

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dottorato di Ricerca in:
GEOFISICA
Ciclo XXIII
Settore Scientifico-Disciplinare di afferenza : GEO/10

**Development of SuperEnsemble Techniques
for the Mediterranean ocean
Forecasting System**

**Presentata da:
Jenny Pistoia**

**Relatore:
Prof. Nadia Pinardi
Co-relatori:
Paolo Oddo
Matthew Collins**

Acknowledgements

I wish to thank my supervisor Prof. Nadia Pinardi for having let me become involved in the research project on Superensemble Techniques, and Dr. Paolo Oddo his very valuable collaboration and overall for his PATIENCE.

I am very grateful to Prof. Matthew Collins for his kind hospitality at Exeter University and for having wisely supervised my work on EOFs/PC Analysis.

During my university years I met the best Professor allover the world, A.F. I wish to thank him for his advices and our talks, which are the best gift I have ever had. And I have to thank him since he let me met Him.

Finally I would like to thank the CMCC for having financed my travel to SLC and for having fully covered my visit to Exeter University.

*... These turbulent fluctuations, embedded in a larger flow, tend to drain the large-scale flow of energy by a variety of mechanical processes and in turn pass the energy to finer scale of motions where viscosity can act directly. This notion of the cascade of energy from the largest to the smallest scale of motion is far from clear and rigorous. [Pedlosky]
But It is the most challenging and appealing phenomenon*

Development of SuperEnsemble Techniques for the
Mediterranean Ocean Forecasting System

Contents

Introduction	i
The Mediterranean ocean forecasting system	i
The problem of uncertainty in ocean modelling	iii
State of art of multi model SuperEnsemble	iii
Study of multi model Superensemble	v
High resolution model	v
Thesis Objectives	vi
Structure of the thesis	vi
1 Uncertainties and model errors	1
Background	1
1.1 Errors in initial conditions and atmospheric forcing	1
1.2 From the closure problem to the sub-grid scale parametrizations	4
1.2.1 Lateral parametrization	8
1.2.2 Parametrization in the vertical direction	9
1.3 The Primitive Equation	10
1.4 Errors of numerical approximations	13
1.4.1 Round-off errors	13
1.4.2 Truncation errors	14
1.5 Error Propagation	14
2 Building the SuperEnsemble dataset	17
Overview	17

2.1	Multiphysics SuperEnsemble Experiment	20
2.1.1	The basic model configuration	21
2.1.2	Creation of the ensemble members	22
2.2	BHM SuperEnsemble Experiment	25
2.3	Multi-Model Multi-Physics SuperEnsemble Experiment	25
2.3.1	Analysis mode	28
2.3.2	Forecast Mode	30
3	Classical SuperEnsemble approach	37
	Overview	37
3.1	Methodology	38
3.2	Results	41
	Verification methods	41
3.2.1	Results Multi-physics experiment	43
3.2.2	Results in the BHM experiment	44
3.2.3	Results MultiModel Multiphysics experiment	49
3.2.3.1	Results in the MMSE experiment Analysis Mode	49
3.2.3.2	Study on the sensitivity of the algorithm	55
3.2.4	Results in the MMSE experiment Forecast Mode	63
3.3	Summary	68
4	Improving the classical MMSE approach	75
4.1	Background	75
4.2	Formulation and computation of EOFs	76
4.3	Results with EOFs analysis	79
4.3.1	Multiphysics Experiment	79
4.3.2	BHM experiment	79
4.3.3	Multimodel Multiphysics experiment Analysis mode	83
4.3.4	Multimodel Multiphysics experiment Forecast mode	88
	Conclusions	95

A Numerical schemes available in NEMO	99
A.1 Vertical Physics	99
A.1.1 KPP turbulent closure scheme	99
A.1.2 TKE scheme	99
A.1.3 The Pacanowski and Philander (PP) scheme	100
A.2 Tracer Advections	100
A.3 Viscosity Operator	101
A.3.1 Laplacian Operator	101
A.3.2 Bilaplacian Operator	102
B High resolution model	105
B.1 Vertical model discretization	107
Bibliography	117

List of Figures

1	ii
1.1	Differences °C the twin experiment simulations	15
2.1	Practical Example corresponding to the idealized " <i>good ensemble</i> "	18
2.2	<i>Mean Variability generated in one month by the Multi-Physics ensemble (left) versus the natural variability in the observations.</i>	24
2.3	<i>Standard Deviation of Multi-physics members (left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).</i>	24
2.4	<i>Standard Deviation of BHM ensemble members (left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).</i>	26
2.5	<i>Mean Variability generated by the BHM ensemble (left) versus the natural variability in the observations (right) during the training period</i>	27
2.6	<i>Mean Variability generated by the MMSE (left) versus the natural variability in the observations (right) during the time period, year 2008</i>	31
2.7	<i>Standard Deviation of MMSE members in analysis mode(left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).</i>	31
2.8	<i>Mean Variability generated by the MMSE in forecast mode (left) versus the natural variability in the observations (right) during the time period, 3day forecast lead time</i>	33
2.9	<i>Standard Deviation of MMSE members in forecast mode(left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).</i>	35
3.1	Anomaly of a simple dataset	40
3.2	resulting maps for the first experiment, SE on the top, SST Satellite-derived(middle) and the worst MP member (bottom)	43

3.3	resulting maps for the first experiment, SE on the top, SST Satellite-derived(middle) and the worst MP member (bottom)	45
3.4	<i>RMSE between truth observation and MP members (coloured bar) and SE (brown bar) for SST field, during the five days of the test period</i>	46
3.5	<i>performance during the training period of the SE prediction(red), OPA BHM members(black lines) and NEMO BHM members(blue lines).RMSE (top panel) Bias Removed RMSE (lower panel)</i>	47
3.6	<i>Domain averaged Bias Removed RMSE during the test period.</i>	47
3.7	<i>BIAS during the test period.</i>	48
3.8	<i>ACC during the test period.</i>	48
3.9	<i>MSESS for the 22 BHM MM members and SE for the parameter SST over Rhode Gyre region during the test period(March 2008). Black bars represent gain respect OPA predictions, and blue bars the gain respect NEMO</i>	48
3.10	<i>A typical example of a reconstructed map of SST, on day five over Rhode Gyre region, valid the 5 March 2008: analysis (top left), lowest skill model (bottom left), SuperEnsemble (top right) and best participating model (bottom right).</i>	50
3.11	<i>Schematic representation of the experiment carried in MMSE analysis mode.</i>	51
3.12	<i>A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of June 2008 reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)</i>	52
3.13	<i>A typical example of predictor for the last day of test period over the Mediterranean area, valid on the 10th of June 2008, reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)</i>	53
3.14	<i>Time serie of the domain average (over the Mediterranean) Bias removed RMSE</i>	53
3.15	<i>Anomaly Correlation Coefficient time series</i>	54
3.16	<i>Bias (Model - Observation) Time serie over the Mediterranean sea)</i>	54

3.17	Domain average RMSE comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line represent shorter training period, yellow line display longer training period	56
3.18	ACC comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line represent shorter training period, yellow line display longer training period	56
3.19	BIAS comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line represent shorter training period, yellow line display longer training period	57
3.20	Bias Removed RMSE mean Skill of the Multi Model superensemble with 9 members against the trainin period legths (green spots) the error bar is the standard deviation	57
3.21	ACC mean Skill of the Multi Model superensemble with 9 members against the trainin period legths(red spots) the error bar is the standard deviation	58
3.22	Bias Skill of the Multi Model superensemble with 9 members against the trainin period legths(black spots) the error bar is the standard deviation	58
3.23	Comparison of the mean skills for BIAS(black spots), ACC (red spots) and Centred RMSE(green spots)	59
3.24	Domain average	60
3.25	<i>A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of June 2008 reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)</i>	61
3.26	<i>A typical example of prediction for the last day of test period over the Mediterranean area, valid on the 10th of June 2008, reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)</i>	62

-
- 3.27 Bias Removed RMSE mean Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are Bias Removed RMSE mean Skill for the subsample A, the green bars are the standard deviations. 63
- 3.28 ACC mean Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bar are the standard deviations. 64
- 3.29 BIAS Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bar are the standard deviations. 64
- 3.30 Domain average RMSE comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C 65
- 3.31 ACC comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C 66
- 3.32 BIAS comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C 67
- 3.33 *A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 10th of November 2008 reached with 21 days of training period, SE prediction (top left), lowest skill model (bottom left), SST Satellite derived (top right) and high resolution output (bottom right) 69*

3.34	centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (21 days) during the year 2008 with forecast period three days.	70
3.35	centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (20 days) during the year 2008 with forecast period five days.	71
3.36	centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (18 days) during the year 2008 with forecast period nine days.	72
4.1	<i>Eigenvalues spectrum for the weights mask in the BHM dataset</i>	78
4.2	<i>Eigenvalues spectrum for the weights mask in the BHM dataset</i>	79
4.3	<i>first prediction day of the classical SE (top right) SST satellite derived (top left), SE EOFs based (bottom left) and one of the model of the ensemble</i>	80
4.4	<i>RMSE between truth observation and MP members (coloured bar) and SE EOFs based (brown bar) for SST field, during the five days of the test period</i>	80
4.5	<i>A typical example of a reconstructed map of SST, on day three over Rhode Gyre region, valid the 3 March 2008: analysis (top left), SE EOFs based (bottom left), SuperEnsemble classic (top right) and best participating model (bottom right). . .</i>	81
4.6	<i>Domain averaged RMSE between the OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period</i>	82
4.7	<i>MSESS for the 22 BHM MM members and SE EOFs for the parameter SST over Rhode Gyre region during the test period (March 2008). Black bars represent gain respect OPA predictions, and blue bars the gain respect NEMO</i>	82
4.8	<i>ACC, OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period</i>	82
4.9	<i>BIAS, OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period</i>	83

4.10	<i>A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of July 2008 reached with 55 days of training period, SE prediction (top left), SE EOFs based prediction (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)</i>	84
4.11	<i>A typical example of prediction for the last(10th) day of test period over the Mediterranean area, valid on the 1st of July 2008 reached with 55 days of training period, SE prediction (top left), SE EOFs based prediction (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)</i>	85
4.12	<i>Performances of SE obtained by 10 days of training period, during year 2008, left panels the classical approach, right side the SE EOFs based approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.</i>	86
4.13	<i>Performances of SE obtained by 30 days of training period during year 2008, left panels the classical approach, right side the SE EOFs based approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.</i>	87
4.14	<i>Geographical distribution of the regression coefficients for the first model member, as the simple anomaly field (top) and for the filtered anomaly field (bottom) after 100 days of training period, note the different colorbars</i>	89
4.15	<i>Statistical coefficient from Krishnamurti et al. (2003)</i>	90
4.16	<i>centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the same training period (9 days) during the year 2008 with forecast period nine days.</i>	91
4.17	<i>centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the training period of 18 days, for the 1st of March 2008, with forecast period nine days</i>	92
4.18	<i>centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the training period of 27 days, for the 1 November 2008, with forecast period nine days</i>	93
B.1	<i>New model bathymetry and domain for the MFS2491 set up</i>	106
B.2	<i>new Bathymetry near Messina Strait</i>	106
B.3	<i>new Bathymetry near Iskenderum Gulf</i>	107

B.4 *Vertical level distribution versus the depth of water column as computed by NEMO-OPA, new one 91 levels in red, in black the old distribution with 72 levels . . .* 109

B.5 *Histogram of relative frequency number of level for each interval depth, in old version in red, while the new in yellow.* 109

List of Tables

2.1	<i>Experiment 1 design: Model configurations used in the multi-physics ensemble. Member number 1 is the model basic configuration (Default version). All the other member are obtained changing only one numerical scheme per time from the Default version.</i>	23
2.2	<i>Set up of BHM SuperEnsemble Experiment, the columns are the members, while the rows are the difference between each member.</i>	26
2.3	<i>Set up of MultiModel SuperEnsemble Experiment obtained by MyOcean database (on the top) and the multi-physics members(on the bottom of the table)</i>	30
2.4	<i>Experiment design for MultiModel SuperEnsemble in forecast configuration, BHM multi-physics members (on the top of the table) and MyOcean database(on the bottom)</i>	34

Introduction

The Mediterranean ocean Forecasting System

The Mediterranean ocean Forecasting System (MFS) is working operationally since 2000 and is being continuously improved through the framework of international projects. The system is part of the Mediterranean Operational Oceanography Network -MOON- and MFS is coordinated and operated by the Italian Group of Operational Oceanography (GNOO) (Tonani *et al.*, 2009). Since year 2009 it is part of the MyOcean system (www.myocean.eu) becoming Med-Monitoring and Forecasting System (Med-MFC) (M.Tonani *et al.*, in press).

The numerical model component of Med-currents is composed by two elements: an Ocean General Circulation Model (OGCM) and a Wave Model. The OGCM code is NEMO-OPA (Nucleus for European Modelling of the Ocean-Ocean PARallelise) version 3.2 (Madec, 2008). The code is developed and maintained by the NEMO-consortium. The model is a primitive equation in spherical coordinates. The Wave Model is based on the WAM (Wave Analysis Model) code. NEMO-OPA has been implemented in the Mediterranean at 1/16 deg. x 1/16 deg. horizontal resolution and 72 unevenly spaced vertical levels (Oddo *et al.*, 2009). Since September 2005 the system produces short-term ocean forecasts for the next ten days and the production is on a daily basis, while before it was weekly and it is coupled off-line with a biogeochemical forecasting system and a wave model. Every day (J) the system produces 10 days forecast from J to J+9, as shown in Figure 1. On Tuesdays, 15 days of analysis are produced, from J-15 to J-1, with the assimilation of all the available satellite and *in situ* data. Med-biogeochemistry 10-day forecast

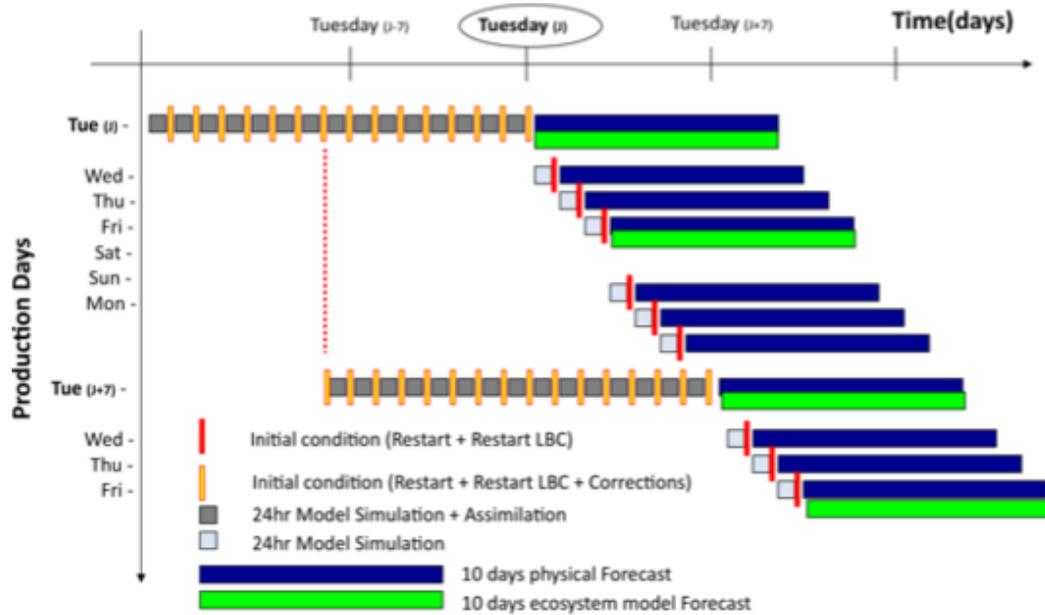


Figure 1:

is produced bi-weekly on Tuesday and on Friday (see Figure 1). All days, except Tuesday, a 24-hour simulation is performed (from $J-1$ to J) in order to get the best initial condition for the forecast. The simulation will be different from the forecast produced the previous day ($J-1$) for the atmospheric forcing, which is an analysis field instead of a forecast. All the products are validated and assessed in near real time via comparison with dependent and semi-independent observations (Tonani et al 2009). A real time network has been developed for this purpose in collaboration with the MOON community (Mediterranean Operational Oceanography Network) in order to collect all the available moored observations for temperature, salinity, currents and sea level. All the information collected via this network are elaborated by and ad hoc-software in order to evaluate the quality of the Med-MFC products (<http://gnoo.bo.ingv.it/myocean/calval>).

The problem of uncertainty in ocean modelling

The first successful numerical weather prediction occurred in the 1950s, and since that date many important improvements have been made in the accuracy of numerical weather prediction models.

Some advances depend on the steady developments in high-performance computing resources which has permitted the increase in the spatial resolution of numerical prediction models permitting to resolve different scales of motion. The accuracy of the numerical prediction models has also benefited from improvements in the way physical processes and motions on the subgrid-scale are represented as well as improvements in the representation of the interaction with the ocean and the land surface. Nevertheless, some elements, such as initial condition uncertainties, atmospheric forcing inaccuracy and model errors, being intrinsically related to truth state of the fluid and the way that we use to describe it, cannot be easily removed, giving an important contribution to errors growth.

State of art of multi model SuperEnsemble

Taking into account the previous statements about the impossibility of eliminate forecast errors, many techniques, such as ensemble forecasting, have been developed recently. The goal of ensemble forecasting is to predict the probability of future weather events as much precisely as possible (Epstein 1969; Leith 1974; Mullen and Baumhefner 1994). Forecasts are sensitive to both small uncertainties in the initial condition (Lorenz 1963) and model error (Harrison et al. 1999), so a deterministic prediction may not be reasonable and then fail. An ensemble forecast system starts by the creation of equally likely analyses of an initial state in which the ensemble mean is the best estimate of the true state of the fluid we want to describe (both atmosphere and ocean). As the forecast time runs the trajectory of ensemble member continue to diverge. For long enough time (depending on the system) they will differ so much that they will be impossible to be distinguished from another randomly chosen states.

Operationally, ensemble prediction systems have been made changing the initial con-

ditions, adding the perturbation field to the analyses. In this way, we assume that forecast errors arise from analysis errors while the model is considered perfect. The National Centers for Environmental Prediction (NCEP) define these perturbations through bred modes (Toth & Kalnay, 1993) which represent the fastest growing perturbations and simulate the development of growing errors in the analysis cycle. The European Center for Medium-Range Weather Forecasts (ECMWF), instead, uses singular vectors (Buizza & Palmer, 1995) which maximize the linear growth of energy over a specified domain. moreover, The Canadian Meteorological Center (CMC) applies the ensemble Kalman filter (EnKF; Evensen, 2003; Houtekamer *et al.*, 1996) which provides an ensemble of initial conditions thanks to the assimilation of an ensemble of perturbed observations.

in order to consider both initial conditions and model errors, multimodel approach, statistical techniques involving the use of different forecast models (e.g. Evans *et al.*, 2000; Krishnamurti *et al.*, 2000; Stensrud, 2001), different physical parameterization schemes (e.g. Houtekamer *et al.*, 1996; Stensrud *et al.*, 1999) or stochastic physical parameterization (e.g. Palmer, 2001; Grell and Devenyi, 2002) have been developed. others use a multimodel formed by a combination of several operational products plus perturbations of these runs, like Fujita *et al.* (2007) in their exploration of various ensembles dealing with the influence of initial conditions and multiphysics or Meng and Zhang (2007) in their test of an EnKF for data assimilation. In spite of the wide range of existing building techniques it is still unclear which one is the best. Numerous studies compare not only different sources of error (model or analyses) but different techniques to deal with each source. For example, Stensrud *et al.* (2000) and Fujita *et al.* (2007) compare initial conditions and model physics uncertainties in an ensemble, Clark *et al.* (2008) compares the contributions of initial/lateral boundary conditions and mixed physics to the spread and skill of 120-h precipitation ensemble forecasts while Meng and Zhang (2007) tested the sensitivity of an EnKF to model error induced by different cumulus parameterizations. in Krishnamurti's approach (Krishnamurti *et al.*, 1999) he used an unbiased linear combination of the available models, optimal (in the least-squares sense) with respect to observations during a training period of a priori chosen length; all observations have

equal importance. Lenartz *et al.* (2010) introduced dynamically evolving weights in a linear combination of models, using data assimilation techniques (Kalman filter and particle filter) adapted to the super-ensemble paradigm. These techniques are able to train the weights on a time-scale corresponding to their natural characteristic time, discarding older information automatically. The weights rate of change is determined by the respective (and evolving) uncertainties of the weights themselves, of individual models and of observations.

Study of multi model Superensemble

Because uncertainties arise from the initial conditions, improper model physics parametrization schemes, and atmospheric forcing inaccuracy, a good ensemble may need to contain all these aspects. The intent of this thesis was to evaluate the impact of model error, initial condition and atmospheric forcing throughout several experiments. The first was addressed to model parametrizations scheme, since presence of model error can often produce both a large bias of the ensemble mean and too little spread, bringing the ensemble algorithm to fail (Meng & Zhang (2011)); in the second experiment, uncertainties have been sampled both in initial conditions and atmospheric forcing by a MFS-BHM winds. Finally, in the third experiment, we used a wider approach addressed to remove the systematic error by using different research institute estimates (analysis and forecast). SuperEnsemble performances have been evaluated in every experiment on the base of every dataset created.

High resolution model

All numerical simulations of atmospheric and oceanic phenomena are limited by the finite spatial resolution, generally requiring a parametrization of effects of motions on unresolved scales on those explicitly resolved. A goal of numerical modellers has been to resolve as many scales of the actual circulation as practically possible. With the recent advent of a new generation of high-performance computing resources some notable thresholds in terms of model resolution have been approached or, in

some cases, surpassed. In this thesis starting from the default version of the Ocean General Circulation employed in MFS, we implemented a higher resolution model in the Mediterranean sea.

Thesis Objectives

This research activity studied how the uncertainties are concerned and interrelated through the multi-model approach, since it seems to be the bigger challenge of ocean and weather forecasting. Moreover, we tried to reduce model error throughout the superensemble approach. In order to provide this aim, we created different dataset and by means of proper algorithms we obtained the superensemble estimate. We studied the sensitivity of this algorithm in function of its characteristics parameters. Clearly, it is not possible to evaluate a reasonable estimation of the error neglecting the importance of the grid size of ocean model, for the large amount of all the sub grid-phenomena embedded in space discretizations that can be only roughly parametrized instead of an explicit evaluation. For this reason we also developed a high resolution model, in order to calculate for the first time the impact of grid resolution on model error.

Structure of the thesis

This thesis is organized as follows: In chapter 1 we will describe the wide range of uncertainties concerning the numerical modelling. The differences between "intrinsic" and "structural" uncertainties. Furthermore, the perfect model assumption must be dropped in real world studies where computational constraint allow us to parametrize some phenomena instead of an explicitly evaluation. We will explain the differences between the Navier-Sokes Equations and Primitive equations. Finally we will give a practical example of the error propagation. A large part of the effort of this thesis was dedicated to method for the creation of suitable ensemble members used in the experiments trained are presented in chapter 2. Furthermore we commented the goodness of our ensemble on the basis of *ad hoc indexes*. In chapter 3

we will show the methodology employed, the Krishnamurti Superensemble Concept, and the results due to the employment of this technique in ocean ensemble datasets. In chapter 4 we try to bring improvement to the classical superensemble approach considering the EOFs/PC Analysis, so after an overview of how does this statistical techniques can decompose a space and time dependent field into a set of spatial patterns and associated time indices, giving the linking between their development and the application in our case.

Chapter 1

Uncertainties and model errors

Background

Uncertainties in numerical modelling arise both from *Intrinsic Uncertainties* such as in the initial conditions and boundary (surface and lateral) conditions, and *Structural Uncertainties* resulting from the fact that some processes are not fully understood. Furthermore, the perfect model assumption must be dropped in real world studies where the computational constraints allow us to parametrize some phenomena instead of explicitly evaluated (hydrostatic approximation and rigid lid/free surface assumption) bringing model error growth. Inadequate representation of physical processes, parametrizations of the sub grid scale physical phenomena, numerical inaccuracies, and truncation errors arrange that numerical output simulations will be different from the true state of the fluid we want to describe.

1.1 Errors in initial conditions and atmospheric forcing

Ocean observations are required for the model initialization, but the platforms and sensors providing a wide range of physical and biological measurements are sparse

both in space and time. The sampling error is one of the most relevant error in oceans sciences. In an ideal world, observation should sample all the physical processes, when using limited area model, observations are needed to prescribe at the lateral open boundary processes not occurring locally but affecting and influencing local dynamics. Ideally we should have one observation in each point of our domain, so assimilation techniques have been developed in order to meld these data with the dynamical models variables and produce the *best estimate* of the current oceanic state. Beyond this native lack of knowledge of the system we want to describe, ocean-earth's system interactions are approximate and ocean boundary conditions are inexact. These sources of uncertainties together with model error (approximations to equations and computational errors) lead to our estimate will be different from the truth, and we will say that our estimate is affected by an error, Lermusiaux (2006). All these sources of error are not really separable since the estimation of the initial conditions involves a numerical model as merger so initial condition error will be affected by model error and grown up in time.

The first theoretical studies on error growth in atmospheric prediction and on the implied limits of a prediction and probabilistic prediction appeared in the late 1950s to early 1970s. This area of meteorological research is referred as *predictability*. Let us focus on initial condition uncertainties first, as we have already said they can be estimated within a certain accuracy. Lorenz (1963) studied the growth of forecast errors due to initial condition uncertainties by looking at the rate at which solutions of the ECMWF numerical weather prediction model diverged, and realized that forced dissipative systems of ordinary differential equations that originate from highly truncated models of fluid flows, "*systematically exhibit sensitive dependence to initial conditions*".

In a deterministic system the state is uniquely determined from initial condition (no stochastic variables). Both atmosphere and ocean are deterministic systems, and they are also perpetually forced by dissipative terms (there are sinks/sources). Sensitivity to initial conditions means that two nearby trajectories separate exponentially fast. This implies that, even if we have a deterministic system two solutions starting from two almost equal states will separate and lose the similarities. It is

the non-linearity of the system that gives raise to the irregular behaviour.

The ocean exchanges mass and energy with the atmosphere through its surface. The surface boundary condition for momentum is:

$$A^{vm} \frac{\partial \mathbf{u}_h}{\partial z} \Big|_{z=\eta} = \frac{\tau}{\rho_0} \quad (1.1)$$

where A^{vm} is the vertical eddy viscosity, $\tau = (\tau_u, \tau_v)$ represents the zonal and meridional wind stress components and $\mathbf{u}_h = (u, v)$.

The water flux boundary condition states that a particle of water can enter or escape the sea surface only through precipitation or evaporation:

$$w = \frac{D\eta}{Dt} - (E - P) \quad (1.2)$$

E and P are evaporation and precipitation, $\frac{D}{Dt} = \frac{\partial}{\partial t} + u_h \Big|_{z=\eta} \cdot \nabla$ is the total derivative.

The water flux is coupled to the salinity boundary condition by the $(E - P)$ term with runoff(R) by:

$$A^{vs} \frac{\partial S}{\partial z} \Big|_{z=\eta} = (E - P - R) S_{z=\eta} \quad (1.3)$$

Finally, the boundary condition for heat flux is:

$$A^{vT} \frac{\partial \mathbf{T}}{\partial z} \Big|_{z=0} = \frac{Q}{\rho_0 C_p} \quad (1.4)$$

where C_p is the ocean heat capacity constant and Q is the heat budget and consists of the solar radiation flux Q_s minus the net long-wave radiation flux Q_B , the latent heat flux Q_E and the sensible heat flux Q_H . Knowledge of all the components of the heat and water budgets at the air-sea interface are key points in ocean modelling. Unfortunately, given the complexity of the processes involved, they must necessarily be parametrized. Several techniques have been historically applied and are widely described in the literature, providing different results depending on their approximations and accuracy. In the nineties and last ten years, heat fluxes started to be parametrized with empirical (bulk) formulae forced by atmospheric data sets rather

then the true measures of fluxes (due to their cost are too sparse both in space and time). The seminal work of Rosati & Miyakoda (1988) described the first OGCM application of this methodology. At present, this approach is the most often used, since it represents the best compromise. Atmospheric forcing data (wind, temperature, humidity, precipitation and cloud cover) originate from a variety of sources with different accuracy, the question arises how accurate the atmospheric data has to be in order to be useful for a realistic simulation of atmosphere state. Recent works have demonstrated a Kinetic Energy (KE) deficiency in global ocean surface winds provided by Numerical Weather Prediction (NWP) system with respect to coincident surface winds retrieved by scatterometer data (Chin *et al.* (1998) and Milliff *et al.* (2004)). A similar behaviour (ECMWF winds usually underestimates wind speed) has been recognized comparing ECMWF analyses and QuickScat data over Mediterranean basin (Milliff (2004)). Bayesian Hierarchical Model for surface winds (Milliff *et al.* (2011) and Pinardi *et al.* (2011)) is a statistical technique able to produce different estimates of winds as a combination of ECMWF products with QuickScat Data which sampling the uncertainties affecting the atmospheric field forcing (see the experiment described in 2.2). Furthermore all the atmospheric variable uncertainties are coupled through the bulk formulae, for example the heat fluxes are affected by wind perturbation through two processes: wind speed is an input in the parametrization scheme of sensible and latent heat, while variations of sea surface temperature due to wind driven circulation affect the parametrization schemes of sensible and latent heat and the net outgoing long-wave flux. For all that reasons air-sea interactions constitute a primary source of model error growth (Gould *et al.* (2001)).

1.2 From the closure problem to the sub-grid scale parametrizations

The numerical ocean modelling involves the seeking of solutions of a coupled set of non-linear partial differential equations (PDEs), called Navier-Stokes equations, able to describe the time-dependent behaviour of properties of a fluid flowing in three-dimensional space and acted upon by various forces, under the constraint of

conservation equations for some scalar properties such as sea temperature and salinity with appropriate source and sink terms. Each numerical model solves the basic conservation equation for mass (eq.(1.5)) and momentum (eq.(1.6)) suitably modified in order to describe the rotation of the reference system and the thermodynamic(1.8) or other dissipative processes which can take place in the ocean¹.

$$\frac{d\rho}{dt} + \rho \nabla \cdot \vec{v} = 0 \tag{1.5}$$

where $\vec{v} = (u, v, w)$ and ρ is the density;

$$\rho \frac{d\vec{v}}{dt} = -\nabla p - \rho \nabla \phi + \mathbf{F}(\vec{v}) \tag{1.6}$$

or that mass per unit volume times the acceleration is equal to the sum of the pressure gradient force, the body $\rho \nabla \phi$ where ϕ is the potential by which conservative body forces can be represented, and the force \mathbf{F} is in principle the frictional force in the fluid. For Newtonian fluid like air or water

$$\mathbf{F}(\vec{v}) = \mu \nabla^2 \vec{v} + \frac{\mu}{3} \nabla (\nabla \cdot \vec{v}) \tag{1.7}$$

where μ is the molecular viscosity. This is an exact representation for \mathbf{F} when μ , in principle a thermodynamic variable is taken as constant over the field of motion.

$$c_p \frac{dT}{dt} + \frac{T}{\rho} \alpha \frac{dp}{dt} = \frac{k}{\rho} \nabla^2 T + Q \tag{1.8}$$

where α is the coefficient of thermal expansion defined by the following relation $\alpha = -\frac{1}{\rho} \left(\frac{\partial \rho}{\partial T} \right)_p$, k is the thermal conductivity, $c_p = T \left(\frac{\partial S}{\partial T} \right)_p$ is the specific heat at constant pressure, Q is the rate of heat addition per unit mass by internal sources. Moreover since $\rho = \rho(T, \rho, p)$ we need an additional constraint on salinity taking in account of the related source/sink terms and diffusivity redistribution of salinity.

$$\frac{dS}{dt} = F(S) \tag{1.9}$$

¹see Pedlosky (1979)

In order to include some effects of the small-scale processes, which are not the focus of our interest but they may influence large scale-flow motions, we can decompose the state variables (scalars or vectorial) in large-scale (and/or long-period) and smaller-scale (and/or shorter-period) components.

$$B(t) = \bar{B}(t) + \hat{B}(t)$$

By convention the over bar ($\bar{\cdot}$) represents the chosen averaging operator and the caret ($\hat{\cdot}$) denotes the deviation from that average, and B is a generic variable. Under the basic assumption that:

$$\begin{aligned}\bar{\hat{B}} &= 0 \\ \overline{\bar{B}} &= \bar{B} \\ \overline{\bar{B}_1 \hat{B}_2} &= 0\end{aligned}$$

and

$$\overline{\bar{B}_1 \bar{B}_2} = \bar{B}_1 \bar{B}_2$$

Here B_1 and B_2 are other different generic variable. So starting from the inviscid momentum equation written in tensorial form (1.10):

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} - f \epsilon_{ijk} u_j = \frac{\partial p}{\partial x_i} - \frac{g}{\rho_0} \rho \delta_{i3} \quad (1.10)$$

We can rewrite each state variable as the sum of mean field and a perturbation:

$$u_i = \bar{u}_i + \hat{u}_i \quad \rho = \bar{\rho} + \hat{\rho} \quad p = \bar{p} + \hat{p}$$

This kind of averaging procedure dates back to the last century and it is referred as *Reynolds decomposition* (Reynolds (1895)).

Substituting in each term of 1.10 its mean field and perturbations, and then averaging we obtain:

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} - f \epsilon_{ijk} \bar{u}_j = -\frac{\partial \bar{p}}{\partial x_i} - \frac{g}{\rho_0} \bar{\rho} \delta_{i3} - \frac{\partial \overline{\hat{u}_i \hat{u}_j}}{\partial x_j} \quad (1.11)$$

1.2 From the closure problem to the sub-grid scale parametrizations 7

The last term is called *Reynolds stress tensor*, its diagonal elements: $\overline{\hat{u}_i \hat{u}_i}$ stand for the mean turbulent momentum fluxes, while the symmetric off-diagonal ($i \neq j$) elements $\overline{\hat{u}_i \hat{u}_j}$ are the shearing stresses. These new variables, for which there are no prognostic equations, arise in the Reynolds averaging procedure, so we cannot resolve the equations unless we specify the stress ($\overline{\hat{u}_i \hat{u}_j}$) tensor directly or writing some kind of relationship of the fluctuating components to the mean field, this is the so called *closure problem*. Experimental evidences support the hypothesis that small scale processes act on large scale flow in a manner that mimics the way in which molecular motions affect macroscopic flow, and so Reynolds stresses are parametrized directly assuming a linear dependence (see eq. 1.12) with large scale flow gradient:

$$\frac{\partial(\overline{\hat{u}_i \hat{u}_j})}{\partial x_i} = \frac{\partial}{\partial x_i} (A^{x_i x_j} \frac{\partial \overline{u}_i}{\partial x_j}) \quad (1.12)$$

In a general approach the tensor A in a 3x3 matrix if space and time varying coefficients,

$$A = \begin{bmatrix} A^{xx} & A^{xy} & A^{xz} \\ A^{yx} & A^{yy} & A^{yz} \\ A^{zx} & A^{zy} & A^{zz} \end{bmatrix} \quad (1.13)$$

In wide use is the much simpler diagonal form:

$$A = \begin{bmatrix} A^h & 0 & 0 \\ 0 & A^h & 0 \\ 0 & 0 & A^v \end{bmatrix} \quad (1.14)$$

This approximation follows from the small aspect ratio ($\delta = \frac{H}{L}$) in the ocean, which suggests a separate treatment of lateral and vertical sub grid scale parametrizations. If the tensor A is assumed to be a constant, the new equation form for large scale flow is identical to ordinary Navier-Stokes equations with an effective viscosity ($\frac{\mu}{\rho} + A$).

1.2.1 Lateral parametrization

Lateral mixing schemes in ocean models usually use a first order closure method. The lateral mixing includes all quasi-diffusive and viscous processes that occur along geopotential surfaces, along surfaces of constant potential or *in situ* density and along the bottom boundary. In an inhomogeneous flow field, turbulent viscosity may clearly depends on the local position, so it plausible that the rate of SGS mixing should vary with location and resolution. A realistic approach relates the coefficients to the ambient conditions in the large scale flow field, so an extension of the constant coefficient concept is used, involving spatially varying mixing coefficients $A(x_i, x_j)$. Usually there are two classes of these realistic schemes: *upstream* (Lin & Walker, 1994) and the Smagorinsky (Smagorinsky, 1993) In the first, the eddy coefficient is proportional to the local flow speed and the grid spacing, while the latter combines a grid size dependence with the deformation of velocity field. Moreover in high resolution simulation, since part of the spectrum of mesoscale eddies is explicitly incorporated, the harmonic approach seems to be too dissipative on the eddy scales, especially where the cut-off wave number on the numerical grid is close to the Rossby deformation radius. So it is used higher order diffusive or viscous operator² e.g.

$$\frac{\partial(\overline{\hat{u}_i \hat{u}_j})}{\partial x_i} = (-1)^{n+1} \frac{\partial}{\partial x_i} (A_{bi}^{x_i x_j} \frac{\partial \nabla^{2n} \overline{u_i}}{\partial x_j}) \quad (n = 0, 1, \dots) \quad (1.15)$$

For $n=0$ we have the basic harmonic operator. If $n=1$ we have the bi-harmonic viscosity/diffusivity term, which can offer a compromise between increased scale selectivity and computational requirement. Since small-scale noise often accumulates at the highest wave-number, and since these flow components are unlikely to be accurately computed in any case, scale selective filtering techniques (like Shapiro filter, see Shapiro, 1970) have been developed. However we have to consider that the principle direction of mixing is therefore neither strictly vertical nor purely horizontal, but a spatial mixture of the two. A rotation of the mixing tensor from the standard horizontal/vertical orientation can be performed (Solomon, 1971 and Redi, 1982) and available in ocean models. Often viscous and diffusive operators have

²see Haidvogel & Beckmann (1999)

the same forms, but different dissipative coefficients. Nevertheless monotonicity and positive definiteness are usually considered important proprieties of the numerical solution for the tracer equation, while they are less stringent for the momentum equations. The choice of diffusive closure is therefore closely related with the form of the horizontal advection operator. *A priori* choice is not always obvious. Some closures are rather *ad hoc*, and the only available justification for their use is the preservation of *smooth* numerical results. Furthermore, the SGS schemes can be used as filter, in order to remove the small scale numerical noise.

1.2.2 Parametrization in the vertical direction

As reviewed in detail by Large (1998) parametrization can be classified into *local* and *non-local* closure schemes. The formers assume that the eddy fluxes depend on the local properties of the large scale flow, while the latter recognize that the turbulent transports at a given level may not depend exclusively on the local properties at that level, but rather on the overall state of the boundary layer. Both these closure schemes are derived by time evolution on the Turbulent Kinetic Energy (TKE), as:

$$\frac{\partial TKE}{\partial t} = A_v^M \left[\left(\frac{\partial \bar{u}}{\partial z} \right)^2 + \left(\frac{\partial \bar{v}}{\partial z} \right)^2 \right] - A_v^T N^2 + \frac{\partial}{\partial z} \left(A_k \frac{\partial TKE}{\partial z} \right) - \epsilon \quad (1.16)$$

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \bar{\rho}}{\partial z} \quad (1.17)$$

N^2 is the *Brunt – Väisälä* frequency and A_k is the vertical diffusivity of the TKE. This two terms are related to the production of the vertical shear and buoyancy. This equation is the basis for a large number of higher-order turbulent closure schemes. Two frequent approach are the so called $k - l$ and $k - \epsilon$ schemes. The first uses a length scale l to close the system, the second gives a definition for l , as $l = c_\epsilon \frac{\sqrt{TKE^3}}{\epsilon}$, and use an additional equation for the rate ϵ . In the absence of high frequency forcing, night-time convection, and with low vertical resolutions, turbulent closures cannot produce high enough mixing at the top layer interface. This explains why non-local parametrisation have been developed, like the KPP scheme (Large *et al.*

(1994) see appendix A.1.1 for a brief description) where the vertical gradients are explicitly retained rather than the assumption of having a well mixed boundary layer.

A different approach, called Pacanowski and Philander (*P.P.*) (R C Pacanowski, 1981), reckons the vertical diffusion coefficients as a function of the local Richardson number neglecting the time-evolution of the Turbulent Kinetic energy (see description in appendix A.1.3). In this scheme, turbulent mixing is treated by a first-order local diffusion approach in which the sub-grid scale turbulent vertical kinematic flux of a quantity x (X as the mean) is assumed proportional to the local property gradient with an appropriate eddy mixing coefficient K . This so-called K theory can be described as

$\overline{w\bar{x}} = -K \frac{\partial X}{\partial z}$ where the upper case (X) represents mean quantities resolved at the model grid and the lower case (x) represents the subgrid-scale variables (also called turbulent fluctuations).

The mixing coefficients can be approximated as

- $K_m = \frac{\nu_0}{(1+\alpha Ri_g)^n} + \nu_b$
- $K_t = \frac{\nu_0}{(1+\alpha Ri_g)^{n+1}} + \kappa_b$

where K_m represents viscosity, K_t represents diffusivity, and the local gradient Richardson number :

$$Ri_g = \frac{N^2}{U_z^2 + V_z^2} \quad (1.18)$$

1.3 The Primitive Equation

The task of solving the resulting PDEs in the most efficient manner possible requires careful attention to the nature of the flow, the available computer resources, and the simplifications that can be made without adversely affecting the solutions sought. To conduct analytical studies Navier-stokes equation, they are subjected to some approximations, furthermore other kind of simplification are due to computational constraint. That simplifications are derived by the *scaling analysis*, task of this process is to eliminate certain terms that will be unlikely to be important.

- **The Boussinesq approximation:** density variations may be neglected except when they are coupled to the gravitational acceleration in the buoyancy force, the variations in volume expansion due to temperature gradients will also be small. Density field can be expressed as a sum of a constant reference value ρ_0 and a smaller, space and time varying perturbation, $\hat{\rho}$:

$$\rho(x, y, z, t) = \rho_0 + \hat{\rho}(x, y, z, t) \quad \rho_0 \gg \hat{\rho} \quad (1.19)$$

As consequence in the momentum equation and the mass conservation, we may substitute the inertial term and the continuity equation with a constant.

- **Incompressibility:** The specification of incompressibility and constant density immediately decouples the dynamic from the thermodynamics and reduce the equation of mass conservation to the condition incompressibility or non divergence field. the three dimensional divergence of the velocity vector is assumed to be zero ; Let's take the conservation equation for mass eq.(1.5 together with the approximation 1.19 it became:

$$\rho_0 (\nabla \cdot \vec{v}) + \hat{\rho} (\nabla \cdot \vec{v}) + \frac{d\hat{\rho}}{dt} = 0 \quad (1.20)$$

The second term is smaller compared to the former, so it can be ignored, furthermore if the characteristic length and time scales of perturbation density are comparable to those of the velocity components, than the three term is also smaller then the first by a factor. So a suitable approximation can be:

$$\nabla \cdot \vec{v} = 0 \quad (1.21)$$

for a quantitative and more accurate demonstration refers to Batchelor (1967).is in the momentum equation where the kinematic viscosity can be assumed constant

- **The Hydrostatic approximation:** considering the oceanic circulation and taking the appropriate typical magnitudes for the variables involved in Navier-

Stokes equation, the vertical momentum equation is reduced to a balance between the vertical pressure gradient and the gravitational force.

$$-\frac{1}{\rho} \frac{\partial p}{\partial z} - g \simeq 0 \quad (1.22)$$

A complete justification of this approximation (completed by demonstration that perturbations to the mean hydrostatic state are themselves hydrostatic, and that scales of motion typical of mesoscale circulation are also hydrostatic, see Holton (1992)).

- **Spherical earth approximation:** The geopotential surfaces are assumed to be spheres so gravity field lines are parallel to the earth's radius, and the equations are written in spherical coordinates (λ, ϕ, z) where λ is longitude, ϕ is latitude and z is depth.

The vector invariant form of the primitive equations as solved by the OceanGeneral Circulation Model employed in our institute became:

$$\begin{aligned} \frac{\partial \mathbf{U}_h}{\partial t} &= - \left[(\nabla \times \mathbf{U}) \times \mathbf{U} + \frac{1}{2} \nabla (\mathbf{U}^2) \right]_h \\ &\quad - f \mathbf{k} \times \mathbf{U}_h - \frac{1}{\rho_0} \nabla_h p + \mathbf{D}^u \\ &\quad \frac{\partial p}{\partial z} = -\rho g \\ &\quad \nabla \cdot \mathbf{U} = 0 \\ &\quad \frac{\partial T}{\partial t} = -\nabla \cdot (T\mathbf{U}) + D^T \\ &\quad \frac{\partial S}{\partial t} = -\nabla \cdot (S\mathbf{U}) + D^S \\ &\quad \rho = \rho(T, S, p) \end{aligned}$$

Here we defined $\mathbf{U} = \mathbf{U}_h + w\mathbf{k}$ (the subscript h denotes the local horizontal vector i.e. over the (i,j) plan), T is the potential temperature, S the salinity ρ is the *in-situ* density. Furthermore ∇ is the generalised derivative vector operator in the $(\mathbf{i}, \mathbf{j}, \mathbf{k})$

directions, t is the time, z is the vertical coordinate f the coriolis acceleration ($f = 2\omega \mathbf{k}$ where ω is the earth angular velocity vector) g is rge gravitational acceleration, \mathbf{D}^u , D^T and D^S are the parametrizations of small scale physics for momentum, temperature and salinity, including surface forcing terms.

1.4 Errors of numerical approximations

There are several potential sources of errors in a numerical calculation. Two sources are universal in the sense that they occur in any numerical computation. They are round-off and truncation errors. Inaccuracies of numerical computations due to the errors result in a deviation of a numerical solution from the exact solution, no matter whether the latter is known explicitly or not.

1.4.1 Round-off errors

Numbers are represented in a computer by a finite number of digits of precision. The simplest variant for hardware implementation is to keep the first n digits and to chop off all remaining digits. A more accurate scheme is to examine the $(n + 1) - st$ digit and to round the $n - th$ digit to the nearest integer. This procedure leads to round-off errors. We studied this kind of error with a simple example: the Leibniz formula for π .

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots = \frac{\pi}{4} \quad (1.23)$$

We truncated the sum for $n = 100$, in the first case we did the integration 100 times on the same processor, in the other cases we did the integration $100/N_{procs}$ for processor, sending the information from one processor to the other and printing the results in double precision.

- RUN on 1 processor: π is approximately: 3.1416009869231254
- RUN on 2 processors: π is approximately: 3.1416009869231241
- RUN on 10 processors: π is approximately: 3.1416009869231249

- RUN on 20 processors: π is approximately: 3.1416009869231245
- RUN on 40 processors: π is approximately: 3.1416009869231258
- RUN on 50 processors: π is approximately: 3.1416009869231254

1.4.2 Truncation errors

Several kind of truncation errors occur in representing a continuous function in term of a discrete set of variables. A first possibility is to obtain a power series for the function $u(x)$ by truncating the Taylor series expansion about some point x_0 under the hypothesis that continuous function $u(x)$ possesses an $(n + 1)$ derivative everywhere on the interval $[x_0, x]$. Taylor series expansion is used in order to evaluate the derivative of a function.

Another example of truncation error is the numerical integration of a function where the exact calculations would require us to calculate the area under the curve by the infinite summation of the subtended rectangles. Since we cannot choose an infinite number of rectangles, we will have truncation error.

Error can come by piecewise interpolation due to the representation of a function defined on a grid in a grid with different grid spacing. it has been shown that maximum error goes like grid spacing squared, Haidvogel & Beckmann (1999). Other times, it is necessary to replaces the series by the $n - th$ order polynomial, this lead the grows of a truncation error of the $n - th$ order.

1.5 Error Propagation

Since different sources of error can affect a simulation we trained a simple twin experiment to test error propagation due to inaccuracy in atmospheric forcing uncertainties and model approximations. European Centre for Medium-Range Weather Forecasts (ECMWF) provides to MFS Wind velocity(10 m), cloud cover, humidity, temperature and pressure(at 2 m) with a time interval of 6 hours once per day. For our purposes, since all those fields come from a global model $1/4^\circ$ (from 2009, before $1/2^\circ$)

they have to be interpolated by an appropriate algorithm (bi-linear or bi-cubic according to the field) to the operational set up regular grid $1/16^\circ$ resolution. ($1/24^\circ$ in the high resolution model (see appendix B)). As long as the NEMO model code ran on a vector machine, a bicubic spline interpolation had been used for the interpolation procedure. During our work thesis, we ported on a scalar machine the old code in order to run with domain decomposition technique. The old spline method was no more a suitable option, hence a simpler bicubic interpolation has been implemented. We studied how a new interpolation method can impact the forecast simulation by setting up the following twin experiment. Two simulations started from the same initial conditions and were forced for 10 days with the same ECMWF files, but have been interpolated with the two different interpolation algorithms. Uncertainties in atmospheric forcing and computational error were propagated by the model. The results are depicted in fig.1.1. Even though the interpolation method ensures the same

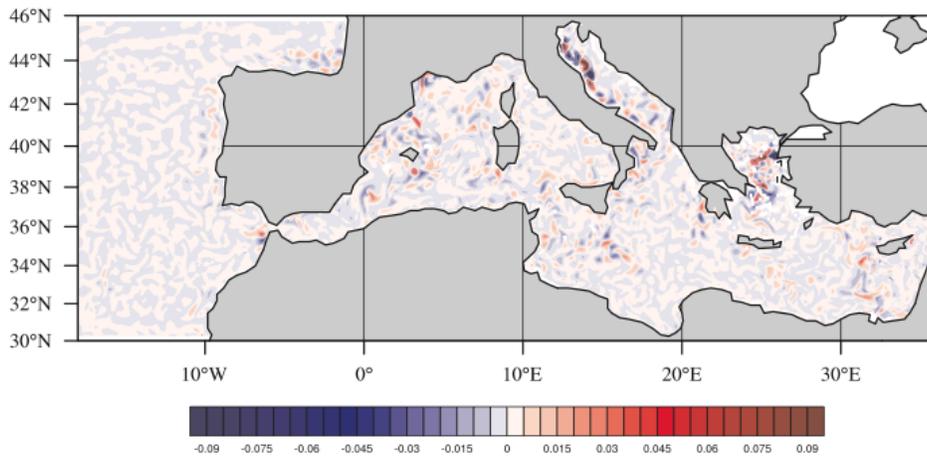


Figure 1.1: Differences $^\circ\text{C}$ the twin experiment simulations

accuracy of evaluation, and the simulations have been carried with same computational design (same machine and domain splitting), we were able to reach significant differences between the simulation only by the choice of an interpolation procedure.

Chapter 2

Building the SuperEnsemble dataset

Overview

For weather and ocean prediction systems, the uncertainties and model errors can be investigated by applying the ensemble techniques designed to sample the error under investigation. If we are interested in model initialization uncertainties the ensemble members will be obtained by the repeated integrations of the model forward in time from slight-perturbed initial conditions, with perturbations designed in order to capture as much as possible of the underlying uncertainty(Kalnay (2003)). Similarly, model error due to physical parametrizations can be addressed by the "*multi-physic*" approach, *i.e.* running the same simulation several times with different settings of the physical parametrization schemes(Pellerin *et al.* (2003)) or with different parametrization schemes. Moreover, in order to take into account both sources of error (initial conditions, model error), both the techniques can be combined to a new approach, the multi-model superensemble concept, in this case the ensemble is built combining several model integrations able to sample a wider range of uncertainties, from different initial conditions, atmospheric forcing inaccuracy and the different physical parametrizations.

As far as uncertainties are concerned and interrelated, no one of the previous approach is clearly better than another, but in this work we focused in the latter approach since it seems to be the bigger challenge of ocean and weather forecasting.

In this chapter, the basic idea beside the set up of each experiment carried out, the tasks addressed and considerations on the goodness of ensemble generated are presented. Since simulations results are indirect measurements of the quantity studied, we can use the same terminology used in experimental physical measurements. Hence, the *accuracy* of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value, while the *precision* is the degree to which repeated measurements under unchanged conditions show the same results. A measurement system can be accurate but not precise, precise but not accurate, neither, or both. For example, if the experiment contains a systematic error, then increasing the sample size generally increases precision but does not improve accuracy. Eliminating the systematic error improves accuracy but does not change precision. For our purposes, the spread around the ensemble mean mimics the accuracy while the difference in generated variability from ensemble and the observed state can mimic the precision. *metti qualcosa che lo spieghi* As evidenced by Kalnay *et al.* (2006) we have built a "good ensemble" if "truth" looks like a member of the ensemble (see figure 2.1)

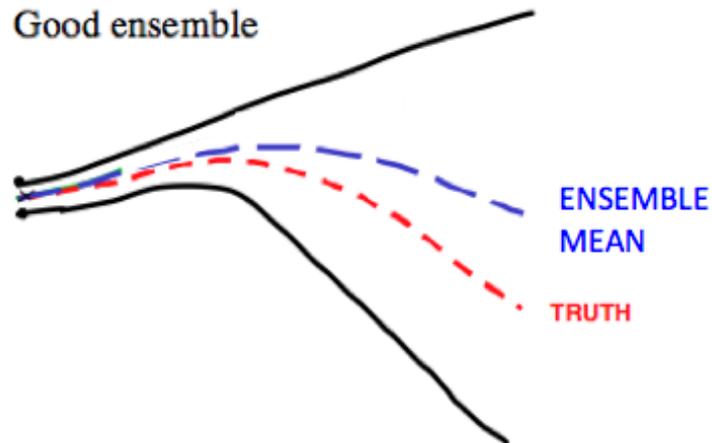


Figure 2.1: Practical Example corresponding to the idealized "good ensemble"

So two main conditions must be satisfied:

- **Spread** around the ensemble mean. As evidenced by Hagedorn *et al.* (2005) before, Weigel *et al.* (2008) and Knutti (2010) later, the feedback spread-bias reduction is challenging since it is impossible to remove a correct bias if the standard deviation around ensemble mean, is less than the mean difference between simulations and truth state. It means we gain in accuracy despite to the precision of the measurement. most correlated patterns are also those that reduce the ensemble spread Feddersen *et al.* (1999)
- **Generated Variability.** This feature can be considered as the average ensemble capability in reproducing the truth observed state variability during a time period. It means that even if we were able to remove the bias, in order to be reliable we should reproduce major phenomena occurring during the period of interest, Johnson & Bowler (2009).

The variability for the ensemble member has been evaluated as:

$$V_{MP} = \frac{1}{M} \sum_i^M \frac{\sum_{t=1}^T (F_{i,t} - \bar{F}_i)^2}{T} \quad \bar{F}_i = E[F_t]_i = \frac{\sum_{t=1}^T F_{i,t}}{T} \quad (2.1)$$

While for the truth state it has been evaluated as:

$$V_O = \frac{\sum_{t=1}^T (O_t - \bar{O})^2}{T} \quad \bar{O} = E[O_t] = \frac{\sum_{t=1}^T O_t}{T} \quad (2.2)$$

The variability generated by the multi-physics ensemble was smaller than the natural variability of the system as depicted in figure 2.2, also the spread defined as:

$$STD_E = \frac{1}{T} \sum_t^T \sqrt{\frac{\sum_{i=1}^M (F_{i,t} - \bar{M}_t)^2}{M}} \quad \bar{M}_t = E[F_i]_t = \frac{\sum_{i=1}^M F_{i,t}}{M} \quad (2.3)$$

was smaller than the mean absolute difference (MAD) $MAD = \frac{1}{T} \sum_t^T | \bar{M}_t - O_t |$

2.1 Multiphysics SuperEnsemble Experiment

Despite there are several kinds of uncertainties and model error affecting a simulation, here we focused our attention only in some specific topic fields. The main idea leading to this experiment derived from the analysis of the Primitive Equations.

$$\frac{\partial \vec{u}}{\partial t} + \vec{u} \cdot \nabla \vec{u} + f \times \vec{u} + \frac{1}{\rho} \nabla p = k_H \nabla^2 \vec{u} + k_z \frac{\partial^2 \vec{u}}{\partial z^2} \quad (2.4)$$

The first term on LHS is the local tendency, the quantity we want to resolve, the second is the advection term, the third the coriolis force and the last term on LHS is the pressure term. All this terms can be resolved explicitly and errors related to the LHS are mostly due to discretization techniques or accuracy of numerical schemes. In the RHS the viscosity term has been splitted into the horizontal (first term on RHS) and vertical components. For these two terms (the closure problem discussed in the previous chapter) some physical assumption and/or parametrization must be used, that inevitably will impact the model results.

Other structural uncertainties arise if we consider the advective part in the advection/diffusion equation for the tracer.

$$\frac{\partial T}{\partial t} + (\bar{u} \cdot \nabla_h) T + w \frac{\partial T}{\partial z} = A_h \nabla_h^2 T + A_z \frac{\partial^2 T}{\partial z^2} \quad (2.5)$$

Since the equation must be solved on computers, a space and time discretization is needed. The choice made in space and time interpolation defines the value of the tracer at the velocity points that can change according to the grid and time stepping used.

Concerning the space discretization, the vertical grid size is usually an increasing function of depth, this leads to the sparsest resolution at the ocean floor. In this way the bottom boundary layer dynamics can not be explicitly represented and a parametrization for the bottom momentum flux is necessary. The previous sources of uncertainties can be sampled building the multi-physics ensemble as a collection of simulations generated using the related physical parametrization schemes available in the Ocean General Circulation Model NEMO (Nucleus for European Modelling

of the Ocean, Madec (2008)), implemented in our institute.

2.1.1 The basic model configuration

The Ocean General Circulation Model (OGCM) code is NEMO-OPA (Nucleus for European Modelling of the Ocean-Ocean PARallelise) version 2.3 (Madec et al 2008). The code is developed and maintained by the NEMO- consortium. The model is primitive equation in spherical coordinates. NEMO-OPA has been implemented in the Mediterranean at $1/16 \times 1/16^\circ$ horizontal resolution (almost 6.5 Km) and 72 unevenly spaced vertical levels and they have a thickness ranging from 3 m at the surface to 300 m at the ocean bottom. Partial steps on the floor, in order to better reproduce the bathymetry, are employed for the vertical grid mesh. This model can be defined as a mesoscale-resolving model for the Mediterranean Sea, since the first internal Rossby radius of deformation is around in summer and for most of the Mediterranean basin (except in the Adriatic Sea). The Digital Bathymetric Data Base Variable Resolution (DBDB-V) has been used to determine the coast line and bathymetry. Furthermore, bathymetry has been manually interpolated along the Croatian coast by a comparison with detailed nautical chart. In order to better describe bottom shape, the vertical column are discretized by partial cells. The model covers the entire Mediterranean Sea and also extend into the Atlantic in order to better resolve the exchanges with the Atlantic Ocean at the Strait of Gibraltar. The model is nested, in the Atlantic, within the monthly mean climatological fields computed from ten years of daily output of the $1/4 \times 1/4^\circ$ global model (Drevillon *et al.*, 2008). In this way the total volume is allowed to change producing improvements particularly evident in the Mediterranean sea level seasonal variability and in the salinity characteristics of the Modified Atlantic Water (details on the nesting technique and other major impacts on the model results are in Oddo *et al.*, 2009). Primitive Equations are intergrated with a time step of 600s.

Momentum, water and heat fluxes are interactively computed by bulk formulae using the 6-h, 0.25° deg. horizontal-resolution operational analyses and forecasts fields from the European Centre for Medium-Range Weather Forecasts (ECMWF). The water balance is computed by Evaporation minus Precipitation and Runoff.

The evaporation is derived from the latent heat flux while the precipitation and the runoff are provided by monthly mean datasets: the Climate Prediction Center Merged Analysis of Precipitation (CMAP) Data (Xie and Arkin, 1997), the Global Runoff Data Centre dataset (Fekete et al., 1999) for the Ebro, Nile and Rhone and the dataset from Raichich (Raichich, 1996) for the Adriatic rivers (Po, Vjose, Seman and Bojana). The Dardanelles inflow is parametrized as a river and the climatological net inflow rates are taken from Kourafalou and Barbopoulos (2003). The horizontal viscosity and diffusion operators are assumed to be bi-laplacian with coefficients of $5 \times 10^9 m^2/s$ and $3 \times 10^4 m^2/s$ for viscosity and diffusion respectively. The vertical diffusion and viscosity terms are dependent upon the Richardson number. The vertical convective processes are parameterized using the enhanced vertical diffusivity parameterization. The advection scheme for active tracers, temperature and salinity, is a mixed up- stream/MUSCL (Monotonic Upwind Scheme for Conservation Laws, Estubier and Lvy, 2000) scheme. This mixed scheme has the capability to switch to a simple up-stream scheme in areas where numerical instabilities might occur such as in proximity of the river mouths, close to the Atlantic boundaries or near the Strait of Gibraltar. Here the large mixing acting, due to the internal wave and tide breaking, is not explicitly resolved. In this area also the vertically diffusivity is artificially increased.

2.1.2 Creation of the ensemble members

All the simulations started from the same initial conditions (climatology), the 7th of January 2004, at the 00 UTC and lasted 30 days. Specifically, (see table 2.1) the multi-physics dataset is the combination of two different viscosity operators (harmonic and biharmonic) for the momentum, two vertical diffusion schemes ($k - \epsilon$ and P.P.), three different advection schemes (*upstream*, Total Variance Dissipation -TVD- and Monotone Upstream Scheme for Conservative Laws (MUSCL) -all those schemes are described in appendix A- and three kinds of bottom friction parametrizations.

In each multi-model database, the distinction between "best effort" simulations (*i.e.* the results from the default version, in this case the operational model in simula-

n°	Vertical scheme		Tracer advection			Bottom friction	Lateral momentum viscosity	
	$k - \varepsilon$	P.P.	UBS	TVD	MUSCL		Laplacian	Bilaplacian
1		×			×	non linear slip		×
2		×	×			non linear slip		×
3		×		×		non linear slip		×
4		×			×	free slip		×
5		×			×	no slip		×
6		×			×	non linear slip	×	
7	×				×	non linear slip		×
8	×			×		non linear slip		×
9	×		×			non linear slip		×
10	×				×	no slip		×

Table 2.1: *Experiment 1 design: Model configurations used in the multi-physics ensemble. Member number 1 is the model basic configuration (Default version). All the other member are obtained changing only one numerical scheme per time from the Default version.*

tion mode used at INGV) and the perturbed physics ensembles is important because can help us in modelling the uncertainties. As Murphy *et al.* (2004) pointed out, the main limit of this kind of procedure is the un-reduction of the *systematic error*, since the ensemble will have the structure of the underlying default model, so if it is not able to reproduce a special process, then no one of the perturbed ensemble member will be able to do it. For instance, if the ensemble is a collection of hydrostatic models, the vertical flow is determined by the incompressibility condition, and has no direct energy source. So total kinetic energy in the hydrostatic case remains bounded, as the horizontal velocity is limited by the available potential energy and the vertical velocity is limited through the incompressibility condition(eq.1.21). Since all convective processes aren't explicitly resolved, neither they can be eliminated, they will be parametrized, in different way, in each member.

A comparison of the resulting simulation as been analysed in terms of spread are depicted in figure 2.3.between the model and the observations(figure 2.3).

Despite the plain results this experiment was helpful since we were able to evaluate

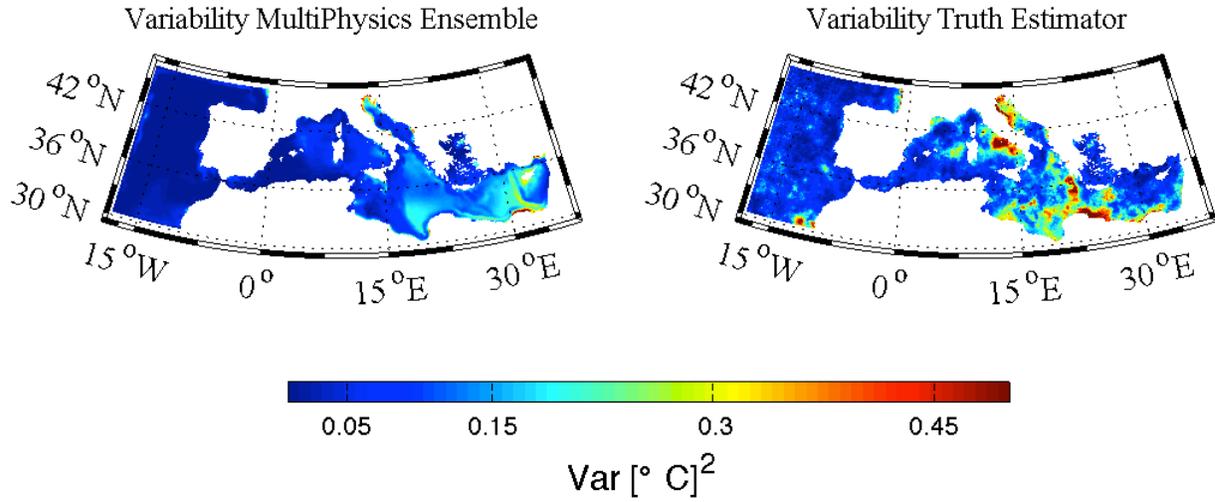


Figure 2.2: Mean Variability generated in one month by the Multi-Physics ensemble (left) versus the natural variability in the observations.

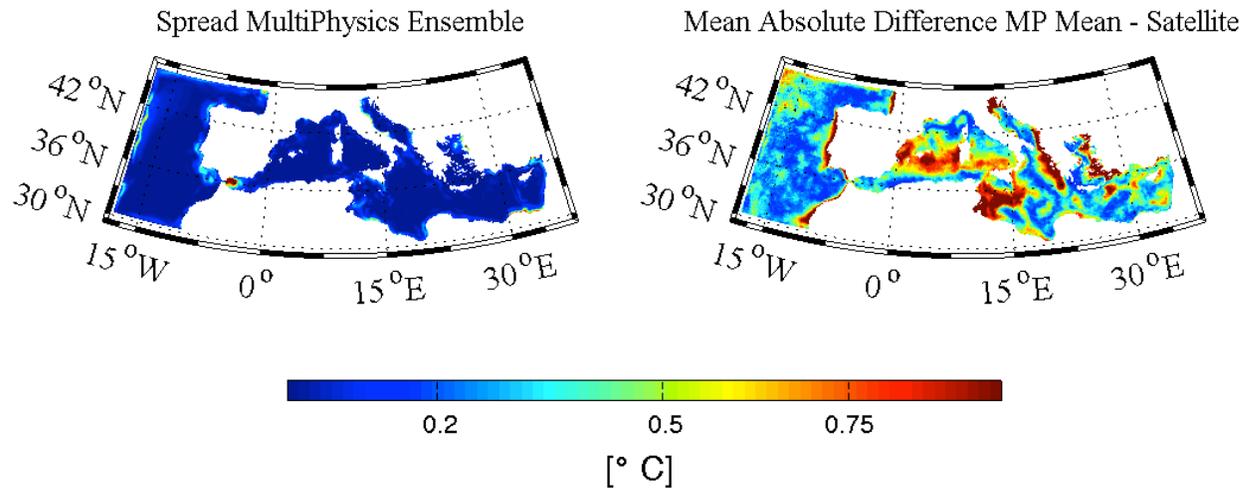


Figure 2.3: Standard Deviation of Multi-physics members (left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).

the capability of the model to generate spatial variability, the rate of formation of the spread and we could studied the impact of each numerical scheme choice.

2.2 BHM SuperEnsemble Experiment

In the second approach we produced 22 ensemble members ensemble that could help us in investigating the atmospheric forcing uncertainties. Usually wind analysis present systematic and persistent errors in representing the external forcing over sea surface and can constitute an ideal starting point to address the problem of ocean forecasting. This is especially true for high resolution model. Quick scat wind constitutes a large data-set to compare the ECMWF products and derive a realistic representation of the errors. This system is not optimal in the sense the reduces the ensemble size that is necessary to consider, but it is a sound representation of a well known source of uncertainty for ocean forecasting [Pinaridi *et al.* (2011)]. Hence, in this case, we suppose a perfect ocean model (in parametrizations choice sense) and model error is due to the wind forcing and initial conditions uncertainties. We decided to address a more classical approach, in order to reproduce Krishnamruti's results for mesoscale resolving ocean simulations. The table 2.2 shows the major differences between the ensemble members. **Target Field:** Horizontal map of temperature from Operational model output with data assimilation (SYS3a2);

In this case we had reached satisfactory results in term of variability, but the map of spread shows low pattern correlations with MAD.

2.3 Multi-Model Multi-Physics SuperEnsemble Experiment

On the base of previous experiments results a third ensemble has been created. The MultiModel Multi-Physics Ensemble dataset uses the simulations with the physical parametrizations set having the largest impact on the ensemble spread of on the first experiment(Multi-Physics Ensemble), in addition to varies initial conditions and estimates of ocean state evolution obtained using by different model results (both

	SYS3a2	OPA 8.2	OPA 9.0
Atlantic Boundaries	Close	Close	Open
Topography	Z-levels	Z-levels	Z-levels & Partial Step
Water Flux	WF=E-P-R	relaxation to climatology	WF=E-P-R
Assimilation Scheme	3Dvar	no assimilation	no assimilation
Atmospheric Forcing	ECMWF	ECMWF + 10 BHM	ECMWF + 10 BHM

Table 2.2: Set up of BHM SuperEnsemble Experiment, the columns are the members, while the rows are the difference between each member.

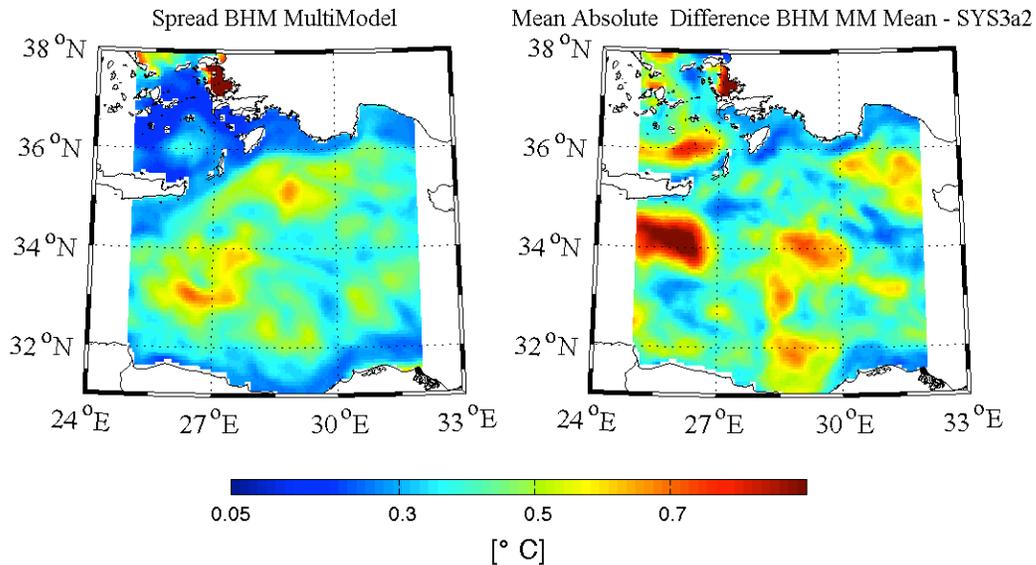


Figure 2.4: Standard Deviation of BHM ensemble members (left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).

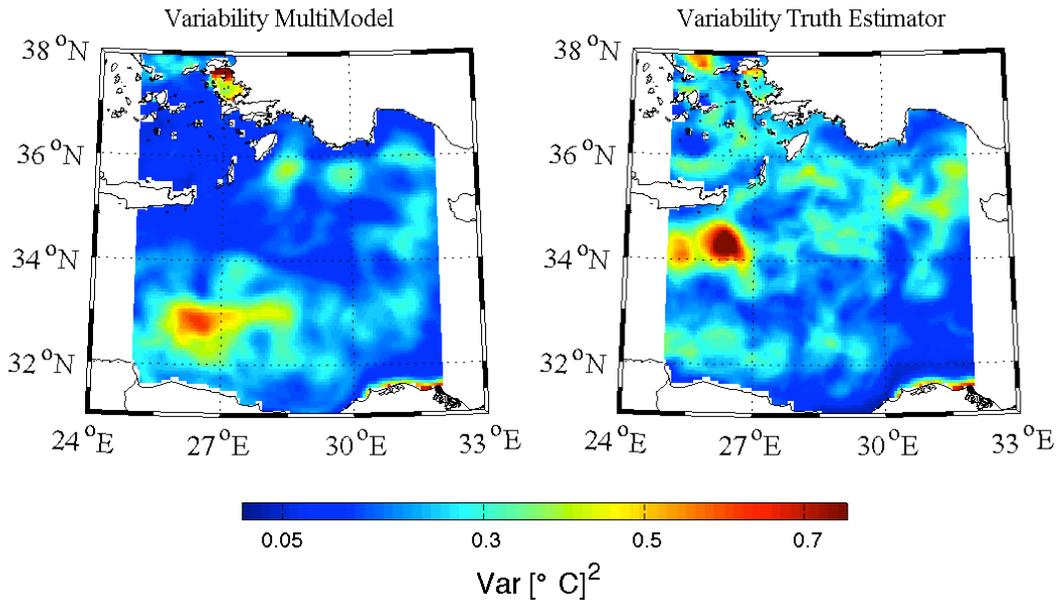


Figure 2.5: Mean Variability generated by the BHM ensemble (left) versus the natural variability in the observations (right) during the training period

analysis and forecasts). This study is based on the extensive data set of analysis produced in the MyOcean project, purposed to evaluate the performances of "standard" Multi-Model Super-Ensemble (MMSE) approaches to get the Mediterranean Sea Surface Temperature best estimate, by combining the operational results from the single members. In the first phase of this research activity, the MMSE have been applied to analysis results to obtain a first guess of the method advantages and limits (studying the sensitivity on training period lengths and number of members in the ensemble). In the second phase, a forecast exercises will be performed evaluating the limit of the applied method in terms of forecast range. The members have been regressed on a truth state produced by SST analysed from satellite available in MyOcean Catalogue. This approach started with a synergic effort from Istituto Nazionale di Geofisica e Vulcanologia (INGV), Hellenic Centre for Marine Research (HCMR) and MERCATOR Ocean; in a second phase of this research United Kingdom Met Office (UKMO) has been contacted. A description of the data set can be found on the following sections, while an explanation of the diagnostic tool and nomenclature is presented in section 3.1.

2.3.1 Analysis mode

An extensive multi-model analysis dataset, produced in the framework of MyOcean project has been used for a complete and comprehensive assessment of multi-model capability in reproducing SST by satellite and to study the benefit of having different operational oceanographic systems in the Mediterranean Sea We explored the impact of model error, initial conditions and atmospheric forcings in analysis configuration. All the members are numerical models solving the the primitive equations with assimilation of real observations. The members used to create the ensemble are:

- INGV¹: SYS3a2 & SYS4a3 Daily Output; Both the system use the same data assimilation technique, 3VAR (Dobricic & Pinardi, 2008). The major differences between this kind of products concern the numerical code used, OPA 8.2 in the first, while NEMO 9.0 in used in the latter. Furthermore the advection scheme for active tracer has been upgraded from the 2nd order centered used in SYS3a2 to a mixed up-stream/MUSCL in SYS4a3. Concerning the parametrization in the Atlantic part of the model, SYS4a3 is nested (Oddo *et al.* (2009)) within the monthly mean climatological fields computed from the daily output of the MERCATOR $1/4^\circ$ resolution global model Drevillon *et al.* (2008).
- HCMR ²: This model is based on the Princeton Ocean model (POM ³), which is a primitive equations free surface ocean model, operating under the hydrostatic and Boussinesq approximations. The model equations are written in sigma-coordinates(Mellor & Blumberg (1985)) and discretized using the centered second-order finite differences approximation. The model domain has a horizontal resolution of $1/10^\circ$ and 24 sigma layers along the vertical with a logarithmic distribution near the surface and the bottom. This model is forced with hourly surface fluxes of momentum, heat and water provided by the weather prediction system Poseidon - ETA high resolution ($1/20^\circ$). The assimilation system is based on local SEEK Singular Evolutive Extended Kalman (SEEK) filter filter(Pham *et al.* (1998)). SEEK is an error subspace extended Kalman

¹<http://gnoo.bo.ingv.it>

²<http://www.hcmr.gr>

³<http://www.aos.princeton.edu/WWWPUBLIC/htdocs.pom/>

filter that operates with low-rank error covariance matrices as a way to reduce the computational burden. The filter uses covariance localization and partial evolution of the correction directions. This assimilation scheme corrects the forecast state of the model on a weekly basis (Korres *et al.*, 2009); . The assimilated observational data set is multivariate including AVISO sea level height, AVHRR sea surface temperature, MEDARGO floats T and S profiles and XBT data. The Mediterranean hydrodynamical model is forced with hourly momentum, heat and freshwater fluxes derived from the POSEIDON weather prediction system based on a $1/20^{\circ}$ resolution, ETA regional non-hydrostatic atmospheric model.

- MERCATOR ⁴ : Mercator Ocean operates with resolution ($1/12^{\circ}$) in the Mediterranean Sea. The system uses the NEMO OGCM. Its assimilation system is based on the SEEK filter base (computed once from an ensemble). Atmospheric forcing is applied by computing interactive fluxes (bulk formulae) of heat, momentum, and freshwater from European Center for Medium- Range Weather Forecasts (ECMWF) operational atmospheric products, applying a precipitation correction using US National Aeronautics and Space Administration (NASA) Global Precipitation Analysis (GPCP) data with a method derived from Troccoli and Kallberg (2004). River runoffs are computed from Dai and Trenberth (2003) climatology. The products are nested in their global ocean ocean model at $1/4^{\circ}$ resolution. The analysis come from two different code version: NEMO 2.3 and NEMO 3.2 and the data assimilation system have been switched from local SEEK Filter + 3DVAR bias correction for Temperature and Salinity with sequential correction in V0 version to a local SEEK Filter + 3DVAR bias correction for Temperature and Salinity with Incremental Analysis Update (IAU) Tranchant *et al.* (n.d.)) in the V1 version.

In order to asses the seasonal dependence of the technique, the dataset cover one year time period, from the 1st of January to 31 December 2008. The atmospheric forcing

⁴<http://www.mercator-ocean.fr>

<i>MMSE Member</i>	<i>Vertical scheme</i>	<i>Diffusion</i>	<i>Viscosity</i>	<i>Assimilation</i>
SYS3a2	<i>P.P.</i>	Bilap.	Bilap.	OceanVAR +SST nudging
SYS4a3	<i>P.P.</i>	Bilap.	Bilap.	3DVAR +SST nudging
Mercator V0	$k - \epsilon$	Bilap.	Bilap.	SAM
Mercator V1	$k - \epsilon$	Bilap.	Bilap.	SAMv2
HCMR(POM)	$k - l$		Bilap.	SEEK filter
NEMO multiphysics	<i>P.P.</i>	Bilap.	Bilap.	no
	$k - \epsilon$	Bilap.	Bilap.	no
	<i>P.P.</i>	Bilap.	lap...	no
	<i>P.P.</i>	lap.	lap.	no

Table 2.3: Set up of MultiModel SuperEnsemble Experiment obtained by MyOcean database (on the top) and the multi-physics members(on the bottom of the table)

were taken from the ECMWF analysis dataset (excluding HCRM which used its own atmospheric field). The ocean initial condition were different for each model except for the multi-physics member which started from the same conditions given by the default model run used in experiment 1, multi-physics ensemble. Ocean observations have been assimilated in all the analysis member with their own assimilation scheme. The main differences in the dataset are shown in table 2.3.

This experiment shows very good results in terms of variability generated by the Multimodel and also very good correlation between the spread and the MAD, in agreement with literature results, probably this good agreement is due to the nice quality of ensemble members.

2.3.2 Forecast Mode

Different works (Krishnamurti *et al.*, 1999 , Krishnamurti *et al.*, 2000 and Vich *et al.*, 2011) have demonstrated the superior forecast skill of superensemble versus single-

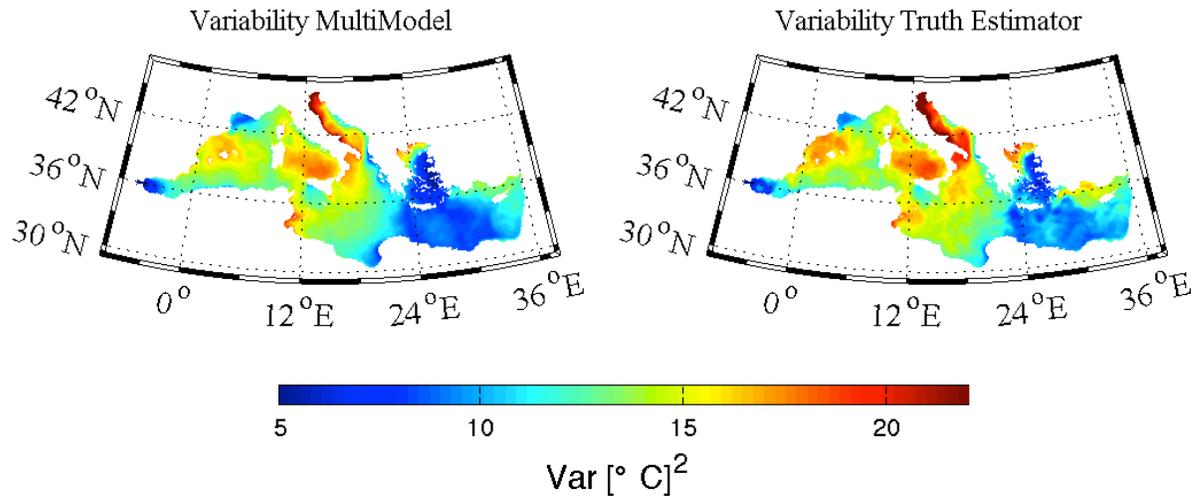


Figure 2.6: Mean Variability generated by the MMSE (left) versus the natural variability in the observations (right) during the time period, year 2008

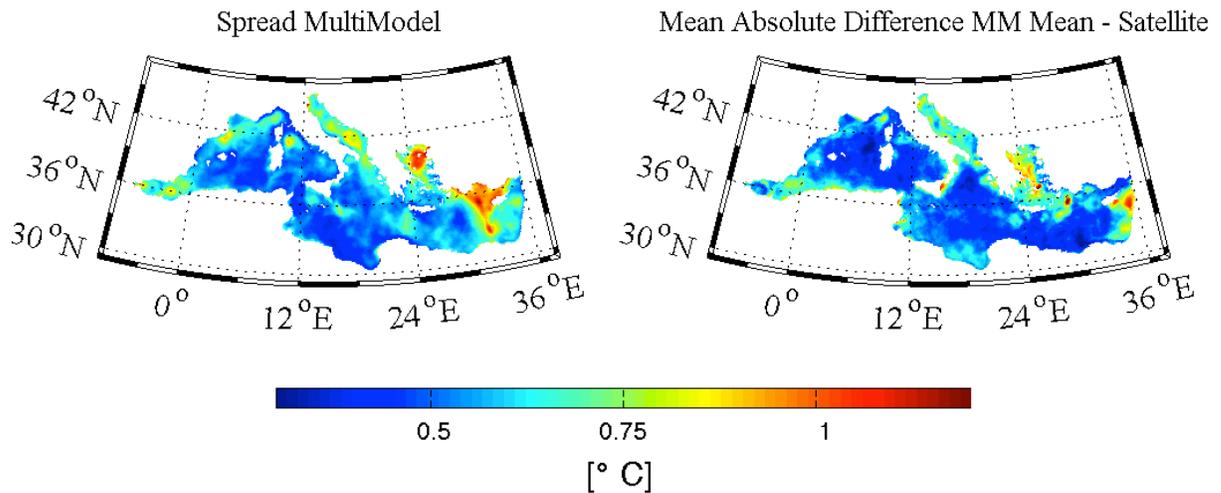


Figure 2.7: Standard Deviation of MMSE members in analysis mode(left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).

model performances, but any substantial work on the success of this technique has been done with forecast members. Motivated by this lack of groundwork, we test the multi-model in a forecast configuration. The dataset comprises three different OGCMs of the following institutions:

- INGV : SYS2b Daily Output lead-time 9 days. The model equations consider an implicit free surface approximation to the primitive equations for oceanic fluids (Roullet et al., 2000). The model salinity and temperature fields along the boundary of the Atlantic box are relaxed at all depth to the climatology. The salt flux is given by a relaxation at the surface toward monthly mean climatological values given by MEDATLAS climatology. A sub-model computes air-sea fluxes of momentum and heat from 6h operational analysis atmospheric forcing in all the Mediterranean Basin. The sub-model computes separately all the terms in the surface heat budget and the choice of bulk formulas follows Castellari et al. (1998, 2000). Different choice has been done for the Atlantic Box where the wind stress are from a monthly mean climatology and the heat flux is given by relaxation at the surface toward the MEDATLAS monthly mean climatology. The data assimilation technique involved for the creation of the best estimate of the initial state was System for Ocean Forecast and Analysis (SOFA), which is a reduced order optimal interpolation intermittent scheme (De Mey, 2002)
- HCMR 3 Day-Forecast (Daily output);
- Met Office ⁵: Numerical model output based on NEMO code, with a resolution of $1/12^\circ$. The model is then forced by 3-hourly forecast winds from the Met Office numerical weather prediction (NWP) system. The horizontal momentum diffusion is a combination of laplacian and bilaplacian operators, and the tracer equations employ a TVD advection scheme. The vertical mixing uses

⁵<http://www.metoffice.gov.uk/>

the TKE scheme. 5 Day-Forecast lead time.

We merged these members with some multi-physics member used in the first experiment, forced with ten different realizations obtained by BHM. For all that members forecast was 9 days lead time. In this way we can investigate the uncertainties in initial conditions, the model error due to the parametrizations and the atmospheric forcing impact. Furthermore, for the first time, we can evaluate the impact of grid resolution in ocean simulations using a high resolution model (see appendix B) $1/24^\circ$ (around 4.5 Km at our latitude), developed during this research work. With this large dataset we could study the time evolution of forecast error in different seasons for year 2008: February-March-April ; July-August; October-November. Again in this case the truth state is represented by map of satellite derived SST. Table 2.4 shows the major difference in each member of the dataset.

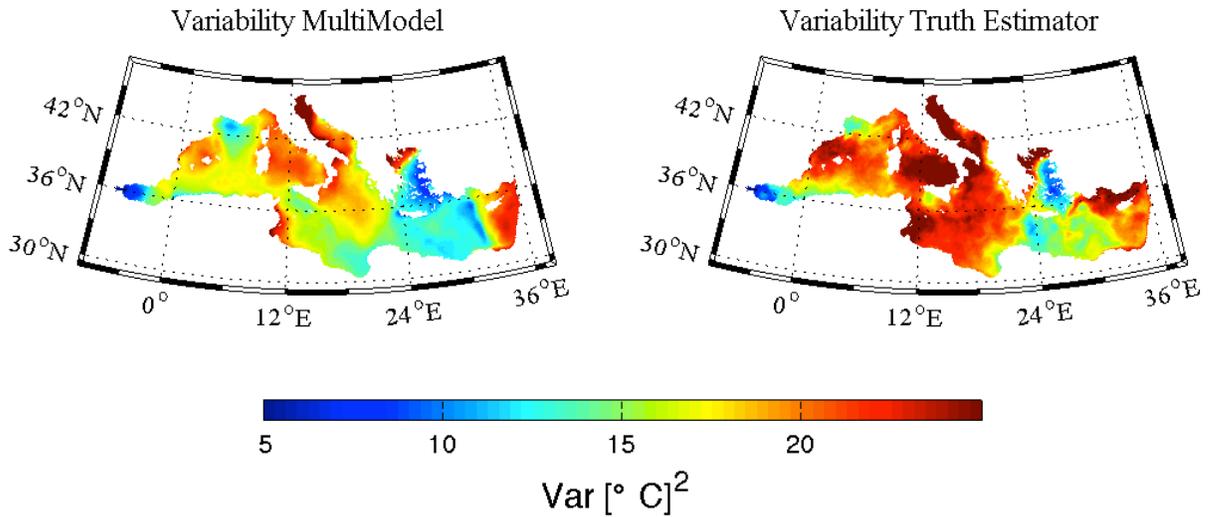


Figure 2.8: Mean Variability generated by the MMSE in forecast mode (left) versus the natural variability in the observations (right) during the time period, 3day forecast lead time

To avoid spurious results due to the forcing lasting, all the comparison could be done only with homogeneous dataset. In this section we will show only the comparisons obtained with 3 days forecast lead time that corresponds to the larger

<i>MMSE Member</i>	Code	<i>Vertical scheme</i>	<i>Diffusion</i>	<i>Viscosity</i>	<i>Forcing</i>	Notes
1	NEMO 2.3	<i>P.P.</i>	Bilap.	Bilap.	0000	
2	NEMO 2.3	<i>P.P.</i>	Lap.	Lap.	0001	
3	NEMO 2.3	$k - \epsilon$	Lap.	Lap.	0002	
4	NEMO 2.3	$k - \epsilon$	Lap.	Lap.	0003	free slip bottom fr.
5	NEMO 2.3	<i>P.P.</i>	Bilap.	Bilap.	0004	
6	NEMO 2.3	<i>P.P.</i>	Lap.	Lap.	0005	
7	NEMO 2.3	$k - \epsilon$	bilap.	bilap.	0006	
8	NEMO 2.3	<i>P.P.</i>	Lap.	Lap.	0007	no slip bottom fr.
9	NEMO 2.3	$k - \epsilon$	Lap.	Lap.	0008	no slip bottom fr.
10	NEMO 2.3	<i>P.P.</i>	Bilap.	Bilap.	0009	free slip bottom fr
11	NEMO 2.3	<i>P.P.</i>	lap.	lap.	0010	
12	NEMO 3.2	<i>P.P.</i>	Bilap.	bilap.	0000	$1/24 Res.$
13	NEMO 2.3	<i>P.P.</i>	Bilap.	bilap.	UKMO	$1/12 Res.$
14	POM	$k - l$	Bilap.	bilap.	HCMR	$1/10 Res.$
15	OPA 8.2	<i>P.P.</i>	Bilap.	bilap.	0000	SYS2b

Table 2.4: *Experiment design for MultiModel SuperEnsemble in forecast configuration, BHM multi-physics members (on the top of the table) and MyOcean database(on the bottom)*

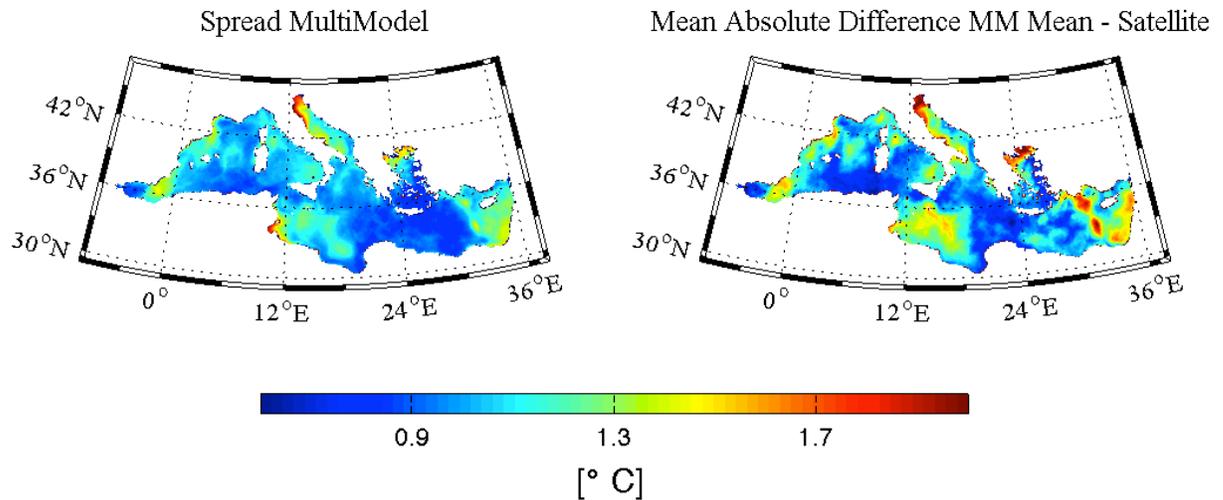


Figure 2.9: *Standard Deviation of MMSE members in forecast mode(left) and the Mean Absolute Difference between Ensemble Mean and the observation(right).*

dataset. For a more complete discussion see section 3.2.4. The results reached in this last experiment are good in terms of correlation in the variability generated by the Multimodel (see 2.8) and the natural variability of the system, but MMSE seems to underestimate variability in some areas (for example over the Adriatic and Tyrrhenian Sea and Sicily Strait), while good correlation and right values, has been found between the spread and the MAD.

Chapter 3

Classical SuperEnsemble approach

Overview

With the availability of ocean predictions produced by several numerical models, multi-model ensemble forecasting has drawn some attention recently (Krishnamurti *et al.* 2000 , Lenartz *et al.* 2010 and Yun *et al.* 2003). Several calibration methods of combining individual forecasts from a group of models to produce an ensemble of predictions have been developed, and in this section we will describe the combination method proposed by Krishnamurti *et al.* (1999). This weighing average methodology will be applied to the ensemble datasets described in the previous chapter. Different verification procedures have been employed to verify the SE results since to assess the quality of a set of forecasts, scalar measures are commonly used, but being scalar measures of a multidimensional problem, they only partially describe the various aspects of the forecasts. The comparison is generally addressed to the degree of similarity between forecast and corresponding Satellite-derived Sea Surface Temperature (SST). Within this study, it has been decided to employ two of the most commonly used indexes: the (bias removed) Root Mean-Squared Error(RMSE), and since we were interested in the spatial comparison between "truth" data and observation in a point, the (Centred) Anomaly Correlation Coefficient, ACC. Some sensitivity tests have been done in order to evaluate the limit of the involved procedure.

3.1 Methodology

The conventional multimodel superensemble forecast (Krishnamurti *et al.* (1999)) can be constructed with bias-corrected data as follows:

$$S = \bar{O} + \sum_{i=1}^N a_i (F_{i,t} - \bar{F}_i) \quad (3.1)$$

where S is the SuperEnsemble (SE) estimate, $F_{i,t}$ is the i -th model forecast for time t , F_i is the mean of the i -th model over the training period, \bar{O} is the observed mean (*Unbiased Estimator*) over the training period, a_i are regression coefficients obtained by an appropriate minimization procedure during the training period, and N is the number of models involved. The systematic errors of ensemble members in eq (3.1) are removed because the anomalies term ($F_{i,t} - \bar{F}_i$) in the equation accounts for each model's own bias. The linear regression technique, used to compute the coefficients a_i , involves a minimization function that acts to limit the spread between the variables of member models and observed state. The regression coefficient (weights) a_i are computed for by a point-wise multiple regression technique in each grid point by minimizing the following:

$$G = \sum_{t=0}^{T_{train}} (S_t - O_t)^2 \quad (3.2)$$

So we have to substitute in 3.2 the 3.1 and obtain:

$$G = \sum_{t=0}^{T_{train}} \left(\left[\bar{O} + \sum_{i=1}^N a_i (F_i - \bar{F}_i) \right]_t - O_t \right)^2 \quad (3.3)$$

The minimum of (3.3) is found by setting its gradient to zero.

Since the equation contains N parameters there will be a system N equations.

So for $j=1\dots N$ we will have:

$$\frac{\partial G}{\partial a_j} = 2 \sum_{t=0}^{T_{train}} \left(\left[\bar{O} + \sum_{i=1}^N a_i (F_i - \bar{F}_i) \right]_t - O_t \right) \left[\sum_{i=1}^N \frac{\partial a_i}{\partial a_j} (F_j - \bar{F}_j) \right] = 0; \quad (3.4)$$

$$\frac{\partial a_i}{\partial a_j} = \delta_{i,j}$$

So the sum in the last term reduces in :

$$(F_j - \bar{F}_j) \equiv b_{j,t} \quad (3.5)$$

$$\sum_{t=0}^{T_{train}} \bar{O} b_{j,t} \equiv \beta_j \quad (3.6)$$

$$\sum_{i=1}^N a_i \sum_{t=0}^{T_{train}} b_{i,t} \cdot b_{j,t} \equiv \sum_{i=1}^N a_i \cdot \gamma_{j,i} \quad (3.7)$$

The covariance $\gamma_{j,i}$ matrix is built with the multi-model members anomaly ($b_{j,t}$):

$$\sum_{t=0}^{T_{train}} O_t b_{j,t} \equiv \varepsilon_j \quad (3.8)$$

With this definitions the system reduces in a set of data as a linear function of input data.

$$\sum_{i=1}^N a_i \cdot \gamma_{j,i} = \varepsilon_j - \beta_j \equiv \phi_j \quad (3.9)$$

or in vectored form:

$$\vec{a} \cdot \Gamma = \vec{\phi} \quad (3.10)$$

Where $\vec{\phi}$ is vector containing the covariances of the observations with the individual models for which we want to find a linear regression formula, Γ is the covariance matrix, and \vec{a} is the vector of regression coefficients (the unknowns). In the conventional superensemble approach, the regression coefficients are obtained using

Gauss-Jordan elimination with pivoting:

$$\vec{a} = \vec{\phi}\Gamma^{-1} \quad (3.11)$$

The covariance matrix Γ and $\vec{\phi}$ are rearranged into a diagonal matrix, and the solution vector is obtained as in eq.(3.11) The Gauss-Jordan elimination method for obtaining the regression coefficients is not numerically robust (Yun *et al.*, 2003). Problems arise if a zero pivot element is encountered on the diagonal of the matrix since the solution procedure involves division by the diagonal elements. Overall to evaluate the inverse of a matrix, the determinant involved is closer to zero when highly correlated measured are used. To avoid ill-conditioning problem we have to remove the degeneracy of covariance matrix enlarging the training period length. Some studies (Krishnamurti *et al.*, 2000) have interpreted the regression coefficients as indicators of the relative model "*reliability*". However, this interpretation needs some specification as Kharin & Zwiers (2002) had done. For instance, consider a simple ensemble dataset of two members, the first M_1 that overestimates, and the latter M_2 which underestimates the true state of the field we want to reproduce, O , as depicted in fig. 3.1 where $M_1 = 1.5 \cdot O + \epsilon_1$ and $M_2 = 0.5 \cdot O + \epsilon_2$. The members

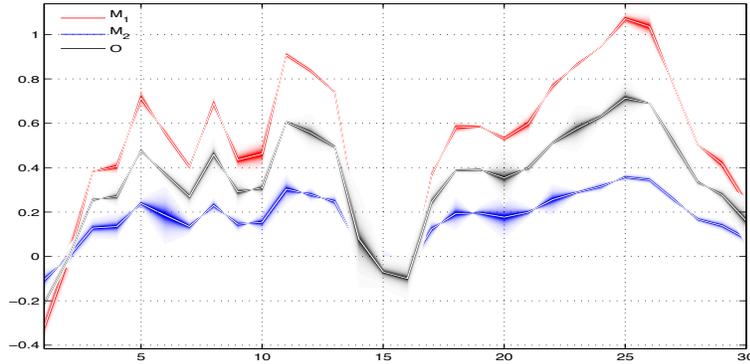


Figure 3.1: Anomaly of a simple dataset

will have the same RMSE during the training period, equal to: $0.5^2 \cdot \sigma_O^2$. If we solve the system 3.11 we can find that $a_2 = 1/4$ while $a_1 = 3/4 = 3a_2$. As shown we can infer that "*equally reliable model outputs may not necessarily be weighted equally when*

combined optimally".

Furthermore, a simple *"toy experiment"* has been done in order to check the value of coefficients, using the truth estimator as an ensemble member. This test, can be considered as the maximum skill that could be achieved with a multi-model and it is the only way to check the perfect multi-linear combination coefficients estimates. As expected, all the regression coefficients are zero, except the weight related to "fake member", which is set as 1. Enlarging or trimming the dataset (adding or removing some members), we noticed that when the training period units are less than the number of multi-model members involved, the algorithm fails, giving wrong values to the coefficients. We can infer that the regression procedure is not able to recognize the "best" member when it is trained for a (relatively to the dataset size) tiny number. Hence a threshold constraint for the training period must be:

$$Length\ Training\ Period \geq Models\ Number \quad (3.12)$$

3.2 Results

Verification methods

A wide range of possible scores are available in order to asses the ranking of a forecast, and for each method we can find the *"best"* and *"worst"* model, according to a chosen score. Following Murphy (1993), a good forecast should satisfy the three *"desirable"* properties and for each one we can set an appropriate score(in brackets):

- Consistency(Anomaly Correlation Coefficient, ACC)
- Quality(Root Mean Squared Error, RMSE)
- Value(Mean Square Error Skill Score, MSESS).

The mean error, *i.e.* the sum of the deviation from the reference value during the test period, is an inadequate measure of the skill of a prediction, since positive-

negative values can compensate in turn. A simple way of avoiding compensation errors is achieved considering the absolute value of the deviances, Mean Absolute Error (MAE), or in turn taking the root mean squared of the deviances. We referred to the latter quantity as Root Mean Squared Error (RMSE), which is described by the following statistic law:

$$RMSE = \sqrt{\sum_i^N \frac{(F_i - O)^2}{N}} \quad (3.13)$$

Since we are accounting on the squares, this measure is more sensitive to large forecast errors than other methods (like the Mean Absolute Error). For this reason we subtracted the estimated model bias to recalibrate it, centring the error of the ensemble members and the observations around their respective mean value.

$$RMSE_c = \sqrt{\sum_i^N \frac{(F_i - \bar{F}_i - O_i - \bar{O})^2}{N}} \quad (3.14)$$

Furthermore, since we are evaluating the performance of an "*expensive*" post-processing method against the "*classical*" mean or the default value from a model, it is straightforward to use (a generalization) of the Mean Square Error Skill Score, which is defined by Murphy & Epstein (1989) as:

$$MSESS = 1 - \frac{MSE_{SE}}{MSE_R} \quad (3.15)$$

MSE_{SE} is the squared of the term evaluated by 3.13 and the denominator is the same but for the reference member. Its maximum value is 1 which indicates the perfect forecast, and decreases to zero as the forecast skill is equal to the climatological (or in our case the reference member). A negative value implies that the model forecast is worst than the reference member. It is dimensionless and increase with the forecast skill. A good method developed for the spatial evaluation of a forecast is the Anomaly Correlation Coefficient, which is the correlation between the forecast and the observed anomaly over a grid. We used the centred version in order to take

in account of the respective bias. In this formulation (3.16)

$$ACC_c = \frac{\left[(f - c) - \overline{(f - c)} \right] \left[(a - c) - \overline{(a - c)} \right]}{\sqrt{\left((f - c) - \overline{(f - c)} \right)^2 \left((a - c) - \overline{(a - c)} \right)^2}} \quad (3.16)$$

Here f is the value of the forecast, c is the climatological value (in our case the truth estimator \bar{O}) and a stands for the behaviour of the validating analysis dataset. ACC is equivalent to the Pearson product moment correlation coefficient of the anomalies, and measures how well a forecast captures the magnitude of the anomalies. As a correlation, it ignores the bias so it is more appropriately considered a measure of potential performance.

3.2.1 Results Multi-physics experiment

The multi-physics ensemble was generated using a variety of physical parametrizations available in NEMO model described in section 2.1. Here we will show the results after the verification procedure. Since observations from ARGO and other *in situ* measurements were too sparse in space and in time, we decide to use the SST satellite derived for the comparison. In this way we set the unbiased estimator as the time mean of SST during the training period (see fig. 3.2) from the 7 January to

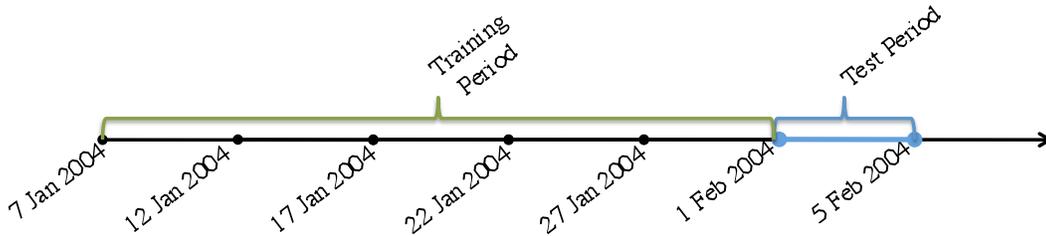


Figure 3.2: resulting maps for the first experiment, SE on the top, SST Satellite-derived(middle) and the worst MP member (bottom)

31 January 2004. During the same period we evaluate regression coefficients according eq.(3.11). The figure 3.3 shows the retrieved SE prediction, truth state and the

output of the worst member corresponding to the first day of the test period. The SE prediction was a highly discontinuous field, where a large fraction of grid points gives unlikely results such as $SST > 50\text{ }^{\circ}\text{C}$ or $SST < 0\text{ }^{\circ}\text{C}$. The performances of the SE and the other members are assessed by the evaluation on RMSE domain average and showed in figure 3.4, each multi-physics member is better than the SE prediction. Hence the SE is the worst predictor, this is due to the *rank-deficiency* of the covariance matrix Γ . The determinant is almost zero, since the rows are almost *small perturbation* of linear combination of each other. Matrix ill-conditioning is a general term used to describe a matrix of values which is unsuitable for any use in a particular analysis. The condition number measures how small perturbations in the data affect the answer. This problem occurs frequently in applications of linear multiple regression when the matrix of correlations for the predictors is singular and thus the regular matrix inverse cannot be computed.

3.2.2 Results in the BHM experiment

The dataset used in this study was generated by 22 runs with MFS BHM-winds over the Mediterranean area and described in section 2.2.

Training period:

- 1 February-31 March for years 2005 2006 2007 and 1-29 February 2008;
- 1 simulation driven by ECMWF analysis +10 runs with MFS-BHM - Winds of Operational Model (OPA 8.2);
- 1 simulation driven by ECMWF analysis +10 runs with MFS-BHM- Winds NEMO (OPA 9.0);

Event Period :

- 1-31 March 2008;
- 1 simulation driven by ECMWF analysis +10 runs with MFS-BHM - Winds of Operational Model (OPA 8.2);

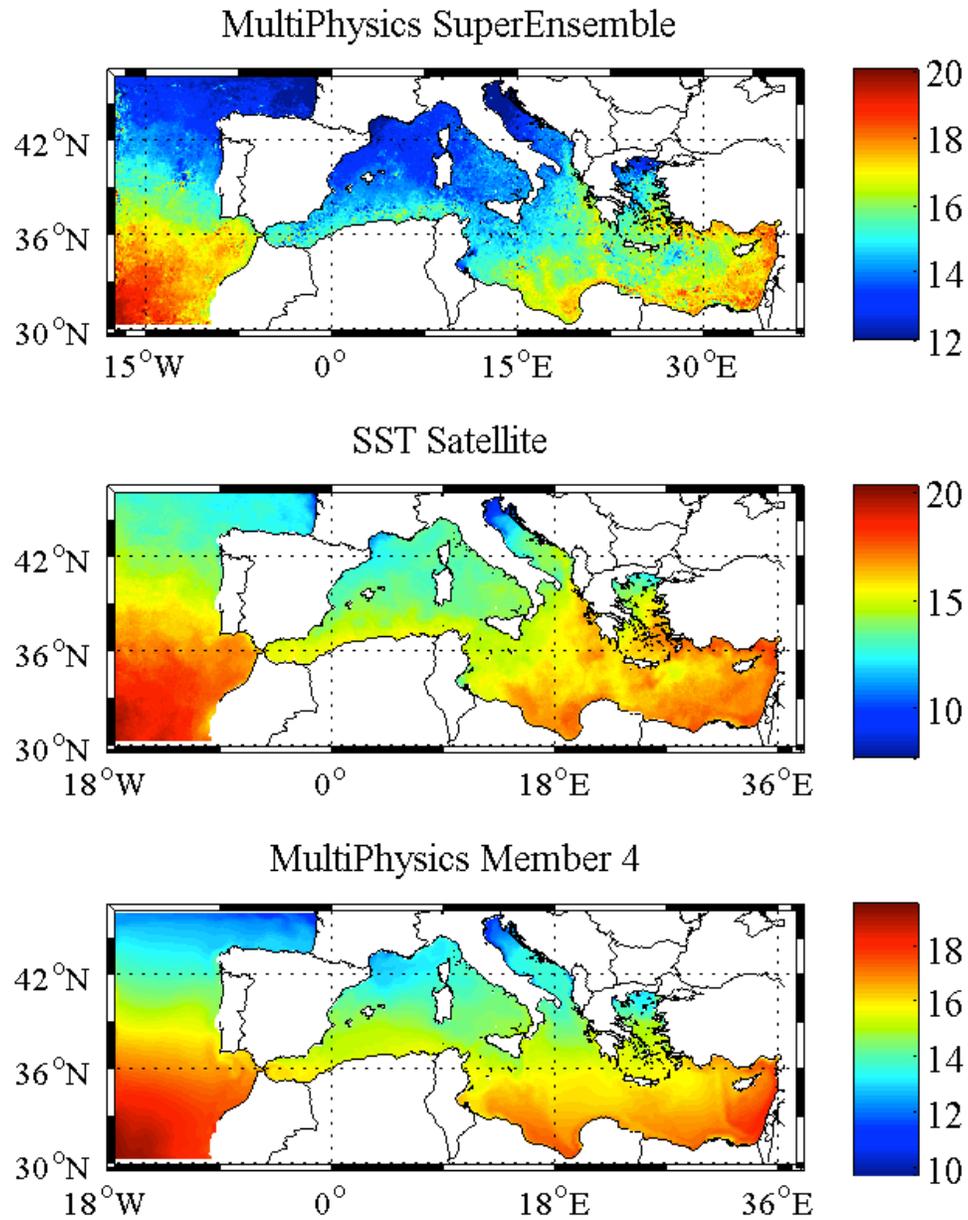


Figure 3.3: resulting maps for the first experiment, SE on the top, SST Satellite-derived(middle) and the worst MP member (bottom)

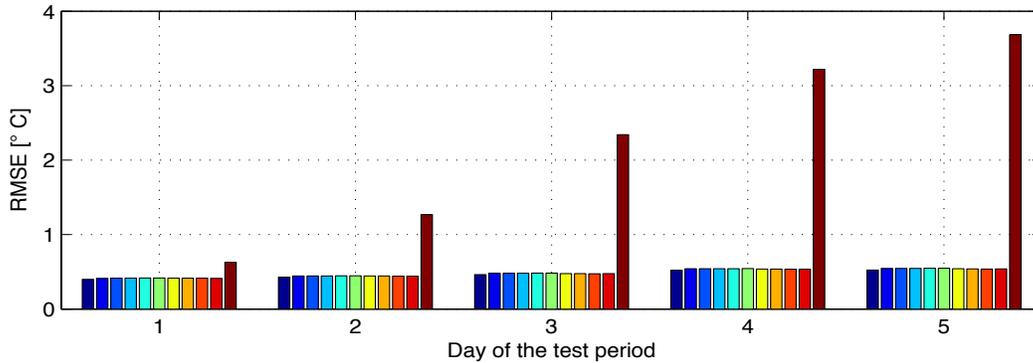


Figure 3.4: *RMSE between truth observation and MP members (coloured bar) and SE (brown bar) for SST field, during the five days of the test period*

- 1 simulation driven by ECMWF analysis +10 runs with MFS-BHM- Winds NEMO (OPA 9.0);

As first, we checked the algorithm behaviour during the training period, to be sure that we had a good dataset, and to avoid any errors in the regression procedure and the results (see fig.3.5) confirm that the SE has been regressed in the optimal way for the most of the training period both in term of Root Mean Squared Error and Bias Removed RMSE.

So we applied the regression coefficients in the test dataset, and we evaluated the RMSE and the Bias Removed RMSE during the test period (fig 3.6). In this case, the SE can outperforms all the models, at least in the first part of the test phase.

The results for the BIAS, (fig. 3.7) indicate that all members involved under-predict temperature (Bias <0). The analysis of ACC, fig.3.8, indicates that BHM MM members show better performance in displacing the anomaly of the temperature during the test period. An important consideration when assessing the impact of the multi-model approach is related to choice of the metric used in diagnostic and training period. The solid bars in fig. correspond to the MSESS between the SE and each MM members, this comparison display a gain in the prediction using the SE since the MSESS is always positive. Usually only one particular diagnostic, the RMSE has been used to demonstrate the out-performance of SE, but as shown in

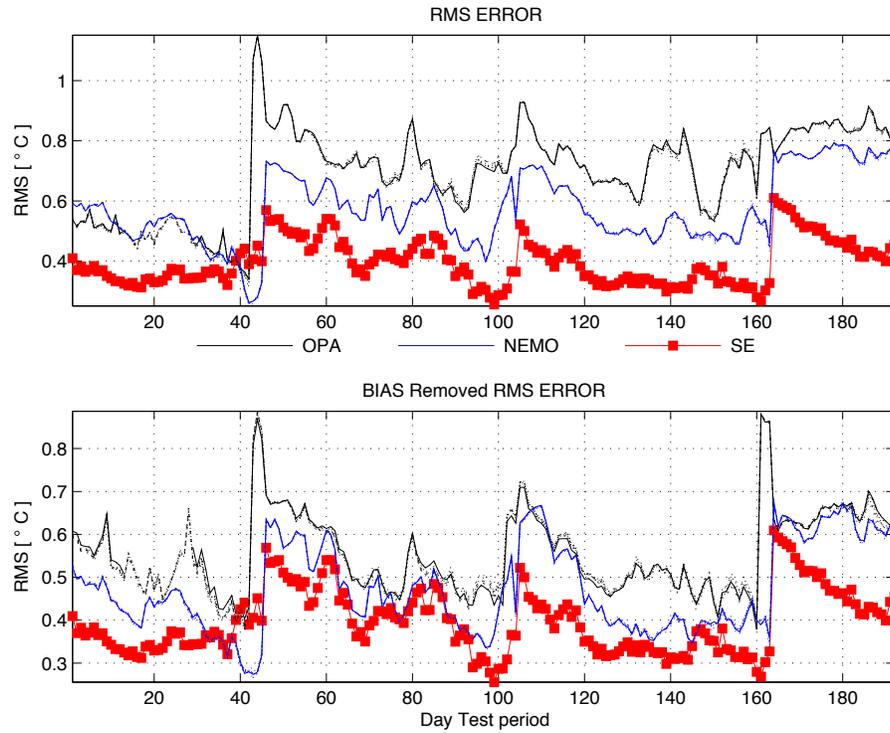


Figure 3.5: performance during the training period of the SE prediction (red), OPA BHM members (black lines) and NEMO BHM members (blue lines). RMSE (top panel) Bias Removed RMSE (lower panel)

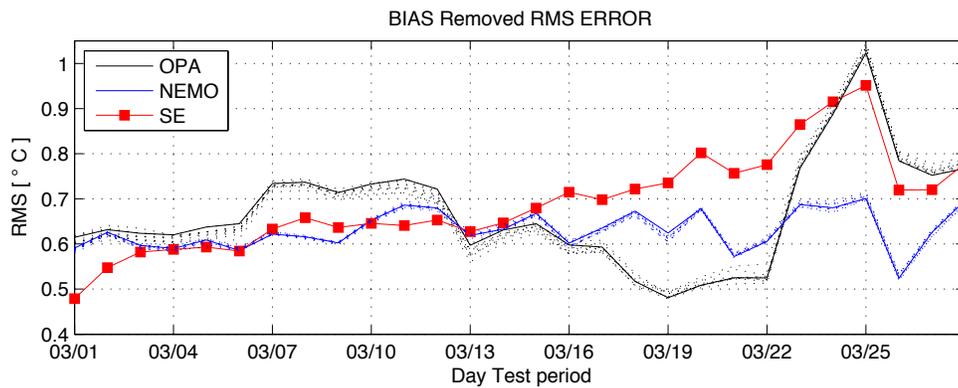


Figure 3.6: Domain averaged Bias Removed RMSE during the test period.

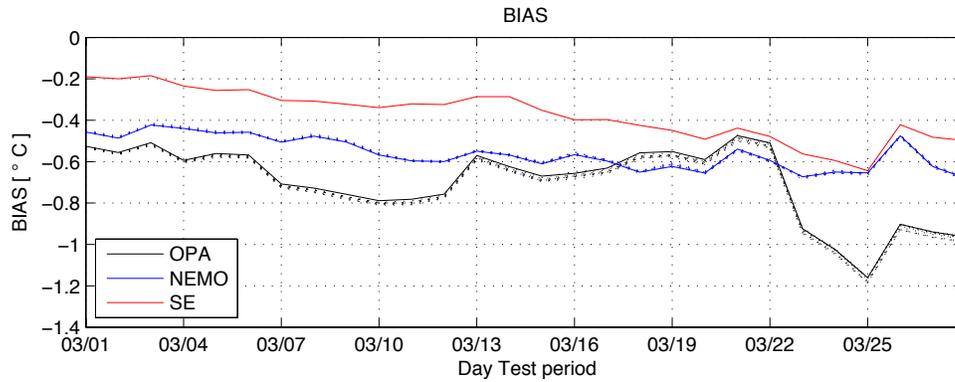
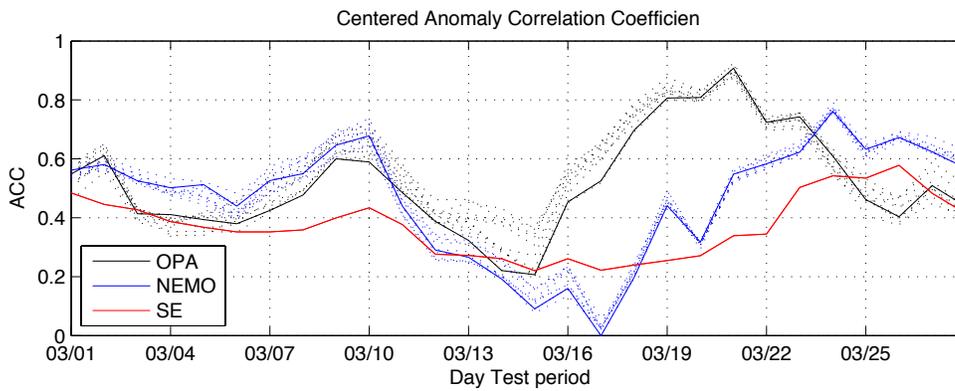
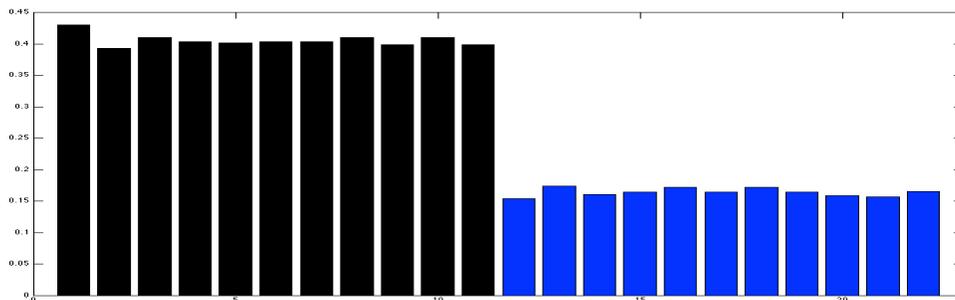
Figure 3.7: *BIAS during the test period.*Figure 3.8: *ACC during the test period.*Figure 3.9: *MSESS for the 22 BHM MM members and SE for the parameter SST over Rhode Gyre region during the test period (March 2008). Black bars represent gain respect OPA predictions, and blue bars the gain respect NEMO*

fig.3.8 we don't reach the same result changing the diagnostic metric. So we can infer that SE prediction trained with RMSE can not skilful in diagnostic ACC metrics. An examination of the fifth day of the prediction over the Rhode Gyre region (fig. 3.10) shows some unrealistic features, confirming the low ACC skill of SE. The map seems to be over-fitted by the regression procedure against the observations. *Overfitting* (Tetko *et al.* (1995)) occurs when a regression model begins to memorize training data rather than learning to generalize from trend. In our case, since the number of parameters is less than the number of observations, the regression model learnt to perfectly predict the training data simply by memorizing them. In this way it fails drastically on unseen data, as it has not learned to generalize at all. To avoid this problem it is necessary to use additional filtering techniques (described in chapter).

3.2.3 Results MultiModel Multiphysics experiment

In the previous experiment we noted that SE predictions can have higher skill compared to all the participating multimodels. This experiment is purposed to study the sensitivity of the regression procedure versus the major impacting factor of its computation: training period length and composition of the dataset.

3.2.3.1 Results in the MMSE experiment Analysis Mode

In this case we have used the extensive dataset provided by MyOcean project described in section 2.3.1. A range of experiments has been carried out to test the performance of the method proposed by Krishnamurti, setting the unbiased estimator as the Sea Surface Temperature from Satellite. To assess the sensitivity on training period lengths, we fixed the length of the test period as 10 days, and we adopted the training period as the last n days from the first day of the prediction, fd . For example, when we try to forecast the SST in the 1st of July 2008, the training is performed using maximum n cases, with training period set as n days before the first day, fd . As we will show in section 3.2.3.2 that the training period length can not be reduced beyond a threshold that can ensure robust optimal weights from the regression algorithm, which in our case has been set as 14 days. So for each day fd

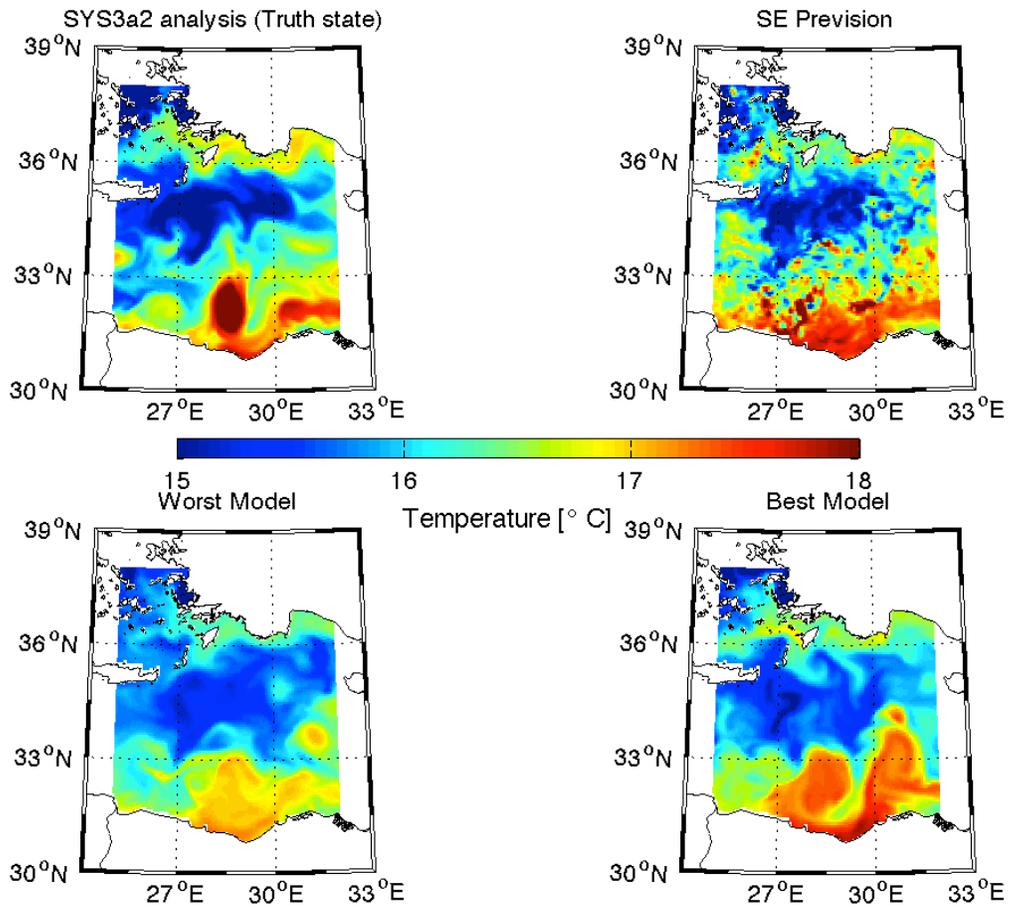


Figure 3.10: A typical example of a reconstructed map of SST, on day five over Rhode Gyre region, valid the 5 March 2008: analysis (top left), lowest skill model (bottom left), SuperEnsemble (top right) and best participating model (bottom right).

we could have 46(60 has been set has the maximum training period length minus the threshold value 14) SE realizations experienced by the corresponding training period.

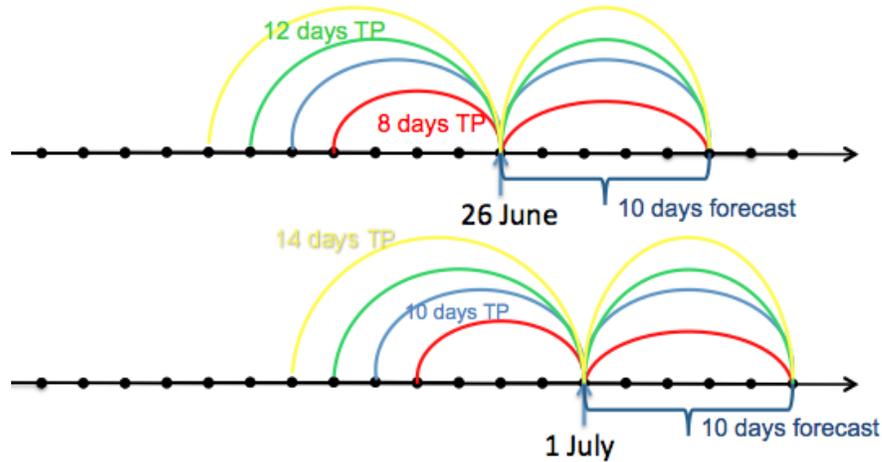


Figure 3.11: *Schematic representation of the experiment carried in MMSE analysis mode.*

To find the optimal training period length we repeated the same regression different times during the year, changing training period from a minimum of 14 days to maximum value of 60 days. Due to large amount of cases, 2532 during the year 2008, we decided to analysed the results in terms of different skill scores rather than changing each day. Generally the reconstructed map of SST show that within a range of 30-40 training days the superensemble is best predictor for the SST at least for the early 4-5 days of the test period, and its performances deteriorates in time, for example fig.3.12 depicts the resulting map of SST for the first day of the test period, 1 June 2008 obtained by 41 day of training period while fig3.13), depicts the resulting maps for the 10 June 2008 also derived trainging the algorithm for 41 days, we can see in the latter figure, the noise is incremented in SE realizations (top left of both figures) which is confirmed by the curve of time-series of the domain average centred RMSE which reach its maximum value in the last day of the test period. On the opposite side ACC curves (fig. 3.15)shows a "U shape", until the second day SE can outperforms all the ensemble members, after the best performances are obtained

by the single-model members. In fig.3.16) we can see the better performances of SE prediction in terms of BIAS (removed during construction) until the fourth day, again a single-model member is the best for the remaining day of the test period (the blue curve, SYS3a2 is closer to zero than the SE, red curve). The ranking of SST performance are almost constant according each metric applied. The best model in centred RMSE term is almost the best also in ACC and BIAS ranking. We can conclude that the features of the ranking are consistent across the whole range of used skill measures.

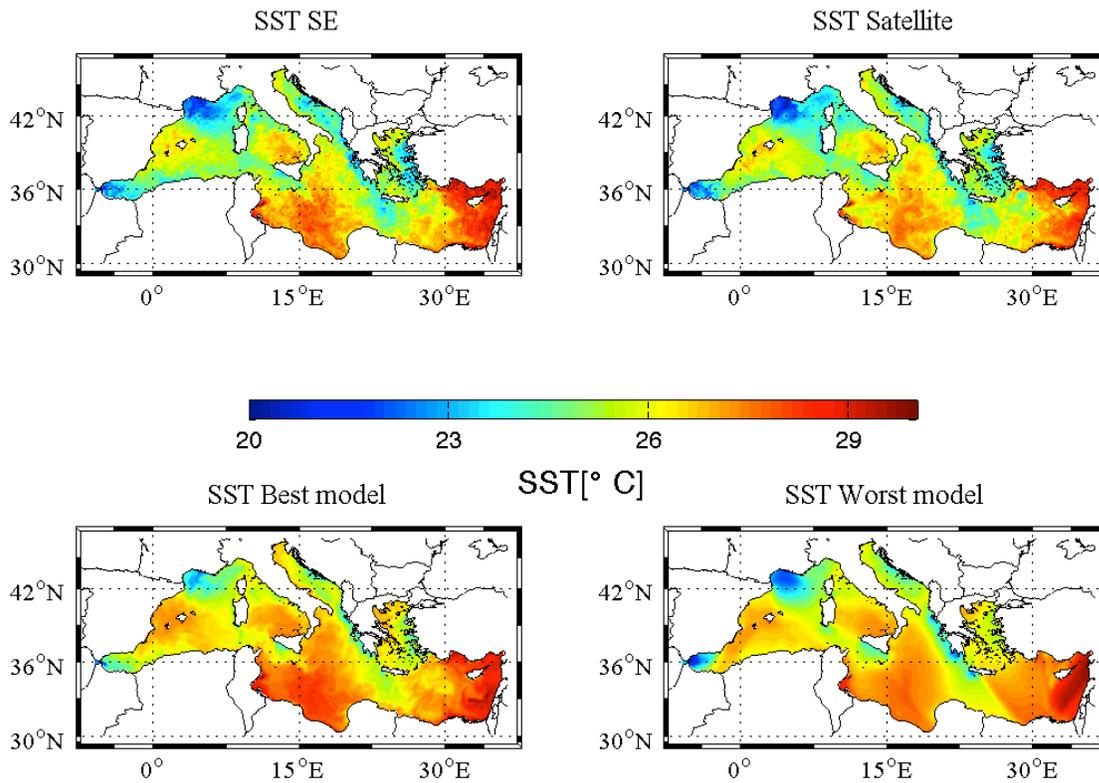


Figure 3.12: A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of June 2008 reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)

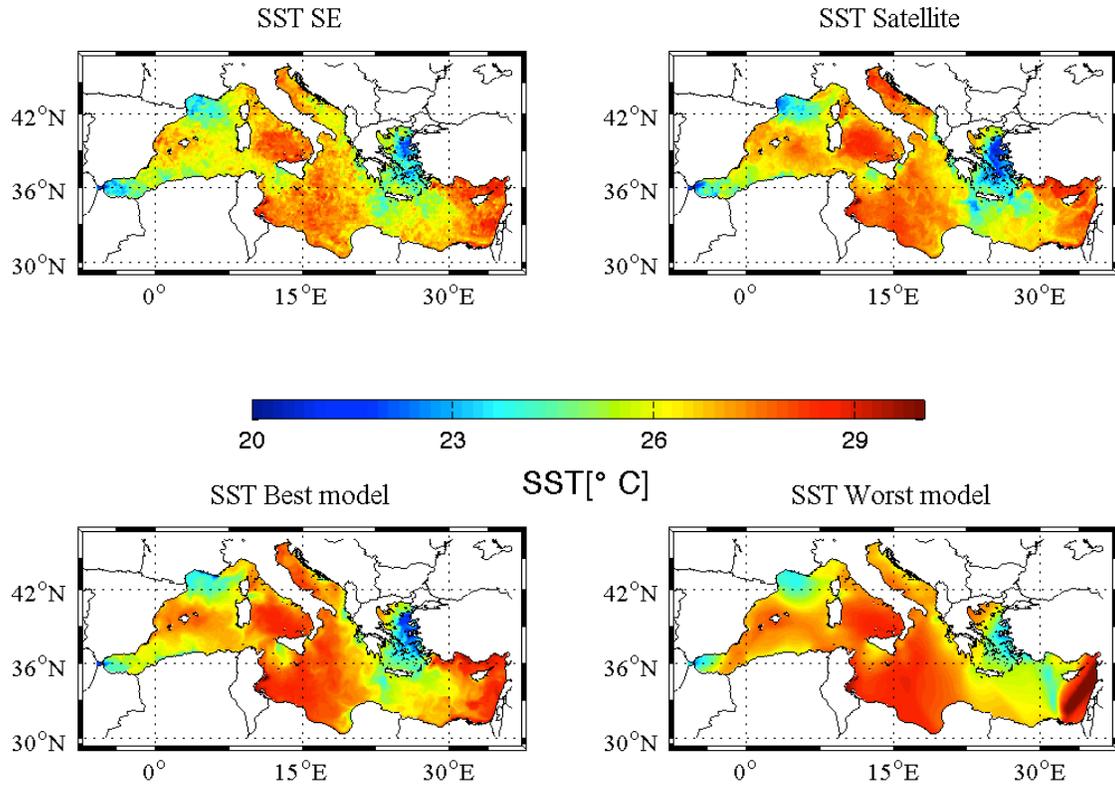


Figure 3.13: A typical example of prediction for the last day of test period over the Mediterranean area, valid on the 10th of June 2008, reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)

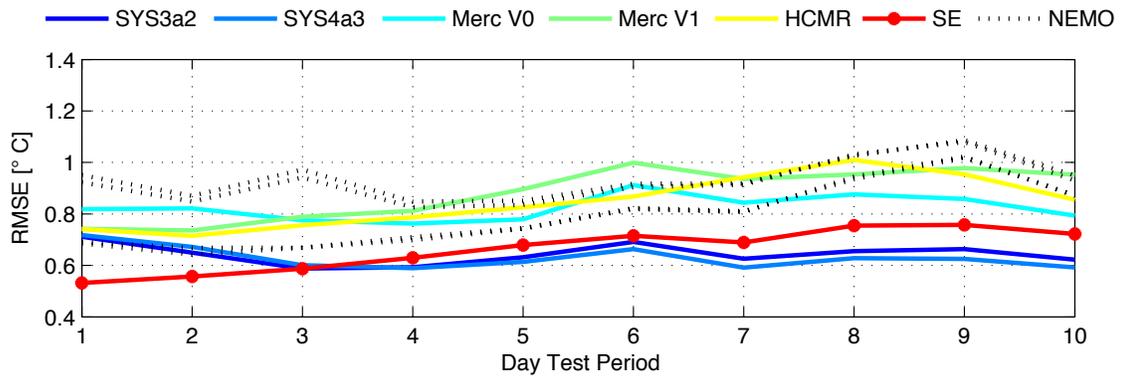


Figure 3.14: Time serie of the domain average (over the Mediterranean) Bias removed RMSE

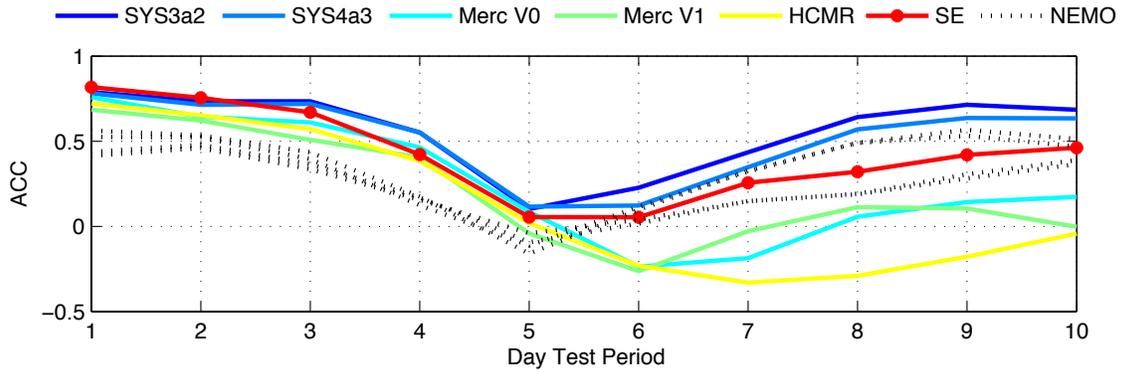


Figure 3.15: *Anomaly Correlation Coefficient time series*

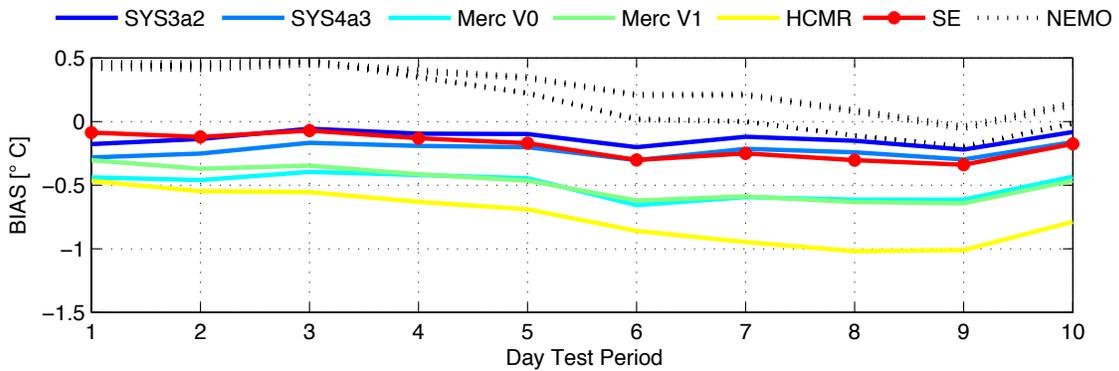


Figure 3.16: *Bias (Model - Observation) Time serie over the Mediterranean sea)*

3.2.3.2 Study on the sensitivity of the algorithm

Sensitivity on the training period length During year 2008, we carried out 2532 experiments, and in each we changed the initial day of the training period and its length. Hence for each test period, we can have several predictions. For example the previous map are the results from the 1 to the 10 th June, reached training the algorithm for 41 days, but during the same test period we have other realizations, obtained training the algorithm for a different period. All those realizations have been over-imposed in the same graph (figures: 3.17, 3.18 and 3.19) in order to asses the dependence with training period, the lines colour, in the box on the right side graph, stand for training period involved. First, all the scores used to asses the prediction show a dependence with the training period length. As expected the most sensitive score is the RMSE, which curve arise more slowly (becoming almost flat) for training period longer that 40 days (see fig. 3.17). An important feature displayed in figures 3.18 and fig. (3.19) is the relative improvement of the multi-model performance defined by other skill scores (ACC and BIAS), enlarging the training period. Furthermore all the experiments show that there is a critical day in the test period, in which the SE performance matches the performance of an other ensemble member, so for all the following days the superensemble won't be the best predictor (fig.3.14), for the following day the best performance are obtained by the single-model member. We refer that day as "*skill*" of the SE prediction, in this way, using only one number, we could study more easily the evolution of the skill with the training period compared to the other ensemble members. All the skills , for RMSE (fig. 3.20), ACC (fig.3.21) and BIAS(fig. 3.22) used to asses the prediction, show a weak dependence with the training period. As expected only the bias is almost constant according the training period length. We over imposed all the curves of skill, in order to identify the different growth rates. As we can see ACC curve grows faster and reach a plateau around 40 days of the training period. On the other hand RMS is almost constant until 35 days of training period and then seems to growth, but as we have shown in BHM experiment this due to the regression procedure, which perfectly predict the training data by memorizing them leading to have unlikely results in the SST maps.

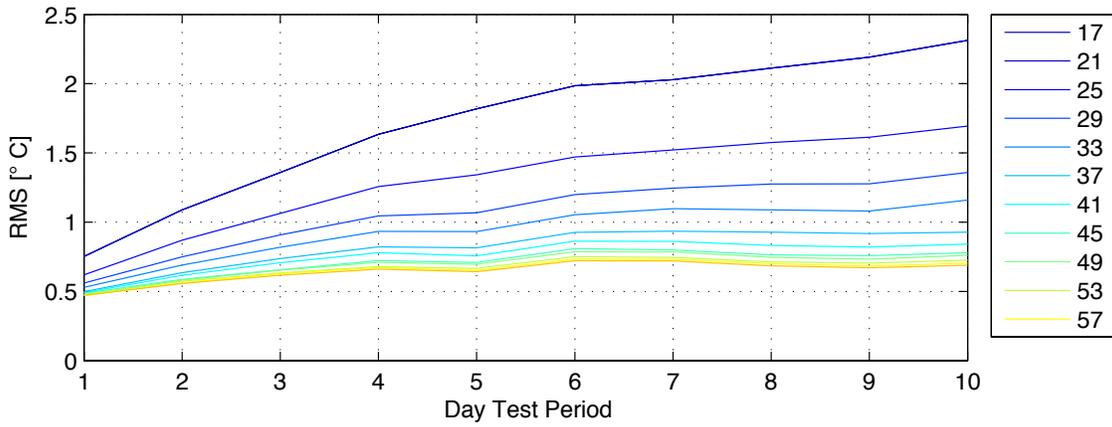


Figure 3.17: Domain average RMSE comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line repret shorter training period, yellow line display longer training period

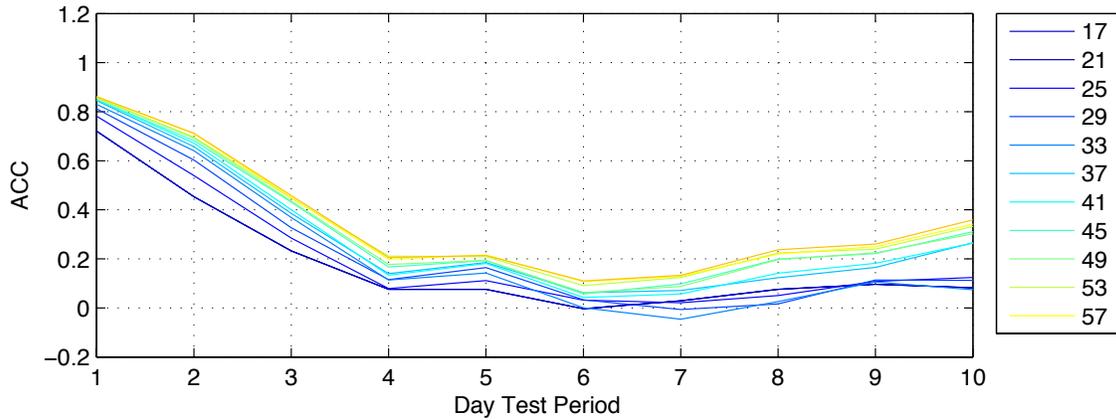


Figure 3.18: ACC comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line repret shorter training period, yellow line display longer training period

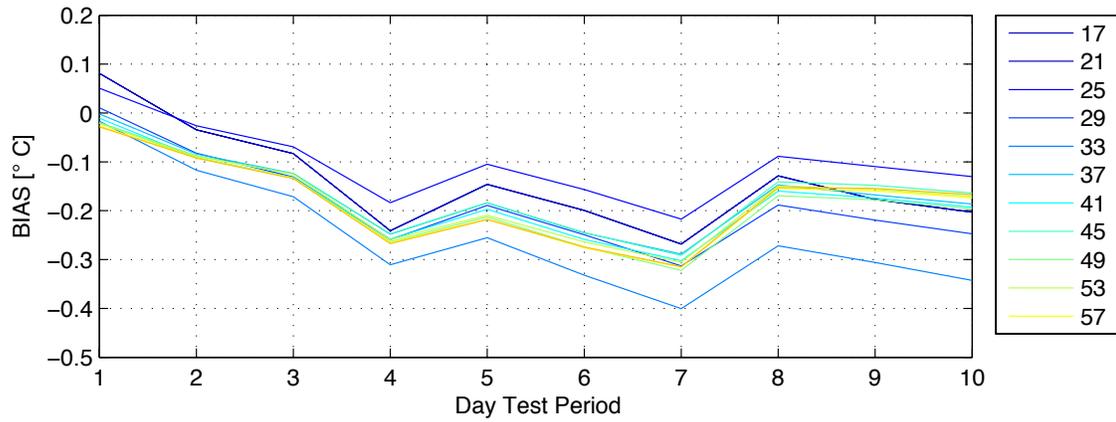


Figure 3.19: BIAS comparison for the same test period(1-10 June 2008), with different training period. Dark Blue line represent shorter training period, yellow line display longer training period

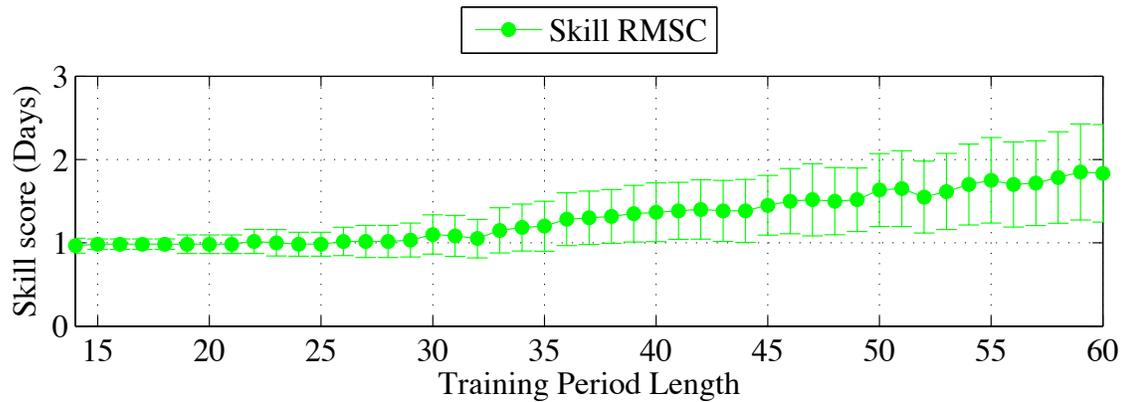


Figure 3.20: Bias Removed RMSE mean Skill of the Multi Model superensemble with 9 members against the trainin period legths (green spots) the error bar is the standard deviation

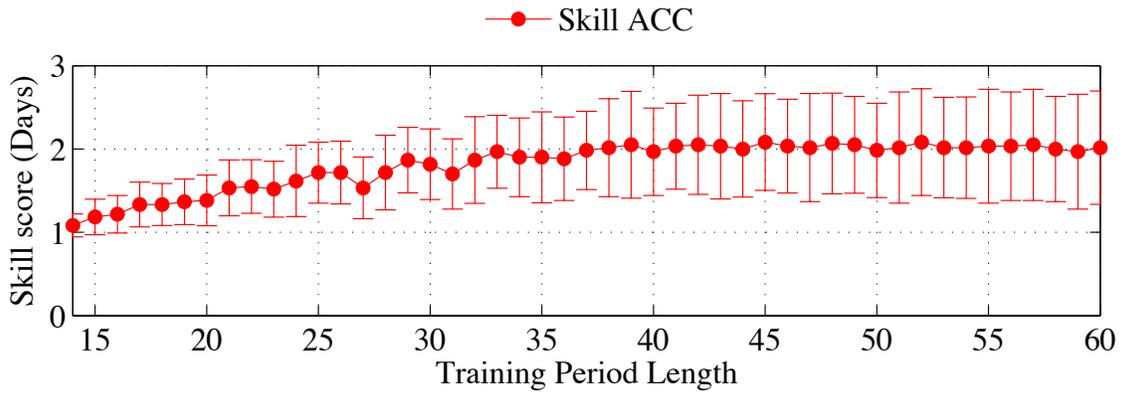


Figure 3.21: ACC mean Skill of the Multi Model superensemble with 9 members against the trainin period legths(red spots) the error bar is the standard deviation

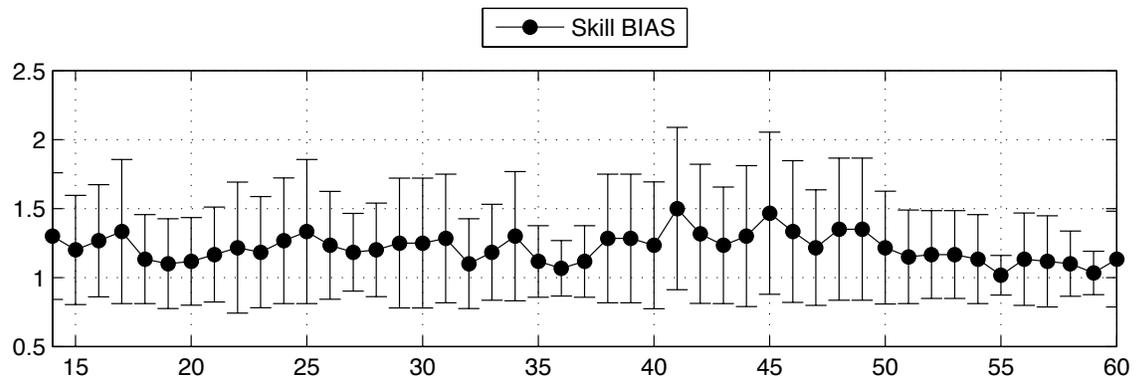


Figure 3.22: Bias Skill of the Multi Model superensemble with 9 members against the trainin period legths(black spots) the error bar is the standard deviation

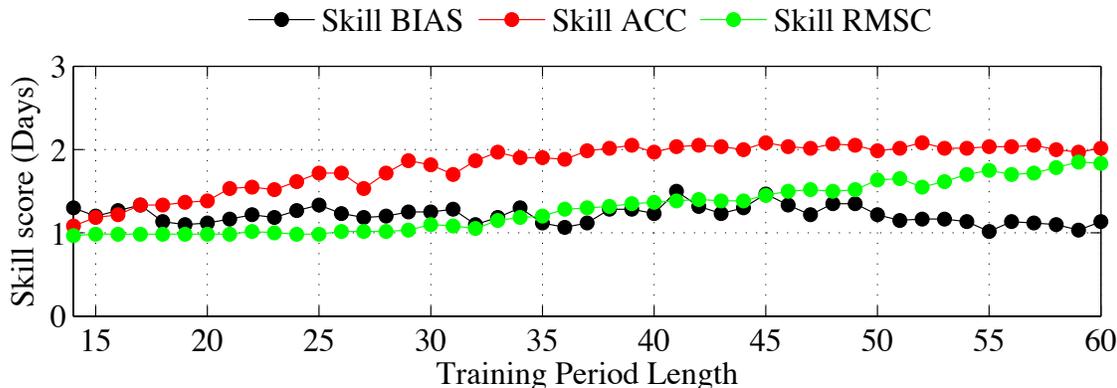


Figure 3.23: Comparison of the mean skills for BIAS(black spots), ACC (red spots) and Centred RMSE(green spots)

Sensitivity on the the dataset The aim of this study is to study the impact in SE performance adding or removing a "bad" or a "good" model (*i.e* a model that consistently performs worse than average of the ensemble, and complementary the best model) in order to identify the condition under which the MMSE approach really enhance the prediction skill. Taylor diagrams (Taylor (2001)) have been introduced for model inter-comparison, providing a way of graphically summarizing how closely a pattern matches observations. The similarity between two patterns is quantified in terms of their correlation, their centred root-mean-square difference and the amplitude of their variations (represented by their standard deviations). From the analysis of the Taylor diagram (3.24) three different sub-samples have been created from the complete dataset in order to test the sensitivity to multimodel goodness and to study the impact of overconfident (Weigel *et al.* (2008)) dataset or too correlated ensemble members:

- **Multimodel constructed from overconfident models ensemble:** Sub-sample A, the members are the operational product, SYS3a2, SYS4a3, Mercator V0, Mercator V1, HCMR and NEMO operational in simulation.
- **Multimodel constructed from well dispersed models ensemble:**Sub-sample B, the members are SYS3a2, HCMR and the four simulations with multi-physics members NEMO.

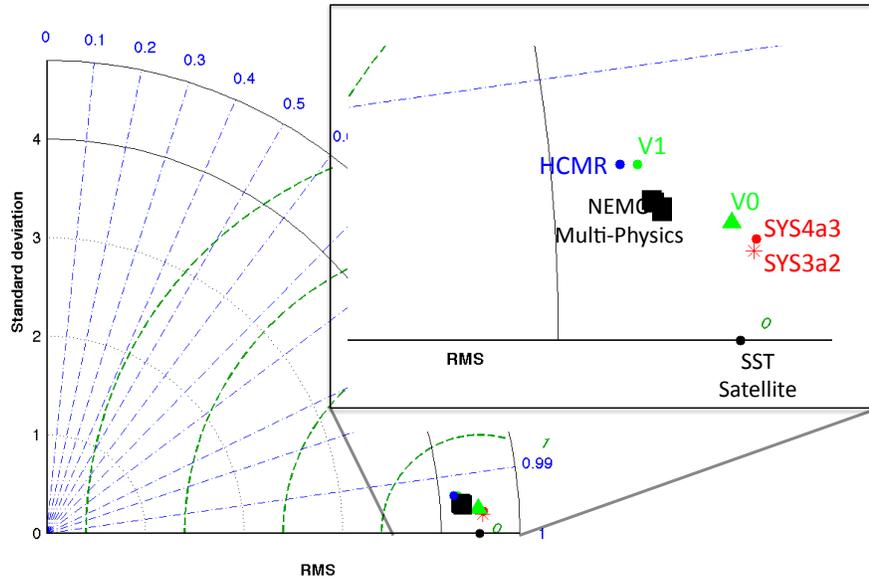


Figure 3.24: Domain average

- **Multimodel constructed from bad dispersed models ensemble:** Sub-sample C, Mercator V1; HCMR and four simulations with multiphysics members NEMO.

For each dataset we carry out again all the experiment done with the original dataset and compare the results using the previous diagnostic metrics. To avoid misleading we compared the skill only between the overconfident dataset and the original because in the other dataset the potential better performance would be a paradox due to deteriorating dataset performances. As we can see the overconfident dataset show better skill scores and faster skill growth versus the training period length respect the original 9 member ensemble but again for longer training period we reach the plateau for ACC, meaning that incur again in overfitting. In figures 3.30, 3.31 3.32 we show the sensitivity on the skill scores versus the training period and the sub sample. As we can see the original dataset isn't the best predictor for short training period, but the differences decrease enlarging the training period, letting the bad dispersed dataset be the worst of the available predictions. As we can see

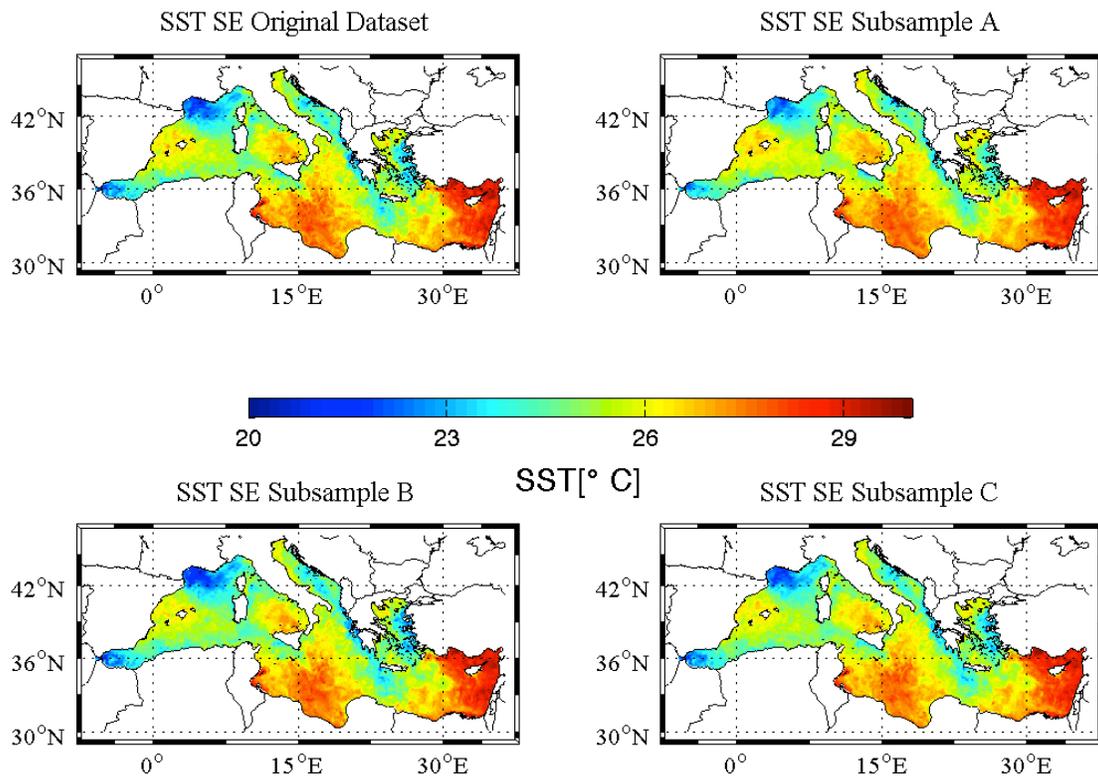


Figure 3.25: A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of June 2008 reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)

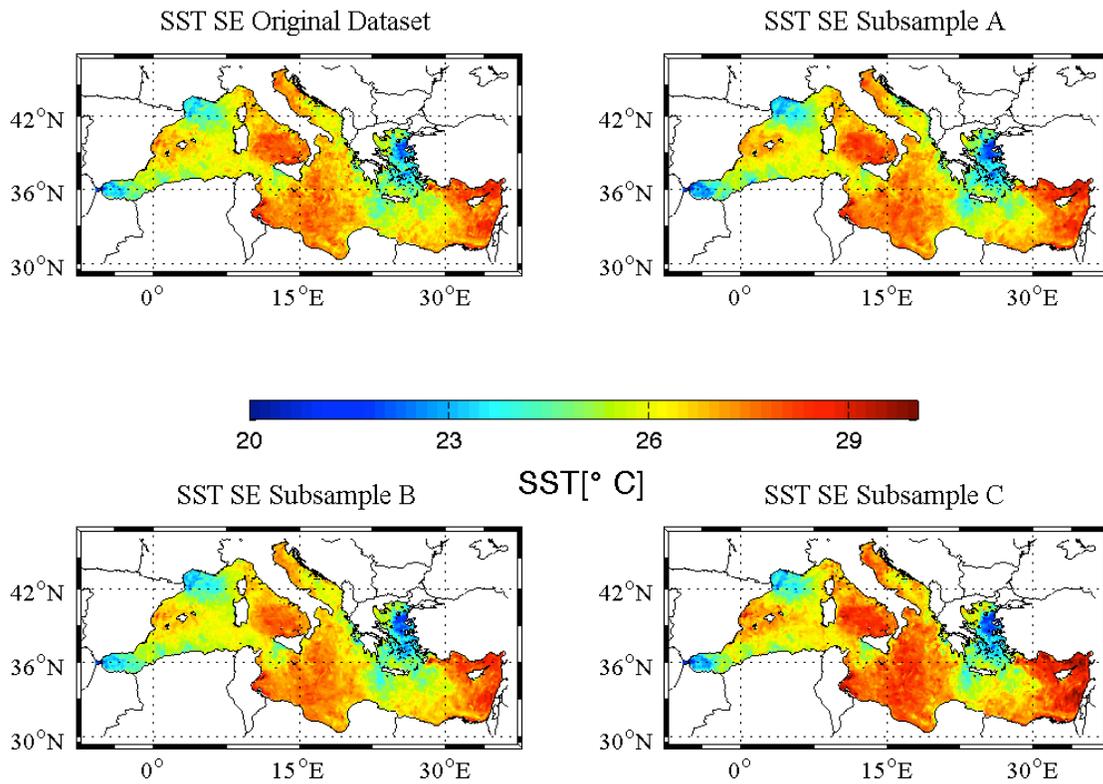


Figure 3.26: A typical example of prediction for the last day of test period over the Mediterranean area, valid on the 10th of June 2008, reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)

in the sub sample A, the overconfident dataset is the best in term of RMSE for longer training period, and also the worst in terms of bias reduction for shorter training period. We can conclude that the algorithm needs time to evaluate the the rights coefficient values, and since all the predictions give the same results for longer training period(3.25 and 3.26), we can infer that regression procedure can choose the best models, neglecting the informations coming from "poor" members only for shorter training period, while for longer training period is common the overfitting of the results, it means that the procedure neglect the informations coming from the predictors, and the SE *learns* only from the past observations.

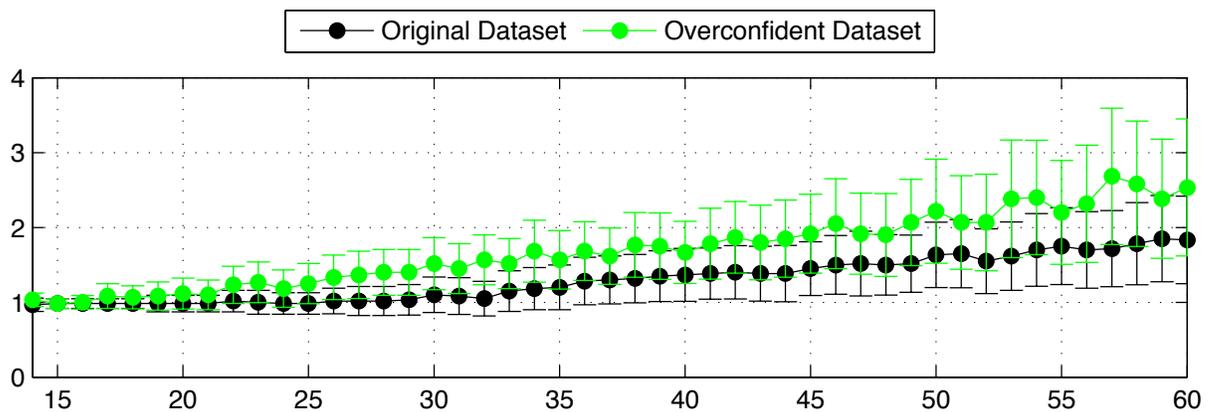


Figure 3.27: Bias Removed RMSE mean Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are Bias Removed RMSE mean Skill for the subsample A ,the green bars are the standard deviations.

3.2.4 Results in the MMSE experiment Forecast Mode

This experiment was aimed to check the performance of the superensemble algorithm in forecast mode, in order to evaluate a potential use in operational forecasting. Since we were interested in a simulation of forecasting, we didn't use long training period and study the results according to the dataset used and the lead time of the forecast. "The importance of this experiment we are using forecast a better performance in term of forecast means a provide a better product".

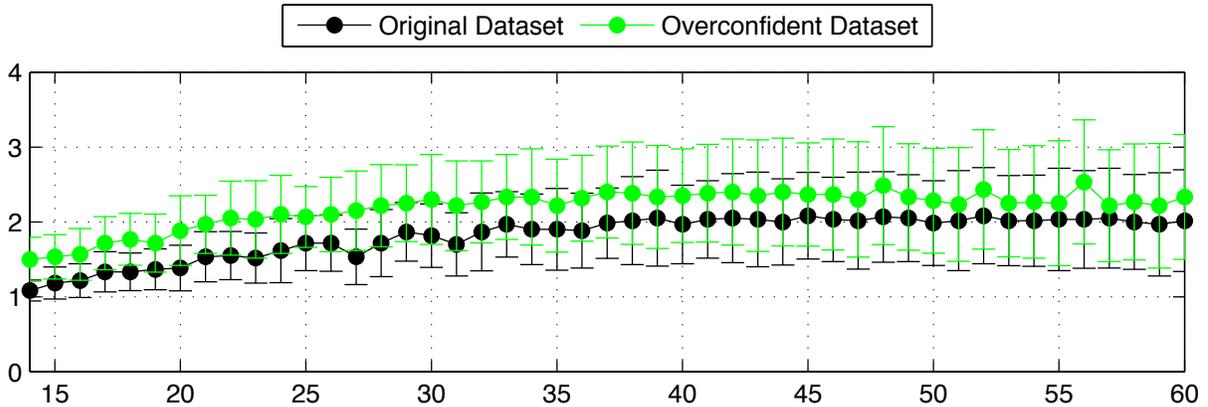


Figure 3.28: ACC mean Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bar are the standard deviations.

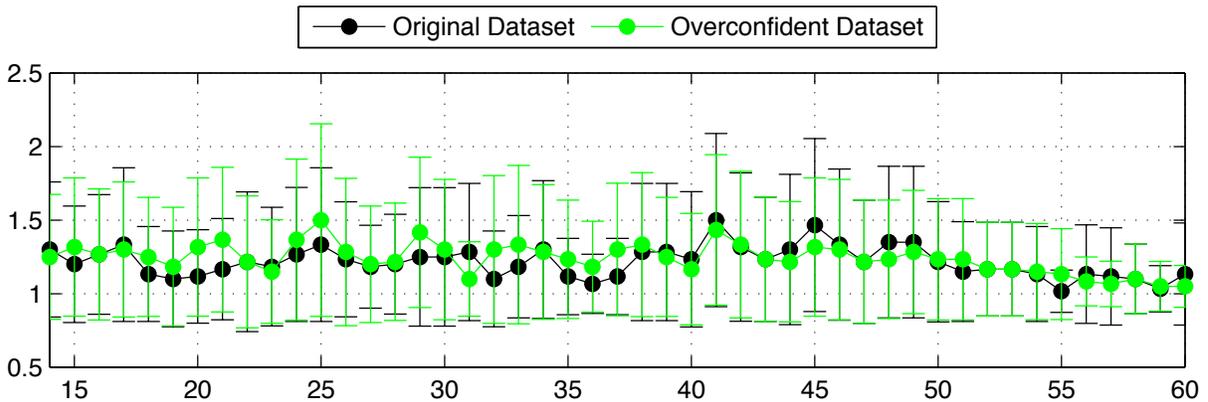


Figure 3.29: BIAS Skill of the Multi Model superensemble with 9 members against the training period lengths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bar are the standard deviations.

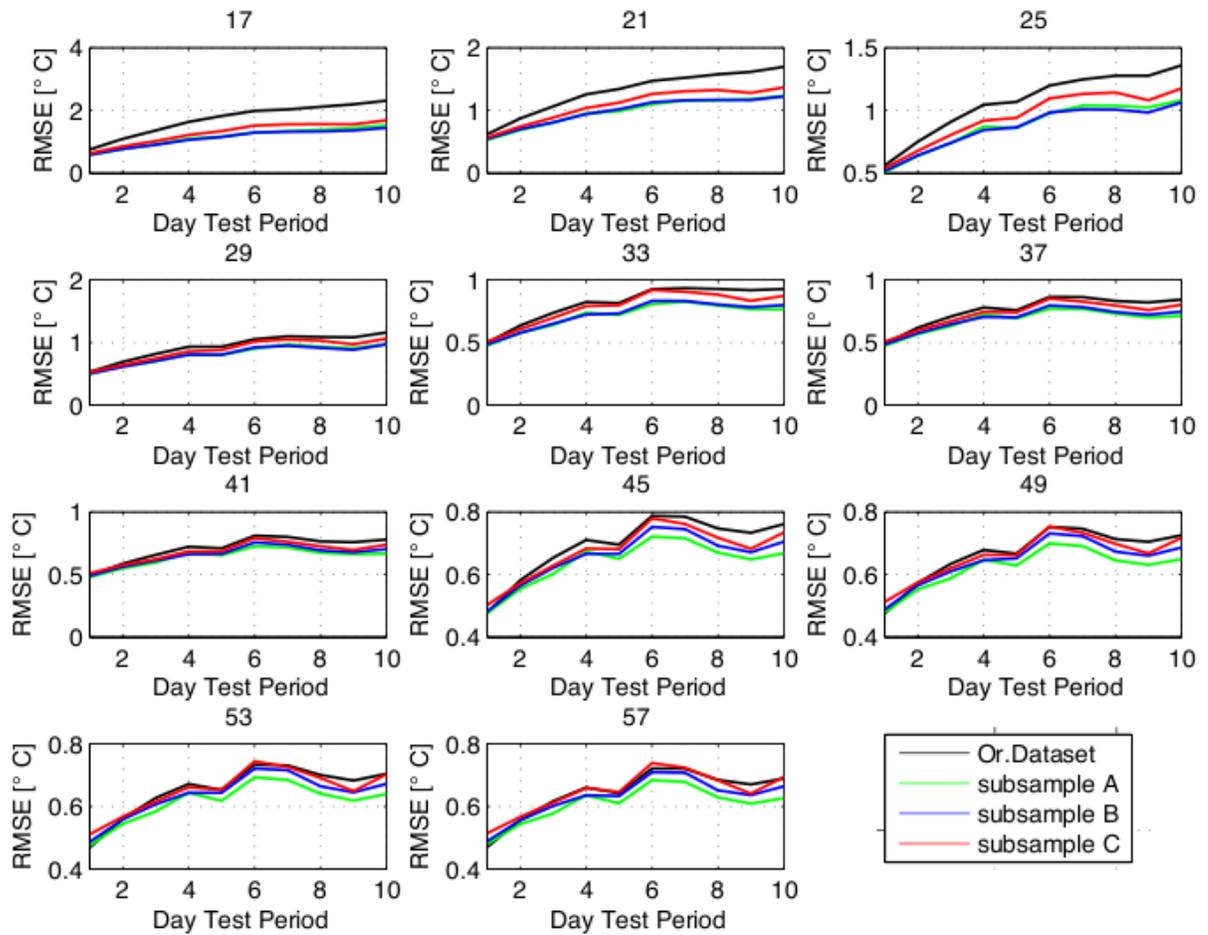


Figure 3.30: Domain average RMSE comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C

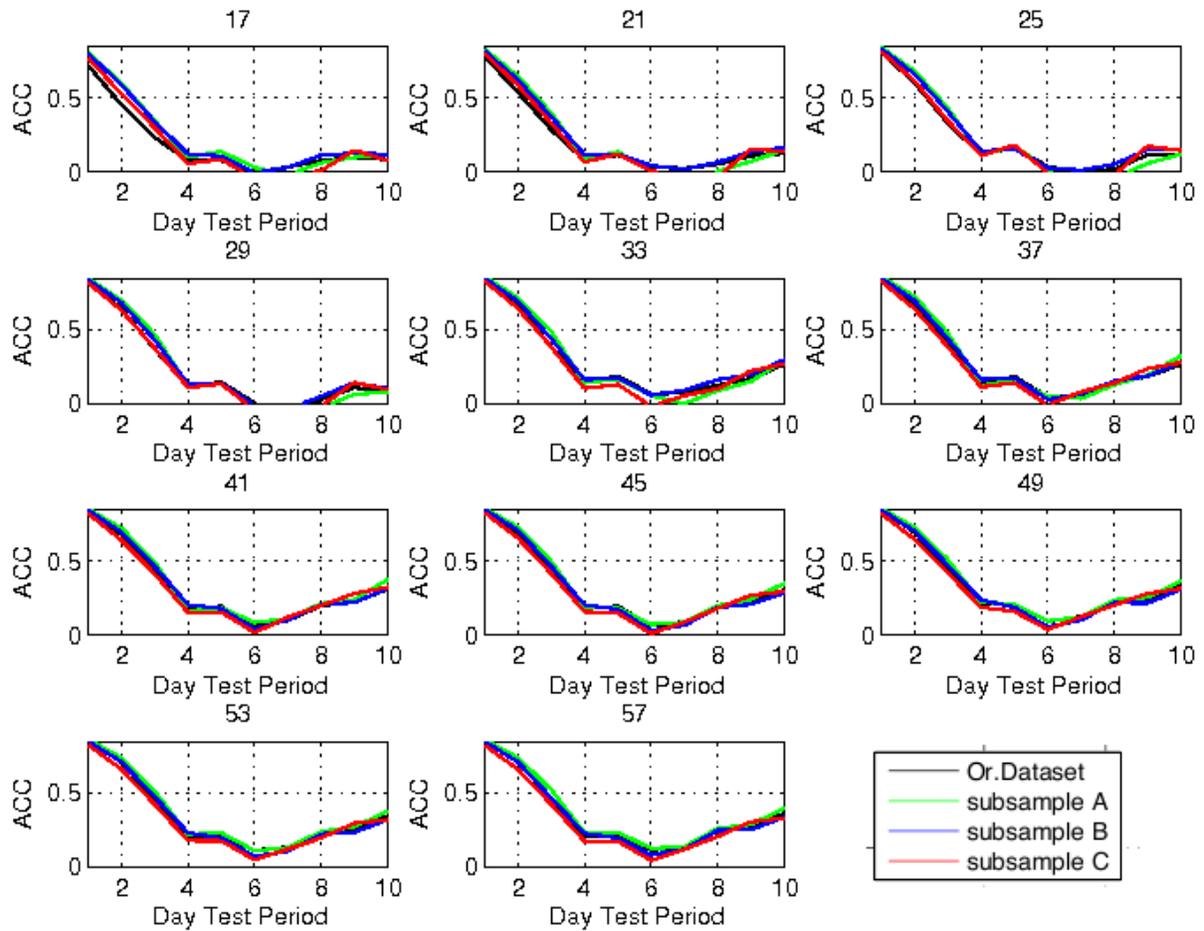


Figure 3.31: ACC comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C

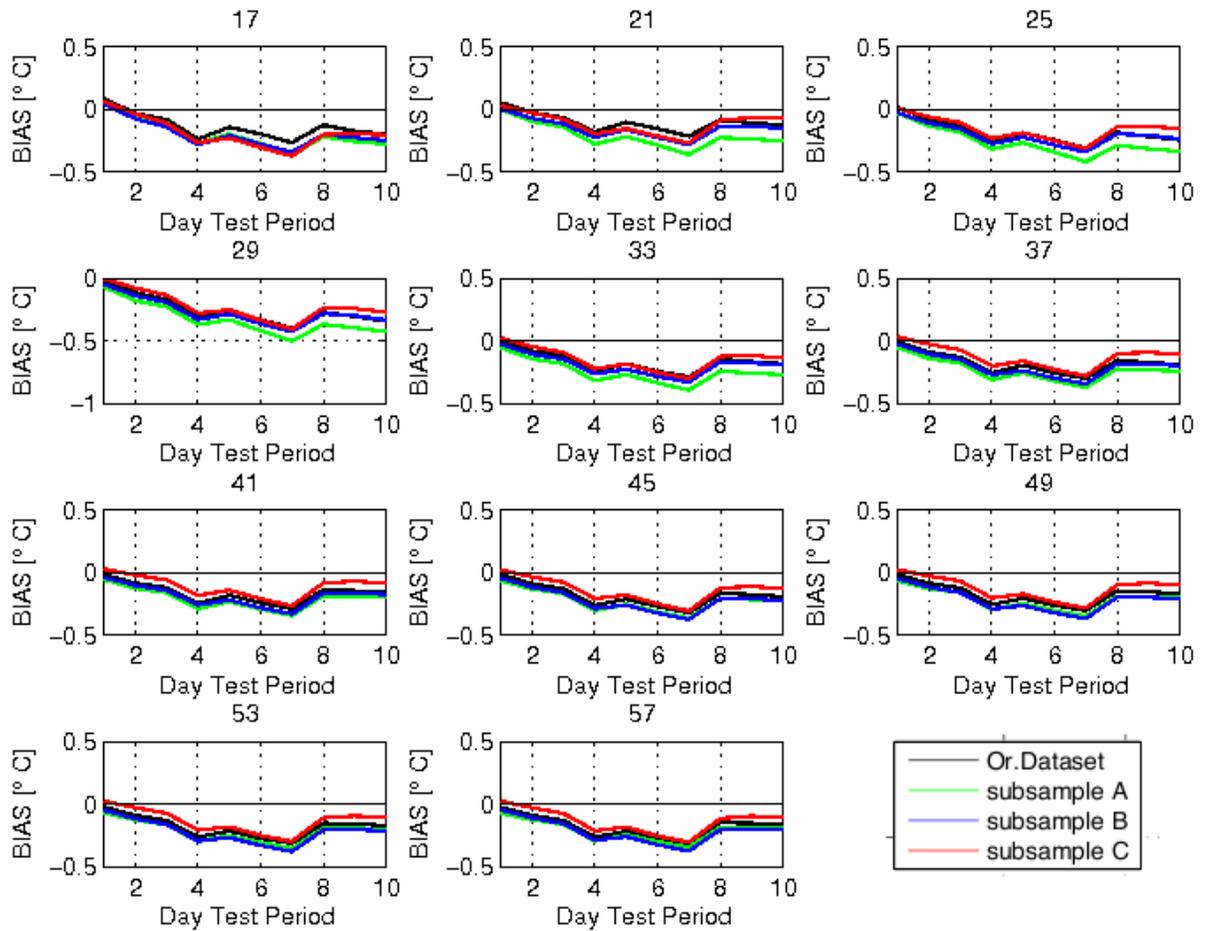


Figure 3.32: BIAS comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C

The training period changed according to the dataset from a minimum of 10 days (with the smaller dataset) to a maximum value of 25 days. We drew 3 sub-sample from the original dataset.

- 3 days of forecast (all the member in table)
- 5 days of forecast (all the member in table except HCMR forecasts)
- 9 days of forecast (all the member in table except HCMR forecasts and Met Office products)

Unfortunately in this case even if we had a good spread and good variability generated by the ensemble the resulting map are physicalness. In all analysed samples SE prediction could not outperform the other members. The noisy results, are reflected in the large values of the RMSE. Again we need to enlarge the training period to have better performance of the SE, but it was out the aim of this experiment. In order to have an homogeneous comparison we will show the results obtained with training period around 20 days, since for the largest dataset we need at least 15 days of training(see eq.3.12).

3.3 Summary

In this chapter we have shown the results obtained with the multi-model techniques. In first place we noted that if the length of the training period (in time units) is less than the ensemble size, the algorithm fails since it is not able to recognize best-worst participating models.

In the first experiment we have seen that lack of spread lead the algorithm fail. In the second experiment we showed that multi-model approach can outperform all the participating model. A careful examinations of the results show that robust optimal weights are difficult to calculate giving short training period samples, while too long training period lead to overfitted predictions. In the third experiment we have studied the sensitivity and the limit of this procedure. The critical point is the choice of the dataset, as long as the individual components are able to make

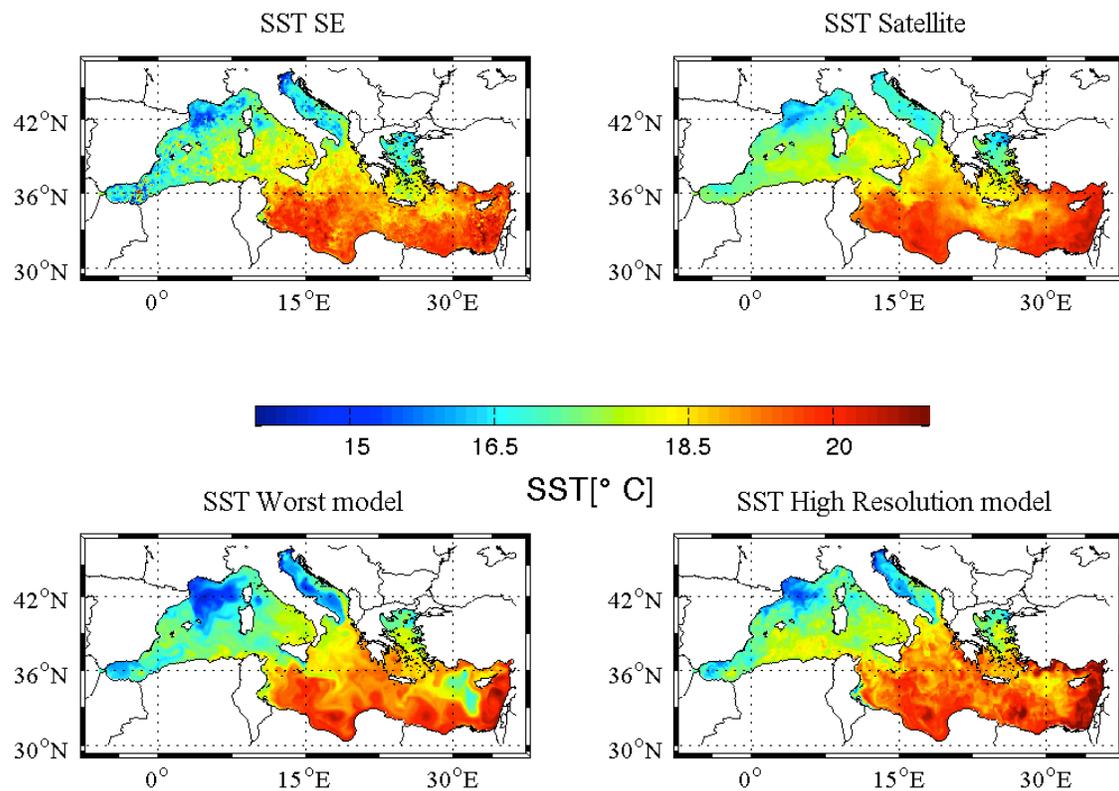


Figure 3.33: A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 10th of November 2008 reached with 21 days of training period, SE prediction (top left), lowest skill model (bottom left), SST Satellite derived (top right) and high resolution output (bottom right)

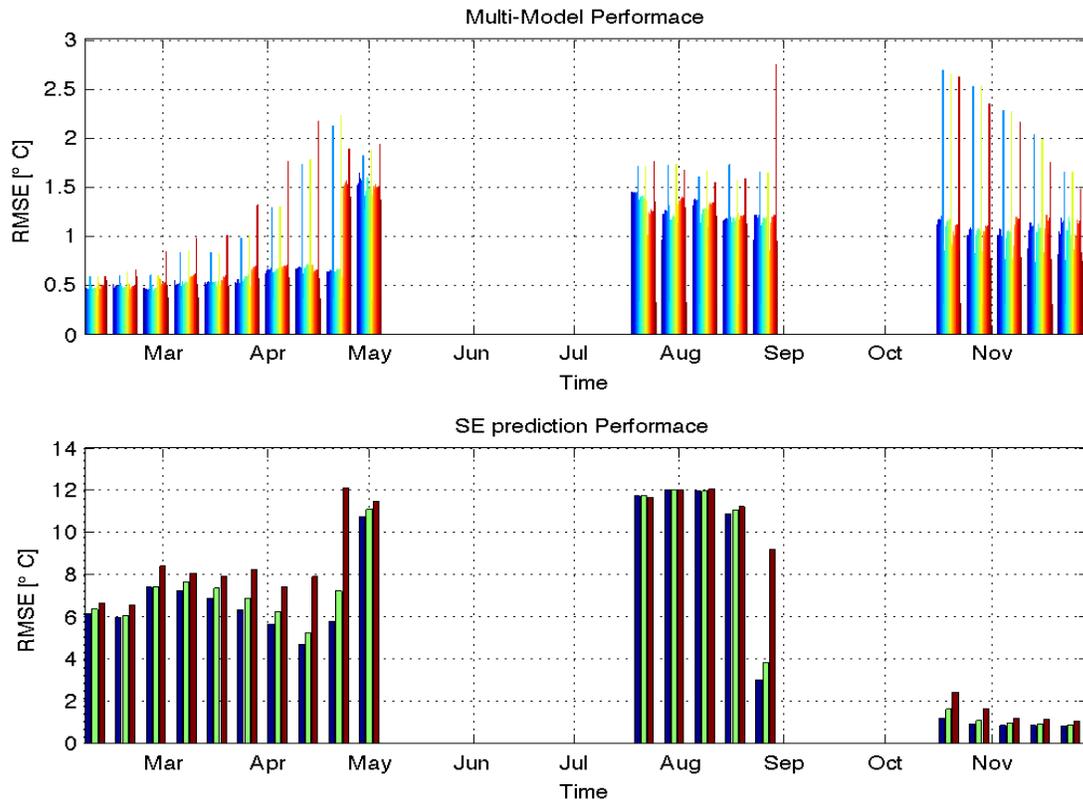


Figure 3.34: centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (21 days) during the year 2008 with forecast period three days.

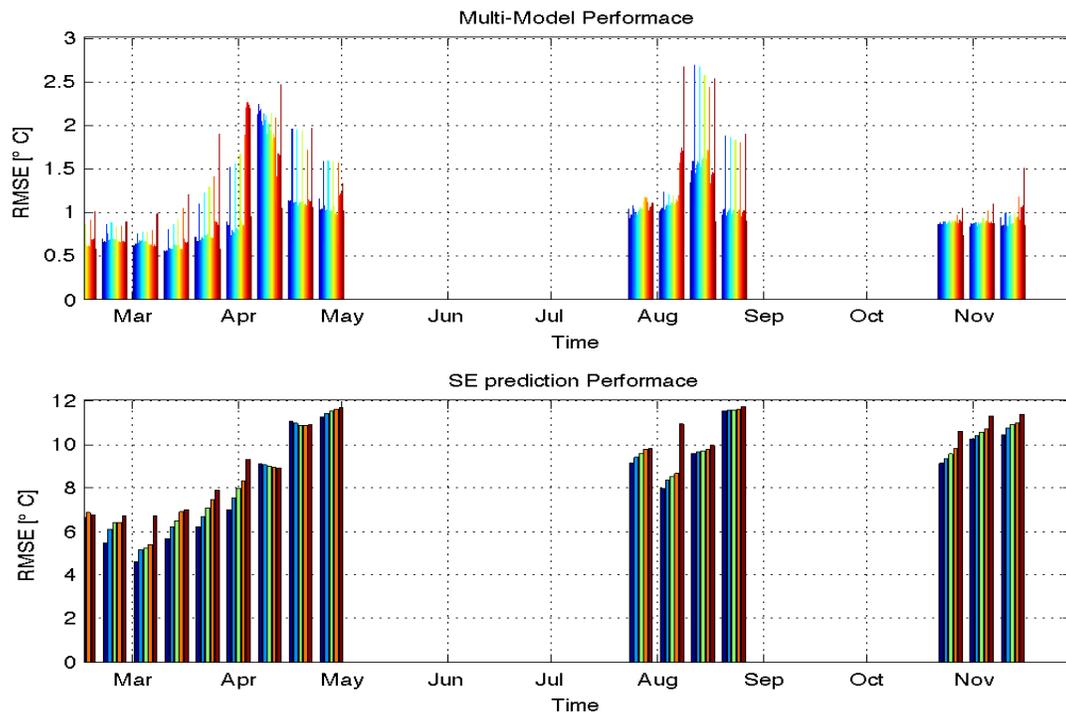


Figure 3.35: centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (20 days) during the year 2008 with forecast period five days.

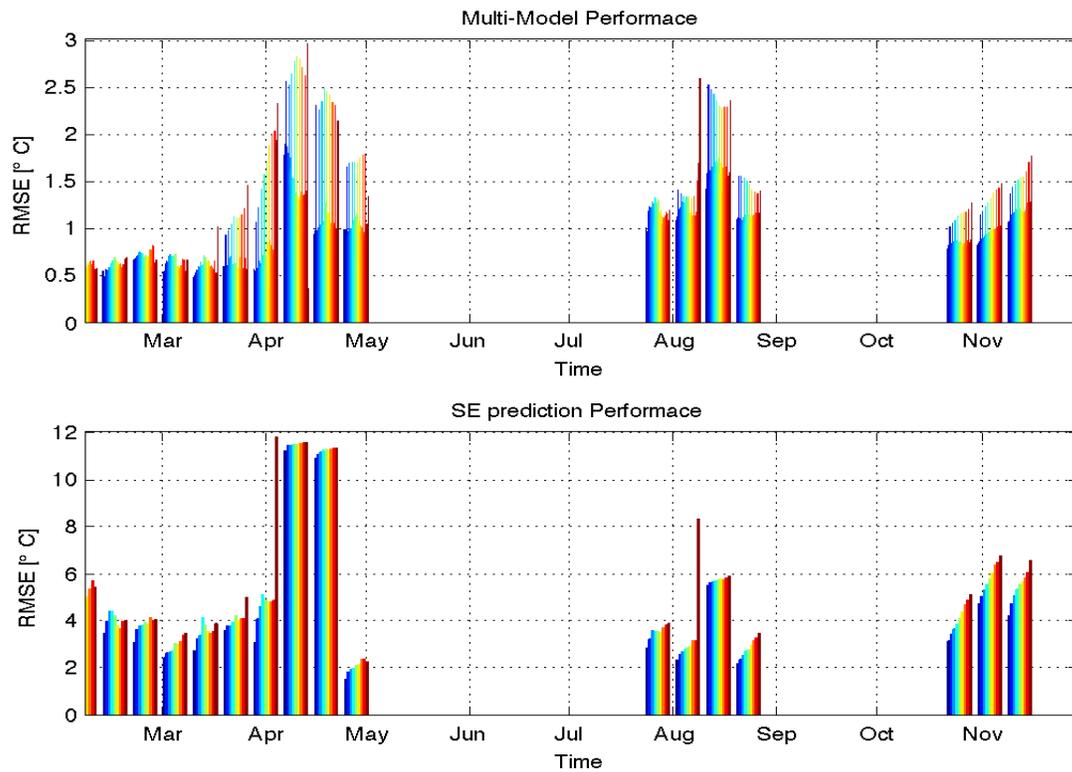


Figure 3.36: centred RMSE for the ensemble members (top) and the superensemble prediction (bottom panel) with the same training period (18 days) during the year 2008 with forecast period nine days.

a positive contributions to a relevant aspect of the prediction, the multimodel will benefit from this additional information. We have shown that better performance are achieved reducing the size of the ensemble, avoiding the worst participating model. In this scenario one model is superior to all the other component, so poorer model couldn't add information. Hence, the key of the success of multi-model concept lies in combining independent and skilful models. Furthermore if the quality assessment detect a single model that is always worst than the others, it should be excluded. Since for long training period all the dataset reach same performances means that there is an upper limit for the predictability, over that limit overfitting problem arise. Moreover in this last experiment we decided to avoid longer training period usages, since the unbiased estimator will be no more "unbiased" being affected by seasonal cycle effects. Finally we use a dataset of forecast members, in order to simulate a potential application in ocean forecasting. In this case SE approach fails at least for shorter training period involved. An area of future work would be to explore the usefulness of EOFs (Empirical Orthogonal Functions) to remove the degeneracy of ill-conditioning matrix when short training period are used (Krishnamurti *et al.* (2003))

Chapter 4

Improving the classical MMSE approach

4.1 Background

Empirical Orthogonal Functions(EOFs) were first used in meteorology in the late 1940s (Obukhov, 1947). This powerful method consisting in a space time-field decomposition into spatial patterns and associated time indices, contributed much in advancing our knowledge of the atmosphere and oceans. The EOFs method is in essence an exploratory (I.E. non model orientated) tool, able to find the spatial patterns of variability, their time variations and gives a measure of the importance of each pattern. The probabilistic prediction obtained from the regression method illustrated in the previous chapter could be not reliable, since the spatial correlation of the field regressed on the observations is not taken into account. Kharin & Zwiers (2002) suggest that that poorer performance of combined multi-model predictions through multiple linear regression is due to overfitting or, in other words, biased estimates of the coefficients. Because co-linearity of predictors may explain part of the failure, principal component regression (von Storch H., 1995.) offers an alternative way of performing the regression with linearly uncorrelated variables.

4.2 Formulation and computation of EOFs

Given any space and time dependent field, EOF analysis finds a set of orthogonal spatial patterns along with a set of associated uncorrelated time series or principal components (PCs). The geometrical constraint characterizing EOFs and PCs can be very useful since the covariance matrix of any subset of retained PCs is always diagonal. Here will be presented a short description of how to obtain EOFs, giving the linking between their development and the application in our case. Once the anomaly data matrix (b_{ij}) is determined:

$$b_{ij} = \begin{bmatrix} (F_{11} - \overline{F_1}) & (F_{12} - \overline{F_1}) & (F_{13} - \overline{F_1}) & \dots & (F_{1n} - \overline{F_1}) \\ (F_{21} - \overline{F_2}) & (F_{22} - \overline{F_2}) & (F_{23} - \overline{F_2}) & \dots & (F_{2n} - \overline{F_2}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (F_{m1} - \overline{F_m}) & (F_{m2} - \overline{F_m}) & (F_{m3} - \overline{F_m}) & \dots & (F_{mn} - \overline{F_m}) \end{bmatrix} \quad \overline{F_m} = \frac{1}{N} \sum_{i=1}^N F_{mi} \quad (4.1)$$

we can determine the covariance matrix of our original data $\Gamma = b' \cdot b$ (for definition check the eq. 3.7).

The EOFs are obtained as the solution of the classical eigenvalues problem

$$\Gamma \vec{x} = \lambda^2 \vec{x}; \quad (4.2)$$

The k -th EOF is simply the k -th eigenvector x_k of Γ . The corresponding eigenvalue λ_k^2 is:

$$\lambda_k^2 = \vec{x}'_k \cdot \Gamma \cdot \vec{x}_k = Var(b\vec{x}_k) \quad (4.3)$$

and hence λ provide a measure of the variance of the data accounted for the direction defined by \vec{x}_k . After finding the eigen elements of the sample covariance matrix the eigenvalues are normally sorted in decreasing order. The projection of the anomaly field b onto the k -th EOF (\vec{x}_k) is the k -th Principal Component, whose elements

$a_k = b\vec{x}_k$ are given by:

$$\sum_{j=1}^p x_{t,j} \cdot u_{k,j} \quad (4.4)$$

It can be shown that is possible to use a powerful tool for very non square matrix, the SVD (Singular Value Decomposition) method, which finds the EOF/PC without solving the complete eigenvalue problem, but simply finding the singular value of the covariance matrix and projecting the anomaly on those singular values. So the initial field b can be expressed in term of the EOF/PC Analysis:

$$b = \sum_{j=1}^p \lambda_k a_k u'_k \quad (4.5)$$

The dimensionality of the initial data can be reduced by truncating the sum at a index M which is less than the rank of the matrix b . However there is no universal rule for truncation, and the choice of the degree of EOF is somehow arbitrary. Usually the truncation order is obtained by fixing the amount of the represented variance, a common value is 90% of variance, hence it is chosen a set of n leading EOFs that can explain at least that amount of variance. The non degeneracy of the eigenspectrum is an important property and can be vary useful in interpreting EOFs. Real physical identification in the pattern retrieved by EOF is not always obvious for their orthogonal nature. Orthogonality translates into the fact that typical patterns appear in secondary (higher order) EOF. Sometimes the resulting optimized modes are difficult to interpret physically, either because the real relation is localized and the EOF, due to new reference system, are spreading it creating artificial non local relations, or because the EOF are so close in terms of eigenvalue separation, that the numerical techniques cannot really distinguish between them. The most common situation in which malfunctioning can arise is when data represent local variances. In this case the EOF will try to fit globally the domain under consideration, with as few modes as possible, generating first EOF(low order modes) with very large structures(A Navarra, 2010).

As first guess we applied the EOF analysis only to the weights mask, in order to filter the noise due to the overfitting procedure involved in the regression of the

anomalies during the long training period. The analysis of the eigenvalues power spectrum yields to a very smooth spectrum, it means that there are not only small modes of variability. In order to remove the degeneracy of the power spectrum we

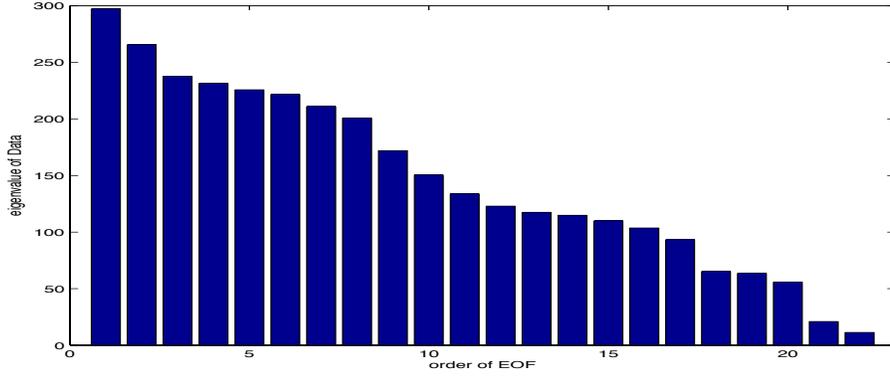


Figure 4.1: *Eigenvalues spectrum for the weights mask in the BHM dataset*

decided to apply the EOF analysis to the multimodel anomaly fields and project the observation anomaly field on the EOF too. Step back to equation 3.10,

$$\Gamma \cdot \vec{a} = \vec{\phi} \quad (4.6)$$

We can substitute in Γ its eigenvalues and solve the following system

$$\vec{x}\lambda\vec{a} = \vec{x}\vec{\phi} \quad (4.7)$$

In this way the regression coefficients vector can be evaluated as:

$$\vec{a} = (\vec{x}\lambda)^{-1}\vec{x}\vec{\phi} \quad (4.8)$$

which is the product of the inverse of the filtered (projected on the EOFs) covariance matrix and the Principal Components of the observations anomaly field. The principal component analysis was performed with each sample in order to reduce the spatial dimension to a number smaller than the original number of grid points. Only the component explaining the 90% of the variance have been retained, which in our

case corresponds to take only the first seven-nine EOFs (according to training period length involved) of the anomaly field, see figure 4.2

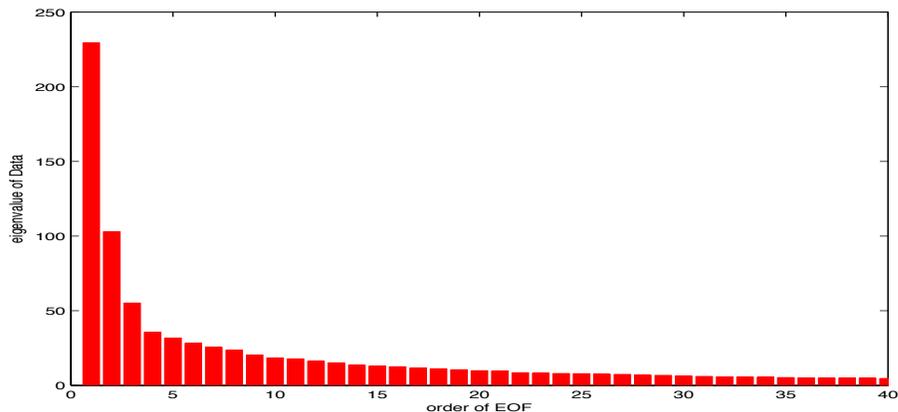


Figure 4.2: *Eigenvalues spectrum for the weights mask in the BHM dataset*

4.3 Results with EOFs analysis

4.3.1 Multiphysics Experiment

We performed the EOFs analysis on the first experiment dataset, the perturbed physic ensemble. The new computation of the superensemble show that EOFs/PC Analysis was able to reduce the degeneracy of covariance matrix due to the high correlation on the ensemble model members (as shown in fig 4.3), nevertheless SE is not the best performer of the ensemble fig.4.4, since the brown bar is taller than the other centred RMSE

4.3.2 BHM experiment

The superensemble was constructed merging the 22 output of MFS BHM-winds described in . The weights have been obtained by the regression of the projections of true observator and multimodel members and retaining the first nine EOFs. The new results are quite appealing. We were able to reduce the noise in the map (see 4.5), and

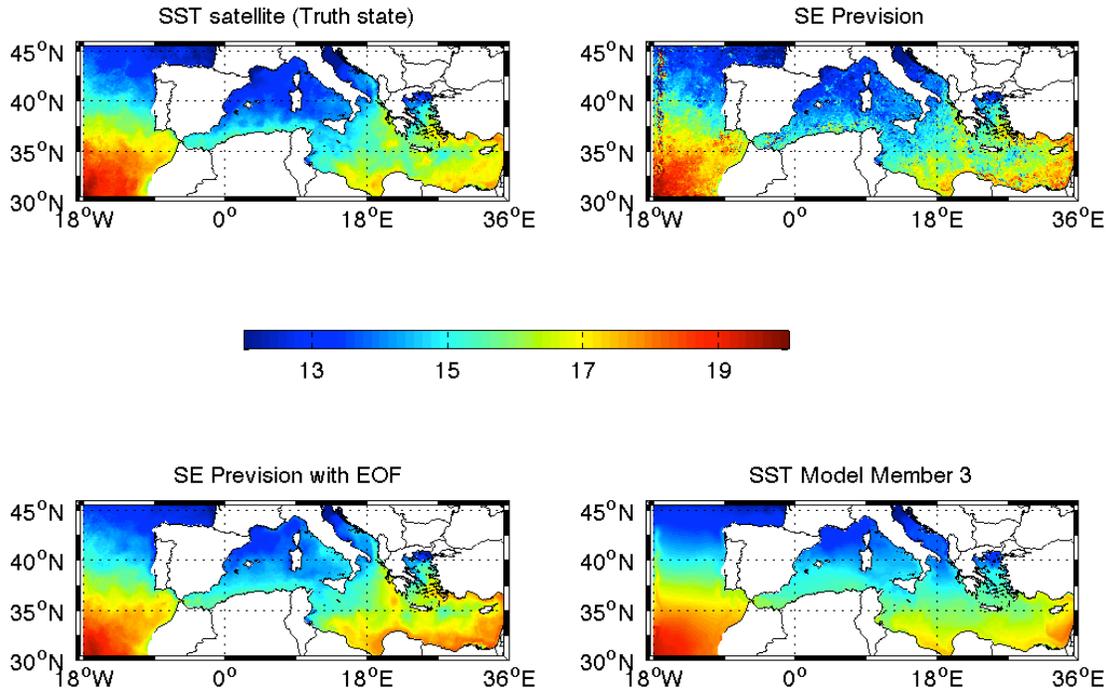


Figure 4.3: *first prediction day of the classical SE (top right) SST satellite derived (top left), SE EOFs based (bottom left) and one of the model of the ensemble*

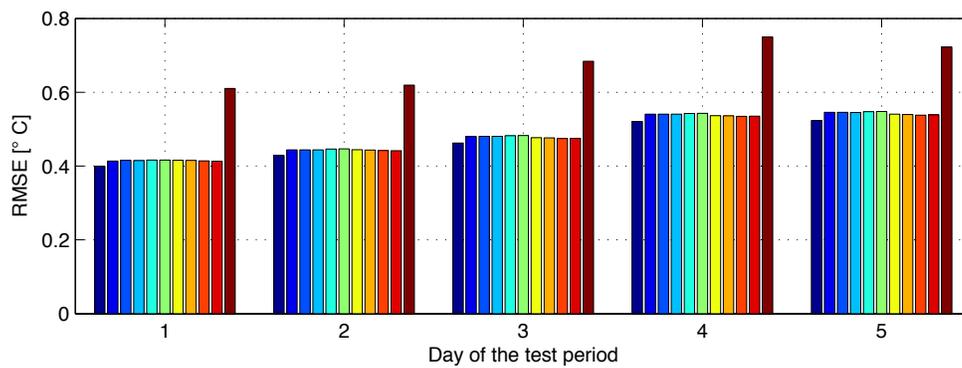


Figure 4.4: *RMSE between truth observation and MP members (coloured bar) and SE EOFs based (brown bar) for SST field, during the five days of the test period*

until the 20th days of test period the SE EOFs-based gives the best results in terms of centred RMSE(see fig.4.6). Only at the end of test period its performance became worst due to an increase of the bias(see fig. 4.9). In term of anomaly correlation coefficient (fig.4.8)superensemble prediction was the best model too(compared to the performance of the classical superensemble).

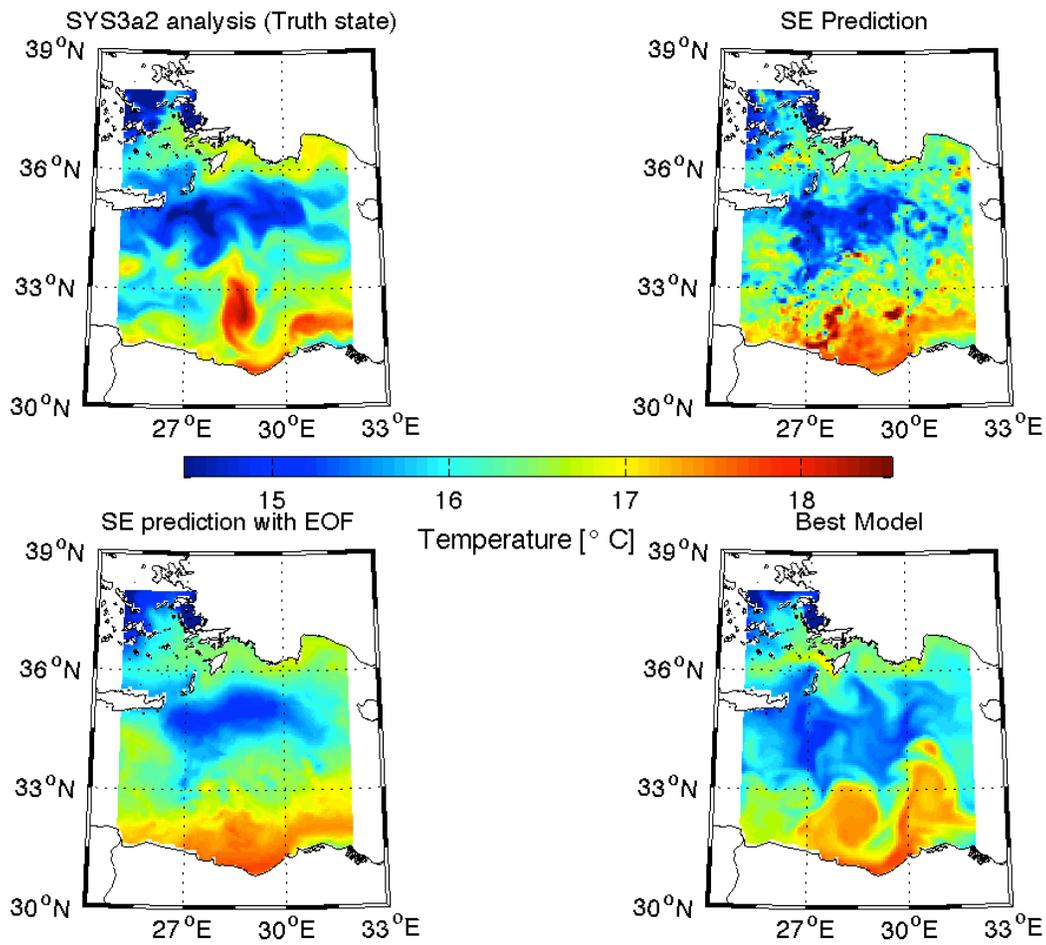


Figure 4.5: A typical example of a reconstructed map of SST, on day three over Rhode Gyre region, valid the 3 March 2008: analysis (top left), SE EOFs based(bottom left), SuperEnsemble classic (top right) and best participating model (bottom right).

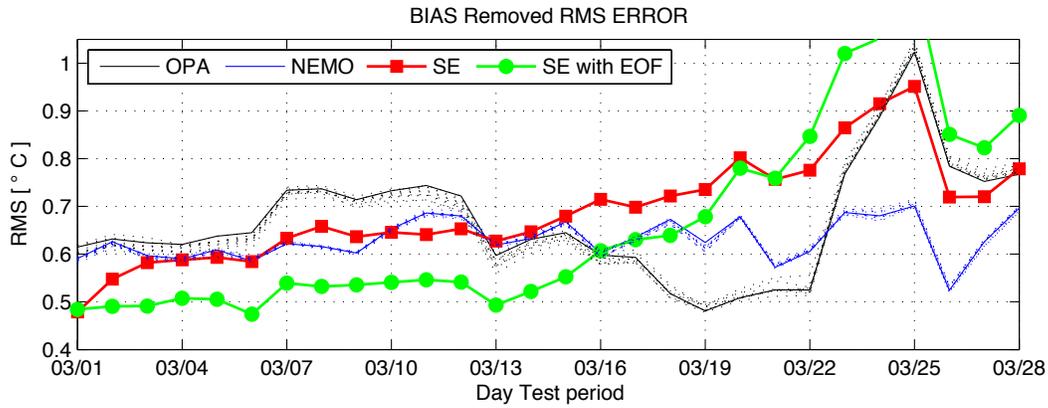


Figure 4.6: Domain averaged RMSE between the OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period

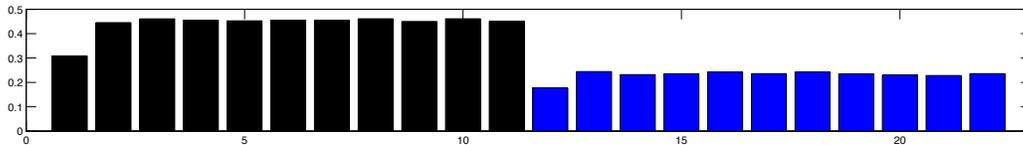


Figure 4.7: MESS for the 22 BHM MM members and SE EOFs for the parameter SST over Rhode Gyre region during the test period (March 2008). Black bars represent the gain respect OPA predictions, and blue bars the gain respect NEMO

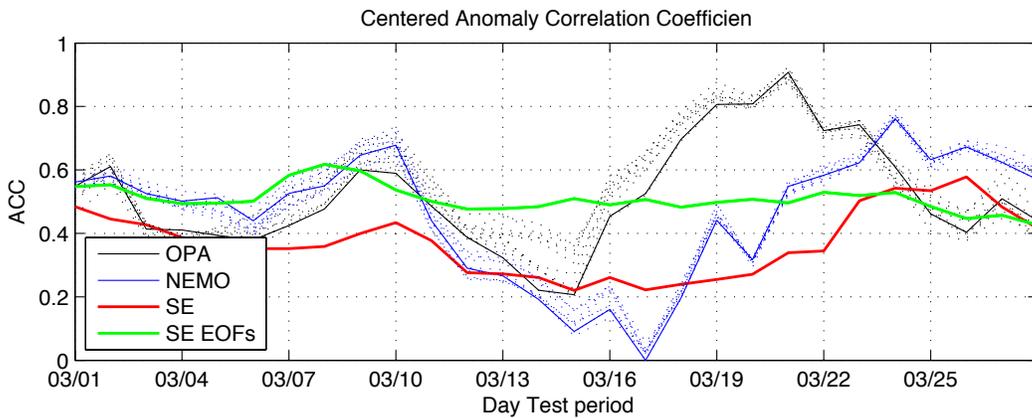


Figure 4.8: ACC, OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period

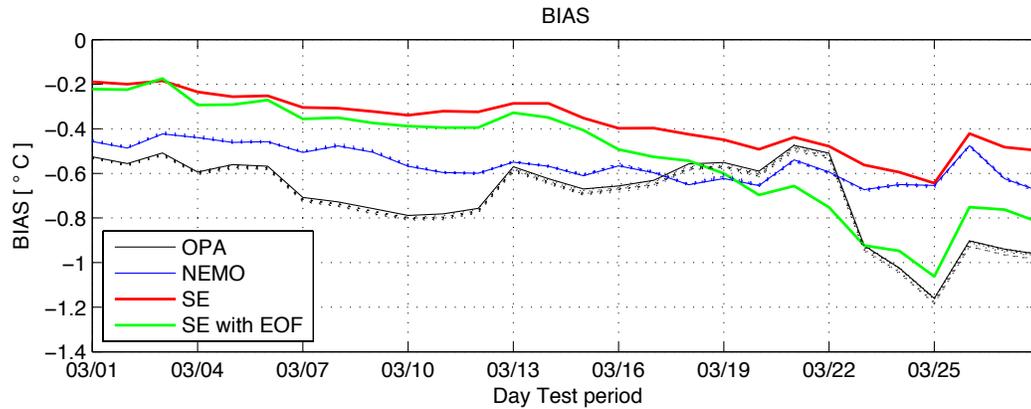


Figure 4.9: *BIAS*, OPA members (black), NEMO members (blue), SuperEnsemble prediction (red) and SE prediction EOFs based (green) during the test period

4.3.3 Multimodel Multiphysics experiment Analysis mode

In this section we present the results on the multimodel Multi-physics experiments, a collection of analyses from different research institutes and four perturbed physics members, covering the year 2008. We performed again all the experiments carried out in previous chapter and in this case we reached different results. Despite the fact that SE prediction EOFs based seems better at first sight, see fig.4.10 and fig.4.11, the performance of SE EOFs based, got worst increasing the lead time of the test period. Figure 4.12 shows on the right side the "classical" scores for the fifth day of the test period during year 2008 for the original 10 members dataset, of the centred RMSE, bias and ACC respectively (top, middle and bottom) and the correspondently results with the superensemble with EOF employment to evaluate the weights, for a training period of 10 days.

Figure 4.13 is the same 4.12, but obtained with longer training period, 30 days. Since in the latter case we have a bigger bias, consequently we have a bigger RMSE. It is important to note that during August since we have the minimum of the bias, RMSE is smaller if SE is evaluated by EOFs analysis (fig.4.12). ACC plots show clearly a worsening of performance of SE EOF based in terms of ACC, nevertheless the ACC in the classical approach is affected by overfitting. It is clear from those results that the new method doesn't bring the desired improvements. A careful analysis of the weights mask shows smaller

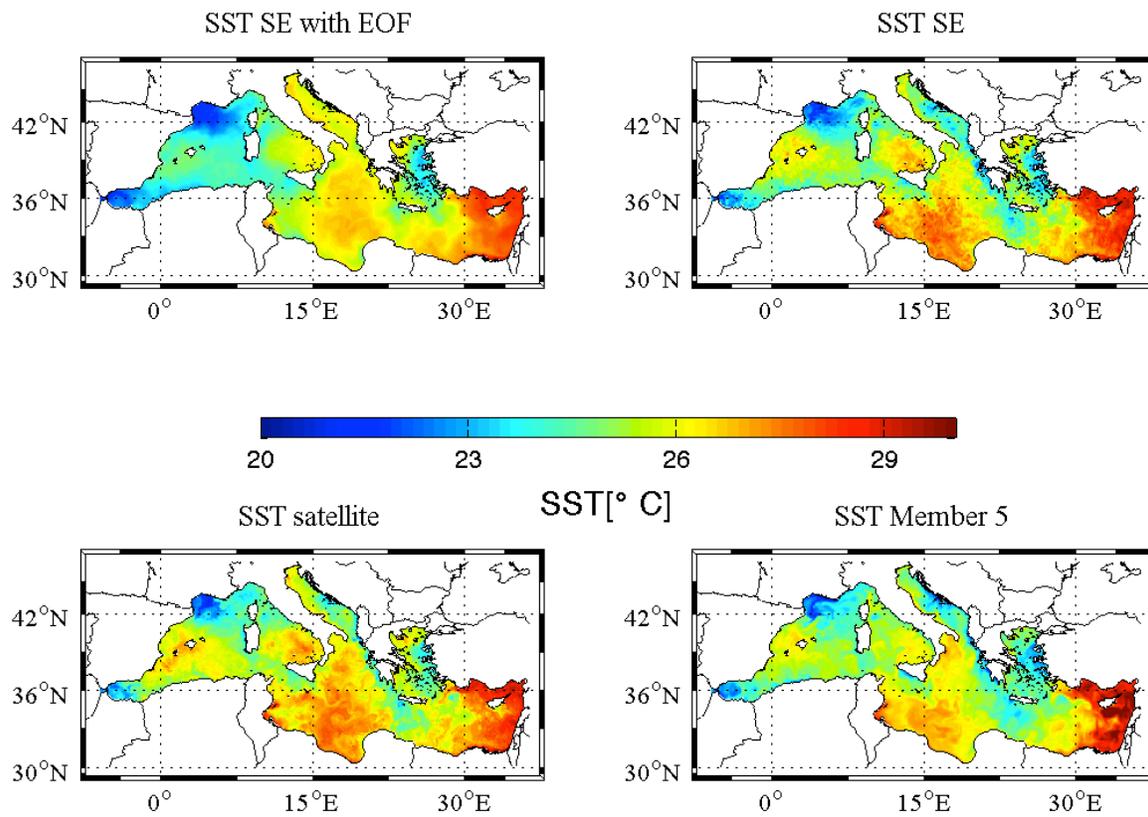


Figure 4.10: A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1st of July 2008 reached with 55 days of training period, SE prediction (top left), SE EOFs based prediction (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)

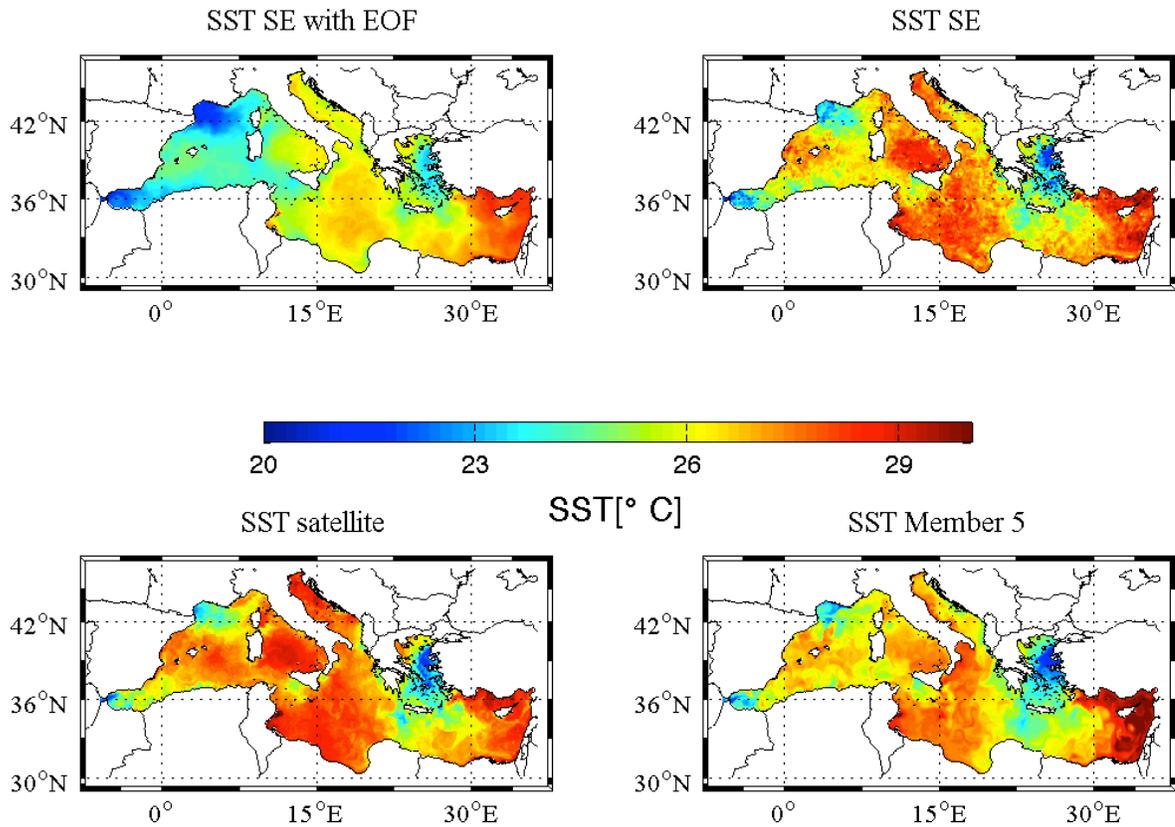


Figure 4.11: A typical example of prediction for the last(10th) day of test period over the Mediterranean area, valid on the 1st of July 2008 reached with 55 days of training period, SE prediction (top left), SE EOFs based prediction (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)

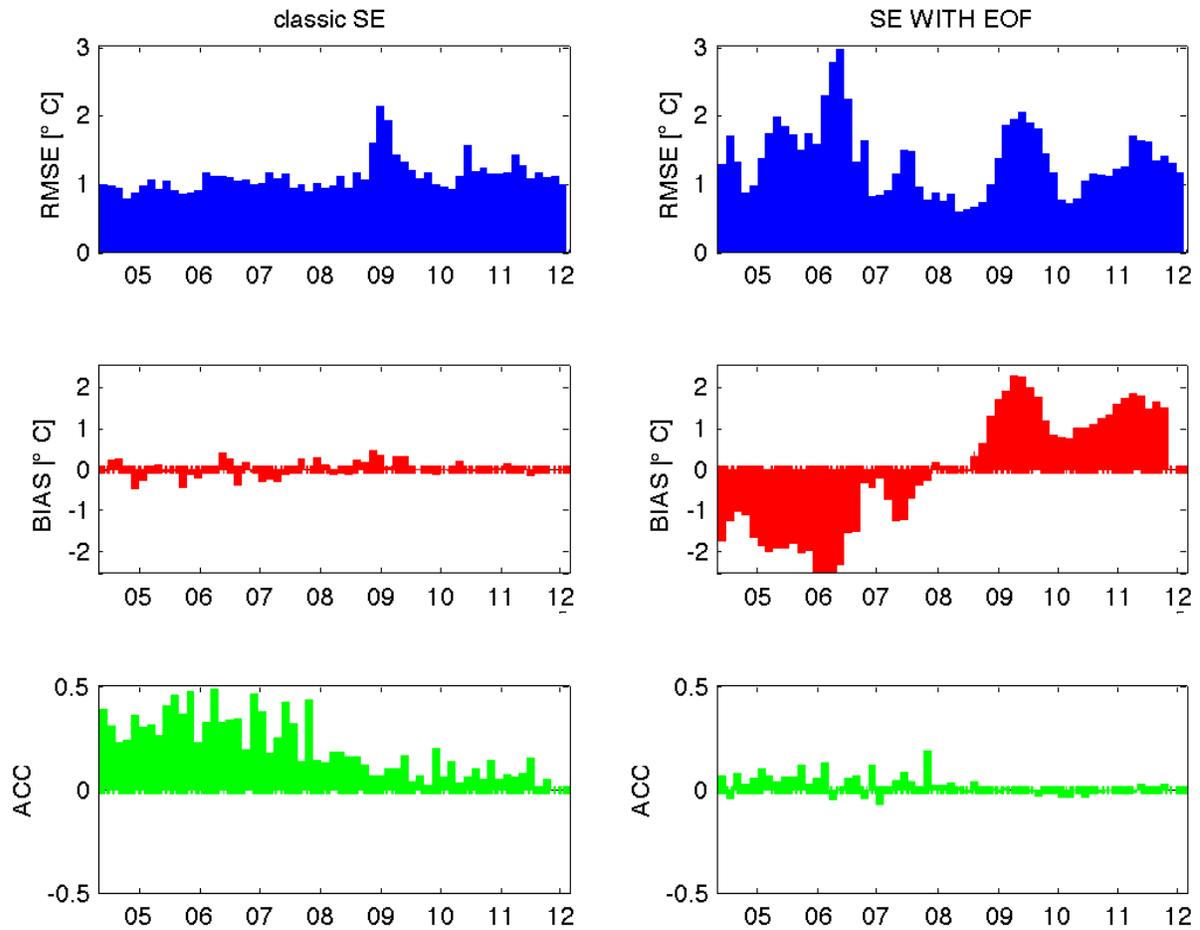


Figure 4.12: Performances of SE obtained by 10 days of training period, during year 2008, left panels the classical approach, right side the SE EOFs based approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.

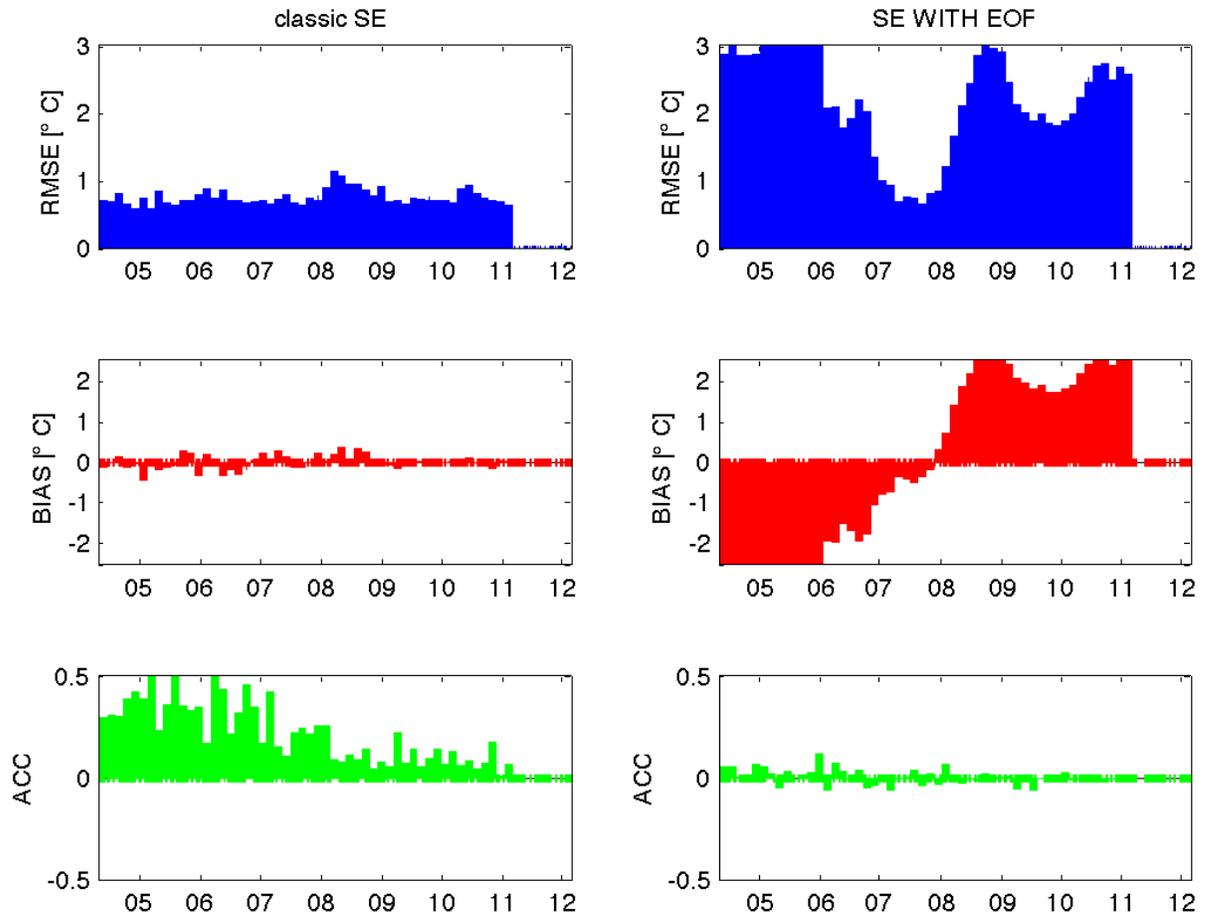


Figure 4.13: Performances of SE obtained by 30 days of training period during year 2008, left panels the classical approach, right side the SE EOFs based approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.

values for the coefficients than in the classical approach (see fig.4.14). Referring to Krishnamurti *et al.* (2003) we try to reproduce a similar set up, using seven members (the subsample A, overconfident dataset) and trained them for 100 days. Krishnamurti in his experiment used a laplacian based superensemble assessing that in that way he could extract some extra skill from gradients and laplacian, avoiding the degeneracy of covariance matrix.

As we can see with the classical approach (top of 4.14) we got greater value (the magnitude order of unit and more), while in the filtered approach the weights exhibit a distribution of positive and negative fractional values. So we can states that:

$$|w_{eof}^m (F_{mn} - \overline{F_m})| < |w^m (F_{mn} - \overline{F_m})| \quad (4.9)$$

EOFs gives value of weights closer to that papers (see fig.4.15), but in our case the unbiased estimator, \overline{O} in no more *unbiased* since it takes in account of the seasonal cycle. Since we were trying to predict during a different test period, the smaller value of the weights are no more able to remove the actual bias during the test period.

It is clear from the inequality (4.9) that smaller weights are not able to reduce the bias in season cycle affected "Unbiased Estimator" if the variability (the term $(F_{mn} - \overline{F_m})$) evaluated by multimodel is the same order of the variability of observations. Hence SE fails due to a wrong BIAS reduction, at least if the superensemble involves fractional weights. As a result SE will be affected by cold bias during spring-summer time and warm bias during fall-winter time. The key point is the kind of field we want to reproduce, since we are considering a faster varying field than Krishnamurti, EOFs based SE evaluated with this kind of "unbiased estimator" was not a good way to perform the regression.

4.3.4 Multimodel Multiphysics experiment Forecast mode

In this section we will show some results obtained using the EOFs based SE in forecast experiments. In this case we didn't find a clearly correlation on training period or dataset, since results changed with any specific correlation for each experiment. For shorter training period we got the irretrievable results due to the satisfaction

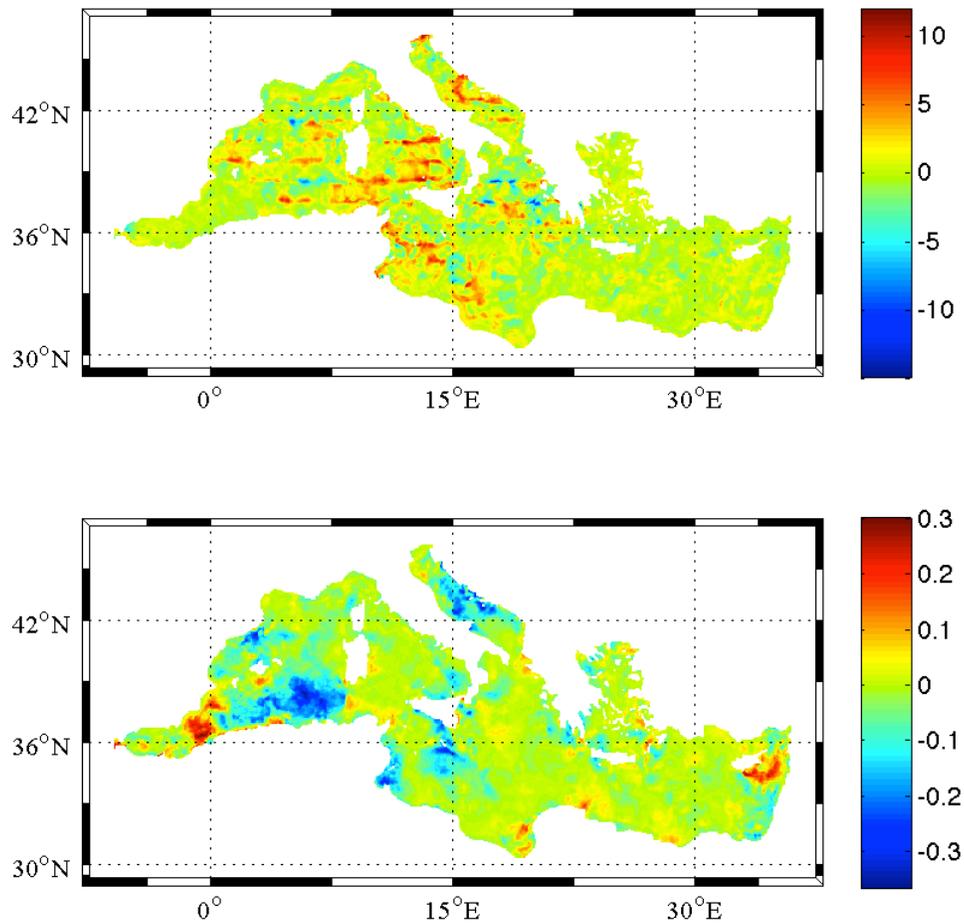


Figure 4.14: Geographical distribution of the regression coefficients for the first model member, as the simple anomaly field (top) and for the filtered anomaly field (bottom) after 100 days of training period, note the different colorbars

1088

MONTHLY WEATHER REVIEW

VOLUME 131

NH – Regression Coefficients for Day 6 Forecast

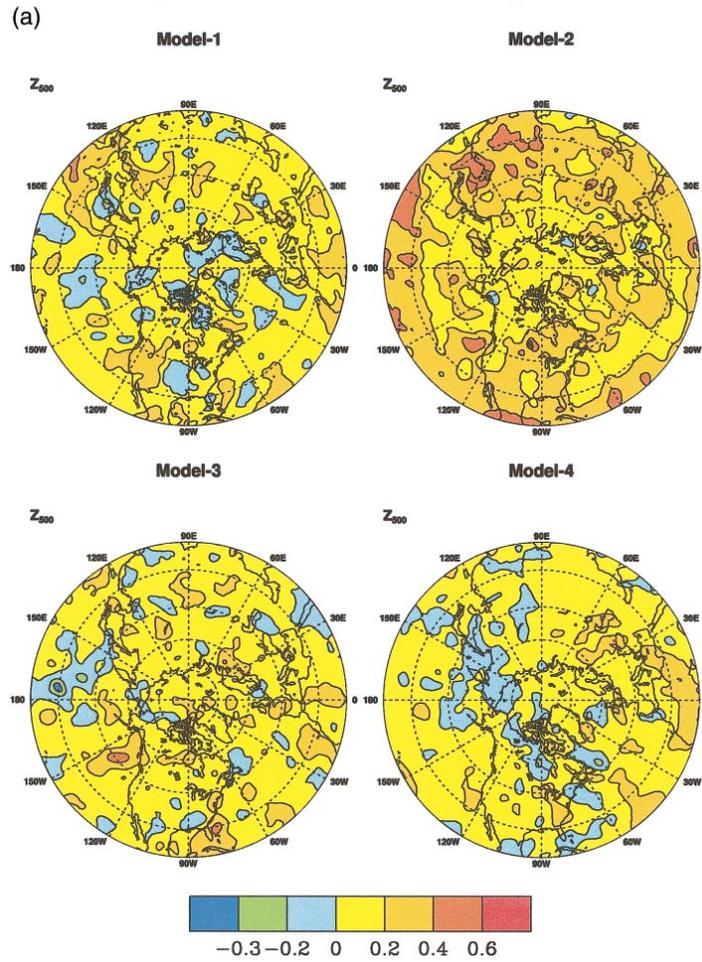


FIG. 5. Geographical distribution of statistical weights for different member models: (a) Northern Hemisphere and (b) Southern Hemisphere. Color scale of the fractional weights is shown at the bottom.

Figure 4.15: *Statistical coefficient from Krishnamurti et al. (2003)*

of inequality 3.12, and showed in figure 4.16, here the brown bar, showing the performance of SE EOFs based in term of centred RMSE are the worst of the entire dataset for all the experiments trained. Enlarging the training period in some cases SE could outperform all ensemble member, but at the end of the test period 4.17. We noted that the rate of growth of multi-model members and SE are different, the first show a faster growth rate than the latter. For this reason, SE can outperforms all the members at the end of the forecast period. It must be noted that when the forecast consist in a unusual situation, it means the forecast forcing have different performance than usual, for example showing a flat histogram of the errors (it means RMSE almost constant during the test period) the superensemble approach fails(4.18). Hence as Krishnamurti *et al.* (2003) noted, training (and the per-

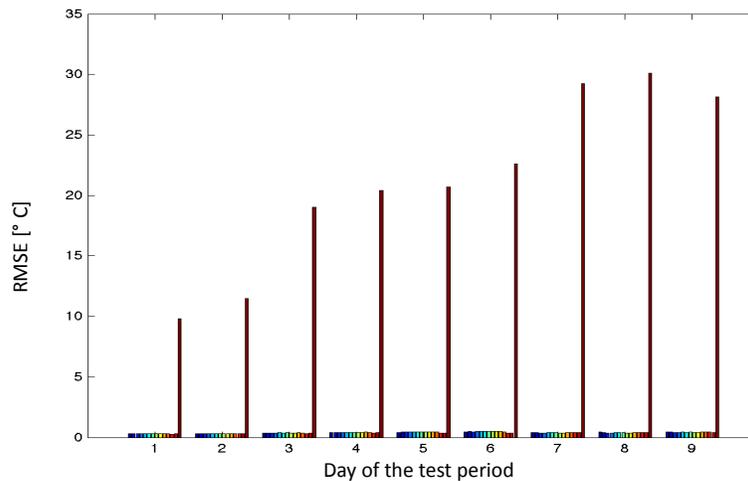


Figure 4.16: centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the same training period (9 days) during the year 2008 with forecast period nine days.

formances of Multimodel member) is a major component of the forecast initiative. Since the SE is trained is a higher variability field, in which the performances of the MM often change, SE can not outperform the originating dataset.

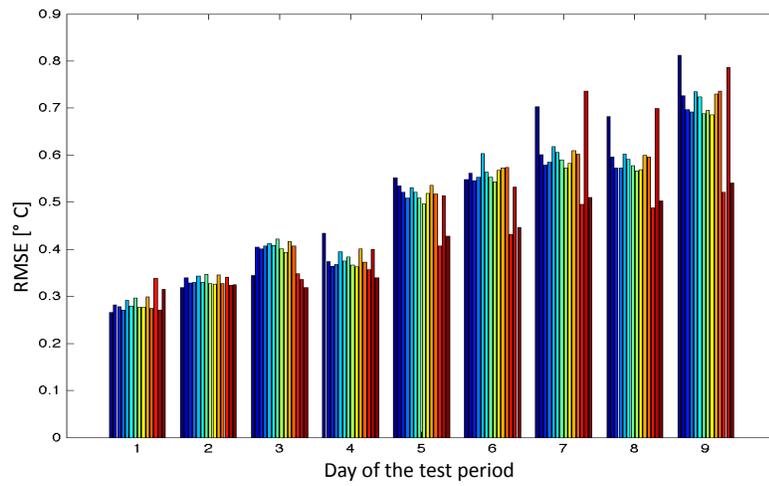


Figure 4.17: centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the training period of 18 days, for the 1st of March 2008, with forecast period nine days

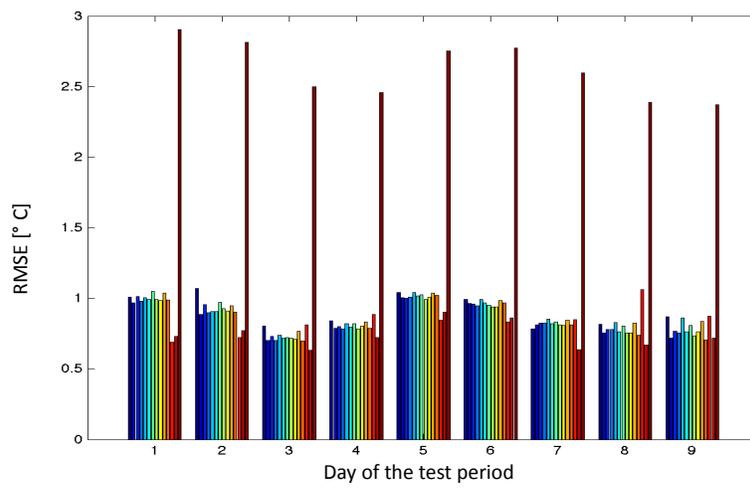


Figure 4.18: centred RMSE for the ensemble members and the superensemble prediction (brown bar) with the training period of 27 days, for the 1 November 2008, with forecast period nine days

Conclusions

Weather , ocean and climate prediction are subjected to many sources of error:

- Uncertainties in model initialization, due, for example due to incomplete data coverage, measurement error and boundary conditions.
- Uncertainties and error in the model itself, because of some physical processes are not fully understood, or also due to the parametrization of physical processes computational limitations.

Multi-model is pragmatic a approach pursued to obtain a a first crude estimate of the range of uncertainties induced by model error , while superensemble (SE) is a weighted mean of different model outputs in which the weights are evaluated by multiple linear regression between the "*truth observator*" and model outputs , which seems to be affected by a reduced model error. Previous employments of this technique (Krishnamurti *et al.*, 2000 and Kharin & Zwiers, 2002) have shown slightly contradictory results.

In this thesis we examined the different conditions under which SE seems to outperforms the generating ensemble. First we noted that for a very short training period, the algorithm fails *a priori*. Hence a first condition concerns the length of the training period which must exceeds the generating ensemble size. As we have pointed out, our ensemble perturbed physic approach is not good method to generate the dataset, since lack of spread causes a degeneracy in covariance matrix, which can not be easily inverted. MFS BHM winds perturbation, acting on the vertical stratification of the fluid, was able to generate a large ensemble variability. In this latter contest the SE approach could give good results. A careful comparison

of the centred RMSE, reduced the outperformances of SE, enhancing the critical point played by the unbiased estimator. Longer training period let the unbiased estimator be a plain estimate of the field we want to reproduce (seasonal cycle affected). Furthermore, due to overfitting of multilinear regression (common problem when the number of observations is bigger than the parameters) the resulting maps show very noisy fields, which neglect the spatial correlations of the field.

As Krishanumrti and Kharin pointed out, the employment of EOFs/PC analysis could be a good method to remove the ill - conditioning of covariance matrix and the overfitting. In this way the results show that multimodel is not able to reduce the bias, since coefficients evaluated from filtered fields are smaller than in the classical approach, and the unbiased estimator is no more the best estimate of the field we want to reproduce. A careful study on the sensitivity of the algorithm on training period length and ensemble composition had been done in the third experiment using the extended dataset provided by MyOcean. We found a weak dependence of the performances of SE versus the length of the training period, and equally Anomaly Correlation Coefficient, ACC, showed a similar dependence but with a different rate, that is faster for short training period and then reaches a *plateau*. Instead, no important correlation have been noticed for the BIAS. We can infer that as soon as the ACC curve reaches the *plateau* and correspondly the RMS starts to diminish, we are in a overfitting regime, anomalies are perfectly "predicted" with any physical meaning but only due a statistical regression . This idea is confirmed even by the sensitivity of the algorithm with the ensemble composition. Following Weigel *et al.* (2008) we built three different sub-samples, an overconfident dataset, a well dispersed dataset and a bad dispersed dataset. For short training period, the tree sub-sample outperform the results of the original dataset, while for longer training period (almost 4 times the ensemble size) there are no particular differences between them, meaning that, again, we incur in overfitting. It must be pointed out that, as expected, better performances are reached with the overconfident dataset, confirming Weigel's suggestion that the dataset quality is a main impacting factor to the out-performance of SE. In this latter approach the employment of EOFs didn't give any improvement. The retrieved coefficients were very small, and bias reduction

problem arose. Histogram of the bias in time clearly show seasonal cycle effects (negative temperature bias in temperature rising period and conversely positive bias in fall-winter time).

In the last experiment we tried to simulate a forecast approach for the superensemble, and for the first time we tried to study the impact of model resolution on model error. As expected the spread was very high, confirming the high importance of the grid resolution on model error. SE results cover a wide range of possibilities: in some cases SE could be a better estimates but in general we drew the same considerations of the third experiment, *i.e.* the strong dependences on ensemble dataset quality and on the training period length.

Appendix A

Numerical schemes available in NEMO

A.1 Vertical Physics

A.1.1 KPP turbulent closure scheme

The KPP turbulent closure assumption for tracer is:

$$\overline{\hat{w}b} = -k \left(\frac{\partial \bar{b}}{\partial z} - \gamma \right) \quad (\text{A.1})$$

where k is the vertical mixing coefficient and b is any prognostic quantity. The non-local transport term is non zero only under convective forcing condition; while γ is proportional to the surface flux and inversely proportional to vertical friction velocity and mixing layer depth.

A.1.2 TKE scheme

The vertical mixing coefficients are computed from a 1.5 turbulent closure model based on a prognostic equation for e , the turbulent kinetic energy, and a closure assumption for the turbulent length scales.

$$\frac{\partial \bar{e}}{\partial t} = \frac{A^{\nu m}}{e_3} [(\partial u / \partial k)^2 + (\partial v / \partial k)^2] - A^{\nu m} N^2 + \frac{1}{e_3} \frac{\partial}{\partial k} \left[\frac{A^{\nu m}}{e_3} \frac{\partial \bar{e}}{\partial k} \right] - c_\varepsilon \frac{\bar{e}^{3/2}}{l_\varepsilon}$$

$$A^{\nu m} = C_k l_k \sqrt{\bar{e}}; \quad A^{\nu T} = A^{\nu m} / P_n$$

- l_ε and l_k are the dissipation and mixing turbulent length scales;
- P_{rt} Prandtl number.

A.1.3 The Pacanowski and Philander (PP) scheme

As in Pacanowski and Philander (1981), the background viscosity $\nu_b = 1 \text{ cm}^2/\text{s}$, diffusivity $\kappa_b = 0.1 \text{ cm}^2/\text{s}$, $\nu_0 = 100 \text{ cm}^2/\text{s}$, $n = 2$ and $\alpha = 5$. Here N represents Brunt-Väisälä frequency and U_z and V_z the vertical shear. For the convection case ($Ri_g < 0$), a maximum value of $1 \cdot 10^6 \text{ cm}^2/\text{s}$ is used in order to mix the heat instantaneously in the vertical to a depth that ensures a stable density gradient.

A.2 Tracer Advections

- Total Variance Dissipation scheme, the tracer at velocity points is evaluated using a combination of an upstream and a centred scheme. For example, in the i-direction :

$$\tau_u^{ups} = \begin{cases} \tau_{i+1} & \text{if } u_{i+1/2} < 0 \\ \tau_i & \text{if } u_{i+1/2} \geq 0 \end{cases}$$

$$\tau_u^{tvd} = \tau_u^{ups} + c_u (\tau_u^{cen2} - \tau_u^{ups});$$

– $0 \leq c_u \leq 1$ flux limiter function;

- Monotone Upstream Scheme for Conservative Laws, the tracer at velocity points is evaluated assuming a linear tracer variation between two T-points. For

example, in the i-direction :

$$\tau_u^{muscl} = \begin{cases} \tau_i + \frac{1}{2} \left(1 - \frac{u_{i+1/2} \Delta t}{e_{1u}} \right) & \widetilde{\partial}_i \tau & \text{if } u_{i+1/2} \geq 0 \\ \tau_{i+1/2} + \frac{1}{2} \left(1 + \frac{u_{i+1/2} \Delta t}{e_{1u}} \right) & \widetilde{\partial}_{i+1/2} \tau & \text{if } u_{i+1/2} < 0 \end{cases}$$

– $\widetilde{\partial}_i \tau$ is the slope of the tracer on which a limitation is imposed to ensure the positive character of the scheme.

- Upstream-Biased Scheme, it is based on the fourth order scheme to which an upstream-biased diffusion term is added. In the i-direction :

$$\tau_u^{ubs} = \begin{cases} \tau_u^{cen4} + \frac{1}{12} \tau_i'' & \text{if } u_{i+1/2} \geq 0 \\ \tau_u^{cen4} - \frac{1}{12} \tau_{i+1}'' & \text{if } u_{i+1/2} < 0 \end{cases} \quad (\text{A.2})$$

$$\tau_i'' = \delta_i [\delta_{i+1/2} \tau]$$

- The 4th order part (as well as the 2nd order part as stated above) has to be evaluated at the now time.
- The diffusion term is a biharmonic operator with an eddy coefficient proportional to the velocity.

A.3 Viscosity Operator

A.3.1 Laplacian Operator

In basin scale models, the smallest spatial scale is often the width of the western boundary current. When it is controlled by laplacian friction it is called a *Munk boundary layer*. The condition that the grid scale Δx be smaller than the Munk layer width results in a minimum bound for viscosity (Smith and McWilliams, 2003):

$$\nu > \nu_M \approx \beta \Delta x^3$$

On the other hand, viscosity cannot be arbitrarily large due to the stability constraint (similar to the CFL criterion for advection). This criterion is more severe in ocean models that use explicit leap-frog time stepping schemes for non-linear advection, with the viscous terms lagged by one time step for stability. For laplacian viscosity :

$$\nu < \frac{\Delta x^2}{8\Delta t}$$

Laplacian operator as modelled in OPA is :

$$D_T^{IT} = \frac{1}{e_{1T}e_{1T}e_{1T}} \left[\delta_i \left[A_u^{IT} \left(\frac{e_{2u}e_{3u}}{e_{1u}} \delta_{i+1/2} [T] \right) \right] + \delta_j \left[A_v^{IT} \left(\frac{e_{2v}e_{3v}}{e_{1v}} \delta_{j+1/2} [T] \right) \right] \right] \quad (A.3)$$

It preserves symmetry and ensures a complete separation between vorticity and divergence parts.

A.3.2 Bilaplacian Operator

For a biharmonic operator the criterion as coded in the OPA model is:

$$\nu < \frac{\Delta x^4}{128\Delta t}$$

For the biharmonic operation the numerical stability criterion is often more stringent than the Munk layer constraint. A decrease of the biharmonic coefficient with the grid spacing is often needed in order to ensure stability on spatially variable grids Smagorinsky (1963) has proposed to make the laplacian viscosity proportional to the deformation rate times the squared grid spacing Δx^2 . Such a parametrisation can be physically motivated in three dimensional turbulence and is used in large eddy simulations. A study by Griffies and Hallberg (2003) suggests that using a biharmonic operator with Smagorinsky-like viscosity is better in eddy permitting simulation when the flow is non homogeneous (in the presence of western boundary currents, for instance) because it allows lower levels of viscosity in the interior. The latter combination is the operational model set-up. The lateral fourth order

bilaplacian operator on tracers is obtained by applying (A.3) twice.

Appendix B

High resolution model

The aim of this study is to give a detailed description of the new implementation of the OGCM employed Mediterranean Sea forecasting. This model has been developed starting from the latest operational model, with a higher resolution of $1/24^\circ$ (almost 4.5Km) and 91 unevenly spaced vertical levels unevenly spaced and having a thickness ranging from 2 m at the surface to 245 m at the ocean bottom. The depth of the first level is 1 m and that of the deepest is 5000 m. The model domain and the bathymetry are shown in Fig.B.1: the coastline resolves 49 islands. The Digital Bathymetric Data Base-Variable Resolution has been used to make the MFS2491 coastlines and bathymetry. DBDB-5at 1' resolution has been used for the Mediterranean basin, whilst for the Atlantic DBDB-5 have been used. The bathymetry file has been manually corrected along the Croatian coast by a comparison with detailed nautical charts. The bathymetry has been interpolated on the model horizontal and vertical grid and manually checked for isolated grid point, islands and straits and passages and it is shown in figureB.1 . With this new grid, Messina Strait (fig. B.2) could be resolved, and as for Gibraltar strait an up-stream scheme is used to avoid numerical instabilities. Other main differences between the old resolution coast line can be found near Iskenderum gulf, that now it is fully resolved (fig.B.3).

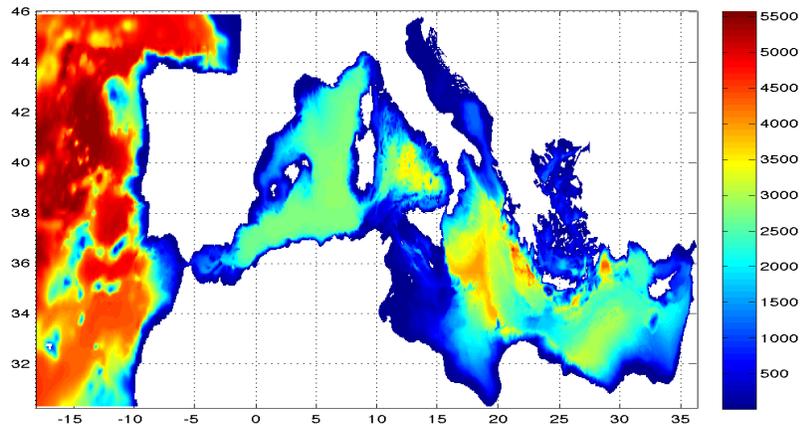


Figure B.1: *New model bathymetry and domain for the MFS2491 set up*

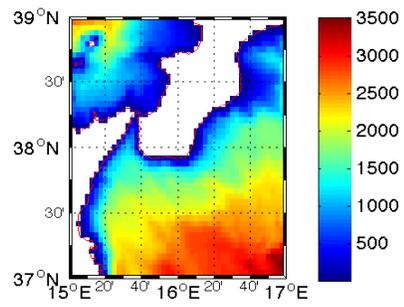


Figure B.2: *new Bathymetry near Messina Strait*

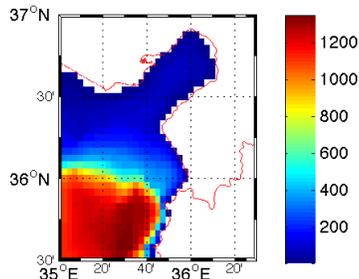


Figure B.3: *new Bathymetry near Iskenderum Gulf*

B.1 Vertical model discretization

All the vertical profiles taken from CTD show the same shape: gradients of density and tracers are often concentrated in the thin layers just below the surface mixed layer or in the thermocline. Because of the computational limitation, a stretched grid in vertical direction (with the maximum of resolution near surface) is best choice to ensure a second-order accuracy scheme (see Treguier *et al.* (1996)) without increasing computational cost.

NEMO vertical mesh is determined by four factors:

1. the bathymetry (in meters);
2. the number of vertical level (in our case 72, in the future it can increase till 100);
3. the analytical transformation of z (depth of level, see equations (B.1) and (B.2)) and the vertical scale factor (vertical derivatives of the transformation); in our case we use z -coordinate with partial step bathymetry. The vertical

distribution follow the equations (see ?) :

$$z_0(k) = h_{sur} - h_0 k - h_1 \log \left[\cosh \left(\frac{(k - h_{th})}{h_{cr}} \right) \right]; \quad (\text{B.1})$$

$$e(k) = -h_0 - h_1 \tanh \left(\frac{(k - h_{th})}{h_{cr}} \right); \quad (\text{B.2})$$

In the current configuration the values are

$$\begin{aligned} h_{sur} &= -110493.9930400577 \\ h_0 &= 1362.526788714143 \\ h_1 &= 1362.915990505609 \\ h_{th} &= 101.8303560439433 \\ h_{cr} &= 30.00000000000000 \end{aligned} \quad (\text{B.3})$$

In figure B.4 we can see the new vertical distribution profile implemented and compared to the previous version. In figure B.5 is the depicted the new relative frequency histogram of the number of vertical levels per interval depth, in yellow, and compared with the old relative frequency, in red.

4. Masking system.

The equation (B.1) allows us to define a nearly uniform vertical location of levels at the ocean top and a bottom with a smooth hyperbolic tangent transition in the water column between. Once chosen the desired resolution in the surface (bottom) about 2m (300m) layer and a range of depth (in our case it varies from 0 to -5000m) we can determine the values of the parameters in expression (B.3). Due to the atmospheric forcing parametrization and assimilation, the first layer must lie under 1m depth.

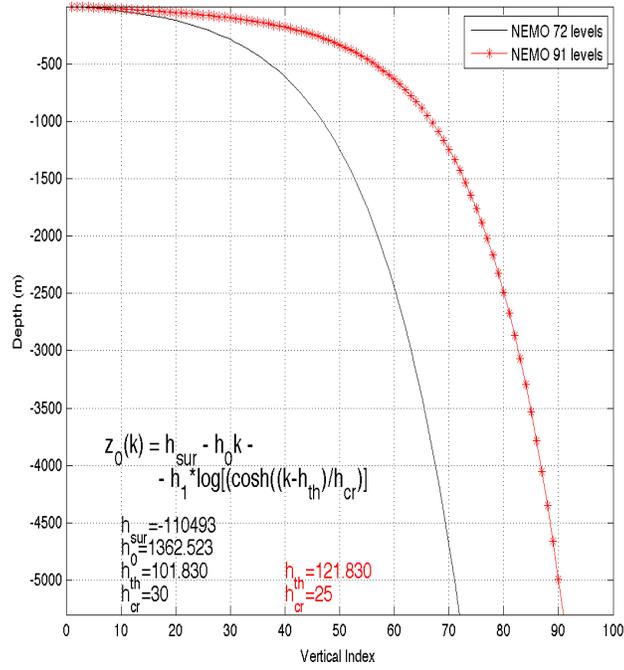


Figure B.4: Vertical level distribution versus the depth of water column as computed by NEMO-OPA, new one 91 levels in red, in black the old distribution with 72 levels

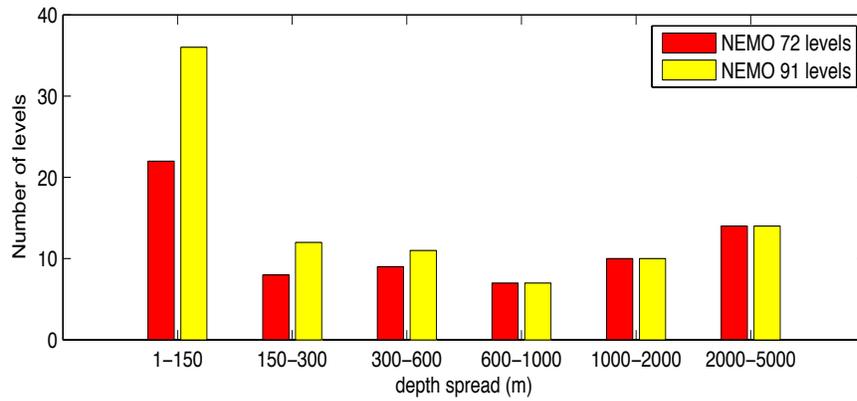


Figure B.5: Histogram of relative frequency number of level for each interval depth, in old version in red, while the new in yellow.

Bibliography

- A Navarra, V Simoncini. 2010. *A guide to empirical orthogonal functions for climate data analysis*. Springer.
- Batchelor, G.K. 1967. *An introduction to fluid mechanics*. Cambridge University Press.
- Buizza, R., & Palmer, T. 1995. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**(9), 1434–1456.
- Chin, T.M, Milliff, R.F, & Large, W.G. 1998. Basin-scale, high-wavenumber sea surface wind fields from a multiresolution analysis of scatterometer data. *Journal of atmospheric and oceanic technology*, **15**(3), 741–763.
- De Mey, P., M. Benkiran In :. 2002. *A multivariate reduced-order optimal interpolation method and its application to the mediterranean basin-scale circulation*. Springer-verlag, berlin heidelberg new york, edn. Page 472pp.
- Dobricic, Srdjan, & Pinardi, Nadia. 2008. An oceanographic three-dimensional variational data assimilation scheme. *Ocean modelling*, **22**(3-4), 89 – 105.
- Drevillon, M., Bourdallé-Badie, R., Derval, C., Drillet, Y., Lellouche, J. M., Rémy, E., Tranchant, B., Benkiran, M., Greiner, E., Guinehut, S., Verbrugge, N., Garric, G., Testut, C. E., Laborie, M., Nouel, L., Bahurel, P., Bricaud, C., Crosnier, L., Dombrosky, E., Durand, E., Ferry, N., Hernandez, F., Le Galloudec, O., Messal, F., & Parent, L. 2008. The GODAE/Mercator-Ocean global ocean forecasting system: results, applications and prospects. *Journal of operational oceanography*, 51–57.

- Evans, R, Harrison, M, Graham, R, & Mylne, K. 2000. Joint medium-range ensembles from the met. office and ecmwf systems. *Mon. wea. rev.*, **128**(9), 3104–3127.
- Evensen, Geir. 2003. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 343–367.
- Fedderson, H, Navarra, A, & Ward, M.N. 1999. Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. climate*, **12**(7), 1974–1989.
- Gould, John., Siedler, Gerold., & Church, John. 2001. *Ocean circulation and climate : observing and modelling the global ocean / edited by gerold siedler, john church, john gould*. Academic, San Diego, Calif. ; London .
- Hagedorn, DOBLAS-REYES, & PALMER. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – i. basic concept. *Tellus a*, Apr, 18.
- Haidvogel, D. B., & Beckmann, A. 1999. *Numerical ocean circulation modeling*. Series on Environmental Science and Management, vol. 2, no. 319 pp. Imperial College Press.
- Holton, J. R. 1992. *An introduction to dynamic meteorology*. 3rd edn. Academic Press, San Diego.
- Houtekamer, PL, Lefaiivre, L, Derome, J, Ritchie, H, & Mitchell, H.L. 1996. A system simulation approach to ensemble prediction. *Monthly weather review*, **124**(6), 1225–1242.
- Johnson, Christine, & Bowler, Neill. 2009. On the reliability and calibration of ensemble forecasts. *Mon. wea. rev.*, **137**(5), 1717–1720. doi: 10.1175/2009MWR2715.1.
- Kalnay, E, Hunt, B, Ott, E, & Szunyogh, I. 2006. Ensemble forecasting and data assimilation: two problems with the same solution. *Predictability of weather and climate*, 157–180.

- Kalnay, Eugenia. 2003. *Atmospheric modeling, data assimilation and predictability*.
- Kharin, V.V., & Zwiers, F.W. 2002. Climate predictions with multimodel ensembles. *J. climate*, **15**(7), 793–799.
- Knutti, R. 2010. The end of model democracy? *Climatic change*, **102**(3), 395–404.
- Korres, G., Nittis, K., Hoteit, I., & Triantafyllou, G. 2009. A high resolution data assimilation system for the aegean sea hydrodynamics. *Journal of marine systems*, **77**(3), 325 – 340.
- Krishnamurti, T, Kishtawal, C, LaRow, Timothy, Bachiochi, David, Zhang, Zhan, Williford, C, Gadgil, Sulochana, & Surendran, Sajani. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**(5433), 1548–1550. 10.1126/science.285.5433.1548.
- Krishnamurti, T, Kishtawal, C, Zhang, Zhan, LaRow, Timothy, Bachiochi, David, Williford, Eric, Gadgil, Sulochana, & Surendran, Sajani. 2000. Multimodel ensemble forecasts for weather and seasonal climate. *J. climate*, **13**(23), 4196–4216.
- Krishnamurti, TN, Rajendran, K, Kumar, TSV Vijaya, Lord, S, Toth, Z, Zou, X, Cocks, S, Ahlquist, J.E, & Navon, I.M. 2003. Improved skill for the anomaly correlation of geopotential heights at 500 hpa. *Monthly weather review*, **131**(6), 1082–1102.
- Large, W. G. 1998. Modeling and parameterization ocean planetary boundary layers. *Journal of physical oceanography*, 81–120.
- Large, WG, McWilliams, JC, & Doney, SC. 1994 (Jan). *Oceanic vertical mixing - a review and a model with a nonlocal boundary-layer parameterization*.
- Lenartz, F, Mourre, B, Barth, A, Beckers, J.M, Vandenbulcke, L, & Rixen, M. 2010. Enhanced ocean temperature forecast skills through 3-d super-ensemble multimodel fusion. *Geophys. res. lett.*
- Lermusiaux, P.F.J. 2006. Quantifying uncertainties in ocean predictions.

- Lin, S.J., W.C. Chao Y.C. Sud, & Walker, G.K. 1994. A class of the van leer-type transport schemes and its application to the moisture transport in a general circulation model.
- Lorenz, Edward N. 1963. Deterministic Nonperiodic Flow. *J. atmos. sci.*, **20**(2), 130–141.
- Madec, Gurvan. 2008 (May). *Nemo ocean engine*. 3.1 edn. Institute Pierre-Simon Laplace, France.
- Mellor, G.L, & Blumberg, A.F. 1985. Modeling vertical and horizontal diffusivities with the sigma coordinate system. *Monthly weather review*, **113**(8), 1379–1383.
- Meng, Zhiyong, & Zhang, Fuqing. 2011. Limited-area ensemble-based data assimilation. *Mon. wea. rev.*, **139**(7), 2025–2045. doi: 10.1175/2011MWR3418.1.
- Milliff, R. F. 2004. Comparing the kinetic energy vs. wavenumber in surface wind fields from ecmwf analyses and the nasa quikscat scatterometer. *Istituto nazionale geofisica e vulcanologia*, **Technical Report**(Mar), 1–19.
- Milliff, R.F, Morzel, J, Chelton, D.B, & Freilich, M.H. 2004. Wind stress curl and wind stress divergence biases from rain effects on qscat surface wind retrievals. *Journal of atmospheric and oceanic technology*, **21**(8), 1216–1231.
- Milliff, R.F, Bonazzi, A, Wikle, C.K, Pinardi, N, & Berliner, L.M. 2011. Ocean ensemble forecasting. part i: Ensemble mediterranean winds from a bayesian hierarchical model. *Q.j.r. meteorol. soc.*
- M.Tonani, G., Teruzzi A.and Korres G.and Pinardi N.and Crise A.and Adani M.and Oddo P.and Dobricic S.and Fratianni C.and Drudi M.and Salon S.and Grandi A.and Girardi, V., Lyubartsev, & S., Marino. in press. *The mediterranean monitoring and forecasting centre, a component of the myocean system*.
- Murphy, A, & Epstein, E. 1989. Skill scores and correlation coefficients in model verification. *Monthly weather review*, **117**(3), 572–581.

- Murphy, Allan H. 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and forecasting*, **8**(2), 281–293.
- Murphy, JM, Sexton, DM, Barnett, DN, Jones, GS, Webb, MJ, Collins, M, & Stainforth, DA. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**(7001), 768–72.
- Obukhov, AM. 1947. Statistically homogeneous fields on a sphere. **Nauk 2:**, 196–198.
- Oddo, P., Adani, M., Pinardi, N., Fratianni, C., Tonani, M., & Pettenuzzo, D. 2009. A nested atlantic-mediterranean sea general circulation model for operational forecasting. *Ocean science*, **5**(4), 461–473.
- Pedlosky, Joseph. 1979. *Geophysical fluid dynamics / joseph pedlosky*. Springer Verlag, New York .:
- Pellerin, G, Lefaivre, L, Houtekamer, P, & Girard, C. 2003. Increasing the horizontal resolution of ensemble forecasts at cmc. *Nonlinear processes in geophysics*, **10**(6), 463–468.
- Pham, D Tuan, Verron, J, & Roubaud, M Christine. 1998. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of marine systems*, **16**(3-4), 323–340.
- Pinardi, N, Bonazzi, A, Dobricic, S, Milliff, R.F, Wikle, C.K, & Berliner, L.M. 2011. Ocean ensemble forecasting. part ii: Mediterranean forecast system response. *Q.j.r. meteorol. soc.*
- R C Pacanowski, S G H Philander. 1981. Parameterization of vertical mixing in numerical models of tropical oceans.
- Redi, Martha H. 1982. Oceanic isopycnal mixing by coordinate rotation. *Journal of physical oceanography*, **12**(10), 1154–1158.

- Reynolds, O. 1895. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical transactions of the royal society of london. a*, **186**, 123–164.
- Rosati, A, & Miyakoda, K. 1988. A general circulation model for upper ocean simulation. *J. phys. oceanogr*, **18**(11), 1601–1626.
- Shapiro, R. 1970 (Jan). *Smoothing, filtering, and boundary effects*.
- Smagorinsky, J. 1993. *Some historical remarks on the use of nonlinear viscosities*. Cambridge University Press.
- Solomon, Harold. 1971 (April). *On the representation of isentropic mixing in ocean circulation models*.
- Taylor, Karl. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. geophys. res.*, **106**(D7), 7183–7192.
- Tetko, I.V, Livingstone, D.J, & Luik, A.I. 1995. Neural network studies. 1. comparison of overfitting and overtraining. *Journal of chemical information and computer sciences*, **35**(5), 826–833.
- Tonani, Pinardi, Fratianni, Pistoia, Dobricic, S, Pensieri, de Alfonso, & Nittis. 2009. Mediterranean forecasting system: forecast and analysis assessment through skill scores. *Ocean science*, **Ocean Science**(5), 649–660.
- Toth, Z., & Kalnay, E. 1993. Ensemble Forecasting at the NMC: The generation of perturbations. *Bull. amer. meteorol. soc.*, **74**, 2317–2330.
- Tranchant, B, Greiner, E, Bourdallé-Badie, R, & Testut, CE. Monitoring the global ocean mesoscale with a global ocean forecasting system at 1/12.
- Treguier, A, Dukowicz, J, & Bryan, K. 1996. Properties of nonuniform grids used in ocean general circulation models. *J. geophys. res*, Jan.

- Vich, M., Romero, R., & Brooks, H. E. 2011. Ensemble prediction of mediterranean high-impact events using potential vorticity perturbations. part i: Comparison against the multiphysics approach. *Atmospheric research*, **102**(1-2), 227–241.
- von Storch H. 1995.. *Spatial patterns: Eofs and cca. in analysis of climate variability: Application of statistical techniques*,. Springer verlag: Berlin; edn. Vol. Springer Verlag:.
- Weigel, A. P, Liniger, M. A, & Appenzeller, C. 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q.j.r. meteorol. soc.*, **134**(630), 241–260.
- Yun, Krishnamurti, & Stefanova. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. *Journal of climate*, **36**.
- plainnat