETS					
	OSD	MBT	XBT	CTD	MRB	Total
Bottom depth (m)	1,451,069	617,108	444,565	352,007		2,864,749
Water color (Forel-Ule color scale)	231,968	12,411	429	1,087		245,895
Secchi disk visibility depth (m)	372,236	12,146	447	1,952		386,781
Wave direction (WMO 0877)	302,044	29,843	30,529	3,853		366,269
Wave height (WMO 1555)	164,546	114,145	47,257	9,768		335,716
Sea state (WMO 3700)	511,282	478,000	53,963	27,368		1,070,613
Wind force (Beaufort Scale)	420,311	14,444	2,199	3,948		440,902
Wave period (WMO 3155 or NODC 0378)	104,604	34,385	37,765	9,439		186,193
Wind direction (WMO 0877)	954,473	653,395	151,590	51,571	348,776	2,159,805
Wind speed (in knots)	531,432	673,101	153,368	46,986	351,301	1,756,188
Barometric pressure (millibar)	658,706	337,933	26,181	46,764	2,546	1,072,130
Dry bulb temperature (° C)	891,584	622,633	135,336	50,644	375,780	2,075,977
Wet bulb temperature (° C)	230,377	495,859	49,841	35,062		811,139
Weather condition (WMO 4501 and WMO 4677)	628,704	514,729	41,754	35,194		1,220,381
Cloud type (WMO 0500)	322,837	25,453	14,321	20,197		382,808
Cloud cover (WMO 2700)	604,530	523,929	25,751	29,515		1,183,725
Horizontal visibility (WMO 4300)	50,952	185,428	857	22,666		259,903
Reference/Sea surface temperature (° C)	24,876	1,172,477	115,942	391	1,421	1,315,107
Absolute air humidity (g m ⁻³)	56,451	1,565		82		58,098
Sea surface salinity		2,556	11,652			14,208



Metadata available in WOD09

Primary header metadata: Such as unique cast number, **ISO country code**, cruise number, date, time, position, and the number and type of variables in the cast.

Character metadata: Originator's cruise, originator's cast, & **Principal Investigator's** codes.

Secondary Metadata

Variable specific metadata

ID	DESCRIPTION	ID	DESCRIPTION
1	NODC Accession Number	35	Digitization Method (NODC 0612)
2	NODC Project Code	36	Digitization Interval (NODC 0613)
3	WOD Platform Code	37	Data Treatment and Storage Method (NODC 0614)
4	Institution Code	38	Trace Correction
5	Cast/Tow number	39	Temperature Correction
7	Originator's station number	40	Instrument for reference temperature (NODC 0615)
8	Depth Precision	41	Horizontal visibility (WMO Code 4300)
9	Ocean Weather Station	45	Absolute Humidity (g/m3)
10	Bottom Depth (meters)	46	Reference/Sea Surface Temperature
11	Cast Duration (hours)	47	Sea Surface Salinity
12	Cast Direction (down assumed)	48	Year in which probe was manufactured
13	High-resolution pairs	49	Speed of ship (knots) when probe was dropped
14	Water Color	54	Depth fix
15	Water Transparency (Secchi disk)	71	Real time
16	Wave Direction (WMO 0877 or NODC 0110)	72	XBT Wait (code no longer used)
17	Wave Height (WMO 1555 or NODC 0104)	73	XBT Frequency (code no longer used)
18	Sea State (WMO 3700 or NODC 0109)	74	Oceanographic measuring vehicle
19	Wind Force (Beaufort scale or NODC 0052)	77	xCO2 in atmosphere (ppm)
20	Wave Period (WMO 3155 or NODC 0378)	84	ARGOS fix code
21	Wind Direction (WMO 0877 or NODC 0110)	85	ARGOS time (hours) from last fix
22	Wind Speed (knots)	86	ARGOS time (hours) to next fix
23	Barometric Pressure (millibars)	87	Height (meters) of XBT launch
24	Dry Bulb Temperature (C)	88	Depth of sea surface sensor
25	Wet Bulb Temperature (C)	91	Database ID
26	Weather Conditions (WMO 4501/4677)	92	UKHO Bibliographic Reference Number
27	Cloud Type (WMO 0500 or NODC 0053)	93	Consecutive profile in a tow segment
28	Cloud Cover (WMO 2700 or NODC 0105)	94	WMO Identification Code
29	Probe Type	95	Originator's Depth unit
30	Calibration Depth	96	Originator's flags
31	Calibration Temperature	97	Water Sampler
32	Recorder (WMO 4770)	98	ARGOS ID number
33	Depth Correction	99	Time stamp (YYYYJJJ, Y=year, J= year day) to
34	Bottom Hit		ASCII version of cast was created

ID	DESCRIPTION
1	Accession number
2	Project
3	Scale
4	Institution
5	Instrument
6	Methods
8	Originator's units
10	Equilibrator type
11	Filter type and size
12	Incubation time
13	CO2 sea warming
15	Analysis temperature
16	Uncalibrated
17	Contains nitrite
18	Normal Standard
	Seawater batch
19	Adjustment

Other metadata:
Plankton & biological
metadata
(not shown)



WODSelect: database select and search

WORLD OCEAN DATABASE SELECT AND SEARCH

Note: new data added after the release of the WOD09 have not gone through the full set of quality control procedures and should be considered preliminary. The set of all casts that fall into this category are also available on the [WOD updates page](#).

The WODselect retrieval system allows a user to search World Ocean Database 2009 and new data added since its release using a user-specified search criteria. A distribution map and cast count of these search criteria will give the user the option to have the data extracted and placed on the NODC FTP site in the WOD09 native and ".csv" data format (more information about [downloading and reading the data files](#)).

Build a Data Retrieval Request Based on Your Choice of Criteria:

To build a user defined search query:

1. Place check mark in front of any number of criteria.
2. Press the "Build a query" button.

(If any criteria below are not checked, the default will apply).

SEARCH CRITERIA: (definitions)	DEFAULT:
☐ GEOGRAPHIC COORDINATES	- whole world
☐ OBSERVATION DATES - e.g., Year(s), Month(s), Day(s)	- all years/months/days
☐ DATASET - e.g., OSD, CTD, XBT	- OSD only
☐ MEASURED VARIABLES - e.g., Temperature, Salinity, Nutrients	- all available variables
☐ BIOLOGY - e.g., Phytoplankton, Zooplankton	- all available plankton
☐ DEEPEST MEASUREMENT	- all depths
☐ COUNTRY	- all countries
☐ SHIP/PLATFORM	- all ships/platforms
☐ CRUISE	- all cruises
☐ ACCESSION #	- all accessions
☐ PROJECT	- all projects
☐ INSTITUTE	- all institutes
☐ DATA EXCLUSION USING WOD QUALITY CONTROL FLAGS	- no exclusion
☐ DATA ADDITIONS	- WOD05 released data

Build a query

Reset

BACK TO BUILD a new query
GEOGRAPHIC COORDINATES: [HELP](#)

Please wait for map to load.



Northern edge 54.5000
Western edge 114.5000 Eastern edge 151.5000
Southern edge 22.0000

Redraw World Map or Get an Inventory

Please, CLICK ONLY ONCE, it may take a while before results are shown.

WODselect > Download Data

BACK TO BUILD a new query

DOWNLOAD DATA:

1. CHOOSE FORMAT

☒ WOD05 native ASCII format

- [output example](#)
- [downloading and reading instructions](#)

☐ Comma Delimited Value (CSV) format

- Ocean Data View does not support csv format
- [output example](#)
- [downloading and reading instructions](#)

☒ Data from each selected instrument in separate file

☐ Data from all selected instruments together

☒ Standard output

- ☐ Small file size output (maximum size of file less than Microsoft Excel number of rows limit)

2. CHOOSE DEPTH LEVEL

☒ Observed level data

- [definition](#)

☐ Standard level data

- [definition](#)

3. EXTRACT DATA

Enter your E-mail address

to

EXTRACT DATA

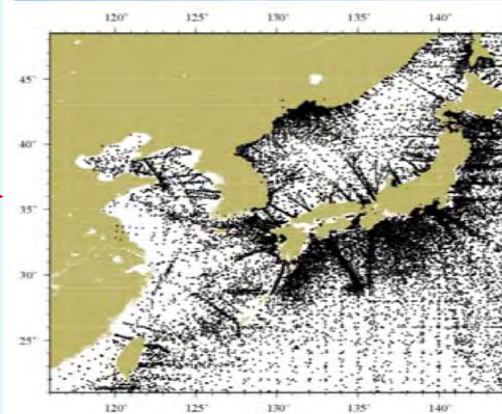
[This email address will only be used to notify you when the extraction is completed. This email will provide information on the file name(s) and instructions (and/or a link) for downloading the data from the NODC FTP site.]

If you encounter any problems, please contact: OCL_help@noaa.gov

BACK TO BUILD a new query

DOWNLOAD DATA

DATA DISTRIBUTION PLOT:



Geographic distribution of casts
Number of casts is: 55915

NOAA/NOOC Ocean Climate Laboratory
<http://www.noaa.gov/OCL/>

WOD Generalities

- 1) Database development and management is labor-intensive.
- 2) Nearly every data set we process has some type of problem that requires human intervention.
- 3) We work with data providers to provide feedback on problems we encounter. Users can provide feedback (OCL.help@noaa.gov)

Online availability of the original and processed data

- *WOD* produced every 3-4 years with full quality control.
- Corrections to data and metadata available on a monthly basis
- New data updates available quarterly (with limited QC)

WOD is a work in progress



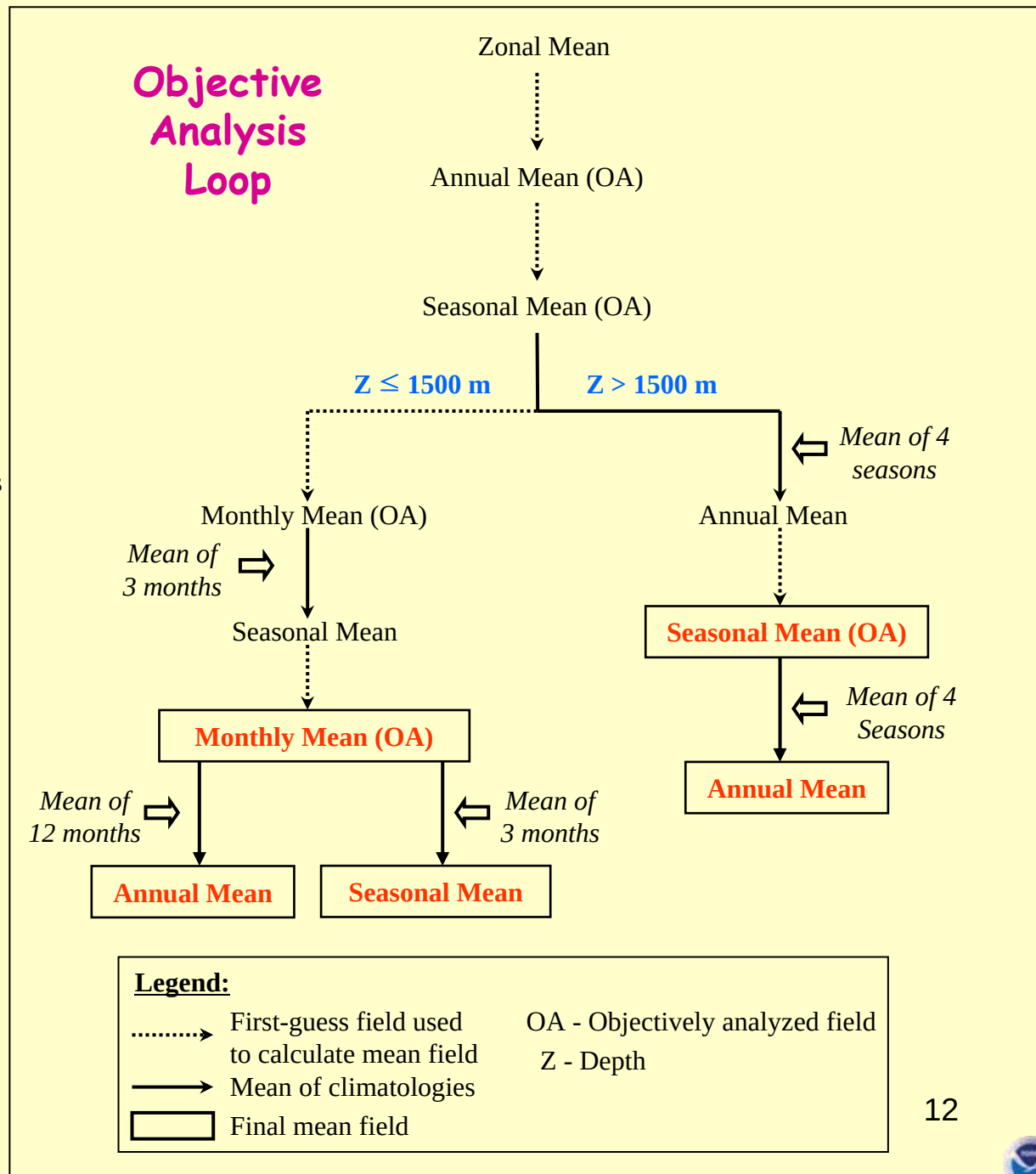
Quality Control

A. QC OF OBSERVED LEVEL DATA

1. Format conversion
2. Check position/date/time
3. Assignment of cruise and cast numbers
4. Speed check
5. Duplicate checks
6. Depth inversion and depth duplication checks
7. High-resolution pairs check
8. Range checks on observed level data
9. Excessive gradient checks
10. Observed level density checks
11. Vertical interpolation method

B. QC STANDARD LEVEL DATA

12. Density inversion & static stability check
13. Statistical analysis
14. Objective analysis
15. Subjective analysis



Definition of WOD Quality Flags (quantitative numerical definitions)

(1) FLAGS FOR ENTIRE CAST (AS A FUNCTION OF VARIABLE)	
0	accepted cast
1	failed annual standard deviation check
2	two or more density inversions (Levitus, 1982 criteria)
3	flagged cruise
4	failed seasonal standard deviation check
5	failed monthly standard deviation check
6	failed annual and seasonal standard deviation check
7	bullseye from standard level data or failed annual and monthly standard deviation check
8	failed seasonal and monthly standard deviation check
9	failed annual, seasonal and monthly standard deviation check
(2) FLAGS ON INDIVIDUAL OBSERVATIONS	
(a) Depth Flags	
0	accepted value
1	duplicates or inversions in recorded depth (same or less than previous depth)
2	density inversion
(b) Observed Level Flags	
0	accepted value
1	range outlier (outside of broad range check)
2	failed inversion check
3	failed gradient check
4	observed level "bullseye" flag and zero gradient check
5	combined gradient and inversion checks
6	failed range and inversion checks
7	failed range and gradient checks
8	failed range and questionable data checks
9	failed range and combined gradient and inversion checks

(c) Standard Level Flags	
0	accepted value
1	bullseye marker
2	density inversion
3	failed annual standard deviation check
4	failed seasonal standard deviation check
5	failed monthly standard deviation check
6	failed annual and seasonal standard deviation check
7	failed annual and monthly standard deviation check
8	failed seasonal and monthly standard deviation check
9	failed annual, seasonal and monthly standard deviation check
(d) Biological data flags (applied only to Comparable Biological Value - CBV Taxa code 27)	
0	accepted value
1	range outlier (outside of broad range check)
2	questionable value ("bullseye flag")
3	group was not reviewed
4	failed annual standard deviation check

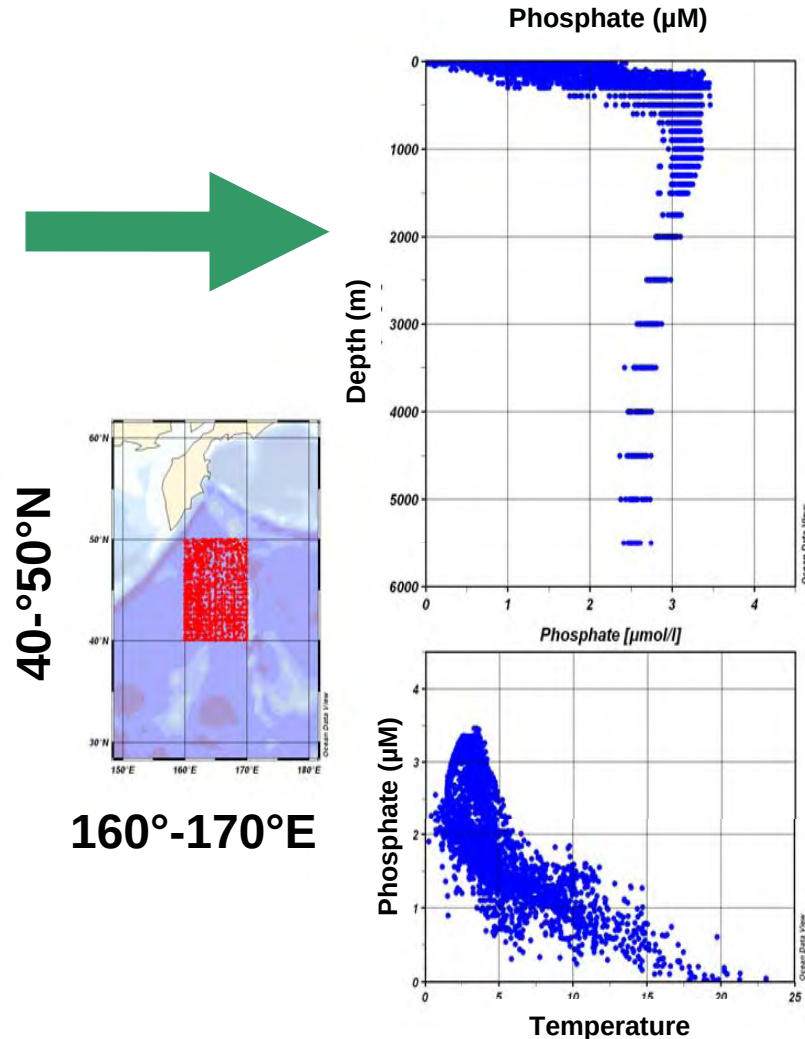
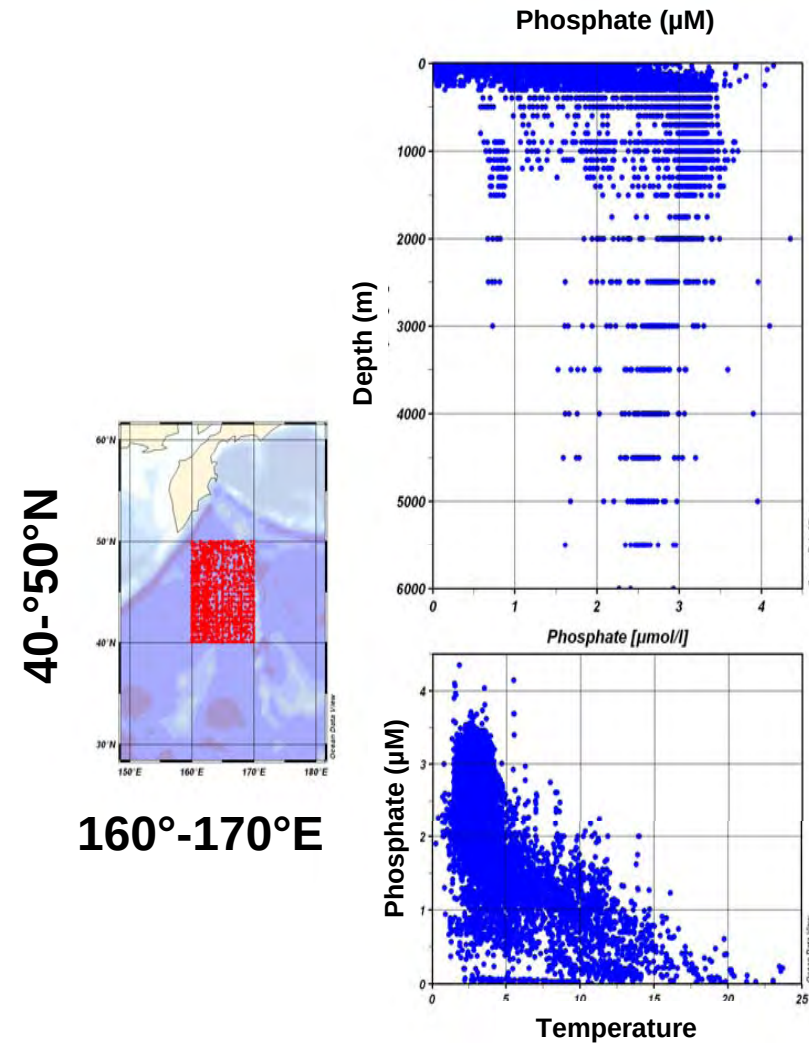
1. User can use or reject WOD QC flags.
2. Mapping of WOD QC flags to ODV
3. Preserves numerical originator's QC flags



Objective and subjective QC: Phosphate example

Before QC

After QC



~17% of the phosphate profiles are marked as “questionable”



Impact of Quality Control

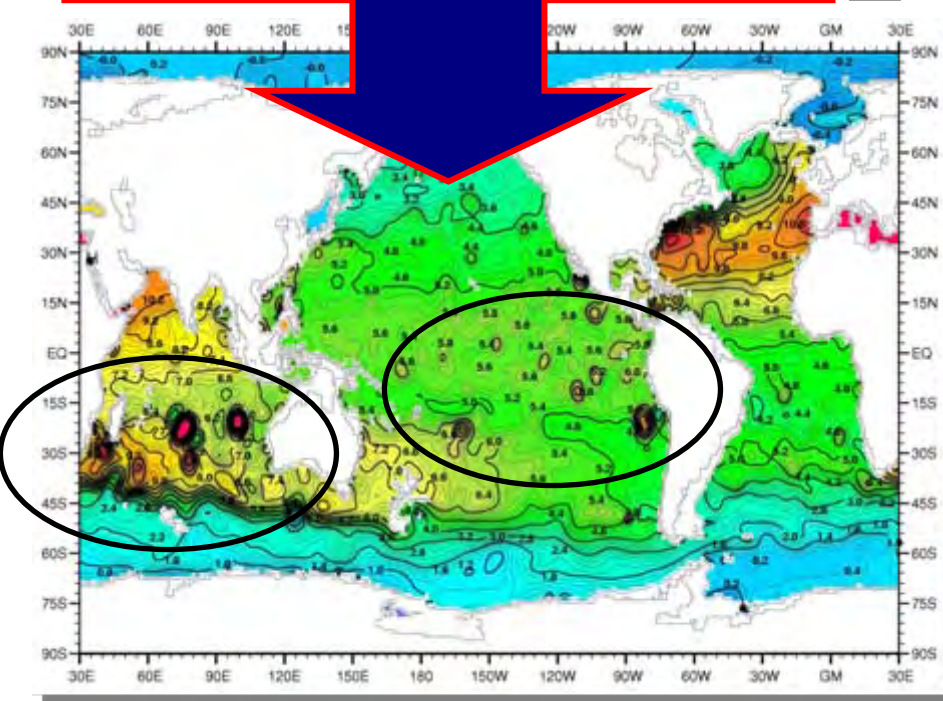
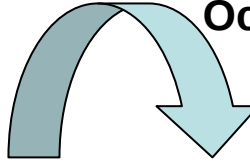
January, 800 m depth temperature climatology

Some automatic checks performed:

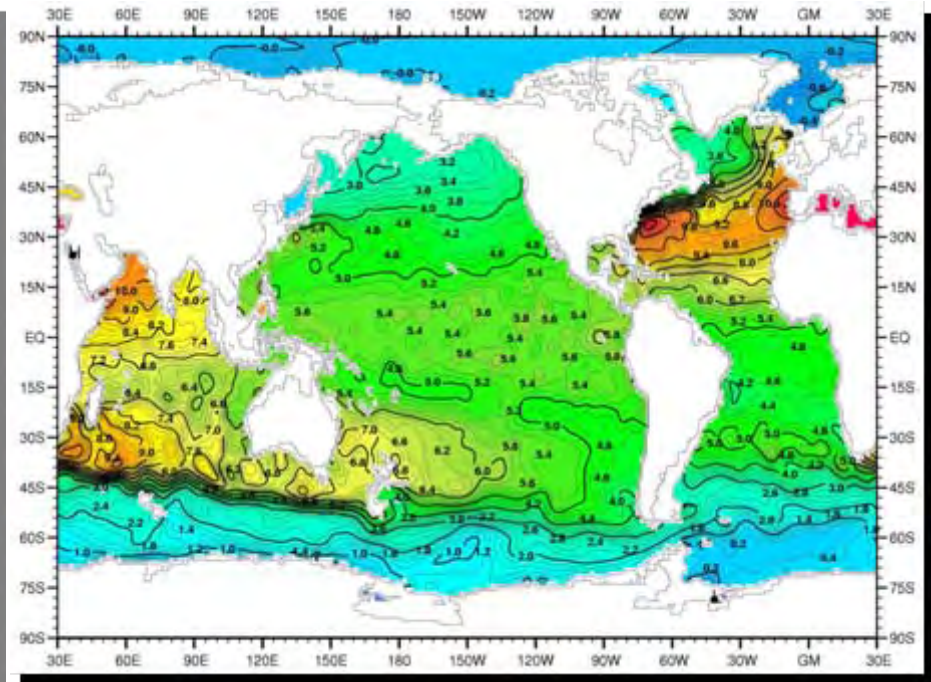
- Standard deviation
- Range
- Vertical gradient
- Spike
- Density inversion
- Cruise speed

Producing WOD series is labor-intensive.

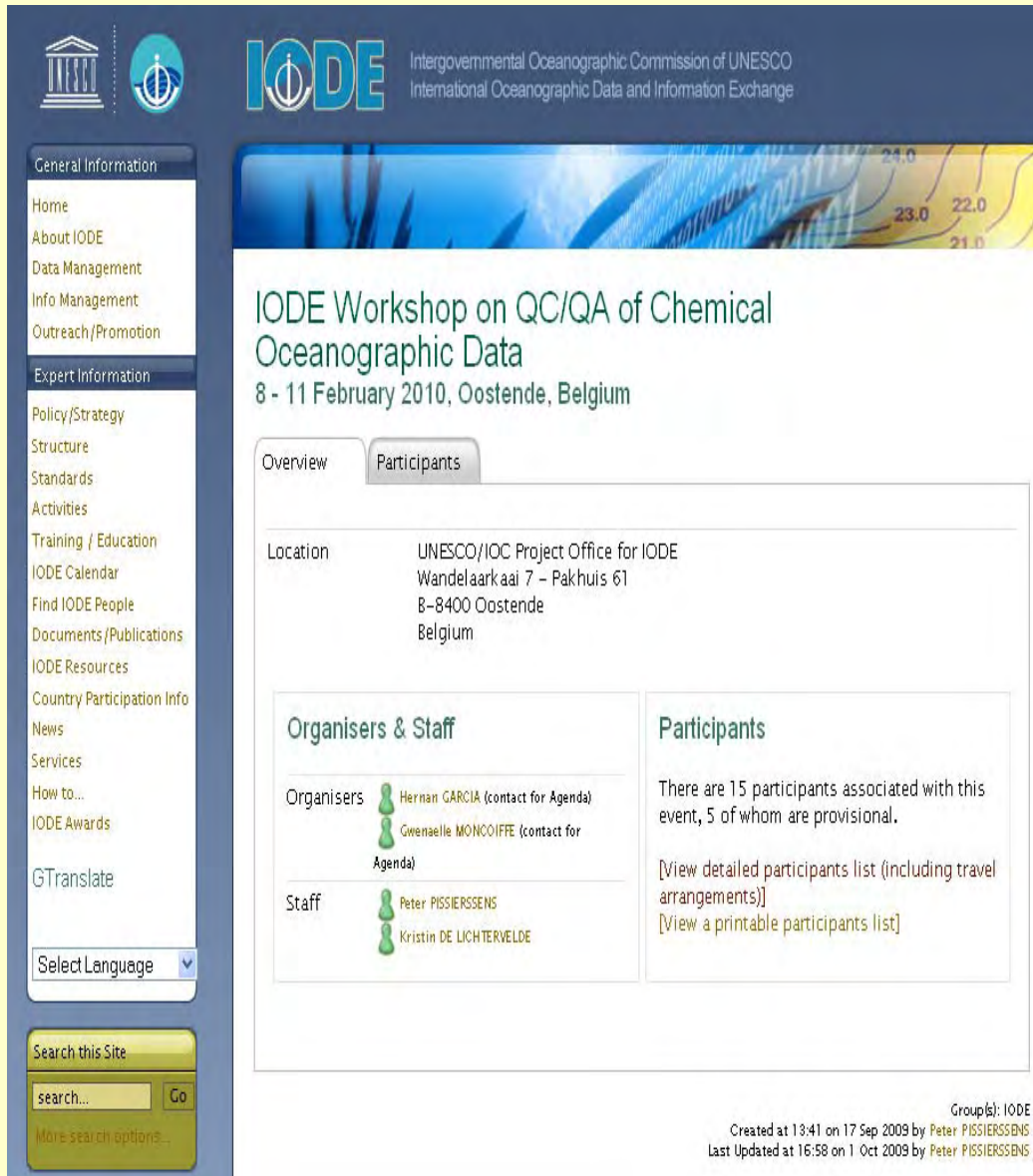
Ocean data are sparse.



After “automatic” QC



After “automatic and subjective” QC



The screenshot shows the IODE website interface. At the top, the IODE logo and full name are displayed. A left sidebar contains navigation links for General Information, Expert Information, and a search bar. The main content area features the workshop title, dates, and location. Below this, there are tabs for Overview and Participants. The Participants tab is active, showing a list of 15 participants, with 5 being provisional. The page also includes a footer with creation and update dates.

General Information

- Home
- About IODE
- Data Management
- Info Management
- Outreach/Promotion

Expert Information

- Policy/Strategy
- Structure
- Standards
- Activities
- Training / Education
- IODE Calendar
- Find IODE People
- Documents / Publications
- IODE Resources
- Country Participation Info
- News
- Services
- How to...
- IODE Awards

IODE Workshop on QC/QA of Chemical Oceanographic Data
8 - 11 February 2010, Oostende, Belgium

Overview **Participants**

Location
UNESCO/IOC Project Office for IODE
Wandelaarkaal 7 - Pakhuis 61
B-8400 Oostende
Belgium

Organisers & Staff

Organisers
Hernan GARCIA (contact for Agenda)
Gwenaëlle MONCOIFFE (contact for Agenda)

Staff
Peter PISSIERSSSENS
Kristin DE LICHTERVELDE

Participants
There are 15 participants associated with this event, 5 of whom are provisional.
[View detailed participants list (including travel arrangements)]
[View a printable participants list]

Group(s): IODE
Created at 13:41 on 17 Sep 2009 by Peter PISSIERSSSENS
Last Updated at 16:58 on 1 Oct 2009 by Peter PISSIERSSSENS

IODE Group of Experts on Biological and Chemical Data Management and Exchange Practices (GE-BICH)

1st Workshop on Quality Control and Quality Assurance of Chemical Oceanographic Data Collections –

Oostende, Belgium, 8-11 February 2010 (sponsored by IOC)

Convenors: Hernan Garcia (NODC, Hernan.Garcia@noaa.gov) & Gwenaëlle Moncoiffe (BODC, gmon@bodc.ac.uk)

Goal: Explore a minimum set of QC technical procedures and QC flags for chemical data starting with *in situ* dissolved inorganic nutrients and dissolved oxygen.



Data Products based on WOA/WOD05 (09 ...)

NOAA Atlas NESDIS 60

WORLD OCEAN DATABASE 2005

Timothy P. Boyer
John I. Antonov
Hernan E. Garcia
Daphne R. Johnson
Ricardo A. Locarnini
Alexey V. Mishonov
Michelle T. Pitcher
Olga K. Baranova
Igor V. Smolyar

Editor: Sydney Levitus

National Oceanographic Data Center
Ocean Climate Laboratory

Silver Spring, MD
July 2006

U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary

National Oceanic and Atmospheric Administration
Vice Admiral Conrad C. Lautenbacher, Jr., USN (Ret.),
Under Secretary of Commerce for Oceans and Atmosphere

National Environmental Satellite, Data, and Information Service
Gregory W. Withee, Assistant Administrator



National Oceanographic Data Center Internal Report 18

WORLD OCEAN DATABASE 2005 Documentation

Daphne R. Johnson, Tim P. Boyer, Hernan E. Garcia, Ricardo A. Locarnini,
Alexey V. Mishonov, Michelle T. Pitcher, Olga K. Baranova,
John I. Antonov, and Igor V. Smolyar

Ocean Climate Laboratory
National Oceanographic Data Center

Silver Spring, Maryland
October 3, 2005



U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary

National Oceanic and Atmospheric Administration
Conrad C. Lautenbacher, Jr.,
Vice Admiral, USN (Ret.)
Under Secretary of Commerce for Oceans and Atmosphere

National Environmental Satellite, Data and Information Service
Gregory W. Withee, Assistant Administrator

NOAA Atlas NESDIS 63

WORLD OCEAN ATLAS 2005 Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation

Hernan E. Garcia, Ricardo A. Locarnini, Timothy P. Boyer,
and John I. Antonov

Ex

Oce
National

Sil



ARTMENT OF
Gutierrez, Seci

Oceanic and At
Lautenbacher, Ji
ral, USN (Ret.)
etary of Comm

Environmental Sa
W. Withee, Assist

NOAA Atlas NESDIS 64

WORLD OCEAN ATLAS 2005 Volume 4: Nutrients (phosphate, nitrate, silicate)

Hernan E. Garcia, Ricardo A. Locarnini, Timothy P. Boyer,
and John I. Antonov

Editor: Sydney Levitus

Ocean Climate Laboratory
National Oceanographic Data Center

Silver Spring, Maryland
September 2006



U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary

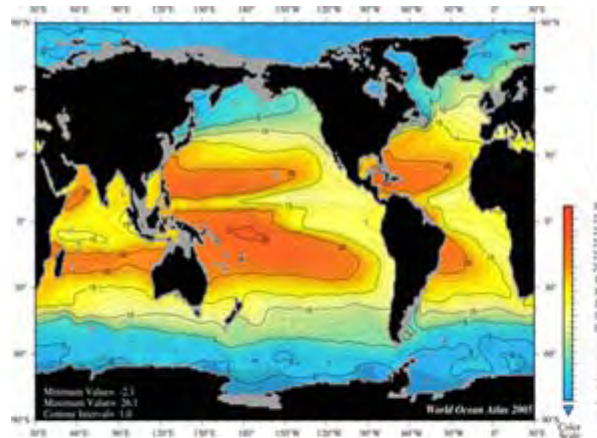
National Oceanic and Atmospheric Administration
Conrad C. Lautenbacher, Jr.,
Vice Admiral, USN (Ret.)
Under Secretary of Commerce for Oceans and Atmosphere

National Environmental Satellite, Data and Information Service
Gregory W. Withee, Assistant Administrator

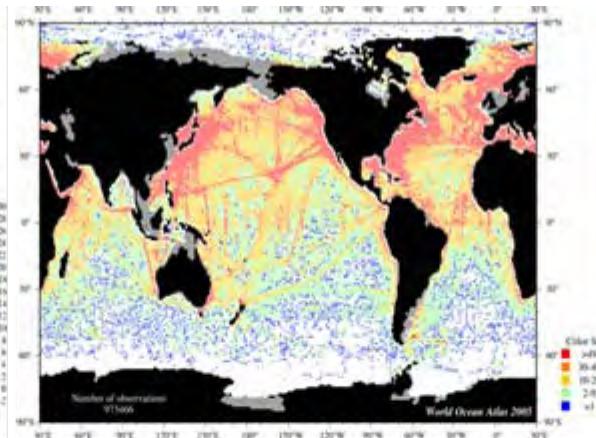


World Ocean Atlas Series

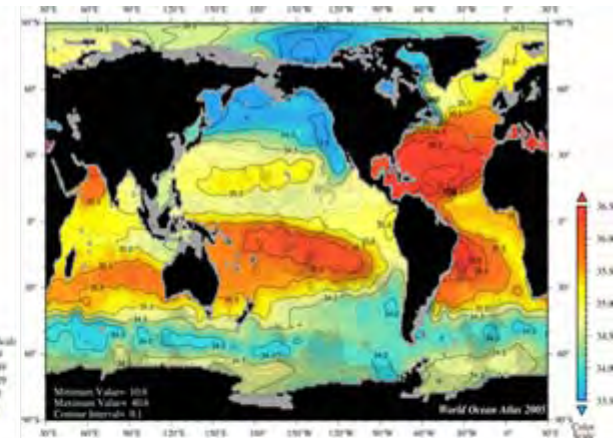
(4a) Climatological summer (Jul-Sep) temperature [°C] at 150 m depth.



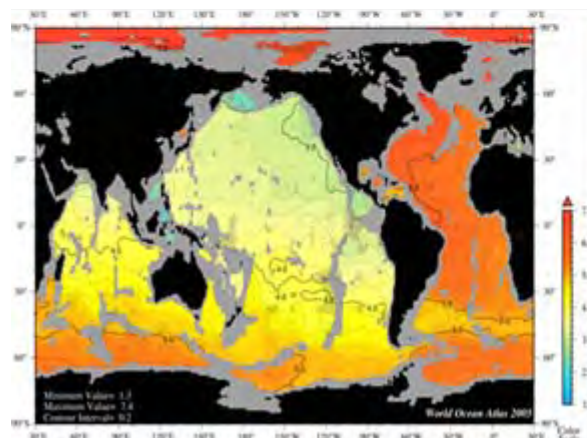
(4b) Summer (Jul-Sep) temperature data distribution at 150 m depth.



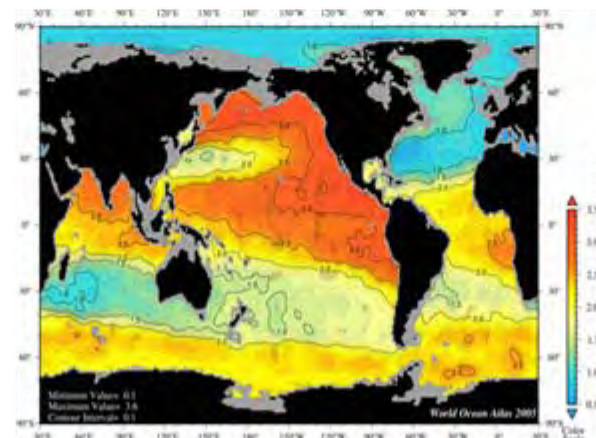
(4c) Climatological summer (Jul-Sep) salinity [PSS] at 150 m depth.



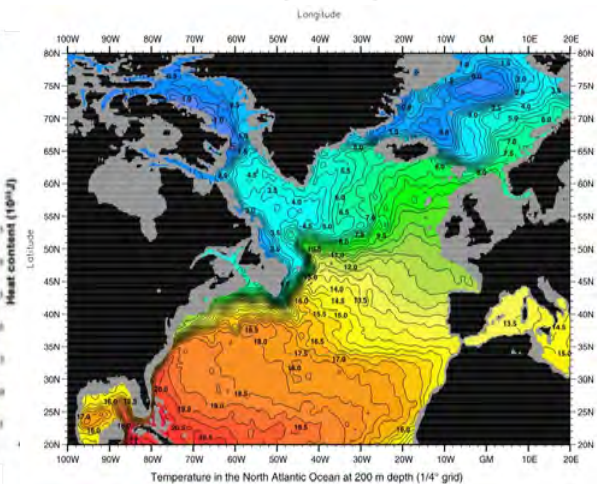
(4d) Climatological annual oxygen [ml/l] at 3500 m depth.



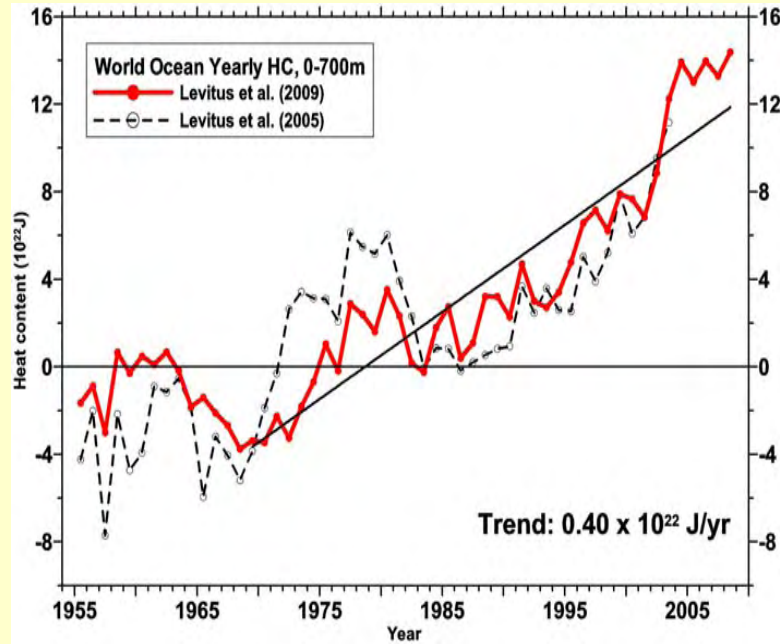
(4e) Climatological annual phosphate [μmol/l] at 600 m depth.



¼° Temperature & Salinity



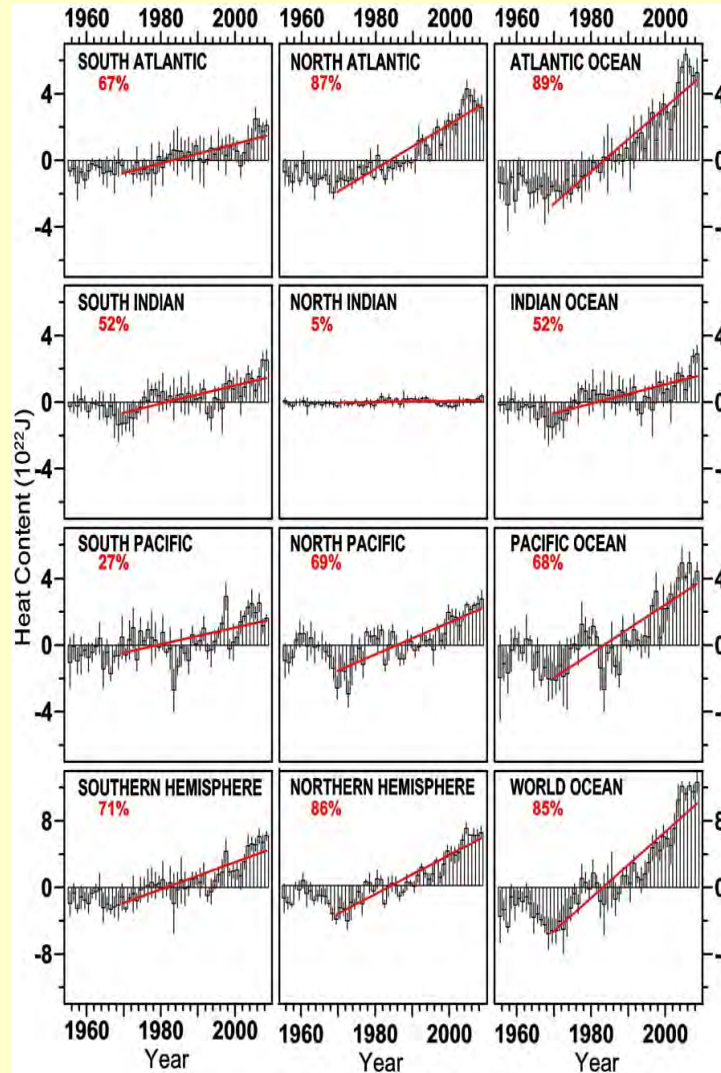
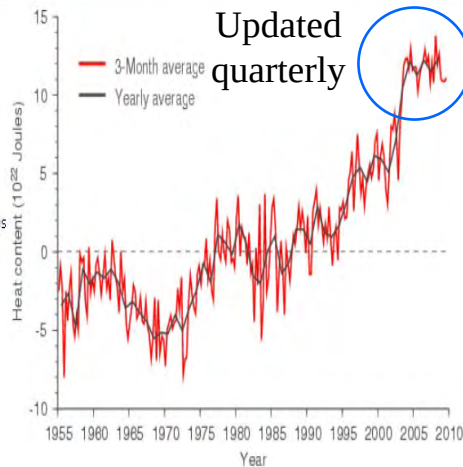
Global ocean heat content (1955-2008)



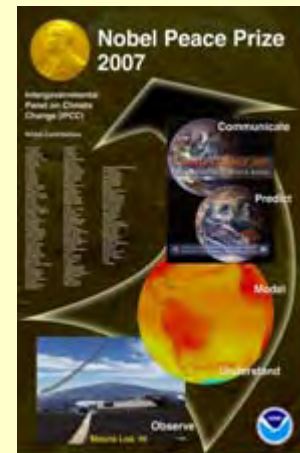
Global Ocean Heat Content [Subscribe to RSS Feed](#)

Data distribution figures, temperature anomaly fields, and heat content fields associated with Global ocean heat content (1955-2008) in light of recent instrumentation problems published in Geophysical Research Letters. See the manuscript below for details.

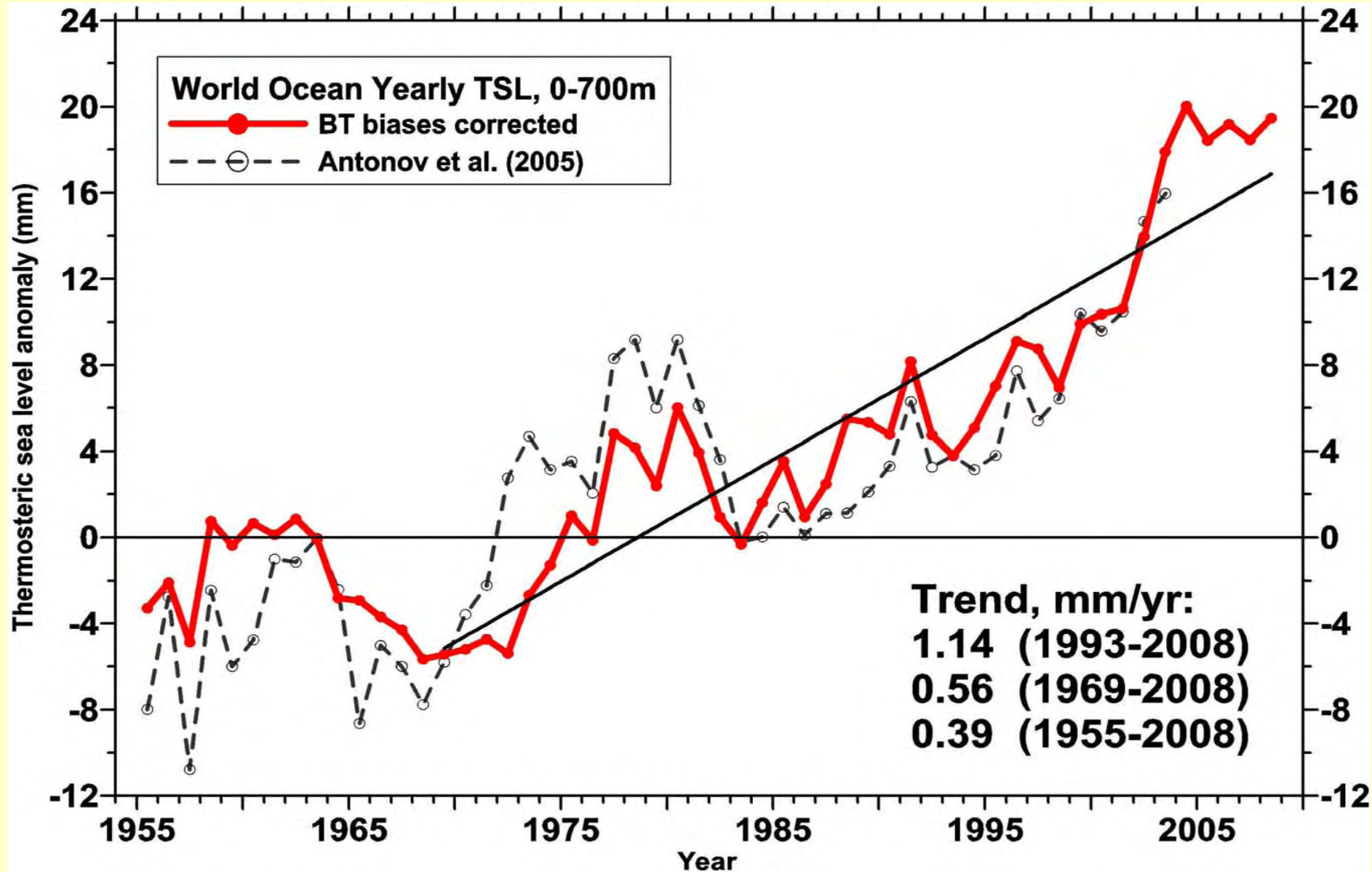
- [Published Paper \(pdf - 1.40 MB\)](#)
- [Temperature data distribution figures](#)
- [Temperature anomaly figures](#)
- [Heat content figures](#)
- [Climatological temperature fields ASCII files](#)
- [Temperature anomaly fields ASCII files](#)
- Heat content fields ASCII files: [global analyzed fields](#) [basin time series](#)



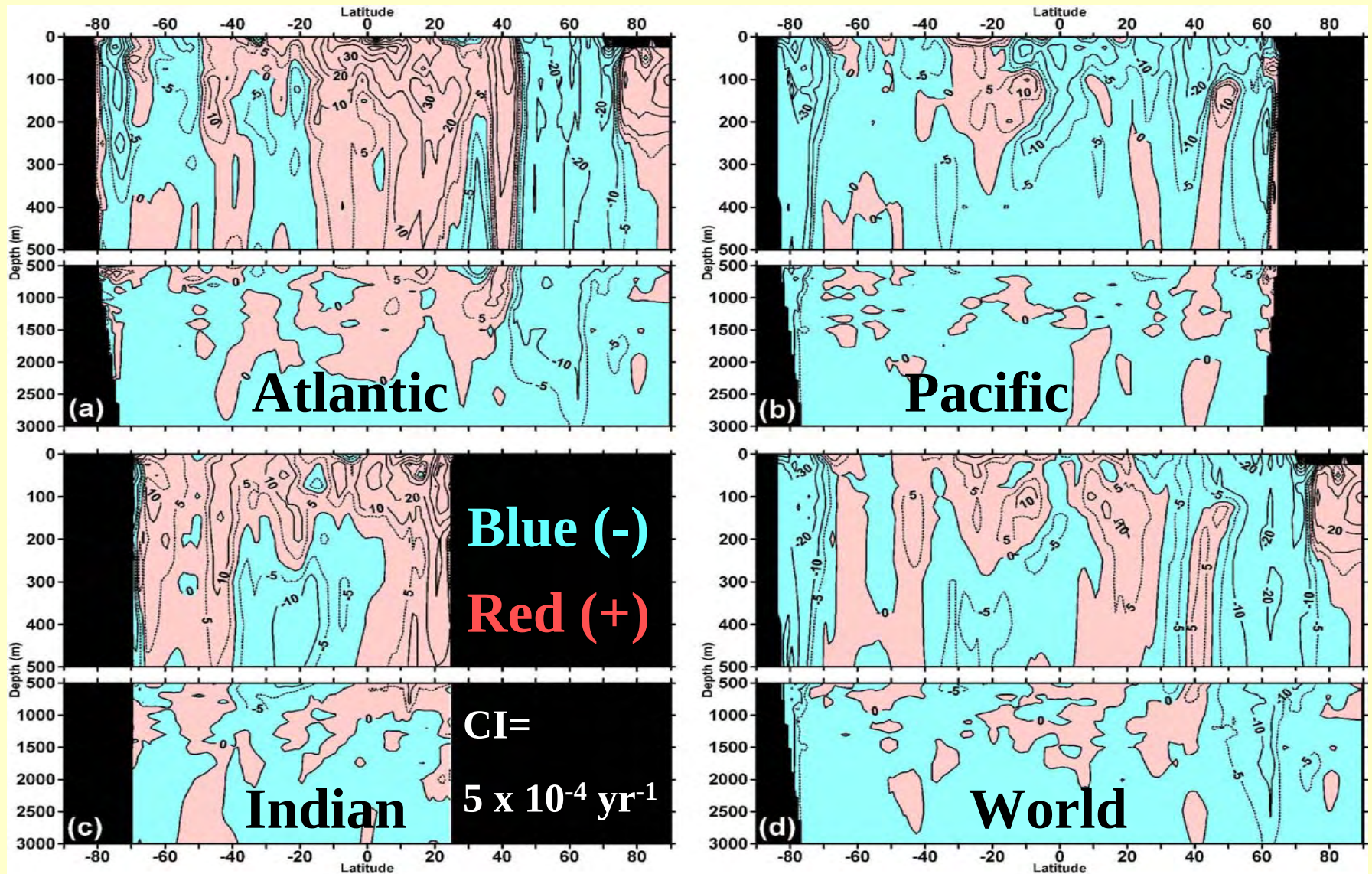
IPCC 2007



Thermosteric Sea Level Variability (1955-2008)

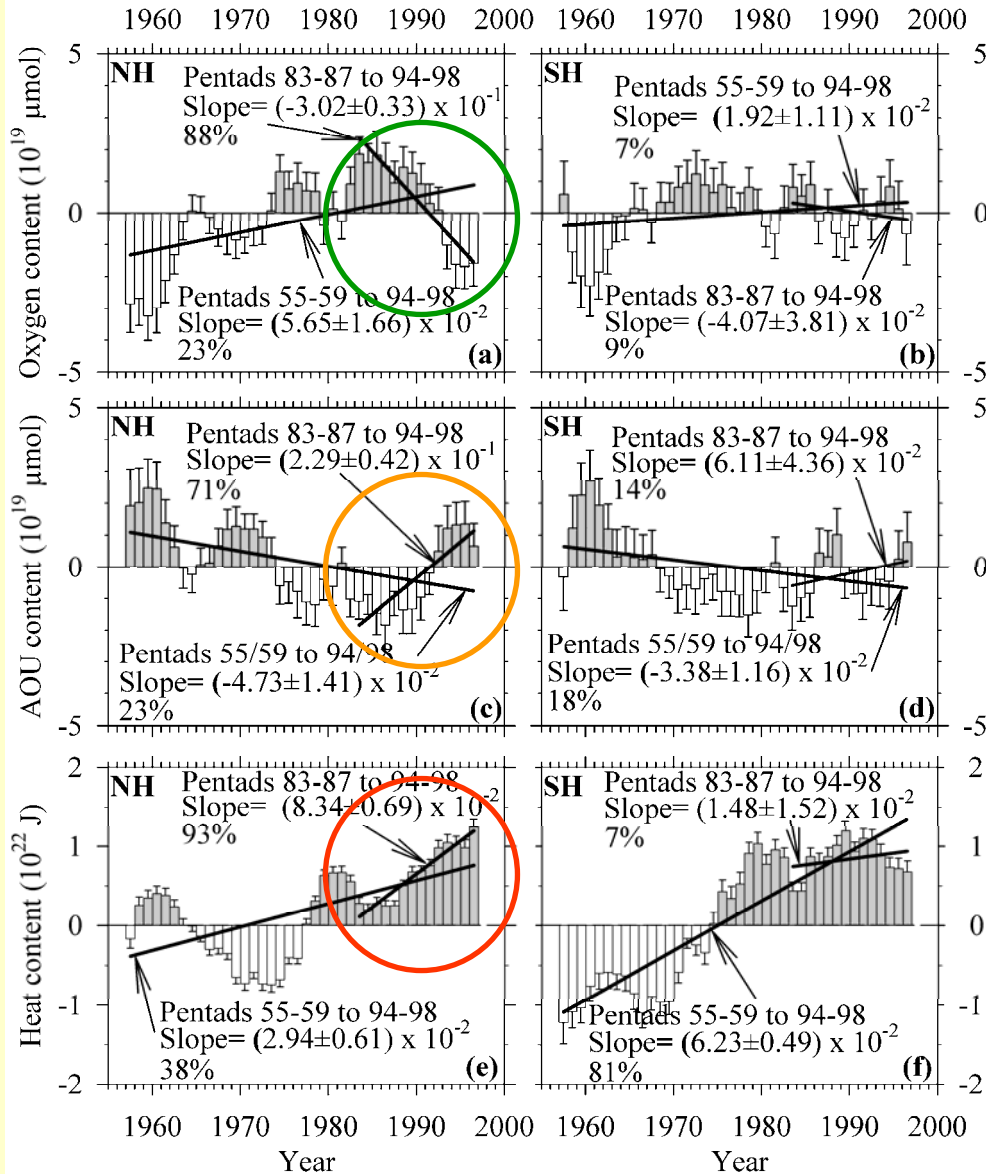


Linear Trend in Salinity ($\times 10^{-4} \text{ year}^{-1}$) of the zonally averaged pentadal salinity anomaly (1955/59 to 1994/98)

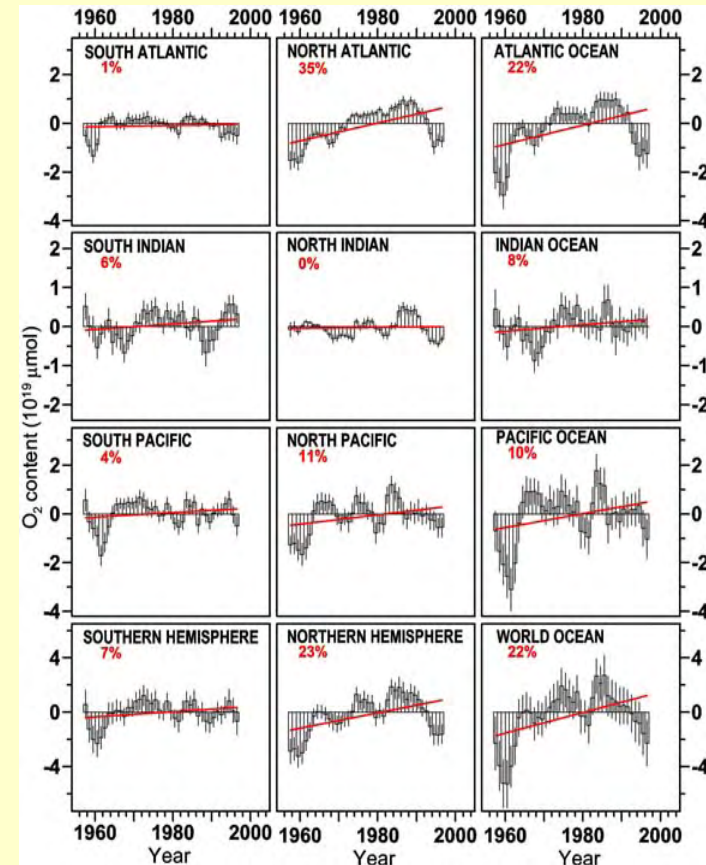


Variability in O₂, AOU, and heat content (0-100 m)

O₂



83-98 NH: -30 Tmol O₂ /decade
 ~ -4 nmol O₂ per J of heat



Garcia et al., 2005

Data sharing and data integration are required to monitor the oceans

1. Make it easier and routine for data providers to submit data to data centers (partnerships).
2. Peer reviewed "data publication" recognition equal to a peer-reviewed journal publication (work credit).
3. Associate research funding with data sharing before and after publication.
4. Increase funding (1) to manage observations and (2) to make the data available in digital usable form.

The Importance of Long-term Preservation and Accessibility of Geophysical Data ([Statement adopted by AGU Council 29 May 1997, Reaffirmed 2001, 2005, 2006, Revised and Reaffirmed May 2009](#)).



PICES Technical Committee on Data Exchange (TCODE)

Editorial (nature), *Nature* 461, 145 (10 | doi:10.1038/461145a; 23 September 2009

nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2009

Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.

More and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community. Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries — especially when data are combined from many laboratories and analysed in ways that the original researchers could not have anticipated.

All but a handful of disciplines still lack the technical, institutional and cultural frameworks required to support such open data access (see pages 168 and 171) — leading to a scandalous shortfall in the sharing of data by researchers (see page 160). This deficiency urgently needs to be addressed by funders, universities and the researchers themselves.

Research funding agencies need to recognize that preservation of and access to digital data are central to their mission, and need to be supported accordingly. Organizations in the United Kingdom, for instance, have made a good start. The Joint Information Systems Committee, established by the seven UK research councils in 1993, has made data-sharing a priority, and has helped to establish a Digital Curation Centre, headquartered at the University of Edinburgh, to be a national focus for research and development into data issues. Other European agencies have also pursued initiatives.

The United States, by contrast, is playing catch-up. Since 2005, a 29-member Interagency Working Group on Digital Data has been trying to get US funding agencies to develop plans for how they will support data archiving — and just as importantly, to develop policies on what data should and should not be preserved, and what exceptions should be made for reasons such as patient privacy. Some agencies have taken the lead in doing so; many more are hanging back. They should all be finding moving forwards vigorously.

What is more, funding agencies and researchers alike must ensure that they support not only the hardware needed to store the data, but

also the software that will help investigators to do this. One important facet is metadata management software: tools that streamline the tedious process of annotating data with a description of what the bits mean, which instrument collected them, which algorithms have been used to process them and so on — information that is essential if other scientists are to reuse the data effectively.

Also necessary, especially in an era when data can be mixed and combined in unanticipated ways, is software that can keep track of which pieces of data came from whom. Such systems are essential if tenure and promotion committees are ever to give credit — as they should — to candidates' track record of data contribution.

Who should host these data? Agencies and the research community together need to create the digital equivalent of libraries: institutions that can take responsibility for preserving digital data and making them accessible over the long term. The university research libraries themselves are obvious candidates to assume this role. But whoever takes it on, data preservation will require robust, long-term funding. One potentially helpful initiative is the US National Science Foundation's DataNet programme, in which researchers are exploring financial mechanisms such as subscription services and membership fees.

Finally, universities and individual disciplines need to undertake a vigorous programme of education and outreach about data. Consider, for example, that most university science students get a reasonably good grounding in statistics. But their studies rarely include anything about information management — a discipline that encompasses the entire life cycle of data, from how they are acquired and stored to how they are organized, retrieved and maintained over time. That needs to change: data management should be woven into every course in science, as one of the foundations of knowledge.

"Data management should be woven into every course in science."

A step too far?

The Obama administration must fund human space flight adequately, or stop speaking of 'exploration'.

After the space shuttle *Columbia* burned up during re-entry into Earth's atmosphere in 2003, the board that was convened to investigate the disaster looked beyond its technical causes to NASA's organizational malaise. For decades, the board pointed out, the shuttle programme had been trying to do too much with too little money. NASA desperately needed a clearer vision and a better-defined mission for human space flight.

The next year, then-President George W. Bush attempted to supply that vision with a new long-term goal: first send astronauts to build

a base on the Moon, then send them to Mars. This idea immediately set off a debate that is still continuing, in which sceptics ask whether there is any point in returning to the Moon nearly half a century after the first landings. Why not go to Mars directly, or visit near-Earth asteroids, or send people to service telescopes in the deep space beyond Earth?

Yet that debate is both counter-productive — a new set of rockets could go to all of these places — and moot, because Bush's vision never attracted the hoped-for budget increases. Indeed, a blue-ribbon commission reporting to US President Barack Obama this week (see page 153) finds the organizational malaise unchanged: NASA is still doing too much with too little. Without more money, the agency won't be sending people anywhere beyond the International Space Station, which resides in low Earth orbit only 350 kilometres up. And even the ability to do that is in question: Ares I, the US rocket that would return

© 2009 Macmillan Publishers Limited. All rights reserved.

145

The vision of the Global Earth Observation System of Systems (GEOSS) is "to realize a future wherein decisions and actions for the benefit of humankind are informed via coordinated, comprehensive and sustained Earth observations and information." ([GEO Data Sharing Principles Implementation](#))