



Canadian Foundation for Climate  
and Atmospheric Sciences (CFCAS)

Fondation canadienne pour les sciences  
du climat et de l'atmosphère (FCSCA)

**GOAPP**

Global Ocean-Atmosphere  
Prediction and Predictability

# **GOAPP Data Management Update**

Fred Woslyng

GOAPP

Department of Oceanography

Dalhousie University

# Contents

- ◆ Introduction
- ◆ CFCAS Data Manager Workshop
- ◆ Legacy Data
- ◆ Theme I Server
- ◆ Serving Data - OPeNDAP
- ◆ Visualization - OGC, Godiva2 & OceanDIVA
- ◆ File Types & Metadata
- ◆ Data Uploading
- ◆ Conclusion

# Introduction

I would like to present recent and ongoing work of the Data Management Committee, as it pertains to the GOAPP Network legacy.

# CFCAS Data Manager Workshop

CFCAS organized a DM workshop, held in November 2009, in Winnipeg, to identify and discuss issues and courses of action, related to data management.

In attendance were representatives of CFCAS, GOAPP, CANDAC, STAR and DRI, and via teleconference, IP3 / WCN2, C-SPARC, CPP and CAFC.

# CFCAS Data Manager Workshop

Discussions revealed many networks face critical issues including the fate of legacy data, data validation and metadata standards.

One result of the workshop was the creation of a group in Google Groups where network data managers share and discuss information and ideas.

# Legacy Data

- ◆ Early in the life of GOAPP, it was decided that Theme II data would be hosted and served by CCCma. The fate of Theme I data was less clear.
- ◆ Since early 2009, the GOAPP DMC has been in discussions with its government partners regarding archiving and serving Theme I data.
- ◆ In late 2009, informal contacts with Fraser Davidson (DFO / C-NOOFS) established a number of common objectives and similar technical approaches.

# Legacy Data

- ◆ A February, 2010 teleconference between GOAPP DMC, DFO ISDM and DFO C-NOOFS representatives, produced positive feedback.
- ◆ Currently, ISDM is accessing the proposed hardware and software configurations.
- ◆ The process is being facilitated by Fraser Davidson and Mathieu Ouellet (DFO ISDM).
- ◆ It is expected that the Theme I archive will be relocated to a DFO data centre for long term data serving.

# Theme I Server Hardware

- ◆ Supermicro 933T-R760B, 3U server chassis
- ◆ Supermicro X8DAL-I motherboard
- ◆ 2 x quad core HT Intel Xeon CPUs, 2.40 GHz
- ◆ 8 GB DDR3 ECC memory
- ◆ 15 x 2 TB Seagate Barracuda SATA II drives
- ◆ 3Ware 9650SE-16ML RAID controller
- ◆ 2 x 1 Gb/s ethernet ports
- ◆ 2 + 1 redundant power supply, 760W

Storage: 23 TB, RAID 6 (2 parity drives)







# Serving Data - OPeNDAP

- ◆ Open source Project for a Data Access Protocol.
- ◆ An web application providing remote access to scientific datasets and metadata, via the Internet.
- ◆ Supports interactive and batch processing.
- ◆ Supports data subsetting and constraints.
- ◆ Supports easy metadata augmentation using simple text files.

# OPeNDAP Architecture

- ◆ A client-server multi-tiered architecture.
- ◆ A wide range of possible clients include: web browsers, spreadsheet applications, Ferret, GrADS, ncdump, ncview and even LAS.
- ◆ An OPeNDAP client exists for Matlab.
- ◆ The latest server, Hyrax (server 4), replaces the Perl and CGI based server 3.

# Hyrax Server Components

OLFS (OPeNDAP Lightweight Front end Servlet)

- ◆ Written in Java.
- ◆ Executes within the Apache Tomcat Servlet / JSP Container.

BES (Back End Server)

- ◆ Written in C++.
- ◆ Utilizes loadable data handlers for serving netCDF, HDF4, HDF5, DBF, ASCII and binary formats.

# Visualization - OGC Standards

- ◆ The Open Geospatial Consortium (OGC) develops publicly available interface standards.
- ◆ WMS (Web Map Service) interface provides images of geo-registered data, via HTTP.
- ◆ WFS (Web Feature Service) interface provides geographical features.
- ◆ KML is an XML language focused on geographic visualization including annotation of maps and images.

## ncWMS / Godiva2

- ◆ Provides interactive visualization of data.
- ◆ Implements OGC WMS protocol.
- ◆ Used in conjunction with Godiva2 web client.
- ◆ Java server application for Apache Tomcat.
- ◆ Relies on Java netCDF interface from Unidata.
- ◆ Supports OPeNDAP server data and netCDF (CF-compliant) data.

Auto-zoom on select

[-] ESSC Web Map Service for environmental data

- NCOF Products
- EU-MERSEA
- EU-ECOOP
- Ocean Hindcasts
- Observations
  - Ocean
    - GODAE SST analyses
    - OSTIA GMPE
    - OSTIA Anomalies
    - OSTIA SST
      - sea\_surface\_temperature
      - estimated error standard deviation of analysed\_sst
    - CNR: GOS High Res SST foundation Med Sea NRT
    - OSTIA SST analysis (up to 2006-11-19)
    - OSTIA SST analysis
    - Ssalto / Duacs Near-Real Time Sea Level Anomalies
  - Atmosphere
  - Land surface
- Other

[User guide](#)

Layer: ESSC Web Map Service for environmental data > Observations > Ocean > OSTIA SST > sea\_surface\_temperature  
Units: kelvin

Date/time: 28 May 2010 12:00:00 UTC [first frame](#) [last frame](#)

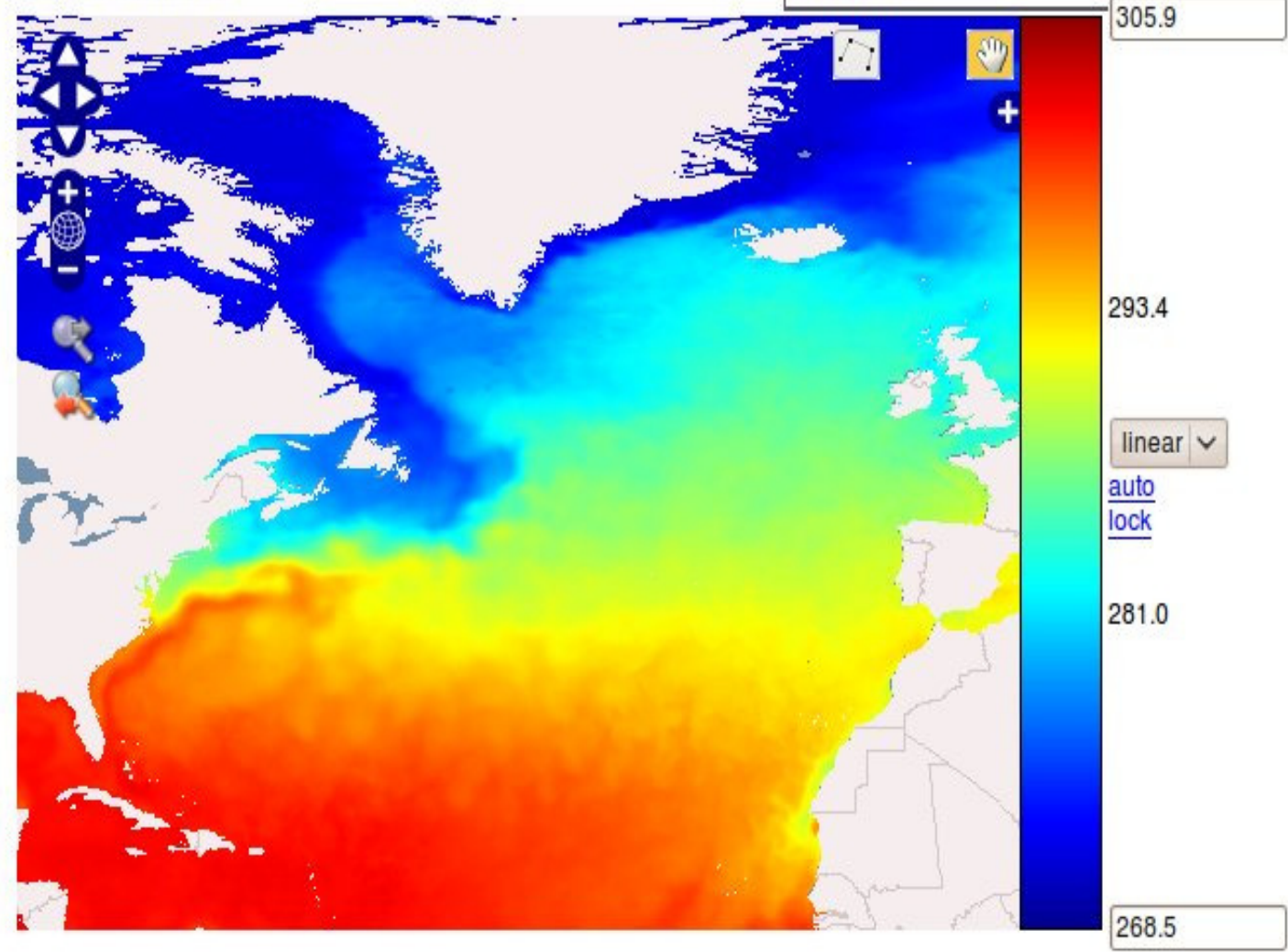
? **May, 2010**

« < Today > »

Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Select date

[Fit layer to window](#)



# OceanDIVA

- ◆ Ocean Data Inter-comparison and Visualization Application from ESSC at University of Reading.
- ◆ Provides an interactive visual comparison of gridded model output and ocean observations using virtual globes.
- ◆ Supports OPeNDAP server data and netCDF (CF-compliant) data.
- ◆ Outputs include probability distributions (png) and KML (for Google Earth).
- ◆ Java web application.



Want help or background on OceanDIVA? Try the [OceanDIVA information page](#).

Observations: EN3 ENACT/ENSEMBLES Dataset

Subset data by:  Instrument Type  Location

Model: World Ocean Atlas (2005) Climatology

Start Date: 1 Jan 2008 End Date: 31 Jan 2008

Vertical coordinate: Depth

Output: Probability Distributions

Regions (default is that N/S/Tr Atl, N/S/Tr Pac and S/Tr Ind regions are selected)

Name displayed in Google Earth: testData

Google Earth labels to include:

Profile Date  Profile id  WMO Instrument type  Max. depth of Profile

Root mean squared (RMS) misfit cutoffs:

Temperature	0.0	1.0	2.0	3.0	4.0	∞
Salinity	0.0	0.1	0.2	0.3	0.4	∞

Colour icons by: Temperature RMS: all depths

Upper: and Lower: values for user-defined RMS interval if desired

Go

### Probability Distribution output is selected.

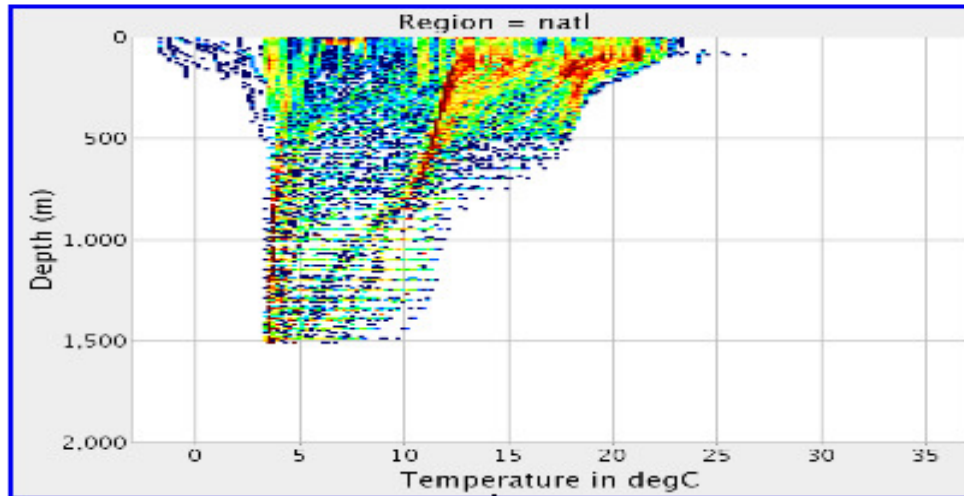
- Plots illustrating the probability of finding a given observed value, and model - observed misfit value on a given level are outputted for standard or user-defined ocean regions.
- Input is from two sources - one observations dataset, and one model dataset should be specified.
- Users should select region(s) of interest from the interface (defaults to North Atlantic).
- Larger regions will contain more data and hence produce less noisy plots.
- The probability distribution of the chosen datasets(s) (e.g. temperature and salinity on depth levels) will be outputted for the given region(s) in a series of PNGs within one HTML document.
- colour-coded values of probability are given as  $\log_{10}$  of the probability of finding that value on that level (i.e. normalized to the data on each level).

EN3 dataset has been selected. This comprises monthly quality controlled data from argo floats, buoys, XBTs and other instruments. Data is available in monthly format from 1950 until 2008 inclusive.

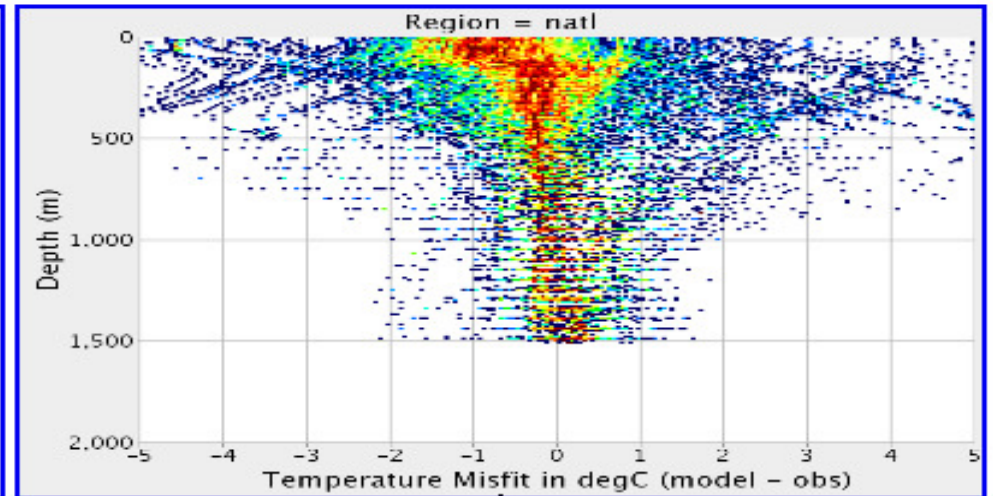
World Ocean Atlas (2005) model data has been selected. This climatology dataset is available for each month of the year.

# Region: natl

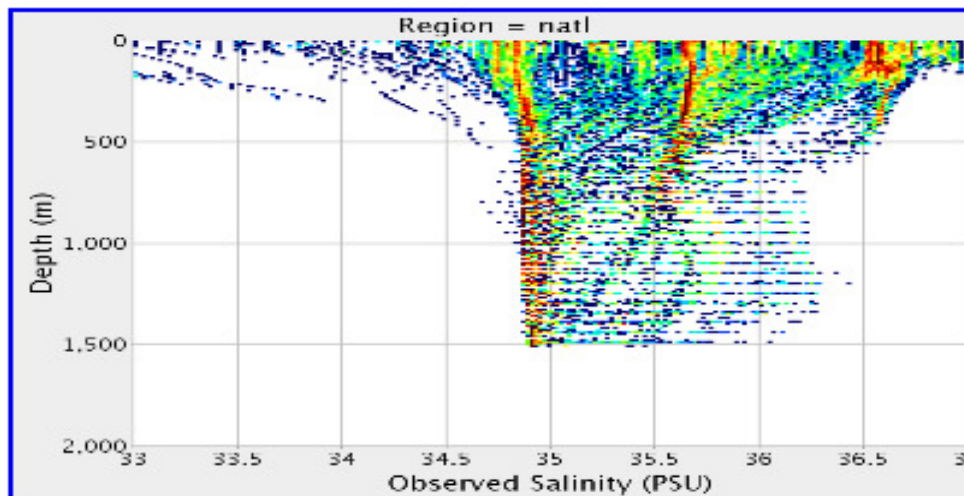
## Observed Temperature



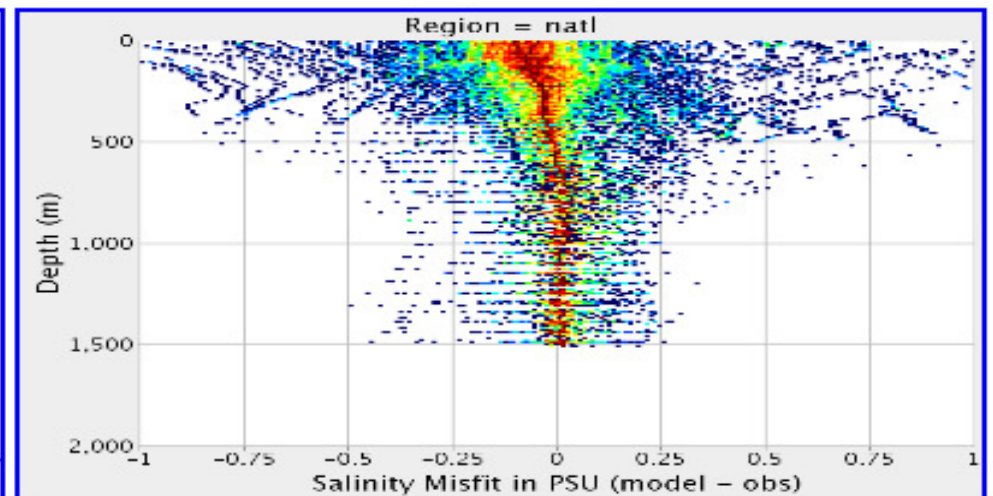
## Temperature Misfit



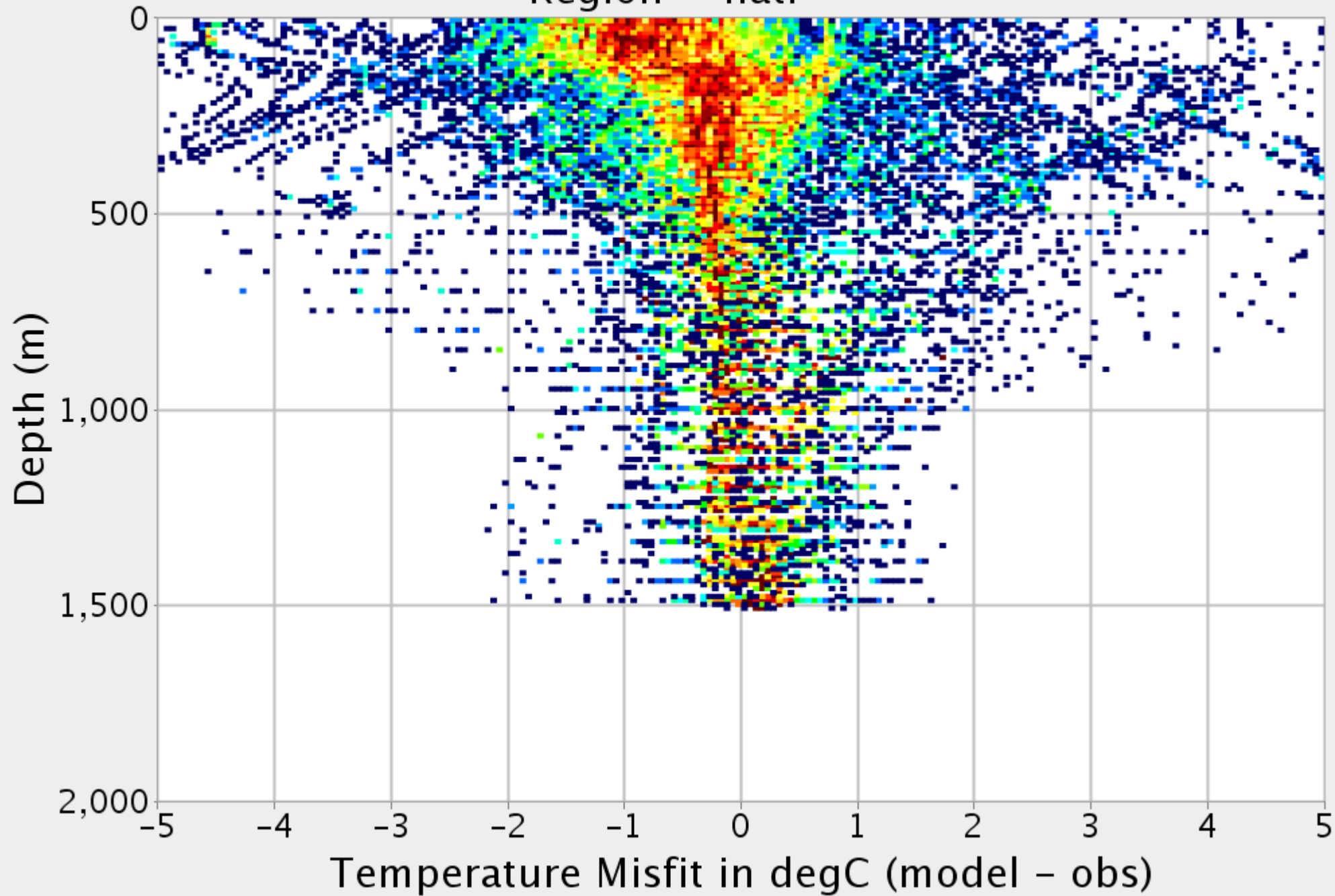
## Observed Salinity



## Salinity Misfit



Region = natl

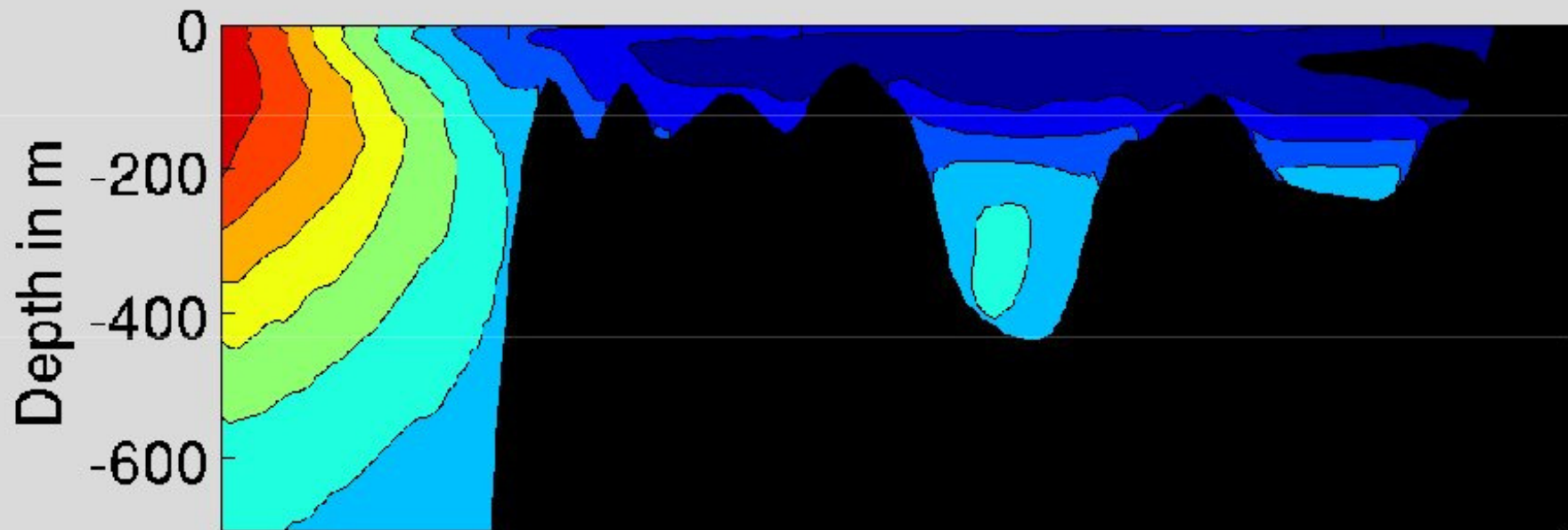
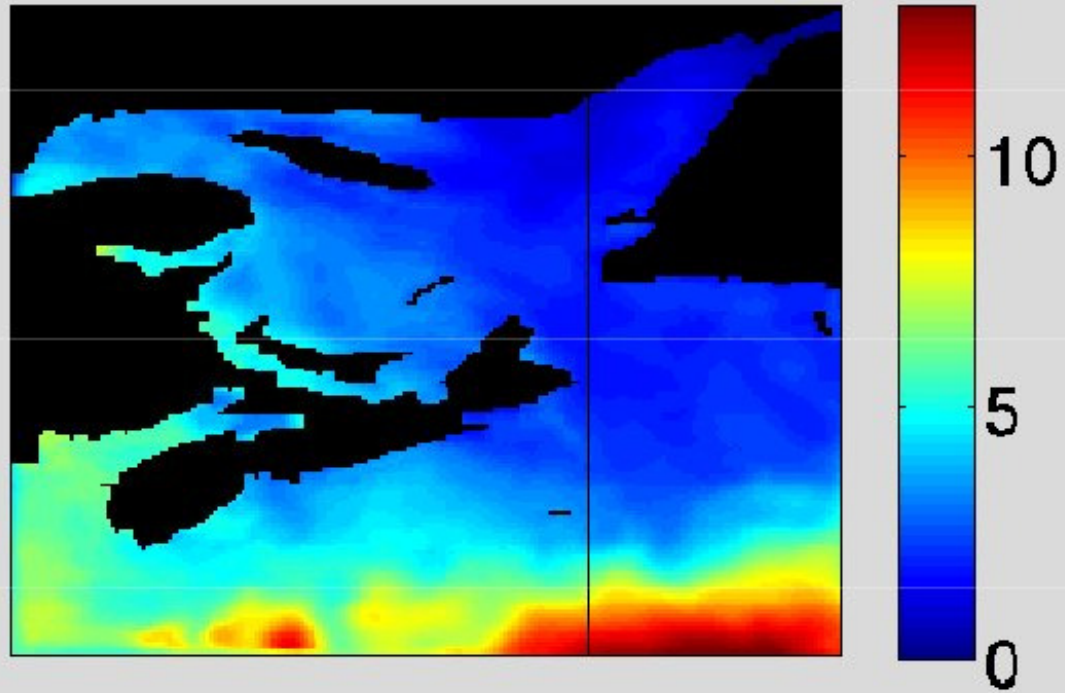


# Visualization Scripts

- ◆ It may be possible to extend the capabilities of the visualization packages using custom scripts.
- ◆ The servers 8 HT cores (16 virtual cores) should provide more than adequate compute power for on-the-fly processing.



# BODAS



# File Types & Metadata

- ◆ For the Theme I server, any GRIB datasets will be converted to netCDF, as neither OPeNDAP nor the visualization packages support GRIB format.
- ◆ The visualization packages require netCDF datasets to be CF-compliant.

# Data Uploading

- ◆ Possible methods (under discussion) for uploading data to the Theme I server include: ftp, sftp, http, DVDs and portable disk drives.
- ◆ Files should be zipped (gzip, bzip2) to reduce file size and to provide file integrity testing. MD5 128-bit checksums could be calculated from source and destination files, and compared, to verify file contents.

# Conclusion

While a number of critical data management issues have been resolved, considerable work remains to be completed. A successful outcome will make valuable GOAPP data readily accessible to researchers and stakeholders, post-GOAPP.



The following are extra slides ...

# Apache Tomcat

- ◆ Tomcat is a Java servlet and JavaServer Pages (JSP) container.
- ◆ Provides web server functionality and authentication and security functions for Java programs.
- ◆ Tomcat was developed by the Apache Software Foundation.

# LAS – Live Access Server

- ◆ Web application for serving metadata, images and scientific datasets via the Internet.
- ◆ Similar to OPeNDAP.
- ◆ Capabilities vary, but may include:
  - ◆ On-the-fly 2-D images
  - ◆ On-the-fly statistics, interpolation, conversion
  - ◆ Data subsetting
  - ◆ NetCDF, ASCII, ArcView format output
  - ◆ Interactive and batch access

# LAS Architecture

- ◆ Three tiered architecture.
- ◆ A web browser is the only LAS client.
- ◆ The back end application defines the types of data formats that may be served.
- ◆ Available back end applications include Ferret, Matlab, NCL, IDL and GrADS.
- ◆ Back end applications exist for serving netCDF, GRIB, HDF, OPeNDAP and other data formats.

# OpenLayers

- ◆ Implements OGC WMS and WFS protocols.
- ◆ Javascript library and API.
- ◆ API similar to Google Maps and MSN Virtual Earth.
- ◆ Add an 'open map widget' to any web page.
- ◆ Free – open source.
- ◆ Project of OSGeo (Open Source Geospatial Foundation).