Final Report to

Final Report for Phase I

Project Number 08121-2801.FINAL

Gulf of Mexico 3-D Operational Ocean Forecast System Pilot Prediction Project (GOMEX-PPP)

May 10, 2012

Principal Investigator:
Christopher N. K. Mooers[1], Research Professor
cmooers@cecs.pdx.edu

Edward D. Zaron[1], Research Assistant Professor

Matthew K. Howard[2], Research Scientist

[1] Department of Civil and Environmental Engineering, College of Engineering and Computer Science, Portland State University, P.O. Box 751, Portland, OR 97207-0751.

[2] Department of Oceanography, Texas A&M University, O&M Building, MS 3146, College Station, TX 77843.

LEGAL NOTICE

This report was prepared by Christopher N. K. Mooers as an account of work sponsored by the Research Partnership to Secure Energy for America, RPSEA. Neither RPSEA, members of RPSEA, the National Energy Technology Laboratory, the U.S. Department of Energy, nor any person acting on behalf of any of the entities:

a. MAKES ANY WARRANTY OR REPRESENTATION, EXPRESS OR IMPLIED WITH RESPECT TO ACCURACY, COMPLETENESS, OR USEFULNESS OF THE INFORMATION CONTAINED IN THIS DOCUMENT, OR THAT THE USE OF ANY INFORMATION, APPARATUS, METHOD, OR PROCESS DISCLOSED IN THIS DOCUMENT MAY NOT INFRINGE PRIVATELY OWNED RIGHTS, OR

b. ASSUMES ANY LIABILITY WITH RESPECT TO THE USE OF, OR FOR ANY AND ALL DAMAGES RESULTING FROM THE USE OF, ANY INFORMATION, APPARATUS, METHOD, OR PROCESS DISCLOSED IN THIS DOCUMENT.

THIS IS A FINAL REPORT. THE DATA, CALCULATIONS, INFORMATION, CONCLUSIONS, AND/OR RECOMMENDATIONS REPORTED HEREIN ARE THE PROPERTY OF THE U.S. DEPARTMENT OF ENERGY.

REFERENCE TO TRADE NAMES OR SPECIFIC COMMERCIAL PRODUCTS, COMMODITIES, OR SERVICES IN THIS REPORT DOES NOT REPRESENT OR CONSTIITUTE AN ENDORSEMENT, RECOMMENDATION, OR FAVORING BY RPSEA OR ITS CONTRACTORS OF THE SPECIFIC COMMERCIAL PRODUCT, COMMODITY, OR SERVICE.

THIS PAGE INTENTIONALLY LEFT BLANK

Signature and Date Stamp

Christopher N. K. Mooers Principal Investigator	Date

Abstract

The project aims to conduct the applied R&D necessary to demonstrate, evaluate, and establish an operational forecast system for ocean currents in the Gulf of Mexico. Such an operational forecast system is comprised of a numerical ocean circulation modeling subsystem, an ocean (satellite and in situ) observing subsystem with real-time components, and a data assimilation subsystem for initializing the forecasts. The project is being conducted in two phases. For the first phase (with a duration of 24 mos), several state-of-the-science, mesoscale eddy-admitting baroclinic ocean circulation numerical models have participated in a series of forecast experiments for assessment of their skill relative to standard metrics that are either science-based or applications-based. For the second phase (with a duration of 18 months), one or more of these models may be advanced as a pilot operational forecasting system in a real-time demonstration. The forecasting systems have been provided by Princeton University, North Carolina State University, University of California at Los Angeles, Naval Research Laboratory, Naval Oceanographic Office, National Ocean Service, and National Weather Service. An independent skill assessment group has been utilized at Texas A&M University and Portland State University. A Scientific Advisory Committee (SAC) has been formed to (1) review the plans, progress, and prospects of the project; (2) build a broad consensus on the skill of the forecast systems; and (3) recommend (at the end of Phase II) a Concept-of-Operations (CONOPS), which defines stakeholders, user requirements, roles and responsibilities, etc. for transitioning the pilot forecasting system to sustained operations. The key deliverables are (1) manuscripts documenting the skill assessment of the forecasting systems, (2) the pilot operational forecasting system, including a Website with real-time products, and (3) a recommended CONOPS, all to be completed in Phase II. The operational forecasting system will provide information that can be used to guide marine operations that are affected by transient currents throughout the water column associated with the Loop Current, the eddies it sheds, and the passage of wintertime cold fronts and summertime tropical cyclones. Such current forecasts will also have collateral benefits for marine emergency managers, environmental managers, and ecological managers.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADCP	acoustic Doppler current profiler
AGU	American Geophysical Union
AMS	American Meteorological Society
AMSEAS	NCOM American Seas
AXBT	airborne expendable bathy-thermograph
BOEM	Bureau of Energy Management
CASE-JIP	a consortium of offshore oil & gas companies
CCAR	Colorado Center for Astrodynamics Research
CONOPS	Concept-of-Operations
CPA	Closest-Point-of-Approach metric
CSDL	Coast Survey Development Laboratory/NOS
CSL	Coastal Sea Level
CWG	CONOPS Development Group, a subcommittee of SAC
DAS	data assimilation system
DBDB2	NRL Digital Bathymetry Data Base 2-minute
DwH	Deepwater Horizon
EOF	Empirical orthogonal function
FNMOC	Fleet Numerical Meteorology and Oceanography Center
GCOOS RA	Gulf of Mexico Coastal Ocean Observing System Regional Association
GOMEX	Gulf of Mexico
GOMEX-PPP	Gulf of Mexico Pilot Prediction Project
GTS	Global Telecommunications System
HMI-EW	Horizon Marine Incorporated Eddy Watch
HYCOM	Hybrid Coordinate Ocean Model
IASNFS	Inter-American Seas Nowcast Forecast System
IOOS	Integrated Ocean Observing System
IOP	Intensive Observing Period
JPL	Jet Propulsion Laboratory
LC	Loop Current
LCE	Loop Current (anticyclonic) Eddy (or, Ring)
MAST	Modeling and Analysis Steering Team (MAST)/IOOS
MITGOM	Massachusetts Institute of Technology Gulf of Mexico
MME	Multi-Model Ensemble
MMS	Minerals Management Service
MODAS	Modular Oceanographic Data Assimilation System
MODIS	Moderate Resolution Imaging Spectrometer
NAVO	Naval Oceanographic Office
NCEP	National Centers for Environmental Prediction/NWS
NCOM	Navy Coastal Ocean Model
NCSU	North Carolina State University
NGI	Northern Gulf Institute
NIMS	Near-Inertial Motions
NMFS	National Marine Fisheries Service/NOAA
NOAA	National Oceanic and Atmospheric Administration

NOGAPS	Naval Oceanography Global Atmospheric Prediction System
NOS	National Ocean Service/NOAA
NRL	Naval Research Laboratory
NWS	National Weather Service/NOAA
OCS	Outer Continental Shelf
PMG	Project Management Group
POM	Princeton Ocean Model
PROFS	Princeton Ocean Forecast System
PSU	Portland State University
PU	Princeton University
R & D	Research and Development
RADS	Radar Altimeter Database System
RFP	Request For Proposal
RMSE	Root-Mean-Square Error
ROMS	Regional Ocean Modeling System
RPSEA	Research Partnership to Secure Energy for America
RTOFS	Real-time Ocean Forecast System
SAC	Scientific Advisory Committee, broad community-based advisory committee provided by the project
SAP	Surface Atmospheric Pressure
SEACOOS	Southeast Atlantic Coastal Ocean Observing System
SECOORA	Southeast Coastal Ocean Observing System Regional Association
SIO	Scripts Institute of Oceanography
SF	Straits of Florida
SOOP	Ships of Opportunity Program
SOW	Statement of Work
SSH	Sea Surface Height
SST	Sea Surface Temperature
SURA	Southeastern Universities Research Association
TAC	Technical Advisory Committee, industry-based project oversight committee provided by RPSEA
TAMU	Texas A&M University
TOS	The Oceanography Society
UCLA	University of California at Los Angeles
UDW	Ultra Deepwater Program
URL	Uniform Resource Locator
USM	University of Southern Mississippi
USCG	United States Coast Guard
YC	Yucatan Channel

Executive Summary

The Gulf of Mexico Pilot Prediction Project (GOMEX-PPP) was undertaken to evaluate several prototype ocean forecast systems for the Gulf of Mexico, with the view towards implementing an operational forecast system in the future. The project was motivated by (1) the offshore industry's need for forecasts of ocean currents for safety and efficiency of routine operations, (2) environmental and emergency managers' needs for forecasts of ocean currents which are a prerequisite to physically-based ecological forecasts, and (3) the Navy and NOAA needs for forecasts of ocean currents to support many of their numerous mandated missions. The GOMEX-PPP approach consists of two phases: (Phase I) the evaluation and skill assessment of several extant models, and (Phase II) a real-time demonstration of a pre-operational system. This report documents the design and outcomes of the Phase I investigation.

GOMEX-PPP Phase I was conducted in the style of a model test bed in which a set of models was evaluated in a series of hindcast and forecast experiments designed by consensus of the participants who represented four academic-research models, two U.S. Navy pre-operational models, one NOAA pre-operational model, and one NOAA operational model. The experiments were performed in the following three steps: (Step 1) a year-long hindcast (i.e., a retrospective nowcast or "analysis" for 2010); (Step 2) a series of 3-month retrospective forecasts, conducted monthly, for 2010; and (Step 3) a series of 3-month forecasts, performed in real-time, for three months in 2011-2012. The experiments provided the setting for evaluating model nowcast skill and forecast skill, and for evaluating the availability of appropriate environmental data for routine operational initialization of models and systematic verification and validation of their outputs.

Deliverables of GOMEX-PPP Phase I include (1) an evaluation of the multiple data-assimilative forecast systems, (2) an assessment of multi-model ensemble forecasts, (3) a demonstration of a suite of prototype mesoscale eddy-admitting ocean prediction systems, (4) the recommendation of a concept of operations (CONOPS) for the pre-operational prototype to be implemented in GOMEX-PPP Phase II, and (5) a recommended demonstration, upgrading, and evaluation program for Phase II.

Table 1, Summary of Quantitative Comparisons, presents an overview of statistics from the quantitative assessments performed. Quantitative metrics are presented separately for nowcasts and forecasts to highlight the differences observed when models were run in nowcast vs. forecast mode.

Several observations may be made from the table. First, no model stands out as being superior in all respects. In part this is because the nowcasts from the modeling systems are all constrained by data assimilation, and are therefore constrained to be within the range of plausible states. But another reason for the lack of distinctions is the relative paucity of data available for verification and validation. In situ data tend to address smaller scales for which models show little skill, and, consequently, no skill differential (e.g., AXBT 20°C Depth).

In spite of its limitations, the summary comparison does reveal a few deficient models. For example, the initial transient and drift of several models is significant, and this is seen in terms of

Florida Current transports where these errors are larger than the natural variability. Some models are noticeably worse than the others with regard to Loop Current (LC) front position, correlation with Florida Current transport, and sea surface height anomaly (SSHA).

The report below discusses these assessments in detail, and identifies an apparently significant time-variation in model nowcast and forecast skill. The variation is consistent with a seasonal cycle, but this cannot be determined from the 1-year duration of the GOMEX-PPP experiments. It is not known whether the variation in skill is due to time-variable GOMEX dynamics or time-variation in the completeness of observations used for assimilation and validation.

Efforts to examine the usefulness of a multi-model ensemble (MME) find that ensemble spread is correlated with forecast error. This positive result demonstrates that a MME estimate of forecast accuracy should accompany long-term forecasts, thus potentially increasing their usefulness. With the proper choice of constituent models MME forecasts show an incremental increase of skill compared to single-model forecasts; however, the number of models and number of time periods for assessing MME skill is too limited to draw any definite conclusions.

The report concludes with a series of findings and recommendations. With the present state of the art and science it is possible to make long-term (3-month) forecasts of the GOMEX. Quantitative assessments find little skill in the form of variance explained, with the accuracy of initial conditions (nowcasts) placing a severe limit on the performance of the forecast systems. It is recommended that future studies consider longer, multi-year, time periods so that the seasonal vs. episodic skill variations can be quantified.

Table 1: Summary of Quantitative Comparisons

		MODEL							
QUANTITY (F-Figure #, T-Table #)	Value, Range, or Units	IASNFS	PROFS	IASROMS 3DVAR	MITGOM	AMSEAS	NGOM	RTOFS	IASROMS NHYCOM
NOWCAST									
Florida Current, Mean (T9)	30.5Sv	26.1	30.2	31.8	25.8	25.2	28.8		28.7
Florida Current, Variability (T9)	3.4Sv	2.9	2.2	3.2	3.0	2.7	1.7		3.2
Florida Current, Correlation (T9)	-1 to 1	0.51	0.56	0.24	0.33	0.56	0.28		0.10
LC/LCE Front CPA, Error (F12)	0-300km	52	61	31	51	36	143	129	78
SSHA, Error (F23)	19cm	12	11	8	11	11	27		15
AXBT 20C Depth, Error (T22)	75m-300m	42	47	40	38	35	57	82	47
FORECAST									
SSHA, 2mo. skill (F31)	-1 to 1	0.1	-0.3	0.1	0.3		-0.6		0.0
Florida Current, transient (F4)	Sv	-2	0	0	-6		0		3
Florida Current, drift (F4)	Sv/mo.	0	0	-3	0.5		0		-1
AXBT 20C Depth (F36)	m	52	52	60	44		57		59

Introduction

Background

Though much effort has been expended over the past decade to develop real-time ocean modeling systems in the Gulf of Mexico (GOMEX; see List of Abbreviations and Acronyms), forecasting of Loop Current/eddies (LC/LCE) remains problematic. Meanwhile, the Offshore Oil & Gas Industry's need to forecast the associated currents, which can extend over the full water column, has become even stronger as the Industry ventures further south in the GOMEX and into so-called ultra-deep water. While there are many existing models, none have been demonstrated to perform at the level of accuracy required. Besides the LC/LCE, other current forecast applications include guiding (1) oil spill response and (2) potential shallow water operations involving hypoxia and produced water.

The Industry loses millions of dollars every year due to drill-rig downtime caused by the presence of LC/LCE. Perhaps more importantly, there have been numerous "near-misses" documented by the Minerals Management Service (MMS) in which rig operators faced potentially serious issues. Accurate forecasts could help operators avoid some of these conditions and events through better planning and operations, and avoid potentially dangerous surprises

Objectives

The Project is focused on evaluating (Phase I; 24 mos) several candidate modeling systems and demonstrating (Phase II; 18 mos) a real-time Pilot Prediction System for the mesoscale (baroclinic) circulation of the Gulf of Mexico. (From another perspective, a multi-modeling system basis is being established in the fashion of a model testbed for the facilitation of approaches to ensemble modeling.)

The general objective is to demonstrate a well-validated operational 3-D modeling system that produces timely, accurate forecasts, nowcasts, and hindcasts of currents across the GOMEX. The aim is to have sufficient accuracy to be useful for a number of applications including LC/LCE forecasts, oil-spill trajectory forecasts, and similar current-dependent forecasts. The numerical products of the modeling system are to be Web-based so that they are, thus, available/accessible to the public and provide substantial benefits for many well-informed users.

More specifically the objectives of this project are to:

- Identify the needs of the user community for various types of prediction outputs.
- Test various methods for quantifying modeling system errors, with special attention on quantifying the errors of relevance to the end users.
- Establish model metrics that will accurately gauge the ability of the modeling system to meet users' requirements.
- Apply model metrics on a routine basis so as to provide a means for continuous monitoring of the modeling system performance in order to help improve the modeling system.
- Better utilize and synthesize on-going observations through data assimilation.

- Develop an archive of hindcasts that can be used for climatological studies, climate change detection, and diagnostic studies of ocean dynamics in the GOMEX.
- Quantify improvements made by single model and multi-model ensemble forecasts.
- Make the modeling system results and ancillary tools easily accessible (Web-based) and usable by subject matter experts.

The Project commenced in the same timeframe as the occurrence of the Deepwater Horizon (DwH) incident (20 APR 2010 and beyond) but with no direct connection to it. The Project was conceived a few years earlier as a partnership between RPSEA and the GCOOS-RA. The focus was on forecasting the position of the Loop Current (LC) and the large (ca. 100km) anticyclonic Loop Current Eddies (LCEs) it sheds every 6-to-24 months, the subsidiary, small (ca. 20km) cyclonic (frontal) eddies (LFEs), the response to tropical cyclones, the continental shelf transient circulation, and their interactions, for example, by entrainment of shelf waters by LCEs and detrainment of LFE waters by shelf flows.

To include the interests of a broad base of potential users, the project is evaluating Pilot Prediction Systems for their ability to meet the environmental prediction needs of both the offshore oil & gas industry and the GCOOS-RA community; i.e., the large, diverse GOMEX community of "super-users" and "end-users" that consumes synoptic environmental information in the spirit of IOOS. The offshore oil & gas industry has several needs for environmental predictions but the primary need is for forecasts of the LC's position and strength and those of the LCEs. The prediction horizon desired by Industry is a few months. The LC transport and position are known to have, on average, a seasonal cycle comprised of broadband annual and semi-annual components, which suggests the LC & LCE system may have useful predictability at the time scale of a few months. The GCOOS-RA community has several needs, too, including Lagrangian trajectory estimates used in search-and-rescue operations, dispersal studies of fish eggs & larvae, and calculations of transport pathways and rates for oil & other contaminant spills, especially over the continental margin. It also needs open boundary conditions for down-scaling GOMEX-scale circulation information to relatively high-resolution shelf & estuarine circulation models.

Scope of Work

The RFP from RPSEA states that the primary project aim is to establish an operational prediction (hindcasting, nowcasting, and forecasting) system for strong currents associated with the Gulf of Mexico (GOMEX) Loop Current (LC) and eddy (LCE) system in support of the oil & gas industry's southward extension of its activities into "ultra deepwater". A full water column capability is required, especially for surface and bottom currents, including those along the lower continental slope and continental rise. Secondary applications include guidance for oil spill response and potential shallow water operations involving hypoxia and "produced water". The offshore oil industry is interested in a forecast horizon of a few months for LC & LCE positions, etc. Past industry experience with forecasting LC & LCE positions has indicated a high degree of dependency on the initial conditions. Thus, attention must be given to assessing the quality of data-assimilative model analyses (or nowcasts) used for initial conditions.

This project is proceeding without the support of IOOS funding at the present time while still trying to help GCOOS-RA (& SECOORA) advance. Collaborative relationships with MMS's ongoing GOMEX modeling and observational activities are anticipated. The RFP calls for two phases to the project: Phase I (R&D and Selection Project; initial 24 mos) is comprised of two competitive experiments, and Phase II (Demonstration Project; final 18 mos) is comprised of pilot operational prediction system implementation and assessment. An early step was to firm-up "user requirements" for a deep-water current prediction system that will have a continental shelf capability, and which, in addition to the oil industry, may include, for example, marine transportation, maritime safety, and commercial and recreational fisheries users. The RFP called for a user survey; GCOOS-RA and SECOORA were asked to help in this regard. The "user requirements" helped establish the skill assessment metrics (e.g., Oey et al. (2005)).

The notion of an "operational current prediction system" is not much developed in the RFP, other than that a Website will be established (to be provided by GCOOS-RA) and kept fresh with forecast products that are meaningful to knowledgeable users. It is not made clear whether the aim is to establish, for example, a federally operated prediction system for the GOMEX as a public service or a for-profit private sector system paid for by the offshore oil industry (and also operated as a public service) or a hybrid. An example of a hybrid system would be for the Federal operational center to provide an operational analysis (nowcast) of the GOMEX on an hourly basis that would be used for initial conditions in a private or academic sector weekly LC & LCE feature model prediction system with a forecast horizon of 3 mos. The RFP allows for the possibility of using multiple models to form ensemble forecasts. In short, a Concept–of-Operations (CONOPS) does not appear in the RFP, so the CWG (a subcommittee of the SAC) will be asked to recommend viable alternatives.

Another early step is to identify a recent "historical or target year" in which interesting LC & LCE events occurred, and for which abundant in situ and remote sensing observational datasets exist. The occurrence of a Northern Gulf land-falling hurricane during the target year would be a bonus attribute. For the first experiment, a forecast was to be started at the beginning of each month of the target year and run for a 3-mos period; altogether, there would be 9 three-month forecasts to be skill assessed against GOMEX analyses. (In other words, this activity will serve as something of a de facto predictability experiment for the GOMEX.) For the second experiment, there would be a 3-mos forecasting test in the blind for a future period, again to be skill assessed against GOM analyses. The results of these experiments were to be the main determinants in selecting the model (s) for Phase II.

Operational, quasi-operational, and research modelers have agreed to participate with their models. For example, the Navy's Global-NCOM and NOAA's RTOFS-Atlantic – HYCOM operational model output were made available. Navy and NOAA were welcome to participate in the multi-week R&D forecast experiments of Phase I and the prototype multi-week operational forecasts of Phase II. Leo Oey (PU/MMS contractor and POM modeler), Yi Chao (JPL & UCLA/DAS developer and ROMS modeler) plus Ruoying He (NCSU and SECOORA/quasi-operational GOMEX ROMS modeler), Dong-Shan Ko (NRL/quasi-operational GOMEX Regional NCOM developer), and Rich Patchen (CSDL/operational GOMEX POM developer) agreed to participate in Phase I (R&D and Selection Project) and compete for Phase II (Demonstration Project). [NOTE: to participate in the proposed project, all of these modelers, of

necessity, leveraged ongoing R&D activities. Also, all of these models already cover the GOMEX Outer Continental Shelf (OCS).] Ann Jochens, Matt Howard, and Steve DiMarco, TAMU were engaged to help with accessing the observational databases, archiving model output, providing a web portal, and conducting skill assessments, together with Ed Zaron and Chris Mooers, PSU.

Project Organization

Sponsors

GOMEX-PPP is sponsored by DOE via RPSEA (Research Partnership to Secure Energy for America, a joint offshore oil & gas industry and university consortium), and CASE-JIP (an offshore oil & gas industry consortium, represented by Cort Cooper, Chevron, and Dave Driver, BP America). RPSEA provides the majority of project funding, and CASE-JIP provided the requisite matching funds (namely, 20% of total project costs).

Participants

Principal Investigators
Dr. Christopher N. K. Mooers, Portland State University (PI)
Dr. Cortis Cooper, CHEVRON (CASE-JIP Co-PI)
Mr. David Driver, BP America (CASE-JIP Co-PI)

Sub-Contractors
Dr. Yi Chao, Jet Propulsion Laboratory & UCLA
Dr. Ruoying He, North Carolina State University
Dr. Matthew Howard, Texas A&M University
Dr. Dong-Shan Ko, Naval Research Laboratory
Dr. Leo Oey, Princeton University
Dr. Edward D. Zaron, Portland State University

Affiliates
Dr. Frank Bub, U.S. Navy, Naval Oceanographic Office/Ocean Modeling Division
Dr. Bruce Cornuelle, Scripps Institution of Oceanography
Dr. Avichal Mehra, NOAA/NWS/NCEP/EMC Marine Modeling and Analysis
Mr. Richard C. Patchen, NOAA/NOS/OCS Coast Survey Development Laboratory (CSDL)

Prime Contractor's Technical Point of Contact
Dr. Christopher N. K. Mooers (503) 954-2772 cmooers@cecs.pdx.edu

Prime Contractor's Contractual Point of Contact
Karen Thomson (503) 725-9652 thomsonk@pdx.edu
DUNS Entity Number: 05-222-6800

Research and Development Participants

The research and development participants are organized according to contractor and sub-contractor model. The prime contractor is PSU (Chris Mooers & Ed Zaron); the subcontractors are NRL (Dong-Shan Ko), NCSU (Ruoying He), JPL/UCLA (Yi Chao), Princeton University (Leo Oey), and TAMU (Matt Howard, Steve DiMarco, and Ann Jochens). The modeling group is comprised of Dong-Shan Ko (IASNFS), Ruoying He and Yi Chao (IASROMS-NHYCOM and IASROMS-3DVAR), and Leo Oey (PROFS); the skill assessment group is comprised of Ed Zaron and Chris Mooers and Matt Howard, Steve DiMarco, and Ann Jochens. Additional, volunteer modeling participants include SIO (Bruce Cornuelle), NOS (Rich Patchen), NWS (Avichal Mehra), NAVO (Frank Bub), NGI (John Harding), and USM (Jerry Wiggert). Hence, there are eight modeling systems participating: five R&D models (IASNFS, IASROMS-NHYCOM, IASROMS-3DVAR, PROFS, and MITGOM), two pre-operational models (AMSEAS, NGOM), and one operational model (RTOFS). The names of participants and represented models are listed in Table 2. Their primary individual attributes are summarized in Table 3 (for more detail, see http://abcmgr.tamu.edu/gomexppp/?page_id=56).

Groups and Committees

Several groups and standing committees were formed to review GOMEX-PPP progress, and to envision the form and function of a future operational forecast system. Group and committee members represent GOMEX-PPP participants; industry, academic, and governmental stakeholders; and subject matter experts on GOMEX circulation, observations, and modeling.

Project Management Group (PMG)

Chris Mooers, PSU, PI/SAC Chair
Cort Cooper, Chevron, Co-PI/TAC Co-Chair
Dave Driver, BP America, Co-PI/TAC Co-Chair

Table 2: Participating Modeling Systems

Number	Name	POC	Institution
#1	IASNFS	Dong-Shan Ko	NRL
#2	PROFS	Lie-Yauw Oey	Princeton University
#3	IASROMS-3DVAR	Yi Chao & Ruoying He	UCLA/JPL/NCSU
#4	MITGOM	Bruce Cornuelle	Scripps Inst. Oce.
#5	AMSEAS	Frank Bub	NAVO
#6	NGOM	Rich Patchen	NOAA/NOS/CSDL
#7	RTOFS	Avichel Mehra	NOAA/NWS/NCEP
#8	IASROMS-NHYCOM	Ruoying He & Yi Chao	NCSU/UCLA/JPL
#9	ForLoop	Cort Cooper	Chevron

Web URLS:
#1 http://www7320.nrlssc.navy.mil/IASNFS_WWW/GOMEX/GOMEX.html
#2 http://www.aos.princeton.edu/WWWPUBLIC/PROFS/
#3 http://ourocean.jpl.nasa.gov/
#4 http://mitgcm.org/
#5 http://www.ngi.msstate.edu/edac/NCOM_AmSeas.php
#6 http://www.nauticalcharts.noaa.gov/csdl/NGOM.html
#7 http://polar.ncep.noaa.gov/ofs/
#8 http://omglnx6.meas.ncsu.edu/sabgom_nfcast/

Scientific Advisory Committee

A Scientific Advisory Committee (SAC) was formed to (1) review the plans, progress, and prospects of the project; (2) build a broad consensus on the skill of the forecast systems; and (3) recommend a Concept-of-Operations (CONOPS), which defines stakeholders, user requirements, roles and responsibilities, etc. for transitioning the pilot forecasting system to sustained operations.

Table 3: GOMEX-PPP Modeling Systems Attributes

Model	Numerics	Assimilation	Tides	Resolution	Assimilation window	Atmospheric forcing	Turbulence submodels
#1	NCOM	MODAS	N	6km x 41 levels	4 weeks	NOGAPS/FNMOC 0.5deg x 3hr	Mellor-Yamada 2
#2	POM	Yin and Oey, 2007	Y	5-10km x 26 sigma levels	72hr	NCEP, 0.5deg x 3hr	Mellor-Yamada 2.5, Sfc Waves, Smagorinsky
#3	ROMS	3d-Var	N	5km x 36 sigma levels	12hr	NCEP Reanalysis, 1.875 deg x 6hr	Mellor-Yamada 2.5
#4	MITGCM	4d-Var	N	4km x ?levels	2mo	NCEP Reanalysis, 1.875 deg x 6hr	Large et al KPP
#5	NCOM	NCODA	Y	3km x 40 levels	72hr	COAMPS, 15km	Mellor-Yamada 2, Smagorinsky
#6	POM	Objective-Analysis; SSH via MODAS	N	4-10km x 37 sigma levels	varies	NAM, 12km x 3hr	Mellor-Yamada 2, Smagorinsky
#7	HYCOM	Nudging (SST); Quasi-3DVAR (SSH)	Y	4-17km x 26 hybrid layers	24hr	GFS/NCEP	Canuto et al 1-pt closure
#8	ROMS	Nudging	N	6km x 30 sigma levels	varies	NCEP Reanalysis, 1.875deg x 6hr	Mellor-Yamada 2.5

For more information, please refer to http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/ModelAttributes/.

The SAC is responsible for reviewing and advising the PI and Co-PIs on the progress and plans of the GOMEX Forecasting Project. For this purpose, the SAC received progress reports and other documents to review for comment. The SAC was asked to attend the three one-and-one-half-day meetings/workshops planned, as separately described, over the course of this 30 mo.-project. Several SAC members with operational ocean forecasting roles were asked to serve on a SAC subcommittee called the Concept-of-Operations Development Working Group, as described below.

(Academia/Research)
John Allen, OSU
Bill Schmitz, TAMU-CC
Bob Weisberg, USF & SECOORA
Bob Leben, CU
Nan Walker, LSU
Alan Blumberg, Stevens Institute and consultant to BP
Patrick Hogan, NAVO/USN
Steve Payne, CNMOC/USN
Gregg Jacobs, NRL/USN
Frank Bub, NAVO
Hendrik Tolman, OMB/NCEP/NWS/NOAA
Frank Aikman, CSDL/NOS/NOAA
Alexis-Lugo Fernandez, BOEME
Robert "Buzz" Martin, Shell

(Industry)
Michael Vogel, Shell
Steve Anderson, Arete Associates
Sergei Frolov, Weather Predict Consulting Inc. (WPC)

TAC CONOPS Development Working Group (CWG)

The CWG, a subcommittee of the SAC, considered the design options for a Concept-of-Operations (CONOPS) for a Gulf of Mexico operational ocean forecast system. The options ranged from a government service providing a "public good," to a non-profit academic service providing a "public good," to a for-profit commercial service providing a "proprietary good," or some hybrid. The design considered user needs; modeling subsystem, observational subsystem, computational, telecommunication, skill assessment, performance assessment, and other technical requirements; and personnel, budgetary, management, administrative, and operational requirements. The management requirements include consideration of the roles and responsibilities of the entities involved as stakeholders, sponsors, operational forecasting service providers, etc. Recommendations of the CWG were vetted by the SAC as a whole and forwarded to TAC and others for further consideration.

Cort Cooper, Co-Chair
Dave Driver, Co-Chair
Steve Payne
Frank Bub
Hendrik Tolman

Frank Aikman
Robert ”Buzz” Martin
Chris Mooers, ex-officio

Timeline and Schedule

DEC 2010-FEB 2011: Test Case for the Experimental Plan
6 JUN 11 – Finalize the EP and start Step 1 run (one-year retrospective nowcast)
1 JUL 11 – Complete Step 1 runs
15 JUL 11– Complete skill assessment of Step 1 and start Step 2 runs (one year of 12 3-month retrospective forecasts)
15 AUG 11 – Complete Step 2 runs and start assessment of individual models and ensembles
1 SEP 11– Complete skill assessment of individual runs for Step 2 and start Step 3 runs (three months of real-time forecasts started every two weeks and run to the end of the original 3-month time horizon). Continue assessment of ensembles from Step 2
1 DEC 11 - Complete Step 3 runs
1 JAN 12- Complete skill assessment of Step 3
1 FEB 12 – Complete draft of final report
17-18 FEB 12 – Hold workshop for subcontractors, SAC, and RPSEA PM to make recommendations for Phase II
9 MAR 12 – Brief RPSEA-TAC on Phase I results and Phase II recommendations to RPSEA

Project milestones and progress are summarized on the Gantt chart, below.

TASK	Owner	Jul '10	Oct '10	Jan '11	Apr '11	Jul '11	Oct '11	Jan '12	Apr '12	Jul '12	Oct '12	Jan '13	Apr '13
1. Project Management Plan & "Kick-Off" Mtg	PSU	x											
2. Technology Status Assessment	PSU	x											
3. Technology Transfer Plan	PSU	x											
4. Monthly Reports	PSU	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
5. Identify user needs	PSU	xxx	xxx										
5.1 Gather input from users				xx									
5.2 Develop model performance specs and evaluation criteria	PSU	xxx			xx								
5.3 Write Task 5 final report	PSU		x		x								
6. Model selection		xxx	xxx	xxx	xxx	xxx	xxx						
6.1 Conduct modeling work			xxx	xxx	xxx	xxx							
6.1.1 Set up model grid, nowcast/forecast June 2010, & deliver model results to TAMU & PSU	modelers		x	x									
6.1.2 Define historical period, set model metrics, & standardize model delivery & format	PSU/TAMU		x	xxx	x								
6.1.3 Develop & test toolbox for model evaluation	TAMU/PSU			xxx	xxx								
6.1.4 Run CASE statistical model	PSU-Cort					xx							
6.1.5 Generate nowcasts for historical	Modelers				xx								
6.1.6 Generate forecasts for historical period	Modelers				x	x							
6.1.7 Generate forecasts for forecast period	Modelers					xx	x						
6.2 Conduct model inter-comparisons & evaluate results					x	xxx	xxx						
6.2.1 Develop 'true' historical picture of LC/eddy fronts	TAMU/PSU				x	x							
6.2.2 Evaluate individual models	TAMU/PSU					xxx	xxx						
6.2.3 Evaluate multimodel ensembles	PSU/JPL(?)					xxx	xxx						
6.3 Write draft then final report	PSU						xx	x					
6.4 Meet with RPSEA & recommend model								x					
***7.* Demonstrate Quasi-Operational Forecasts** **8. Model Finalization and Technology Transfer**									xxx	xxx	xxx	xxx	x

Design of the GOMEX-PPP Experiment

Introduction

The GOMEX-PPP experiment involved evaluating model performance over different time periods. In order to describe the experiment components, it is useful to establish the following terms of reference for describing the outputs and modes of operation of the participating modeling systems:

(1) The term, "nowcast," refers to an estimate of the oceanic state produced by a modeling system using data acquired prior to the time of the nowcast. Each of the GOMEX-PPP modeling systems is capable of producing a nowcast of the oceanic state by combining ocean measurements with a prior forecast using methods of data assimilation. So-called "retrospective nowcasts" are nowcasts for dates in the past which use data prior to the date of the nowcast. A retrospective nowcast differs from a "verifying analysis," in that the latter usually involves additional data not available leading up to the nowcast, including data collected after the nowcast. A "reanalysis" is an estimate of the oceanic state at a series of prior times, made using a consistent methodology, which may consist of either nowcasts or verifying analyses. In this document the term, "hindcast," is used interchangeably with "retrospective nowcast."

(2) The term, "forecast," usually indicates a prediction of the oceanic state at a date in the future; however, we shall use the term "retrospective forecast" to refer to an estimate for the oceanic state at a past date, obtained by integrating an ocean model from an earlier retrospective nowcast. Thus, it is possible to describe Step 2 of GOMEX-PPP as "a series of twelve retrospective 3-month forecasts produced at 1-month intervals," which implies that twelve retrospective nowcasts were created and used as initial conditions for a set of corresponding model runs of 3-month duration.

GOMEX-PPP Phase I consisted of the following four experimental components: (0) Step 0, an interoperability test case, (1) Step 1, a retrospective nowcast for 2010, (2) Step 2, twelve 3-month retrospective forecasts for 2010, at monthly intervals; and (3) Step 3, seven 3-month forecasts for the fourth quarter of 2011, at bi-weekly intervals. In all cases the model output consisted of 25-hour average fields, produced daily. The sequence of the four Phase I Steps is summarized in Table 4.

The experimental plan was developed by consensus of the R&D participants, and it sought to incrementally demonstrate modeling system capability. Considerations in the design of the experiments included, (1) the need to demonstrate model-data interoperability, (2) an evaluation of all-season capability, (3) the desire to exploit the opportunistic intensive observation period associated with the DwH accident, (4) the concern for the adequacy of initial conditions (Steps 1 and 2) and (4) the interest in making unambiguous forecast evaluations in a real-time operation (Step 3).

Table 4: Summary of Steps 0-3 in Phase I

Name	Simulation Start Date	Simulation End Date	Notes
Step 0	Jun 1, 2010	Jun 31, 2010	Interoperability test case.
Step 1	Jan 1, 2010	Dec 31, 2010	Retrospective nowcasts
Step 2			Retrospective 3-mo. forecasts at 1-mo. intervals
S2-01	Jan 1, 2010	Apr 30, 2010	
S2-02	Feb 1	May 31	
S2-03	Mar 1	Jun 30	
S2-04	Apr 1	Jul 31	
S2-05	May 1	Aug 31	
S2-06	Jun 1	Sep 30	
S2-07	Jul 1	Oct 31	
S2-08	Aug 1	Nov 30	
S2-09	Sep 1	Dec 31	
S2-10	Oct 1	Jan 31	
S2-11	Nov 1	Feb 28	
S2-12	Dec 1	Mar 31	
Step 3			3-mo. (14-week) forecasts at 2-week intervals
S3-01	Sep 16, 2011	Dec 23, 2011	
S3-02	Sep 30, 2011	Jan 1, 2012	
S3-03	Oct 14, 2011	Jan 20, 2012	
S3-04	Oct 28, 2011	Feb 3, 2012	
S3-05	Nov 11, 2011	Feb 17, 2012	
S3-06	Nov 25, 2011	Mar 2, 2012	
S3-07	Dec 9, 2011	Mar 16, 2012	

The 2010 time period was chosen because it coincided with the unusually large amount of in situ observations available, above and beyond routine operational observations, in association with the societal response to the DwH event. To facilitate inter-comparisons between models and comparisons with observations, it was decided to render the model output on a standard analysis grid with 1/20-degree horizontal resolution (~5km) and 22 vertical levels (20m resolution from surface to 100m, 50m resolution from 100 to 300m, 100m resolution from 300 to 1000m, and 1000m resolution below). To reduce archival volume by sub-sampling and suppress consequent aliasing from near-inertial motions (NIMs) and tides (present in some models), fields were time-averaged within a 25hr window centered on 12UTC. Modelers were free to use the bottom topography, atmospheric forcing, river runoff forcing, and open boundary forcing of their choice. Tidal forcing was optional. The forcings were to be documented and made available to others in the Project. The modeling systems were run in both the nowcast and forecast modes.

Step 0: Interoperability Test Case

Step 0, the interoperability test case, was performed to demonstrate model-data interoperability and align the requirements and expectations of the model assessment groups at TAMU and PSU with the participating modeling groups. It provided the opportunity to assess observational data and availability, and it motivated the development of a protocol for the exchange of model outputs within GOMEX-PPP.

The detailed design for Step 0 is available at (http://abcmgr.tamu.edu/gomexppp/localdocs/GOMEX-PPP_TCEP_20Dec10.doc), "Test Case for GOMEX-PPP Experimental Plan (TCEP) by Chris Mooers/20 DEC 10." Representative results for TCEP are illustrated and summarized at http://maki.cee.pdx.edu/~ezaron/TCEP/main.pdf. The purpose of Step 0 was solely to test the software machinery, and its interoperability, for the modeling systems and observational archives. In other words, no scientific or operational conclusions or decisions were to be derived from the exercise.

However, some preliminary collective findings are worthy of mention. For example, the various modeling systems exhibited the major circulation features of GOMEX, but yielded substantially different phase and amplitude results; hence, they are not absolutely redundant. When compared at face value to observations, there were large discrepancies in model results for vertical profiles of temperature and horizontal velocity, perhaps, in part, due to temporal or spatial aliasing. As another example, not all modelers were able to conform to the requested protocols for data transfer, etc.

NOTE: Though treating the topic of near-inertial motions (NIMS) was not required by the contract, NIMs can be very intense and propagate from the surface to the bottom and, thus, can influence mixing processes throughout GOMEX, perhaps especially over deep topography and under the LC. It was suggested that the modeling groups might provide high-frequency output to examine the significance or dynamics of NIMs in the models. During Step 0, Model #6 provided high frequency (hourly) model output. These data have been archived; however, they have not been analyzed at this time.

Step 1: Retrospective Nowcasts for 2010

Because they are used for initialization of forecasts, the quality of the data-assimilative nowcasts is a critical component of modeling system skill. Hence, the logical first step was to establish and evaluate the retrospective nowcasts for 2010. Among other things, it was hoped that the evaluation of the nowcasts would lead to a consensus scenario of major circulation events occurring in GOMEX during 2010. This factor was especially significant in the case of the GOMEX circulation where the present observing system does not generally allow estimation of the synoptic state by observations alone.

Since the project goal is the evaluation and demonstration of a potential operational prediction system, only operational observations were used in the nowcasts. The operational observations available in GOMEX are mainly satellite altimetric SSHs and satellite radiometric SSTs, as supplemented by coastal sea levels (CSLs) (mainly from coastal tide gauges), vertical profiles of horizontal velocity (from moored or rig-mounted ADCPs that telemeter data which subsequently

receive QA/QC in real-time), and surface current maps from (pre-operational) coastal HF radars. Observations available from the non-operational, opportunistic observing systems provide valuable independent data for skill assessment, especially during the extended de facto IOP, from MAY through SEP 2010.

As was the case for Step 0, the modelers were free to choose, from operational sources, their atmospheric, river run-off, and open boundary forcing. Tidal forcing was optional. The modelers were asked to document their forcing and make it available to the Project.

Table 5 contains a summary of the model outputs submitted for Step 1. Note that multiple versions of Model #2 and Model #4 were submitted early in the project. In December 2011, Model #3 submitted a new version of Step 1, which corrected a bug in the original submission. Likewise Model #6 submitted a revised Step 1, in which the data assimilation strategy was significantly modified. These new model submissions have not been thoroughly analyzed. The analysis contained in this report refers to the original versions unless it is explicitly stated otherwise.

Table 5: Step 1 Inventory

Model #	Status or versions
#1	Complete
#2	hcastNoT – no tides hindcast hcastT.horcon0p2 – tides + reduced horizontal diffusivity hcastT – tides
#3	ROMS_IAS_V2_2010_Nowcast_Step1 – revision, submitted Dec 2011. IAS_ROMS_V2_3DVAR_2010 – original, submitted Aug 2011.
#4	Version 1 Version 2
#5	June-December
#6	Complete
#8	Complete
#9	Hindcast is same as HMI-EW fronts.

Step 2: Retrospective Forecasts for 2010

During Step 2 a series of retrospective forecast experiments were conducted to build incrementally on the results of Step 1. The design of the forecast experiment accounted for the relevant time scales; e.g., the ocean mesoscale time scale is a few months and may be roughly tracked with a 30-day interval between forecast re-initializations. (NOTE: not coincidentally, there is a significant update of satellite altimetric SSH on the 10-day time scale.) The time scale of weather systems is several days to a few weeks and much of the consequent ocean response can be tracked with a 1-day archival rate, especially with use of daily averages to suppress aliases of the near-inertial motions engendered. Forecasts covered an entire year in order to

include seasonal effects, which may be substantial. Hence, on the first day of every month, beginning 1 JAN 2010 and ending 1 DEC 2010, a forecast was made for 90 days, with daily-averaged forecast fields archived. Altogether, for each prediction system, twelve 90-day forecasts were made and sub-sampled to daily resolution.

Table 6 lists the submissions received for Step 2.

Step 3: Real-Time Forecasts for Q4 of 2011

Step three consisted of a real-time test of the participating systems in quasi-operational mode. The modeling groups produced a series of 90-day forecasts, initialized at 2-week intervals, for the last quarter of 2011, extending into 2012. The purpose of this experiment was to exercise the modeling systems in a manner unambiguously analogous to operational mode, in which forecasts were to be produced on schedule, and there was no possibility of "future data" leaking into the forecasts. Furthermore, it was anticipated that the evolving state of the LC would provide an interesting benchmark of forecast skill during a LCE-formation event.

Submissions received for Step 3 are listed in Table 7**.** Note that some modeling groups submitted additional forecasts beyond the 7 planned for Step 3.

Table 6: Step 2 Inventory

Model	1	2	3	4	5	6	7	8	9	10	11	12
#1	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
#2	part	ok	ok	ok	ok	ok	ok	ok	ok	part	ok	ok
#3	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
#4	x	x	ok	x	ok	x	ok	x	ok	x	ok	x
#5												
#6	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
#8	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
#9	ok	ok	ok	ok	ok	ok	ok	ok	ok			

Legend:
"ok" indicates model output received and archived.
"part" indicates only a portion of the run was archived due to a file transmission error.
"x" indicates a model run which was archived, but later withdrawn from consideration.
(blank) indicates that no output was submitted or archived.

Table 7: Step 3 Inventory

Model	9/16/11-12/23/11	9/30/11-1/6/12	10/14/11-1/20/12	10/28/11-2/3/12	11/11/11-2/17/12	11/25/11-3/2/12	12/9/11-3/16/12	12/23/11-3/30/12	1/6/12-4/13/12
#1	ok	ok	ok	ok	ok	ok	ok	ok	ok
#2	ok	ok	ok	ok	ok	ok	ok	ok	
#3	ok	ok	ok	ok	ok	ok	ok	ok	
#4	ok	ok	ok	ok	ok	ok	ok		
#5									
#6		ok	ok	ok	11/10 start [and 11/18 start]	ok	12/02 start	12/17 start	
#8	ok	ok	ok	ok	ok	ok	ok		
#9	9/21 start	9/26 start		10/24 start [and 10/31 start]	11/13 start [and 11/14 start]		12/12 start		

Legend:
"ok" indicates model output received and archived.
"mm/dd start" indicates actual start date of the forecast where it differs from standard.
(blank) indicates that no output was submitted or archived.
Entries colored green are the core Step 3 forecasts. Additional forecasts received after 1/23/11 are listed but have not been analyzed.

Attributes of Participating Modeling Systems

Participants in GOMEX-PPP include researchers and ocean forecasters from institutions in education, the U.S. Navy, and NOAA. Table 2 lists the four core funded participants (models #1, #2, #3, and #8), and four affiliated participants (models #4, #5, #6, and #7), together with their institutions. The modeling systems are all based on a similar set of physical assumptions; however, they differ in many details regarding numerical methods, assimilated data and assimilation methodology, river input and atmospheric boundary conditions, and lateral oceanic open-boundary conditions.

All of the modeling systems utilize three-dimensional primitive-equation models that solve the Reynolds averaged Navier-Stokes equations subject to the hydrostatic and Boussinesq approximations. The models use different vertical resolution and coordinate systems; however,

in every case, resolution is enhanced in the upper ocean and turbulence sub-models are used in an effort to represent vertical mixing and mixed-layer processes. Air-sea fluxes of momentum, heat, and freshwater are also utilized, these being derived from the output of operational atmospheric models. Horizontal resolution of the models ranges from 3 to 10km in the GOMEX, hence, all models well-resolve oceanic mesoscale eddy processes at scales larger than 50 to 100km.

Model #1 (NRL), the Intra-Americas Sea Nowcast Forecast System (IASNFS), is based on the U.S. Navy Coastal Ocean Model (NCOM), which is derived from the Princeton Ocean Model (POM). The system has been run in a research and quasi-operational mode since 2002, and it has been described in numerous conference proceedings and the peer-reviewed literature (e.g., Ko et al, 2003; Ko et al, 2008). Satellite altimeter and MODIS SST data are assimilated via an incremental adjustment scheme that updates interior temperature and salinity fields; and in situ data are incorporated as well (Cummings, 2005; 2010). Bottom topography is based on the 2-minute resolution NRL DBDB2 topography and NGDC data, and the coastline is assigned a depth of 5m. Atmospheric forcing is obtained from the Fleet Numerical Meteorology and Oceanography Center (FNMOC) operational NOGAPS system. Long-range (3-month) forecasts computed for GOMEX-PPP utilize an annual climatology computed from 18 years of NOGAPS archived output. Tides are not included in IASNFS.

Model #2 (Princeton), the Princeton Ocean Forecast System (PROFS), is built on POM. The system has been run since approximately 1993 and has been the basis for many studies in the peer-reviewed literature (e.g., Oey, 1996). Altimetry and other data are assimilated using a combination of nudging and subsurface projection using correlations computed from a model climatology (Lin et al. 2007). Bottom topography is derived from 2-minute DBDB2 topography and NOS bottom topography, with the coastline at 5m depth, but includes an optional wetting-and-drying feature. Atmospheric forcing is computed from the NCEP GFS winds. For long range forecasts the atmospheric forcing zero after 7 days. Tides are included in Step 2 and Step 3 runs.

Model #3 (JPL/UCLA), Inter American Seas Regional Ocean Model System 3D-VARiational assimilation (IASROMS 3DVAR) is a descendent of Model #8. Like Model #8, the development of IASROMS 3DVAR was initiated during GOMEX-PPP to add 3-dimensional multi-scale variational assimilation to assimilate SSH, SST, and subsurface data (e.g., Li et al, 2009).

Model #4 (Scripps), the MITgcm Gulf of Mexico model (MITGOM), is based on the MIT general circulation model (MITgcm). The model and assimilation system have recently been developed in the context of a collaboration between BP and the Scripps Institution of Oceanography (SIO). MITGOM uses a four-dimensional variational assimilation method in which the control variables (initial conditions and air-sea fluxes) are optimized to obtain agreement with both surface and subsurface observations over a 1-month window leading up to the long-range forecast period. A consequence of the assimilation approach is that the Step 1 outputs from MITGOM are not retrospective nowcasts as defined previously. The Step 1 outputs of MITGOM are reanalysis products that depend on all data within the assimilation window, including dates before and after the nominal nowcast. Furthermore, only a subset of the submitted Step 2 model runs were retrospective forecasts, the others being reanalyses. Step 3 outputs from MITGOM are true forecasts, and are directly comparable with the other model

results. Bottom topography is derived from the 2-minute resolution ETOPO2 dataset, with a minimum depth of 10m at the coastline. Atmospheric forcings are computed from the NCEP reanalysis product for hindcasts. MITGOM is a non-funded participant in GOMEX-PPP.

Model #5 (NAVO), the American Seas model (AMSEAS), is a pre-operational version of the relocatable Navy Coastal Ocean Model (NCOM). The AMSEAS model has been running routinely since May 25, 2010, and it is currently undergoing evaluation as an operational modeling system for NAVO and NOAA. Observations are assimilated using the NRL Coupled Ocean Data Assimilation (NCODA) system, a multivariate optimal interpolation (Cummings, 2005), which assimilates in situ and satellite data available on the Global Telecommunications System (GTS) after thorough quality control (Cummings, 2006). In many respects AMSEAS is similar to IASNFS, but AMSEAS runs at higher, 3km, resolution and utilizes tidal forcing which is omitted from IASNFS. AMSEAS is a non-funded participant in GOMEX-PPP. Because long-range forecasts were not available, AMSEAS results are presented only for Step 1 of GOMEX-PPP.

Model #6 (NOS/CSDL), the NOS Gulf of Mexico Nowcast/Forecast Model (NGOM), is based on the POM model. NGOM has been run since 2004 and is currently being developed in a partnership between NOS and Dynalysis of Princeton, the site of extensive POM development. SSH and SST are assimilated using the Navy's Modular Ocean Data Assimilation System (MODAS; Fox et al, 2002), and the system is currently being upgraded with new methods for assimilating SSH data. Topography for NGOM is based on bathymetry from NGDC, Texas A&M, and NOS, interpolated on the model grid, and then smoothed based on model diagnostics. Atmospheric forcing is computed from the Navy Coupled Ocean/Atmosphere Prediction System (COAMPS), but it is intended to use NOAA's North American Mesoscale (NAM) model in the future. NGOM is a non-funded participant in GOMEX-PPP.

Model #7 (NWS/NCEP), the Real Time Ocean Forecast System (RTOFS), is based on the HYCOM model. RTOFS has been operationally producing 120 hour forecasts since May, 2006, within a domain that covers the entire North Atlantic (Mehra and Rivin, 2010). Due to computational instability outside the GOMEX domain, the RTOFS forecasts for 2010 were restarted several times in 2010, leading to discontinuities in the forecast fields within the GOMEX. After this initial analysis, the RTOFS outputs were not analyzed further in GOMEX-PPP.

Model #8 (NCSU), Inter American Seas Regional Ocean Model System Nudge to HYCOM, (IASROMS NHYCOM), is based on ROMS and is the direct descendent of the SABGOM model (Hyun and He, 2010). Although the SABGOM model has been used for several years in process-studies and forecasts, the development of a new system, IASROMS NHYCOM, was initiated during GOMEX-PPP to add rudimentary assimilation capabilities. Bottom topography is based on DBDB2. Atmospheric forcing is obtained from the NCEP Reanalysis. Long-range (three month) forecasts computed for GOMEX-PPP utilize a monthly climatology computed from 48 years of NCEP archived output. Tides are not included. IASROMS NYHCOM does not directly assimilate observations, instead the model state is nudged towards the data assimilative solution of global HYCOM (Chassignet et al, 2009). Thus, IASROMS NHYCOM runs as a prognostic

model whose solution is blended with the lower-resolution fields from the global data-assimilative model.

In summary, the modeling systems used in GOMEX-PPP were all based on state-of-the-art numerical ocean models, and incorporate realistic atmospheric forcing. Data and data assimilation methodologies differ in detail amongst the models, but all are based on some form of Gauss-Markov smoothing or nudging. Model #4 is unique amongst participants in that the assimilation methodology is not entirely sequential, instead an optimization of model control parameters is performed over a 1-month time window. Model #2 has the most extensive literature connected with process studies in the GOMEX. Non-comprehensive validation studies exist for models #1, #2, #6, and #7. Two of the systems, #3 and #8, were undergoing active development during GOMEX-PPP. Table 2 and Table 3 summarize the attributes of the modeling systems.

Attributes of Available Observing Systems

Introduction

Observing systems and the data they produce may be categorized as operational or non-operational.

Operational data products receive sustained institutional support, are documented with calibrations standards, and are part of a recognized production suite for a federal agency. Operational data are intended to be used on a routine (e.g., hourly, daily, or weekly) basis for weather and ocean forecasting, safe navigation, other marine operations, emergency management, environmental management, fisheries management or other purposes.

Non-operational data are the product of research, pre-operational, or proprietary (commercial, national security, etc.) programs or systems. Examples of non-operational data include products provided in a delayed mode, which may be delayed for a variety of reasons: (1) proprietary use by the originator (e.g., HMI drifters), (2) application of additional quality control to an operational data stream (e.g., SOOP XBT data), (3) anomalies in the processing of an operational data product, (4) data acquired without real-time telemetry (e.g., sensors attached to moored buoys or autonomous profilers working in the research mode), or (5) data acquired for a limited term experiment and/or over a limited area. The key features of non-operational data products are lack of support for sustained observations, lack of calibration or validation of instruments or methodologies, or lack of provision for real-time access. Because the same data product may be distributed through alternate channels, the operational versus non-operational characteristics of a data product may not be easily distinguished.

Some modelers use operational data obtained via a non-operational channel, such as the Internet (see http://polar.ncep.noaa.gov/mmab/oper.shtml, which notes that the Internet is not officially part of the operational data distribution system). Transitioning to use an operational channel, for example, the GTS, may require significant effort. Another consideration is that the distinction between operational and non-operational data may not be evident at the time data are provided. For example, it may sometimes occur on the GTS that apparently distinct measurements

represent duplicate data provided with different levels of quality control. Cummings (2011) discusses the challenges of data quality control and assessment in an operational setting.

Operational Data

The key data in the operational data stream relevant to GOMEX-PPP are (1) SST measurements from the NOAA Geostationary Operational Environmental Satellites (GOES) and the AVHRR instrument; (2) water level measurements, atmospheric measurements, and some subsurface temperatures from several coastal tide gauges, C-MAN stations, and a DART buoy in the Caribbean, (3) water current vertical profiles from more than a dozen moored BOEM/offshore industry ADCPs; (4) surface meteorology, and other data from several NDBC buoys; and (5) XBT data from the Ships of Opportunity Program (see Table 8).

Non-Operational Data

There are several sources of routinely-available, real-time, or near-real-time data which are also available: (1) SST measurements from NASA (e.g., MODIS-AQUA) and EUMETSAT satellites (e.g., METOP IASA); (2) SSH measurements from the NASA/CNES JASON-1 and JASON-2 satellites, and the CNES Envisat and ERS satellites; (3) altimetric SSH data merged from several satellite missions and presented in an objectively analyzed map format by AVISO; (4) coastal HF radar systems; and (5) in situ data available in delayed mode from various sources such as NODC, BOEM/SAIC (including moored profilers, current meters, and surface drifters), GCOOS-RA, and AOML/RSMAS (including AXBT flights) (see Table 8).

In addition to the above-mentioned data, which are publically available, there is also an extensive commercial program of periodic observations (e.g., surface drifters, CTD profiles, ship-mounted ADCPs, AXBT flights) conducted by Horizon Marine, Inc. (HMI).

NOTE: in recognition of their importance to numerical ocean prediction, the satellite altimetric missions are basically pre-operational in nature and presumably headed toward a fully operational status; hence, the SSH data are treated as operational in GOMEX-PPP.

Table 8: Observing System Attributes

Data Type	Platform	Detailed ID	Averaging	Data Source
Satellite Data				
NRT SSHA	Jason-1	tracks: 15, 91,167, 204, 26, 102	none	RADS
NRT SSHA	Jason-2	tracks: 15, 81, 167, 128, 204, and 26	none	RADS
SST	AVHRR/AMSR (Reynolds v.2)	rectangle: (90W,22N) to (82W,30N)	daily average	NOAA
In Situ Currents				
MMS ADCP	Brutus Genesis Holstein Auger Boomvang Ursa Mars Thunder Horse Blind Faith Na Kika Horn Mountain Ram-Powell	42362 42372 42370 42361 42373 42365 42363 42887 42385 42375 42374 42364	25-hour average centered on 12UTC; vertical interpolation to analysis grid.	NDBC [based on list provided by Cort Cooper]
SAIC-CM (MAY09-JUL10)				
SAIC-ADCP (MAY09-JUL10)				
In Situ T/S				
ARGO drifters				AOML
XBT/AXBT				AOML
Lagrangian Surface Drifters				
Coast Guard				
GDPAC				
NAVO				Frank Bub
Ocean Circ. Group				
SLDMB				Dave Driver
GTS				Dave Driver

RADS: http://rads.tudelft.nl/rads/rads.shtml
NOAA: ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/NetCDF/2011/AVHRR-AMSR/
CO-OPS: http://tidesandcurrents.noaa.gov/
NDBC: http://www.ndbc.noaa.gov/
TAMU: http://lighthouse.tamucc.edu/pq/195/
AOML: http://www.aoml.noaa.gov/data/

Results and Discussions

Overview of GOMEX Conditions During GOMEX-PPP Timeframe

2010

During the start of the year, LCE Ekman was present in the western Gulf. As the year progressed, the LC intrusion pushed further northward until reaching about 27°N. In late May through July, several large undulations in the LC led to transient formations of LCE Franklin. Thermal imagery shows large scale cyclonic eddies, particularly to the NE of the LC/LCE. Conditions evolved rapidly through this time period. There is an intriguing "puff" of water westward which looks like an eddy dipole in the thermal imagery.

Summary:

Processes involved in 2010 LCE Franklin Separation as inferred from animations of potential vorticity on 17.25°C surface:

(1) Model #1: LCE separation is preceded by cyclone formation to NE, which merges with a cyclone to SW.
(2) Model #2: A cyclone to NE seems to come from further north than in Model #1, and merges with a weak cyclone from SW. Subsurface PV seems to be attached to Florida shelf as in Model #1. Separation appears to be initiated by PV anomaly from SW.
(3) Model #3: A cyclone from the N becomes attached to LC intrusion, while a small cyclone crosses from YC. The PV anomaly appears to push across from the E.
(4) Model #4: A cyclonic PV anomaly is carried along the LC intrusion or initiated by its instability. The strength and location of PV anomaly on Florida shelf is quite unlike other models.
(5) Model #5: n/a
(6) Model #6: Unclear where first cyclones start, but they are advected rapidly around LC intrusion. LCE Franklin seems to be initiated by positive PV anomaly at NE. Oscillations of LCE are noticeably larger in this model compared to others.
(7) Model #8: Similar to model #6, in that LC intrusion seems to oscillate and a cyclonic anomaly is advected towards LCE separation point. Separation event appears to originate with PV anomaly from SE.

Animations of Ertel potential vorticity from Step 1 illustrate LCE genesis and separation, http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/PV17.25_Step1/.

Animations of frontal locus from Step 1 illustrate LCE movement, http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/.

Atmospheric Forcing

Introduction
Fluxes of momentum (wind stress), heat, and mass (fresh water) across the ocean surface are drivers of diverse processes in the Gulf. The degree to which LC and LCE processes are directly influenced by air-sea fluxes is currently unknown, but it is thought that both internal variability and the air-sea momentum flux play a significant role in the dynamics of LCE formation (Chang and Oey, 2010; Chang and Oey, 2011).

Step 1
Snapshots of wind stress and atmospheric pressure at two-month intervals for Model #1 are examined (Figure 1). Large scale patterns of surface pressure and wind stress are represented in the atmospheric analysis, in this case, from the FNMOC NOGAPS atmospheric model. Surface wind stress is also available from Model #2. It is similar, though not identical, to the Model #1 wind stress.

Models #3 and #8 also provided wind stress and pressure fields. Wind stress is similar to the above models, but surface pressure appears to be spurious. Spot-checks against the NCEP GFS data from which it is derived indicate that these models use the "surface" atmospheric pressure (which varies with land orography) rather than atmospheric pressure reduced to mean sea level. The interpolation schemes appear to be defective in these models, and pressure over the ocean is contaminated by land values.

Additional figures may be found at
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Step1_AtmForcing/.

Steps 2 and 3
Atmospheric forcing used in the forecast runs of Step 2 and 3 was not examined. It would be useful to verify that the forcing used in Step 2 was obtained from forecast, climatology, or persistence, rather than an atmospheric reanalysis.

Summary
Surface pressure was examined for models #1, #3, and #8. Model #4 does not use atmospheric pressure to force the model, and surface pressure was not provided for models #2 and #6. The atmospheric pressure used in models #3 and #8 is spurious; while this is believed to have little impact on the present analysis (which is focused on LC and LCE properties), it may contribute to increased mismatch with altimetric SSH if the latter includes an inverse barometer correction when it is assimilated into the models.

Surface wind stress was provided by groups #1, #2, #3, and #8. It is grossly similar amongst the models.

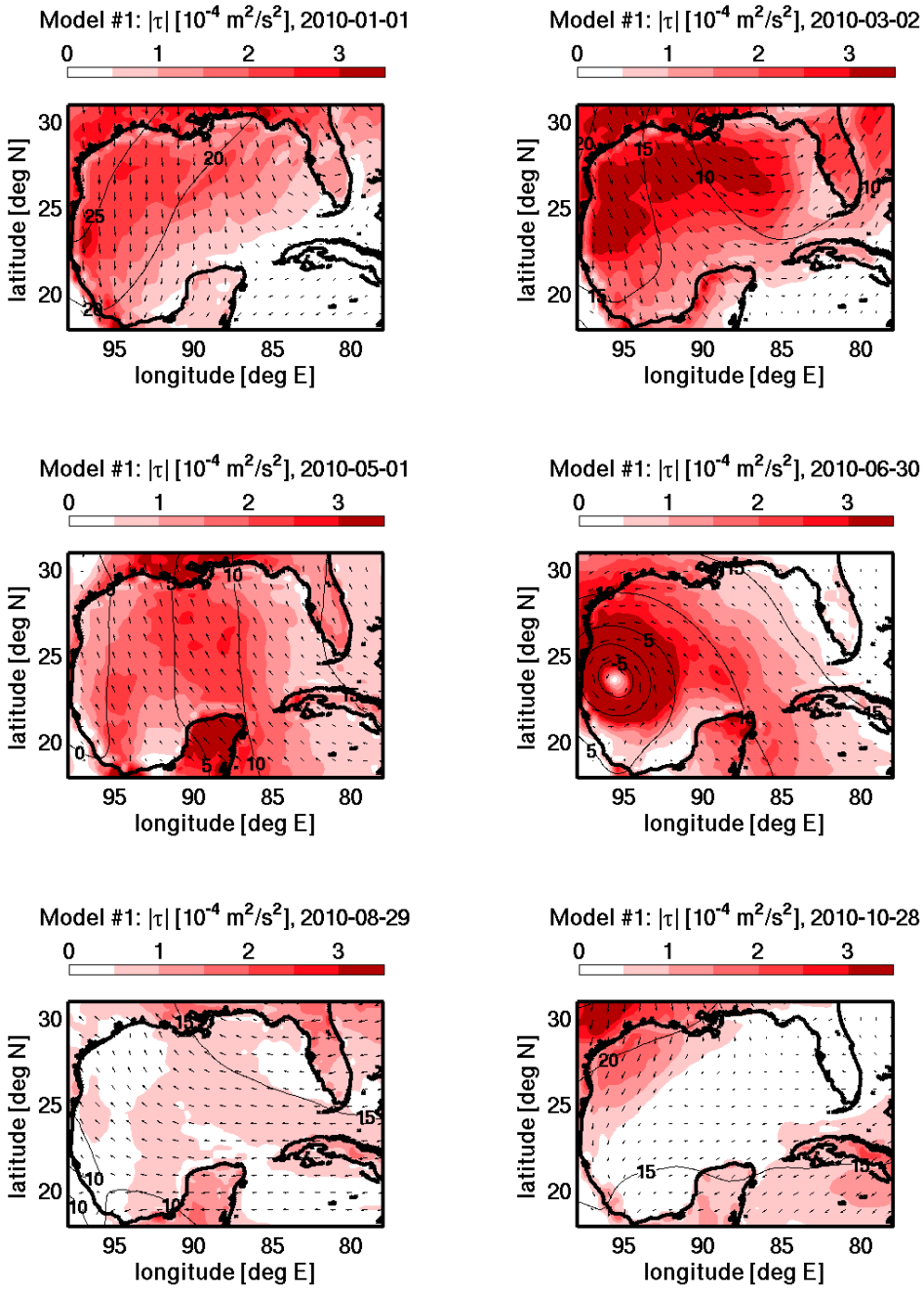


Figure 1: Atmospheric Forcing from Model #1.
Wind stress direction (arrows) and magnitude (color) are shown in relation to sea-level pressure (contours). Note the presence of a tropical cyclone passing through the western Gulf in late June.

Transport and Flow Distribution Across Selected Zonal Transects

Introduction

The Loop Current is part of a larger current system along the western boundary of the North Atlantic which includes the North Brazil Current, the Antilles Current, the Yucatan Current, and the Florida Current, all parts of the Gulf Stream system. The average northward volume transport of the wind-driven component of the Loop Current is determined by the curl of the wind stress to the east of the Gulf of Mexico with possible influences due to complex topography, non-linear recirculation, and other processes. The thermohaline contribution to the transport is not well known (Lee et al, 1996). The transport of the western boundary current has been monitored since 1982 at the Straits of Florida, between the east coast of Florida and the Bahamas at 27°N (Larsen and Sanford, 1985; Shoosmith et al, 2005). The western boundary current is called the Florida Current at this site, and the fidelity of the models at reproducing and forecasting the Florida Current transport is significant because recent evidence suggests a possible upstream trigger mechanism for the formation of LCE (Sturges et al, 2010). Mooers et al (2005) noted upstream (equatorward) phase propagation of sea-level anomaly in coastal tide gauge stations and in the Navy's real-time global NCOM model.

Previous analyses of LC penetration and LCE formation have implicated the flux of vorticity through the Yucatan Channel (Candela et al, 2002; Oey et al, 2003). Thus, modeled flow characteristics in the Yucatan Channel may also be significant in predicting LCE formation and separation.

There is uncertainty about the relationship between volume transport through Yucatan Channel and the Straits of Florida. Hamilton et al (2005) found that the transport at 27°N is not a precise measure of LC transport or transport through Yucatan Channel, finding time-variable differences of roughly 5 Sv, which is similar to the magnitude of variability of the Florida Current transport. In contrast, more recent observations based on longer time series and direct measurements at both sites appear to show balance in the transports (Rousset and Beal, 2010).

The methodology for the analysis here involves comparison of the observed Florida Current Transport at 27°N (data obtained from http://www.aoml.noaa.gov/phod/floridacurrent/index.php) with time series of depth-integrated meridional transports through Straits of Florida (26° 40'N) and Yucatan Channel (21° 45'N, 86° 45'W to 84° 15'W). Vertical integrals of model fields were computed by the trapezoid rule, and horizontal integration was performed using the rectangle method with data centered on grid points (i.e., the midpoint rule). Comparison of IASNFS and PROFS volume transports, computed on native grids by the modelers, showed differences of as much as 1.5 Sv in the time-average transport. Correlations with cable data are reduced by about 0.06 when using transport from the re-gridded fields. Alternate methods of computing vertical integrals lead to transports which differ by as much as 6 Sv, with the trapezoid rule giving closest agreement with native grid results.

Autocorrelation is computed using the biased estimator which normalizes the sum of products by the length of the record, rather the number of terms in the sum (eq. 5.3.8, Jenkins and Watts, 1968). The correlation time, tau, is defined as the lag of the first zero-crossing. The integral time scale is not reported since correlations do not decay to zero sufficiently fast, suggesting a lack of

stationarity or correlation times greater than 1 year. Spectra are computed using Welch's method with a Hann window on annual time series which have been divided into seven segments with 50% overlap. Although the spectra are relatively steep, no significant differences were noted when the time series were pre-whitened. The squared coherence and phase function were computed from the cross-spectrum using Welch's method and a Hann window on 1-month data segments with 50% overlap. Approximate confidence intervals are computed by assuming the arctan of the coherence error is normally distributed and the tangent of the phase error is normally distributed (pp. 379—380, Jenkins and Watts, 1968).

Step 1

Florida Current

Results for Step 1 are summarized in Table 9. Models #3, #4, #6, and #8 have less than 0.5 correlation with the observed time series of Florida Current transport. Among the models with correlation greater than 0.5, one notes autocorrelation time (tau) ranging from 10 to 61 days, reflective of the wide range of frequency content amongst the models. Note that coherency is quite variable, and some poorly-correlated models have large coherency within a small frequency band. Models with better correlation to the observed time series tend to have larger coherency at low frequencies.

Figure 2 summarizes the Step 1 comparison in a Taylor diagram (Taylor, 2001). Model #2 has the lowest root-mean-square deviation (RMSD) of model minus data, a measure of its good correlation with the data and its similar variance. Models #1 and #5 achieve nearly equal performance.

Table 9: Florida Current Transport, Step 1

Model	Mean [Sv]	Std dev [Sv]	Tau [day]	Correlation	Max coherency
Cable Data	30.5	3.4	54	n/a	n/a
#1	26.1	2.9	42	0.51	0.70
#2	30.2	2.2	61	0.56	0.70
#3	31.8	3.2	46	0.24	0.65
#4	25.8	3.0	48	0.33	0.60
#5	25.2	2.7	10	0.56	0.65
#6	28.8	1.7	61	0.28	0.40
#8	28.7	3.2	11	0.10	0.80

Notes:
(1) Model #2 comparison for April 1 -- December 31, when winds were "turned on."
(2) Model #2 shows evidence of a constant phase lag, following the observed data by approximately 12 hours.
(3) Model #5 comparison is for May 25 -- December 31, when model output is available.
(4) Model #5 shows evidence of a constant phase lag, following the observed data by approximately 6 hours.

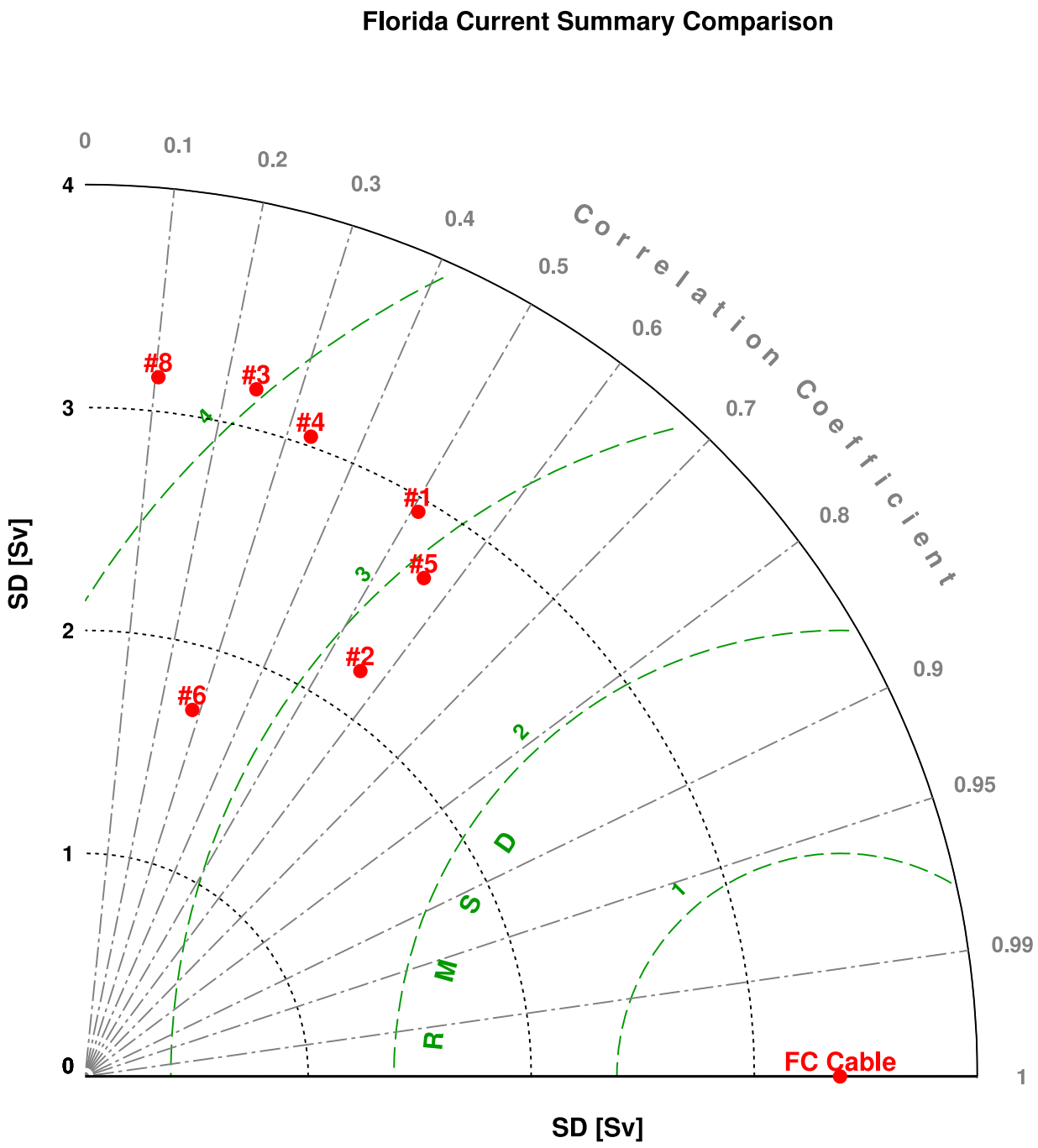


Figure 2: Florida Current Taylor Diagram, Step 1.
The Taylor diagram displays model summary statistics (red dots, labeled #1-#8) in polar coordinates. The angular coordinate is correlation, with zero representing a perfect correlation. The radial coordinate is standard deviation. Distance between the model and Florida Current data (red dot, labeled FC Cable) is the standard deviation of the error.

Figure 3 shows a complementary summary with a target diagram (Jolliff et al, 2009). Model #2 has the smallest bias compared to the cable data; hence it has the lowest RMS error (distance to the origin). The dashed green line on the target diagram indicates the standard deviation of the observed transport, which is a combination of temporal variability and measurement error, the latter estimated at approximately 1 Sv (Larsen and Sanford, 1985). Overall, the hindcast transport is significantly less variable than the observed transport, and the mean transport is much too low in models #5, #4, and #1; although, the discrepancies may be partly caused by re-gridding error. After correcting for the bias, a skill score may be defined as (1 - RMSD/sigma_d)^2, and the best model (#2) obtains a skill of less than 0.125.

Additional plots from the Step 1 analysis, including EOFs of the meridional velocity field, are found at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Step1_FC_EOF/. Significant observations from the transects are noted, below:

(1) Model topography differs substantially. Models #1, #3, and #8 are similar, with maximum depth of about 775m and a 600m-deep side channel to the east of the main channel. Models #2 and #4 contain no side channel, and the maximum depth is slightly less than 700m. The topography in Model #6 is different again, and reaches a maximum depth of 600m
(2) The haline fields are different in the models, indicating different water mass sources for the transported water. Models #1, #4, and #6 have a well-defined subsurface maximum in salinity towards the east end of the transect. Salinity decreases monotonically with depth in models #3 and #8. Salinity is nearly vertically-uniform in Model #2, but increases to the east.
(3) The structure and explained variance of the zonal velocity EOFs differ greatly. In models #4 and #6, the first EOF explains approximately 70% of the variance, but the horizontal structure is mode-0 in Model #4 (no zero-crossings) and mode-1 in Model #6 (one zero-crossing). In the other models, the leading EOF explains less than 50% of the variance, and the horizontal structure again differs amongst the models.

Yucatan Channel

Table 10 summarizes the mean and standard deviation of the transport through the Yucatan Channel in the models. Additional plots are found at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Step1_YC_EOF/. In every case the transport variability is higher in the Yucatan Channel than in the Straits of Florida.

Comparisons of time series and power spectra at Yucatan Channel and Straits of Florida are at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Step1_Delta/. In all the models, except Model #6, there is significant coherency between transport at both sites, with maximum coherence occurring at periods between 5 and 20 days. An interesting feature of the Step 1 results is that, in all models except Model #2, the phase of the transport at Straits of Florida leads the transport at Yucatan Channel, consistent with the results of Mooers et al (2005).

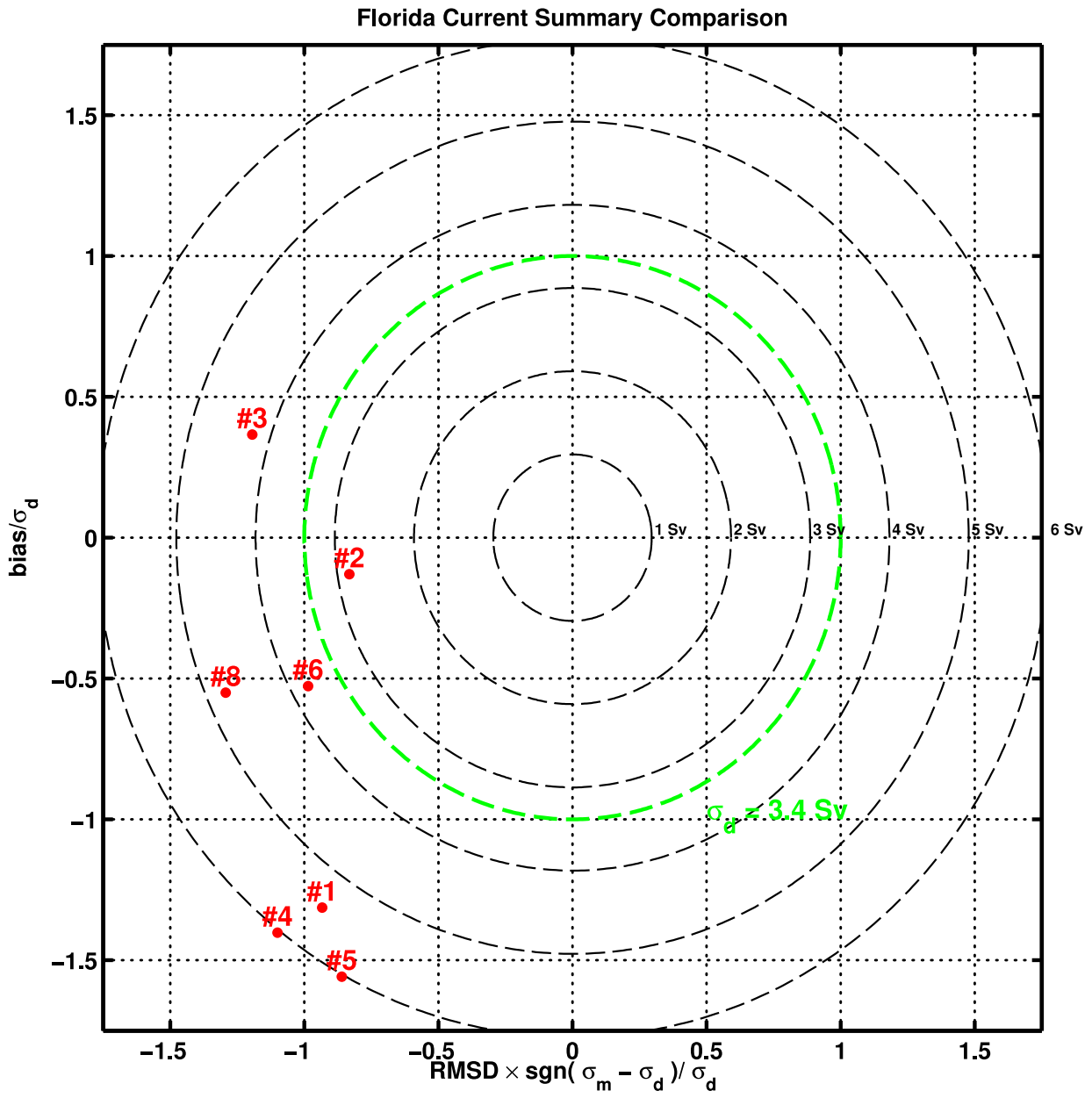


Figure 3: Florida Current Target Diagram, Step 1.
The target diagram displays model summary statistics (red dots) in terms of standard deviation of error (x-axis) and bias (mean error, y-axis). Thus, distance from the center of the diagram is root-mean-square error. The dashed green circle indicates the standard deviation of the observed Florida Current transport.

Table 10: Yucatan Channel Transport, Step 1

Model	Mean [Sv]	Std dev [Sv]	Tau [day]
Cable Data (Florida Current)	30.5	3.4	54
#1	22.3	3.2	69
#2	30.0	3.4	65
#3	32.6	4.1	45
#4	24.6	4.5	61
#5	25.1	3.5	42
#6	29.4	2.0	50
#8	28.8	3.5	12

Step 2

Transport through the Straits of Florida is related to winds (Lee and Williams, 1985), and we expect that the transport will have limited predictability due to the associated lack of predictability of winds beyond several days. This expectation is confirmed by the data in Table 11, which shows the correlation coefficient between observed and modeled transport through the Straits of Florida, averaged over the twelve Step 2 forecasts. Because both modeled and observed transport time series are individually autocorrelated in time, significance levels for the cross-correlations must take into account the estimated degrees of freedom in the time series. Reliably estimating the degrees of freedom is difficult; however, because the variance in each signal appears to be non-stationary (see below). The variance of the sample cross-correlation is approximately (1+a^2)/(1-a^2)/N, where dt is the sample spacing, dt/(1-a) is the decorrelation time, and N is the number of data points (p. 338, Jenkins and Watts, 1968). Taking dt/(1-a)=10 days as a lower limit of the model decorrelation time, suggests that cross-correlation larger than about 0.2 should be considered significant at the 95% level. These values have been shaded green in the table.

As was mentioned above, the variance in the forecasts appears to be nonstationary. This feature is illustrated most clearly in Figure 4, which shows the Florida Current transport anomaly in each of the models as a function of forecast time (i.e., time is referenced to starting time, and transport is referenced to starting transport). Heavy solid lines show the average predicted transport anomaly, and lighter lines show the 12 individual forecasts. Models #4 and #8 have initial transients in excess of 3 Sv, Model #1 appears to have a smaller initial transient, and Model #3 is characterized by a large drift. Note that gaps in the average transports (heavy lines) occur where missing data occur (Model #2) or where model anomalies prevent computation of transport (land/sea mask changes in Model #3).

Table 11: Florida Current Transport -- Cross Correlations, Step 2

	Cable	#1	#2	#3	#4	#6	#8
Cable	1	0.29	0.42	-0.03	0.26	-0.06	-0.07
#1		1	0.05	-0.11	0.24	0.05	-0.07
#2			1	0.12	-0.02	0.06	0.03
#3				1	-0.31	0.14	0.31
#4					1	0.11	-0.18
#6						1	-0.03
#8							1

Green shading indicates correlation significantly different from zero. Forecasts from models #1, #2, and #4 display statistically significant correlation with the observed Florida Current transport.

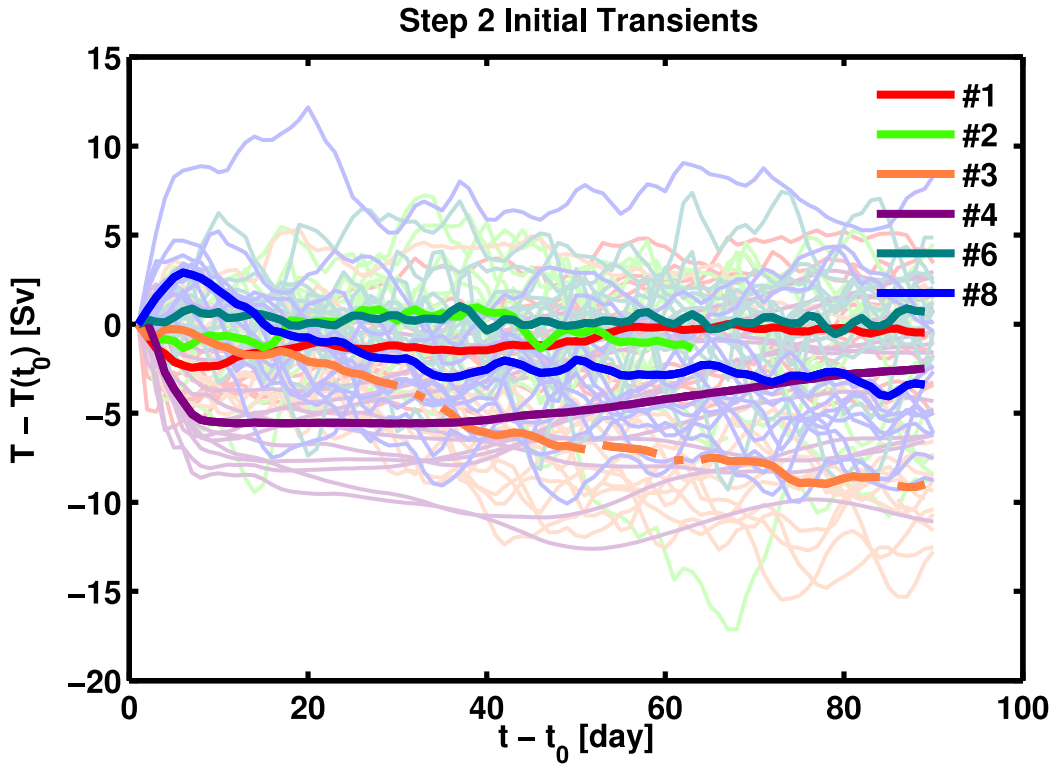


Figure 4: Florida Current Forecast Transients, Step 2.
Averaged over all Step 2 forecasts, models #3, #4, and #8 appear to display transient response or drift larger than the standard deviation of the Florida Current transport.

Step 3
Because Step 2 forecasts displayed low correlation with observations, no attempt was made to analyze Step 3 correlation.

Summary
Step 2 forecasts of the Florida Current transport made with models #1, #2, and #4 were correlated with the observations. The correlation coefficient was small, less than 0.5, but appears to be significant at the 95% level. Forecasts of models #3 and #8 were correlated with each other, but uncorrelated with observations. Models #1, #2, #5, and #6 had positive skill in the Step 1 hindcast of Florida Current transport, although, skill level (fraction of explained variance) did not exceed 0.125. Model #2 had the lowest hindcast bias (less than 1 Sv), lowest root-mean-square error, and highest correlation (greater than 0.4) compared with observed transport.

Loop Current and Loop Current Eddy Fronts

Introduction
Strong upper ocean currents associated with the Loop Current (LC) and Loop Current Eddies (LCE) significantly impact marine operations in the Gulf of Mexico. For example, currents in excess of 0.75m/s cause vortex-induced vibrations which accelerate structural fatigue and impede the safe operation of oil-well platforms, and strong currents make mooring lines and vessels difficult to control. Speeds within the LC and LCE can easily exceed the 0.75m/s threshold; hence, there is great interest in forecasting the location and intensity of strong currents.

Present practice within the offshore industry relies on analysis of nowcasts of the LC and LCE fronts, as inferred from satellite sea-surface temperature, satellite altimetry, and in situ observations obtained with drifters. For example, the Horizon Marine Incorporated Eddy Watch product (hereafter referred to as HMI-EW), is an estimate of the location of LC/LCE fronts identified by the nominal 0.75m/s isotach. Because it is widely used, the HMI-EW frontal analysis provides a reference for the intercomparison of model forecasts below, where the hindcasts and forecasts from 2010 are analyzed in detail using archived HMI-EW analyses provided by HMI (proprietary data provided by Cooper 2011; http://www.horizonmarine.com/eddywatch/ew_gulf.php).

Because of the importance of the frontal predictions to stakeholders in the offshore industry, several different approaches to the frontal analysis were conducted. Working independently in parallel, the analysis groups at PSU and TAMU developed objective (PSU) and subjective (TAMU) methods for mapping LC and LCE fronts within the models, and compared these fronts to HMI-EW. The objective approach, described first, used SSH- and subsurface temperature-based frontal definitions. The subjective approach, described second, used a human analyst to visually locate fronts by viewing animations of model output.

The objective approach was explored in detail. It was found that results are sensitive to the precise space and time domain of intercomparison of model vs. HMI-EW fronts. Model vs. HMI frontal comparisons are reported here for two spatial domains: Domain A, consisting of points within 83-96°W and 22-29°N and water deeper than 500m; and Domain B, consisting of points

within 83-92°W and 22-29°N and water deeper than 500m. Three time periods were considered: June 2010, December 2010, and entire year 2010. Model-based frontal definitions derived from analysis of Step 1 vs. HMI-EW were used in subsequent analysis of Step 2 and Step 3.

The subjective approach required approximately 1 minute of operator time per frame of daily model output. Thus, analysis of 1 year of output from 8 models required approximately 50 man-hours of effort. Hence, the subjective approach was applied to the Step 1 hindcasts, but not Step 2 or Step 3, except to identify the timing of LCE separation events.

Methodology #1: Objective Approach

To provide a benchmark for subsequent comparisons, we seek an objective procedure for identifying frontal locations in a manner compatible with HMI-EW. Previous efforts have utilized different, subjectively-determined, criteria for frontal position based on temperature at 200m depth (Oey et al., 2005; Leben and Corcoran, 2005; Kantha et al, 2005), referred to here as T200; sea-surface-height, SSH (Leben and Born, 1993; Leben and Corcoran, 2005; Leben, 2005); and surface current speed (Leben and Corcoran, 2005). Henceforth we shall let Z denote either T200 or SSH in the context of the determination of frontal location.

Two types of frontal attributes are considered in relation to HMI-EW fronts. The first type consists of the average of quantities along the model-derived front. This is applied to determine, for example, the SSH value associated with maximum near-surface speed. The second type of attribute consists of a minimum value of a certain quantity, which is used to find the closest-point-of-approach (CPA) distance between the fronts and a set of seven locations used previously for this purpose (Oey et al, 2005).

The method for computing model-derived frontal averages is as follows. Given a trial value for the front, Z=Z_f, an along-front average is defined as the spatial average over all points in the analysis domain where Z is in the set [Z_f,Z_f+dZ] for some small increment, dZ. The procedure is analogous to computing a histogram, in that the range of trial values of Z are chosen and divided into bins of width dZ. The values of the attribute corresponding to the given bin range are then spatially averaged, accounting for latitude-dependent area elements in spherical-polar coordinates. The advantage of this approach over, for example, arc-length based averaging, is that it avoids the complex topological problem of parameterizing the time-dependent curves in order to integrate along them. This method is applied to compute along-front averages of the following attributes: surface speed, near-surface current speed (average over top 200m), and average distance to the HMI-EW front.

Figure 5 illustrates the computation of a frontal average quantity. The colormap shows distance to the HMI-EW contour within the analysis region. The HMI-EW curve is the heavy black line, and the distance contours (100km and 200km) are shown by the thin black lines within the analysis region. The SSH-based frontal patch, Z in [Z_f,Z_f+dZ], is shown in gray. The computed metric is a measure of how much the model-based front would need to be deformed in order to overlap the HMI-EW front. Other metrics could be defined, but this one has the advantage of being applicable to frontal estimates of arbitrary topology and resolution.

The method for computing minimum values along the front utilizes the coordinates of the frontal contour, Z=Z_f, directly. In this case, the quantity of interest is computed separately at each frontal coordinate, and the minimum of these values is used as the frontal attribute. The model-derived frontal coordinates are determined with an arc-length resolution of 5 to 7km, while the HMI-EW frontal coordinates are typically provided at a resolution of 7 to 10km. The just-described procedure is used to compute the closest-point-of-approach metric (CPA).

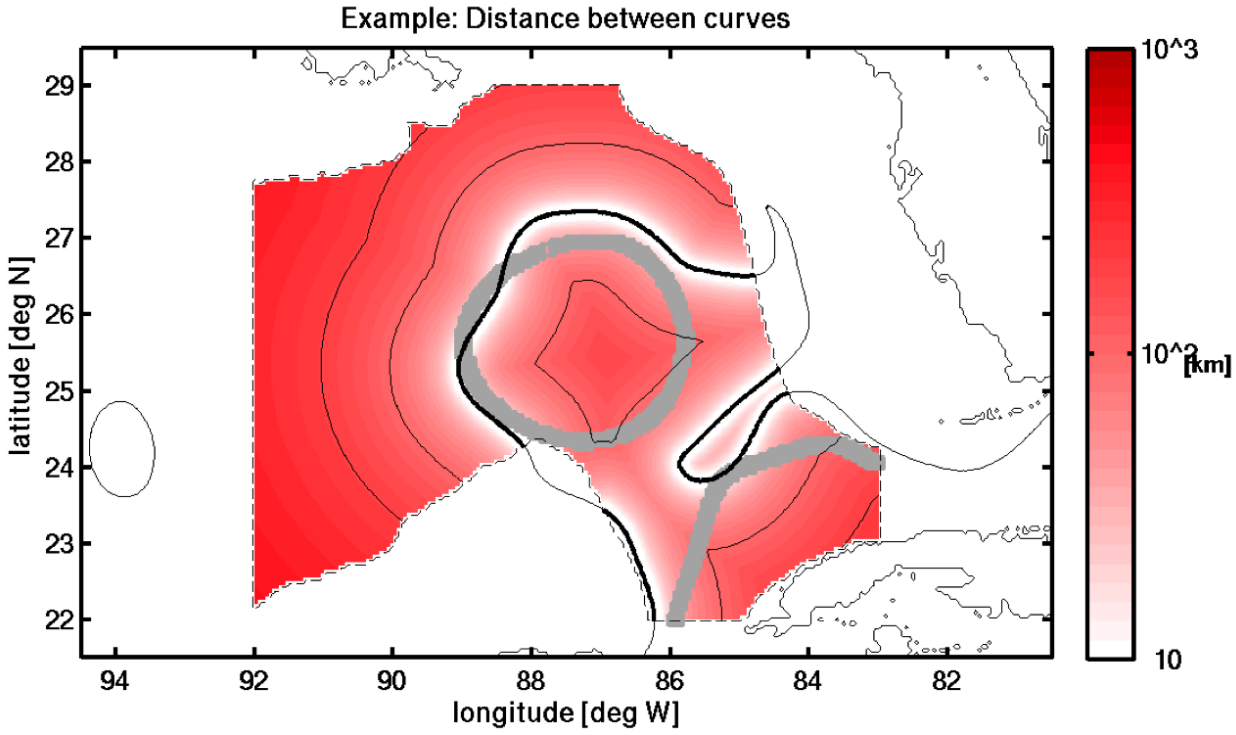


Figure 5 Determination of Distance to Front.
The distance to the closest point along the HMI frontal locus (heavy black curve) is shown with the color scale and contours. The distance between the SSH-based frontal patch (gray area) and the HMI frontal locus is defined as the average, within the SSH-based frontal patch, of the distance to the HMI frontal locus. The analysis region is restricted to the area enclosed by the dashed contour, which is defined by the 200m isobaths between 92W and 83W. Other analysis domains were used for this metric and are discussed in the text.

Figure 6 presents an illustration of the CPA metric for a SSH-based frontal definition. The quantity to be determined, $CPA_{Model}(Z\_f,n)$, is the CPA distance between the SSH=Z_f contour and the n-th reference point (located at the corners of the oil and mineral lease domain; Oey et al 2005). $CPA_{HMI}(n)$ is the corresponding CPA distance for the HMI-EW front. The dashed polygon is the analysis region, and only frontal contours within this region are used. The solid black line, thicker within the analysis region, is the HMI-EW frontal analysis for a given date (June 4, 2010). The colormap shows the modeled SSH field. The grey line is the SSH-based

frontal contour, thicker within the analysis region. Black "+" marks indicate the n=1,…,7 reference locations for the CPA distance computations, and the white and black lines emanating from these points connect to the CPA locations along the HMI-EW and model fronts, respectively. It is the average of the difference, $E_n = CPA_{HMI} - CPA_{Model}$, which is reported in tabular form, below.

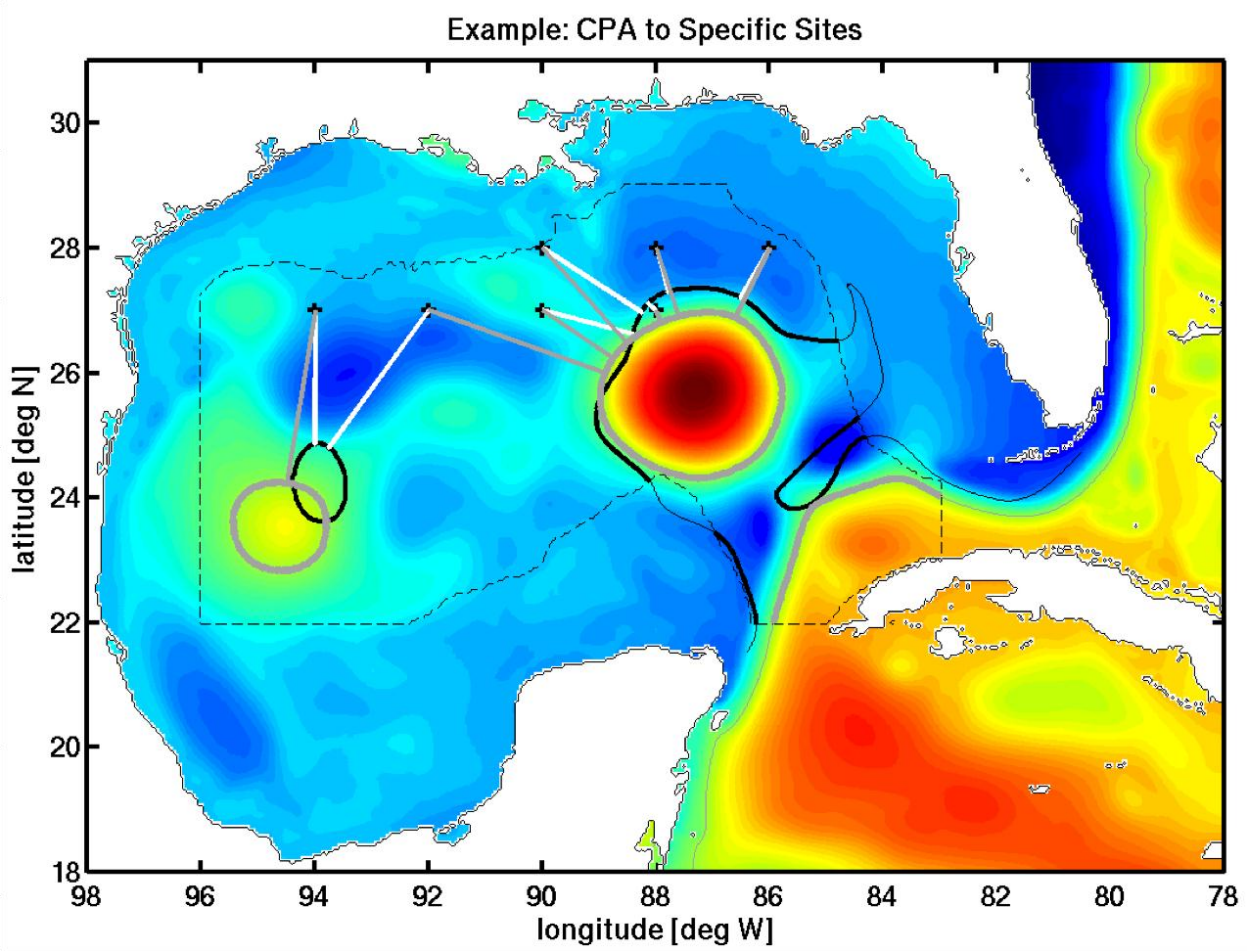


Figure 6: Determination of Closest Point of Approach (CPA).
The HMI frontal locus (black line) and SSH-based model front (grey curve) have been compared by computing distance to a set of seven reference locations (black crosses). Straight line segments indicate the closest point of approach between the seven sites and the HMI analysis (white) and the SSH-based front (gray). The dashed line indicates the analysis region. For reference, the SSH field is shown in color. The animation linked below displays a time series of frontal error metrics together with the CPA line segments: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/2010Fronts/Movies/iasnfs_T200cpa.mov (Note: the front in the animation is defined by the 18C isotherm at 200m depth)

The approach has been to compute attributes of fronts defined by either SSH or T200, and use these attributes to guide the selection of the defining frontal quantity for each model. The advantage of using frontal definitions based on SSH and temperature at a level is that these are

two-dimensional scalar fields, which, generally, discriminate well between Gulf and non-Gulf water masses. Hence, with these scalars, it is usually possible to identify an unambiguous and simply-connected contour that coincides with the edge of the LC or a LCE.

Why are unique frontal definitions required for each forecast system? Each system utilizes slightly different open boundary conditions and covers a somewhat different spatial domain; furthermore, while all the models assimilate satellite SSH, each uses a different methodology. Hence, average SSH in the Gulf differs amongst the models, and unique reference levels must be utilized for each model. Temperature at 200m is somewhat more constrained amongst the models since it is a property of the water masses and is directly related to the fluid buoyancy. Nonetheless, the particular location of the temperature-derived front is a sensitive function of temperature, and it was decided to use a unique temperature to obtain best agreement with HMI-EW for each model. Other dynamics-based definitions of the frontal location are possible, but the SSH- and temperature based approaches were selected due to their simplicity of implementation and for comparison with previous work.

Step 1: Definition of the Front

It was not known whether SSH-based or subsurface-temperature-based fronts would provide a reliable comparison to HMI-EW fronts, so both were used. Time dependence of the intercomparison was investigatd by looking separately at June 2010, December 2010, and all of 2010. The stability of the frontal definitions was also examined by repeating the analysis in two domains, Domain A and Domain B, which used different western boundaries for the analysis domain.

The Figures at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/June2010Fronts/ show the properties of both SSH-based and subsurface-temperature-based fronts for the suite of GOMEX-PPP models for June 2010 (Domain B). The layout of each page of figures is the same:

(1) The left panels show attributes of SSH-based fronts, while the right panels show attributes of T200-based fronts.
(2) The top row shows the current speed as a function of SSH (left) or T200 (right), with surface speed being shown by the blue line, and near-surface speed (average speed of the top 200m) shown by the red line.
(3) The middle row shows the distance between the HMI-EW and model-derived front. Heavy blue line is the time-average, and thin blue lines are +/- one standard deviation. In some plots the standard deviation lines are off-scale.
(4) The bottom row shows $\langle E_n \rangle$, the average CPA error to seven selected sites. Heavy blue line is the time average, and thin blue lines are +/- one standard deviation.

Figures at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Dec2010Fronts/ show the comparison based on December, 2010, of Step 1. And figures at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/2010Fronts/ show the comparison based on the entire year of Step 1. Table 12 summarizes the SSH-based attributes of the fronts, including the specific values of SSH which optimize the Dmin and $\langle E_n \rangle$ metrics. Table 13 presents similar information for attributes of the T200-based fronts.

Table 12: SSH-Based Front Definitions and Attributes

DOMAIN B [Small domain + >500m depth + CPA defined wrt seven sites]

Model	SSH_Dmin [cm] JUN / DEC	Dmin [km] JUN / DEC	SSH_<E> [cm] JUN / DEC	Max U_SFC [m/s] JUN / DEC
#1	-2 / -14	36 / 60	-6 / -12	1.2 / 0.7
#2	-2 / 10	32 / 44	-12 /902	0.7 / 0.7
#3	10 / 10	40 / 46	4 / 8	0.9 / 0.5
#4	-2 / -7	32 / 38	-10 / -7	1.0 / 0.8
#5	2 / -7	30 / 52	-2 / -7	1.0 / 0.6
#6	10 / 20	55 / 87	>20 / >20 o.s.	1.8 / 1.8
#8	2 / -10	40 / 52	14 / -14	1.1 / 0.8

o.s. indicates "off scale"

Table 13: T200-Based Front Definitions and Attributes

DOMAIN B [Small domain + >500m depth + CPA defined wrt seven sites]

Model	T200_Dmin [cm] JUN / DEC	Dmin [km] JUN / DEC	T200_<E> [cm] JUN / DEC	Max U_SFC [m/s] JUN / DEC
#1	16.4 / 16.6	38 / 51	16.3 / 17.4	1.2 / 0.7
#2	17.2 / 17.5	29 / 30	17.2 / 17.6	0.7 / 0.7
#3	18.2 / o.s.	56 / o.s.	18.9 / o.s.	0.5 / 0.3
#4	17.4 / 17.7	50 / 50	16.8 / 17.5	1.0 / 0.8
#5	17.2 / 18.0	32 / 54	17.5 / 18.1	1.0 / 0.6
#6	19.0 / 17.8	54 / 88	>20 / >20 o.s.	1.8 / 1.7
#8	16.6 / 17.0	38 / 64	18.2 / 16.6	1.1 / 0.8

o.s. indicates "off scale"

Observations:
(1) The frontal jet is tilted: the peak subsurface speed occurs at higher SSH and temperature than the peak surface speed. In other words, the modeled fronts are significantly baroclinic on average, as observed.
(2) In almost every case the best frontal definition based on CPA is further inside the Gulf than the best frontal definition based on agreement with HMI-EW. In other words, the SSH level with minimum $<E_n>$ is less than the SSH level closest to HMI-EW.
(3) The surface speed of Model #6 is nearly twice the surface speed of the other models.
(4) The thermal structure of Model #3 is anomalous and the surface LC/LCE front does not occur within the same range of subsurface temperatures as the other models.
(5) Based on the June 2010 comparison, the optimized SSH-based frontal definition is within 40km of the HMI-EW front for all models, except Model #6 which has larger error.
(6) Based on the June 2010 comparison, the optimized T200-based frontal definition is within 40km of the HMI-EW front for models #1, #2, #5, and #8. Models #2 and #5 are within 32km of the HMI-EW front.

Robustness of Frontal Definitions
Robustness is defined here as the stability or lack of sensitivity of frontal attributes to confounding or incidental factors. Because of the importance given to LC/LCE forecasts, these are a key element of the GOMEX-PPP multi-model comparisons, and it is highly desirable to have a robust frontal metric for the intercomparisons. In this section the sensitivity of frontal definitions are examined with respect to the time period of the comparison, the domain of comparison, the defining attribute of the front (SSH vs. T200), and the optimization criteria of the attribute (Dmin vs. $<E_n>$).

Consider first the observation that *the relationship between SSH or subsurface-temperature and surface speed is time dependent.* Comparison of the frontal analysis plots in June vs. December (http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/June2010Fronts/ vs. http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Dec2010Fronts/) shows that surface and near-surface speeds for June are higher and more peaked than the December and annual comparisons. Thus, the relationship between the jet-like characteristics of the front and the proposed frontal definition (contours of SSH or T200) is time dependent.

The time dependence of frontal characteristics also manifests as a time-dependent trend in the agreement of the model-derived vs. HMI-EW fronts. For example, Figure 7 shows a time series of $<E_n>$ when the frontal location is based on SSH optimized to minimize $<E_n>$ during June, 2010. Errors are relatively small and negative from February through May (days 40 to 150). Error increases and grows from June through October (days 150 to 300), and remains elevated in most models. Model #1 shows particularly unstable results after day 300, when the error jumps between -100km and -300km; this is due to the presence of small eddies (defined by closed contours of the given SSH criterion) near the edge of the analysis domain which, when present, greatly increase $<E_n>$.

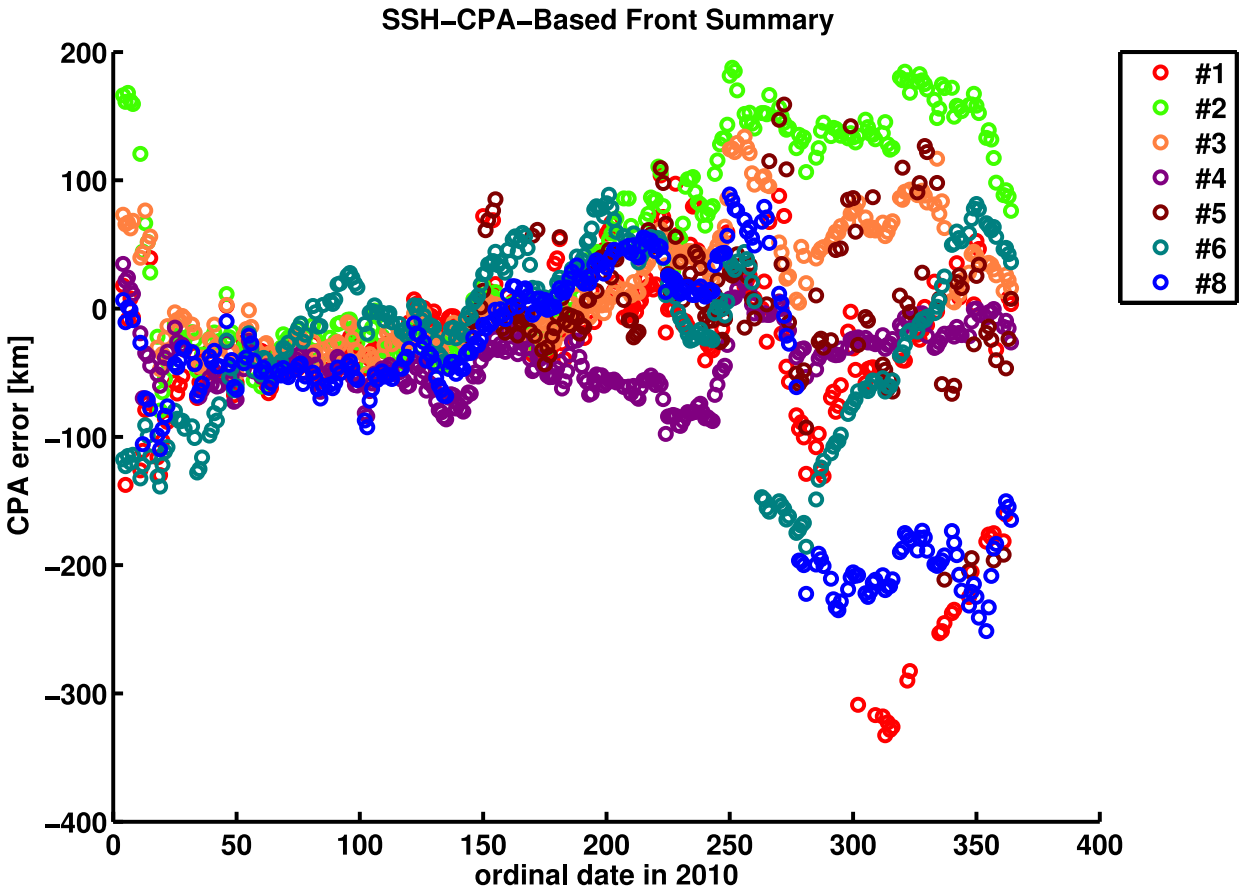


Figure 7: SSH-CPA Based Front Time Series, Step 1.
The SSH-based front used here is that which optimizes <E_n> metric, i.e., it is the front which minimized the CPA error compared with HMI-EW. The frontal metric becomes unstable after day 250.

If the frontal criterion is instead chosen to minimize Dmin, the same general trend is evident (see Figure 8). The front in Models #3 and #4 is stably estimated; however, error in the frontal definition for other fronts varies strongly throughout the year.

Another consideration with respect to timing concerns how to choose the optimal frontal definition. Both error metrics, Dmin and <E_n>, are more sensitive to the frontal definition in December than in June. In other words, the penalty for choosing a suboptimal frontal definition is much worse in December, regardless of whether the front is based on SSH or T200, and whether it optimizes the Dmin or the CPA metric.

Furthermore, the optimal frontal location based on T200 is less variable, between June and December, than the optimal frontal location based on SSH. In other words, the T200-based front appears to be a more robust indicator of frontal location than the SSH-based front. This result makes sense because the frontal velocity field is strongly-sheared and geostrophic; hence, via the

thermal wind relation, there should be a strong correspondence between thermal fronts and velocity jets.

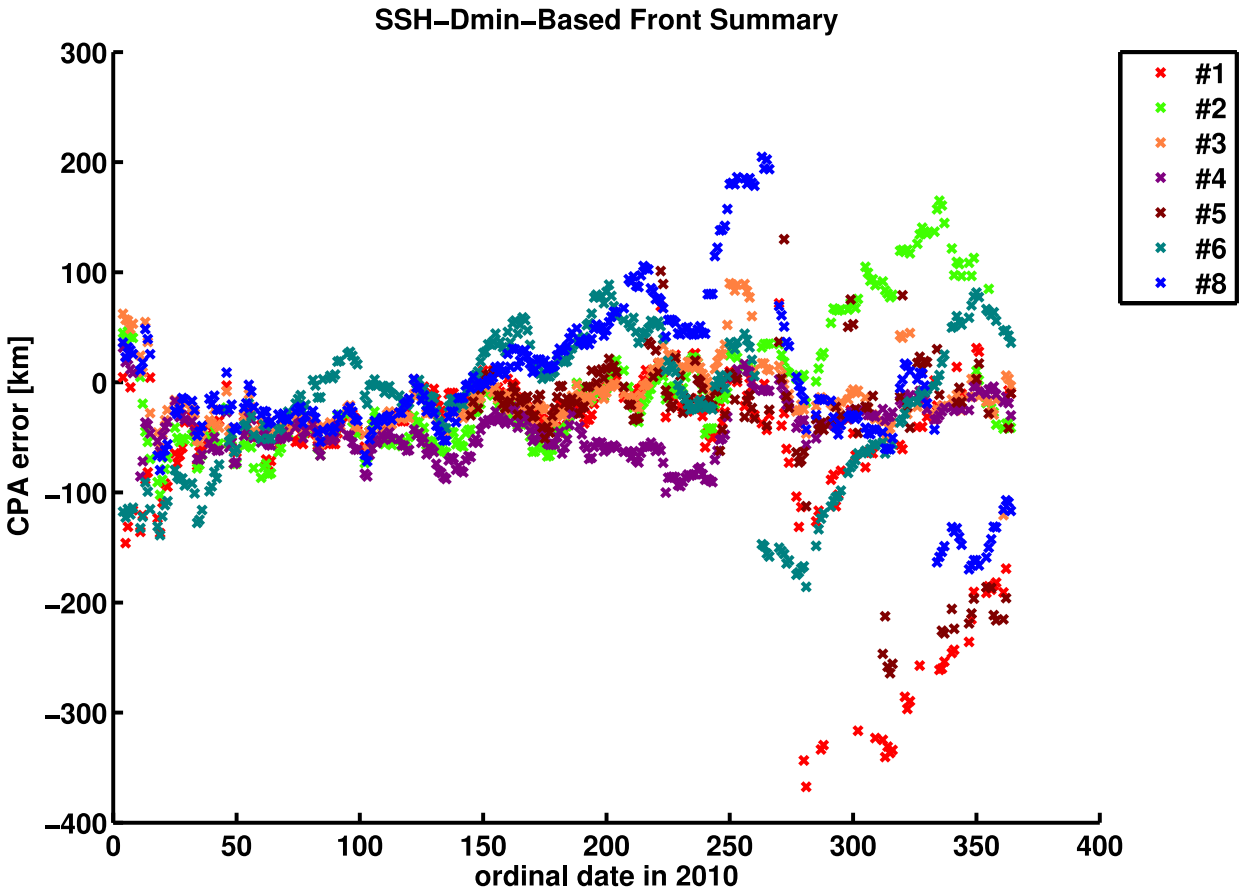


Figure 8: SSH-Dmin Based Front Time Series, Step 1.
The SSH-based front used here is that which minimizes Dmin, i.e., it is the front closest to the HMI-EW front. The frontal metric becomes unstable after day 250.

Due to the relative constancy of deep water properties T200 is less variable than SSH, the latter being influenced by basin-wide mass balance and barotropic processes unrelated to the LC/LCE.

Although the December, 2010, T200-based frontal estimator is arguably more robust than the SSH-based estimator, T200 cannot be used for making all-inclusive model intercomparisons because the thermal structure of Model #3 is anomalous, and no T200-based frontal definition could be identified. Thus, in order to include Model #3 in the frontal intercomparisons for Step 2, it was decided to use the SSH-based estimators for the fronts. Furthermore, since agreement between model-HMI fronts is best in June, the SSH-based front is determined by optimization with respect to the June 2010 hindcast.

Step 1: Timing of Separation Events

The HMI web site states that LCE Franklin separated in May 2010 (http://www.horizonmarine.com/loop_current_eddies.php). Analysis of HMI-EW frontal contours (proprietary data provided by C. Cooper) shows that the separation is not well-described as a single event, but it may have occurred through a series of attachment and detachment events, as has been suggested in previous model studies (Sturges et al, 1993). Independent satellite SST images are consistent and indicate that precise timing cannot be determined due to cloudiness and progressively-degraded signal-to-noise ratio in the GOMEX as the warm season progressed. Figure 9 shows 8-day composite images for May and June, 2010. These images have been reprocessed to enhance contrast, and the color scale is different between the images (original data are from the Level 3 browser at http://oceancolor.gsfc.nasa.gov/). Representative snapshots using a quantitative color scale are shown in Figure 10 during May and June using data from NOAA GOES satellites (http://marine.rutgers.edu/mrs/sat_data/).

Based on the optimized frontal location metrics in Table 12, we can define LCE separation dates for each of the models. The SSH-based frontal location is used because the subsurface-temperature-based frontal location cannot be defined for Model #3. From Table 14, the separation date ranges from May 13 to June 18, depending on whether the "event" is defined by SSH_Dmin (SSH level closest to HMI-EW) or SSH_CPA (SSH level with average CPA, $\langle E_n \rangle$, closest to HMI-EW), with most dates falling in the latter half of May. The timings based on SSH_Dmin and SSH_CPA are in good agreement. However, the timing of the LCE separation "event" in Model #2 and Model #8 is sensitive to the precise definition of the frontal criterion.

The following animations of frontal location are offered:
(1) All SSH_CPA-based fronts on a single animation:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_animate_SSHcpa_Step1/allfronts.mov
(2) All SSH_Dmin-based fronts on a single animation:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_animate_SSHdmin_Step1/allfronts.mov
(3) SSH_CPA-based fronts (solid contour) and T200_CPA-based fronts (dashed contour) for each model:
Model #1
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/iasnfs.mov
Model #2
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/profs.mov
Model #3
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/iasroms2.mov
Model #4
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/mitgom.mov
Model #5
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/amseas.mov
Model #6
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/ngom.mov
Model #8
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/iasroms.mov

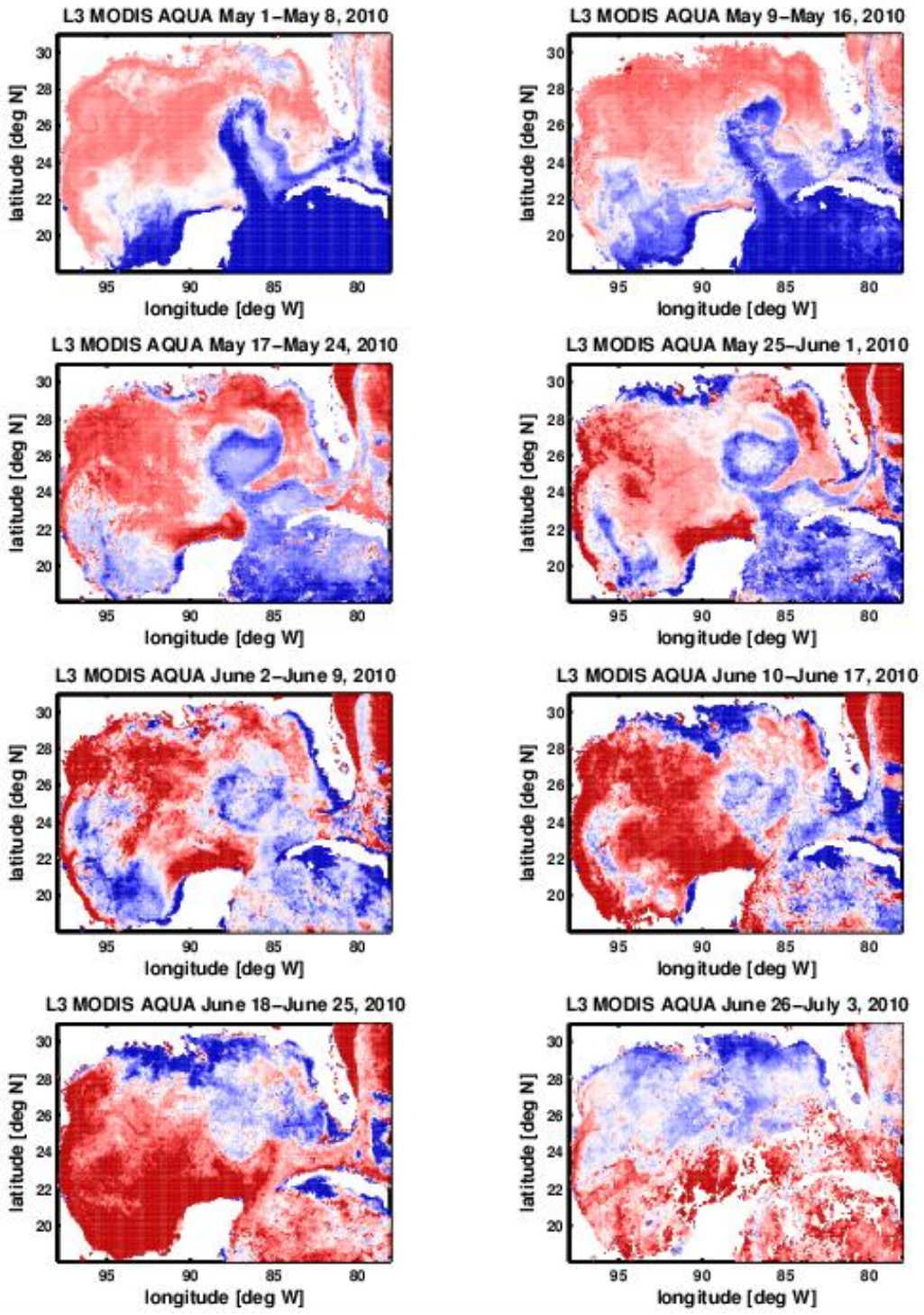


Figure 9: L3 MODIS-AQUA SST, Step 1.
The color scale of each panel is modified to enhance LC/LCE contrast and is not quantitative.

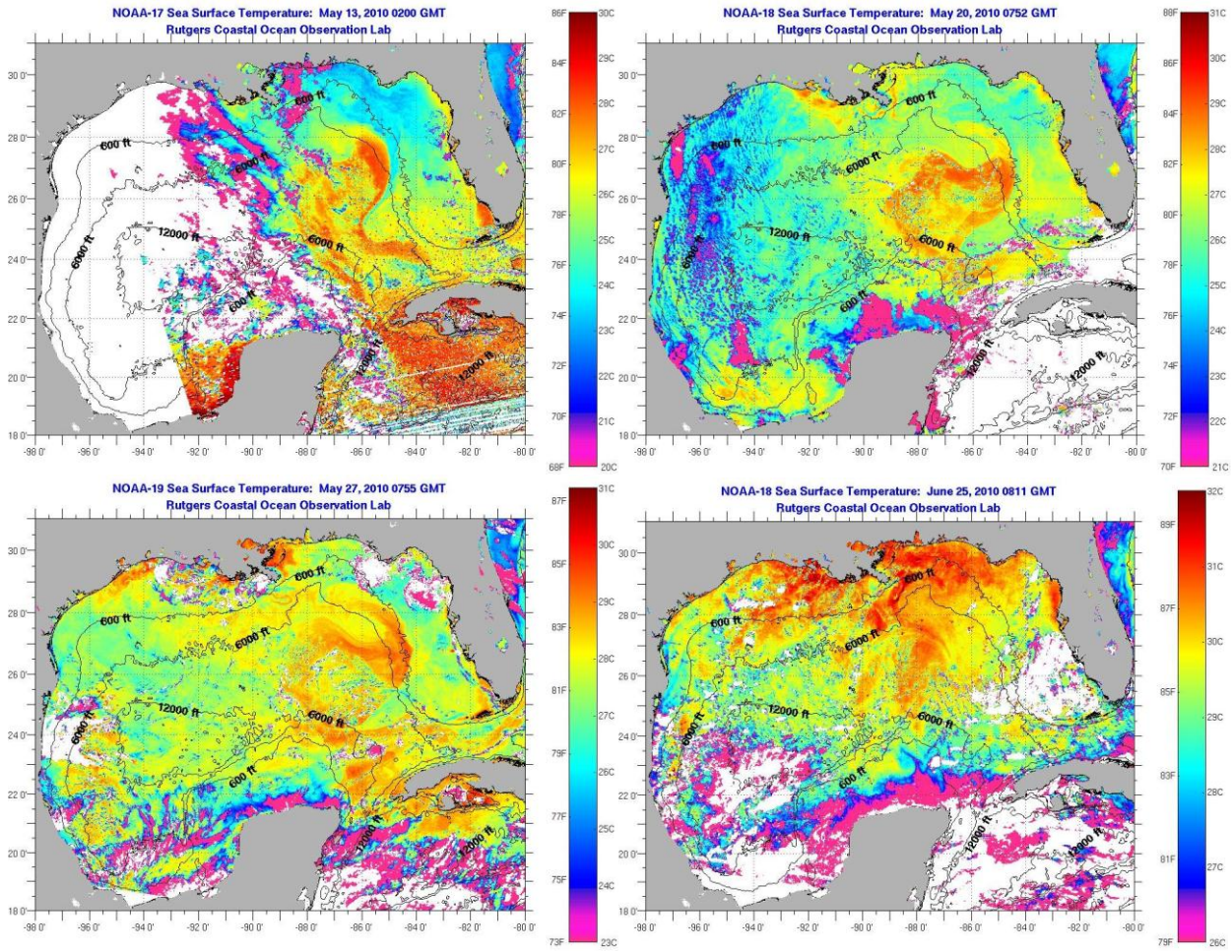


Figure 10: Rutgers-NOAA SST.

Table 14: LCE Separation Events, Step 1
[Based on animations in
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/
and in http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_cpa_Step1rev/]

Model	SSH_Dmin-based	SSH_CPA-based	Notes
#1	May 14	May 15	
#2	Jun 20	Jun 18	Ephemeral. Sensitive to precise definition of front.
#3	May 25	May 25	
#4	May 27	May 31	
#5	n/a		Separated prior to May 25
#6	May 18	May 18	Reattached May 29-July 19
#8	Jun 7	Jun 5	Ephemeral near-separation. Sensitive to precise definition of front. Reattached June 9-August 13

(4) SSH_Dmin-based fronts (solid contour) and T200_Dmin-based fronts (dashed contour) for each model:
Model #1
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/iasnfs.mov
Model #2
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/profs.mov
Model #3
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/iasroms2.mov
Model #4
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/mitgom.mov
Model #5
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/amseas.mov
Model #6
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/ngom.mov
Model #8
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_dmin_Step1rev/iasroms.mov

Step 2: Timing of Separation Events
As noted above, the Step 1 model runs result in a range of dates for the LCE separation event in 2010. In order to assess the predictability of the separation events. The Step 1, retrospective nowcast, separation date is compared with the forecast separation date, using the 3-mo. forecasts from Step 2. Because the models were initialized monthly and integrated forward for three months, there are three separate forecasts to compare with a given Step 1 separation date.

The timing of the LCE separation events was identified from animations of the frontal location in each model. The animations are found in the following directories:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_S2-01/

through
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_S2-12/

The animations display frontal contours defined by the June 2010 SSH-Dmin (solid line) and T200-Dmin (dashed) criteria. The colormap displays SSH with the color scale centered on the frontal definition. Thus, the animations of different models are comparable in spite of differing mean SSH.

Table 15 summarizes the LCE separation dates as inferred from the June 2010 SSH-Dmin-based frontal definitions. The hindcast LCE separation date occurs in May and June in the models. Thus, Step 1 is compared with Step 2 forecasts S2-03, S2-04, S2-05, and S2-06.

No models forecast a LCE separation event in S2-03. Models #3, #8, and #9 did not forecast LCE separations. Model #2 forecast ephemeral LCE separations in S2-04 and S2-05. Only Model #6 stably forecast the LCE separation event with both 1mo. and 2mo. lead time (S2-04 and S2-05).

Step 2: Front Predictability

Front predictability is assessed using the methodology of Oey et al (2005). Each 3-mo. forecast is compared with a persistence forecast, the latter being defined by persistence of the first HMI-EW frontal analysis in the forecast window. Error with respect to HMI-EW is defined as in Oey et al (2005), where $E_n = CPA(Model)_n - CPA(HMI\text{-}EW)_n$ is the CPA forecast error for site n, $\langle E_n \rangle$ is the average across all seven sites, $rms(E_n) = \langle E_n^2 \rangle^{(1/2)}$ is the root-mean-square of E_n, $P_n = CPA(HMI\text{-}EW\ initial)_n - CPA(HMI\text{-}EW)$ is the persistence error, and $\langle P_n \rangle$ and $rms(P_n)$ are defined as for E_n.

Figure 11 shows an example of these statistics for Model #1 during the S2-04 forecast. The persistence error <P_n> (green x) and rms(P_n) (black x) are both initially zero, rise to about 60km at day 30, and continue to rise to more than 100km at day 90. In contrast, the model error <E_n> (green o) and rms(E_n) (black o) are both initially near 40km; the mean error, <E_n>, varies between -30km and 70km over the forecast, while the rms error, rms(E_n), varies from 20km to 70km. In this forecast the model beats persistence after approximately 20 days.

Table 15: LCE Separation Events, Step 2
[Information in this table is based on the animations at
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_S2-01/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Fronts_timing_S2-02/
etc.]
Fronts are defined by revised June 2010 SSH- and T200-based definitions optimized for the minimum Dmin metric.

Model	2010 Hindcast	S2-3	S2-4	S2-5	S2-6	Notes
#1	May 14	none	none	(May 18)	<Jun 1	
#2	Jun 20	none	(Jun 16)	(May 16)	Jun 6	Very sensitive to precise definition of frontal SSH.
#3	May 25	none	none	none	<Jun 6	
#4	May 27	May 23	n/a	(May 31)	n/a	
#6	May 18	none	May 21	May 19	(July 4)	Reattachment June 4
#8	Jun 7	none	none	none	none	Very sensitive to precise definition of frontal SSH.
#9		none	none	none	none	

Dates in parenthesis indicate a detachment/reattachment event.
Model #4 results for S2-4 and S2-6 are not forecasts and not included.

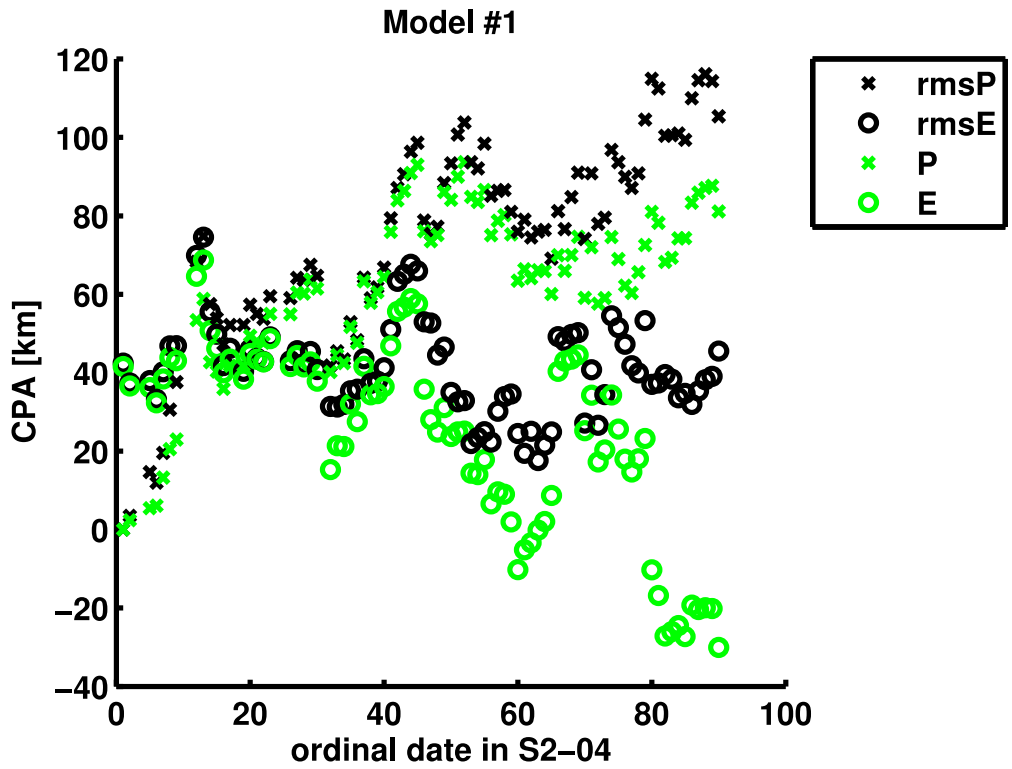


Figure 11: Example Front CPA Metrics, Step 2-04.

Figure 12 shows the errors during the S2-07 forecast for the same model (note change of ordinate scale). In this case, the model performs somewhat worse than persistence. Figure 13 shows rms(P_n) (black) and rms(E_n) (red) for the entirety of 2010. During the first half of the year the model consistently beats persistence, but during the second half of the year, more often than not, persistence beats the model. In Figure 14, the twelve Step 2 forecast statistics are summarized as a function of forecast time. Individual forecasts are shown by the thin lines, and Step 2 averages are shown by the thick lines. Results are noisy, but, on average, Model #1 beats persistence after about 60 days.

A comprehensive summary of the model results is shown in Figure 15, which shows rms(E_n) for each model (colors) and rms(P_n) (black), both quantities averaged within 7-day windows. Initial model errors range from 40km to 100km, with subsequent errors growing less slowly than the persistence error. An examination of the HMI-EW over the entire year found that the maximum persistence error was approximately 180km.

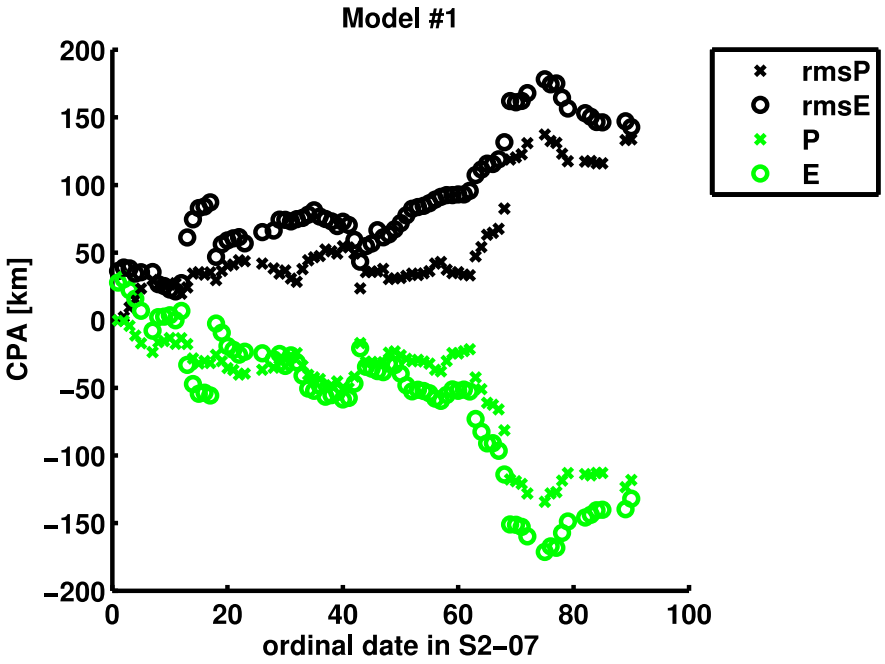


Figure 12: Example Front CPA Metrics, Step 2-07.

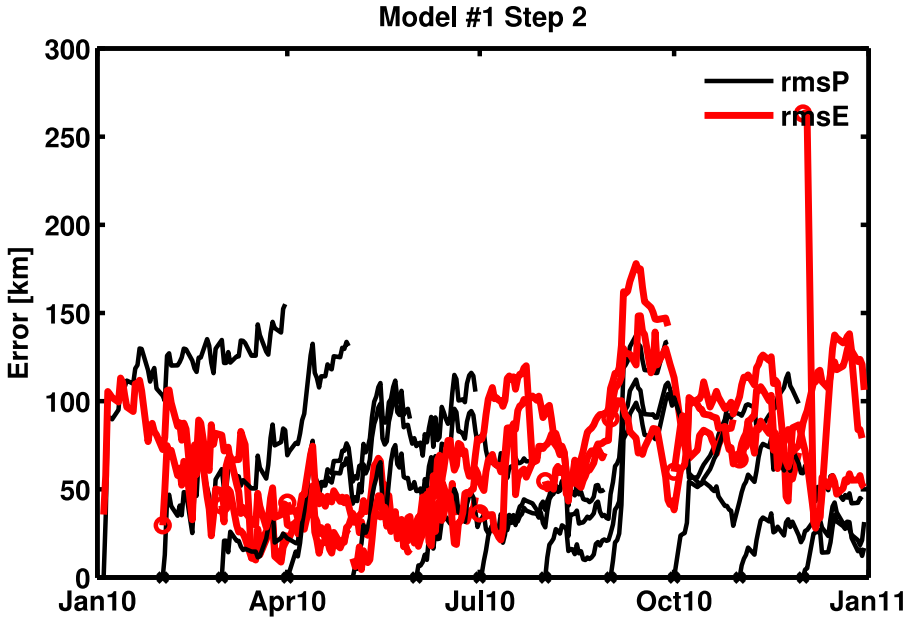


Figure 13: Example Front CPA Metrics, Step 2.

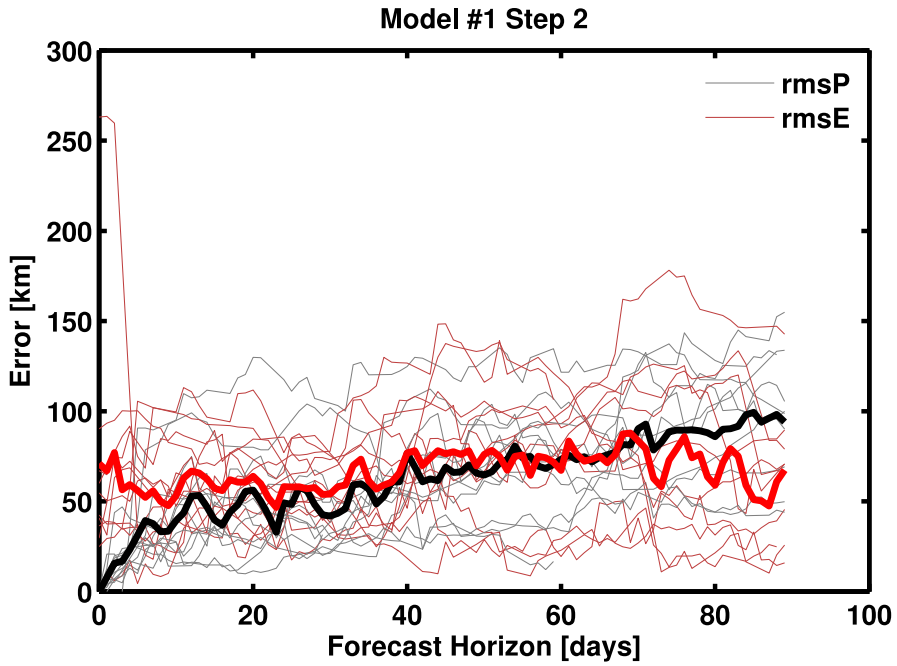


Figure 14: Example Front CPA Metrics, by Forecast Lead-Time, Step 2.

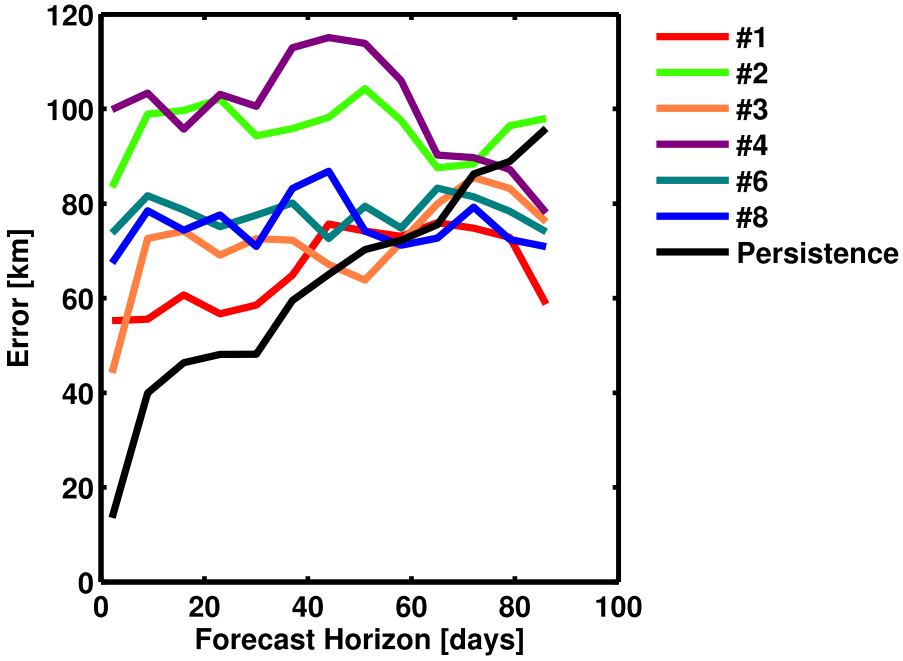


Figure 15: SSH-Based Front CPA Error Summary, Step 2.

The foregoing analysis of frontal forecasts used the SSH-based criterion to optimize Dmin, based on the June 2010 hindcast (Table 12). The previous discussion of time-variability in the frontal definitions highlighted possible benefits of an alternative definition based on December 2010 T200.

Figure 16 shows the error as a function of forecast time for the December 2010 T200 frontal definition. Results differ somewhat from Figure 15 in that model errors are more clustered, except for Model #3. Apparently, the nature of the thermal anomaly in this model is a drift or incorrect initialization.

The results summarized in Figures 16 may be directly compared with the comparison in Oey et al (2005). The persistence forecast error in the present study grew somewhat more slowly in Step 2 than in the study of Oey et al. (2005). Initial forecast errors were similar (for Model #2); however, the forecast error grew more rapidly in Step 2 than in Oey et al. (2005).

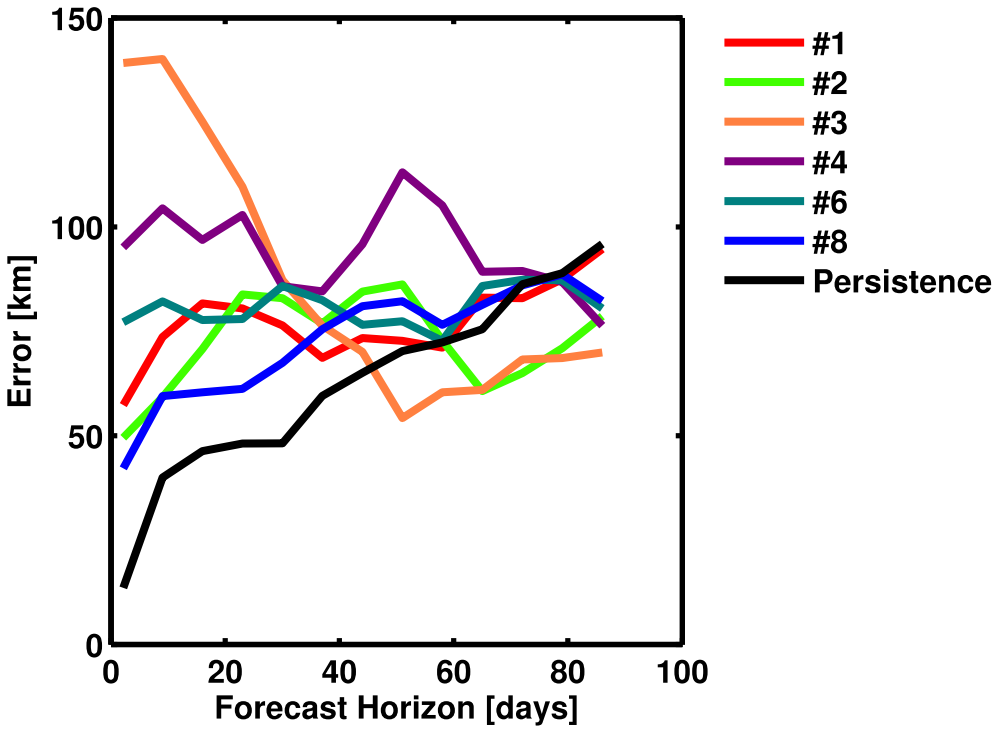


Figure 16: T200-Based Front CPA Error Summary, Step 2.

Step 3: Front Animations

Because of uncertainty in the robustness of the objectively-defined frontal location in the fall and winter (Figure 7), a quantitative assessment of predictability was not made in the Step 3

forecasts. A qualitative assessment of frontal behavior may be obtained from animations of frontal positions, which are found at the following URLS:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-1/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-2/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-3/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-4/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-5/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-6/
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/FrontsS3-7/

Methodology #2: Subjective Approach

The "objective approach," described previously, sometimes finds more fronts than desired even when the search is constrained to a certain geographical sub-region. Consequently, a "subjective approach" was also employed, in which a human would locate and trace the fronts by eye on a computer display of composited model fields. Color-filled contours of sea surface height combined with sea surface velocity vectors (the velocity field is not shown (Fig. 17) to make the figure more legible) were presented. Using a computer mouse, an operator traced the LC front and fit an ellipse to any LCEs. The LC tracings were clipped and smoothed using splines at 101 equidistant control points between the entrance and exit of the Gulf of Mexico. The spline curves were written to disk as latitude-longitude pairs and the elliptical parameters were written to separate disk files for later reconstruction and use.

As a typical example of the tracing result, red lines indicatethe LC and LCE as determined by the operator eye. The northwest portion of the LC tracing does not lie exactly over the SSH contours there. The reason is that the appearance of surface velocity vectors (not shown) in that gap influenced the operator causing he frontal location to be located a bit more to the northwest than if the SSH fields alone were considered. This is an example of where the subjective approach results differ from an automated approach. In this example one would be tempted to use a computer search; however, the LC in this figure is fairly simple. The task is decidedly harder when the LC is spooled deep into the Gulf and has indistinct margins. The yellow lines are the frontal locations as determined by HMI-EW. (The yellow lines in this figure are fictitious in accordance with the confidentiality agreement with HMI). HMI provided frontal locations snapshots for 265 days in 2010 in ESRI shape file formats. They were converted them to ASCII and with a little bit of manipulation objects were tagged as LC or LCE. During the tracing process, the operator did not see the HMI fronts so as not to influence the results. The tracing process was tedious and took about 3 hours to trace 365 days of model output. Eight models were processed and sometimes reprocessed if the results did not look right upon review. A great time savings can be gained by knowing which days had HMI-EW frontal analyses available and then only tracing model output for those days, a 28% reduction in the number of fields to trace.

The next step in the analysis was to find the CPA of model frontal locations to the seven locations marked in the figure with a white X. These locations were used in Oey et al (2005). The CPAs of the HMI fronts were also computed. Each "X" in the figure has a line going to the closest point on a model front and to the closest point on an HMI front. Although the LC and LCE are separate features they were combined when determining the CPA. In this way, whichever feature, LC or LCE, is closest and presumably the most influential on "X" is used.

Figure 17 shows that for the eastern most "X" the model LC is closer than the model LCE. In this instance, for this station, a model LC is compared to a HMI LCE. Originally the comparisons were constrained to be LCE-LCE and LC-LC but this made little sense when the LCE had detached in the model and not according to HMI.

Step 1: Comparison with HMI-EW

The model-data metric used here is the error E_n (or bias) defined as the average value over the seven locations, $E_n = \langle(CPA_{Model} - CPA_{HMI})\rangle$, and root-mean-square of E_n. In some cases the location of the model and observed fronts are quite different. For this and other reasons, the comparisons make most sense if they start with the 09-Jun-2010 LCE separation event as determined by HMI and run to the end of the year.

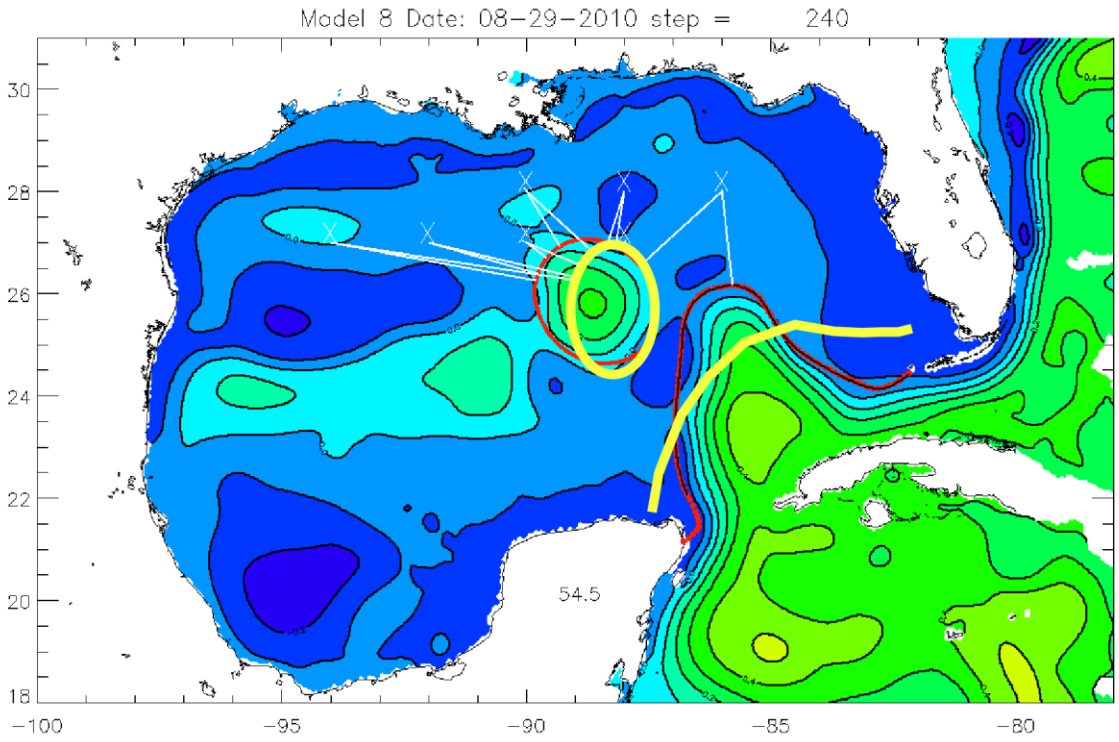


Figure 17: Determination of Front CPA, Subjective.
The CPA metric was computed in the subjective approach using the distance between the seven reference points (white 'X') and the model frontal locus (red line) and the HMI-EW frontal locus (yellow). The model frontal locus was determined by human operator from a subjective assessment of SSH contours (colorscale) and surface velocity vectors (not shown). Line segments indicating the CPA (white) were located by hand by the operator, as well. The processing of each (daily) image required about 1 minute of operator time.

The time series of E_n and RMS(E_n) results are examined for the eight models (Fig. 18). The time averages of E_n and RMS(E_n) are displayed and tabulated in Table 16. Model 6 does not do well primarily because the LCE drifted rapidly westward when the HMI-EW LCE lingered in the separation area. Model 7 had issues throughout the evolution and did not return for Steps 2 and 3. Model 3 had the best performance with a very small positive bias. The error of the method (e.g., placement of the trace or ellipse) may have a larger error than this. Models 1 and 5 had biases of similar magnitude but of different sign. It would be interesting to compare their assimilation data sets and their methods of assimilation. The differences in Model 3 and Model 8 are larger than had been expected.

Step 3: Timing of Separation Events

In this section forecast SSH fields are compared with remotely-observed SSH fields for timing of an eddy detachment and a reattachment event which nominally occurred in mid-November 2011.

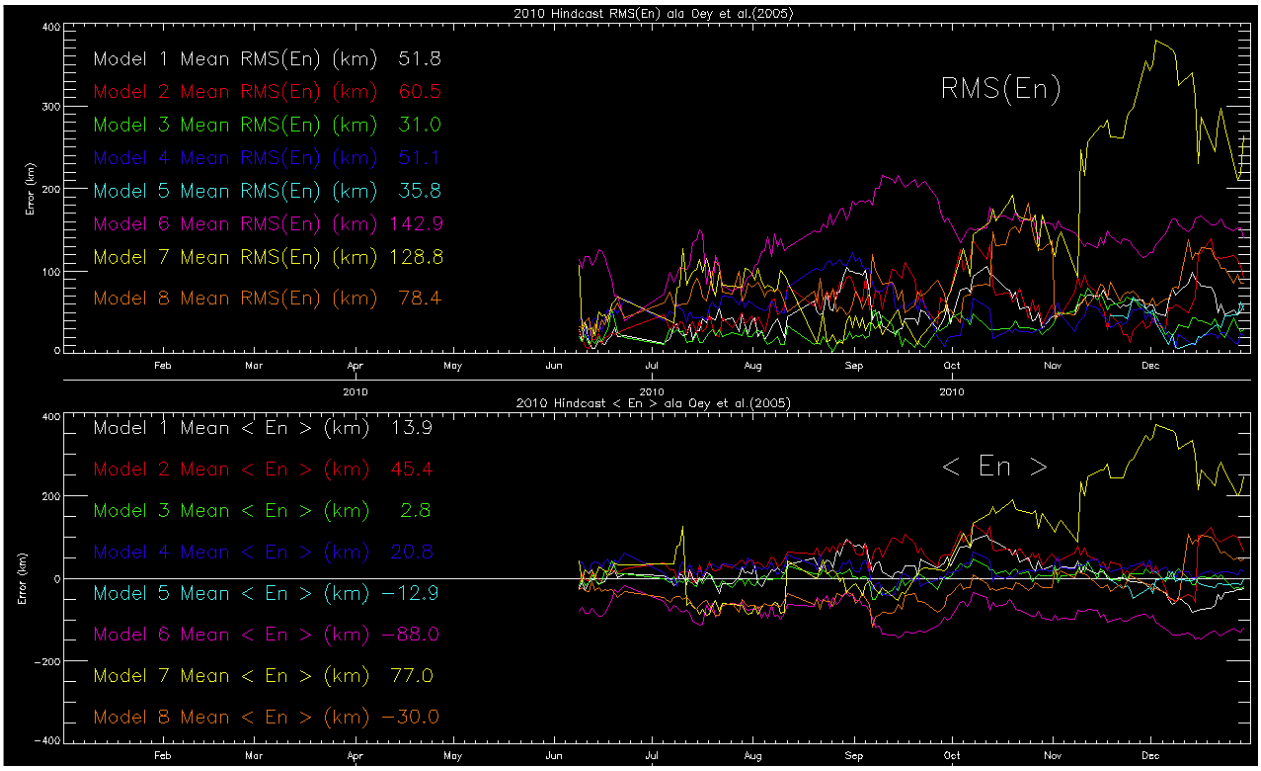


Figure 18: Subjective Fronts, CPA Time Series, Step 1.
Upper panel: RMS CPA error averaged over the seven reference sites, RMS(En). Lower panel: average CPA error, <En>. Because CPA error, En, is a signed quantity, cancellations occur, leading to small average errors in the lower panel.

Table 16: Front CPA Error Metrics -- Subjectively-Defined Fronts

	<En> [km]	RMS(En) [km]
Model 1	13.9	51.8
Model 2	45.4	60.5
Model 3	2.8	31.0
Model 4	20.8	51.1
Model 5	-12.9	35.8
Model 6	-88.0	142.9
Model 7	77.0	128.8
Model 8	-30.0	78.4
Mean	3.6	72.5

Results for Step 1 Loop Current and Loop Current Eddy Fronts (subjective). These are the bias and root-mean-square errors for the CPA for combined LC and LCE to the seven stations listed in Oey et al (2005).

Dr. Robert Leben at the Colorado Center for Astrodynamics Research (CCAR) provided daily SSH fields derived from blended satellite altimetry products to the public on a routine basis. This product has a 10-day averaging window which should be taken into consideration when using it to establish LC spatial extent and LCE separation events. These SSH fields were acquired from the CCAR web site for near Real-Time Altimetry, and they were visually compared with the model output from the Step 3 forecasts during the last quarter of 2011 (dates listed in Table 4).

The CCAR SSH fields are examined at fortnightly intervals (Fig. 19) from 15 SEP 2011 to 20 JAN 2012. They indicate a westward migration of a LCE and a possible detachment of a new LCE around 11 NOV 2011. This LCE appears reattached in the 25 NOV image. Examination of images (not shown) surrounding 11 NOV indicate no detachment on 9-10 NOV and substantial reattachment on 17 NOV. Images from 12-16 NOV look essentially the same as that of 11 NOV.

Animations of side-by-side comparisons of model SSH for the six models and seven forecast experiments are available on the GOMEX-PPP website under Results for Step 3. Model SSH fields on 11 NOV are shown in Figure 20.

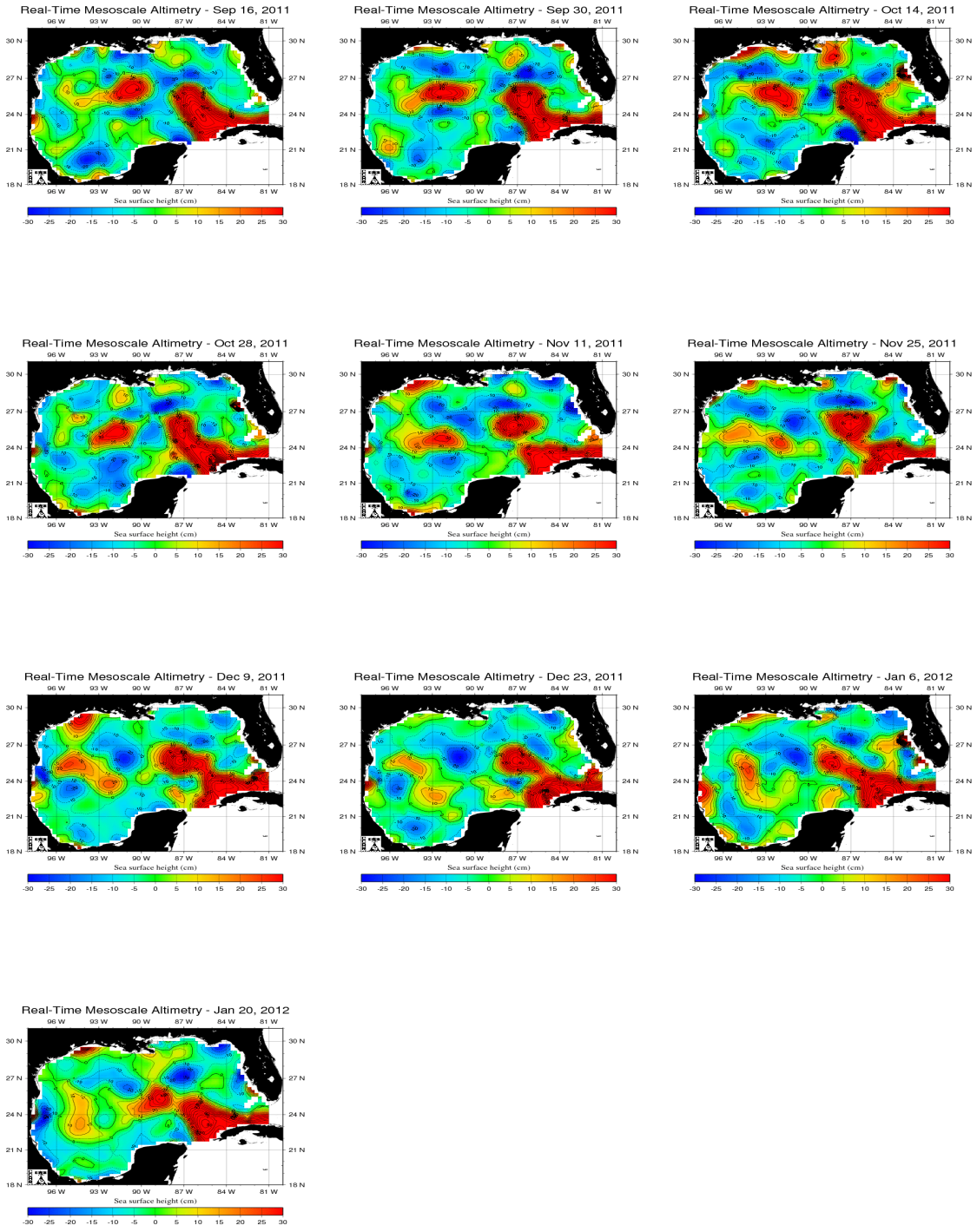


Figure 19: Real-Time SSH from CCAR.

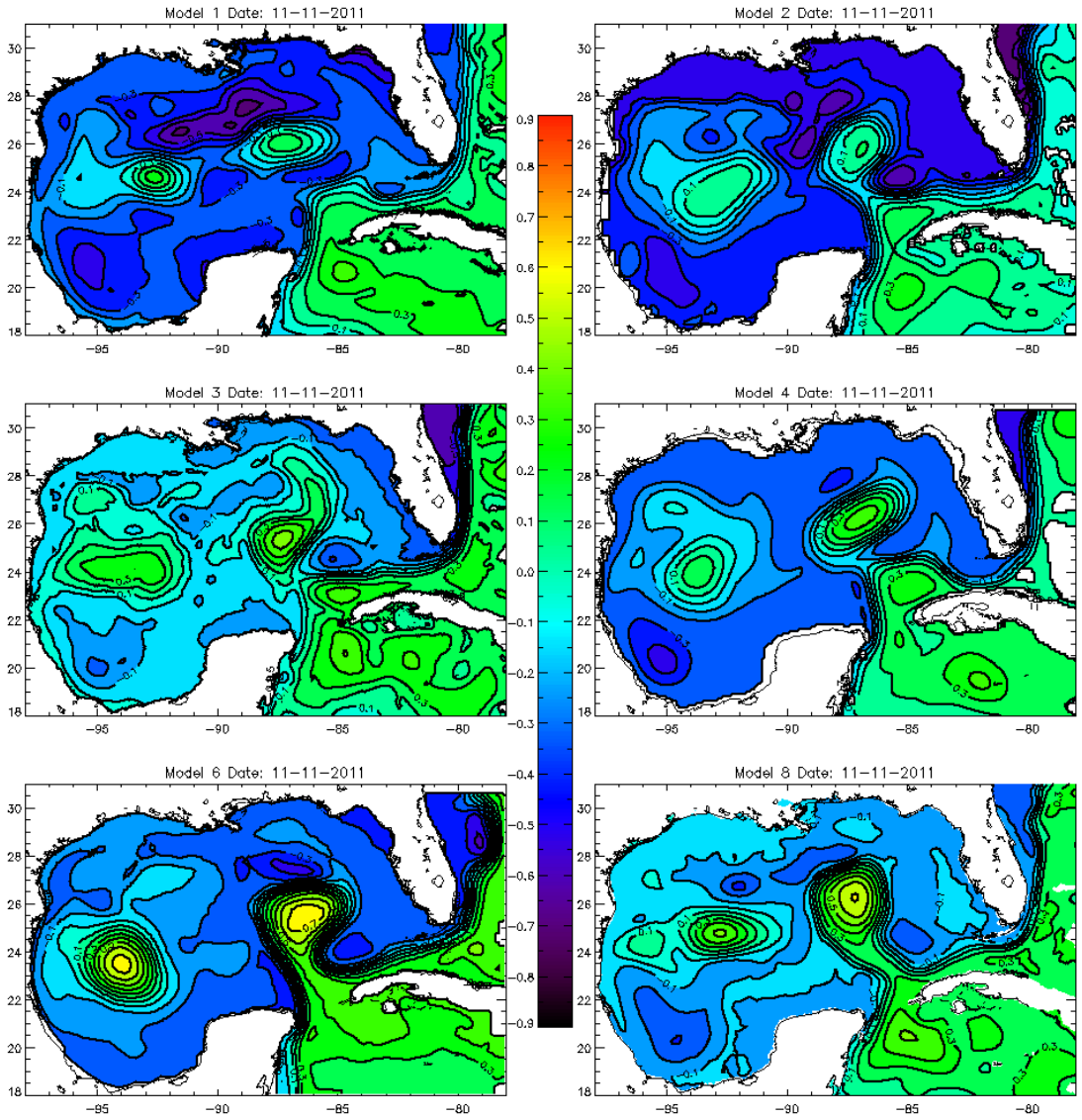


Figure 20: Nowcast SSH for November 1, 2011.

Separation dates estimated from the animations of model forecasts are summarized in Table 17, which gives the dates when the modeled eddy first appeared to be substantially detached. Model 1 gets the detachment date at the right time 8 weeks (S3-01) prior to the event and remains consistent as the forecast horizon goes to zero (S3-05). At six weeks prior (S3-02), Model 1 and Model 6 have the timing approximately correct. At four weeks prior (S3-03) Models 1, 2, 4, and 6 have the possible detachment event isolated to within a few days of the nominal date.

Table 17: LCE Separation Events – Subjective SSH-Based, Step 3
Green shading indicates separation date within 5 days of observed separation on November 11, 2011.

Model	S3-01 (8 wks)	S3-02 (6 wks)	S3-04 (4 wks)	S3-05 (2 wks)	S3-05 (0 wks)
#1	Nov 8	Nov 9	Nov 13	Nov 12	< Nov 11
#2	Oct 30	Dec 22	Nov 14	Nov 11	Nov 14
#3	Nov 12	Nov 27	Nov 20	Nov 10	< Nov 11
#4	Dec 5	Dec 8	Nov 18	Nov 17	Nov 12
#6	N/A	Nov 19	Nov 18	Nov 18	Nov 17
#8	Oct 14	Oct 28	Oct 23	Nov 16	Nov 15

Table 18 summarizes the dates the model eddies reattached. In the CCAR SSH reattachment occurred on 17-November. Forecasting reattachment proved to be more challenging that forecasting detachment and in some cases the detached ring never reattached. Model 6 did the best job of forecasting reattachment. Model 1 did consistently well even 8 weeks prior to the event but did poorly in the most proximal run (S3-05).

Table 18: LCE Reattachment Events – Subjective SSH-Based, Step 3
Green shading indicates reattachment date within November 11-25 window. Based on CCAR SSH, reattachment occurs before Nov 25.

Model	S3-01 (8 wks)	S3-02 (6 wks)	S3-04 (4 wks)	S3-05 (2 wks)	S3-05 (0 wks)
#1	Nov 20	Dec 3	Nov 23	Nov 16	Jan 8
#2	Nov 13	Dec 30	Dec 19	Nov 22	Nov 20
#3	-	Dec 4	Dec 9	Dec 18	Dec 17
#4	-	Dec 27	-	-	Dec 4
#6	N/A	Nov 21	Nov 23	Nov 21	Nov 19
#8	Oct 24	Oct 31	-	Dec 7	Jan 5

Summary
The most significant finding of the frontal analysis is that the optimal definition of the model front from either SSH or T200 is time-varying. Both methods of frontal definition appear to be robust in late winter through early summer, but not reliable in late summer through early winter, at least in the single year of observations considered here. It is possible that the lack of robustness in the frontal definition is not a property of the models, but is instead a property of the HMI-EW comparison data. It may be that HMI-EW is not reliable during the late summer to early-winter time period. Future work should focus on quantitative development and validation of a non-proprietary frontal analysis. Such a product, even if it were available only in hindcast mode, would provide a useful benchmark for model validation.

Forecasting the timing of LCE separation events appears to be infeasible with even 1-mo. forecast horizon in the majority of models. Only Model #6 stably forecasts the LCE separation event with greater than 1-mo. lead time. However, Model #6 displays westward LCE propagation dissimilar to other models and to HMI-EW, and its frontal speeds are nearly twice the values seen in other models.

With the appropriate frontal definition, Models #2 and #5 outperform other models in hindcast mode, based on agreement with HMI-EW, and based on the June 2010 frontal attributes. In Model #6, the optimal frontal location is unambiguously defined by SSH, but the errors with respect to HMI-EW location are largest amongst the models. The thermal structure of Model #3 was anomalous, and no reliable T200-based frontal definition could be obtained.

GOMEX Sea-Surface Height Field

Introduction
Sea-surface height (SSH) is a dynamically significant manifestation of the pressure field within the ocean. The geostrophic component of near-surface current flows along contours of constant SSH, and SSH is approximately equal to the (scaled) streamfunction for the geostrophic flow. Thus, correctly forecasting SSH (and surface Ekman layer properties) is a prerequisite for accurate forecasting of transport in the upper ocean.

Measurements of the SSH field are available from tide gauge stations in the coastal ocean and at a few sites in deep water. Much of the deep water GOMEX is not monitored in situ, instead, the constellation of satellite altimeters is the source of SSH measurements.

Satellite altimeters operate by precisely measuring the travel time of electromagnetic pulses between the ocean surface and the satellite antenna. Interpretation of these travel time measurements as SSH requires carefully accounting for the index of refraction of the atmosphere (the wet- and dry-troposphere corrections and the ionospheric correction), the scattering of the RADAR pulse off the ocean surface (the sea-state bias correction), and the satellite orbit and geoid height relative to a geocentric reference frame (Fu and Cazenave, 2001). Furthermore, in the present application, the GOMEX-PPP models forecast the subtidal (25-hour average) SSH, so it is necessary to remove the diurnal and semi-diurnal tidal signals from the altimeter-derived SSH.

Quantitative assessment of the GOMEX-PPP models has been performed using along-track data from the Jason-1 and Jason-2 satellite altimeters, with raw data and corrections extracted from the Radar Altimeter Database System (RADS; Naeije et al, 2002). The AVISO absolute dynamic topography product, which is a mapping of multiple-satellite data by Gauss-Markov smoothing in space and in time (SSALTO/DUACS 2011), has also been used.

The comparison of satellite-derived SSH with modeled SSH is complicated by errors in the geoid and mean-sea-surface near the LC. The geoid is one surface amongst the family of hypothetical time-mean gravitational equipotential surfaces which would coincide with the ocean surface if the latter were completely at rest. The geoid is computed from observations and models of the Earth's gravity field, and it is uncertain at scales less than 100km (Kenyon et al, 2007, mention 15cm error level; Pavlis et al, 2004 present an estimated error map with significant structure in the Gulf). Thus, in practice, altimeter data products do not utilize the geoid directly, using instead an estimate of the mean sea-surface obtained from long time series of satellite altimeter measurements combined with in situ data. For example, the AVISO absolute dynamic topography product is constructed by Gauss-Markov smoothing of SSH anomalies from multiple-satellites with respect to independent estimates of the time-mean SSH (SSALTO/DUACS, 2011). A geoid model, hydrographic data, and archived SSH data are utilized to estimate the mean dynamic topography (MDT, the mean-sea-surface relative to the geoid) which is then added back to the gridded SSH anomaly fields (Rio and Hernandez, 2003).

In an area like the GOMEX, with a large, strong time-mean current (the LC), the geoid and the mean sea-surface are very different, and an accurate estimate of the mean dynamic topography is needed to interpret absolute dynamic topography. Based on the latest estimate of the mean dynamic topography used in AVISO products (Rio et al, 2011) in the GOMEX (Fig. 21). One particular configuration of the LC is represented in the time-average, and the SSH anomaly associated with the LC is about 40cm. This anomaly is similar to what is seen in the instantaneous LC in the models (except Model #6, which has a jump of almost 1m), and it is surprising that the MDT does not show a smoother rendering of the LC.

In previous studies, the mean-sea-surface of the model and the SSH data have been removed independently (Leben, 2005); however, this is not possible in the present case since only one year of data from each GOMEX-PPP model is available. [Some of the more mature systems have been archived for several years, and it may be possible to compute their time-average SSH.] In principle, one could compute the time-average SSH from the models and the data using just the modeled time period. But this approach is inappropriate considering the range of time scales of variability in the GOMEX, and would likely contribute to error due to the small number of degrees of freedom in the average.

Present efforts have utilized two approaches. In the first approach, the observed and modeled SSH are compared by computing anomalies with respect to a 1st-order polynomial fit of along-track SSH within the GOMEX. This approach is suitable for removing long wavelength geoid error or orbit errors from the absolute dynamic topography. The second approach utilized collinear SSH differences (equivalently, sea surface slope), which highlights the short wavelength components of the SSH field. Both approaches are explained in detail, below.

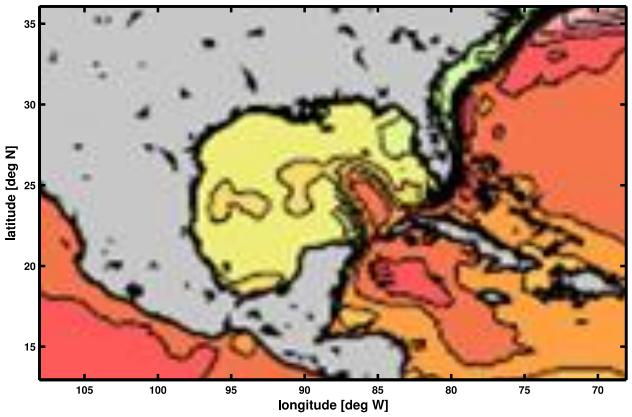


Figure 21: AVISO Mean Dynamic Topography.

Methodology

The AVISO Delayed-Time Merged Absolute Dynamic Topography (DT MADT) product was downloaded from the AVISO website, http://www.aviso.oceanobs.com/index.php?id=1271. This product is produced by Gauss-Markov smoothing of the SSH anomaly from multiple satellite altimeters (Jason-1, Jason-2, ERS-2, and Envisat), with care taken to ensure a consistent multi-satellite, multi-mission, cross-calibration (SSALTO/DUACS, 2010), and a separately-computed mean-sea-surface is added to produce absolute dynamic topography (i.e., SSH relative to the geoid). The native-resolution product (approximate 1/3-degree horizontal resolution on a Mercator grid, 7-day time resolution) was bilinearly interpolated onto the 1/20-degree daily-resolution GOMEX-PPP analysis grid, keeping the 7-day resolution.

The along-track Jason-1 and Jason-2 altimeter data were extracted from the Radar Altimetry Database System (RADS; Naeije et al, 2002) and standard corrections were applied for the orbit altitude, dry troposphere, wet troposphere, ionosphere, inverse barometer, solid earth tide, ocean tide, load tide, and sea-state. Some experiments with the mean-sea-surface and geoid corrections were performed, and it was found that the best agreement between AVISO and single-satellite along-track data was obtained using the EGM2008 geoid, rather than the default mean-sea-surface correction (CLS01). GOMEX-PPP models differ on whether the atmospheric pressure is included in the surface boundary conditions but the impact of this correction was found to be negligible in 2010; however, it may be important to reconsider this in future studies of GOMEX response to tropical cyclones or validation of coastal sea level forecasts.

Comparisons with Jason-1/2 along-track data utilize model output linearly interpolated to the satellite ground-tracks at the times and locations of the satellite passes. Gaps in the satellite data are not filled. Satellite ground tracks passing through the region (90°W,22°N) to (82°W,30°N) are used, see Figure 22. To exclude shelf variability, comparisons are performed only at sites deeper than 500m.

Comparisons of both SSH and collinear SSH slope were performed. The latter is directly related to the surface geostrophic velocity component perpendicular to the satellite ground track, and it

was computed using an optimal 9-point stencil with a half-power point of approximately 64km (Powell and Leben, 2004). Using this methodology, the uncertainty of collinear SSH difference is approximately 0.3cm, which corresponds to a geostrophic velocity error of 8cm/s at 25°N, where an instrumental range error of 1.7cm and along-track spacing of 5.8km are assumed. Following Powell and Leben (2004), no environmental corrections are applied to SSH data when collinear SSH slope is computed.

Comparisons of SSH must account for the offset in mean-sea-level between models and data. Hence, a linear fit to along-track SSH is computed for the period under consideration, and for each track, similar to the loess-type filter used by Leben (2005). This adjustment was performed only for the SSH comparisons, not the collinear SSH slope comparisons.

Step 1

The above-described methodology was used to compare the GOMEX-PPP models, and the AVISO MADT product, with Jason-1/2 data. Statistics were accumulated over non-overlapping 10-day windows for the mean and standard deviation of the difference between the model and satellite SSH or SSH slope. Results for the individual models are found in the following directories:

(1) SSH comparisons: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step1/
(2) Collinear SSH slope comparisons:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/AltimetrySlope9pt_Step1/

Overall, the rms-errors increase in all the models (and AVISO) in the June-July period when the LC intrusion into the Gulf is largest and the LCE is forming. Figure 23 illustrates the results in summary form. AVISO has the best comparison with the Jason data, with an rms error of about 8cm. This error is considerably larger than the instrumental error of about 2cm, reflecting the fact that the AVISO product has filtered out most SSH variability at spatial scales smaller than 150km and at time scales less than 20 days (Chelton, personal communication). In addition, there is an error contribution due to the uncertain mean-sea-surface.

Note that the majority of models (except #1 and #6) have less SSH variance than the Jason data, about 17cm vs. 19cm, and this difference is, again, larger than can be ascribed to instrumental error (sqrt(19^2 – 17^2) = 8cm).

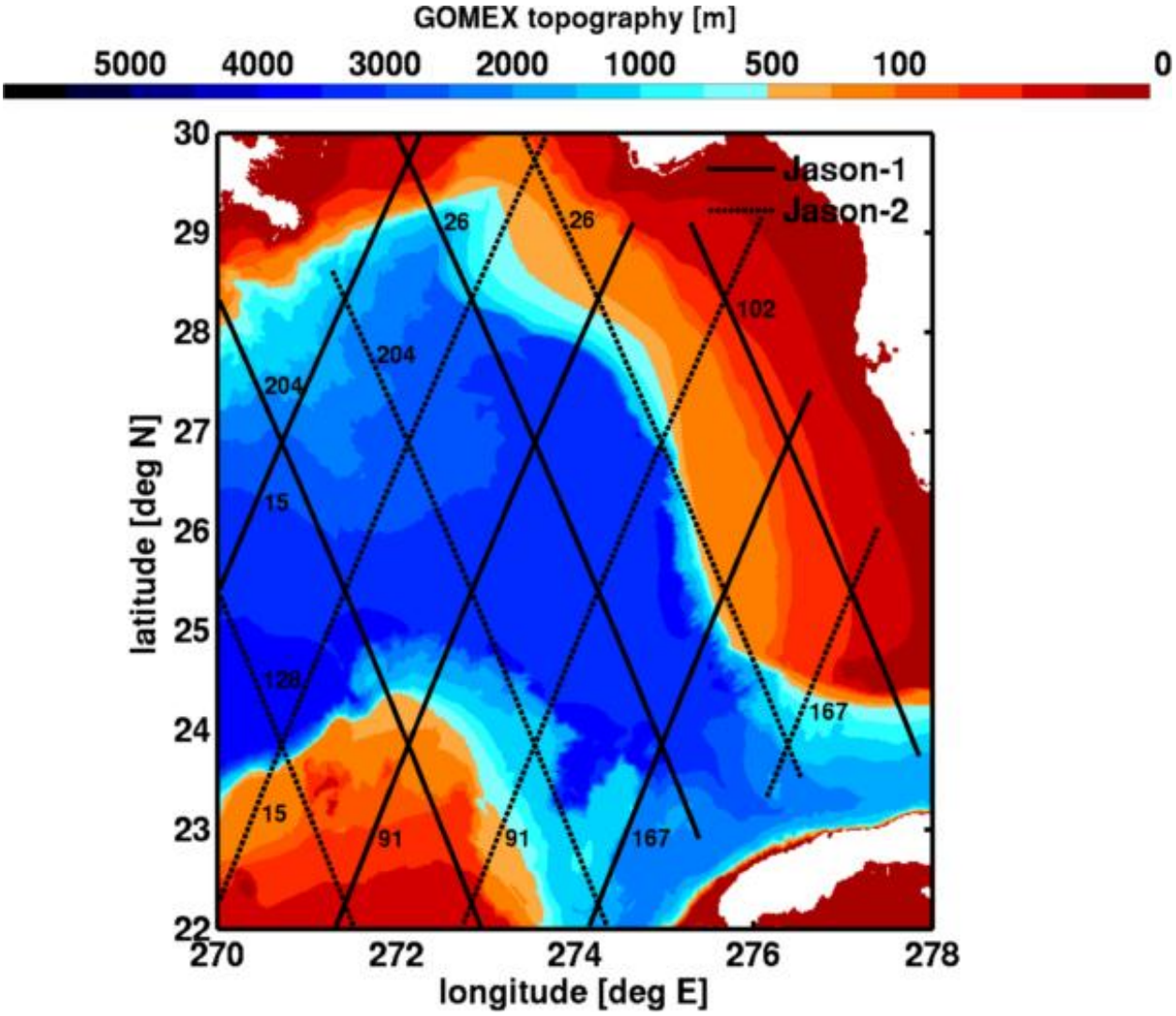


Figure 22: Altimeter Ground Tracks.

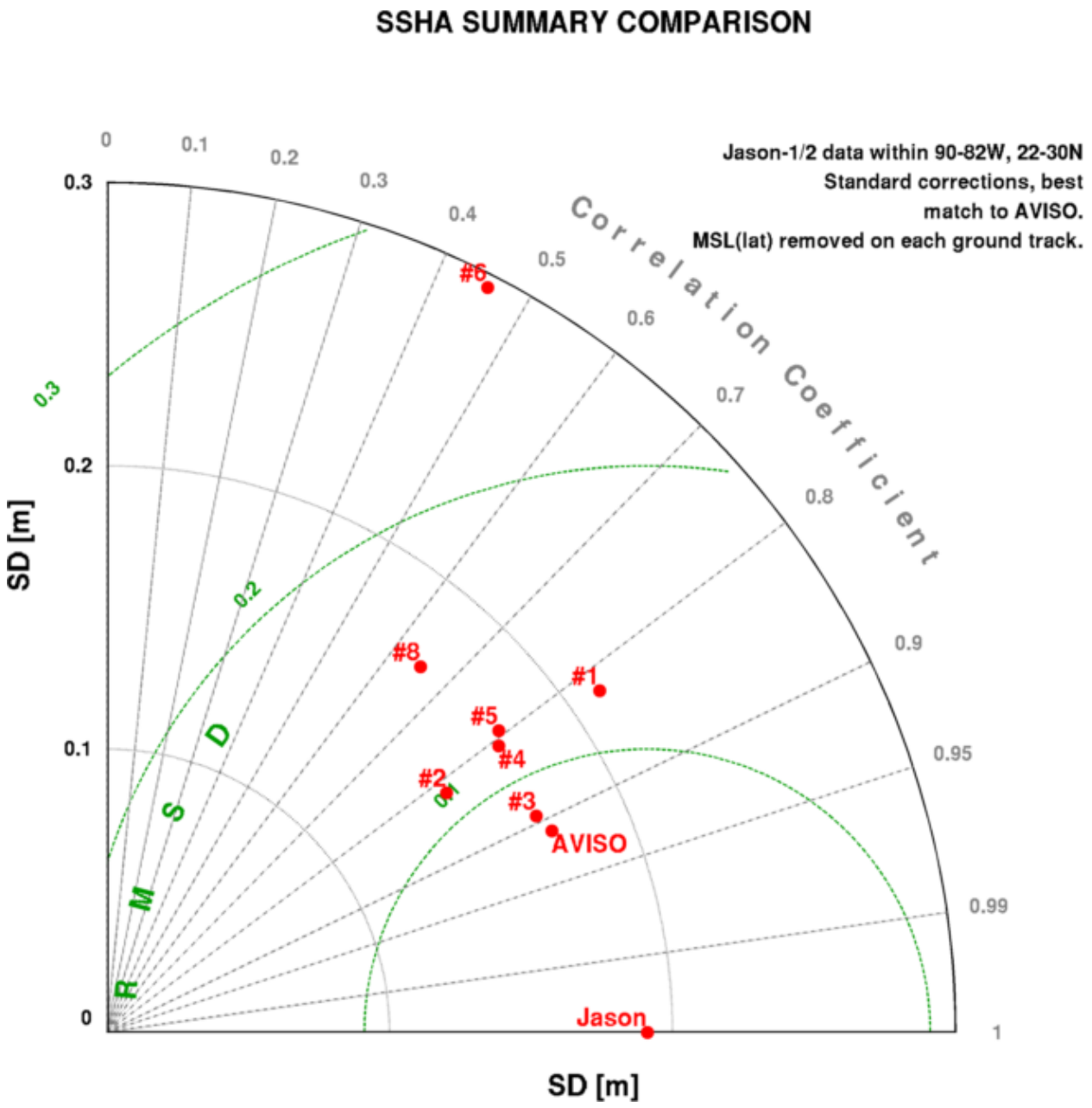


Figure 23: SSH Taylor Diagram, Step 1.

A summary of collinear SSH slope is shown in Figure 24. Results are similar to the SSH comparison in that AVISO displays the best correlation with the data, and the model performance is generally clustered except for Model #6. The numerical values are expressed as collinear SSH difference between consecutive points. The nominal spacing of ground track points is about 5.8km, so a collinear SSH error of 0.01m corresponds to a sea-surface slope error of about 1.7e-6. And slope error of this magnitude corresponds to a geostrophic velocity error of 28cm/s. In fact, the model slope errors tend to be associated with errors in the location and intensity of the LC/LCE fronts, so error velocities are actually much larger. Figure 25 shows the time series of observed and modeled collinear SSH differences. Large errors in SSH slope are present at orbit cycles 12 through 20 (days 110 through 190) which appear to be frontal features in the SSH along the Jason-1 ground track (pass 26, left panel). Note that similar features are absent in the nearby Jason-2 ground track (pass 204, right panel), indicating the importance of front orientation and ground track spacing to the detection of SSH fronts by altimetry. The slope error on Jason-1 orbit cycle 15 at 24.5°N corresponds to a geostrophic velocity error of about 1.5m/s. Figure 26 presents a histogram of collinear SSH slope (model vs. observed, left column) and slope error (model-observed vs. observed, right column) for Model #1. Top row shows comparison on Jason-1 tracks, and bottom row shows comparison along Jason-2 tracks. While the model and observations are evidently correlated, the slope of the modeled vs. observed SSH is less than one, indicating a substantial bias. Modeled SSH slopes, i.e., geostrophic near-surface velocity, is consistently under-predicted. A similar picture could be presented for the other models, except for Model #6, which displays little correlation with observations.

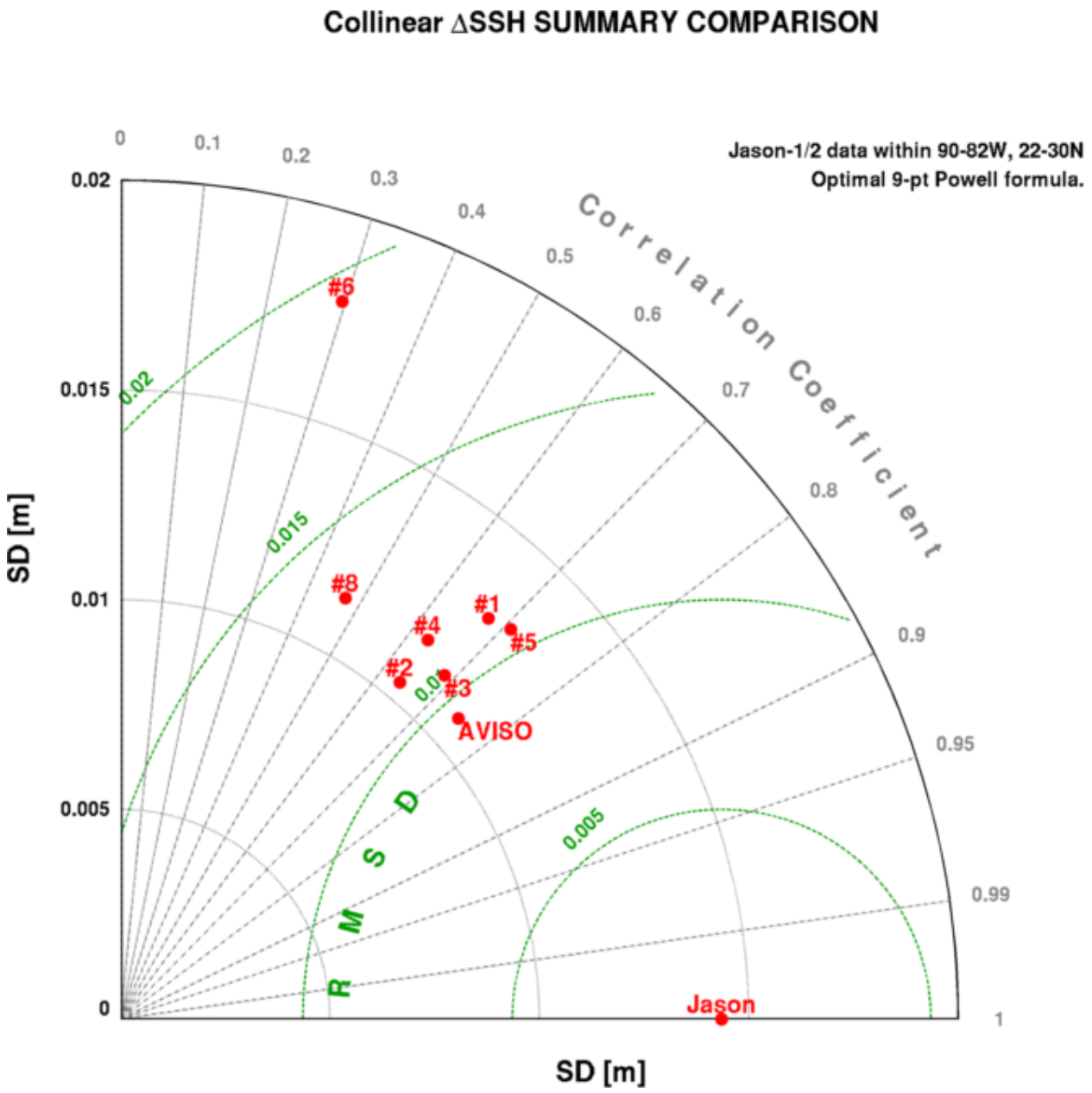


Figure 24: Collinear SSH Differences Taylor Diagram, Step 1.

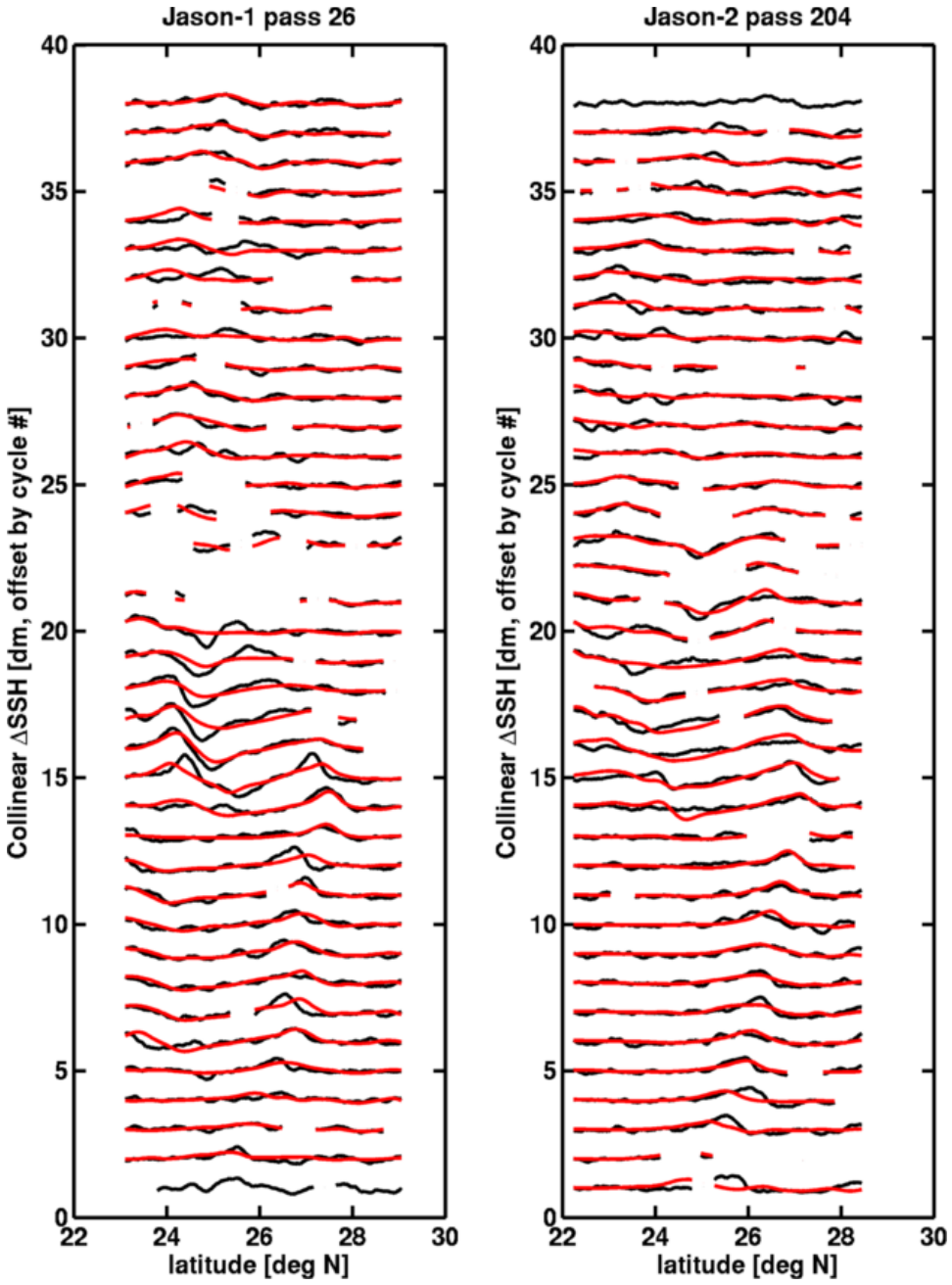


Figure 25: Model vs. Observed Data, Collinear SSH Difference, Step 1.

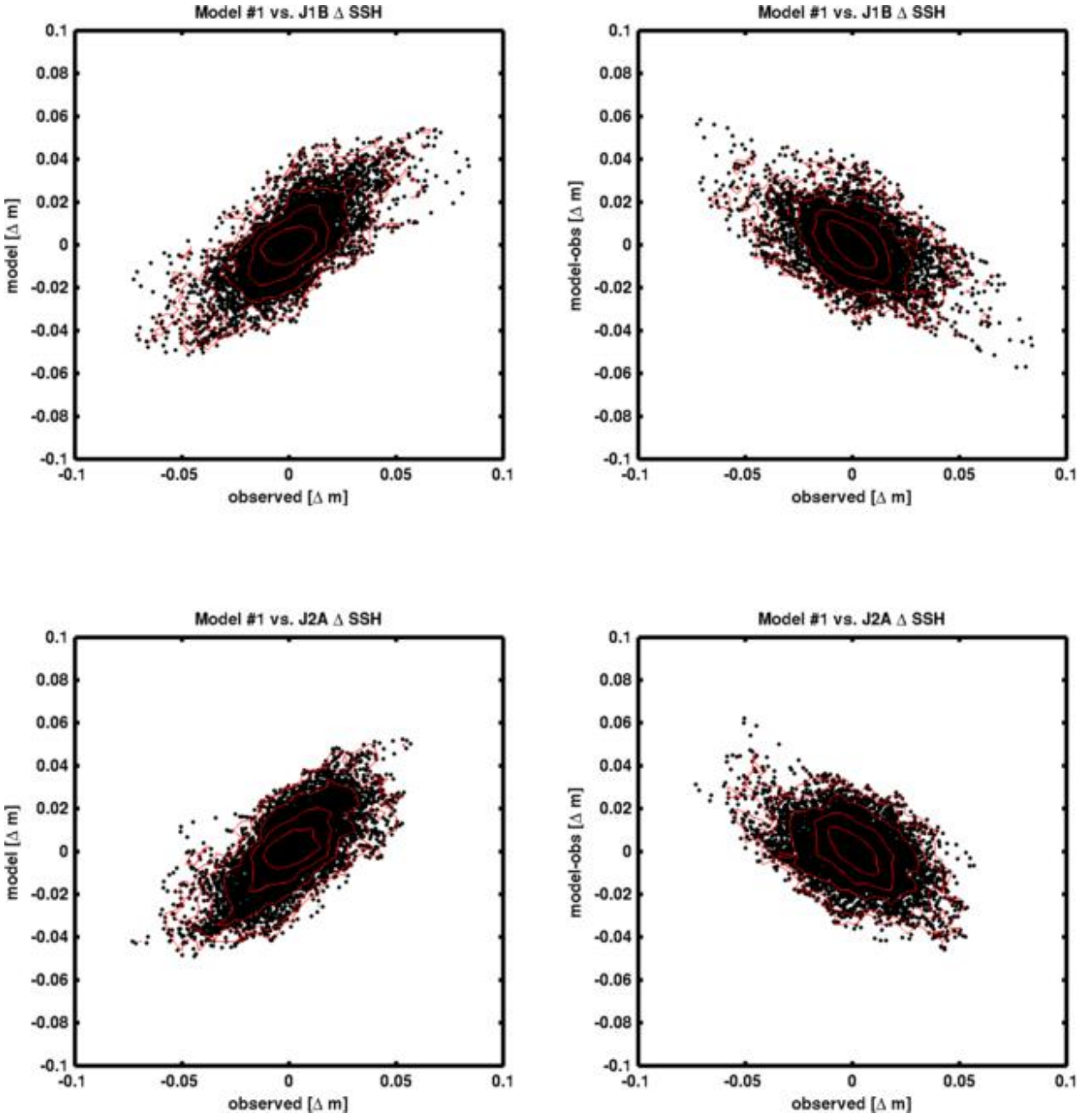


Figure 26: Model vs. Observed Data, Collinear SSH Differences, Step 1.

Step 2

The retrospective forecasts of Step 2 have been analyzed in a manner similar to Step 1. Model and observed SSH are compared along the Jason-1 and Jason-2 ground tracks, and the statistics of the differences are averaged over 10-day windows. The objective of this comparison is to quantify the accuracy of the model forecasts, compared to satellite altimetry data.

The performance of Model #1 is illustrative (Fig. 27). Two sets of curves, model forecast (red) and model persistence (black), where the initial time of the forecast is marked with a circle are examined. The persistence and model forecasts do not exactly overlap, because the persistence error is based on the model state at the initial time, while the forecast error is averaged over a 10-day window beginning at the initial time. The rms difference between model and observed data ranges from approximately 7 to 36cm during 2010, with the largest error occurring in mid-year in association with the LCE formation. Generally, the forecast error is smaller than the persistence error.

The difference between the model and observed SSH is the sum of a mean-sea-surface error (steady in time), and along-track average error, and a time- and space-variable anomaly. The mean-sea-surface computed from the Step 1 hindcast has been subtracted, but there is a time-variable, large-scale difference in SSH which contributes to the rms error shown. Thus, the rms error is a combination of both steady and unsteady components.

An alternative view of the model skill is shown in Figure 28, which presents same data as Figure 29 in terms of a skill score. Skill score is a metric of the explained variance in the observed SSH, defined as

$$SS = 1 - \text{mean\_square(model-data)}/\text{mean\_square(data)}.$$

The skill score defined here, being based on the mean square, includes both bias and variance in order to facilitate the intercomparison with respect to a constant mean-sea-surface throughout the year. The forecast skill is best, and decays most slowly, in the forecasts begun in the January—March and September—November time periods. The skill of both model and persistence decrease rapidly in the April, May, and July forecasts.

Every model shows increased rms error in the June—July time period; although, the models differ in detail with respect to size of initial condition errors and error growth rate. Results for the individual models are available on the Web as follows:
rms error:
#1: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/iasnfs_mse.pdf
#2: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/profs_mse.pdf
#3: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/iasroms2_mse.pdf
#4: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/mitgom_mse.pdf
#6: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/profs_ngom.pdf
#8: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/profs_iasroms1.pdf

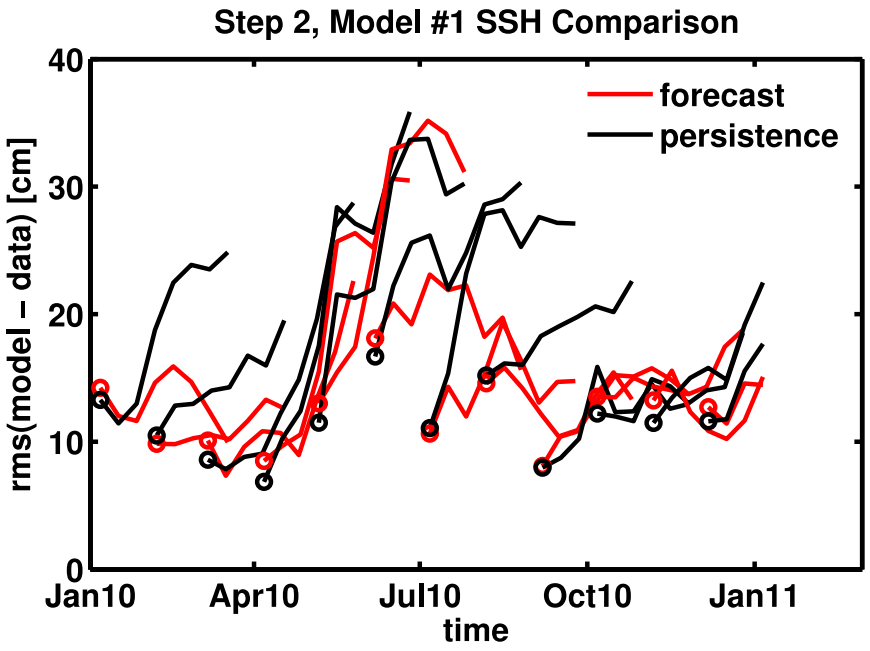


Figure 27: Model-Data SSH Error, Step 2.

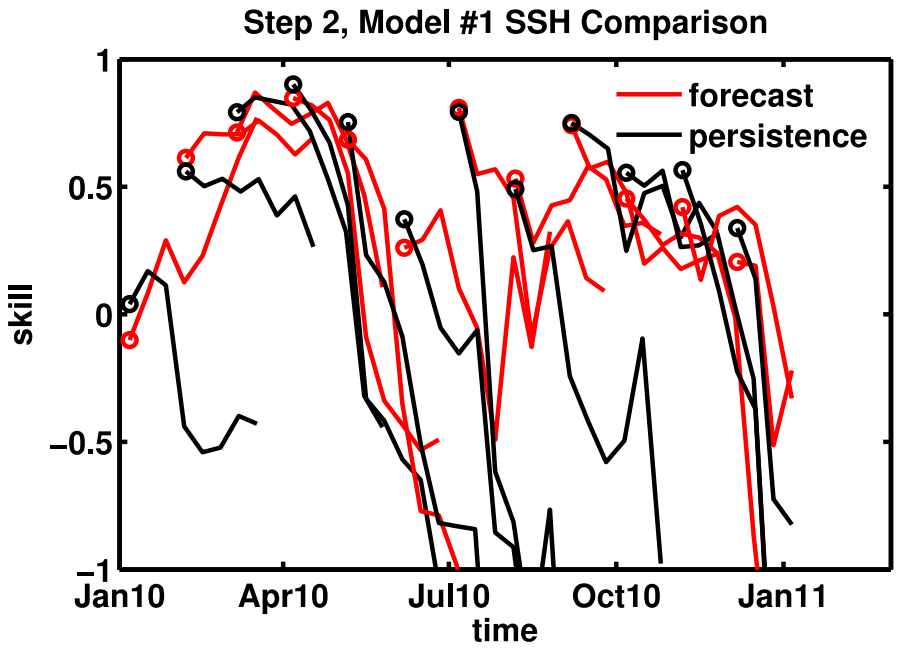


Figure 28: Model SSH Forecast Skill, Step 2.

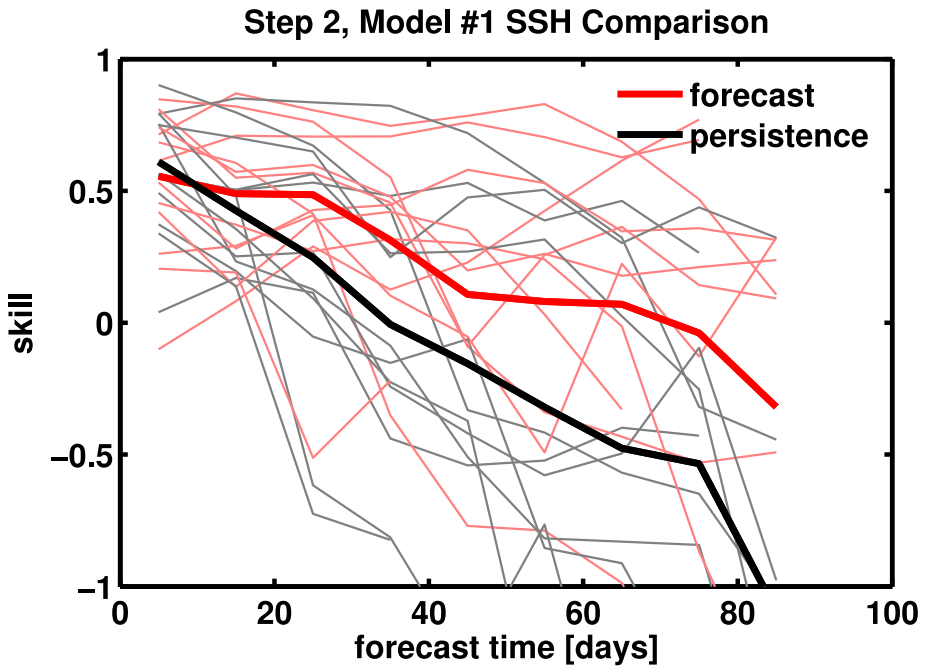


Figure 29: SSH Forecast Skill vs. Forecast Time, Step 2.

skill scores:
#1: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/iasnfs_skill1.pdf
#2: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/profs_skill1.pdf
#3: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/iasroms2_skill1.pdf
#4: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/mitgom_skill1.pdf
#6: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/ngom_skill1.pdf
#8: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2/iasroms1_skill1.pdf

It is evident that there is considerable month-to-month variation in the forecast skill. Nonetheless, differences amongst the models are evident if averages over the twelve forecasts are computed. Figure 29 shows the skill of Model #1 computed by averaging as a function of forecast time. Heavy lines show the average skill of the forecast (red) and persistence (black), while the thin lines show each of the twelve realizations. The scatter is considerable, but the model forecast shows more skill than persistence. Because the variance of the SSH field is inhomogeneous and associated with strong current features, such as the LC and LCE, placing confidence limits on the skill score is problematic. Parametric methods (e.g., t-test) are unlikely to be reliable due to inhomogeneity and the uncertainty in the number of degrees of freedom in the skill statistic; and non-parametric (e.g., bootstrap) estimates are also unreliable due to the small sample size.

The model skill during Step 2 is summarized (Fig. 30) by the average forecast skill vs. forecast time for each model (bold colored thick lines) in addition to average persistence skill vs. forecast time (light colored thin lines). Because there is no single "verifying analysis" at the initial time,

each model forecast is compared with persistence of the model initial conditions. All models, except for Model #6, have initial skill in the range of 0.4 to 0.7, and the skill of persistence decays at a rate of about 0.5/mo.

Forecast skill differs amongst the models. Models #2 and #6 have negative skill at the end of the 90-day forecast, meaning that the mean square error between model and observed data is twice as large as the mean square of the data. After 20 to 50 days, the remaining models display more skill than persistence. However, whether the models achieve useful levels of skill cannot be evaluated in isolation from the end-use of the forecast.

Step 3

The above-described methodology has also been used to evaluate forecast vs. persistence skill during the Step 3 runs. Because the forecasts were created in real-time, there is not yet observational data spanning the entire forecast period. The results, using Jason-1 and Jason-2 observations through January 12, 2012, are examined (Fig. 31). Persistence loses skill at a rate of 0.75/mo., faster than in Step 2. This increased rate of skill loss appears to be connected with the seasonal cycle of LC/LCE evolution, considering that Step 2 results also degraded at a faster rate during the latter part of the year.

Forecasts from Models #2 and #6 lose skill at the fastest rate, but the considerable scatter between Step 2 and Step 3 suggests that results are not likely significant.

Summary

A comparison of SSH from Step 1 hindcasts was performed to assess the GOMEX-PPP models' skill at nowcasting SSH. The modeling systems each assimilate SSH data, so the comparison essentially measures how much of the SSH signal is utilized by the assimilation systems. Comparisons were performed to assess two aspects of the SSH, and statistics were computed for both SSH and collinear SSH slope, the latter being directly related to the component of geostrophic velocity perpendicular to the satellite ground tracks. RMS differences for both quantities exceed estimated data error by about a factor of 3. Considering the skill score conventionally defined as 1-(mean square error)/(mean square signal), i.e., fraction of variance explained, model skill ranges from -0.98 to +0.67 for SSH, and -0.82 to +0.46 for collinear SSH slope. Low skill scores (Table 19) obtained for collinear SSH differences are consistent with apparent under-utilization of SSH data in GOMEX-PPP assimilation systems. Large-amplitude, temporally-coherent features in along-track SSH gradient are absent from the models.

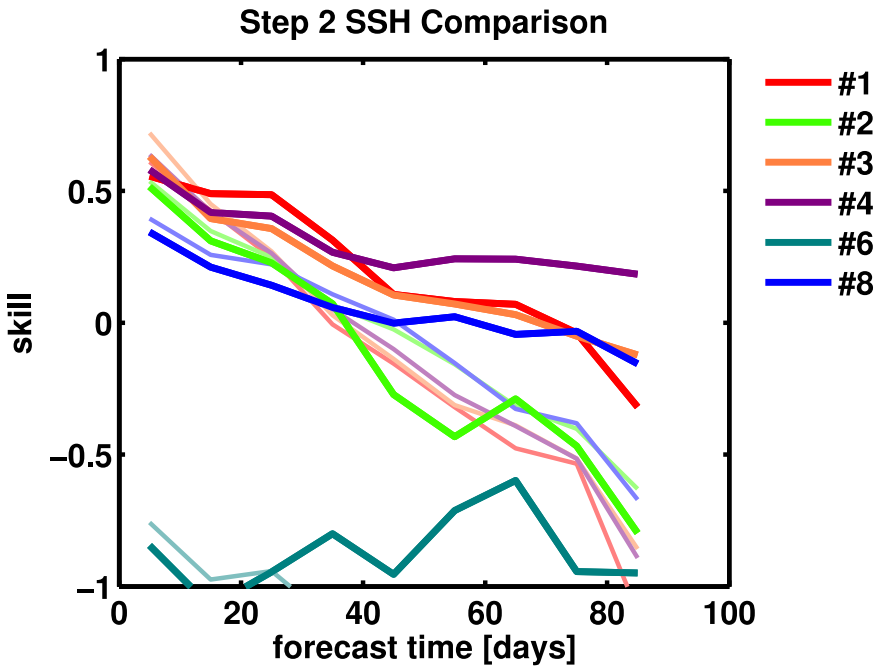


Figure 30: SSH Forecast Skill Summary, Step 2.

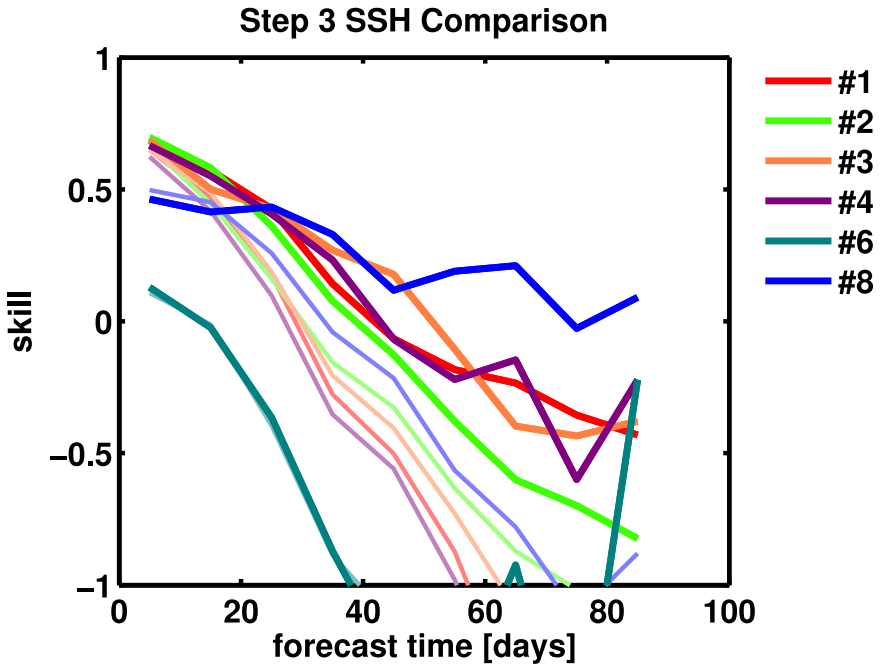


Figure 31: SSH Forecast Skill Summary, Step 3.

Table 19: SSH Initial Skill Scores, Step 1

Skill is defined as 1 – (mean square error)/(mean square data), i.e., fraction of variance explained.

Model	SSH Skill	Delta SSH Skill
#1	0.59	0.41
#2	0.67	0.40
#3	0.78	0.46
#4	0.65	0.37
#5	0.61	0.46
#6	-0.98	-0.82
#8	0.37	0.12
AVISO	0.83	0.56

The same methodology was used to assess forecast skill as a function of forecast lead time during Step 2 and Step 3. Except for Model #6, which had very poor initial skill, the initial skill of the models was similar, in the 0.4 to 0.7 range. Model forecast skill exceeded the skill of persistent initial conditions in nearly all cases.

GOMEX Sea-Surface Temperature Field

Introduction

Sea-surface temperature (SST) is an active tracer of near-surface processes. In-so-far as temperature is a conserved property of the fluid, the SST can serve as a useful marker for distinct water masses, e.g., central GOMEX vs. Caribbean water, which makes it useful for tracking the evolution of fronts associated with GOMEX currents. But SST is not conserved, its dynamics are determined by air-sea heat exchange and subsurface mixing, so there is no one-to-one relationship between a water mass and SST. Also, because the water density is determined in part by temperature, SST is related to the pressure field, geostrophic currents, and acceleration of the fluid. Accurate modeling of SST is a challenge due to the range of processes on which SST depends.

Satellite-based infrared and microwave remote sensing of SST are capable of producing GOMEX-scale views of SST, but both measurement techniques are sensitive only to the very near-surface temperature of the water which may differ considerably from the bulk mixed layer temperature resolved by ocean models. Remote-sensing methodologies are sensitive to interference from a number of environmental processes unrelated to SST.

Step 1

A series of intercomparisons of the modeled and observed SST have been performed for the Step 1 retrospective nowcasts of 2010. Observed SST is taken from the Global High-Resolution SST (GHRSST) analysis produced by NAVO at daily intervals on 0.1-degree resolution grid by merging data from several instruments and a climatology (NAVO-L4HR1m-GLOB-K10_SST, available via http://data.nodc.noaa.gov/opendap/ghrsst/L4/GLOB/NAVO/K10_SST). Several GHRSST products are currently available, see https://www.ghrsst.org/data/real-time-2b/, and the NAVO product was chosen from a cursory examination of fields from several alternatives. A thorough intercomparison of GHRSST products is beyond the scope of this report, but the NAVO product appeared to be produced on a more timely basis and contain fewer artifacts (e.g., visible sensor ground tracks) than several alternatives.

The time average SST during Step 1 is shown at http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/SST_EOF_Step1/mean.png. The observed SST (upper left panel) is labeled "Model NAVO." In every case, the modeled SST is less than the observed. The pattern of the LC intrusion into the Gulf is present in all models; however, SST contrast between, say, Yucatan Channel and western Campeche Bank differs between models and observed data. In particular, the model temperature gradient is stronger than observed in Models #2 and #6.

Figure http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/SST_EOF_Step1/std.png shows the standard deviation of SST. The large-scale features of observed SST variability are reproduced in the models; however, the SST variability of the Caribbean waters is noticeably small in Models #2 and #6.

The spatial patterns of SST have been examined in two ways. Figure http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/SST_EOF_Step1/corr.png shows the correlation of model and observed SST in the LC intrusion area (marked with a "+"). Observed SST is remarkably correlated with model SST at this site; however, this result is likely due to averaging over the entire year, including summer months when remotely-sensed skin-temperature is likely to deviate strongly from bulk mixed layer temperature. The fact that Model #2 SST correlation is unlike the consensus of other models and observed is not necessarily symptomatic of problems in Model #2; it may mean that the other models are over-fitting observed SST data which are not representative of the bulk mixed layer.

Figure http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/SST_EOF_Step1/eof1.png illustrates the first mode EOF of SST for the models. [Once again, recall that Model #5 is not comparable in this figure due to lack of complete 1-year time series.] The bulk of the SST variance (over 90% in all models) is connected with the north-south gradient of temperature. The second mode EOF is shown in Figure http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/SST_EOF_Step1/eof2.png; it essentially depicts an out-of-phase relationship between the near-coast SST and offshore SST. The fact that mode-2 contains in-phase variability between near-coast and the LC is not particularly significant, as it is based on a single year of data and does not suggest a dynamical linkage between the processes. GOMEX-scale views of SST variability contain many different modes of

variability, and it is suggested that a series of regional analyses would provide more meaningful comparison with observations.

Step 3

The Step 3 model forecast SST fields are compared with remotely-sensed SST fields during the brief eddy detachment/reattachment event in the November 11-25 time period. Comparison of numerical values of SST between the models and the observed data were not made.

Dr. Chuanmin Hu of the Optical Oceanography Laboratory (OOL) at the University of South Florida produces operational SST products for various regions including the Gulf of Mexico. We used his "7DAY SST MEAN" product derived from MODIS/Terra and MODIS/Aqua data for the Gulf of Mexico region. They are composite images based on data from the preceding 7 days yielding cloud-free images. The images used here are available from the OOL website.

Based on the OOL SST fields at fortnightly intervals from 15 SEP 2011 to 20 JAN 2012, the eddy is still attached on 11 NOV 2011 (Fig. 32). Images from nearby days (not shown) indicate that the eddy is substantially detached from 17 to 19 NOV (Fig. 33). This detachment occurred approximately a week after the event was seen in the CCAR SSH fields. The eddy begins to reattach on 20-November which is about three days later than when the SSH reattachment occurred.

Animated side-by-side comparisons of model SST for the six models and seven Step 3 forecast experiments are available on the GOMEX-PPP Website under Results for Step 3. The model SSTs on 17 NOV 2011 from the run initialized on 11 NOV 2011 (S3-05) are examined (Fig. 34). The color table was stretched to span 23-30 degrees-C to make it easier to discriminate the edges of the LC and LCEs. Consequently, shallow water may fall outside the color bar range during cool weather.

Individual Step 3 model animations were used to identify when the brief detachment and reattachment occurred (Table 20 and Table 21). The results are subjective and the dates could easily shift a day or two during subsequent estimates or by different analysts. Alternate color schemes might yield smaller variability in estimates. Model 1 and Model 6 were the most consistent in forecasting the detachment even 8 weeks out. Model 4 has the date in the 4-week forecast. All models are close in the two week forecast (S3-04).

The reattachment events (Table 21) were much harder to detect by eye than the separation events. The results are unclear. Model 1 and Model 6 results are fairly consistent with their SSH results but only Model 6 seemed to have the date correct in all runs. Model 4 did not show a reattachment event.

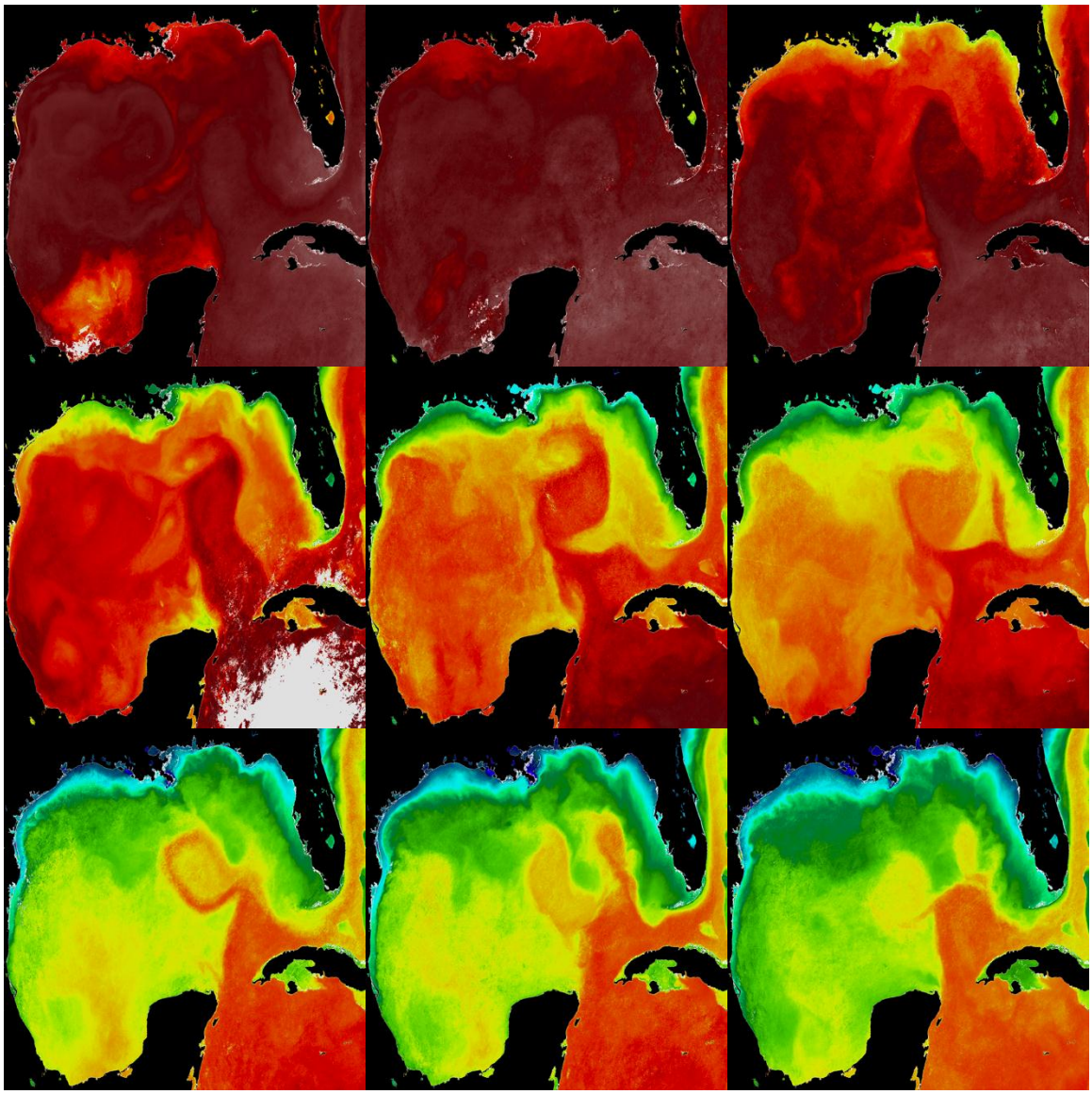


Figure 32: Fortnightly SST, 15 SEP 2011 – 20 JAN 2012, Step 3.
The progression of seasonal cooling is from September (top left) to January (lower right) influences the accuracy of the identification of frontal location from SST.

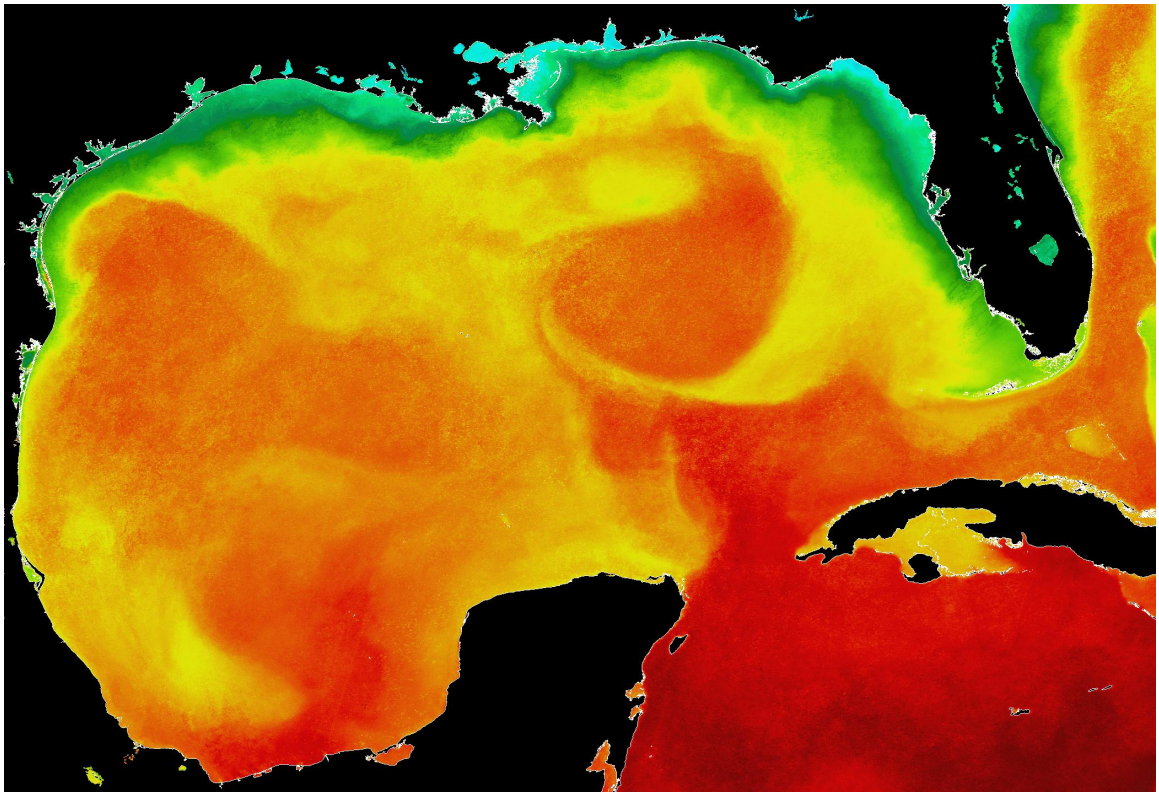


Figure 33: SST on 17 NOV 2011.

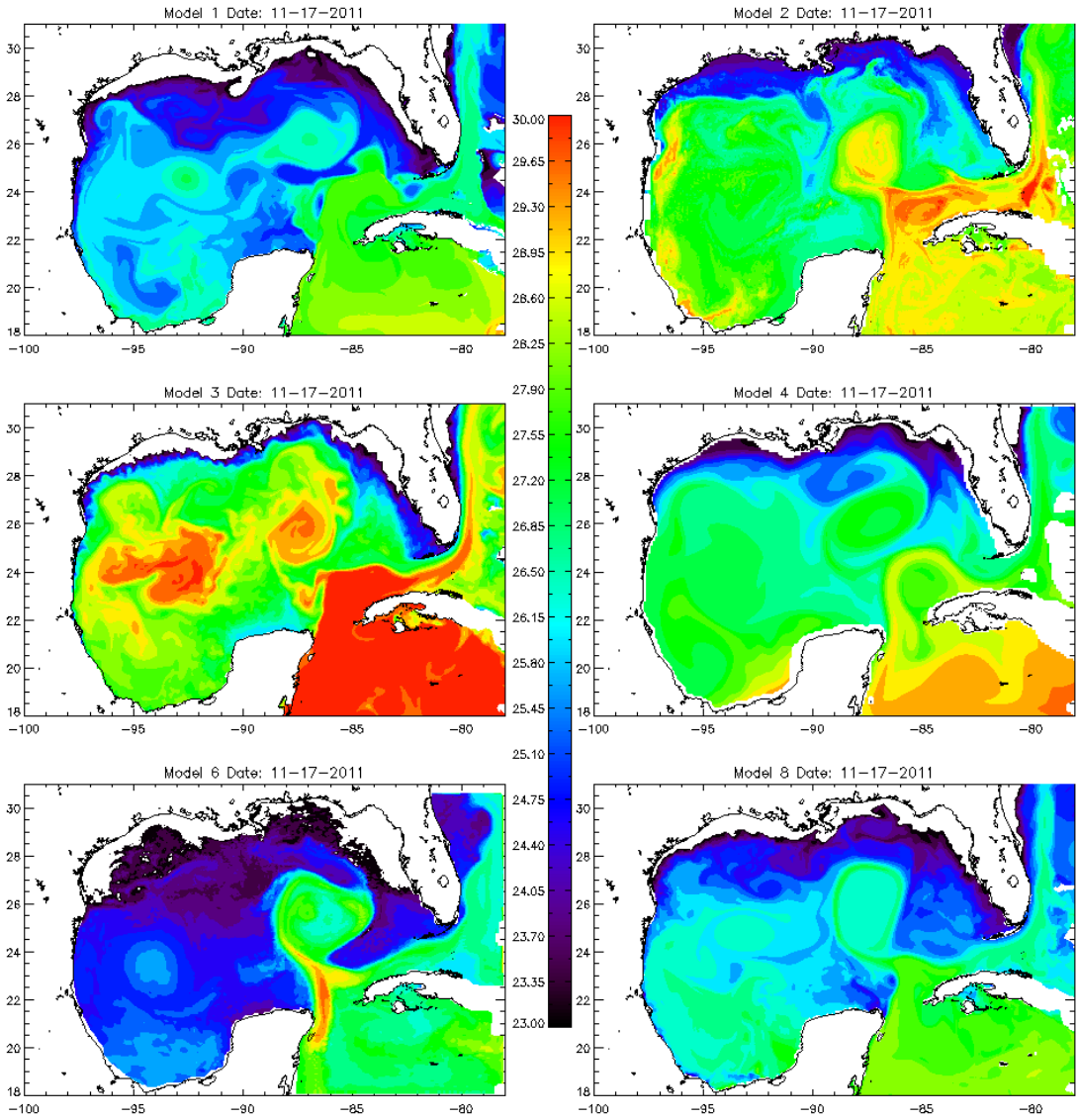


Figure 34: Model SST on 17 NOV 2011, S3-05, Step 3.

Table 20: LCE Separation Events -- Subjective SST-Based, Step 3

Based on OOL SST, separation occurs 11-17 NOV 2011.

Model	S3-01 (8 wks)	S3-02 (6 wks)	S3-04 (4 wks)	S3-05 (2 wks)	S3-05 (0 wks)
#1	13 Nov	19 Nov	13 Nov	11 Nov	17 Nov
#2	2 Nov	9 Nov	12 Dec	18 Nov	16 Nov
#3	22 Oct	11 Nov	26 Nov	16 Nov	17 Nov
#4	28 Nov	6 Dec	15 Nov	17 Nov	12 Nov
#6	N/A	15 Nov	17 Nov	19 Nov	19 Nov
#8	18 Nov	30 Oct	30 Oct	19 Nov	20 Nov

Table 21: LCE Reattachment Events -- Subjective SST-Based, Step 3

Based on OOL SST, reattachment occurs 21-25 NOV 2011.

Model	S3-01 (8 wks)	S3-02 (6 wks)	S3-04 (4 wks)	S3-05 (2 wks)	S3-05 (0 wks)
#1	16 Nov	1 Dec	22 Nov	16 Nov	24 Dec
#2	10 Nov	15 Nov	16 Dec	3 Dec	24 Nov
#3	-	20 Nov	16 Dec	22 Dec	25 Dec
#4	-	-	-	-	-
#6	N/A	23 Nov	23 Nov	21 Nov	25 Nov
#8	24 Nov	7 Nov	10 Nov	-	-

GOMEX Sub-Surface Thermal Field

Introduction

The subsurface thermal field of the GOMEX is the result of advection and mixing of the temperature field at the ocean surface and the thermal structure carried into the GOMEX at its open boundaries. Together with the salinity, the temperature is one component of the water mass identity used by oceanographyers. The dynamical significance of subsurface temperature is its contribution to the buoyancy field, which is the vertical gradient of the pressure field. The buoyancy field also contributes to the hydrodynamic stability of oceanic flows through its contribution to the gradient Richardson number and other buoyancy-related stability parameters. These dynamical factors, as well as other biological significance of temperature to marine organisms make the accurate forecasting of subsurface temperature a top priority for GOMEX modeling systems.

The subsurface thermal field of the GOMEX is, in principle, accessible to measurement via autonomous underwater vehicles, expendable temperature profilers (XBT and AXBT), and conventional, ship-based, CTD measurements. During 2010 there are an unusually large number of subsurface temperature measurements available, which are used here to quantify model skill.

Step 1

Model subsurface temperatures have been compared with the AXBT data described in Shay et al. (2011) in support of the Deepwater Horizon incident. Between May and July 2010, at 7 to 9-day intervals, AXBTs were deployed in an approximately rectangular grid to the southeast of the Deepwater Horizon site, with approximately 100km resolution. Nine nearly synoptic snapshots of subsurface temperature were thus obtained over a large area encompassing the LC, which had penetrated northward to about 27°N.

The nine surveys were conducted on the dates shown below. The number of AXBT records used from each survey are shown in parentheses.

S1: 08 May 2010 (46)	S4: 28 May 2010 (37)	S7: 18 Jun 2010 (23)
S2: 18 May 2010 (28)	S5: 03 Jun 2010 (33)	S8: 25 Jun 2010 (53)
S3: 21 May 2010 (41)	S6: 11 Jun 2010 (48)	S9: 09 Jul 2010 (54)

Fifty-six contour plots of the differences (model-AXBT) of the 20°C isotherm are shown at http://abcmgr.tamu.edu/gomexppp/?page_id=203. This temperature level was chosen because it is typically below the surface mixed-layer, and it is present at very different depths in central GOMEX (100m) vs. Caribbean waters (250m).

Table 22 contains a quantitative summary of the modeled minus observed isotherm depths for each airborne survey. All except Model 7 have a mean error less than 20m and root-mean-square error less than 60m. Most models have a negative mean bias. Negative bias indicate the 20°C isotherm was found deeper in the model than in the observations. A negative spatial bias can be interpreted as regions where observed LC or LCEs were not found in the model. With this interpretation most of the bias occurs in the region between the newly separated LCE and the LC in early June, which is thought to be due to the model LCE separating earlier than observed.

Table 22: Depth of 20C isotherm.
The expected error <E> (Model-Observed) in meters of the depth of the 20°C isotherm averaged over AXBT data for nine AXBT surveys (S1-S9) for eight models (M1-M8). Also shown is the root-mean-square of <E> and the average values over all models and all surveys.

Model #	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
	<E>	<E>	<E>	<E>	<E>	<E>	<E>	<E>	<E>	
M1	3.6	-10.8	-13.8	-10.5	-10.6	-18.2	-21.4	-24.1	-29.4	-15.0
M2	-10.9	-19.2	-16.8	-19.9	-12.9	-24.3	-29.0	-22.9	-20.9	-19.6
M3	-19.8	-15.9	-17.7	-16.2	-14.5	-11.0	-17.3	-5.8	-6.8	-13.9
M4	4.3	-3.0	-4.7	-4.7	-1.1	-9.0	-8.7	-5.2	-5.3	-4.2
M5	n/a	n/a	n/a	-4.3	2.1	-5.3	-9.2	-7.3	-14.4	-6.4
M6	25.9	15.8	14.5	9.0	8.1	11.7	-7.5	-11.9	8.3	8.2
M7	10.7	10.1	9.7	15.5	44.1	62.1	58.3	95.4	127.9	48.2
M8	-10.0	-18.2	-24.0	-20.6	-12.8	-13.0	-15.6	-11.6	-8.5	-14.9
Mean	0.5	-5.9	-7.5	-6.5	0.3	-0.9	-6.3	0.8	6.4	
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
M1	20.4	36.6	41.4	36.6	46.9	52.4	42.8	48.7	51.7	41.9
M2	39.4	44.1	45.0	47.2	43.8	52.6	50.6	50.8	51.4	47.2
M3	31.1	48.8	48.6	32.3	41.1	50.2	32.9	37.8	39.6	40.3
M4	42.2	39.5	40.3	36.6	35.2	34.7	30.3	41.6	44.0	38.3
M5	n/a	n/a	n/a	33.4	45.2	33.5	28.0	35.3	38.0	35.6
M6	46.4	50.8	45.1	35.3	55.0	80.5	75.2	64.7	55.9	56.5
M7	59.4	72.4	64.1	45.3	60.0	83.1	77.8	113.3	163.0	82.0
M8	33.6	53.7	56.7	40.6	42.9	49.1	45.9	51.1	50.7	47.1
Mean	38.9	49.4	48.7	38.4	46.3	54.5	47.9	55.4	61.8	

Another comparison was performed using the temperature difference between 30 and 360m. This metric attempts to measure how well the stratification of the main thermocline in the LC and LCE were represented by the models. Survey-averaged mean and root-mean-square errors are in the range of +/-3°C, which corresponds to vertical displacements of approximately 75m at 300m depth (Table 23). Depending on the horizontal structure of the thermal field, this could correspond to a horizontal error of 50 to 200km displacement of an eddy or frontal feature. Thus, it seems that the models have little skill in nowcasting the structure of the thermal field. Spatially, the largest errors occur between the LCE and LC following the June separation event and preceding the event in the region to the NE of the LC where HMI reported an eastward meander that was not generally reflected in the models.

Table 23. Thermocline thickness.
Mean Error <E> = < $((T_{30}-T_{360})_{AXBT} - (T_{30}-T_{360})_{Model})$ > and Root-Mean-Square Error in degrees C for nine synoptic AXBT surveys(S1-S9) and eight models (M1-M8).

M#	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
	<E>	<E>	<E>	<E>	<E>	<E>	<E>	<E>	<E>	<E>
M1	-0.6	-1.4	-1.4	-0.6	-0.8	-1.3	-1.3	-1.4	-2.2	-1.2
M2	0.1	0.4	0.9	0.8	0.6	0.1	0.0	0.1	-0.1	0.3
M3	-0.1	0.2	0.7	0.8	0.5	1.0	1.0	1.2	1.0	0.7
M4	-0.2	-0.1	0.0	0.1	-0.1	-0.6	-0.5	-0.5	-0.7	-0.3
M5	n/a	n/a	n/a	-0.7	-1.1	-0.9	-1.4	-0.6	-0.7	-0.9
M6	0.3	1.3	1.6	2.1	1.7	2.9	2.6	2.7	3.1	2.0
M7	-0.9	-1.6	-1.5	-1.2	-0.5	-1.2	-1.2	-0.9	0.3	-1.0
M8	-0.5	-0.5	-0.3	-0.2	-0.3	-0.3	-0.3	0.2	0.0	-0.2
Mean	-0.3	-0.2	0.0	0.1	0.0	0.0	-0.1	0.1	6.4	
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
M1	1.5	1.9	2.1	1.4	1.7	2.1	2.1	2.1	2.8	2.0
M2	1.2	2.1	2.8	3.0	2.4	2.9	2.9	2.6	2.4	2.4
M3	0.8	1.8	1.9	1.4	1.5	1.9	1.9	2.2	2.0	1.7
M4	1.2	1.4	1.9	1.5	1.7	2.1	2.2	2.1	1.8	1.8
M5	n/a	n/a	n/a	1.6	2.0	1.8	1.9	1.5	1.4	2.8
M6	1.2	2.0	2.1	2.5	2.7	3.9	3.5	3.5	3.8	2.8
M7	2.9	3.9	3.7	3.0	2.0	3.2	4.0	3.9	5.7	3.6
M8	1.4	1.7	1.7	1.5	1.4	1.5	2.0	1.7	2.2	
Mean	1.5	2.1	2.3	1.8	1.9	2.4	2.6	2.5	2.8	

Step 2
The results described above carry over into Step 2. The models have little skill in Step 1 nowcasts which are used to initialize the Step 2 forecasts. Consequently, the skill of forecasts does not degrade with time.

For example, the mean and rms error of the depth of the 20°C isotherm as a function of forecast lead time is examined (Fig. 35). Because of the 3-mo. forecast horizon in Step 2, with forecasts initialized monthly, there are three forecasts for the 20°C isotherm depth at the time of each AXBT survey. Model #3 displays a significant initialization transient, and mean error reduces as a function of forecast lead time (it is speculated that initial errors are reduced by damping towards climatology as the forecast proceeds in this model). Errors are smallest in Model #4, but, given the scatter, it cannot be concluded that the difference is significant. The failure of the error to grow as a function of forecast lead time suggests that the models have no forecast skill relative to persistence.

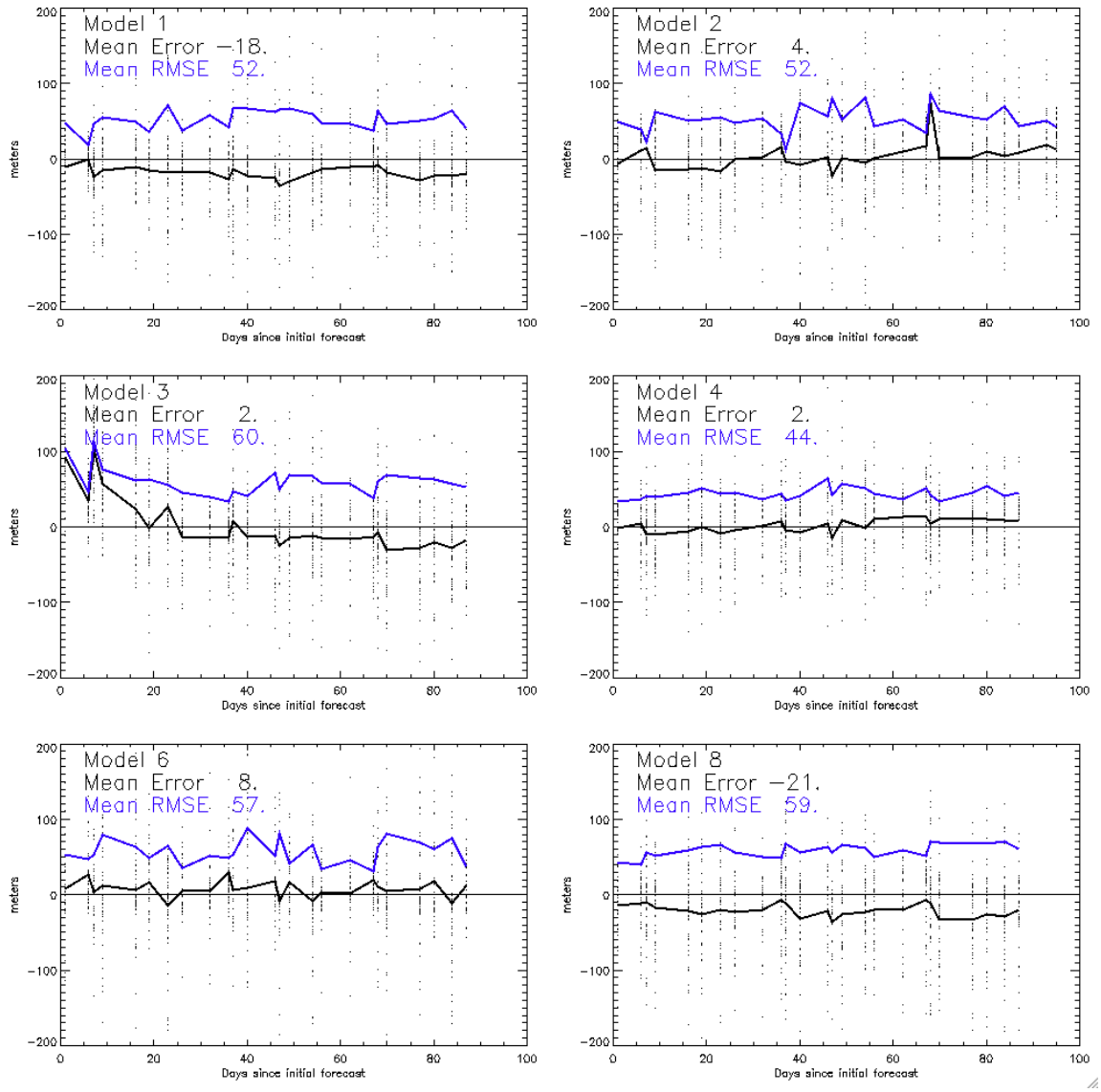


Figure 35: Subsurface Thermal Field — depth of 20°C isotherm, Step 2.

Similar results were obtained from a comparison of 30 to 360m temperature difference. Time series of indicate no skill (Fig. 36), e.g., the ratio of RMSE/StdDev unity or larger.

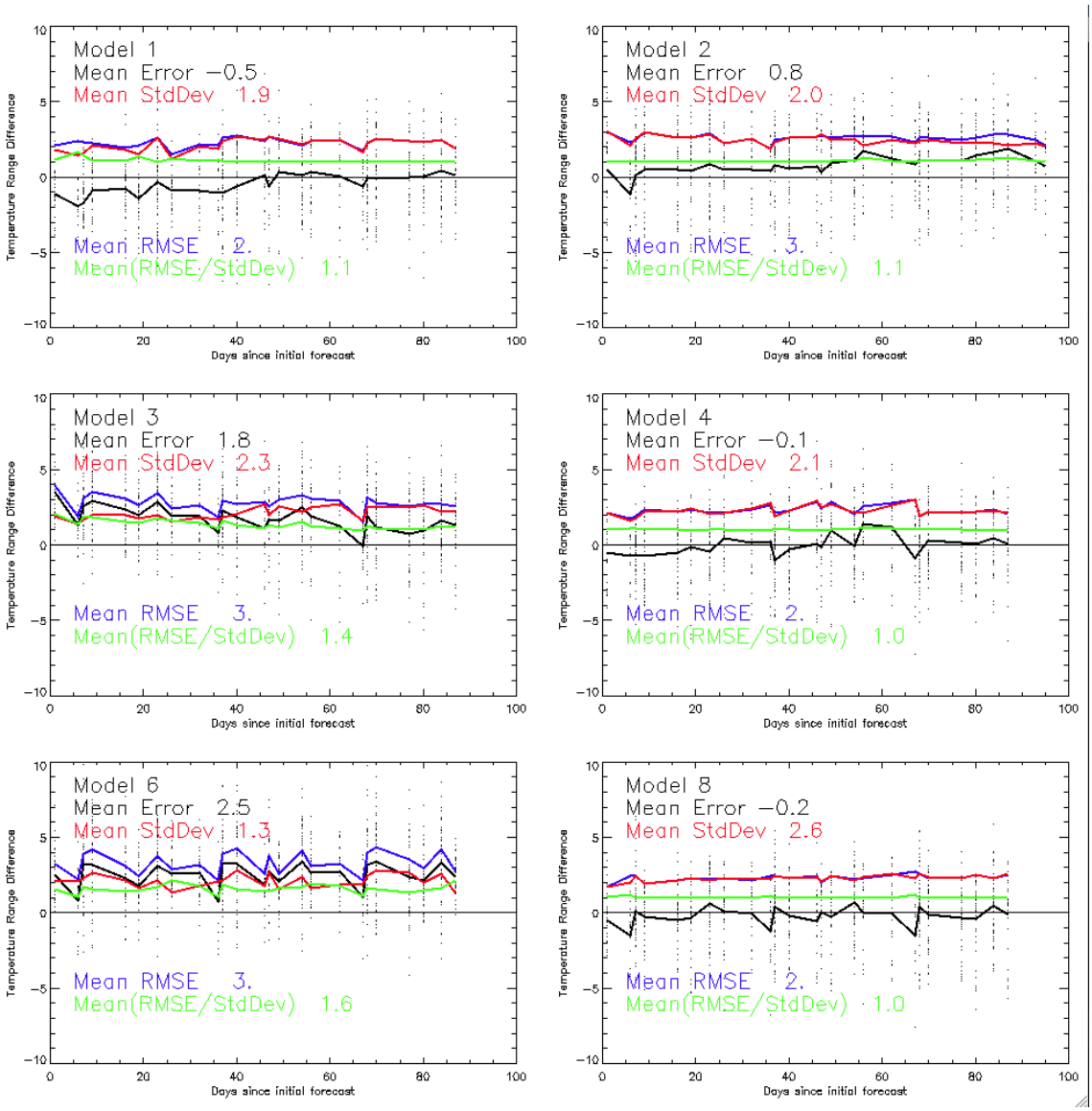


Figure 36: Subsurface Thermal Field — T(30m)-T(360m) temperature difference, Step 2.

Step 3

Due to the low model skill observed in Step 2 forecasts, no attempt was made to validate the models subsurface thermal fields in Step 3.

Summary

An examination of the subsurface thermal fields from the Step 1 retrospective nowcasts indicates that all the models, except Model #7, share the same gross thermal structure as the observations, in that the average depth of the 20°C isotherm is realistic. Root-mean-square error in the depth of that isotherm is 30 to 50m in the models, which may be compared with the approximately 120m change in isothermal depth that occurred between the periphery and center of the LC/LCE

surveyed around the Deepwater Horizon incident. Based on the configuration of the anti-cyclonic LC/LCE in Shay et al (2011), this error corresponds to a horizontal error of some 20 to 200km, depending on where it occurred.

A second intercomparison using the temperature difference between upper and lower layer water is compatible with the error estimated from a single isotherm. The comparison fails to identify the model with known anomalous thermal structure, though, making its usefulness uncertain. If the metric is taken as a measure of the models' ability to accurately locate subsurface thermal structure, and associated fronts, it would appear that the nowcasts of all the models have little skill.

Comparisons of Step 2 forecasts with the AXBT observations are consistent with the above conclusion. Errors in the initial conditions for the subsurface thermal field are so large that forecasts have no skill. Although gross features of the subsurface thermal field are present, i.e., the temperature contrast between the GOMEX and Caribbean waters is roughly accurate, forecasting the location of the LC/LCE thermal front, and smaller-scale cyclonic eddies, is not possible with the GOMEX-PPP systems considered.

GOMEX Surface Velocity Field

Introduction

Comparisons of the model velocity field with observations during the interoperability test, Step 0, were not favorable. Project participants described similar findings in previous experiments. The velocity field is variable on smaller space scales and at shorter time scales than the modeling systems can credibly resolve, and it was not clear what would constitute a fair comparison of model and data. Nonetheless, the velocity field is in some ways the most significant prognostic quantity in an ocean forecast, since advection is the dominant process for redistributing momentum, heat, salt, mass, and passive constituents within the subsurface GOMEX. The velocity field is used as input to oil spill dispersion models that are used by managers and stakeholders (see e.g., Liu et al. 2011 and other articles in that volume). Thus, nowcast and forecast accuracy of the velocity field is of direct societal relevance.

Two types of velocity data were used to compare with model velocity fields: Lagrangian surface drifters and moored current sensors.

Lagrangian drifter trajectories

Fifty-two drifting buoy (drifter) trajectories were compared to synthetic Lagrangian drifter trajectories generated using daily model surface velocity fields. The real drifters were all drogued at or very near the sea surface and location data were available hourly. Due to a data extraction artifact, the drifter records were truncated to the end of August 2010. However, due to the intense sampling following the Deepwater Horizon incident, there were many multi-month records to use in this study. There is little to be gained from drifter records longer than about two weeks as the separation distances become much larger than correlation length scales. Drifter trajectories that were obviously bad, such as being onboard a ship, were removed. No other quality control or drifter correction/filtering were applied.

Much attention has been directed toward improving modeled trajectories. For example, interpolation of daily model fields to 3-hourly time series with a fourth-order Runge-Kutta scheme for integration. However, in the interest of time, the synthetic drifters were advected with daily model velocities interpolated to the locations of the synthetic drifters using a polynomial interpolation scheme. The seeding location was determined by finding the real drifter's location at the first noon following deployment. Model output is a daily block average centered at noon.

The time series of separation distance were computed between the real and synthetic drifters for every day the real drifter trajectory was available in Step 1 and for every day the drifter trajectory was present within the three-month window of Step 2. The evaluation of the metric is simple. Smaller is better and zero separation distance means the model is perfect.

Lagrangian trajectories of surface drifters were computed using the GOMEX-PPP models, and the subsequent separation of the observed and model-based trajectories was calculated for the Step 1 nowcast (Fig. 37). In every case the separation between the nowcast-based and observed trajectory grows to 200km in about 5 days. Trajectory divergence is roughly linear during this phase, and one can infer a trajectory separation speed on the order of 0.5m/s. Currents of 2m/s can occur in association with the LC/LCE front; although, values of 0.2 to 0.5m/s are more typical away from the main jet (Shay et al., 2011).

It is possible that the observed average growth rate of the separation is dominated by errors in the location of the main LC/LCE jet, which could account for 0.5m/s speed differences. The lack of spatial or temporal stationarity of the velocity field statistics suggests that Lagrangian path statistics might be more fruitfully analyzed within specific regions or during specific events. Drifter observations may not be dense enough to adequately sample the velocity field on the scales at which it can be accurately modeled by any of the GOMEX-PPP systems.

Composite results for all drifters for Step 1 (black lines) and for the three runs of Step 2 with April (red), May (green), and June (blue) initialization dates were determined (Figure 38). Note: Models 5 and 7 did not participate in the STEP 2 trials. The drifters were deployed at various times beginning in May. The simultaneous termination at the end of August is due to the data extraction artifact previously mentioned. The simultaneous terminations of the STEP 2 (colored lines) are due to the end of the 3-month long STEP 2 forecasts. Unless a drifter leaves the Gulf of Mexico there is a limit on how large separation distances can get. In any event, the results are probably only useful 10 days forward before the separation distances grow rapidly.

The mean separation distance for Step 1 and three Step 2 model runs were determined, and the results plotted for each drifter (Fig. 39). The blue line appearing in June is due to drifters which were deployed prior to the STEP 2 model initialization time. These drifters are re-seeded at their location at the first step of the forecast and separation distance begins again at zero. Separation distances due to model error are expected to be smaller the closer to the model initialization time. Thus one might expect the order of the colored lines for small to large separation distance would be black < red < green < blue. Interestingly, for early time the hindcast separations are always larger than the forecasts (Fig. 40) Model 5 has the best performance, by a small margin, over the first 10 days.

The normalized cumulative Lagrangian separation distance introduced by Liu and Weisberg (2011) was computed for Step 1 synthetic trajectories. In this metric the cumulative separation distance is normalized by the cumulative trajectory length with the reported advantage of being useful in evaluating models with both strong and weak dynamic regions such as the Loop Current and the West Florida shelf. Smaller numbers are better. Values of S greater than 1 indicate no model skill. Based on a 3-day accumulation (Table 24) the S values from GOMEX-PPP are larger than those reported by Liu and Weisberg (2011) for the same period and region for the Global HYCOM hindcast output. It is not clear if the present method of generating synthetic drifter trajectories is inferior or if other issues explain the relatively larger numbers. In relative terms, judging by the mean scores all models, except for Model 7, did equally well. None of the mean values, however, were less than 1.

A velocity vector correlation coefficient (ρ^2_v) defined by Crosby et al. (1993) was computed between the velocity time series of 47 drifting buoys and the modeled surface velocities at the drifters' locations for Step 1. The code used to perform the calculation was adapted from Crosby et al. (1990). The definition is a generalization of the square of the one-dimensional correlation coefficient and has a simple interpretation in terms of correlation, $\rho^2_v = 0$ means uncorrelated, $\rho^2_v = 2$ means perfect correlation. The results for Step 1 (Table 25) show little correlation for any model, with values well under 2.0.

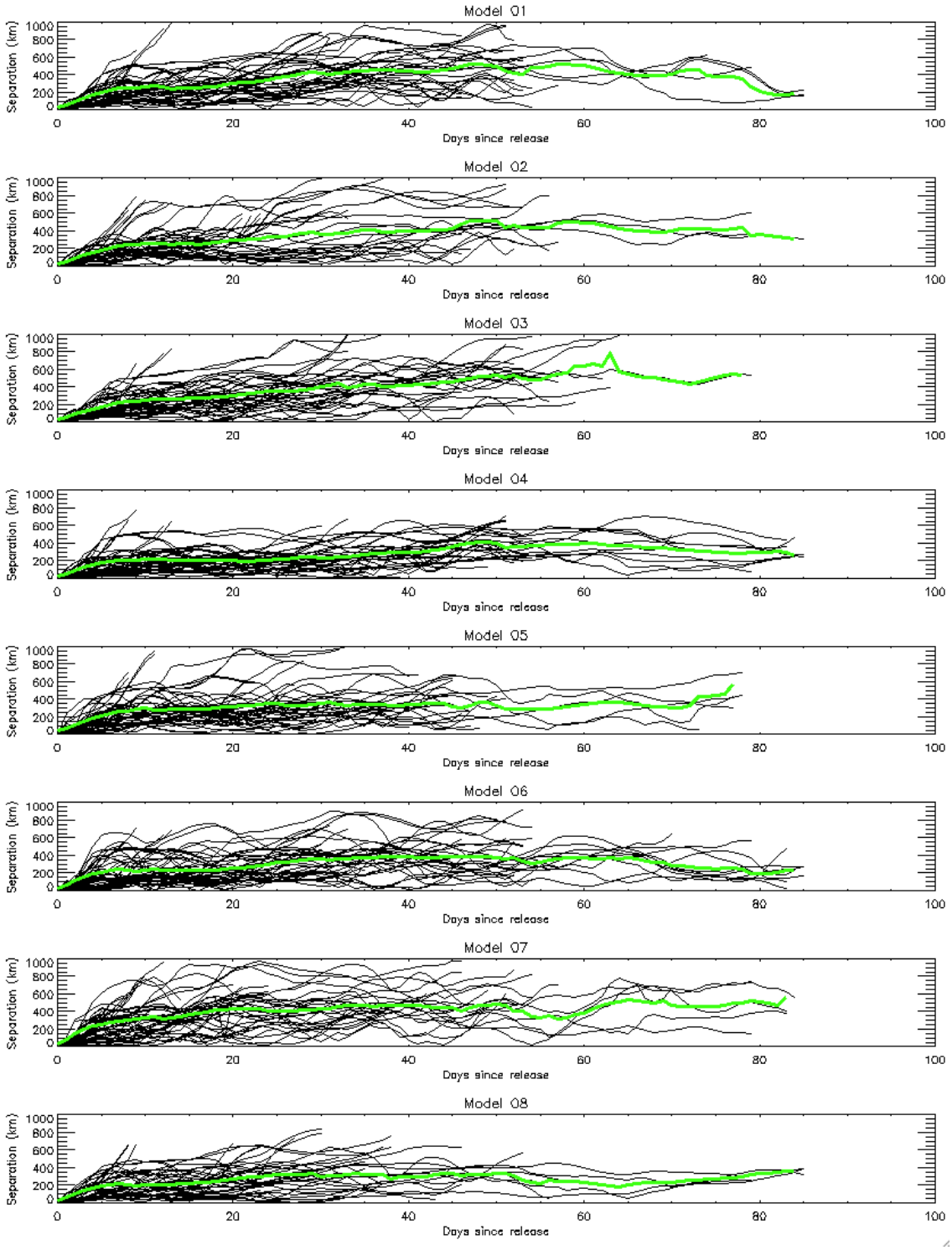


Figure 37: Observed vs. Modeled Lagrangian Trajectory Separation, Step 1. Green curves are the average of the individual trajectories (black).

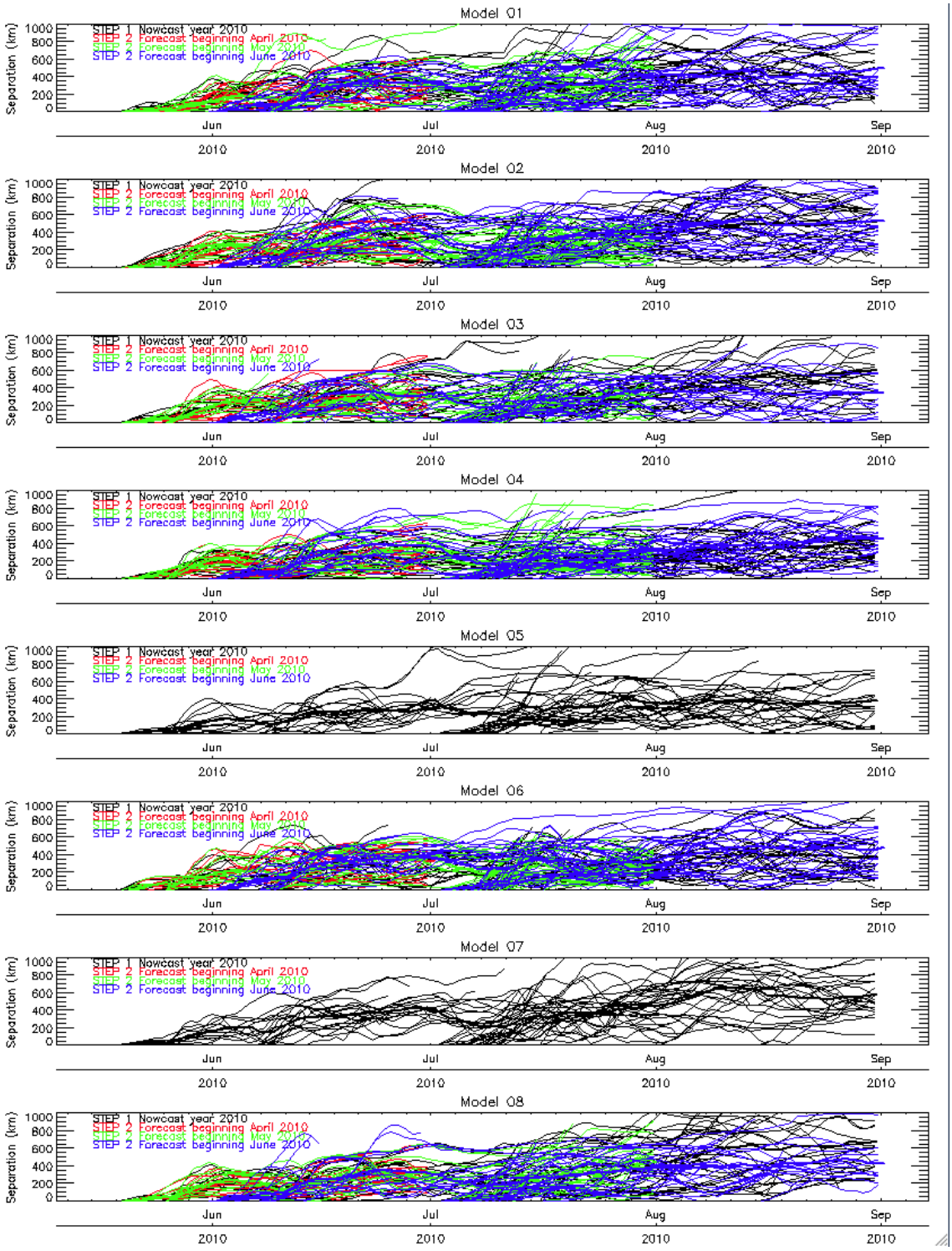


Figure 38: Composite results for Step 1.
Separation statistics for all drifters (Step 1) and three runs of Step 2 with April, May, and June initialization dates.

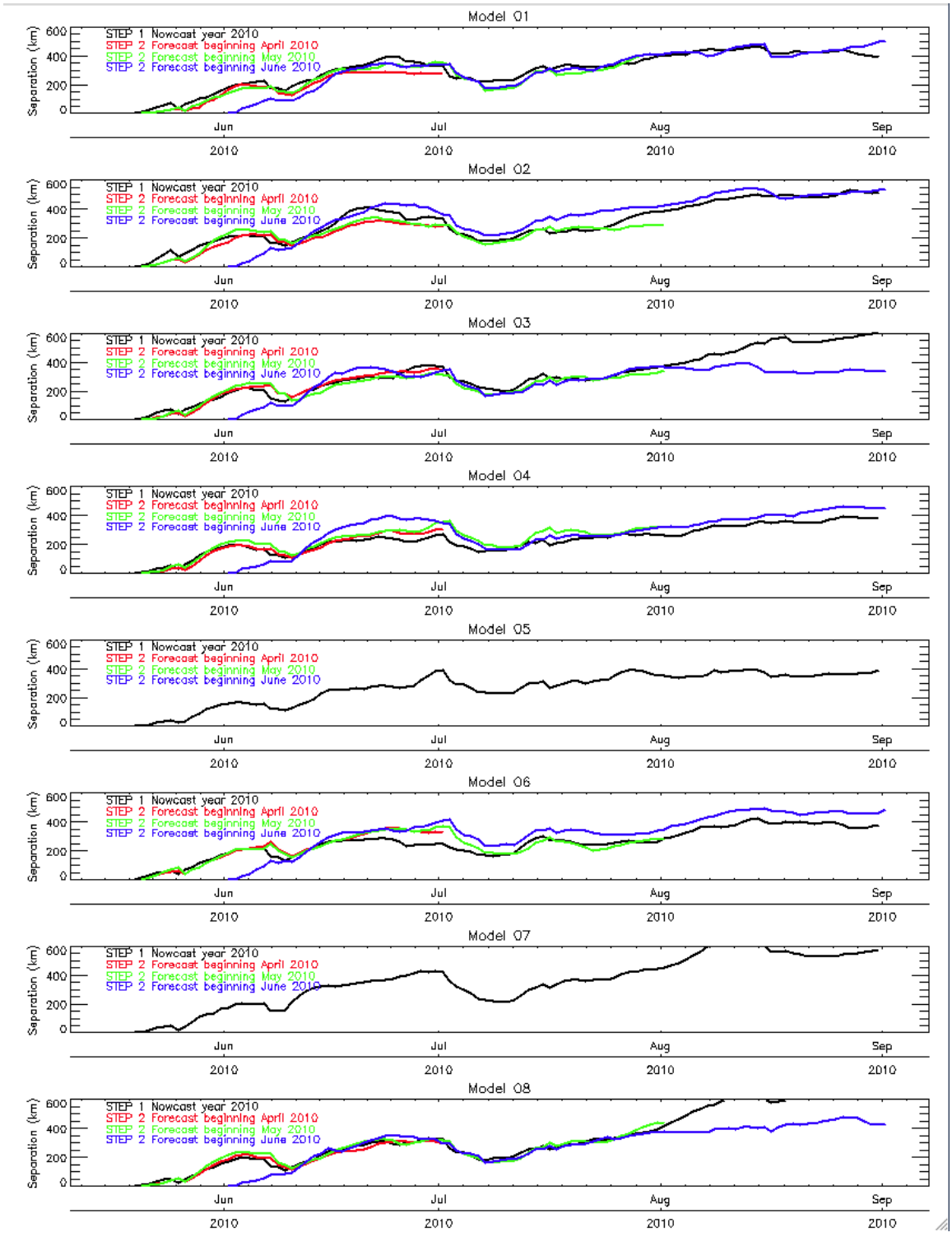


Figure 39: Mean separation distances for the Step 1 and three Step 2 results.

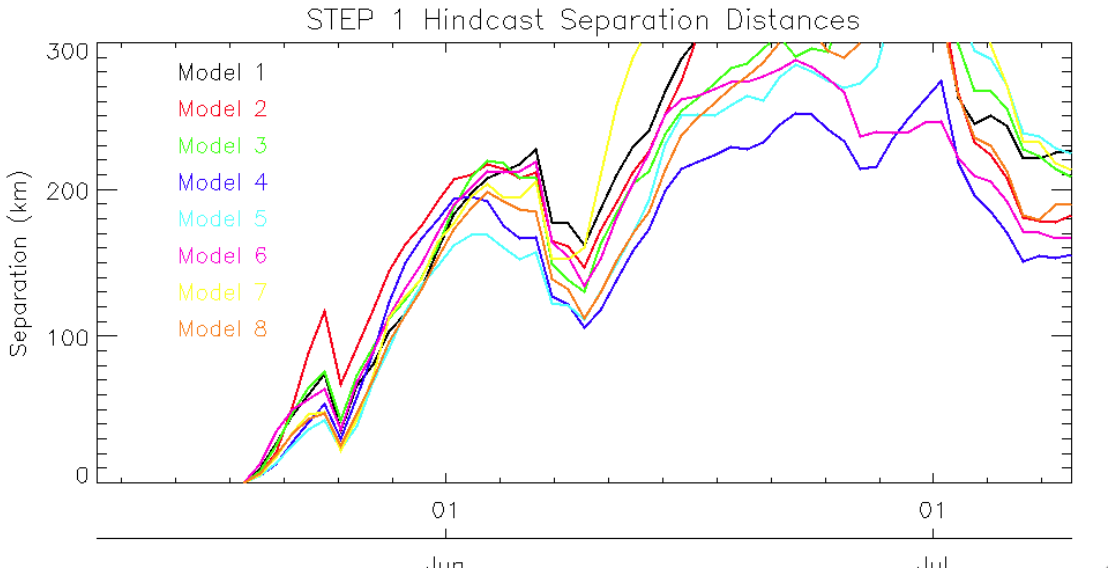


Figure 40: Step 1 results for all eight models for almost 60 days.

Table 24: Drifter separation metric.
Normalized cumulative Lagrangian separation distance metric "S" based on Liu and Weisberg (2011) for 47 drifters 3 days after release for STEP 1 model forecasts.

ID	M1	M2	M3	M4	M5	M6	M7	M8	Mean
42503	4.99	6.17	4.52	1.65	4.87	1.62	16.28	8.30	6.05
42504	1.60	1.96	1.51	2.07	1.30	1.36	3.06	1.73	1.82
42505	0.89	0.80	1.82	0.86	1.41	2.71	1.74	2.23	1.56
42506	1.73	1.58	1.52	1.17	2.17	1.29	2.58	2.03	1.76
42508	2.64	1.96	2.67	1.85	2.46	3.50	3.63	1.64	2.54
42512	1.58	1.96	1.45	2.07	1.25	1.32	2.85	1.71	1.77
42513	0.97	2.58	2.22	1.69	1.76	2.82	1.49	0.94	1.81
42514	0.97	2.57	2.22	1.69	1.75	2.80	1.49	0.94	1.80
42515	1.15	1.46	1.40	1.22	1.07	1.45	1.76	1.49	1.38
42516	0.84	1.05	0.92	1.03	0.95	2.36	1.39	1.39	1.24
42518	1.74	1.14	1.30	1.76	1.48	1.70	2.17	1.61	1.61
42519	1.06	1.40	1.32	1.13	1.01	1.37	1.69	1.41	1.30
42520	1.88	2.27	2.42	2.49	3.33	1.62	3.28	1.24	2.32
42521	1.69	1.65	1.97	2.07	2.46	2.68	3.56	7.02	2.89
42524	0.70	0.70	0.54	0.52	0.58	1.20	1.84	3.30	1.17
42525	1.72	1.58	1.51	1.17	2.17	1.29	2.53	2.02	1.75
42526	4.83	6.17	4.34	1.41	4.85	1.54	16.29	8.34	5.97
42533	2.36	1.90	2.29	1.48	1.39	2.48	1.38	1.71	1.87
42535	1.59	1.29	1.62	0.42	2.67	1.44	3.19	2.26	1.81
42536	5.11	6.76	4.34	2.46	5.68	3.61	9.46	2.37	4.97
42538	1.93	1.04	1.58	2.05	1.37	1.75	2.79	1.57	1.76
42540	1.20	1.72	1.70	1.57	1.78	1.51	2.78	1.27	1.69
42542	1.76	1.16	1.26	1.77	1.47	1.69	2.16	1.61	1.61
42546	3.48	2.30	4.18	1.38	3.09	1.07	3.78	1.89	2.65
42547	1.86	1.00	1.48	1.97	1.31	1.69	2.66	1.52	1.69
42548	1.75	2.20	2.26	2.36	3.27	1.75	3.54	1.27	2.30
42549	3.34	4.30	1.48	4.43	1.23	4.77	2.36	0.79	2.84
42550	3.58	3.01	2.99	3.04	4.09	3.57	3.77	1.37	3.18
42551	1.77	3.15	1.68	0.95	1.17	1.21	4.59	2.17	2.09
42552	1.73	1.34	1.78	0.38	2.99	1.51	3.58	2.52	1.98
42554	2.25	2.54	1.48	1.63	1.84	1.02	2.63	2.39	1.97
42555	0.86	1.15	0.92	1.06	0.95	2.30	1.39	1.40	1.25
42556	6.95	4.85	5.28	4.08	8.43	4.94	20.89	4.62	7.50
42557	5.56	1.38	2.91	1.74	2.42	1.09	9.08	3.07	3.41
42558	5.96	7.60	5.86	2.89	7.08	4.73	13.27	2.44	6.23
42559	0.63	0.24	0.60	0.57	0.40	1.45	1.49	3.40	1.10
42560	2.17	3.03	2.21	0.58	1.37	1.13	2.82	2.35	1.96
42570	1.20	1.24	1.32	1.55	1.37	1.30	1.46	1.26	1.34
42571	2.55	4.47	2.37	1.55	1.26	2.81	0.96	1.91	2.24
42572	2.92	2.34	2.89	1.47	1.69	3.91	1.73	1.90	2.36
42573	3.48	3.75	2.61	1.27	1.11	1.21	5.50	2.82	2.72
42574	1.47	1.33	1.26	1.06	1.81	1.39	1.72	1.27	1.41
42575	1.80	1.24	1.05	1.92	1.56	2.99	0.72	1.95	1.65
42576	6.21	1.82	8.14	18.45	2.44	3.55	30.88	9.77	10.16
42577	4.21	1.86	2.38	6.93	1.86	3.93	3.10	2.33	3.33
42578	2.70	4.40	2.85	1.70	1.68	3.34	1.93	1.50	2.51
42579	4.70	3.29	4.08	1.49	1.98	4.42	3.27	3.60	3.35
Mean	2.51	2.44	2.35	2.13	2.25	2.26	4.61	2.50	

Table 25: Velocity correlation.
Velocity vector correlation coefficients (ρ^2_v)between drifting buoys and co-located 2010 retrospective nowcasts (STEP1) model output for eight numerical models (M1-M8). The coefficient definition is from the method of Crosby et al. (1993). $\rho^2_v = 0$ means uncorrelated. $\rho^2_v = 2$ means perfect correlation.

Drifter	M1	M2	M3	M4	M5	M6	M7	M8	Mean
15659	0.46	0.42	0.75	0.68	0.30	0.67	0.67	0.82	0.60
31754	0.59	0.73	0.72	0.14	0.64	0.12	0.12	0.14	0.40
42503	0.54	0.19	0.19	0.19	0.19	0.19	0.19	0.14	0.23
42504	0.76	0.82	0.90	0.66	0.92	0.47	0.47	0.64	0.70
42505	0.35	0.53	0.59	0.59	0.34	0.27	0.27	0.43	0.42
42506	0.55	0.59	0.52	0.39	0.78	0.43	0.43	0.38	0.51
42508	0.76	0.56	0.73	0.54	0.95	0.54	0.54	0.63	0.66
42512	0.42	0.42	0.62	0.69	1.02	0.32	0.32	0.50	0.54
42513	0.18	0.06	0.03	0.03	0.72	0.44	0.44	0.19	0.26
42514	0.59	0.50	0.75	0.18	0.70	0.44	0.44	0.17	0.47
42515	0.56	0.54	0.74	0.18	0.79	0.33	0.33	0.22	0.46
42516	0.85	0.78	0.95	0.94	0.96	0.86	0.86	0.90	0.89
42518	0.80	0.94	0.83	0.28	0.92	0.53	0.53	0.33	0.64
42519	0.46	0.54	0.63	0.43	0.74	0.59	0.59	0.45	0.55
42520	0.46	0.47	0.41	0.40	0.48	0.34	0.34	0.39	0.41
42521	0.09	0.23	0.13	0.25	0.15	0.16	0.16	0.26	0.18
42524	0.51	0.61	0.48	0.38	0.65	0.33	0.33	0.34	0.45
42526	0.68	0.49	0.72	0.25	0.73	0.46	0.46	0.48	0.53
42533	0.31	0.29	0.28	0.24	0.20	0.40	0.40	0.32	0.31
42536	0.57	0.70	0.65	0.29	0.65	0.54	0.54	0.36	0.54
42538	0.13	0.41	0.19	0.20	0.52	0.26	0.26	0.38	0.29
42540	0.12	0.17	0.09	0.12	0.12	0.10	0.10	0.13	0.12
42542	0.61	0.74	0.77	0.38	0.76	0.36	0.36	0.31	0.54
42546	0.34	0.44	0.35	0.26	0.49	0.21	0.21	0.41	0.34
42547	0.30	0.34	0.44	0.33	0.47	0.15	0.15	0.34	0.31
42548	0.13	0.13	0.12	0.13	0.13	0.14	0.14	0.12	0.13
42549	0.96	0.90	0.96	0.80	0.93	0.80	0.80	0.92	0.88
42550	0.43	0.75	0.73	0.63	0.32	0.48	0.48	0.53	0.54
42551	0.11	0.19	0.12	0.35	0.20	0.27	0.27	0.34	0.23
42552	0.32	0.34	0.28	0.13	0.34	0.36	0.36	0.17	0.29
42554	0.19	0.10	0.05	0.26	0.06	0.15	0.15	0.27	0.15
42555	0.99	1.24	1.23	1.24	1.17	1.18	1.18	1.27	1.19
42556	0.64	0.89	0.62	0.41	0.64	0.41	0.41	0.33	0.54
42557	0.71	0.65	0.65	0.55	0.63	0.53	0.53	0.56	0.60
42558	0.36	0.62	0.63	0.24	0.70	0.64	0.64	0.22	0.51
42559	0.87	0.89	0.98	0.22	0.97	0.82	0.82	0.43	0.75
42560	0.41	0.54	0.55	0.42	0.41	0.41	0.41	0.32	0.43
42570	0.74	0.75	0.82	0.45	0.80	0.83	0.83	0.50	0.71
42571	0.08	0.08	0.11	0.11	0.05	0.11	0.11	0.08	0.09
42572	0.21	0.20	0.13	0.30	0.13	0.15	0.15	0.29	0.19
42573	0.19	0.23	0.16	0.10	0.25	0.40	0.40	0.21	0.24
42574	0.53	0.46	0.52	0.24	0.30	0.42	0.42	0.28	0.40
42575	0.79	0.72	0.83	0.47	0.89	0.58	0.58	0.48	0.67
42576	0.59	0.46	0.60	0.26	0.55	0.56	0.56	0.20	0.47
42577	0.90	0.82	0.98	0.65	0.80	0.58	0.58	0.50	0.73
42578	0.41	0.34	0.12	0.28	0.21	0.42	0.42	0.24	0.31
42579	0.47	0.18	0.59	0.46	0.53	0.43	0.43	0.30	0.42
Mean	0.49	0.51	0.54	0.38	0.56	0.43	0.43	0.39	

Moored Currents

SAIC Inc. deployed nine tall moorings in the region of the Loop Current under contract to BOEM. Each mooring held single point current meters at various depths and an upward-looking ADCP (Acoustic Doppler Current Profiler) at 450m. The moorings were deployed in May 2009, serviced in July or November 2010, and recovered in November 2011. Data were recorded internally at hourly intervals and were not available in near real-time. GOMEX-PPP was given permission by BOEM to use preliminary data from the service cruises for the model-data comparisons. SAIC estimated that data from the final recovery cruise would not be available until February 2012 — too late for use in this comparison. The current velocity data from the single-point, upward looking ADCP instruments for the period January to July 2010 were compared to the eight Step 1 model outputs. The metric computed, "M", is the root-mean-square error (model-data) normalized by the mean square value of the observed velocity component $<(u^2 + v^2)>$ (Table 26). The numbers are a time average over the period and, in the case of ADCP data, also include averaging over the vertical extent of the ADCP record. ADCP data occur in the first nine rows of the table where the instrument depth equals 450m. The remainder of the table entries are for single-point measurements. Model means for each current meter and model means over all instruments appear in the right-most column or the bottom row, respectively.

The metric M is defined as follows, where Um and Vm are model velocity components, Ud and Vd are observed velocity components and < > indicates the mean value.

$$M = \sqrt{\frac{< (Um - Ud)^2 + (Vm - Vd)^2 >}{< (Ud^2 + Vd^2) >}}$$

This can be interpreted as the magnitude of the error vector over the magnitude of the observed current vector. The perfect model would receive a score of zero. Model skill can be defined as 1-M. A skill of 1 is a perfect model, zero means model errors are as large as the signal. Negative values indicate the model had little to no skill. In the following table smaller is better and numbers less than 1 indicate a model with some skill. In general, models did better above 500m, which may be due to the assimilation of SSH data which dominates the surface fields. Model 1, 4, and 5 did better than others overall depths.

Table 26: Model skill.
Metric M Differences Between SAIC Meters/ADCP and Models

Record	Lon	Lat	IDepth	WDepth	M1	M2	M3	M4	M5	M6	M7	M8	Mean
aa151c_d	-88.05	25.96	450.0	3035.0	0.84	1.17	1.05	1.03	0.88	1.03	1.45	0.98	1.05
aa251c_d	-87.55	25.82	450.0	3200.0	0.61	1.03	0.69	1.24	0.98	0.91	1.31	0.88	0.96
aa351c_d	-87.05	25.68	450.0	3293.0	0.88	0.89	1.02	1.42	0.99	1.01	2.06	1.29	1.19
aa451c_d	-86.56	25.49	450.0	3262.0	1.36	1.10	1.31	1.12	1.02	1.41	1.77	1.25	1.29
ab151c_d	-87.32	26.25	450.0	3049.0	1.14	1.12	1.17	1.01	0.92	0.99	1.53	0.92	1.10
ab251c_d	-86.84	26.11	450.0	3132.0	0.69	1.09	0.76	1.28	0.90	0.82	1.70	0.66	0.99
ab351c_d	-86.36	25.93	450.0	3156.0	0.84	0.89	0.91	1.15	1.00	0.95	1.11	0.82	0.96
ac151c_d	-86.14	26.38	450.0	3182.0	0.67	0.89	0.81	0.77	1.14	0.80	1.06	0.77	0.86
ac251c_d	-85.65	26.18	450.0	3250.0	0.61	0.63	0.61	0.58	1.01	0.79	0.88	0.64	0.72
ac291c_1	-85.65	26.18	900.0	3250.0	8.48	8.04	7.40	5.08	1.84	7.32	9.52	6.50	6.77
aa191c_1	-88.05	25.96	900.0	3035.0	2.46	4.59	6.23	1.95	1.69	3.02	3.54	2.32	3.23
ab391c_1	-86.36	25.93	900.0	3156.0	1.64	3.97	4.86	1.88	1.24	2.17	3.83	2.84	2.80
ab291c_1	-86.84	26.11	900.0	3132.3	5.17	6.07	10.15	4.66	1.58	2.63	6.09	5.00	5.17
aa391c_1	-87.05	25.68	900.0	3293.0	2.30	2.69	4.93	2.47	1.11	2.21	2.19	3.66	2.70
ab191c_1	-87.32	26.25	900.0	3048.9	1.94	2.70	5.53	1.61	3.45	1.63	2.74	2.70	2.79
aa3b1c_1	-87.05	25.68	1300.0	3293.0	3.85	7.70	9.38	3.46	1.05	4.82	5.17	6.47	5.24
aa4b1c_1	-86.56	25.49	1300.0	3262.0	3.22	7.96	9.90	4.00	1.30	4.76	4.52	6.21	5.23
ab1b1c_1	-87.32	26.25	1300.0	3048.9	2.37	3.26	7.28	2.02	2.74	2.17	4.70	3.56	3.51
aa1b1c_1	-88.05	25.96	1300.0	3035.0	2.44	5.35	6.28	3.05	2.90	3.89	3.74	2.87	3.81
ab2b1c_1	-86.84	26.11	1300.0	3132.3	3.26	3.80	6.24	2.27	1.39	1.72	3.93	3.20	3.23
ab3b1c_1	-86.36	25.93	1300.0	3156.0	2.19	6.54	6.66	2.73	1.49	2.51	4.46	4.47	3.88
ab1d1c_1	-87.32	26.25	1990.0	3048.9	1.48	1.83	2.58	1.22	1.93	1.82	2.56	1.57	1.87
ac2d1c_1	-85.65	26.18	2000.0	3250.0	1.58	2.72	2.36	1.51	1.21	1.74	2.66	1.37	1.89
ac1d1c_1	-86.14	26.38	2000.0	3182.0	1.38	2.70	1.93	1.90	0.97	1.49	2.58	1.56	1.81
ab2d1c_1	-86.84	26.11	2000.0	3132.3	1.81	2.54	2.57	1.50	1.04	1.65	2.37	1.60	1.89
aa4d1c_1	-86.56	25.49	2000.0	3262.0	1.76	4.17	4.57	2.12	0.85	2.85	3.13	2.91	2.79
ab3d1c_1	-86.36	25.93	2000.0	3156.0	1.34	3.76	2.26	1.83	1.24	1.48	2.58	1.76	2.03
aa3d1c_1	-87.05	25.68	2000.0	3293.0	2.12	4.29	4.03	2.18	0.94	2.83	3.33	2.65	2.80
aa2d1c_1	-87.55	25.82	2000.0	3200.0	1.65	4.37	2.41	1.57	1.51	2.42	3.34	1.83	2.39
aa1d1c_1	-88.05	25.96	2000.0	3035.0	1.67	4.41	3.35	2.34	1.46	3.07	3.37	1.85	2.69
ad111c_d	-87.85	26.42	2727.0	2827.0	1.28	1.99	2.28	1.14	1.39	1.36	2.34	1.44	1.65
ad121c_d	-87.85	26.40	2728.0	2828.0	1.34	2.05	2.63	1.20	1.36	1.43	2.54	1.59	1.77
ad211c_d	-87.13	26.70	2785.0	2885.0	1.35	1.74	1.55	1.20	1.75	1.23	1.95	1.23	1.50
aa1e1c_1	-88.05	25.96	2935.0	3035.0	1.52	4.24	2.76	2.20	1.24	2.60	3.56	1.73	2.48
ab1e1c_1	-87.32	26.25	2948.9	3048.9	1.69	2.08	3.79	1.74	1.76	3.15	3.63	2.08	2.49
ab2e1c_1	-86.84	26.11	3032.0	3132.3	2.01	2.99	0.99	1.85	0.00	1.59	2.98	1.62	1.75
ac2e1c_1	-85.65	26.18	3150.0	3250.0	1.64	2.75	2.17	1.49	1.16	1.51	3.05	1.22	1.87
ab3e1c_1	-86.36	25.93	3056.0	3156.0	1.34	3.36	1.83	2.27	0.00	1.28	2.67	1.61	1.79
ac1e1c_1	-86.14	26.38	3082.0	3182.0	1.37	2.54	1.07	1.66	0.00	1.46	2.74	1.42	1.53
aa2e1c_1	-87.55	25.82	3100.0	3200.0	2.02	3.46	1.41	1.69	0.00	2.22	3.92	1.93	2.08
ad511c_d	-85.86	25.74	3133.0	3233.0	1.04	1.62	1.35	1.26	0.97	1.34	2.13	1.28	1.37
ad411c_d	-86.08	25.30	3157.0	3256.0	1.33	2.08	1.38	1.66	1.04	1.50	2.27	2.18	1.68
aa4e1c_1	-86.56	25.49	3162.0	3262.0	1.92	3.58	3.44	1.88	0.76	2.32	3.11	2.33	2.42
aa3e1c_1	-87.05	25.68	3193.0	3293.0	2.23	4.11	2.63	1.94	1.28	2.54	3.79	2.23	2.59
ad811c_d	-85.15	26.00	3209.0	3309.0	2.00	3.76	1.95	2.08	0.92	0.00	3.01	2.14	1.98
ad711c_d	-85.38	25.55	3211.0	3311.0	2.01	3.29	3.00	2.55	0.96	2.19	4.14	1.68	2.48
				Mean	1.91	3.22	3.34	1.93	1.23	2.06	3.07	2.24	

GOMEX-PPP Web Portal

The project Website is located at http://gomex-ppp.org (Figure 41 shows the main page). The Website is open to the public except for a section containing materials from the project kick-off meeting. The main sections of the website are Models, Project Participants, Data Sets, and Results. The Model page contains links to documents listing model attributes(e.g., model resolution, turbulence closure schemes, boundary conditions etc.) and a summary table of these including links to model Webpages. The Project Participants page lists the members of the management team, subcontractors, other modelers, and the Science Advisory Committee. The "Data Set" pages provide access to downloadable files containing SSH/3D fields from the Global NCOM, monthly climatologies for NOGAPS, daily research grade and operational grade SSH fields from Dr. Bob Leben, river discharge from the 1900s to 2010 for 55 Gulf of Mexico rivers, and hourly water level data for 70 Gulf of Mexico stations.

The bulk of the Website contents are in the "Results" pages. Sub-menus lead to results from the Test Case (June 2010), the main experiments and to PSU's webpages. The "Result" pages contain some of the material appearing in this report. They also contain numerous animations of results which of course could not appear in this report. Animations include side-by-side comparisons of model-data and model-model comparisons. Animated content includes:

- Model SSH/Currents vs drifter trajectories
- Model Currents vs BOEM and SAIC ADCP profiles
- Model Temperature vs AXBT profiles
- Model vs Model potential vorticity
- Model vs Model SSH and SST forecasts

One of the deliverables was to provide Web access to the side-by-side comparisons of fortnightly forecast (Step 3) results together with available observations rapidly after the model results were made available by the modelers — within a day if possible. This was delivered through static Web pages with observed SSH and SST (e.g., Fig. 42). There was discussion in the early part of the project about developing a real-time portal. The purpose was to display model output and data together but with a high-degree of user-driven control over what times, model parameters and data were displayed. Adoption of existing software was considered, but it was decided that software was not mature and difficult to port to another machine. Adoption of software under development by the SURA Testbed activity was also considered, but that activity remained developmental into December of 2011 – too late for use by GOMEX-PPP.

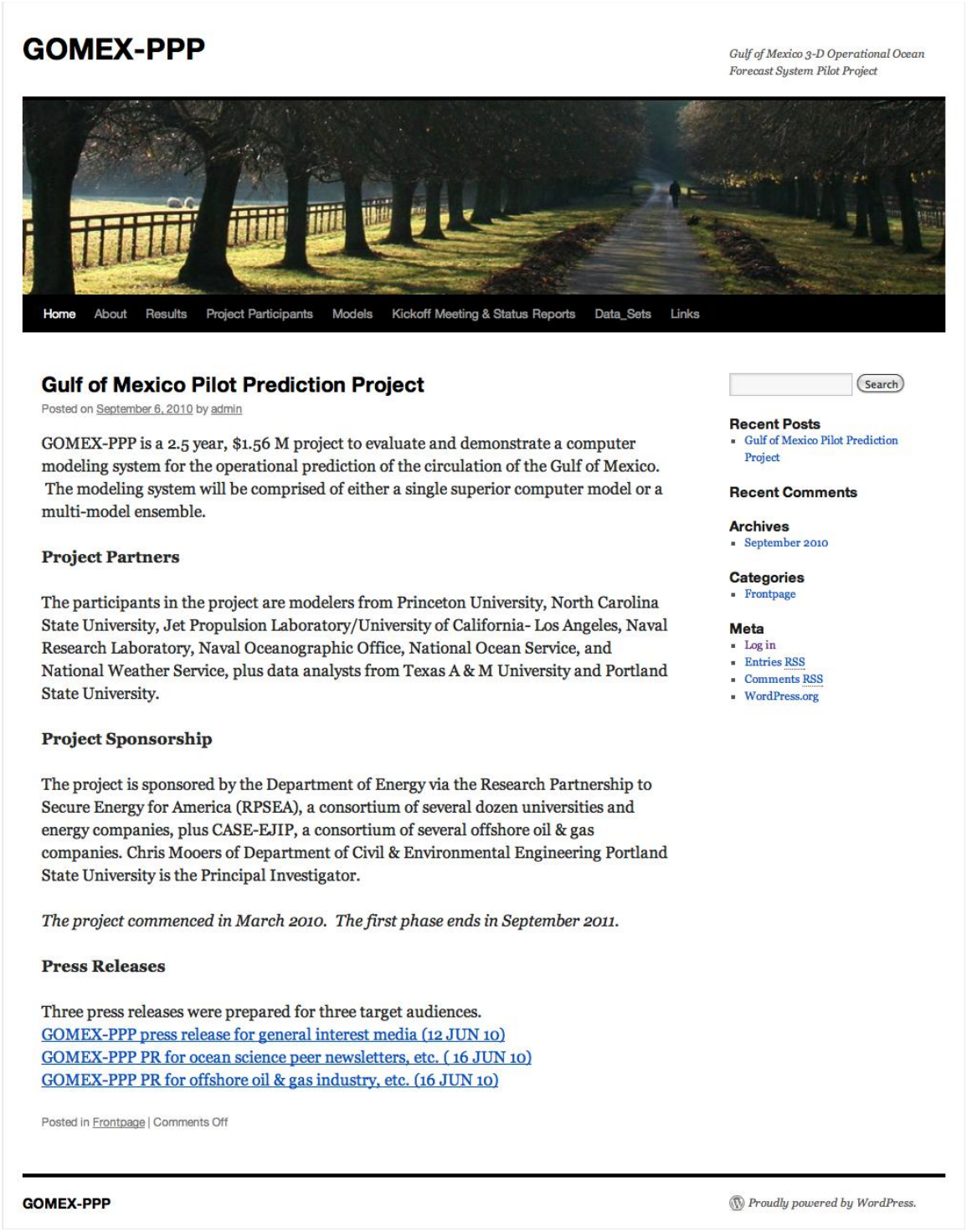


Figure 41: The main page of the GOMEX-PPP Website, http://gomex-ppp.org. It was constructed using WordPress and is hosted on a desktop server at Texas A&M University.

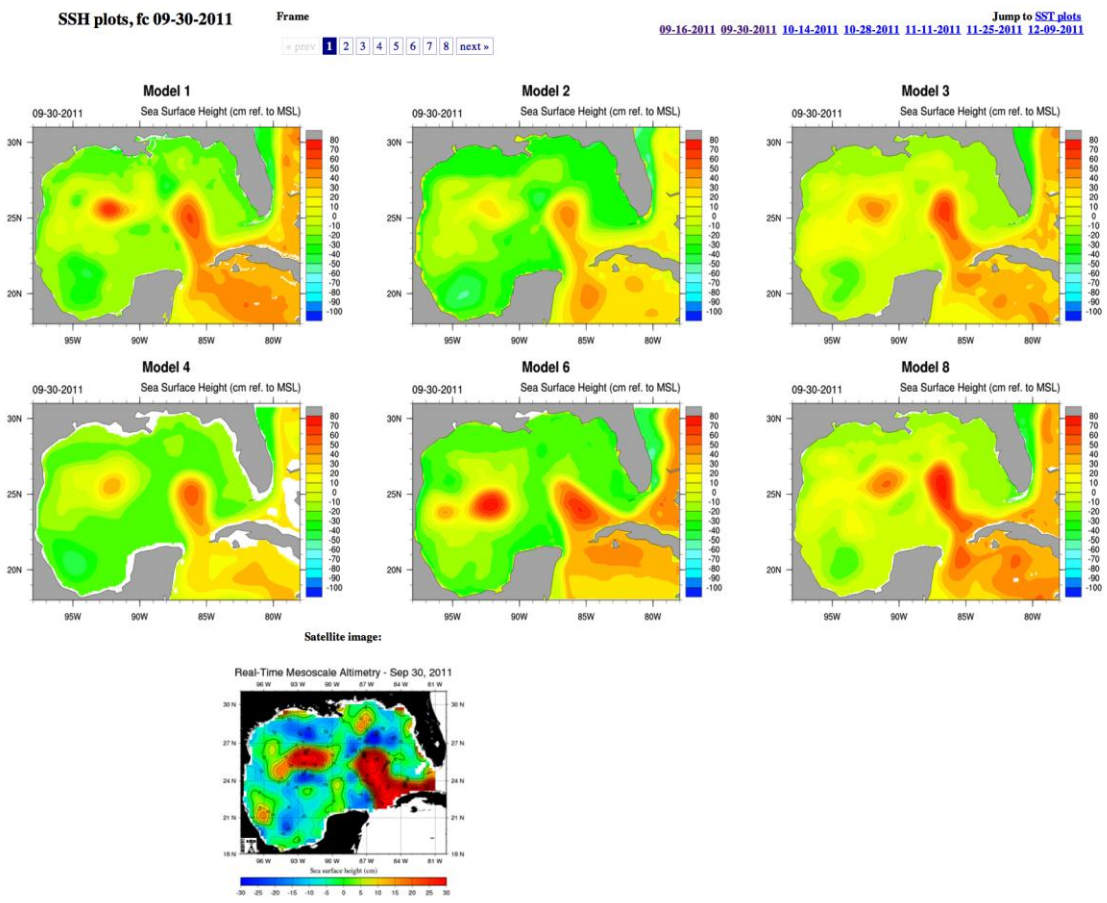


Figure 42: One frame of fortnightly plots of STEP 3 results.
Sea surface height for the eight participating models use shown (upper two rows).
Observed sea surface height for the same date (30-Sept-2011) provided by CCAR (bottom panel).

Impact to Producers: Identification of User Needs and Model Assessment Criteria

Introduction

A survey was conducted of 54 users (24 responded) of environmental predictions for the Gulf of Mexico (GOMEX). The survey respondents represented the offshore oil & gas industry, plus many diverse user types. The focus was on super-users (e.g., marine forecasters) rather than end-users, who are served by super-users, each of whom may serve 10 to a 1,000 end-users. Typical issues concerned prediction variables, and their required accuracy, space-time resolution, and forecast horizons. A further description of the survey and a summary of its results are found in Appendix 1. The survey was apparently unprecedented for GOMEX, and it was conducted jointly by the GOMEX-Pilot Prediction Project (GOMEX-PPP) and the Gulf of Mexico Coastal Ocean Observing System – Regional Association (GCOOS-RA). The survey should be repeated by GCOOS-RA with an updated design in a few years, after a first-generation operational ocean prediction system is established for GOMEX.

The results of the survey indicate an interest in a broad set of products (e.g., those that address continental shelf as well as deepwater GOMEX currents) and a rather vague articulation of qualitative and quantitative requirements, which is understandable until a first-generation operational prediction system is established and evaluated. Hence, a classic "chicken-and-egg" situation exists until a first-generation prediction system becomes operational.

In view of the above ambiguity, professional judgment is utilized to design "model skill assessment criteria" or, more commonly, "model metrics". The metrics are grouped into two categories: (1) scientific, process-oriented metrics that are typically used in the physical oceanographic community as standards for judging the physical validity or fidelity of the model physics and (2) pragmatic, applications –oriented metrics that are tailored to specific user communities.

Model Metrics

As mentioned above, there are two broad categories of metrics considered here: fundamental scientific, process-oriented metrics and pragmatic, application-oriented metrics. The first category relates to understanding the underlying LC & LCE dynamics that presumably control the predictability of the physical (dynamical) system. The second category relates to user needs at "face value".

Examples of the first metric category include:

- Time series of surface wind stress (STX,STY) and surface atmospheric pressure (SAP) maps used in each of the models and their decorrelation.

- Time series of volume transport through the Yucatan Channel (YC) and the Straits of Florida (SF), and their decorrelation times and energy spectra, as compared to the submarine cable-based volume transport estimates from the SF at 27N between Florida and the Bahamas.

- Time series of cross-sections of alongstream flow and mass fields through YC and SF, their low-order 2D-EOFs, and their decorrelation times.

- Time series of SSH, SST, and SSS maps and their decorrelation times, with SSHA compared to AVISIO & "Leben".

- Time series of the spatial mean bias and rmse of the depth of the 20C isothermal surface as compared to AXBT data.

- Time series of the spatial mean bias and rmse of the temperature difference between 30 and 360m as compared to AXBT data.

- Time sequence of LC & LCE 3D-thermal structure analyzed from the sequence of RSMAS/AOML synoptic AXBT surveys compared to model equivalents.

- Ensemble comparisons, including vertical EOFs, of vertical profiles of temperature and salinity from CTD surveys (e.g., on Warlord II tracks and gliders) and model equivalents.

- Animated diagnostics of SSH and Ertel's absolute potential vorticity (PV); e.g., maps of PV on a selected isothermal surface and vertical transects of PV at YC and SF.

- Time series of vertical profiles of currents across the shelf and slope along three cross-shelf transects off the NW Shelf, NE Shelf, and W Florida Shelf, to be assessed from the perspective of potential downscaling activities, and compared to "good" industry and other ADCP current profiles where available.

Examples of the second metric category include:

- Time series of the Closest-Point-of-Approach (CPA) of the LC to DwH and rigs elsewhere, and categorical forecasts, such as forecasts of the date of frontal arrival at a given site. The methodology utilized will be similar to that described in Oey et al., 2005.

- Time series of the difference between Lagrangian observed (e.g., surface drifters) and calculated particle positions, leading to estimated rmse growth rates.

- Comparison of positions of LC & LCE surface fronts and those of Horizon Marine Incorporated (HMI)-determined frontal edges. NOTE: the HMI weekly Eddy Watch maps and frontal edge data will be available only for skill assessment and not data assimilation; hence, only designated members of the analysis groups at TAMU and PSU will have access to this information.

- Overlay of nowcast/forecast versus observed energy spectra for deep currents, especially using SAIC deep current meters.

- Time series of positional parameters for the Mississippi-Atchafalaya River plume.

Conclusion

A representative suite of metrics has been identified that will focus the skill assessment on matters of concern to a broad set of regional users of GOMEX environmental predictions. What can actually be accomplished will depend, in part, on the suitability of the available independent observational data, which is yet to be fully determined. However, as a consequence of the DwH spill, there is an uncommon amount of opportunistic *in situ* environmental data available in 2010, some of which is suitable to model skill assessment. As indicated in the Statement of Work, the set of specific metrics used in the skill assessment is expected to evolve over the course of the project. Overall, it appears that the prospects for rigorous skill assessment of the GOMEX-PPP prediction systems are high.

Technology Transfer Efforts

Background
RPSEA requires that at least 2.5% of the total funding (i.e., ca. $39K, which is rounded up to $40K for convenience below) be assigned to technology transfer.

Activities
The technology transfer activities of GOMEX-PPP are comprised of several elements:

(NOTE: * designates a RPSEA mandate.)

*(1) attend the briefings of RPSEA TAC thru UDW Meetings every four months (YR1/Q2, YR1/Q3, YR1/Q4, YR2/Q2, YR2/Q3, YR2/Q4, YR3/Q2, and YR3/Q3);

(2) make presentations at annual GCOOS-RA and/or SECOORA meetings in YR1/Q2, YR2/Q2, & YR3/Q2;

*(3) present peer-reviewed papers at annual MTS/IEEE OCEANS Meetings in YR2/Q2 and YR3/Q2 and OTC Meetings in YR2/Q1 and YR3/Q1;

(4) submit peer-reviewed papers to the *Journal of Geophysical Research*, *Journal of Operational Oceanography*, and/or *Marine Technology Society Journal* at the end of Phase I (YR2/Q2) and again at the end of Phase II (YR3/Q2);

(5) conduct a GCOOS-RA & SECOORA Stakeholder & Super-User Workshop in YR3/Q1;

(6) maintain and enhance the GOMEX-PPP Website via the GCOOS Website, continuing from YR1/Q1; and

(7) provide a final technical report with recommendations for an operational ocean prediction system for the Gulf of Mexico circulation in YR3/Q2.

Summary
The status of technology transfer activities through Phase I is summarized below:
(1) Accomplished.
(2) Accomplished for GCOOS-RA but not SECOORA because there is enough overlap and cross-talk between GCOOS-RA and SECOORA to make it unnecessary to brief SECOORA separately.
(3) Has not begun to be accomplished for OCEANS and OTC Meetings; however, instead (for stronger peer review), nine oral & poster presentations (see Appendix II) were made at the American Meteorological Society (AMS) and The Oceanography Society (TOS), American Geophysical Union (AGU), and Association for the Sciences of Limnology and Oceanography, Annual & Biennial, resp., Meetings in 2012/Q1.

(4) An overview manuscript covering Phase I is being planned for the Bulletin of the American Meteorological Society (BAMS) now that the Phase I results are in hand; a follow-on manuscript is planned for the end of Phase II.
(5) The Stakeholder & Super User Workshop is still planned for late in Phase II.
(6) The GOMEX-PPP Website is maintained at TAMU and is synergistically linked to the GCOOS-RA Website. It is steadily evolving into a real-time Web-Portal, a Phase II objective.
(7) The final technical report with recommendations regarding an operational ocean prediction system for GOMEX is a Phase II objective.

Prospects for Multi-Model Ensembling

GOMEX-PPP has brought together a suite of modeling systems and applied them to make simultaneous forecasts of the GOMEX. As the foregoing results have shown, the multiple models are not redundant and each makes unique nowcasts and forecasts for the GOMEX state. Can the multiple model nowcasts and forecasts be combined in a manner which permits increased fidelity? This is the challenge of multi-model ensemble state estimation and forecasting.

Four possible roles are envisioned for ensemble techniques: (1) identifying a "consensus forecast" or other measure of the central tendency of the nowcasts/forecasts, (2) estimating forecast errors, (3) generating probabilistic nowcast/forecasts, and (4) generating new forecasts via ensemble smoothers.

Identifying the Consensus Forecast

If the constituent models generate equally plausible realizations of the GOMEX present and future states, then a measure of the central tendency of the ensemble will provide an improved state estimate, provided the probability distribution of GOMEX states is not multi-modal. In other words, if the models are skillful, and if the state evolution is sufficiently linear, then the average, or consensus, forecast should perform better than any constituent forecast.

A priori, one might expect that the approach just described will be unsuccessful in the GOMEX. Nonlinearity inherent in long-range forecasting may cause the mean state to be less likely than any particular realization. For example, assume one has two ensemble members, #1 and #2, which forecast the SSH field on a certain date. Suppose #1 predicts the LC front will be at 26N, while #2 predicts it will be at 28°N. Even if both forecasts are equally likely, a mean forecast formed by simply averaging the SSH fields from #1 and #2 will be very unlikely, as it would consist of a LC with two less-intense fronts at both 26°N and 28°N.

The counter argument to the above is that, on some scales, GOMEX dynamics are very linear. Hence, the multiple models might effectively sample a range of plausible states representing, say, linear barotropic response to strong winds. Forming the mean (or weighted mean, if the model skills vary) may indeed yield improvement for those scales or processes which are governed by linear dynamics.

The skill of consensus forecasts during Step 3 has been investigated by computing that linear combination of constituent models which approximately optimizes the forecast skill. If it is assumed that the variance of the models and observed data are stationary, if the constituent models are equally correlated with the observations, if the variance of the models equals the data variance, and if the model errors are uncorrelated, then the consensus forecast with maximum skill is the equally-weighted average of the constituent models. The Taylor diagrams (Fig. 23 and 24) indicate that models #1, #2, #3, and #4 fall within the same range of correlation with respect to observations of sea-surface height (0.8 to 0.9) and sea-surface slope (0.6 to 0.8). The variance of the constituent models is low, on average, but it does not differ greatly amongst these models. With the caveat that the forecast may be slightly suboptimal, the equally-weighted consensus forecast has been computed during Step 3.

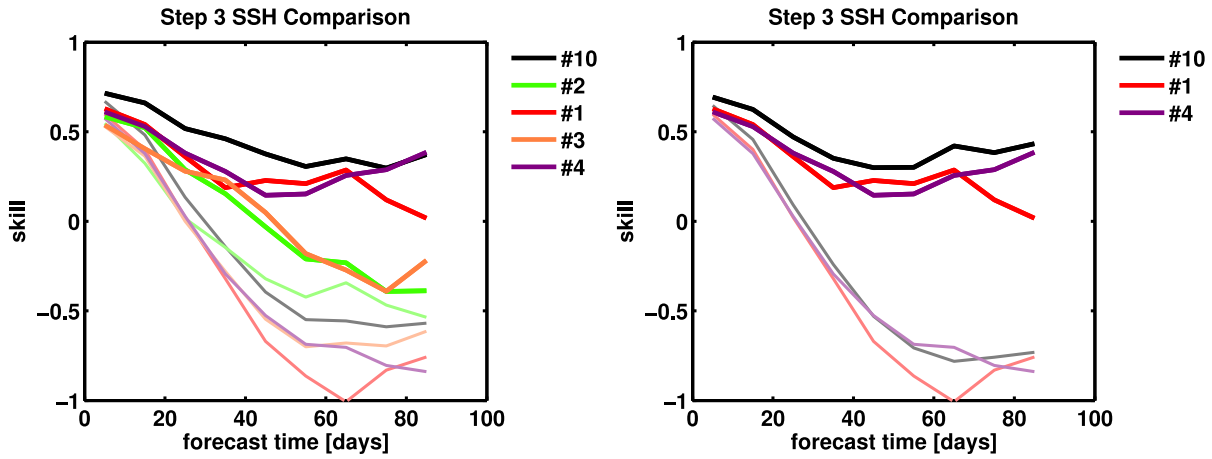


Figure 43: Skill of Consensus Forecasts, Step 3.
The consensus forecast (Model #10, black line), computed as the equally-weighted average of models #1, #2, #3, and #4 (left panel), displays significantly more skill than the constituent models through forecast day 60. Most of the incremental increase in skill may be attributed to the combination of models #1 and #4 (right panel). Bold colors indicate forecast skill (explained variance) with respect to altimeter SSH observations within the LC/LCE region. Light colors indicate persistence skill.

The multi-model consensus forecast has more skill, and explains about 10% more observed SSH variance, than any of the constituent models through day 60 of the forecast (Fig. 43, left). For a given level of useful skill, the duration of a useful forecast is significantly increased with the consensus forecast. For example, if the useful skill level is taken to be 0.3, which corresponds approximately to a pattern correlation of 0.6 here, the useful forecast duration has been extended from about 30 to 50 days.

The analysis also indicates that not all the models contribute equally to the increased skill of the consensus forecast. An ensemble consisting of models #1 and #4 retains much of the long-range forecast skill (Fig. 43, right).

These preliminary results highlight the potential benefits of an ensemble forecast, but more work should be done to verify and extend the results. For example, the forecast skill in Figure 43 is based on the seven forecast experiments in Step 3; thus, there are too few degrees of freedom to make a definitive statement about the usefulness of the results. The upward trend in forecast skill beyond day 45 for models #1 and #4 would likely disappear in a larger sample. Note that Step 2 forecasts were not analyzed with respect to the full multi-model ensemble due to the presence of more model anomalies (e.g., partial runs of model #2, spurious thermal structure of model #3, only 3 forecasts from model #4). Also note that the above analysis utilized Step 1 SSH observations as the basis for constructing an equally-weighted consensus forecast; other forms of observational data should be incorporated for both developing and validating the consensus forecasts.

Estimating Forecast Error

If one regards the ensemble as a small sample from the probability distribution of GOMEX state, new forecast products may be generated which provide likelihood information or confidence limits. Chang et al (2011) utilize a single-model ensemble to assess Lagrangian trajectories in the Gulf, and find that the ensemble forecasts compare favorably with the observed spreading of oil in the DwH event given a 1 to 2-mo. forecast horizon. However, their single-model ensemble approach was based on a large ensemble derived from a single 8-year simulation. Other recent approaches to quantifying GOMEX forecast spread are found in MacFayden et al (2011) and Barker (2011). Counillon and Bertino (2009a) demonstrate that (single-model) ensemble spread is correlated with forecast error in 14-day GOMEX forecasts using small, 10-member, ensembles.

Findings similar to Counillon and Betino (2009a) have been obtained during the twelve 3-mo. Step 2 forecasts. The multi-model ensemble spread (MME spread) has been computed as a function of time for each Step 2 forecast (Fig. 44, dashed black lines). Root-mean-square model-data error (rmse, colored lines) differs amongst the models, but it grows from small values at the start of the year and peaks in the period from May-July for the all the models. Likewise, the MME spread is small at the start of the year and peaks during the same period as the peak forecast errors. The rmse is different for each model, but, on average, the rmse is approximately four times ensemble spread. This result demonstrates that multi-model ensemble spread can provide quantitative guidance for predicting forecast accuracy.

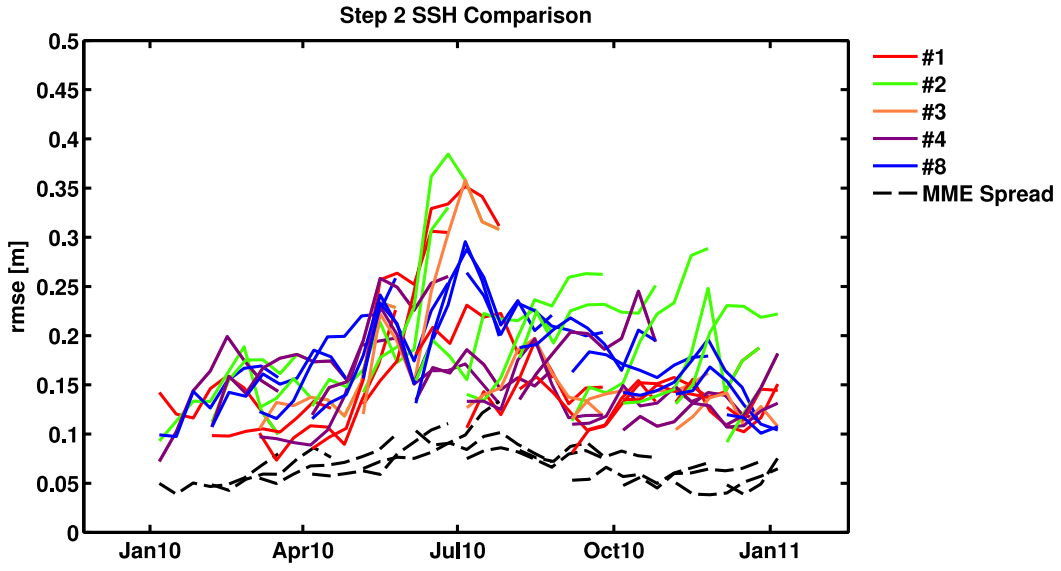


Figure 44: Multi-Model Ensemble Spread and Forecast Error, Step 2.
Forecast error is shown for each model (colored lines) with respect to satellite altimeter data in the LC/LCE region during the twelve Step 2 forecast experiments. Multi-model ensemble spread (dashed line) increases during the middle of the year, coincident with increased forecast error. Note that model #6 has been excluded from the analysis due to its anomalous SSH (c.f., Fig. 23).

Generating Probabilistic Forecasts

Certain forecast products are best presented as probabilities or likelihoods. For example, an oil-well operator might want to know, "What is probability that water speeds in excess of 0.75m/s will occur within the coming weeks?" Figure 45 illustrates one possible answer to such a query by showing the fraction of models with speed in excess of 0.75m/s within a 7-day period centered on October 21, as forecast on October 14 (forecast S3-03). Access to multiple models and their time-histories in GOMEX-PPP demonstrates the feasibility of answering such questions using existing technology.

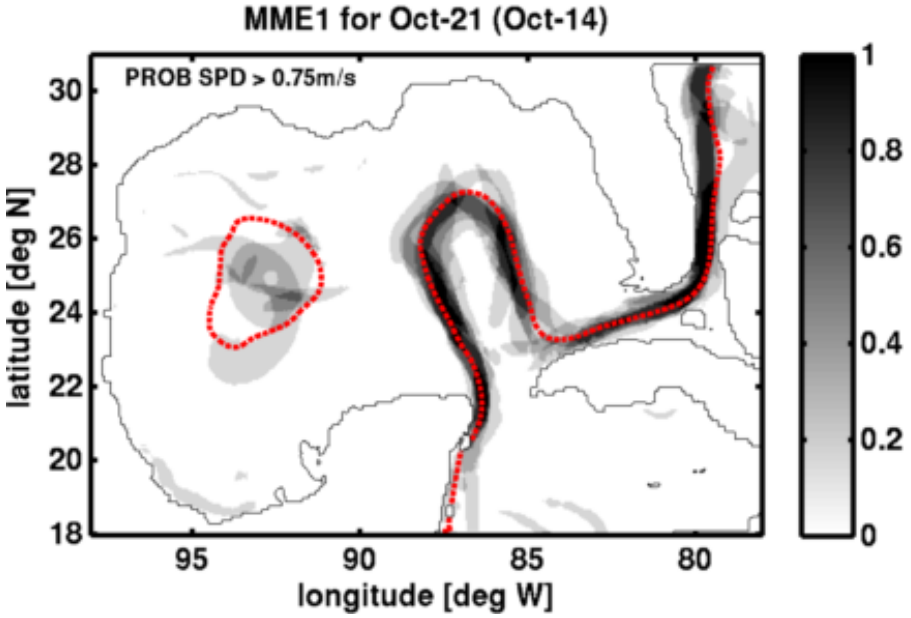


Figure 45: Forecast Probability of Speed in Excess of 0.75m/s.
Example of a probabilistic forecast from Step 3. Shading indicates likelihood of speed exceeding 0.75m/s in 7-day window centered on the 21 OCT 2011, as forecast on 14 OCT 2011 (S3-03). Dashed red line is a consensus forecast of the LC front based on forecast SSH.

In Figure 45, locations with darkest shading are highly likely to encounter a high-speed frontal jet, while lighter shades indicate less likelihood. The central forecast (ensemble mean) for the frontal position is shown by the dashed red line. In this short range, 1-week, forecast, the consensus frontal location (dashed red line) generally coincides with the forecast jet, as inferred from the shading. One can contrast this picture with a long-range, 60-day, forecast of the same quantity (Fig. 46), where the consensus position of the front is not collocated with the darkly shaded regions. In this example, the probabilistic information (shading) directly communicates likely outcomes more effectively than the consensus forecast (dashed red).

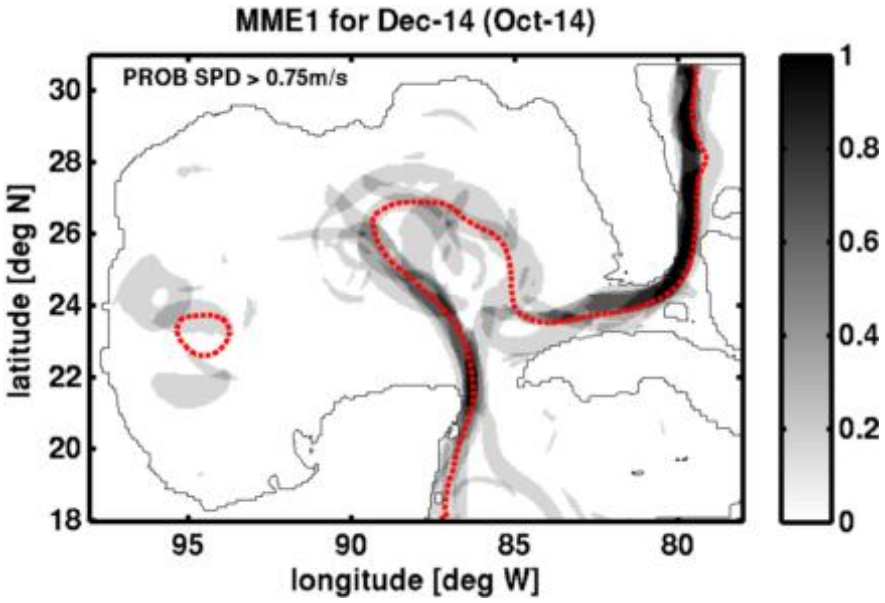


Figure 46: Forecast Probability of Speed in Excess of 0.75m/s.
As in Figure 45, except the forecast is centered on 14 DEC 2011, as forecast on 14 OCT 2011 (S3-03). For this 60-day forecast the consensus frontal location (dashed red line) and strong surface currents (dark shading) are not closely associated in the LC meander.

Generating New Forecasts with Ensemble Smoothers

The analysis of collinear SSH differences above appears to show that the SSH data are under-utilized in the GOMEX-PPP modeling systems. This may be seen in the model-data comparisons which display slowly evolving, apparently non-random, residual variability (cf., Fig. 25, Jason-1 pass 26, orbit cycles 15-20). There could be several reasons why this residual signal is present: some assimilation systems impose limits on the amplitude of SSH signals assimilated (e.g., NCODA; Cummings, 2011), or the systems may be incorrectly modeling the SSH measurement error or their own background error.

Whatever the cause, the above observation suggests that there is additional information in the SSH observations which might be used to improve forecasts. The GOMEX-PPP multi-model approach enables the possibility of performing a "poor-man's" assimilation based on the existing forecasts for a given period produced by the GOMEX-PPP modelers. The objective would be to utilize the residual signal in SSH or other data not incorporated by the original modeling groups. Such an approach could be based on an ensemble smoother which utilizes a set of 3-mo. forecasts as follows. At time 0, the forecasts to +3mo. are received from modelers. New observational data are acquired from time 0 to +1mo. When enough data are assembled, an ensemble smoother algorithm (van Leeuwen and Evensen, 1996) would be used to smooth the data from 0 to +1mo., and to create a new forecast for +1mo. to +3mo. Techniques of ensemble modulation could be used to expand the effective ensemble size, following Bishop and Hodyss (2011).

The success of this approach would depend on whether the sample state covariance computed from the multi-model ensemble is representative of the background errors in the state. The single-model ensemble approach described in Counillon and Bertino (2009a, 2009b) finds that even small, 10-member, ensembles are sufficient to improve initial conditions for relatively short, 7 to 14 day, forecasts. The positive outcome suggests that small ensembles can be effective. An analysis of single-model and multi-model state (co)variance ought to be conducted to assess whether the proposed approach is likely to yield benefits.

Addendum [Post-Phase I Workshop, FEB 2012]

Several questions and comments arose during the Workshop held FEB 2012 in Salt Lake City, UT, at the conclusion of GOMEX-PPP Phase I. Several technical questions were raised which are addressed below.

(1) Leben suggested that it was unnecessary to subtract an along-track linear trend from satellite and model SSH, as this was originally done in Leben (2005) as a correction for along-track orbit error, which is now essentially negligible in modern altimeter products. Instead, he suggested to remove any large-scale bias by subtracting a time-dependent mean sea level determined simply by area averaging. When this approach is used, the dynamic range of the skill scores decreases compared to what is shown above (Fig. 30), essentially because the fraction of explained variance is slightly increased in all models. In other words, all the models capture the along-track trend which was previously removed. See revised skill scores for Step 2: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2rev2/Skill_Summary.pdf

(2) Cornuelle inquired if the analyses of GOMEX-PPP skill are consistent with the RMSE statistics compiled for MITGOM. A detailed comparison indicates that the results are consistent and correct. Note that there are differences in the quantitative values reported for MITGOM because the GOMEX-PPP analysis was restricted to the LC/LCE region of the GOMEX (Fig. 22). The figure at the following link displays RMSE averaged over Step 2 forecasts: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/Altimetry_Step2rev2/RMSE_Summary.pdf

(3) Ko displayed time series of anomaly correlation for Model #1 which indicated correlation decaying to 0.5 at approximately 50-day forecast time. This result appeared to contradict skill score near zero at 50-day lead time (Fig. 31). These results are consistent when model and data have nearly the same variance, as in this case; then the skill score (SS) can be expressed in terms of anomaly correlation (R) as SS=2R-1. In other words, zero skill corresponds to 0.5 anomaly correlation. Skill score is zero because the variance correlated between the model and data is exactly offset by the uncorrelated variance.

(4) Cooper and Oey suggested that the analysis of frontal properties was unnecessarily complicated, and a simpler criterion such as T200=18C should be used since the 18C isotherm is a reliable boundary between interior GOMEX and Atlantic water masses. When this criterion is used, the frontal CPA error metrics for all the models are increased during Step 2: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/GRL2005rev_Summary/Summary.pdf

(5) Cooper inquired about how well the Model #9 (ForLoop) performed in the frontal error analysis. It had not previously been analyzed because it showed little development over the course of the experiment. In spite of this lack of dynamics, the time course of the frontal error is quite similar to what was found in the dynamical models, i.e., its error quickly grows to the 50-70km range, where it stays for the duration of the forecast: http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/GRL2005rev_Summary/forloop_3X.png

(6) Cooper suggested additional changes to the methodology for computing frontal metrics. For example, he suggested computing CPA error metrics only for those sites where the HMI front

was a “threat” to the reference site; e.g., when the front was located to the east of the site. Experimentation with this approach did help to stabilize the CPA metric time series, and it collapsed the frontal CPA error curves, illustrating the similar performance of the models (with the exception of model #3). Step 2 forecast CPA error metrics computed in this manner are here shown for the optimal T200-based fronts:
http://maki.cee.pdx.edu/~ezaron/GOMEX/Figures/GRL2005rev_Summary/SummaryX.pdf

Findings and Recommendations

FINDING 1: State-of-the-Art

Modeling systems exist which can provide forecasts for the GOMEX using realistic, physically-based models.

Modeling systems employ advanced data assimilation strategies to maximize the positive impact of available observational data on model initialization.

Forecasts are routinely initialized with a combination of operational data, climatological data, and background model nowcasts. Ocean forecasts are produced with or without the benefit of long-range (i.e., greater than two weeks) atmospheric forecasts.

Forecast skill defined in terms of explained variance for observed quantities can be quantified; however, the definition of "useful skill" is ill-defined and depends on the evolving end-use context.

LC/LCE frontal position (or, equivalently, the jet position) is a high-value forecast product; however, there are insufficient non-proprietary data (operational or non-operational) to assess whether the forecast products are sufficiently skillful to be of value.

RECOMMENDATION 1: Initialization Improvement

For improved initialization, validation, and verification of mesoscale forecasts in GOMEX, an enhanced real-time observing system needs to be designed, implemented, evaluated, and managed. Particular attention should be given to making fuller use of satellite imagery to delineate fronts and jets.

FINDING 2: Implementation of GOMEX-PPP, Phase I

Model outputs were interpolated to a common grid (5km resolution and 22 vertical levels) and time step (daily average) for the purpose of comparisons.

Some anomalies occurred in the file submission process (e.g., interrupted file transfers), in file content (e.g., inconsistent variable masking and use of missing value attributes), and some participants were unable or unwilling to conform to file standards. While these anomalies did not materially impact project results, they resulted in some inconsistencies in the model/data comparisons.

The organization of the project into three steps (i.e., retrospective nowcasts, retrospective forecasts, and real-time forecasts) permitted assessment of the models in an environment incrementally approaching the operational context of real-time nowcasting and forecasting.

The several forecast experiments conducted in GOMEX-PPP can now be evaluated with respect to time scales of variability in the GOMEX. It appears that the number of forecast experiments performed (N=12 in Step 2 and N=7 in Step 3) was insufficient to observe statistically significant differences among the models' forecast skill.

In spite of the large seasonal variability in circulation (sampled once), large, apparently significant, differences in nowcast skill were observed among the models (based on SSH comparisons within approximately N=36 independent 10-day windows in Step 1).

RECOMMENDATION 2

The statistics of GOMEX fields appear to be non-stationary on the seasonal and annual time scales. Obtaining statistical significance of model comparisons requires replicated experiments under nominally identical conditions. Replicate experiments in an operational environment are normally not feasible, unless the full suite of operational data, forcing fields, and other required model inputs can be archived in total. Hence, it is recommended that pre-operational systems be configured to archive the necessary inputs for a minimum of 3 years, so that three nominally independent replicates will be available for future model comparisons.

FINDING 3: Design of GOMEX-PPP, Phase I

The organization of the project in three steps permitted assessment of the models in an environment incrementally approaching the operational context.

RECOMMENDATION 3

To provide more rapid feedback for modelers regarding the compliance of output files and basic observational constraints, the model output submission process should be automated with a Web-based application for uploading model files, verifying file format, providing immediate feedback (e.g., an animation of model SSH), and providing a quantitative comparison with near-real-time or archived operational data.

FINDING 4: GOMEX Circulation Forecast System

One metric of LC/LCE predictability is the time for a forecast to lose all skill, when model-data error variance equals data variance. Based on satellite observations of SSH during Step 3 the median time to zero skill was about 35 days for persistence and about 50 days for forecast.

RECOMMENDATION 4

A full-time post-doc is needed for three years to pursue such relevant questions as the following:

* Nature and significance of variability on the GOMEX open boundaries (Yucatan Channel & Straits of Florida)
* Response to summertime tropical cyclone and wintertime subtropical frontal passages
* Passive surface particle trajectory prediction
* Mississippi-Atchafalaya River Plume exchanges with deep water GOMEX
* Nature and significance of shelf and deep-water exchanges throughout GOMEX
* Horizontal and vertical structure of the velocity and temperature fields associated with LC/LCE fronts and jets

FINDING 5: Results of GOMEX-PPP, Phase I

GOMEX modeling systems span a wide range of maturity, robustness, and resiliency.

Model #1 has a record of peer-reviewed diagnostic and skill-assessment activities; it was run routinely during GOMEX-PPP and files were submitted in conformance with project specifications; and it exhibited a small initialization transient.

Model #2 has the longest record of peer-reviewed publications and analyses; it was run routinely during GOMEX-PPP, including multiple versions of Step 1, but files required special handling to make them conformant with project specifications; and it did not display any obvious bias or initialization transients.

Model #3 is a research and development system with no direct record of publications, although components have been documented in the peer-reviewed literature in other settings; it participated in all experimental Steps, but revised Step 1 results have been submitted; model results contained a significant problem with upper-ocean temperature, and the output displayed significant drift or initialization anomaly.

Model #4 is a research and development system with no direct record of publications, although components have been documented in the peer-reviewed literature in other settings; it participated in all experimental Steps, including multiple versions of Step 1; and it exhibited a noticeable initialization anomaly.

Model #5 is a pre-operational system which is currently undergoing extensive verification efforts. It did not participate in Step 2 and Step 3 experiments.

Model #6 is a pre-operational system with a few publications, although components have been documented in the peer-reviewed literature in other settings; it participated in most experimental Steps, including multiple versions of Step 1, but model outputs required re-gridding to bring them into conformance with project specifications; and in some respects its quantitative performance was significantly lower than other models.

Model #7 is an operational system which was withdrawn from the study due to obvious anomalies resulting from model instability.

Model #8 is a research and development system with several peer-reviewed publications, and all components have been documented in the peer-reviewed literature in other settings; it fully participated in all experimental Steps, and model outputs files complied with project specifications; however, it exhibited a noticeable initialization anomaly.

RECOMMENDATION 5: Approach to Phase II

There was a divergence of opinion about how to proceed to Phase II.

The PSU participants (Mooers and Zaron) recommended that Model #1 (IASNFS) proceed to Phase II. This recommendation was based on the stability and maturity of the modeling system as evidenced by its background of peer-reviewed publications, consistency of performance throughout Steps 1, 2, and 3 and lack of any anomalies or deficiencies. Most of these features are also shared by Model #2; however, Model #1 appears to be significantly more robust and skillful in SSHA forecasts than Model #2 (cf., Fig. 30).

The participants in the SAC Workshop (FEB 2011) voted with near-consensus to proceed to Phase II with all modeling groups contributing to MME development, evaluation, and demonstration. The majority opinion was that no single modeling group demonstrated superior skill in all aspects, and each system would benefit from further development in the context of a common project.

Subsequently, the leading TAC participants (Cooper, Driver, and Vogel) recommended that Phase II proceed with another model/modeling group, led by Pat Hogan at NRL. This recommendation was based on the conclusion that none of the funded modeling groups had improved on the state-of-the-art found in the 2005 DeepStar project, and the forecast error in all GOMEX-PPP models is severely limited by poor model initial conditions. Furthermore, Hogan's group has demonstrated experience with multi- and single-model ensemble forecasts, has strong connections with the NAVO operational center, and is utilizing more advanced data assimilation for model initialization than the funded Phase I GOMEX-PPP participants.

Independent of the funded GOMEX-PPP effort, the contributions of Model #4 showed good forecast skill. It is recommended that it continue to run in tandem with GOMEX-PPP efforts through Phase II.

References

Barker, C.H. 2011. A statistical outlook for the Deepwater Horizon oil spill. In: Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise, Geophysical Monograph Series, vol. 195. Washington, D.C.: American Geophysical Union. Pp. 237–244.

Bishop, C.H. and D. Hodyss. 2011. Adaptive Ensemble Covariance Localization in Ensemble 4D-VAR State Estimation. Mon. Wea. Rev. 139:1241–1255.

Candela, J., J. Sheinbaum, J. Ochoa, A. Badan, and R. Leben. 2002. The potential vorticity flux through the Yucatan Channel and the Loop Current in the Gulf of Mexico. Geophys. Res. Lett. 29:2059–2062.

Candela, J., S. Tanahara, M. Crepon, B. Barnier, and J. Sheinbaum. 2003. Yucatan Channel flow: Observations versus CLIPPER ATL6 and MERCATOR PAM models. J. Geophys. Res. 108(C12):3385.

Chang,Y.-L. and L.-Y. Oey. 2010. Why can wind delay the shedding of Loop Current eddies? J. Phys. Oceanogr 40:2481-2495.

Chang, Y.-L. and L.-Y. Oey. In press. Why does the Loop Current tend to shed more eddies in summer and winter? Geophys. Res. Lett., doi:10.1029/2011GL050773.

Chang, Y.-L., L. Oey, F.-H. Hu, H.-F. Lu, and A. Fujisaki. 2011. 2010 oil spill: trajectory projections based on ensemble drifter analyses. Ocean Dynamics 61(6):829-839.

Chang,Y.-L. and L.-Y. Oey. In press. Loop Current Cycle: coupled response of Loop Current and deep flows. J. Phys. Oceanogr.

Chassignet, E.P., H.E. Hurlburt, E.J. Metzger, O.M. Smedstad, J. Cummings, G.R. Halliwell, R. Bleck, R. Baraille, A.J. Wallcraft, C. Lozano, H.L. Tolman, A. Srinivasan, S. Hankin, P. Cornillon, R. Weisberg, A. Barth, R. He, F. Werner, and J. Wilkin. 2009. U.S. GODAE: Global Ocean Prediction with the Hybrid Coordinate Ocean Model (HYCOM). Oceanography 22(2):64-75.

Chen, K. and R. He. 2010. Numerical Investigation of the Middle Atlantic Bight Frontal Circulation using a High Resolution Ocean Hindcast Model. Journal of Physical Oceanography DOI: 10.1175/2009JPO4262.1

Chin, T.M., M.J. Turmon, J.B. Jewell, and M. Ghil. 2007. An ensemble-based smoother with retrospectively updated weights for highly nonlinear systems. Monthly Weather Review 135:186-202.

Counillon, F. and L. Bertino. 2009a. High-resolution ensemble forecasting for Gulf of Mexico eddies and fronts. Ocean Dynamics 59:83-95.

Counillon, F. and L. Bertino. 2009b. Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. Tellus 61A:296-308.

Crosby, D.S., L.C. Breaker, and W.H. Gemmill. 1990. A definition for vector correlation and its application to marine surface winds. National Meteorological Center Office Note No. 365, NOAA/National Weather Service, Washington, D.C., 50 pp.

Crosby, D.S., L.C. Breaker, and W.H. Gemmill. 1993. A Proposed Definition for Vector Correlation in Geophysics: Theory and Application. J. Atmos. Sci. 10:355-367.

Cummings, J.A. 2005. Operational multivariate ocean data assimilation. Quarterly Journal of the Royal Meteorological Society 131:3583-3604.

Cummings, J.A. 2006. The NRL real-time ocean data quality control system. NRL Technical Note (available from cummings@nrlmry.navy.mil).

Cummings, J.A. 2011. Ocean Data Quality Control. In: Schiller, A. and G.B. Brassington, eds. Operational Oceanography in the 21st Century. New York: Springer-Verlag. Pp. 91-121.
Ezer, T., L.-Y. Oey, W. Sturges, and H.-C. Lee. 2003. The variability of currents in the Yucatan Channel: Analysis of results from a numerical ocean model. J. Geophys. Res. 108(C1):3012.
Fox, D.N., C.N. Barron, M.R. Carnes, M. Booda, G. Peggion, and J.V. Gurley. 2002. The modular ocean data assimilation system. Oceanography 15:22-28.
Fu, L.-L. and A. Cazenave. 2001. Satellite Altimetry and Earth Sciences. International Geophysics Series, Vol. 69. San Francisco: Academic Press. 463 pp.
Hamilton, P., J.C. Larsen, K.D. Leaman, T.N. Lee, and E. Waddell. 2005. Transports through the Straits of Florida. J. Phys. Oceanogr. 35:308-322.
Hyun, K.H. and R. He. 2010. Coastal Upwelling in the South Atlantic Bight: A Revisit of the 2003 Cold Event Using Long Term Observations and Model Hindcast Solutions. Journal of Marine Systems 83:1-13.
Jenkins, G.M. and D.G. Watts. 1968. Spectral Analysis and Its Applications. San Francisco: Holden-Day. 525 pp.
Jolliff, J.K., J.C. Kindle, I. Shulman, B. Penta, M.A.M. Friedrichs, R. Helber, and R.A. Arnone. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, Journal of Marine Systems 76:64-82.
Kantha, L., J.-K. Choi, K.J. Schaudt, and C.K. Cooper. 2005. A Regional Data-Assimilative Model for Operational Use in the Gulf of Mexico. In: Circulation in the Gulf of Mexico: Observations and Models. Geophysical Monograph Series 161. Pp. 165-180.
Kenyon, S., J. Factor, N. Pavlis, and S. Holmes. 2007. Towards the Next Earth Gravitational Model. Society of Exploration Geophysicists 77th Annual Meeting, San Antonio, Texas. http://earth-info.nga.mil/GandG/wgs84/gravitymod/new_egm/papers.html.
Ko, D.S., R.H. Preller, and P.J. Martin. 2003. An Experimental Real-Time Intra Americas Sea Ocean Nowcast/Forecast System for Coastal Prediction. Proceedings, AMS 5th Conference on Coastal Atmospheric & Oceanic Prediction & Processes.
Ko, D.S., P.J. Martin, C.D. Rowley, and R.H. Preller. 2008. A real-time coastal ocean prediction experiment for MREA04. J. Marine Systems 69:17-28.
Larsen, J.C. and T.B. Sanford. 1985. Florida Current volume transports from voltage measurements. Science 227:302–304.
Leben, R.R. 2005. Altimeter-derived Loop Current metrics. In: Sturges, W. and A. Lugo-Fernandez, eds. Circulation in the Gulf of Mexico: Observations and Models. Washington, D.C.: American Geophysical Union. Pp. 181–201.
Leben, R.R. and G.H. Born. 1993. Tracking Loop Current eddies with satellite altimetry. Adv. Space Res. 13(11):325-333.
Leben, R.R., G.H. Born, B.R. Engebrath. 2002. Operational altimeter data processing for mesoscale monitoring. Marine Geodesy 25:3-18.
Leben, R.R. and K. Corcoran. 2005. Horizon Marine Inc. Numerical Ocean Model Nowcast and Forecast Skill Assessment in the Gulf of Mexico. Horizon Marine, Inc. http://www.horizonmarine.com/library/Horizon%20Marine%20Inc.%20Numerical%20Ocean%20Model%20Nowcast%20and%20Forecast%20Skill%20Assessment%20in%20the%20Gulf%20of%20Mexico.pdf. 31 pp.
Lee, T.N., and E. Williams. 1988. Wind-forced transport fluctuations of the Florida Current. Journal of Physical Oceanography 18:937-946.

Lee, T.N., W.E. Johns, R.J. Zantopp and E.R. Fillenbaum. 1996. Moored observations of western boundary current variability and thermohaline circulation at 26.5N in the subtropical North Atlantic. Journal of Physical Oceanography, 26:962-983.

Li, Z., Y. Chao, J.C. McWilliams, and K. Ide. 2009. A Three-Dimensional Variational Data Assimilation Scheme for the Regional Ocean Modeling System. Journal of Atmospheric and Oceanic Technology 25:2074-2090.

Lin, X.-H., L.-Y. Oey, and D.-P. Wang. 2007. Altimetry and drifter assimilations of Loop Current and eddies. Journal of Geophysical Research 112:C05046.

Liu, Y. and R.H. Weisberg. 2011. Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. J. Geophys. Res. 116:C09013.

Liu, Y., R.H. Weisberg, C. Hu, and L. Zhang. 2011. Trajectory forecast as a rapid response to the Deepwater Horizon oil spill. In: Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise, Geophysical Monograph Series, vol. 195. Washington, D.C.: American Geophysical Union. Pp. 153-165.

MacFadyen, A., G.Y. Watabayashi, C.H. Barker, and C.J. Beegle-Krause. 2011. Tactical modeling of surface oil transport during the Deepwater Horizon spill response. In: Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise, Geophysical Monograph Series, vol. 195. Washington, D.C.: American Geophysical Union. Pp. 167-178.

Mehra, A. and I. Rivin. 2010. A Real Time Ocean Forecast System for the North Atlantic Ocean. Terr. Atmos. Ocean. Sci. 21(1):211-228.

Mooers, C.N.K., C.S. Meinen, M.O. Baringer, I. Bang, R. Rhodes, C.N. Barron, and F. Bub. 2005. Cross validating ocean prediction and monitoring systems. Eos, 86(29): 269-273.

Naeije, M., E. Doornbos, L. Mathers, R. Scharroo, E. Schrama, and P. Visser. 2002. Radar Altimeter Database System: exploitation and extension (RADSxx). NUSP-2 report 02-06, NUSP-2 project 6.3/IS-66. Delft Institute for Earth-Oriented Space Research (DEOS), Delft, Netherlands.

Oey, L.-Y. 1996. Simulation of mesoscale variability in the Gulf of Mexico: Sensitivity studies, comparison with observations, and trapped wave propagation. J. Phys. Ocean. 26(2):145-175.

Oey, L.-Y., T. Ezer, and W. Sturges. 2004. Modeled and observed Empirical Orthogonal Functions of currents in the Yucatan Channel, Gulf of Mexico. J. Geophys. Res. 109:C08011.

Oey, L.-Y., H.-C. Lee, and W.J. Schmitz. 2003. Effects of winds and Caribbean eddies on the Loop Current eddy shedding: A numerical model study. J. Geophys. Res. 108:3324.

Oey, L.-Y., T. Ezer, G. Forristall, C. Cooper, S. DiMarco, and S. Fan. 2005. An exercise in forecasting loop current and eddy frontal positions in the Gulf of Mexico. Geophys. Res. Lett. 32:L12611.

Pavlis, N.K., S.A. Holmes, S.C. Kenyon, D. Schmidt, and R. Trimmer. 2004. A Preliminary Gravitational Model to Degree 2160. Paper presented at the IAG International Symposium on Gravity, Geoid and Space Missions 2004 (GGSM2004) Porto, Portugal. http://earth-info.nga.mil/GandG/wgs84/gravitymod/new_egm/papers.html.

Powell, B.S. and R.R. Leben. 2004. An Optimal Filter for Geostrophic Mesoscale Currents from Along-Track Satellite Altimetry. J. Atm. and Ocean. Tech. 21:1633-1642.

Rio, M.-H. and F. Hernandez. 2003. A Mean Dynamic Topography computed over the world ocean from altimetry, in-situ measurements and a geoid model. J. Geophys. Res. 109:C12032.

Rio, M.H., S. Guinehut, and G. Larnicol. In press. The New CNES-CLS09 global Mean Dynamic Topography computed from the combination of GRACE data, altimetry and in-situ measurements. J. Geophys. Res., doi:10.1029/2010JC006505.

Rousset, C. and L.M. Beal. 2010. Observations of the Florida and Yucatan Currents from a Caribbean Cruise Ship. J. Phys. Oceanogr. 40:1575-1581.

Shay, L.K., B. Jaimes, J.K. Brewster, P. Meyers, E.C. McCaskill, E. Uhlhorn, F. Marks, G.R. Halliwell Jr., O.M. Smedstad, and P. Hogan. 2011. Airborne ocean surveys of the Loop Current complex from NOAA WP-3D in support of the Deepwater Horizon oil spill. In: Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise, Geophysical Monograph Series, vol. 195. Washington, D.C.: American Geophysical Union. Pp. 131-151.

Shoosmith D.R., M.O. Baringer, and W.E. Johns. 2005. A continuous record of Florida Current temperature transport at 27°N. Geophys. Res. Lett. 32:L23603.

SSALTO/DUACS. 2011. SSALTO/DUACS User Handbook: (M)SLA and (M)ADT Near-Real Time and Delayed Time Products. Issue 2.6, CLS-DOS-NT-06-034, Nomenclature : SALP-MU-P-EA-21065-CLS. 57pp.

Sturges, W., J.C. Evans, S. Welsch, and W. Holland. 1993. Separation of warm-core rings in the Gulf of Mexico. J. Phys. Oceanogr. 23:250-268.

Sturges, W., N. Hoffmann, and R. Leben. 2010. A Trigger Mechanism for Loop Current Ring Separations. J. Phys. Oceanogr. 40:900-913.

Taylor, K.E. 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106: 7183-7192.

van Leeuwen, P.J. and G. Evensen. 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. Mon. Wea. Rev. 124: 2898-2913.

Zhao, Y. and R. He. In Press. Cloud-free Sea Surface Temperature and Color Reconstructions for the Gulf of Mexico: 2003-2009. Remote Sensing Letters.

Appendix I: Summary of User Survey for GOMEX-PPP

Prepared by Cort Cooper/CHEVRON, Chris Simoniello/GCOOS-RA, and Chris Mooers/PSU.

Executive Summary

The long-term viability of an operational forecast model will depend strongly on satisfying the needs of critical users. With this in mind, GOMEX-PPP developed a 15-question survey and distributed it widely to major potential users. There were 24 responses from the 54 invitations. This report/annex summarizes the survey results. In general, there is often no dominant single answer to a survey question and a closer look suggests this is probably due to the large diversity in user organizations. However, the sample size is too small to gather any statistically robust insight into the correlation between survey answers and the user organization.

That said, some clear trends do emerge. Almost all users have a high interest in near-surface currents generated by storms and the Loop Current and are universally interested in nowcasts. Though somewhat less universal, there is strong interest in forecasts updated daily with errors of less than 20%. Those in the oil and gas sector would also like to see forecasts with time horizons of several months and are willing to accept errors of up to 100%. There is considerable interest by most users in hindcasts of 10 years duration or longer.

The survey described in this report should be viewed as a first step in an ongoing process of gauging user needs and interests. Subsequent surveys should attempt to cast the net wider and catch a larger number of respondents in order to obtain more statistically robust results and especially to allow a better understanding of how user needs vary by organization type. In particular, a more concerted effort should be made to engage the NWS Marine Forecasters located in the region, the value-added industry operating in the region, and other super-users employed in various agencies. It will be easier to communicate with users once there is a first-generation information system in place for them to critique.

Introduction

The survey was sent to 54 individuals representing a wide breadth of interest in the Gulf of Mexico. Completed surveys were submitted by 24 individuals. A list of names is included in Appendix A, while a copy of the survey is included in Appendix B. The next section summarizes the results and is followed by a discussion section. The final section shows plots of the results and includes a brief summary of the main points in each plot.

Results

Results are summarized by question in the ‘Details’ Section below. All results have been plotted as bar charts. For many of the questions, each bar shows an average plus the standard deviation. The top of the bar thus represents roughly the 90% nonexceedence level (henceforth referred to as the ‘P90’). For most of the questions, the respondent was allowed to weight the answers: ‘blank’ for no interest, ‘1’ for high interest, and ‘2’ for moderate interest. In developing the bar

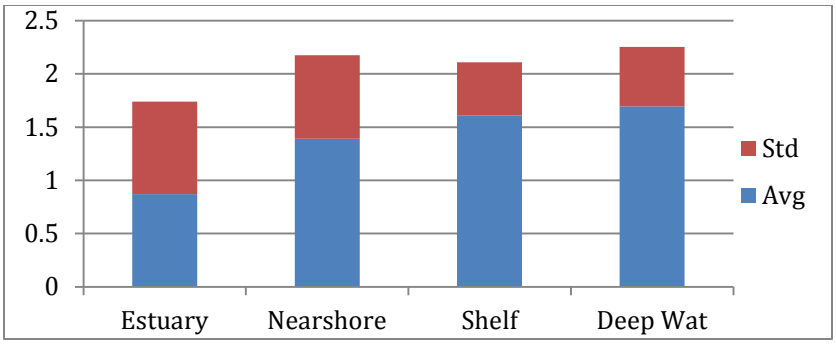
charts, the answers were assigned 0, 2, and 1, respectively, so that the higher the score, the stronger the interest.

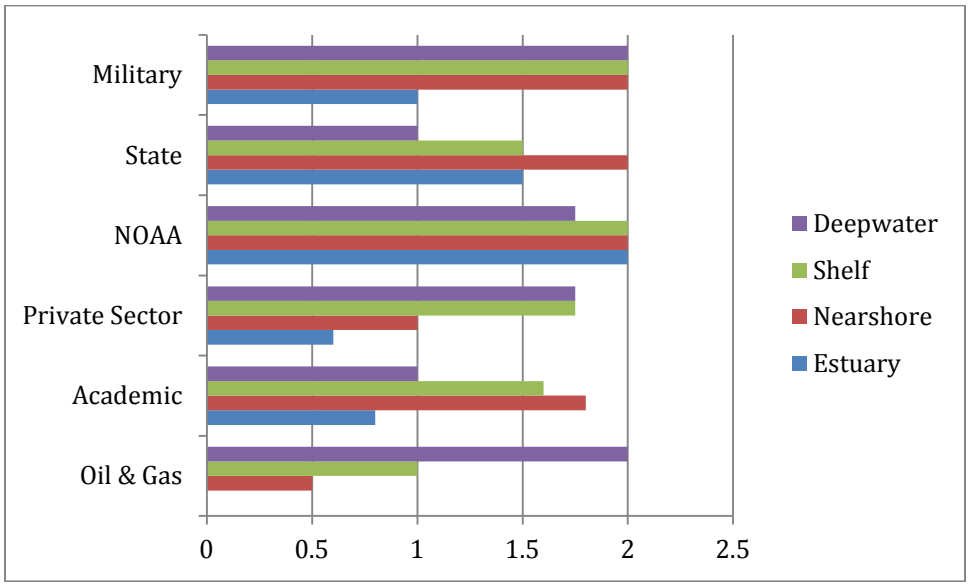
A summary of the results follows:

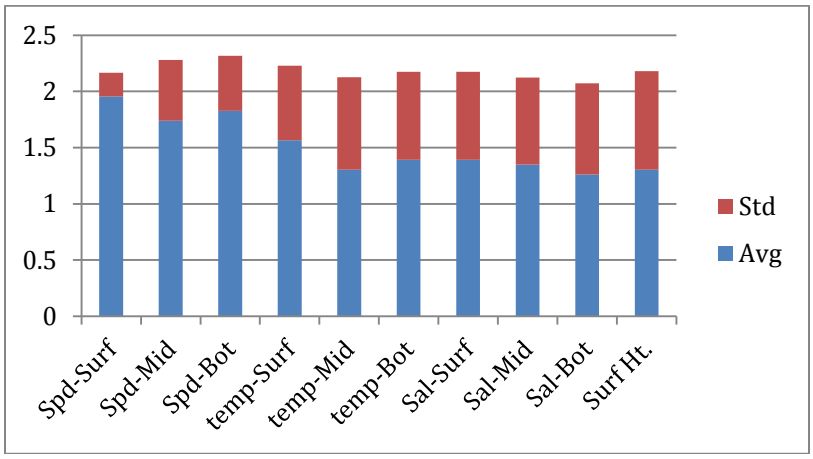
1. **Processes of interest**: Storms and the Loop Current/eddies were of greatest interest.
2. **Region of geographical interest**: There was universal interest in Louisiana and Texas. There was only slightly less interest in other GOMEX States. There is little interest in Cuban or Mexican waters but this almost certainly reflects the US-centric geographic distribution of respondents.
3. **Water Depths**: All water depths except estuaries were of interest. Two academic modelers interested in estuaries declined to take the survey, citing lack of ability to resolve boundary conditon issues as the reason.
4. **Variables of interest**: Surface water velocity was the highest priority, but there was also substantial interest in velocity deeper in the water column and surface water temperatures.
5. **Time frames of interest**: Nowcasts were of greatest interest, with less interest in hindcasts and forecasts, and little interest in simulations.
6. **Duration of hindcasts**: Survey takers were most interested in timeframes longer than 10 years.
7. **Forecast requirements:** The highest priority was daily updates with better than 20% accuracy, but users are almost equally split on the desired time horizon (2 days-2 months). Users are willing to accept larger errors for the longer time horizons.
8. **Desired products**: There was nearly equal preference for raw output, site-specific time series, synoptic maps, and forecast error estimates.

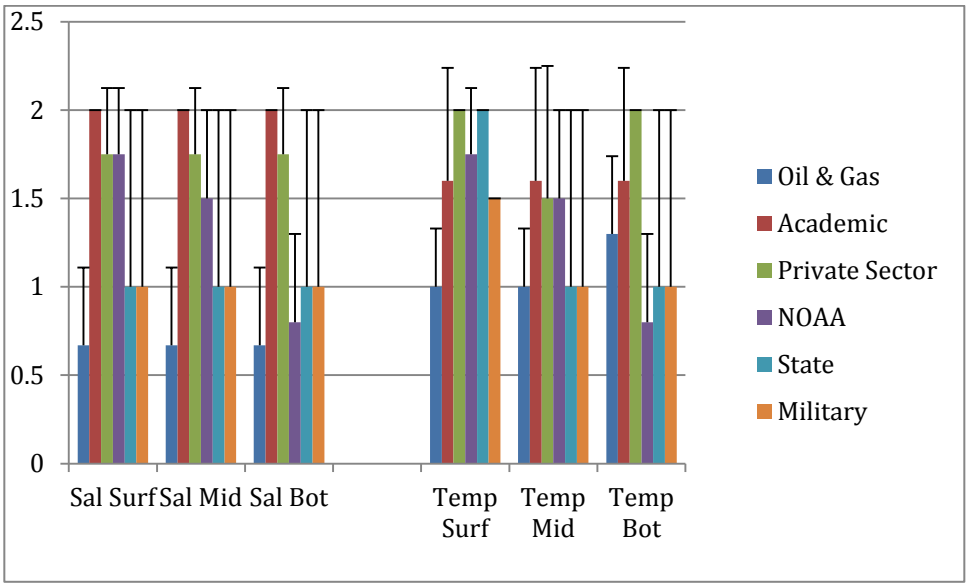
Discussion

For many of the questions there is a fairly small difference in results, especially when considering the P90, but a closer look reveals some fairly strong differences that are highly correlated to the respondent' s work organization. A case in point is Question 4 which seeks to identify the water depth of interest. Overall results indicate about equal interest at the P90 level in nearshore, shelf, and deep water (see first figure below) and only a modest difference in the averages. A closer look at the breakdown of the responses by organization type (see 2nd figure below) shows that modelers from the oil & gas industry are overwhelmingly interested in deep water while the academic and state government modelers are more interested in shallower waters. Likewise, results of question five suggest equal interest in salinity and temperature measurements by all participants. However, closer examination demonstrates how the results are influenced by the survey participants. Pooling responses for surface, middle, and bottom water increases the sample size for each category, and shows the academic community has significantly more interest than the oil and gas industry in both salinity (Student's t-Test, $p<0.01$), and water temperature (Student's t-Test, $p<0.03$) measurements. Similar comparisons can be made between and across other user groups. However, larger sample sizes are needed for meaningful statistical comparisons.


2.5
2
1.5
1
0.5
0
Estuary
Nearshore
Shelf
Deep Wat
Std
Avg

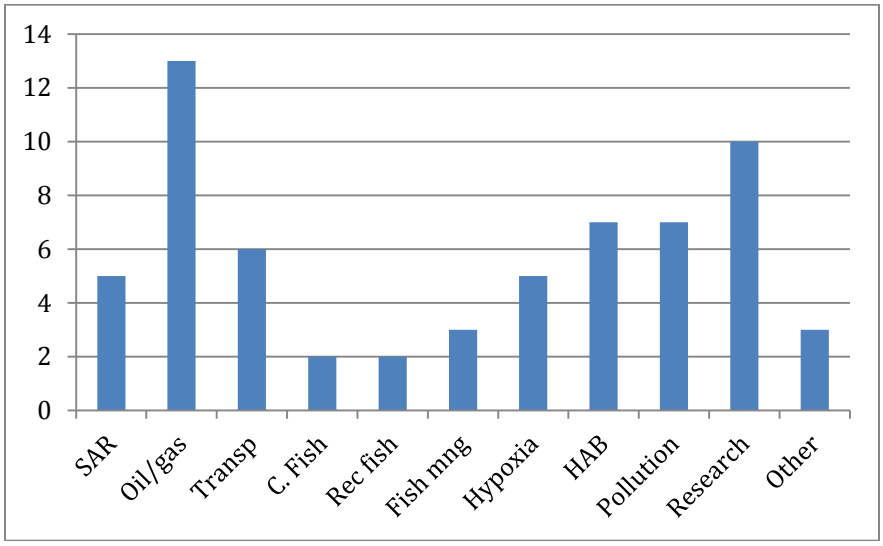

Military
State
NOAA
Private Sector
Academic
Oil & Gas
0
0.5
1
1.5
2
2.5
Deepwater
Shelf
Nearshore
Estuary


2.5
2
1.5
1
0.5
0
Spd-Surf
Spd-Mid
Spd-Bot
temp-Surf
temp-Mid
temp-Bot
Sal-Surf
Sal-Mid
Sal-Bot
Surf Ht.
Std
Avg

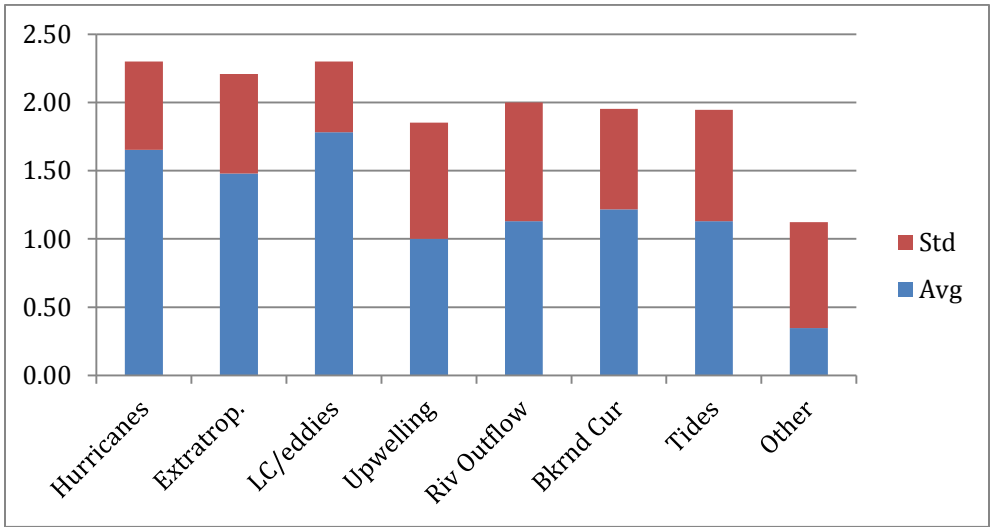

2.5
2
1.5
1
0.5
0
Sal Surf
Sal Mid
Sal Bot
Temp Surf
Temp Mid
Temp Bot
Oil & Gas
Academic
Private Sector
NOAA
State
Military

Details

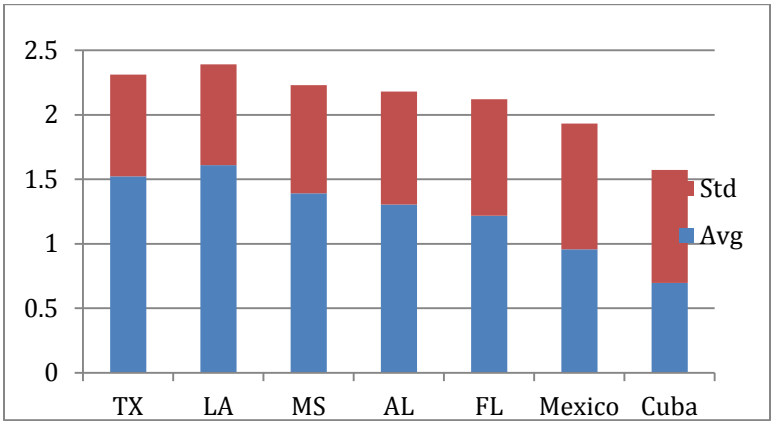
Q1: Which of the following best describes your organization's activities in the Gulf of Mexico? By far, the largest number of respondents indicated interest in the oil and gas sector even though only five of the respondents worked for an oil company. "Research" was the second highest with 10.



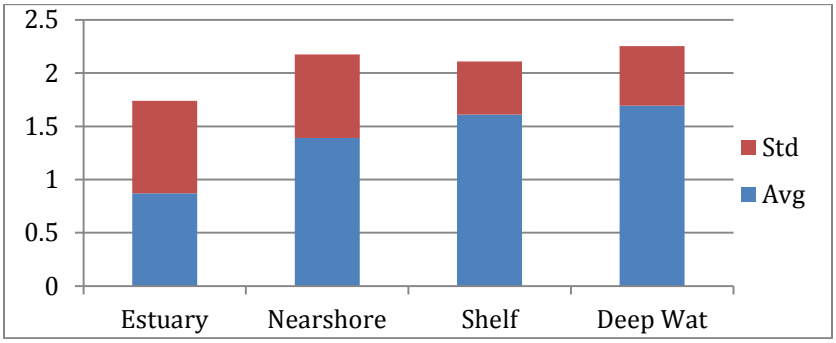
Q2: In which processes are you interested (max score of 2)? Most are interested in storms and the Loop Current. The others were weighted about equally.



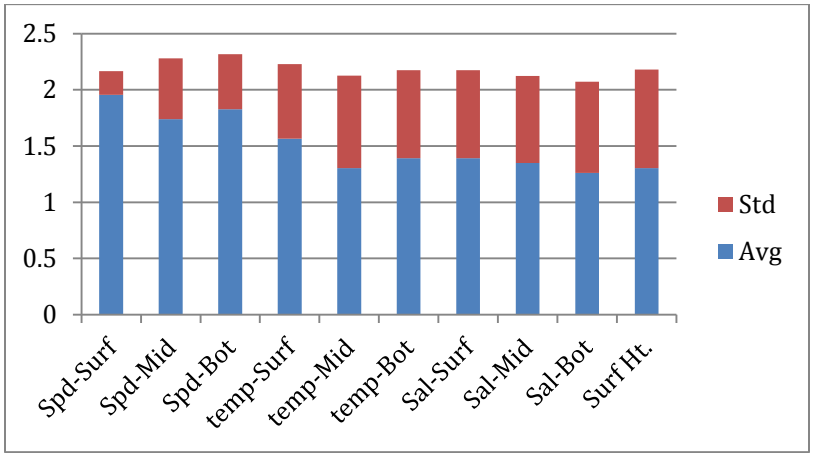
Q3: In which of the following Gulf of Mexico areas are you interested (max score of 2)? Most are interested in Texas and Louisiana but not a great deal more than the other Gulf Coast States.



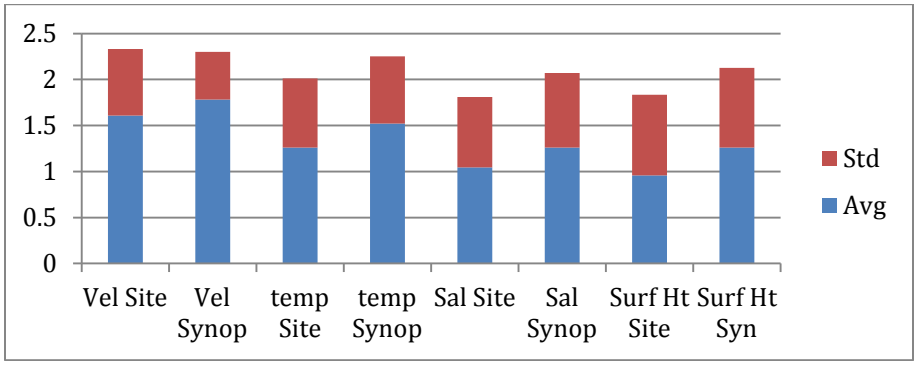
Q4: What water depth regime of the Gulf of Mexico is of interest to you (max score of 2)? The highest interest is in deep water but shelf and nearshore are close seconds. There is little interest in estuaries though this probably reflects the bias in the user sample.



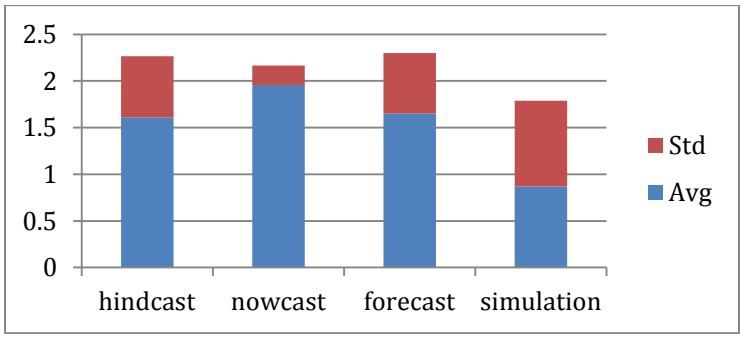
Q5: In what part of the water column are you interested (max score of 2)? All users expressed interest in water velocity but there is considerable interest in the other variables as well.



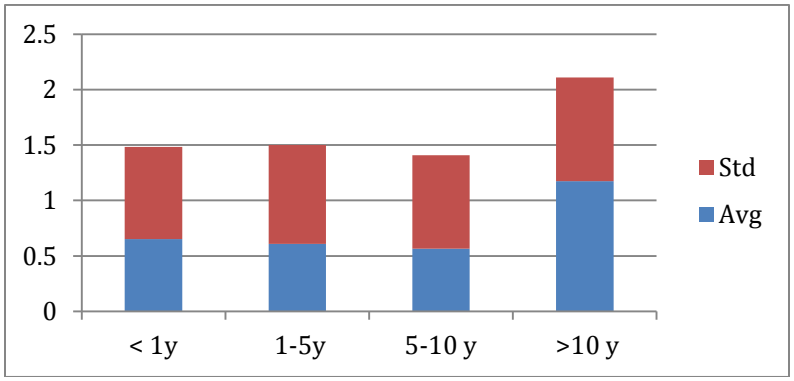
Q6: In which variables and spatial scales are you most interested (max score of 2)? Synoptic velocities are of highest interest with slightly less interest for site velocity and synoptic temperature.



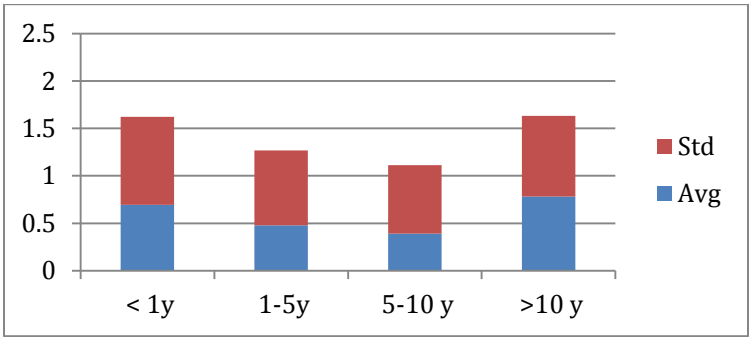
Q7: In which time frames are you interested (max score of 2)? Almost everyone is interested in nowcasts though this is slightly less true for non-velocity variables. There is little interest in simulations.



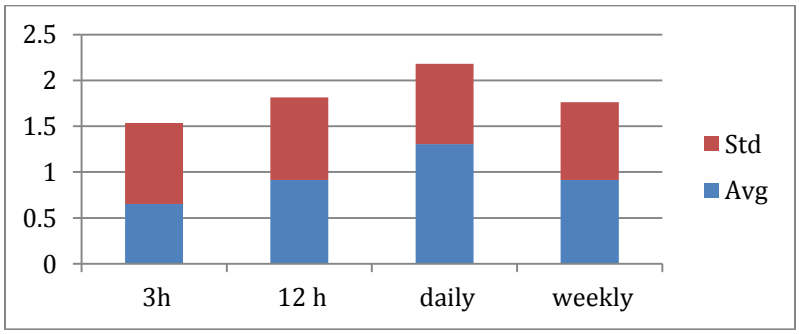
Q8: If you indicated interest in 'hindcasts' in Question 7, what duration would you need (max score of 2)? For velocity, interest is in > 10 yr hindcasts but for other variables it is more even.



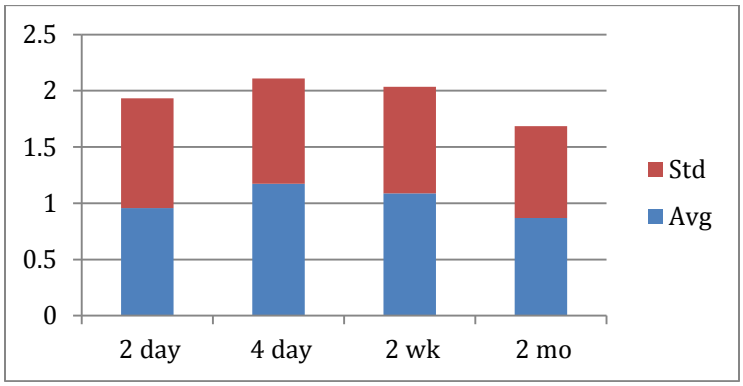
Q9: If you indicated interest in 'simulations' in Question 7, what duration would you need (max score of 2)? For all variables, there is high variability, but this question probably isn't very important given the lack of interest in "simulations".



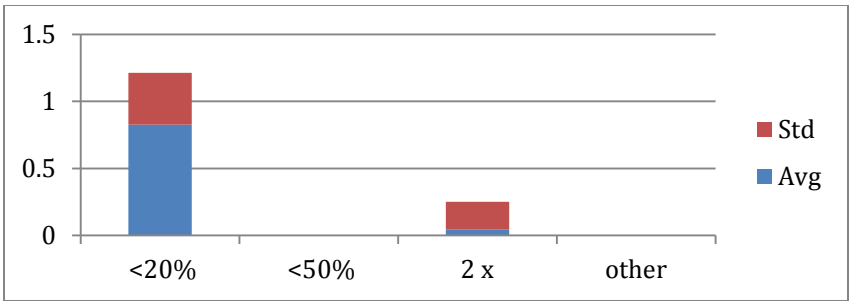
Q10: If you indicated interest in 'forecasts' in Question 7, how often do you need updates (max score of 2)? Daily updates are desirable for forecasts (velocity shown below).



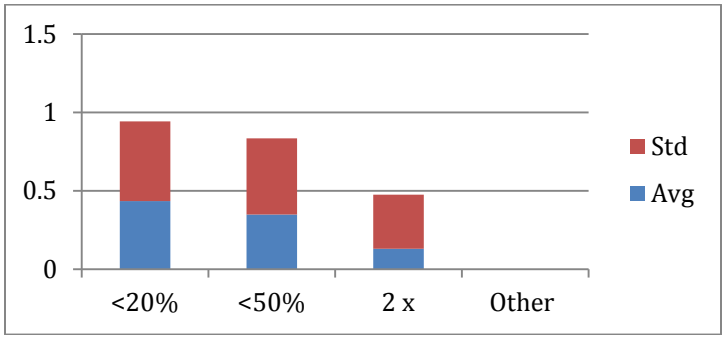
Q11: If you indicated interest in 'forecasts' in Question 7, in what time horizon are you interested (max score of 2)? There is interest in all time horizons though some expressed skepticism about the accuracy of a 2 mo forecast. (Note the high variability in the answers which suggests differences in the average are not significant.)



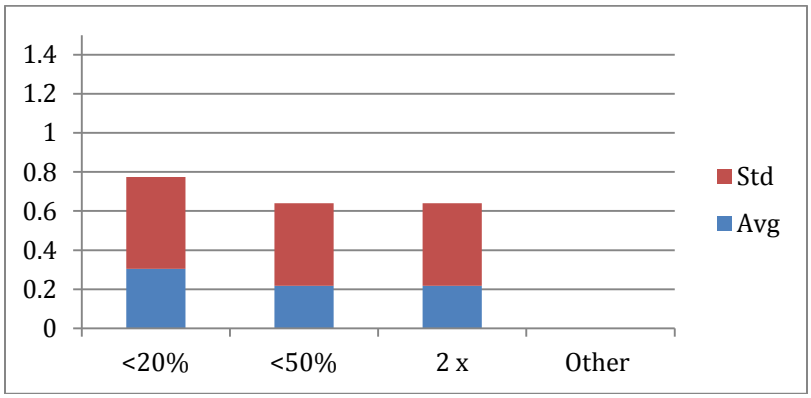
Q12: If you indicated interest in ' forecasts' in Question 7, what *maximum* uncertainty in the amplitude of each variable of interest is required at the end of two days to make the forecast useful (max score of 1)? For velocity, almost everyone wants better than 20% accuracy. For the other variables, there is a bit more willingness to accept less accuracy.



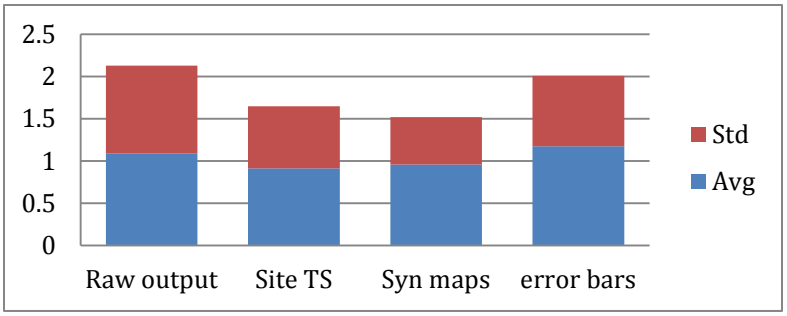
Q13: If you indicated interest in ' forecasts' in Question 7, what *maximum* uncertainty in the amplitude of each variable of interest is required at the end of 1-2 weeks to make the forecast useful (max score of 1)? Most want < 20% but some will tolerate <50% and a few will even tolerate 2x error.



Q14: If you indicated interest in ' forecasts' in Question 7, what *maximum* uncertainty in the amplitude of each variable of interest is required at the end of 1-2 mo to make the forecast useful (max score of 1)? The majority still want <20% error but there is a noticeable increase in those willing to tolerate up to 2x. Indeed, given the high variability, there is statistically little difference in the averages.



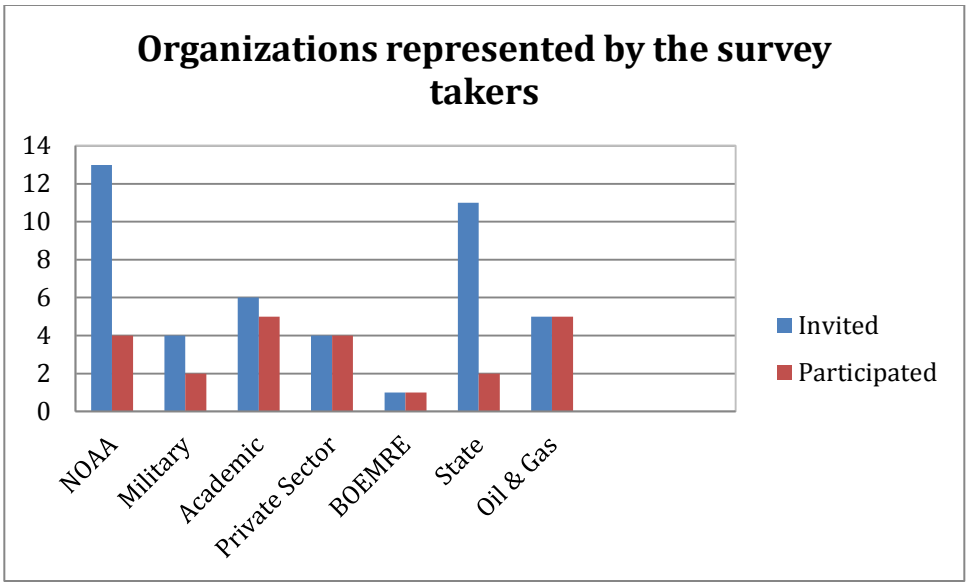
Q15: What kind of products are you interested in accessing from a forecast model (max score of 2)? All products listed were desired just about equally.



Survey Takers and Affiliations

Name	Organization
Frank Aikman	NOAA/National Ocean Service/Coast Survey Development Laboratory
Art Allen	U.S. Coast Guard/Search and Rescue
Adam Bangs	BHP Billiton Petroleum
Peter Brickley	Horizon Marine Inc.
Kjersti Broserud	Statoil
Cortis Cooper	Chevron
David Driver	BP America
Shejun Fan	FugroGEOS
Amy Godsey	Florida Department of Emergency Management
Peter Hamilton	SAIC
John Harding	Northern Gulf Institute
Ruoying He	North Carolina State University
Rob Hetland	Texas A&M University
Pat Hogan	Naval Research Laboratory
Doug Levin	NOAA/Integrated Ocean Observing System (IOOS)
Alexis Lugo-Fernandez	Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE)
Jim O'Brien	Florida State University/Center for Ocean-Atmospheric Prediction Studies
Claire Paris	University of Miami/Rosenstiel School of Marine and Atmospheric Science
Debbie Payton	NOAA/Office of Response and Restoration
David Peters*	Conoco Phillips
Mitch Roffer	Roffer's Ocean Fish Forecasting Service, Inc. (ROFFS)
Charles Sun	NOAA/National Oceanographic Data Center
Steve Wolfe	Florida/Department of Environmental Protection
Chris Yetsko*	Conoco Phillips

*One survey completed jointly was submitted



Fifty six individuals representing 44 organizations were invited to take the survey. Of these, 24 individuals, representing 23 organizations submitted completed surveys. There were 11 individuals who did not reply to the survey. The low participation from state programs is not because there was no interest, but rather, at least for the Gulf States, state climatologists and emergency managers have no in-house capability for numerical modeling. Except for the Florida Emergency Manager who has modeling skills, each referred us to individuals representing academia or NOAA, who were already included in the survey invitation and identified as their source of information. Likewise, of the 13 NOAA organizations invited to participate, there was overlap in the individuals from different branches who share modeling expertise. If individuals within a particular NOAA office did not possess the skill set, colleagues in other divisions were referred--most of whom were already included in the survey invitation. In one instance, a modeler declined to participate stating the lack of model resolution in estuaries and the near-shore environment as the reason.

User-Survey: The questionnaire sent to the 54 invited prospective survey participants

Operational Ocean Current Forecast Model in the Gulf of Mexico

A survey conducted on behalf of the Research Partnership to Secure Energy for America (RPSEA)

Background

With the support of DOE, RPSEA is sponsoring, as part of its Ultra-Deepwater Program, the "Gulf 3-D Operational Current Model Pilot Project" to demonstrate a state-of-the-art prediction system that can be transitioned to operational use. (Here, the term operational prediction system is used in the standard fashion: real-time, routine, automated, continually operating, skill-assessed through a transitional process, and sustained support.)

The prediction system addresses the circulation (i.e., 3-D currents, temperature, and salinity) over the full water column of the entire Gulf of Mexico, especially the Loop Current and its associated large and small mesoscale phenomena (i.e., meandering jets, eddies, and fronts) and the response of the Gulf of Mexico to the passage of summertime tropical cyclones and wintertime cold fronts. The ocean prediction interests of the offshore oil & gas industry, and those of the GCOOS-RA, are taken into consideration. (Note: here, prediction systems for surface gravity waves, tides, storm surges and coastal inundation and tsunamis are not addressed. However, NOAA has well-established operational prediction systems for such phenomena.)

The project is not building a numerical model from scratch but instead it is utilizing well-established modeling systems. More specifically, it is evaluating several R&D quasi-operational modeling systems together with Navy and NOAA operational prediction systems. A candidate operational ocean prediction system for the Gulf of Mexico will be evaluated, demonstrated, and recommended. It may be based on a single model or an ensemble of models.

A Concept of Operations (CONOPS) will be part of the recommendation. As entities involved with routine marine operations, emergency marine management, or marine ecosystem and living marine resource management in the Gulf of Mexico, you are requested to respond to the below questions.

Responses received by COB Friday 23 February 2011 will be assured consideration in the design of skill assessment metrics and other prediction system attributes. Respondents will be invited to a workshop where the outcome of the project will be briefed.

Name:_____

Organization:_____

Position:_____

Date:_____

Email:_____

Instructions: You may print this and fill it out with a pen/pencil. If so, please circle your answers. If you complete the form electronically, please bold or highlight your answers.

Please email the completed survey to simo@marine.usf.edu or mail to: Chris Simoniello, Gulf of Mexico Coastal Ocean Observing System, 4740 14 Avenue North, St. Petersburg, FL 33713

Note: For all questions that have "other" as an option, please specify your answer.

Question 1

Which of the following best describes your organization's activities in the Gulf of Mexico? Circle all that apply.

a. Search and rescue
b. Offshore oil/gas operations
c. Marine transportation
d. Commercial fishing
e. Recreational (fishing, sailing, diving, etc.)
f. Fisheries management
g. Hypoxia monitoring or forecasting
h. Harmful algal bloom monitoring or forecasting
i. Environmental impact of various pollutants
j. Ocean research (please specify)_____
k. Other (please specify)_____

Question 2

In which processes are you interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Rank	Process
	a. Summer storms (tropical cyclones and easterly waves)
	b. Winter storms (extratropical cyclones and cold fronts)
	c. Loop Current and its eddies and fronts
	d. Coastal upwelling and downwelling
	e. River outflow plumes and fronts
	f. Background current (periods >24 hrs)
	g. Tides
	h. Other (please specify)_____

Question 3

In which of the following Gulf of Mexico areas are you interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Rank	Area
	a. Texas
	b. Louisiana
	c. Mississippi
	d. Alabama
	e. Florida
	f. Mexico
	g. Cuba

Question 4

What water depth regime of the Gulf of Mexico is of interest to you?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Rank	Area
	a. Estuary
	b. Nearshore (< 30 ft)
	c. Shelf
	d. Deep water (>600 ft)

Question 5

In what part of the water column are you interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

	Surface	Middle	Bottom
Application			
Current velocity			
Water Temperature			
Water Salinity			
Sea Surface Height			
Other-specify			

Question 6

In which variables and spatial scales are you most interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	Specific Site	Synoptic View (see footnote)	Other (specify)
Current velocity			
Water Temperature			
Water Salinity			
Sea Surface Height			
Other-specify			

By 'Synoptic' scale we mean maps of main features like the boundary of the Loop Current

Question 7

In which time frames are you interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	Hind.	Nowcast	Forecast	Sim.	Other
Current velocity					
Water Temperature					
Water Salinity					
Sea Surface Height					
Other-specify					

By 'Simulation' we mean hypothetical simulations to investigate process origins, changes to the environment, etc.

Question 8

If you indicated interest in 'hindcasts' in Question 7, what duration would you need?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	< 1 yr	1-5 y	5-10 y	>10 y
Current velocity				
Water Temperature				
Water Salinity				
Sea Surface Height				
Other-specify				

Question 9

If you inidicated interest in 'simulations' in Question 7, what duration would you need?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	< 1 yr	1-5 y	5-10 y	>10 y	Other
Current velocity					
Water Temperature					
Water Salinity					
Sea Surface Height					
Other-specify					

Question 10

If you indicated interest in 'forecasts' in Question 7, how often do you need updates?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	3-h	12-hr	Daily	1x/week
Current velocity				
Water Temperature				
Water Salinity				
Sea Surface Height				
Other-specify				

Question 11

If you inidicated interest in 'forecasts' in Question 7, in what time horizon are you interested?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Application	1-2 days	out to 4 days	1-2 weeks	1-2 months	Other
Current velocity					
Water Temperature					
Water Salinity					
Sea Surface Height					
Other-specify					

Question 12

If you indicated interest in ' forecasts' in Question 7, what ***maximum* uncertainty in the amplitude**

of each variable of interest is required at the end of two days to make the forecast useful?

Enter 'x' only for vairables of interest

Application	<20%	<50%	Factor of 2	Other
Current velocity				
Water Temperature				
Water Salinity				
Sea Surface Height				
Other-specify				

Question 13

If you indicated interest in ' forecasts' in Question 7, what *maximum* uncertainty in the amplitude

of each variable of interest is required at the end of 1-2 weeks to make the forecast useful?

Enter 'x' only for vairables of interest

Application	<20%	<50%	Factor of 2	Other
Current velocity				
Water Temperature				
Water Salinity				
Sea Surface Height				
Other-specify				

Question 14

If you indicated interest in ' forecasts' in Question 7, what *maximum* uncertainty in the amplitude

of each variable of interest is required at the end of 1-2 months to make the forecast useful?

Enter 'x' only for vairables of interest

Application	<20%	<50%	Factor of 2	Other
Current velocity				
Water Temperature				
Water Salinity				
Sea Surface Height				
Other-specify				

Question 15

What kind of products are you interested in accessing from a forecast model?

Please rank 1=high; 2=medium; 'blank'=little or no interest.

Rank	Product
	a. Raw model output
	b. Plotted time series at specified sites
	c. Synoptic maps at specified time periods
	d. Plotted profiles
	e. Uncertainty limits/error bars
	f. Other (please specify)_____

Thank you for participating in this survey!

Appendix II: GOMEX-PPP Technology Transfer

Peer-Reviewed Publications:

(1) Chang, Y.-L., and L.-Y. Oey, 2012:Why does the Loop Current tend to shed more eddies in summer and winter? *Geophys. Res. Lett.*, doi:10.1029/2011GL050773, in press.

(2) Zhao, Y. and He, R., Cloud-free Sea Surface Temperature and Color Reconstructions for the Gulf of Mexico: 2003-2009, *Remote Sensing Letter*, in press.

Presentations at Scientific Meetings:

AMS Annual Meeting 2012 held in New Orleans, Louisiana, 22 to 26 JAN 12:

Mooers, C.N.K. and E. Zaron. 2012. Multi-Model Comparisons of the Mesoscale Circulation in the Gulf of Mexico in 2010. Presented at the AMS Annual Meeting, 22-26 January 2012, New Orleans, LA. (a contribution to the embedded AMS 10th Symposium on the Coastal Environment).

AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting 2012 held in Salt Lake City, Utah, 20 to 24 FEB 12:

The presentations below comprised one-half of those made in the Special Session on Gulf of Mexico Circulation and Ecosystem Numerical Modeling.

Chang, Y. L. and L.-Y. Oey. 2012. Why Does the Loop Current Have Seasonal Preferences for Shedding Eddies? Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Farrara, J.D., Y. Chao, Z. Li, X. Wang, H. Zhang, P. Li, R. He, and H. Qian. 2012. A ROMS-based Data Assimilating Ocean Forecast System for the Gulf of Mexico. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Gopalakrishnan, G., B. Cornuelle, I. Hoteit, D. Rudnick, and W. Owens. 2012. State Estimates and Forecasts in the Gulf of Mexico. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Howard, M.K., E. Zaron, C. Mooers, Y. Chao, B. Cornuelle, R. He, D. Ko, L. Oey, A. Mehra, and R. Patchen. 2012. Gulf of Mexico Pilot Prediction Project (GOMEX-PPP): Model-Data Comparisons. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Ko, D.S. 2012. A Long-Term Ocean Forecast Experiment for Gulf of Mexico Applying IASNFS. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Mooers, C.N., E.D. Zaron, M. Howard, Y. Chao, B. Cornuelle, R. He, D.S. Ko, L. Oey, A. Mehra, and R. Patchen. 2012. The Gulf of Mexico Pilot Prediction Project (GOMEX-PPP). Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Wiggert, J.D., J.M. Harding, F.L. Bub, P.J. Fitzpatick, and K.C. Woodward. 2012. Evaluation of the AMSEAS Gulf of Mexico/Caribbean Regional Forecast System: A SURA Super-Regional Modeling Testbed Activity. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Xue, Z., R. He, K. Fennel, W. Cai, and S. Lohrenz. 2012. Modeling Seasonal and Interannual Variability of Circulation and Biogeochemical Processes in the Gulf of Mexico. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.

Zaron, E.D., C.N. Mooers, M.K. Howard, Y. Chao, B. Cornuelle, R. He, D.S. Ko, A. Mehra, L.Y. Oey, and R. Patchen. 2012. Gulf of Mexico Pilot Prediction Project (GOMEX-PPP): Forecast Skill and Model Intercomparisons. Presented at the AGU, TOS, & ASLO Sponsored Ocean Sciences Meeting, 20-24 February 2012, Salt Lake City, UT.