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

Il coordinatore:
Michele Dragoni

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

Esame Finale anno 2012

... 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

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.

Contents

List of Figures

List of Tables

Introduction

The Mediterranean ocean Forecasting System

The Mediterranean ocean Forecasting System (MFS) is working operationally since 2000 and is beeing continuolsy improved througth 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.*, 2008). 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 weekely 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

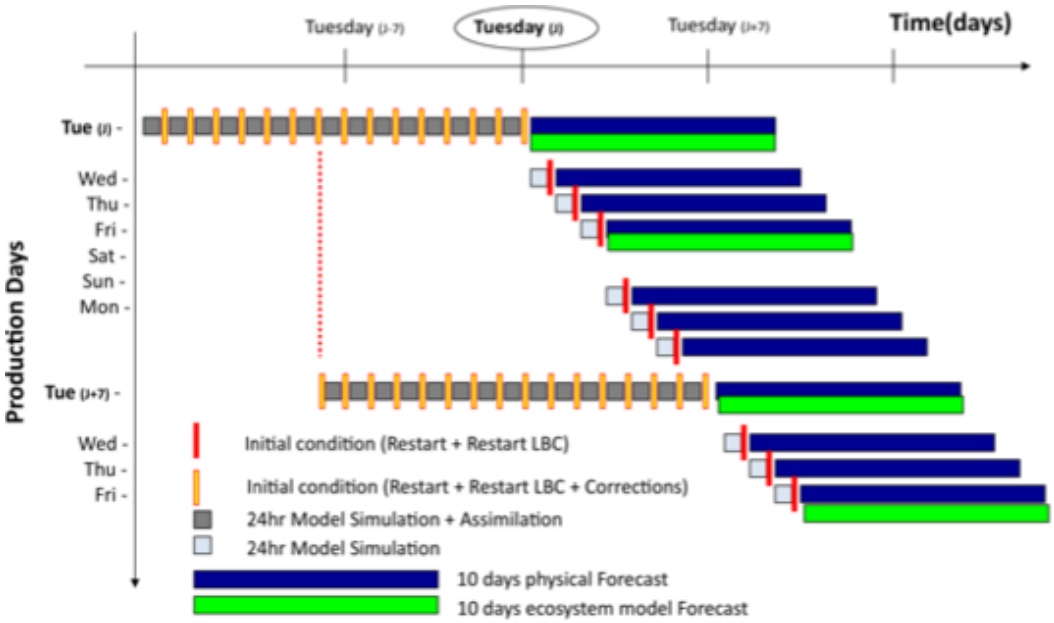

Figure 1: *MFS production cycle. Every Tuesday (J) a ten-day forecast (d1, d2, d3, d4, d5, d6, d7, d8, d9, d10) is produced. It is initialised by an analysis generated by the past sequence of 15 intermittent daily data assimilation cycles.*

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 Baum- hefner 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 superensamble 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 affecting the numerical simulations and the differences between "intrinsic" and "structural" uncertainties. Furthermore, since perfect model assumption must be dropped in real world studies where computational constraints, 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 the creation of suitable ensemble members which are presented in chapter 2. Furthermore, we will comment 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 will present a way to improve to the classical superensemble approach, considering the EOFs/PC Analysis. After an overview of how these statistical techniques can decompose a space and time dependent field into a set of spatial patterns and associated time indices, we will give the linking between their formulation 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 from *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 evaluating explicitly (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,end 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 in both 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 limited area model are used, observations are needend to prescribe at the lateral open boundary processes not occuring locally but affecting and influencing the 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 defined *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 by 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 loose 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:

$$A^{vm}\frac{\partial \mathbf{u_h}}{\partial z} \mid_{z=\eta} = \frac{\tau}{\rho_0} \tag{1.1}$$

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:

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

E and P are evaporation and precipitation, $\frac{D}{Dt} = \frac{\partial}{\partial t} + u_h \mid_{z=\eta} \cdot\nabla$ is the total derivative.
The water flux is coupled to the salinity boundary condition by the $(E - P)$ term with runoff(R) by:

$$A^{vs}\frac{\partial S}{\partial z} \mid_{z=\eta} = (E - P - R)S_{z=\eta} \tag{1.3}$$

Finally, the boundary condition for heat flux is:

$$A^{vT}\frac{\partial \mathbf{T}}{\partial z} \mid_{z=0} = \frac{Q}{\rho_0 C_p} \tag{1.4}$$

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 necessarily must 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 during the last ten years, heat fluxes started to be parametrized with empirical (bulk) formulae forced by atmospheric data sets rather than 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 used, since it represents the best compromise. Atmospheric forcing data (wind, temperature, humidity, precipitation and cloud cover) are originated from a variety of sources with different accuracy, so it arises the question about 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)). 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 ocean[1].

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

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

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

or that mass per unit volume times the accelaration 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 $\boldsymbol{F}$ is in principle the frictional force in the fluid. For Newtonian fluid like air or water

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

where μ is the molecular viscosity. This is an exact representation for $\boldsymbol{F}$ 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 into account of the related source/sink terms and diffusivity redistribution of salinity.

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

In order to include some effects of the small-scale processes, which are not the focus of our interest but still 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.

[1] see Pedlosky (1979)

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

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

$$\bar{\hat{B}} = 0$$
$$\bar{\bar{B}} = \bar{B}$$
$$\overline{\bar{B}_1 \hat{B}_2} = 0$$

and

$$\overline{\bar{B}_1 \bar{B}_2} = \overline{B}_1 \overline{B_2}$$

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

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

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:

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

The last term is called *Reynolds stress tensor*, its diagonals 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:

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

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

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

In wide use is the much simpler diagonal form:

$$A = \begin{bmatrix} A^{h} & 0 & 0 \\ 0 & A^{h} & 0 \\ 0 & 0 & A^{v} \end{bmatrix} \tag{1.14}$$

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 is 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 class, 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 operator[2] e.g.

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

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 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. Therefore, the choice of diffusive closure is closely related to the form of the horizontal advection operator. *A priori* choice is not always obvious. Some

[2]see Haidvogel & Beckmann (1999)

closures are rather *ad hoc*, and the preservation of *smooth* numerical results is the only available justification for their use. 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:

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

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \overline{\rho}}{\partial z} \tag{1.17}$$

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 approaches are the so called $k - l$ and $k - \epsilon$ schemes. The first approach 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 it uses 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 the reason why non-local parametrisation has been developed, like the KPP scheme (Large *et al.* (1994) see appendixA.1.1 for a brief description) where the vertical gradients are explicitly retained rather then the assumption of having a well mixed boundary layer.

A different approach, called Pacanowski and Philander(*P.P.*) (Pacanowski & Philander, 1981), reckons the vertical diffusion coefficients as a function of the local

Richardson number(Ri). In this scheme, turbulent mixing is treated by a first-order local diffusion approach. Measurements of ν (vertical eddy viscosity) and κ (vertical eddy diffusivity), show that these parameters vary considerabily in the ocean,and usually they have a large values in the mixed surface layer, but they have very small values below the termocline. Empirical studies (Robinson, 1966) indicates that the shear dependence of ν (vertical eddy viscosity) and κ (vertical eddy diffusivity) should be of the following form:

- $\nu = \frac{\nu_0}{(1+\alpha Ri)^n} + \nu_b$
- $\kappa = \frac{\nu}{(1+\alpha Ri)} + \kappa_b$

where ν_b and κ_b are the background dissipation parameters and ν_0, α and n are adjustable parameters derived after appropriate measurements and laboratory studies. The local Richardson number(Ri) is evaluated as follows:

$$Ri = \frac{N^2}{U_z^2 + V_z^2} \tag{1.18}$$

Here N represents Brunt-V*äisälä* frequency and U_z and V_z are the vertical shear.

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 about Navier-stokes equations, they are subjected to some approximations, furthermore other kinds of simplification are due to computational constraint. Those simplifications are derived by the *scaling analysis*, and the 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 accelation 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, $\hat{\rho}$:

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

As conseguence in the momentum equation and the mass conservation, we may sustitute the inertial term and the consituity equation with a constant.

- **Incompressibility**:The specification of incompressibility and constant density immediately decouples the dynamic from the thermodynimics and reduce the equation of mass conservation to the condiction 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 that together with the approximation 1.19, it became:

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

The second term is smaller in comparison 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, then the third term is also smaller then the first by a factor. So a suitable approximation can be:

$$\nabla \cdot \vec{v} = 0 \tag{1.21}$$

for a quantitative and more accurate demonstration refers to Batchelor (1967).

- **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.

$$-\frac{1}{\rho}\frac{\partial p}{\partial z} - g \simeq 0 \tag{1.22}$$

For 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 earthś 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:

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

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. Futhermore ∇ 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 explicitly known 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.

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

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 $^{100}/_{N_{prcs}}$ 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 replace the series by the $n-th$ order polynomial, this leads 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 uncertanies 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 °(from 2009, before 1/2) they have to be interpolated by an apposite algorithm (bi-linear or bi-cubic according to the field) to the operational set up regular grid 1/16 °resolution. (1/24 °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 had the porting on a scalar machine,of the old code in order to ran 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 they 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

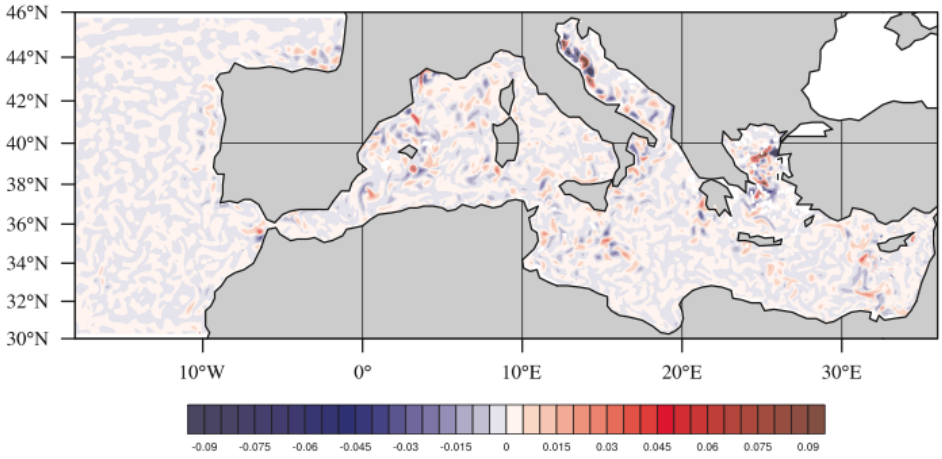

Figure 1.1: Differences °C the twin experiment simulations

interpolation methods ensure the same 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

The SuperEnsemble dataset: preparation and analysis

Introduction

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 slightly-perturbed initial conditions, with perturbations designed to capture as much as possible the underlying uncertainty (Kalnay, 2003). Similarly, model errors due to physical parametrizations can be addressed by the *"multi-physics"* 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. If the uncertanties assosiaciated with model numerics and implementation choices are to be studied, then a multi-model approach can be considered where different numerical models and discretization choices (bottom topography, horizontal resolution and vertical levels) can be *addressed*.
As far as uncertainties are concerned and interrelated, no one of the previous approach (multi-model or multi-physiscs superensemble) is clearly better than another, and in this work, we will build superensemble datasets derived from both.

Since simulations are representation of reality as observations with different error characteristics, we can use the same terminology used in experimental physics. Hence, the *accuracy* of a measurement system is the level of closeness of measurements of a quantity to that quantity's actual (true) value, while the *precision* is the extent to which repeated measurements under unchanged conditions give 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 increases precision but does not improve accuracy. Eliminating the systematic error improves accuracy but does not change precision. For our purposes, the standard deviation around the ensemble mean, also referred as *spread*, mimics the accuracy, while the difference in the variability from the ensemble members and the observed state, estimates the precision.
A justification of the last statement can be given by analysing figure 2.1. As evi-

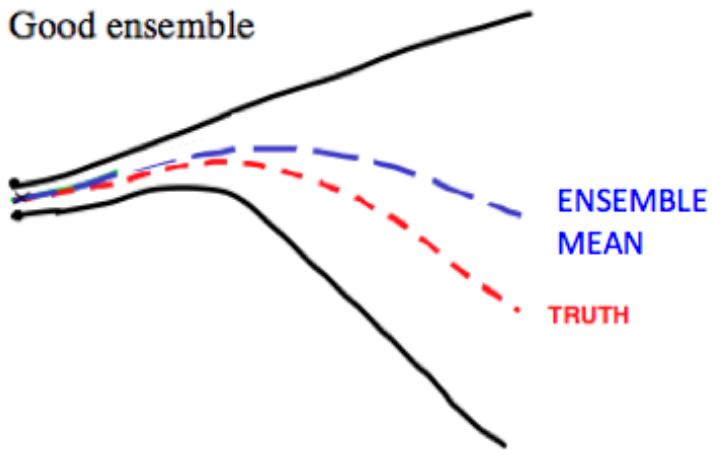

Figure 2.1: Practical example corresponding to the idealized *"good ensemble"* (from Kalnay *et al.*, 2006)

denced by (Kalnay *et al.*, 2006), we have built a *"good ensemble"* for prediction when the "truth" looks like a member of the ensemble. This situation is achieved when the spread around the ensemble mean, is bigger than the mean differences between model simulations and observations. As evidenced by Hagedorn *et al.* (2005), Weigel

et al. (2008) and Knutti (2010) later, the feedback spread-bias reduction is challenging since it is difficult to remove a correct bias from simulations, when the standard deviation around the ensemble mean is less than the mean error between simulations and true state of the field. Let's consider the situation depicted in fig.2.1, the mean difference of the ensemble mean and true state is the area between those (blue and red) lines. Increasing the spread corresponds to enlarge the area between the black border lines. Therefore, the best representation of truth is more likely predicted by a single member of the ensemble when ensemble dataset has a wide spread. On the other hand, when the spread is low, all the members converge in the description of an expected state. As stated by Feddersen *et al.* (1999), the low ensemble spread is likely to be produced by correlated models. In this configuration, all models are affected by similar systematic errors, and possibly no one of them is better than the other to give a reasonable representation of the true state. Hence, multi-model approach seeks a gain in accuracy despite to the low precision of the measurement system.

Furthermore, for a reliable system, we require that an ensemble member should have the same statistical properties as the truth. Within this assumption, the true state can be considered as a member of the ensemble. Let's assume we have a time series of observations, indicated by O_t, and numerical simulations or forecasts for the $i-th$ model at $t-th$ time indicated by $F_{i,t}$, hence, the variance of the observations σ_O^2, can be easily decomposed in a sum of the variance around the ensemble mean and the squared bias between ensemble mean and observations:

$$\sigma_O^2 = E_t\left[\left(O_i - \overline{O}\right)^2\right] = E_t\left[\left(O_i - M_i\right)^2\right] + E_t\left[\left(\overline{O} - \overline{M}\right)^2\right] \tag{2.1}$$

where the operator $E[\,]$ denotes the expectation, or average, over time and

$$M_i = E\left[F_i\right]_t = \frac{\sum_{i=1}^{M} F_{i,t}}{M} \tag{2.2}$$

is the ensemble mean for the $i-th$ day. While

$$\overline{M} = E_t\left[\bar{M}_i\right] = \frac{\sum_{i=1}^{T}\overline{M_i}}{T} \tag{2.3}$$

is the time mean of the ensemble mean. Hence the first term, in the right hand side of eq.2.1 is the variance between the ensemble mean and the observations, while the second stands for the square difference between the means, or the difference in bias. We can decompose in a similar manner the variance of each simulation member:

$$\sigma_F^2 = E_t\left[\left(F_i - \overline{F}\right)^2\right] = E_t\left[(F_i - M_i)^2\right] + E_t\left[\left(\overline{F} - \overline{M}\right)^2\right] \tag{2.4}$$

Since members, observations and simulations are assumed to belong to the same probability density function, we assume the following condition:

$$E_t\left[(O_i - M_i)^2\right] = E_t\left[(F_i - M_i)^2\right] \tag{2.5}$$

Hence, as evidenced by Johnson & Bowler (2009), in a perfect ensemble, the Mean Squared error (MSE) of the ensemble mean must be equal to the average ensemble variance, so that the spread is representative of the uncertainty in the ensemble mean. The variability for the simulation ensemble members will be evaluated as:

$$V_E = \frac{1}{M}\sum_{i}^{M}\frac{\sum_{t=1}^{T}\left(F_{i,t} - \overline{F}_i\right)^2}{T} \qquad \overline{F_i} = E\left[F_t\right]_i = \frac{\sum_{t=1}^{T}F_{i,t}}{T} \tag{2.6}$$

while for the observed state it will be evaluated as:

$$V_O = \frac{\sum_{t=1}^{T}\left(O_t - \overline{O}\right)^2}{T} \qquad \overline{O} = E\left[O_t\right] = \frac{\sum_{t=1}^{T}O_t}{T} \tag{2.7}$$

Here T is the length of the time period, in our case ten days, M is the number of ensemble members, The spread of the ensemble is defined as:

$$STD_E = \frac{1}{T}\sum_{t}^{T}\sqrt{\frac{\sum_{i=1}^{M}\left(F_{i,t} - \overline{M}_t\right)^2}{M}} \tag{2.8}$$

The mean absolute difference (MAD)

$$MAD = \frac{1}{T}\sum_{t}^{T} \mid \overline{M_t} - O_t \mid \tag{2.9}$$

Here, $\overline{M_t}$ is the ensemble mean for the $t - th$ time.

2.1 Ensemble member datasets

2.1.1 Members with different numerical schemes and parametrizations

So far, operational ensemble predictions have been based on the assumption that the numerical models involved are perfect (Toth & Kalnay, 1993 and Palmer & Anderson, 1994). Hence model numerical schemes and parametrizations are not considered to be a source of errors, while, normally, the ensemble predictions focus on initial condition uncertainties, coming from observational sampling errors or imperfect data assimilation schemes (Houtekamer *et al.*, 1996).

In the ocean, where observations are scarcer than the atmosphere, different methods have been used to test initial condition uncertainties (Evensen, 2003 and Dobricic *et al.*, 2007) and multi-model and multi-physics superensemble techniques have not been as widely studied yet. Here, firstly, we will describe a multi-physics ensemble approach done to to sample the uncertainties related to the choice of a particular physical parametrization scheme or the effect of a numerical scheme selection. These sources of uncertainties are sampled building the multi-physics ensemble as a collection of simulations with the corresponding schemes modified. The model used in this thesis is based on NEMO-OPA code (Nucleus for European Modelling of the Ocean-Ocean PArallelise, Madec, 2008), version 2.3, which is developed and maintained by the NEMO-consortium, adequately modified for the Mediterranean domain (Oddo *et al.*, 2009).

In the standard model formulation, the Mediterranean domain is composed by 72 vertical levels and has a horizontal resolution of $1/16 \times 1/16$ °(approximately 6.5Km at our latitude). Momentum, water and heat fluxes are interactively computed by bulk formulae using the 6-h, $1/2$ °horizontal-resolution operational analyses fields from the European Centre for Medium-Range Weather Forecasts (ECMWF). The horizontal viscosity and diffusion operators are assumed to be bi-laplacian with coefficients of $5 \times 10^9 m^4/s$ and $3 \times 10^9 m^4/s$, respectively. The vertical diffusion and viscosity terms are dependent upon the Richardson number, while the vertical convective processes are parametrized using the enhanced vertical diffusivity parametrization. The advection

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: *MPE 1 ensemble model configuration: Member number 1 is the model standard configuration. All the other members are obtained changing the specific numerical scheme or parametrization.*

scheme for active tracers, temperature and salinity, is a mixed up-stream/MUSCL (Monotonic Upwind Scheme for Conservation Laws, Estubier & Lévy, 2000) scheme. This hybrid 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.

In the multi-physics ensemble dataset we modified some of the assumptions in this standard model. Specifically, (see table 2.1) the multi-physics dataset is the combination of models with two different viscosity operators (laplacian and bilaplacian) for the momentum, two vertical diffusion schemes ($k-\epsilon$ and P.P.), three different advection schemes (*upstream*, Total Variance Dissipation -TVD - and MUSCL) and three kinds of bottom friction parametrisations (all those numerical schemes and parametrizations are briefly described in appendix A). All the simulations started from the same initial conditions (climatology), the 7th of January 2004, at the 00 UTC and lasted 30 days. We will refer to this dataset as MPE 1 (Multi-Physics Ensemble number 1).

Resulting simulations have been analysed in terms of variance and ensemble spread for the last ten days of the experiment period, and compared to SST derived by satellite. Comparisons are depicted in figures 2.2 and 2.3. Probably the big differences in the variability maps are initial conditions effects: since we were starting from climatology, a motionless steady field, all the ensemble members will reproduce this variance-lapse. Regarding the spread, it is almost zero in the considered domain, except in some regions (Gibraltar Strait, river mouths and open boundaries). Here, we have large instability areas, so small differences can grow and propagate during the investigation time. Despite the low V_E results, this study was helpful in order to evaluate the capability of the model to generate spatial variability as a function of the numerical schemes and parametrization.

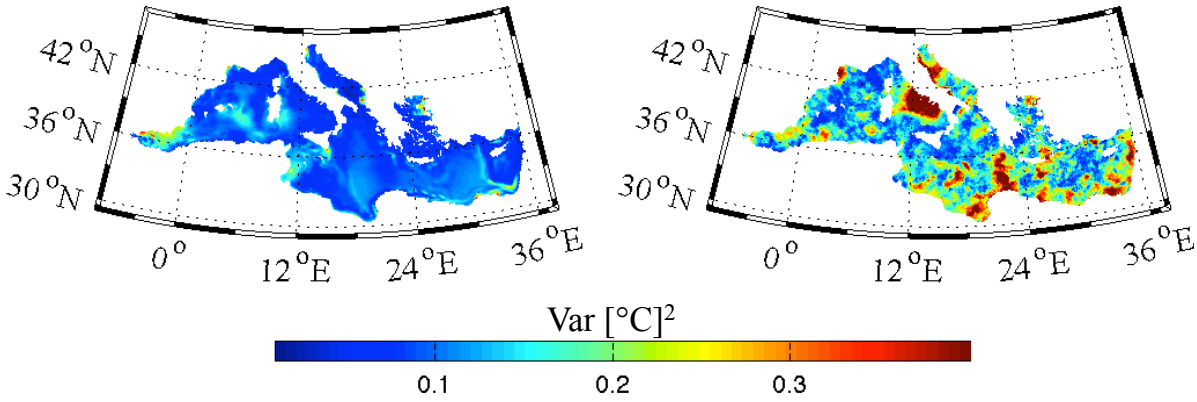

Figure 2.2: *Mean Variability V_E in SST generated Multi-Physics ensemble, MPE-1 (left) versus the natural variability in the observations, the SST from satellite, from the 27th January to 5 February 2004*

As we have already noted the main limit of this MPE dataset is the *systematic error*. Since the ensemble will have the structure of the underlying standard model, if the standard member is not able to reproduce a specific 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 only by the incompressibility condition without direct energy source. The 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

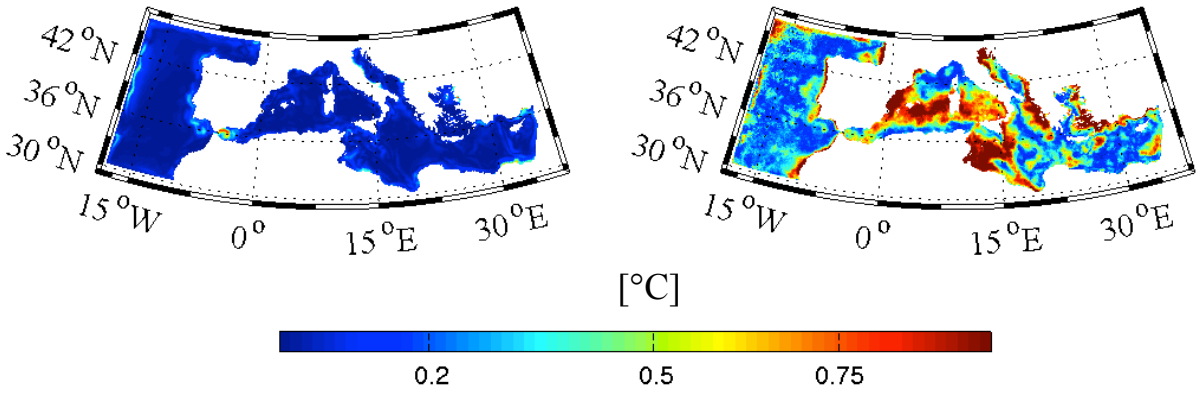

Figure 2.3: *Standard Deviation, STD_E, of Multi-physics ensemble members MPE-1 (left) and the Mean Absolute Difference (MAD) between Ensemble Mean and the SST from satellite (right) from the 27th January to 5 February 2004*

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.

2.1.2 Members with different atmospheric forcing

Atmospheric forcing uncertainties in ocean model simulations and analyses are an important source of errors and thus we choose here to use a wind perturbation technique that in turn modifies air-sea interaction fluxes that force the model simulation. 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, sampling the uncertainties affecting the atmospheric forcing field.

The basic idea behind this approach lies in the usage of atmospheric variable uncertainties to produce changes in air-sea physics parametrization. As said before, the heat fluxes in the standard model are computed with bulk formulas (Pettenuzzo *et al.*, 2010) that contains the winds. Thus different winds will produce different heat fluxes that in turn will produce different model simulations. The major effect

of wind uncertainty will propagate into the momentum flux too, that is the major forcing of the ocean circulation. Hence, in this case, we assume a model error in air-sea physics parametrizations due to the wind forcing. We will refer to this dataset as Multi-Physics Ensemble 2 (MPE-2) since different winds distribution generates different physical exchanges in air-sea interactions.

	Target field	Ensemble members
Product	Measurements by satellite infrared sensors	Output simulation
Time	Mid night *"Instantaneous"* field	Daily mean centred at mid night
Diagnostic	All data available in 10 days, using a symmetrical time window for the statistical interpolation	Daily mean time step 600 sec

Table 2.2: *Set up of Multi-Physics Ensemble MPE 2 Experiment, first row describes the target field and the second the 11 ensemble members.*

In order to compare the spread and variability growth rate, we carried out an experiment with a similar set up to the previous.Table 2.2 lists the main differences between the ensemble members, a total amount of11 output simulations obtained with the BHM scheme -the standard, plus 10 different wind realisations- starting from the same initial conditions, and the target field, satellite SSTs measurements. Again, we compared the ensemble behaviour in term of variance, shows in figure 2.4,and we can note that we get bigger variability that the previous MPE-1, but we were not able to reproduce as much variability as the observations. Furthermore concerning the observed field, we can note that in this period, from 20 of February to the first of March year 2008, the observations present less variability than in the end of January of 2004, the time period of the previous experiment. Concerning the ensemble spread, depicted in figure 2.5, we were able to reach a larger spread than before. Despite this good result, ensemble spread is still minor than the mean observation STD, indicating that our standard simulation is affected by a considerable bias. We can

conclude that BHM for surface winds is a suitable method for generating spread, but in order to be reliable we need to choose a different standard model.

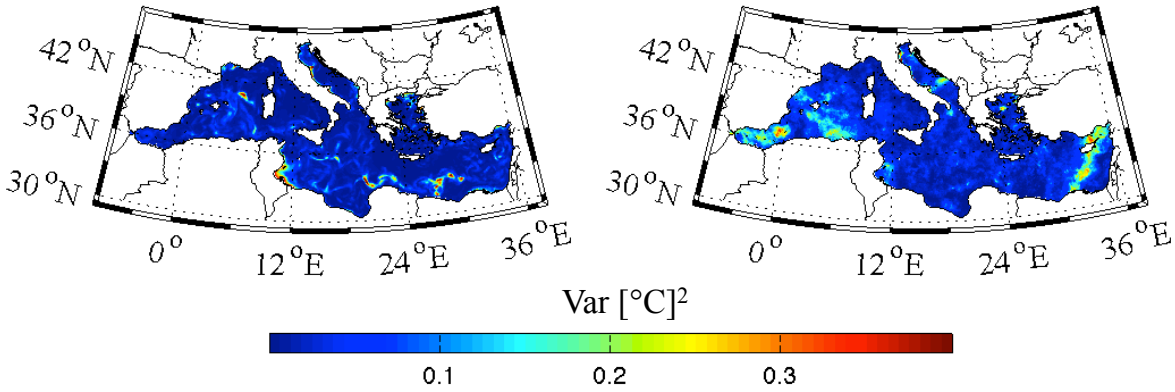

Figure 2.4: *Mean Ensemble Variability V_E for MPE-2 ensemble (left) and the observation variability, V_O, in the observations from the 20th of February to the 1st march 2008.*

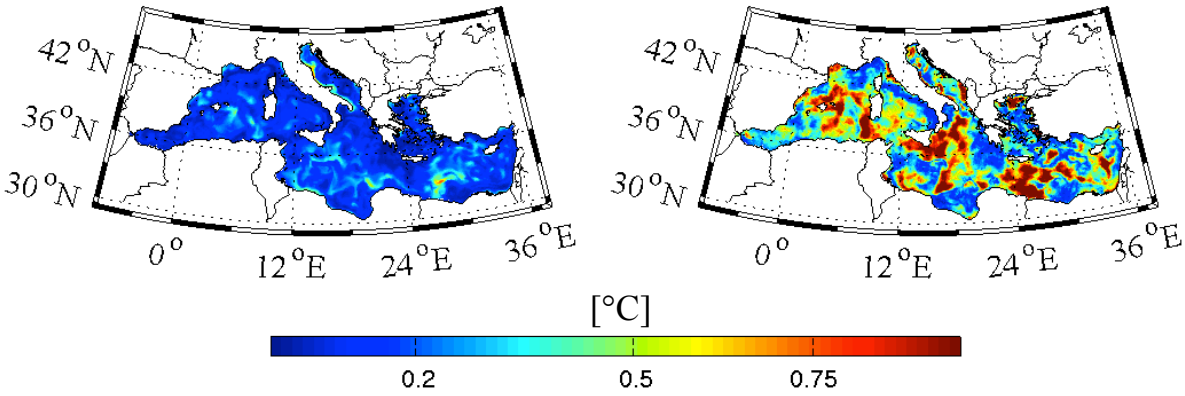

Figure 2.5: *SST Standard Deviation, STD_E, of MPE-2 ensemble (left) and the Mean Absolute Difference between Ensemble Mean and the Satellite SST (right), from the 20th of February to the 1st march 2008*

2.1.3 Multi-Model Ensemble members

A third ensemble dataset that considers different models and implementations has been created basing on the extensive data set of analyses produced in the MyOcean project (www.myocean.eu). The members of the dataset, briefly listed in Table 2.3, come from: Istituto Nazionale di Geofisica e Vulcanologia (INGV), Hellenic Centre for Marine Research (HCMR) and MERCATOR Ocean. The data set build here is called Multi-Model Ensemble 1 dataset, MME1, and a description of the different models available in the consortia follows below. All those numerical models are solving the primitive equations with assimilation of real observations but with different choice of codes, vertical and horizontal discretization and bathymetry. Specifically:

- INGV SYS3a2 & SYS4a3 Daily mean Output; Both the system use the same data assimilation technique, 3DVAR (Dobricic & Pinardi, 2008). The major differences between these kinds of products regard 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 °resolution global model Drevillon *et al.* (2008).

- HCMR : This model is based on the Princeton Ocean model (POM), which is a primitive equations free surface ocean model, operating under the hydrostatic and Boussinesq approximations. Details on the implementation and data assimilation scheme can be found in Korres *et al.* (2009).

- MERCATOR: Prototype SYstem 2 Version 3 (PSY2V3) and Prototype SYstem 2 Version 4 (PSY2V4) daily mean output have been provided by Mercator Ocean Team (Toulouse, France). The main differences between the members concern the version of the numerical code and the boundary conditions (closed and relaxed to climatology in PSY2V3, while PSY2V4 is nested in the global

MMSE Member	*Vertical scheme*	*Diffusion*	*Viscosity*	*Assimilation*
SYS3a2	*P.P.*	Bilap.	Bilap.	3DVAR + SST nudging
SYS4a3	*P.P.*	Bilap.	Bilap.	3DVAR + SST nudging
Mercator V0	$k-\epsilon$	Laplacian.	Bilap.	SAM
Mercator V1	$k-\epsilon$	Laplacian	Bilap.	SAMv2
HCMR(POM)	$k-l$	Laplacian	Laplacian	SEEK filter

Table 2.3: *Set up of MultiModel Ensemble MME1 obtained by MyOcean database*

model). Details on the operational systems set up can be found in Tranchant *et al.* (2011).

All members have been interpolated on the same reference grid ($^1/_{16}$ resolution). All the simulations have been forced by ECMWF analysis fields except HCMR which was forced with hourly momentum, heat and freshwater fluxes derived from the POSEIDON weather prediction system based on a 1/20 °resolution, ETA regional non-hydrostatic atmospheric model.

In figure 2.6 are presented the variance for the 5-members ensemble and the observations. The main differences are related to the amplitude; In some area, like as Gibraltar Strait and the Dardanelles, the variability is underestimates by the Multi-model. Reversing, we can note the strong ensemble variability near Gulf of gulf of Gabès and in proximity of Po mouth advected and spreaded by the Western Adriatic Current.

Generally, the spread (see fig.2.7) over the considered domain is larger than the Mean Differences, confirming that we need to use the multi-model ensemble in order to reduce the systematic error.

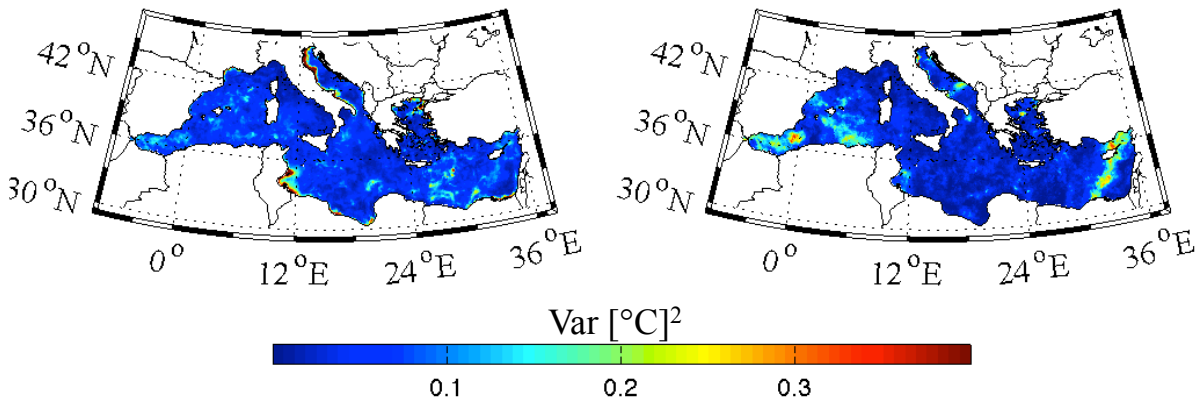

Figure 2.6: *Mean Variability V_E in the Multi-model ensemble MME1 (left) compared with the natural variability V_O in the observations (right) for the SST field, from the 20th of February to the 1st march 2008.*

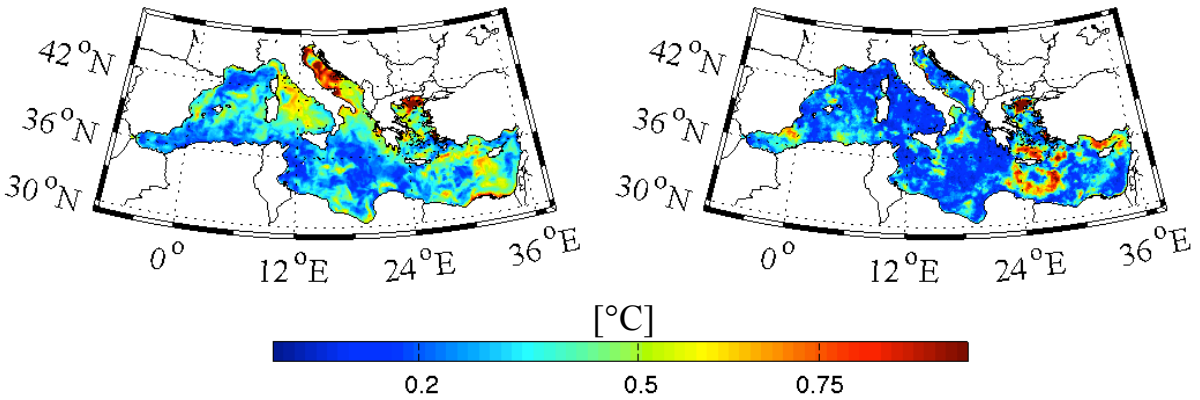

Figure 2.7: *Standard Deviation STD_E of Multi-model members MME1 (left) and the Mean Absolute Difference between Ensemble Mean and the observation, SST from satellite, (right) from the 20th of February to the 1st march 2008.*

2.2 Multi-Model Multi-Physics Superensemble

A mixture of the previous datasets has been done to improve the reliability of the ensemble datasets beforehand described, and to increase the ensemble spread.

2.2.1 Multi-Model Multi-Physics Superensemble Data Set 1

The dataset is created using two different models together with the multi-physics ensemble produced by the BHM method described in section 2.1.2. Concerning the standard ensemble member, one is the previous operational numerical model implemented at INGV, OPA 8.2, in simulation mode (without assimilation of true observation) and forced by analysis of ECMWF forcing (Tonani *et al.*, 2008). The latter member is the actual operational model in simulation mode, NEMO 2.3 (Oddo *et al.*, 2009). Starting from the analysis of ECMWF we produced 10 different realizations of wind using the MFS-BHM winds. The table 2.4 shows the major differences between the ensemble members, hereafter called MPMM-1(Multi-Physics Multi-Model number 1).

	Target	Ensemble Members	
Product	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 2.4: *Set up of Multi-Physics Multi-Model Ensemble MPMM1 Experiment, the columns are the members, while the rows are the difference between each member.*

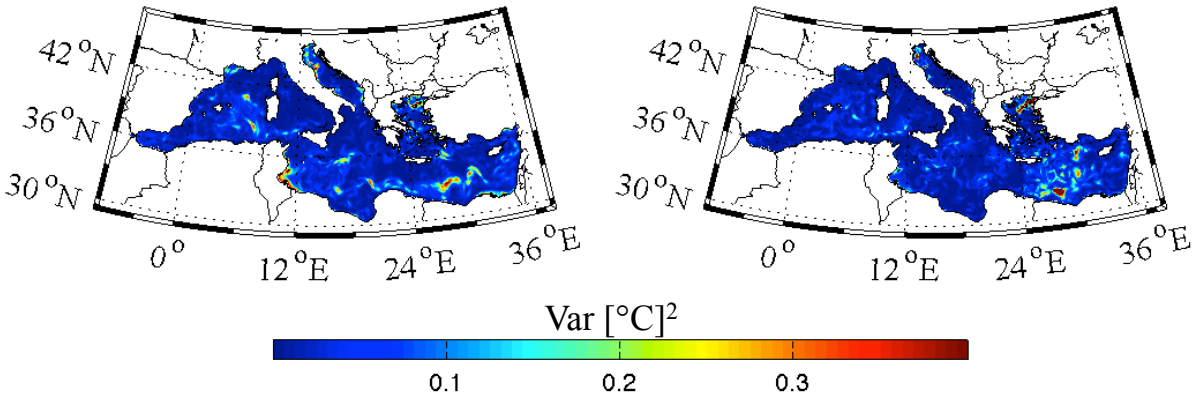

Figure 2.8: *Mean Ensemble Variability V_E in MPMM1 dataset (left) versus the variability in SYS3a2 (right) from the 20th of February to the 1st march 2008*

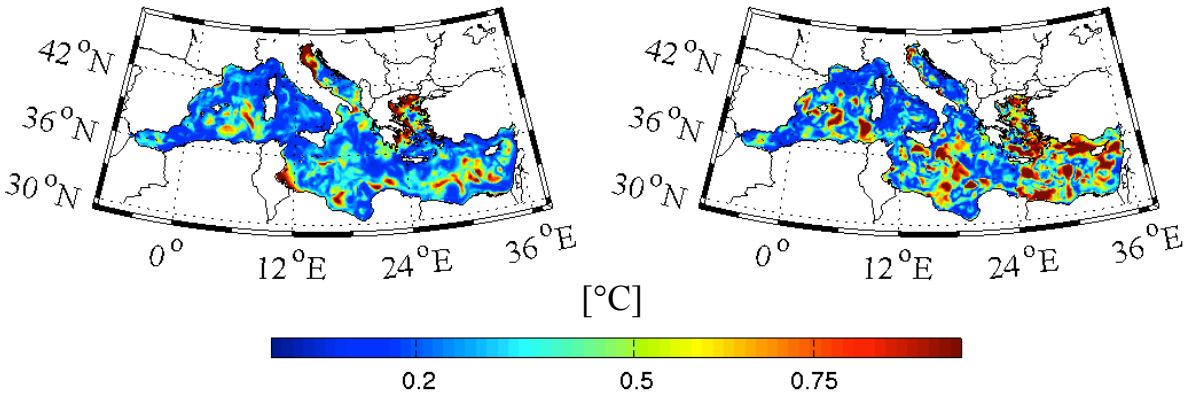

Figure 2.9: *Standard Deviation STD_E of MPMM1 ensemble (left) and the Mean Absolute Difference between Ensemble Mean and SYS3a2 SST (right) from the 20th of February to the 1st march 2008.*

The map of ensemble variability shows in figure 2.8 is pretty similar to the variability of the target field, which in this case is the operational SST produced by SYS3a2. We can note that for this experiment, the strong variability areas are correctly reproduced on the ensemble variability map, and in some isolated cases, for example near Gulf of gulf of Gabès, ensemble variability is bigger than the SYS3a2 variability. The map of spread (fig. 2.9) has the right magnitude (bigger than the previous experiments) but shows low pattern correlations with Mean Absolute differences, for example over Mersa-Matruh or near Rhode area, there are big differences between the ensemble mean and the observations and over the same regions, the ensemble could not generate enough spread.

2.2.2 Multi-Physics Multi-Model Superensemble Data Set 2

The dataset is the merging of the extensive dataset provide by MyOcean project and some *ad hoc* members expressly created with the Multi-Physics approach. The atmospheric forcing were taken from the ECMWF analysis dataset (again, excluding HCRM which used its own atmospheric field). The ocean initial condition were different for each model except for the multi-physics members that started form the same conditions given by the default model run used in the multi-physics ensemble.

Ocean observations have been assimilated in each member with its own assimilation scheme. The main differences in the dataset are shown in table 2.5, hereafter called MMMP-2 (Multi-Physics Multi-Model number 2). However in order to have an homogeneous comparison we evaluate variability and spread in the same period of the previous cases.

The comparisons of the results are appealing. Multi-model Multi-Physics Ensemble 2 is able to reproduce the variance of the observations in the last ten days of February 2008, see fig.2.10. The maps are very similar with a very good pattern correlation. Very good correlation can be found even (fig.2.11) between the spread and the Mean Absolute Differences, and in some region, for example in the Levantine basin, the spread in the ensemble is bigger the difference between observation and ensemble mean, in this way we should supply the lack in variability observed in the same area. This is a practical example in which the multi-model and multi-physics

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

Table 2.5: *Set up of Multi-Physics Multi-Model Experiment MPMM 2 obtained by MyOcean database (on the top) and the multi-physics members (on the bottom of the table) from 20th of February to the 1st march 2008*

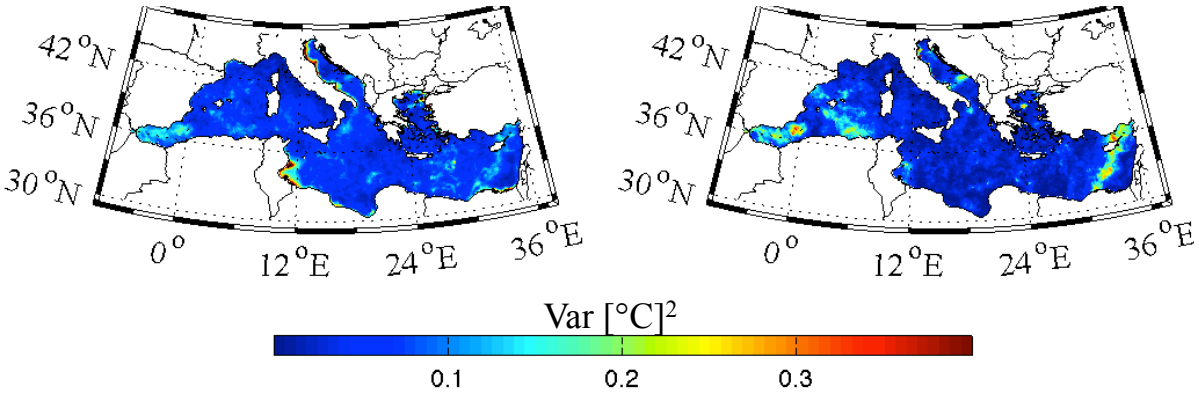

Figure 2.10: *MPMM 2 Mean Ensemble Variability V_E (left) and the natural variability V_O in the observations (right) from the 20th of February to the 1st march 2008*

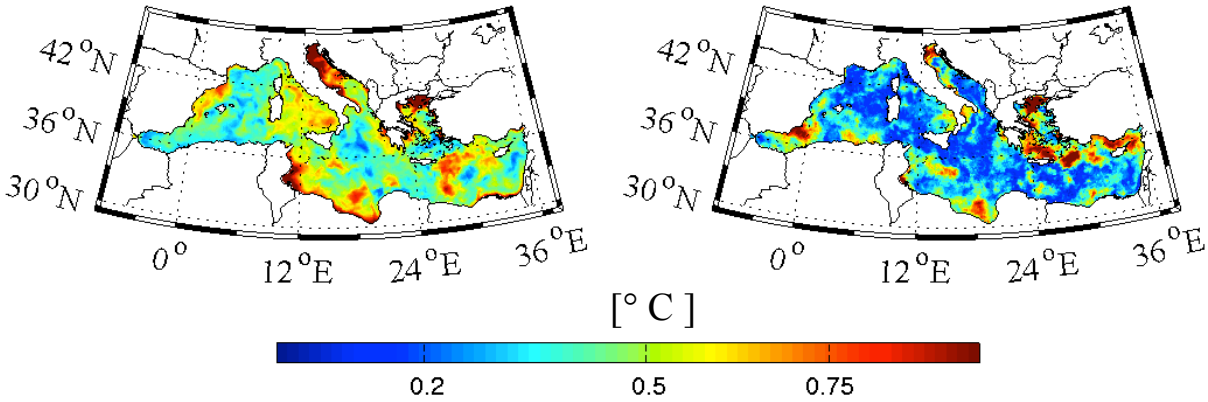

Figure 2.11: *Standard Deviation of MPMM 2 ensemble members(left) and the Mean Absolute Difference between Ensemble Mean and the observation(right), from the 20th of February to the 1st march 2008 .*

members can be combined in order to reduce their relative error. In this condition we should be able to proceed with a reliable bias reduction operation.

2.3 Discussion

We have presented a scheme for the generation of ensemble of analysis that could be used to develop a superensemble estimate. Usually SE performances are evaluated in a test configuration, hence this the first time that an assessment on the potential skill of the dataset is being proposed, studying the dataset behaviour during the training phase. We have described the ensemble components and examined how errors from different sources get mixed in a non linear fashion. To obtain an estimate of datasets strengths and imperfections, we studied how well the ensemble variability and spread corresponded to the observed values, evaluating the correlation between the maps. Since good correlation does not guarantee accurate prediction, we also calculated the difference between the mean values, which is a sort of bias measure. In a perfect ensemble, we should have correlations one for both, zero difference in the mean difference for the variability and the discrepancy in spread and mean error bigger than zero but only with the MPMM ensembles we were able to satisfy this option.

In general we observed an underestimation of the variance during the period

of interest, this situation is also confirmed from the analysis of figure 2.12, where differences between the mean variability, $V_E - V_O$, are always negative, meaning that all the ensembles are underestimating its real magnitude, supporting the low-pass filter behaviour in numerical models. As expected the MPMM 2 dataset is the best ensemble since it has the biggest correlation coefficients and the smaller bias for variability and bigger value for the difference with spread and mean ensemble error. Quite surprising , we can note that the best performance in variability correlation is reached in the MPE 1, even if it shows the bigger bias between them. Furthermore we can note the impact for the ensemble built with different atmospheric forcing, adding other 11 realizations with a different model, we get an increase in pattern correlation between the spread and mean error abut we penalize the correlation in variability and the true values of the scores since we get bigger mean values difference between ensemble and observations.

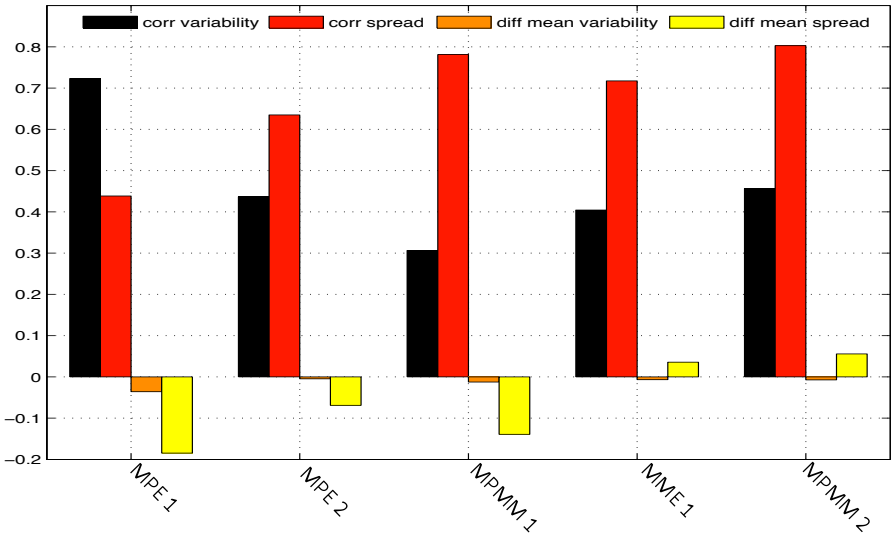

Figure 2.12: *Correlation coefficients between the enseble variability and observations variability(black bars), correlation coefficients between the enseble spread and MAD(red bars) difference in the domain avarage ensemble variability and observations variability (orange bars) and difference in the domain avarage spread and MAD(yellow bars).*

Chapter 3

Classical SuperEnsemble method

3.1 The Krishnamurti method

Upon the availability of ocean predictions produced by several numerical models, multi-model ensemble forecasting has attracted a lot of attention recently (Krishnamurti *et al.* 2000 , Yun *et al.* 2003, Berliner & Kim 2008 and Lenartz *et al.* 2010). Several methods combining different model forecasts were developed to produce a forecast ensemble estimate, but in this section we will describe in details the SuperEnsemble (SE) method proposed by Krishnamurti *et al.* (1999) hereafter called Krishnamurti method.

Let's call S_t the SE estimate of a model state variable, $F_{i,t}$ and let's define two different periods, the training and the test periods. The former is a time period preceding in time the target period of estimation, while the second is the period where the SE estimate is needed. Krishnamurti method is then defined as:

$$S_t = \bar{O} + \sum_{i=1}^{N} a_i \left(F_{i,t} - \bar{F}_i \right) \tag{3.1}$$

$$\overline{O} = \frac{1}{M} \sum_{t=0}^{M} O_t \tag{3.2}$$

where $F_{i,t}$ is the $i-th$ model forecast available both for the training and test period, t, $\bar{F}_i$ is the mean of the $i-th$ model over the training period, $\bar{O}$ is the mean of the observations O_t over the training period, a_i are regression coefficients obtained by an appropriate minimization procedure during the training period, N is the number of models and M is the number of training period units. The systematic errors of ensemble members in equation (3.1) are removed because anomalies $\left(F_{i,t} - \bar{F}_i\right)$ are used for each model. The regression coefficients (weights) are computed by minimizing the following cost function:

$$G = \sum_{t=1}^{M} \left(S_t - O_t\right)^2 \tag{3.3}$$

Substituting (3.1) we obtain:

$$G = \sum_{t=1}^{M} \left(\bar{O} + \sum_{i=1}^{N} a_i \left(F_{i,t} - \bar{F}_i\right) - O_t \right)^2 \tag{3.4}$$

The minimization of eq.3.4 with respect to the unknowns values a_i is found by setting $\frac{\partial G}{\partial a_i} = 0$. Since equation (3.4) contains N parameters, there will be a system of N equations.

$$\frac{\partial G}{\partial a_j} = 2 \sum_{t=0}^{M} \left(\bar{O} + \sum_{i=1}^{N} a_i \left(F_{i,t} - \bar{F}_i\right) - O_t \right) \left[\sum_{i=1}^{N} \left(F_{j,t} - \bar{F}_j\right) \right] = 0; \tag{3.5}$$

Let's define the covariances of the ensemble members as:

$$\sum_{t=1}^{M} \left(F_{i,t} - \overline{F_i}\right) \left(F_{j,t} - \overline{F_j}\right) \equiv \Gamma_{i,j} \tag{3.6}$$

and the vector containing the projection of anomaly observations with the individual models anomalies as:

$$\sum_{t=1}^{M} \left(F_{j,t} - \overline{F_j}\right) \left(O_t - \overline{O}\right) \equiv \phi_j \tag{3.7}$$

Thus (3.5) reduces to a linear system:

$$\vec{a} \cdot \Gamma = \vec{\phi} \tag{3.8}$$

where $\vec{a}$ is the vector of regression coefficients (the unknowns). In the conventional superensemble approach, the regression coefficients are obtained using Gauss-Jordan elimination with pivoting:

$$\vec{a} = \vec{\phi}\Gamma^{-1} \tag{3.9}$$

The covariance matrix Γ and $\vec{\phi}$ are rearranged into a diagonal matrix, and the solution vector is obtained as in eq.(3.9) 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 or in other words the determinant of the matrix is close to zero. This happens when highly correlated observations are used. We noted that enlarging the training period we could reduce the degeneracy of the covariance matrix since this is equivalent to introduce some noise in the input data.

We note that, if there are fewer equations than unknowns, the regression coefficients are under-determined. A simple *"toy experiment"* has been done using the truth as one of the ensemble members. This test, can be considered as the maximum skill that could be achieved with a multi-model, and it is also the only way to control the perfect multi-linear combination coefficients estimates. As expected, all the regression coefficients are zero, except the weight related to "truth" member, which is set to 1. Trimming the dataset (removing members) we noticed that when the training period units (M) are less than the number of multi-model members involved(N), the algorithm fails, giving incorrect values for the coefficients. So, we can conclude that the regression procedure is not able to identify the "best" member when it is trained for a short time period. Hence a minimum constraint for the training period units must be:

$$M \geq N \tag{3.10}$$

Some studies, (Krishnamurti *et al.*, 2000), have interpreted the regression coefficients as indicators of the relative model "*reliability*". However, this interpretation

needs some clarifications as Kharin & Zwiers (2002) enunciated. 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. In this specific case we set O as 30 days of SST from satellite over the Mediterranean Sea, from the 1st March 2008 to 30th march 2008. We built M_1 and M_2 as $M_1 = 1.5 \cdot O + \epsilon_1$ and $M_2 = 0.5 \cdot O + \epsilon_2$. In this situation,$\epsilon_1$ and ϵ_2 represent the unpredictable internal variability that is present in the corresponding model estimates. Without any loss of generality, we can assume that these are independent random variables with 0 time mean and the variance equal to σ_O^2. The members will have the same RMSE during the training period, equal to: $0.25 \cdot \sigma_O^2$. If we solve the system 3.9 we can find that $a_2 = 1/4 \cdot \sigma_O^2$ while $a_1 = 3/4 \cdot \sigma_O^2 = 3a_2$ Thus we infer that *"equally reliable model outputs may not necessarily be weighted equally when combined optimally"*.

3.2 Verification scores

A wide range of possible scores are available in order to asses the quality of a SuperEnsemble estimate, 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):

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

Root Mean Squared Error (RMSE) is defined as:

$$RMSE = \sqrt{\sum_i^N \frac{(F_i - O_i)^2}{N}} \tag{3.11}$$

Since RMSE contains the square quantities it is more sensitive to large forecast errors than other scores. For this reason it is possible to subtract the estimated model bias centering the error of the ensemble mean, i.e.:

$$RMSE_c = \sqrt{\sum_i^N \frac{\left(F_i - \overline{F}\right)^2 - \left(O_i - \overline{O}\right)^2}{N}} \tag{3.12}$$

The generalization of the Mean Square Error:

$$MSE = \sum_i^N \frac{\left(F_i - O_i\right)^2}{N} \tag{3.13}$$

is also (see Murphy & Epstein, 1989) the Mean Square Error Skill Score,

$$MSESS = 1 - \frac{MSE_{SE}}{MSE_R} \tag{3.14}$$

$MSESS$ equal to 1 indicates the perfect forecast, and it is zero as the forecast skill is equal to the climatology. A negative value implies that the SE forecast is worst than the other ensemble members. It is dimensionless and increase with the forecast skill. Another score is the Anomaly Correlation Coefficient, which is the correlation between the forecast and the observed anomaly. We used the centered version of it (3.15) in order to take in account the bias:

$$ACC_c = \frac{\overline{\left[\left(F_i - \overline{O}\right) - \overline{\left(F_i - \overline{O}\right)}\right]\left[\left(O - \overline{O}\right) - \overline{\left(O - \overline{O}\right)}\right]}}{\sqrt{\left(\left(F_i - \overline{O}\right) - \overline{\left(F_i - \overline{O}\right)}\right)^2 \left(\left(O - \overline{O}\right) - \overline{\left(O - \overline{O}\right)}\right)^2}} \tag{3.15}$$

3.2.1 Multi-Physics SuperEnsemble retrieved from MPE 1

Here we will apply the Krishnamurti method to the multi-physics ensemble members generated using a variety of physical parametrizations as described in section 2.1.1. Here we will show the results after the verification procedure. We decided to use the SST satellite-derived measurements (Marullo *et al.*, 2007) as the target field. In this

way we set the unbiased estimator as the time mean of those SSTs during a training period chosen to be from the 7 January to 31 January 2004. During this period,

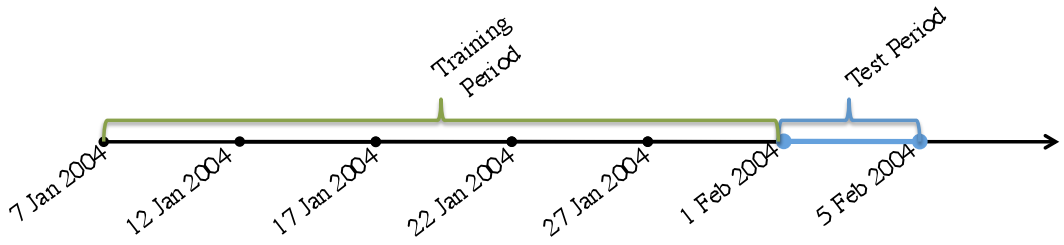

Figure 3.1: Schematic representation of the MPE 1 experiment set up.

we estimated regression coefficients according to eq.(3.9). For all the days during the test period,1-5 February 2004 (see fig.3.1), SE estimates were very noisy and unrealistic (not shown). We assessed the performance of the SE using the RMSE according to the formula (3.12). For each day in the test period, every multi-physics member is better than the SE prediction (not shown).
The bad performance of SE predictor is due to the *rank-deficiency* of the covariance matrix Γ. The determinant is approximately zero, since the members differ only by *small amplitude anomalies*. Thus we conclude that the MPE1 dataset has not enough spread to recover estimates of the regression weights in Krishanmurti algorithm.

3.2.2 Multi-Physics Multi-Model SuperEnsemble MPMM 1 Experiment

The dataset used in this experiment was generated by 22 runs with MFS BHM-winds over the Mediterranean area as described in section 2.1.2.
The experiment set up is the following:

1. The target field is chosen to be:

   - Horizontal maps of temperature, the SSTs from operational model output SYS3a2 (Tonani *et al.*, 2008), which are the temperature of the first

layer as evaluated by the model, corrected throughout satellite based SST values.

2. The training period is chosen to be:

   - 1 February-31 March for years 2005 2006 2007 and 1-29 February 2008;

3. The test period is chosen to be:

   - 1-31 March 2008;

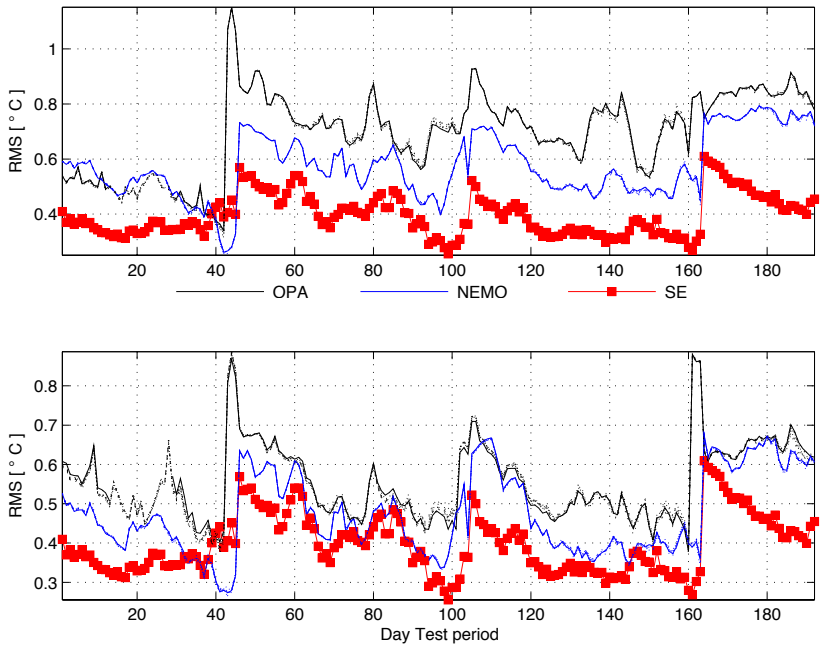

Figure 3.2: *Domain average RMSE (top panel) $RMSE_c$ (lower panel) values for the SE experiment done with MPMM1 dataset.*

First we checked that the algorithm work well during the training period. The results (fig.3.2) confirm that the SE weights were calculated well for most of the training period both in terms of Root Mean Squared Error and $RMSE_c$. The domain used to calculate the average is the Rhode Gyre region in the Levantine Basin, between 26-32 °E and 30.25-38 °N.

During the test period the RMSEc and BIAS are shown in figure 3.3 and 3.4 respectively. These results indicate that the SE estimates is comparable to each member RMSEc and BIAS.

The solid bars depicted in fig.3.6 correspond to the MSESS between the SE and

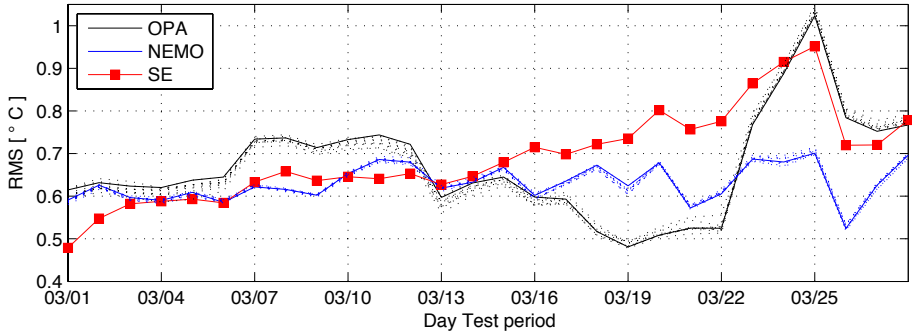

Figure 3.3: *Domain averaged $RMSE_c$ during the test period, 1-31 March 2008, for the MPMM1 ensemble.*

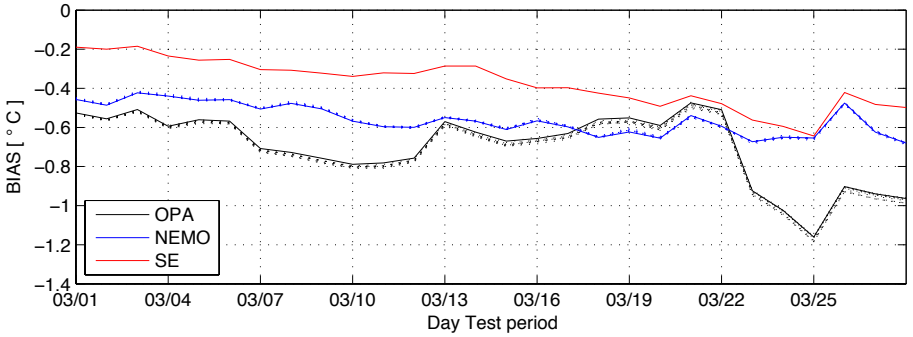

Figure 3.4: *BIAS during the test period,1-31 March 2008, for MPMM1 ensemble.*

each MM members, this comparison displays that for the test period under study, 1-31 March 2008, SE performs better, in term of SST prediction, than all the other ensemble members, since the MSESS is always positive. Usually only one particular diagnostic, the RMSE has been used to demonstrate the good performance of SE, but as shown in fig.3.5 we don't reach the same result changing the diagnostic metric. Hence, a crucial consideration to assess the impact of the multi-model approach, concerns the choice of the diagnostic metric during training period. We can argue

that SE prediction in which the coefficients are retained by least squares method during a training period, can not skilful in diagnostic by ACC metrics.

An examination of the fifth day of the prediction over the Rhode Gyre region

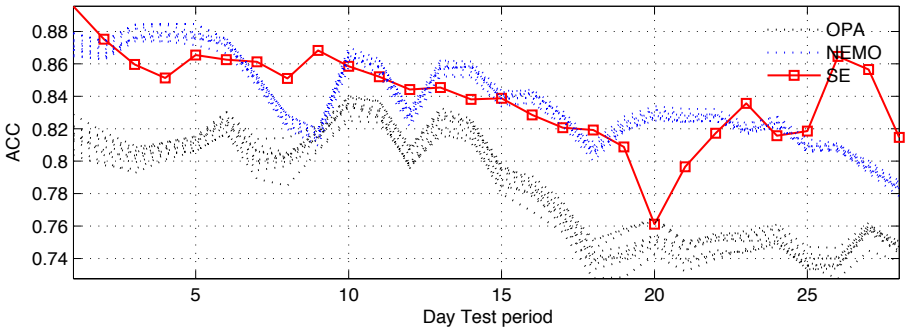

Figure 3.5: *ACC during the test period,1-31 March 2008, for MPMM1 ensemble.*

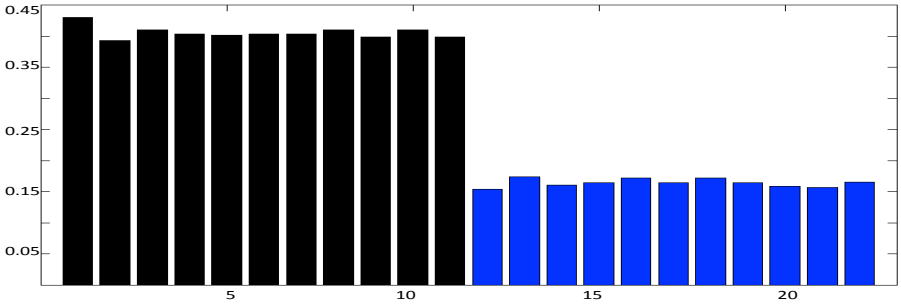

Figure 3.6: *MSESS for MPMM1 ensemble for the SST estimates over Rhode Gyre region during the test period(1-31 March 2008). Black bars represent gain respect OPA predictions, and blue bars the gain respect NEMO. Each bar is the MSESS as evaluated by3.14 with reference the MSE of each member.*

(fig. 3.7) shows some unrealistic features of the SE estimate. The map seems to be affected by overfitting of the ensemble members 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. In our experiment we can note that in the unbiased estimator map some sub-basin scale

features, for example the Mersa-Matruh anticyclonic gyre, (Robinson *et al.*, 2001) *"disappeared"*. When the training period covers different years, is common to observe a shifting (and consequently a compensation if we evaluate a mean of cyclonic and anti-cyclonic gyres) in circulation patterns, which can not be any more reproduced in the unbiased estimator map. Additional filtering techniques are needed (and described in the following chapter 4) to avoid this problem.

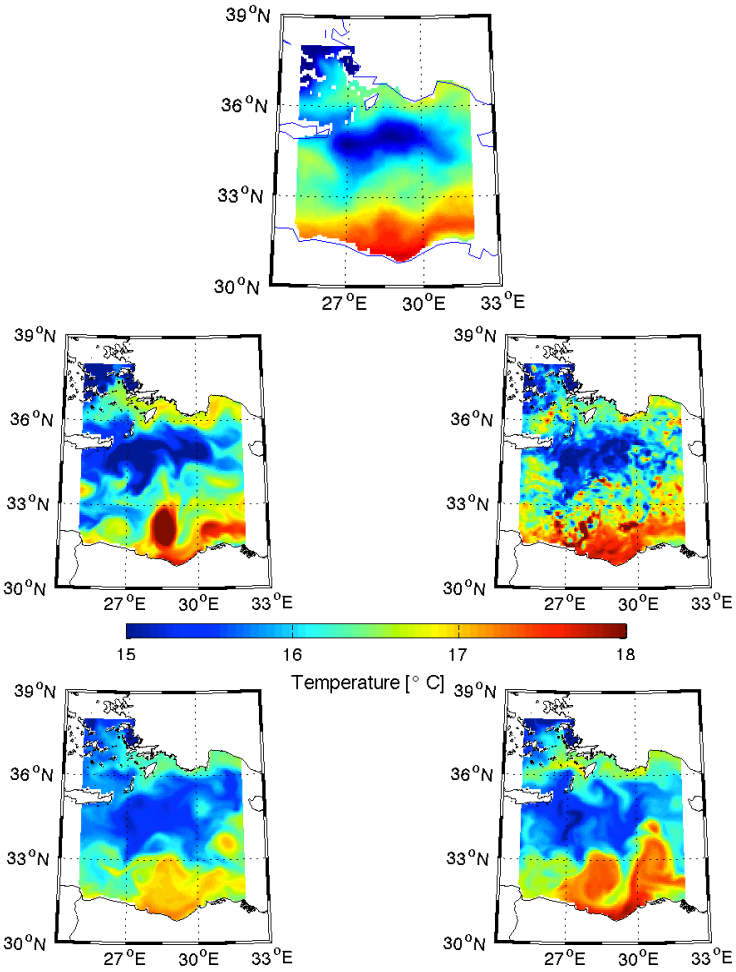

Figure 3.7: *An example of a SE SST estimate on the fifth day of the test period, over Rhode Gyre region:Observational Mean (top center), analysis (middle left), lowest skill model (bottom left), SuperEnsemble (middle right) and best ensemble member (bottom right).*

3.2.3 Multi-Physics Multi-Model SuperEnsemble MPMM2 Experiment

This experiment had been set up to investigate the sensitivity of the algorithm to the training period length and the ensemble member dataset quality. In this experiment, we have used the MPMM2 ensemble dataset described in section 2.2.2.

1. The target field :

   - Horizontal maps of temperature, the Sea Surface Temperature from Satellite.

   We fix the test period as ten days, we call the first day of the target period TP1(Test period 1) and so on to TP10=TP+1. Furthermore, we call the number of days in the training period TR. With this nomenclature we set:

2. The training period:

   - TP1-TR to TP1-1;

3. The test period:

   - TP1 to TP10 ;

In order to assess the seasonal dependence of the technique, the dataset time period covered a whole year, from the 1st of January to the 31 December 2008. For example, when we want to predict the SST the 1st of June 2008, the exercise was performed using maximum n cases, with training period set as TR days before the first day, TP1. As we had shown at the end of section 3.1, 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 $TP1$ we could have 46 (60 has been set as the maximum training period length, minus the threshold value, 14) SE realizations trained by the corresponding training period(see 3.8).

The SST SE estimate, figure 3.9, show that, a range of 30-40 training days gives the SE is best results, at least for the first 2-3 days of the test period. The SE performance

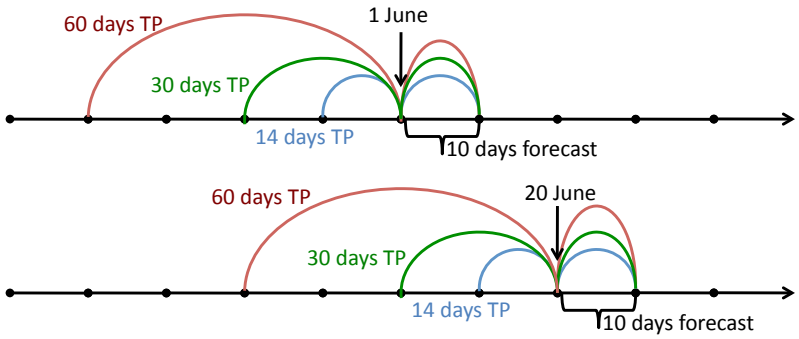

Figure 3.8: *Schematic representation of the experiment carried with the MPMM2 ensemble to find the optimal training period length.*

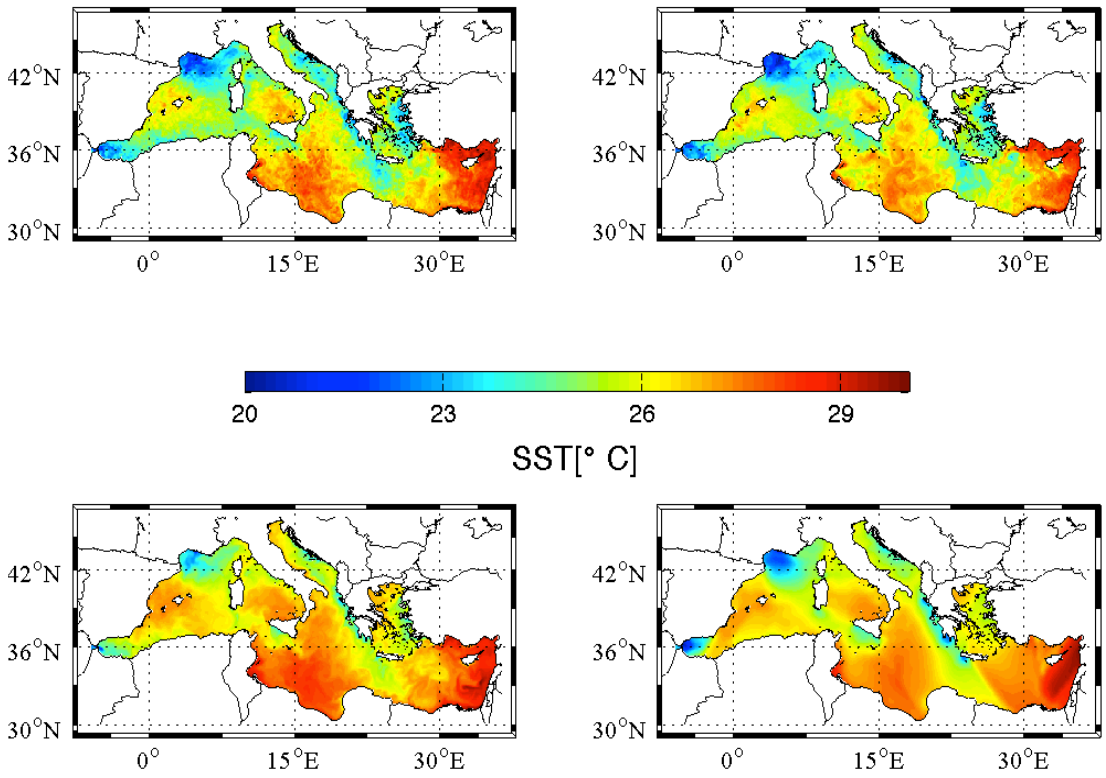

Figure 3.9: *A typical example of prediction for the first day of test period over the Mediterranean area, valid on the 1^{st} of June 2008 reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)*

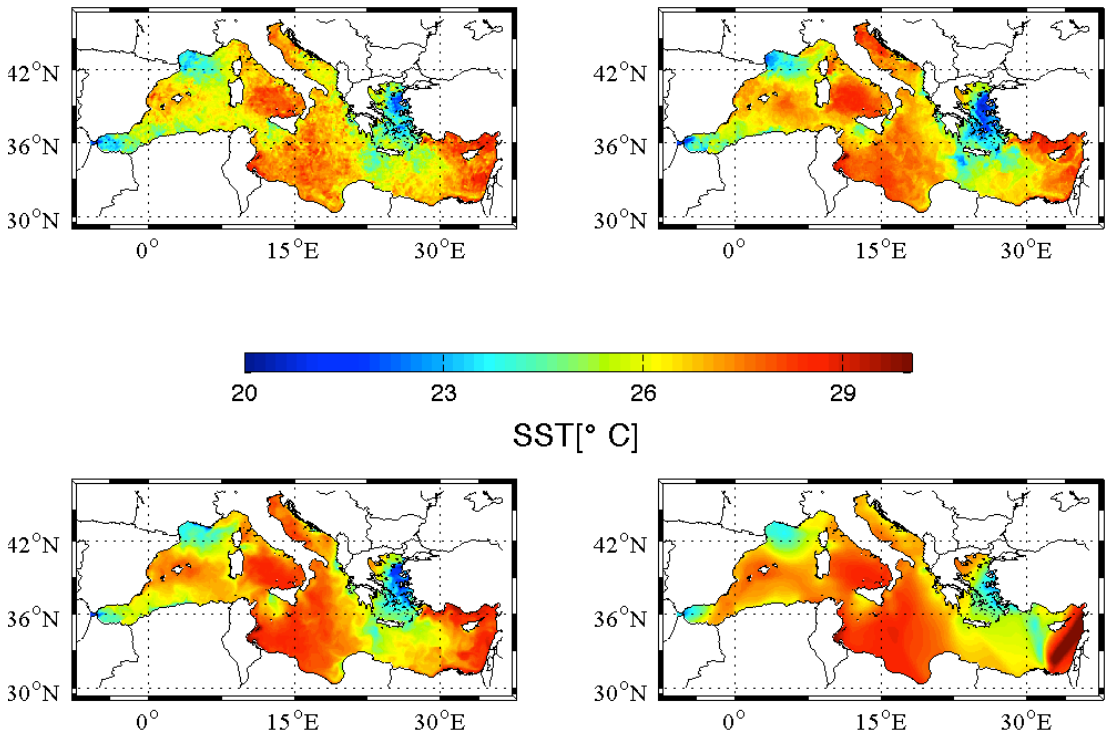

Figure 3.10: *A typical example of prediction for the last day of test period over the Mediterranean area, valid on the* 10^{th} *of June 2008, reached with 41 days of training period, SE prediction (top left), best participating model (bottom left), SST Satellite derived (top right) and lowest skill model (bottom right)*

deteriorates in time; for example, depicts the resulting map of SST for the first day of the test period, 1 June 2008, obtained with 41 days of training period, while figure 3.10 depicts the resulting maps for the 10 June 2008. We can observe that, in the latter figure, the noise increased in SE realization and the SE is not any more the best estimate. This is reflected in the domain average centered RMSE, which shows a higher skill for the SE estimate up to the day 3 with respect tp the RMSE of the ensemble members (fig.3.11). The ACC score (fig.3.12), show again that up to the second day of the test period, SE can be better than all the ensemble members. In fig.3.13 we can see the better performances of SE prediction in terms of bias until the fourth day. The best model in centred RMSE term is almost the best also in ACC and BIAS ranking. We can conclude that, for MPMM2 ensemble, the features of the ranking are almost consistent across the whole range of used skill measures. The previous results were reached for an experiment obtained with 41 training days. We studied the average behaviour of SE predictions analysing all the experiments carried out during the whole year 2008. The first thing to know is that as expected, for all the participating ensemble members, their performances are constant during the test period (see figure3.14). Average of SE predictions can reproduce the performances of the best participating models, but the associated standard deviation, let us infer that the performances are highly variable according the training period. For example, the average SE predictions when obtained with 60 days of the period (fig.3.15) shows that SE can outperform all the participating models for the first two days of the test period. Conversely, for shorter training period, 14 days (fig3.16), the SE is the worst predictor during the whole test period. As for RMSE, we evaluated the average of ACC (fig.3.17) the averages obtained with 60 or 14 days of training period (fig.3.18 and fig.3.19) and the and BIAS time series during year 2008(fig.3.20). In order to study the variability, we also evaluated the average of the experiments obtained with the maximum and minimum training period (fig.3.21 and fig.3.22). ACC show a similar behaviour to the RMSE performances. While for bias no particular inference can be drawn from the analyses of SE performances for year 2008. We can only confirm the unbiased behaviour of SE prediction. We can argue that, the best performances (as shown in Krishnamurti *et al.*, 2000) of SE predictions were due

to the unbiased nature of SE prediction. Since Mean Squared Error takes in to account of the standard deviation of the errors and the bias, SE can outperform all the ensemble members if the latter term is approximately zero.

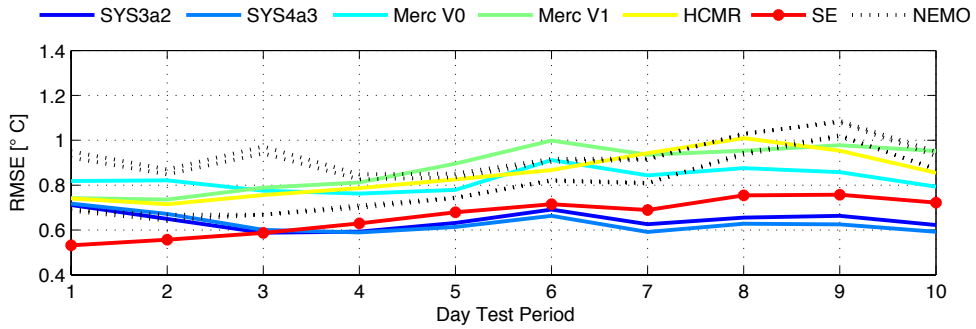

Figure 3.11: *Domain average (over the Mediterranean)Time serie of $RMSE_c$, from the 1st to the 10th June 2008*

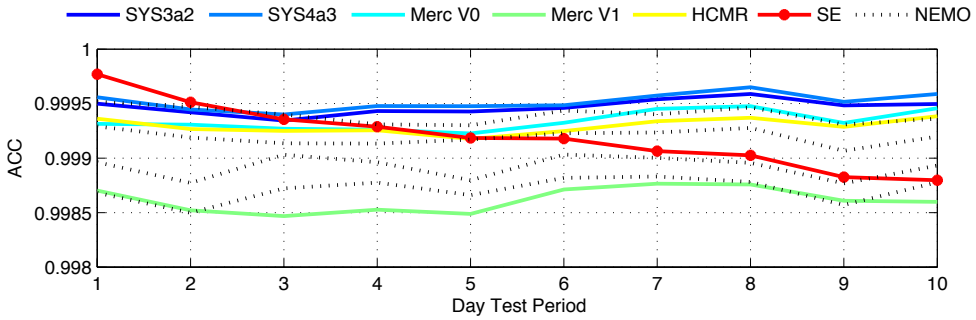

Figure 3.12: *Anomaly Correlation Coefficient time series, from the 1st to the 10th June 2008*

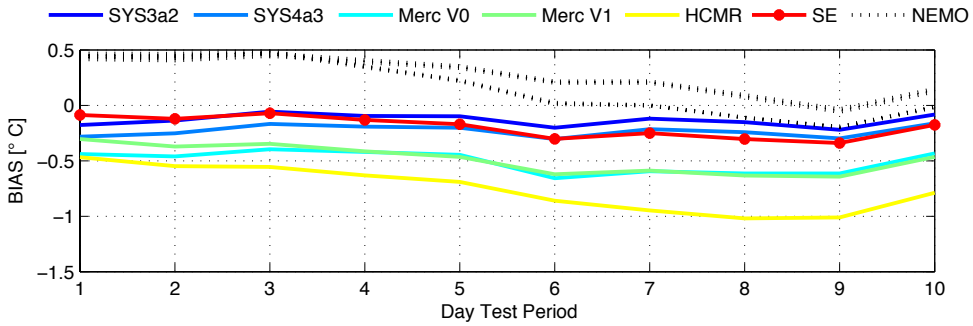

Figure 3.13: *Bias (Model - Observation) Time serie over the Mediterranean sea, from the 1st to the 10th June 2008)*

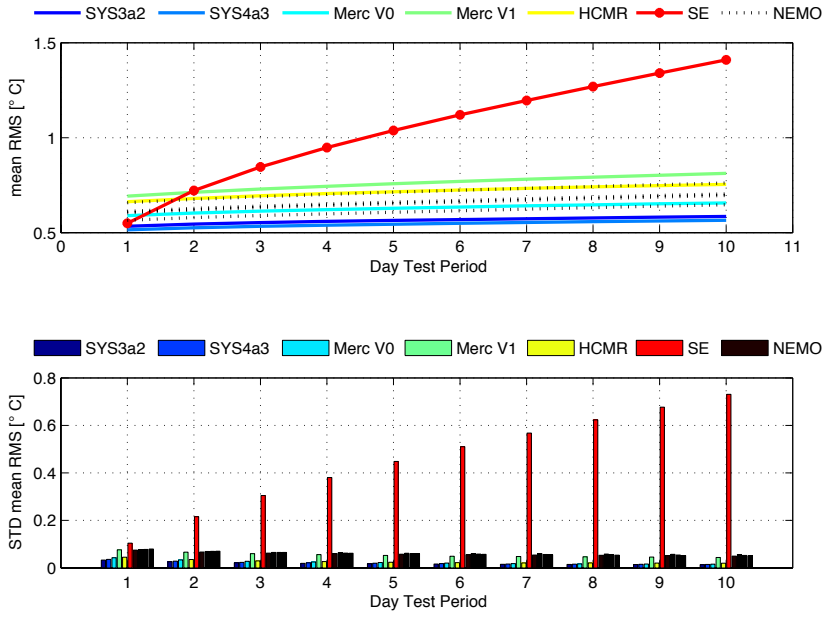

Figure 3.14: *Domain average (over the Mediterranean) time serie of $RMSE_c$ (whole 2008, all the experiments)*

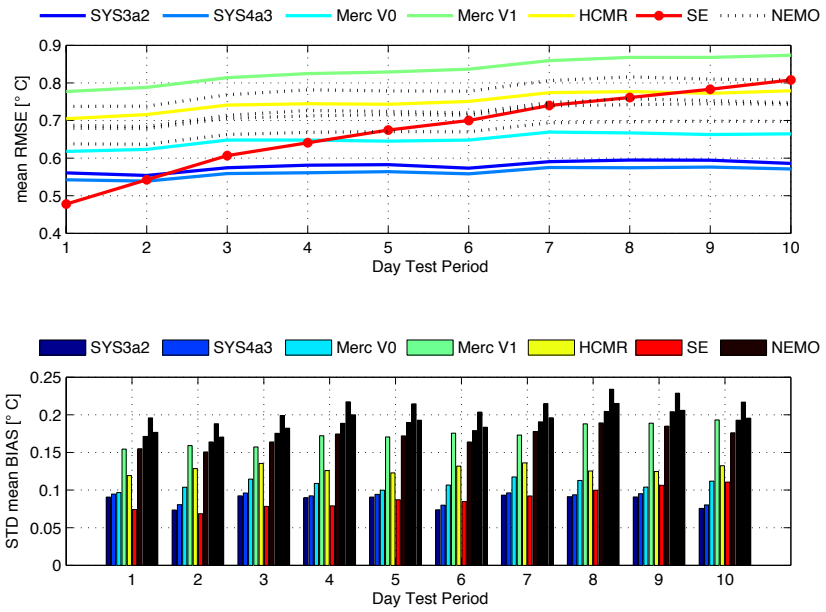

Figure 3.15: *Domain average (over the Mediterranean) time serie of mean $RMSE_c$ all experimets trained with 60 days (whole 2008,all the experiments)*

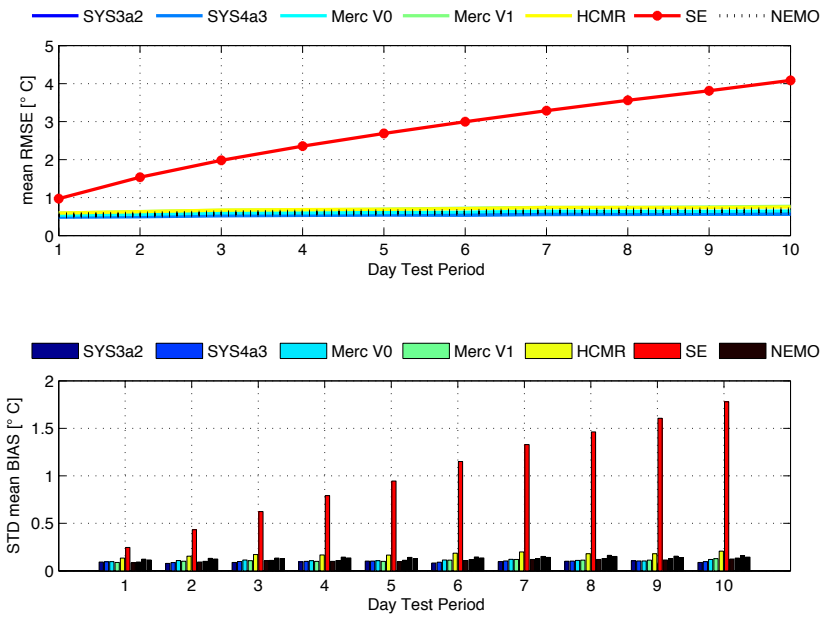

Figure 3.16: *Domain average (over the Mediterranean) time serie of mean $RMSE_c$all experiments trained with 14 days (whole 2008)*

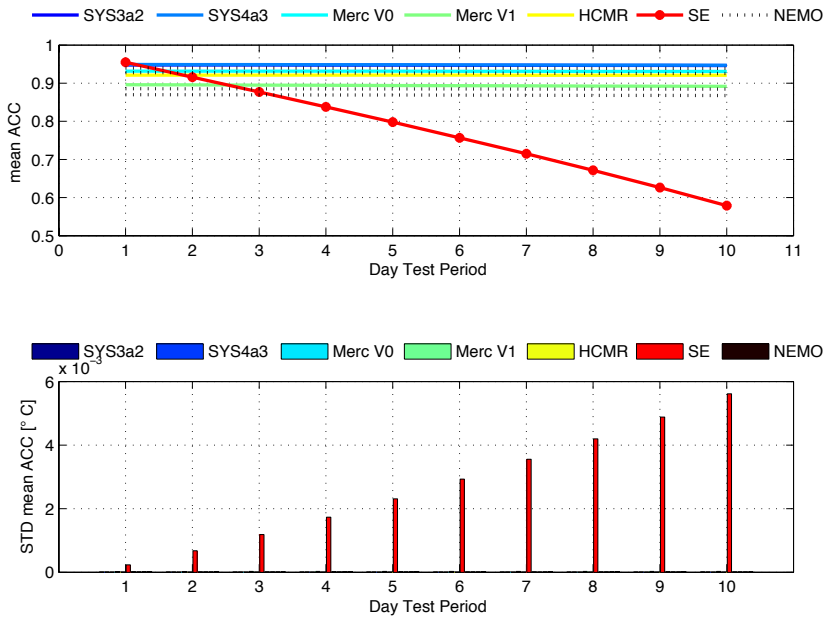

Figure 3.17: *Domain average (over the Mediterranean) time serie mean ACC (whole 2008, all the experiments)*

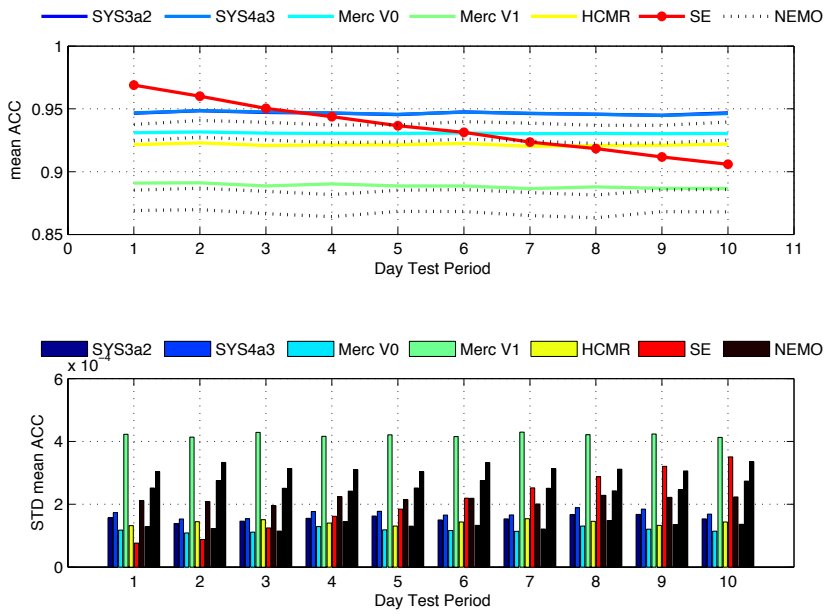

Figure 3.18: *Domain average (over the Mediterranean) time serie mean ACC, all experiments trained for 60 days (whole 2008)*

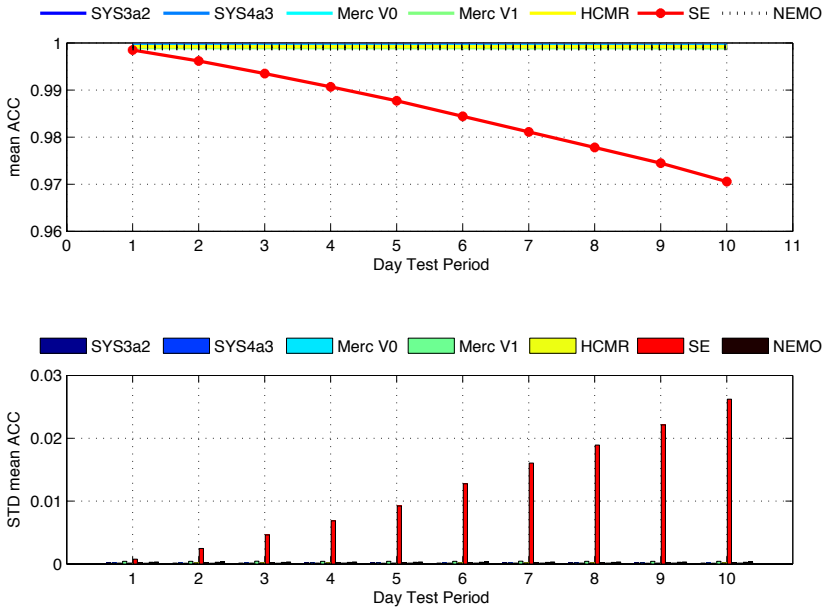

Figure 3.19: *Domain average (over the Mediterranean) time serie mean ACC, all experiments trained for 14 days (whole 2008)*

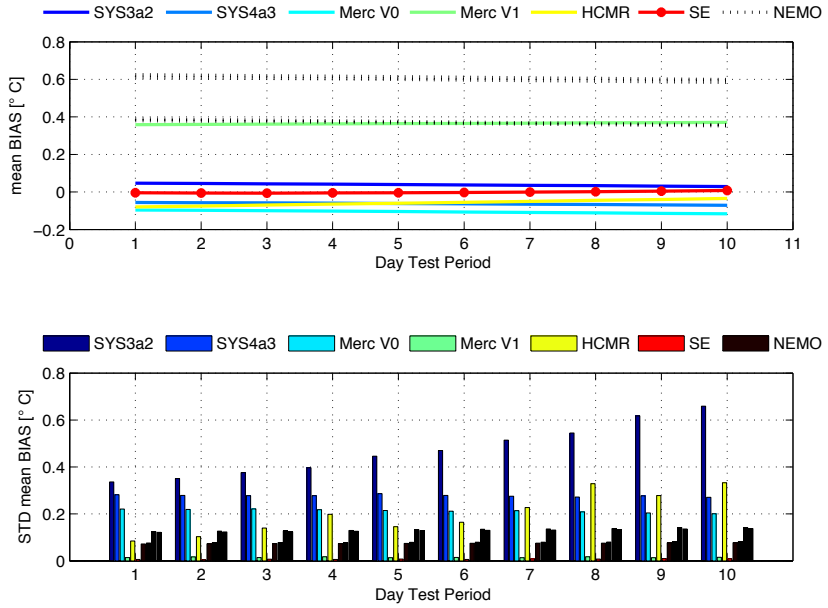

Figure 3.20: *Domain average (over the Mediterranean) time serie mean BIAS (whole 2008, all the experiments)*

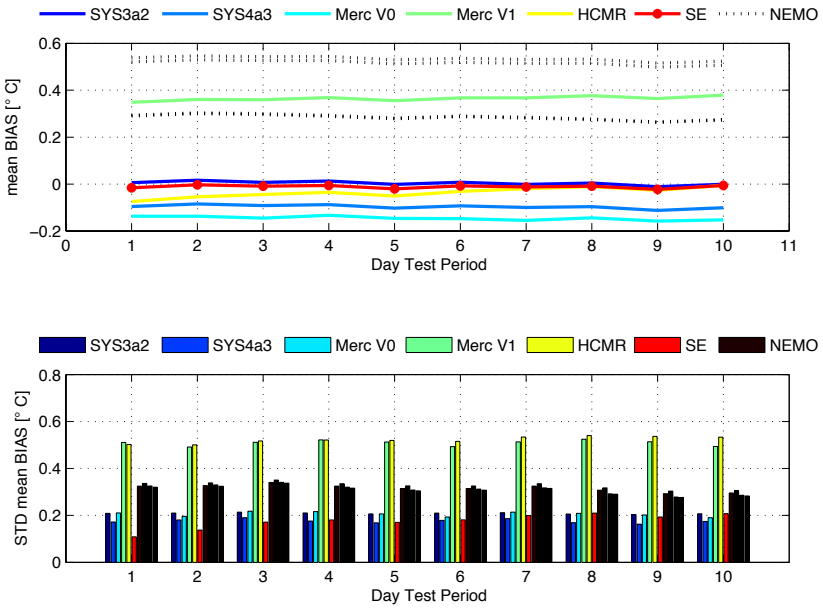

Figure 3.21: *Domain average (over the Mediterranean) time serie mean BIAS, all experiments trained for 60 days (whole 2008)*

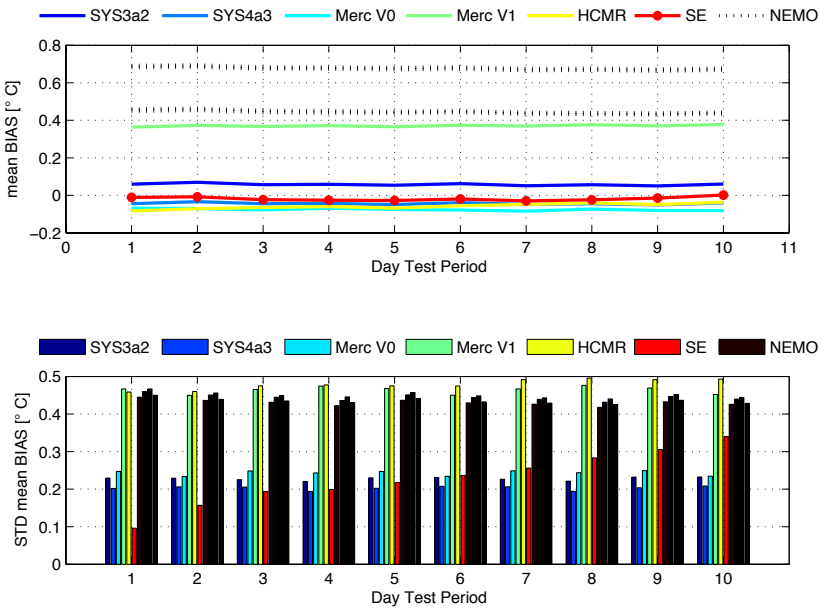

Figure 3.22: *Domain average (over the Mediterranean) time serie mean BIAS, all experiments trained for 14 days (whole 2008)*

3.3 Sensitivity Studies

3.3.1 Sensitivity on the training period length

During year 2008, we carried out 2532 experiments, and in each, we changed the starting day of the training period and its length. Hence, for each test period, we could have several predictions. For example, in the previous plots (for example figure **??**) which show the results for several test periods covering the whole 2008, the SE predictions have been trained for 45 days. During the same test period, we have also other realizations, obtained training the algorithm for a different sample. All those realizations have been over-imposed on the same graph (figures: 3.23, 3.24 and 3.25) in order to assess the dependence with training period. The first important consideration when assessing the prediction scores, is the dependence with the training period length.

As expected the most sensitive score is the RMSE, which curves arise more slowly (becoming almost flat) for a training period longer than 40 days (see fig. 3.23). A key feature displayed in figure 3.24 and figure3.25 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, and for all the following days, the superensemble will not be the best predictor (check fig.3.11). In the remaining last days of the test period, the best performance are obtained by the single-model member. We indicate that day as *"skill"* of the SE prediction, with this definition, using only one index, we could study more easily the evolution of the skill with the training period when compared with other ensemble members. All the skills, for RMSE (fig. 3.26), ACC (fig.3.27) and BIAS(fig. 3.28) used to assess 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 (depicted in figure 3.29), in order to determine the different growth rates. As we can see depicted in figure 3.29, the ACC curve grows faster and reaches a plateau around 40 days of training period.

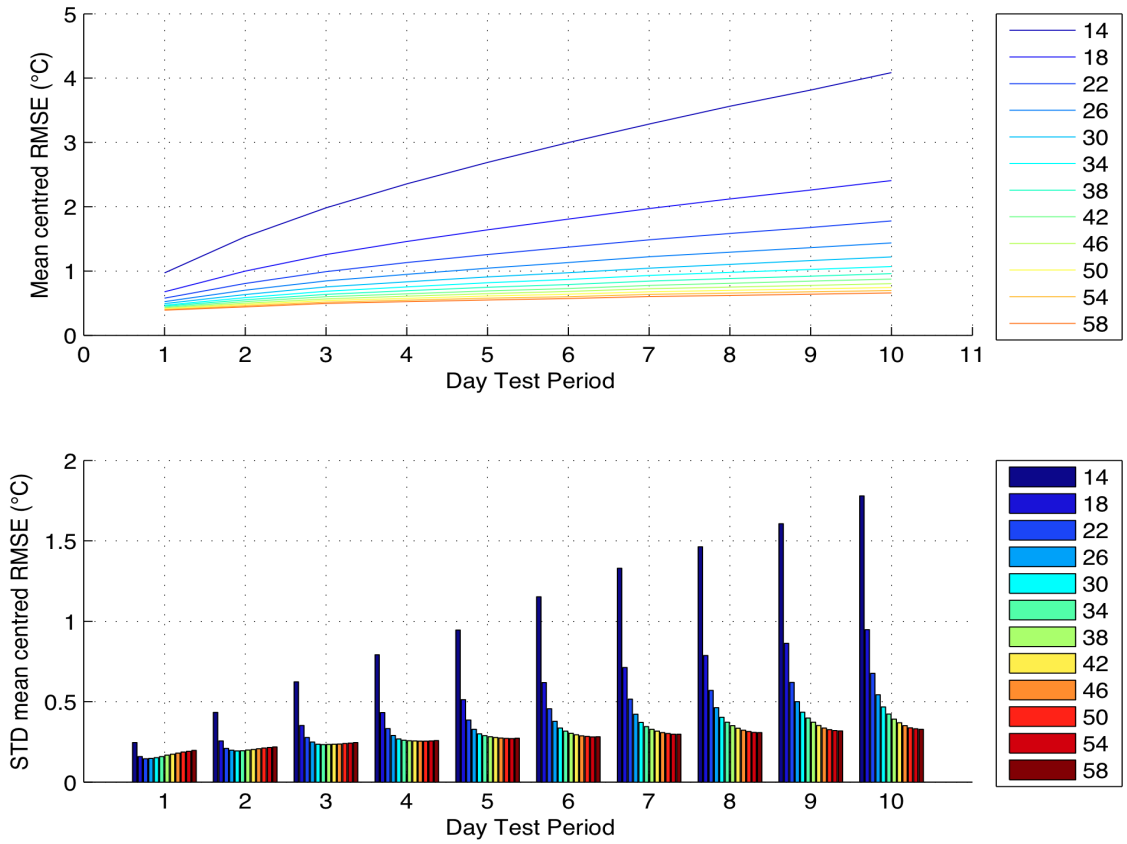

Figure 3.23: *Domain average RMSE comparisons during the whole year 2008, with different training period. Dark Blue lines represent shorter training period, red line displays the longest training period; the associated standard deviation on the bottom panel*

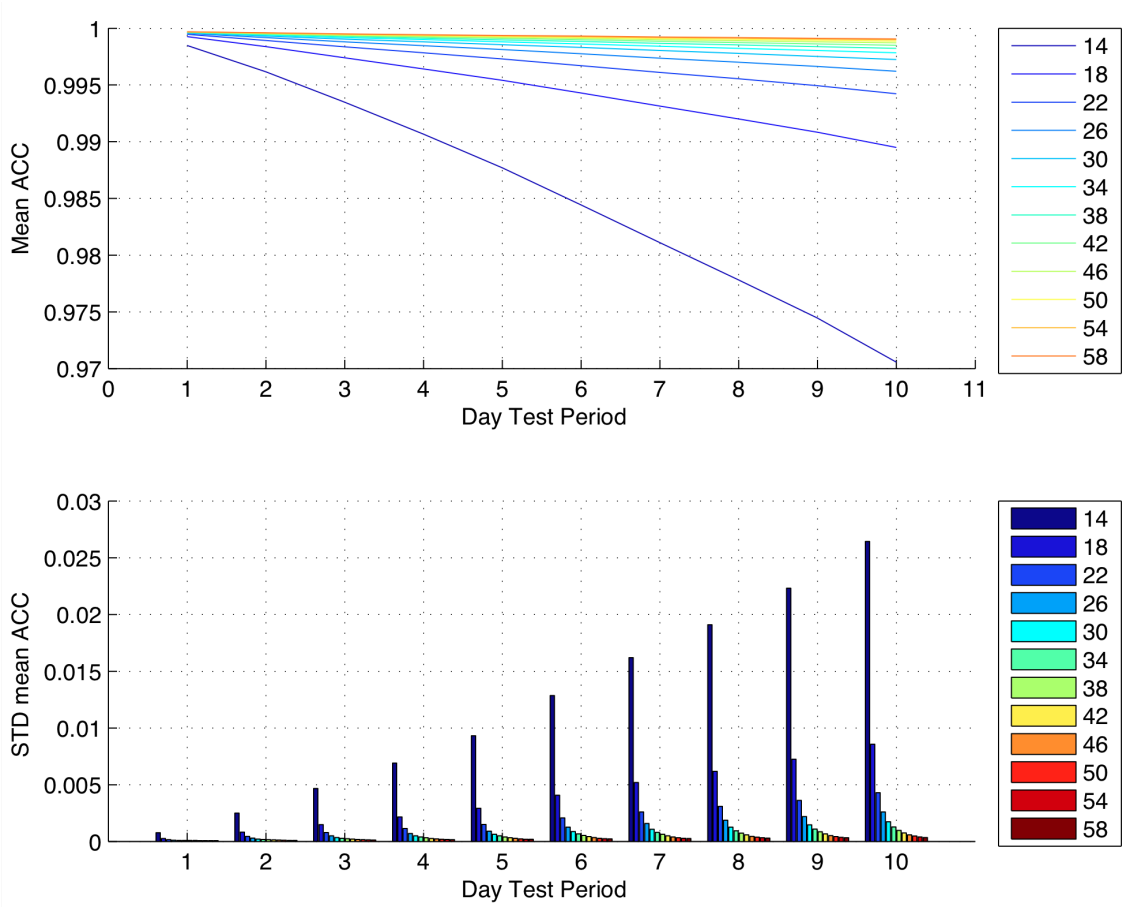

Figure 3.24: *ACC comparisons during the whole year 2008, with different training period. Dark Blue lines represent shorter training period, red line displays the longest training period; the associated standard deviation on the bottom panel*

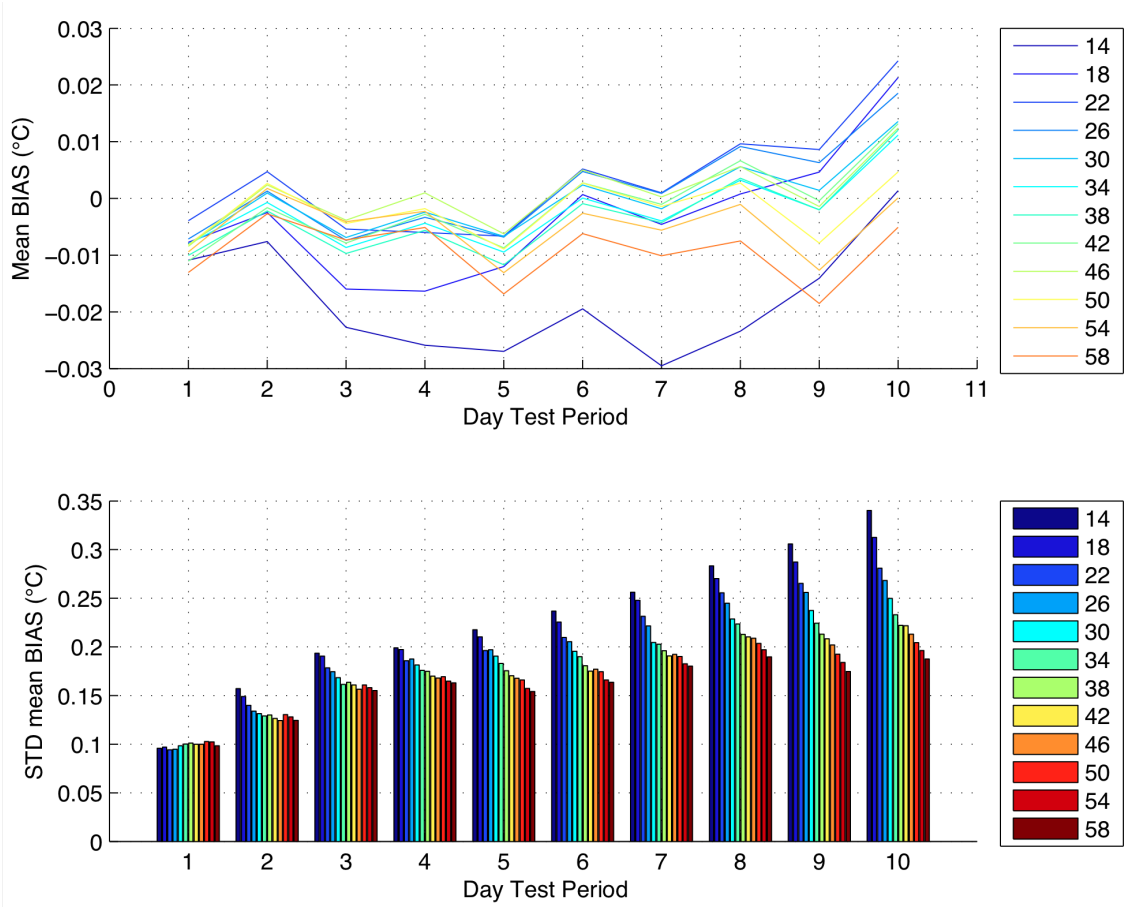

Figure 3.25: *BIAS comparisons during the whole year 2008, with different training period. Dark Blue lines represent shorter training period, red line displays the longest training period; the associated standard deviation on the bottom panel*

On the other hand, RMS is almost constant until 35 days of the training period and then it seems to growth. As we already shown in MPMM 1 experiment, this due to the overfitting of the regression procedure, which perfectly predict the training data by memorizing them leading to have unlikely results in the SST maps.

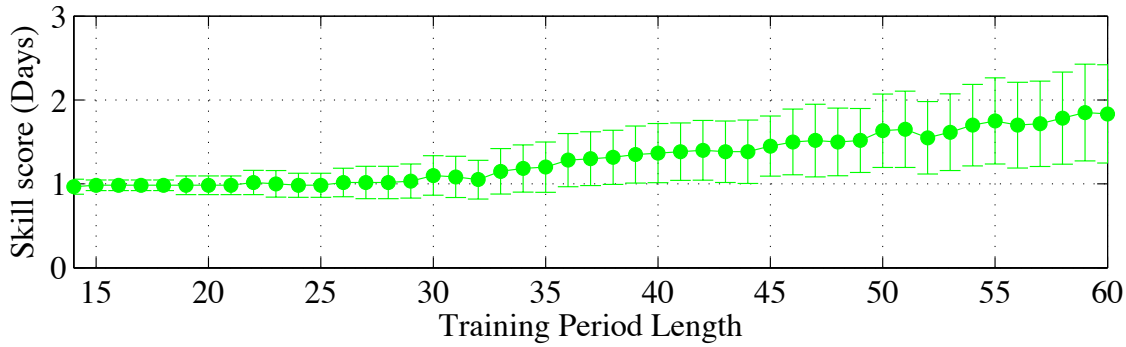

Figure 3.26: *Bias Removed RMSE mean Skill of the Multi Model superenseble with 9 members against the trainin period legths (green spots) the error bar is the standard deviation*

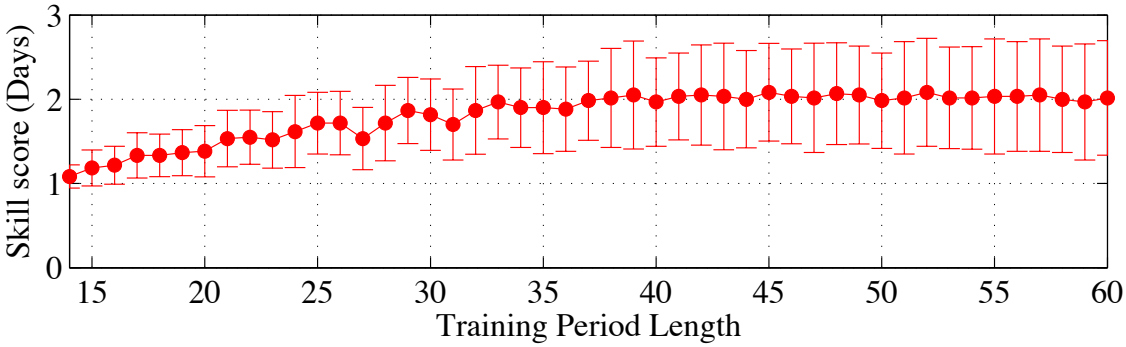

Figure 3.27: *ACC mean Skill of the Multi Model superensemble with 9 members against the trainin period legths(red spots) the error bar is the standard deviation*

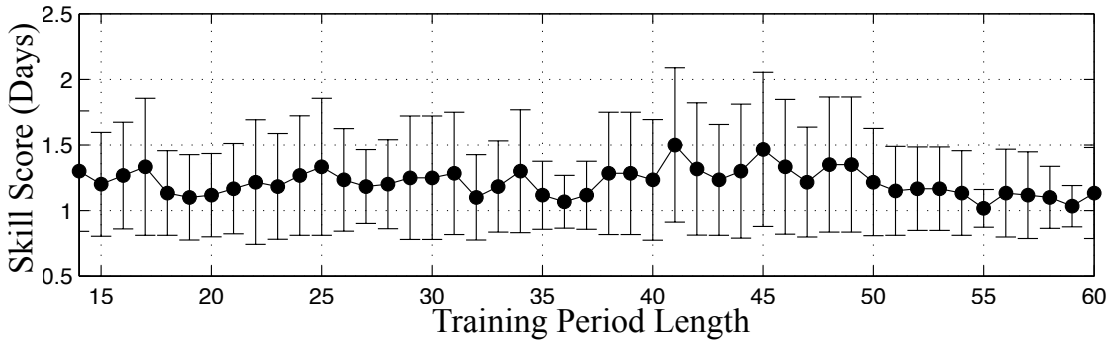

Figure 3.28: *Bias Skill of the Multi Model superenseble with 9 members against the trainin period legths(black spots) the error bar is the standard deviation*

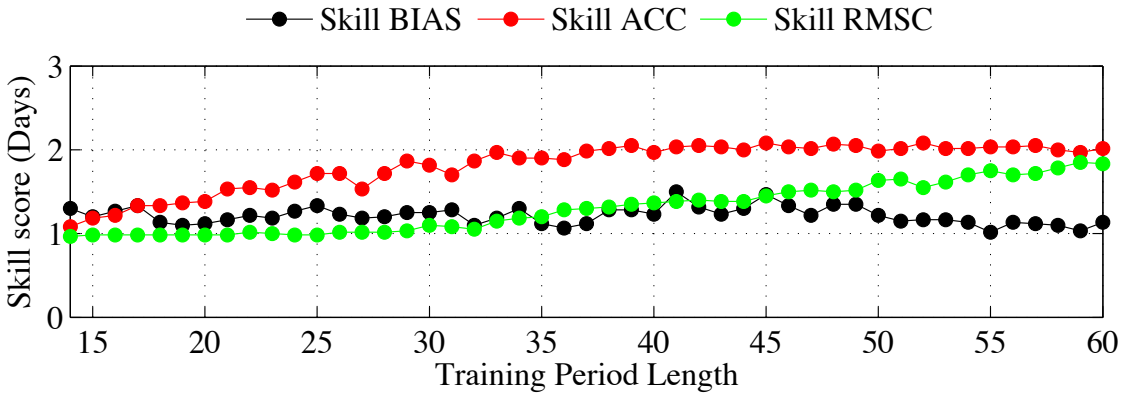

Figure 3.29: *Comparison of the mean skills for BIAS(black spots), ACC (red spots) and Centred RMSE(green spots)*

3.3.2 Sensitivity on the dataset

The aim of this study is to examine the effect on SE performance adding or removing a "bad" or a "good" model (respectively, a model that consistently performs worse than average of the ensemble, and complementary the best model) in order to identify the circumstances under which the MPMM ensemble system really enhances the prediction skill. Taylor diagrams (Taylor, 2001) have been introduced for model inter-comparison, providing a method 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). The analysis of

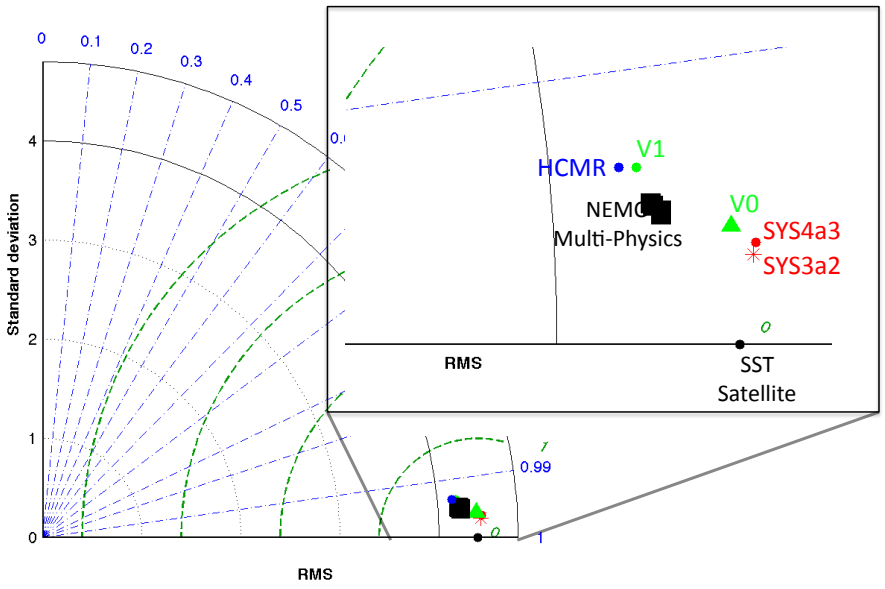

Figure 3.30: *Taylor diagram of the ensemble members employed in MPMM 2 experiment vs. SST satellite-derived.*

the Taylor diagram (fig. 3.30) let us create three different sub-samples from the complete dataset, in order to assess the sensitivity to ensemble goodness and to study the impact of overconfident (Weigel *et al.*, 2008) dataset or too correlated ensemble members. The drawn sub-samples are:

$MMSE$ $Member$	$Vertical$ $scheme$	$Diffusion$	$Viscosity$	$Assimilation$
SYS3a2	*P.P.*	Bilap.	Bilap.	3DVAR
SYS4a3	*P.P.*	Bilap.	Bilap.	3DVAR
Mercator V0	$k-\epsilon$	Laplacian.	Bilap.	SAM
Mercator V1	$k-\epsilon$	Laplacian	Bilap.	SAMv2
HCMR(POM)	$k-l$	Laplacian	Laplacian	SEEK filter
NEMO	*P.P.*	Bilap.	Bilap.	no

Table 3.1: *Overconfident Dataset Set up, obtained by MyOcean database*

$MMSE$ $Member$	$Vertical$ $scheme$	$Diffusion$	$Viscosity$	$Assimilation$
SYS3a2	*P.P.*	Bilap.	Bilap.	3DVAR
HCMR(POM)	$k-l$	Laplacian	Laplacian	SEEK filter
NEMO multiphysics	*P.P.*	Bilap.	Bilap.	no
	$k-\epsilon$	Bilap.	Bilap.	no
	P.P.	Bilap.	Laplacian	no
	P.P.	Laplacian	Laplacian	no

Table 3.2: *Well-dispersed Dataset Set up, obtained by MyOcean database*

- **Multimodel constructed from overconfident models ensemble**: Subsample A, the members are the operational product, SYS3a2, SYS4a3, Mercator V0, Mercator V1, HCMR and NEMO operational in simulation.(see table 3.1)

- **Multimodel constructed from well dispersed models ensemble**:Subsample B, the members are SYS3a2, HCMR and the four simulations with multi-physics members NEMO (see table 3.2).

- **Multimodel constructed from bad dispersed models ensemble**:Subsample C, Mercator V1; HCMR and four simulations with multi-physics members NEMO (see table 3.3).

$MMSE$ $Member$	$Vertical$ $scheme$	$Diffusion$	$Viscosity$	$Assimilation$
Mercator V1	$k-\epsilon$	Laplacian	Bilap.	SAMv2
HCMR(POM)	$k-l$	Laplacian	Laplacian	SEEK filter
NEMO multiphysics	$P.P.$	Bilap.	Bilap.	no
	$k-\epsilon$	Bilap.	Bilap.	no
	$P.P.$	Bilap.	Laplacian	no
	$P.P.$	Laplacian	Laplacian	no

Table 3.3: *Bad-dispersed Dataset Set up, obtained by MyOcean database*

For each dataset, we carried out again all the experiments done with the original, complete, dataset and compared the results using the previous diagnostic metrics.

The reconstructed maps of SST from the sub-samples for the first day of the test period (see fig.3.31), are pretty similar to the complete dataset prediction. Differences arise during the lead time of the test phase. For the last day of the test period (see fig.3.32), the maps show bigger differences originated by the increase of the noise due to overfitting. To avoid misleading, we compared the time-series of skills only between the overconfident dataset and the original, because with the other sub-samples, the possible greater skills would be a paradox due to the deteriorating dataset performances.

Different growth rates can be distinguished for RMSE and ACC, respectively by analysing figures 3.33 and 3.34. While, there is no growth rate for the bias with the over-confident dataset (see fig. 3.35). We can infer that the overconfident dataset has better skill scores and faster skill growth versus the training period length respect the original nine members ensemble, but for longer training period, the plateau in ACC curve means that, we incur again in overfitting. A careful analysis of the performances diagnostic metrics curves show that, the original dataset is not the best predictor for short training period (figures 3.36, 3.37 and 3.38), but the differences decrease enlarging the training period, letting the bad dispersed dataset, sub-sample C, be the worst of the available predictions. This latter sub-sample, is the best in terms of bias for longer training period (see fig. 3.38). These figures also shows that

the sub-sample A, the overconfident dataset, is the best in term of RMSE and ACC for longer training period (as we have already seen before), and it 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.
Since all predictions, derived by the different sub-samples, give similar results for longer training period, we can argue that, the 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. The regression procedure ignore the informations coming from the predictors, and the SE *learns* only from the past observations.

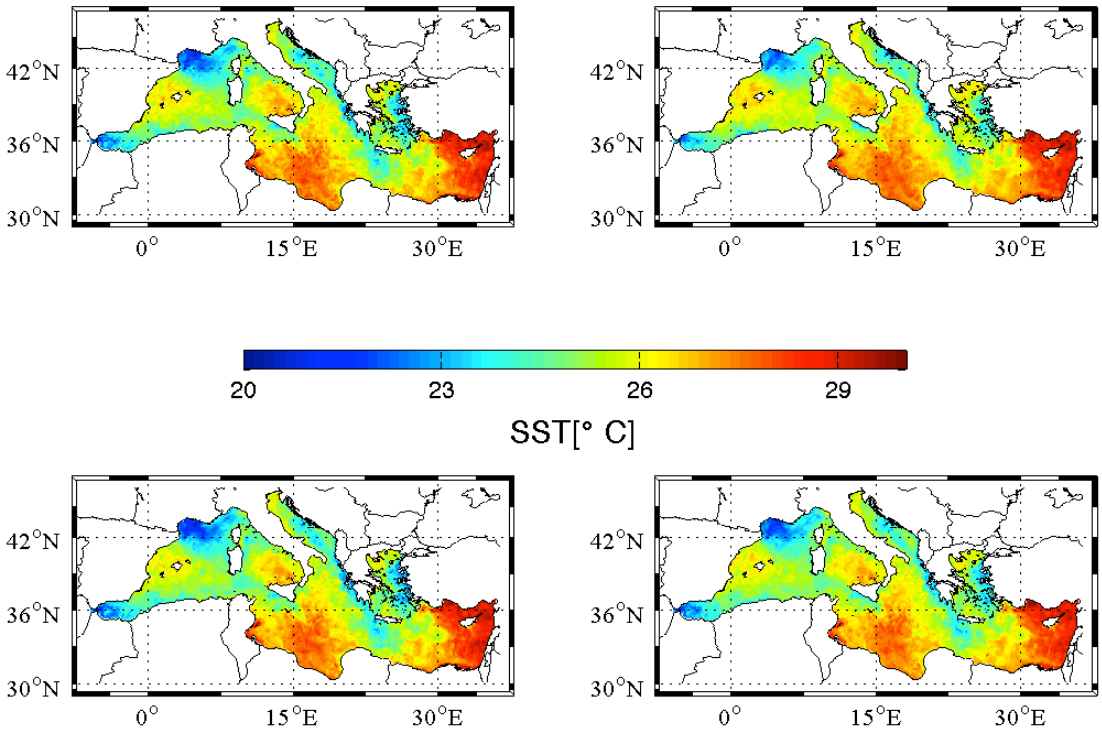

Figure 3.31: *A typical example of prediction for the first day of test period over the Mediterranean area, valid on the* 1^{st} *of June 2008 reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)*

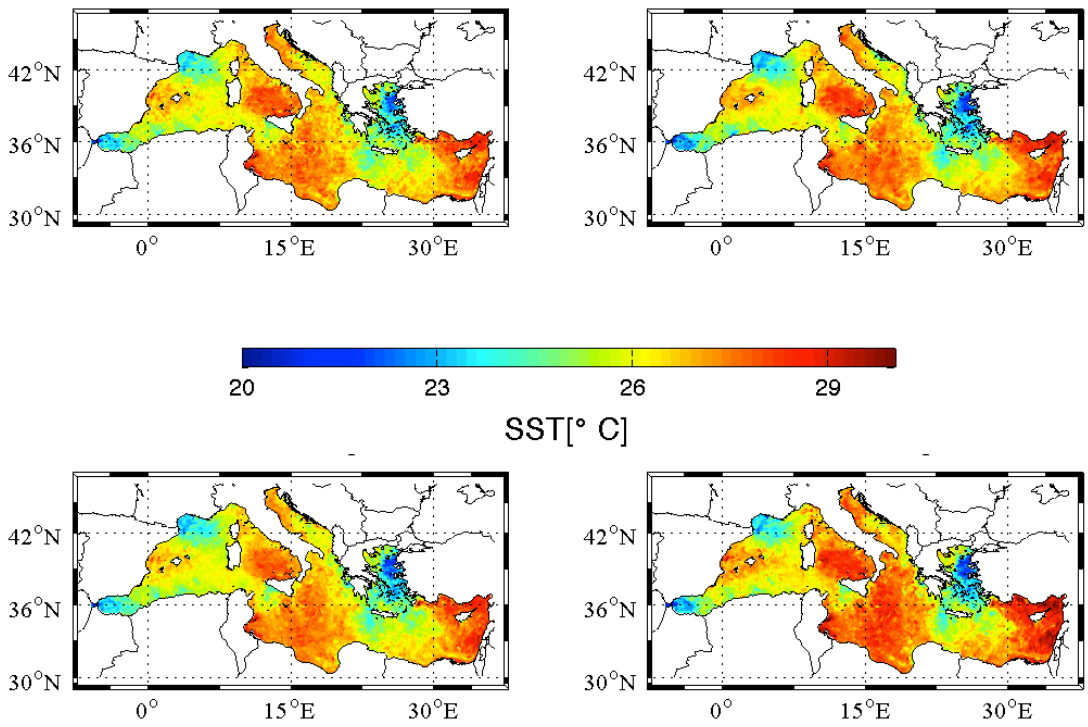

Figure 3.32: *A typical example of prediction for the last day of test period over the Mediterranean area, valid on the* 10^{th} *of June 2008, reached with 41 days of training period, SE prediction from the original dataset (top left), SE prediction from the well dispersed dataset (bottom left), SE prediction from the overconfident dataset (top right) SE prediction from the well dispersed dataset (bottom right)*

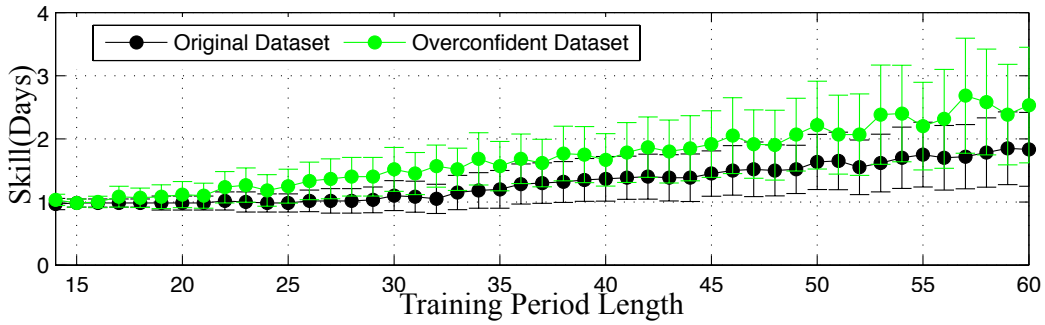

Figure 3.33: *Bias Removed RMSE mean Skill of the Multi Model superenseble with 9 members against the training period legths (black spots) the black error bar is the standard deviation, green spots are Bias Removed RMSE mean Skill for the subsample A ,the green bars are the standard deviations.*

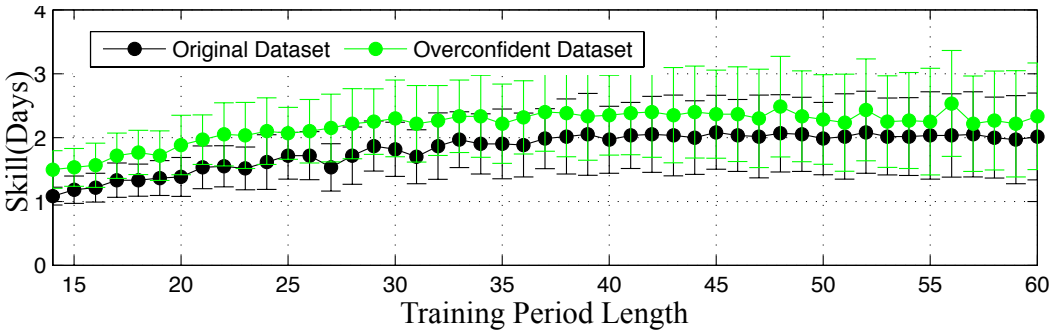

Figure 3.34: *ACC mean Skill of the Multi Model superenseble with 9 members against the training period legths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bars are the standard deviations.*

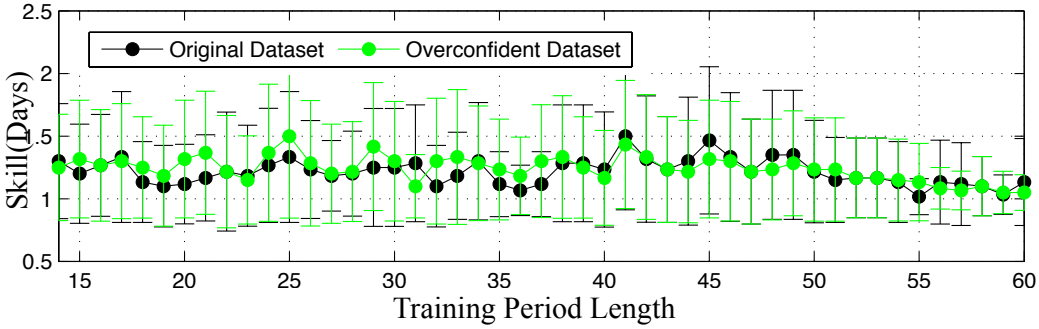

Figure 3.35: *BIAS Skill of the Multi Model superenseble with 9 members against the training period legths (black spots) the black error bar is the standard deviation, green spots are ACC mean Skill for the subsample A the green bars are the standard deviations.*

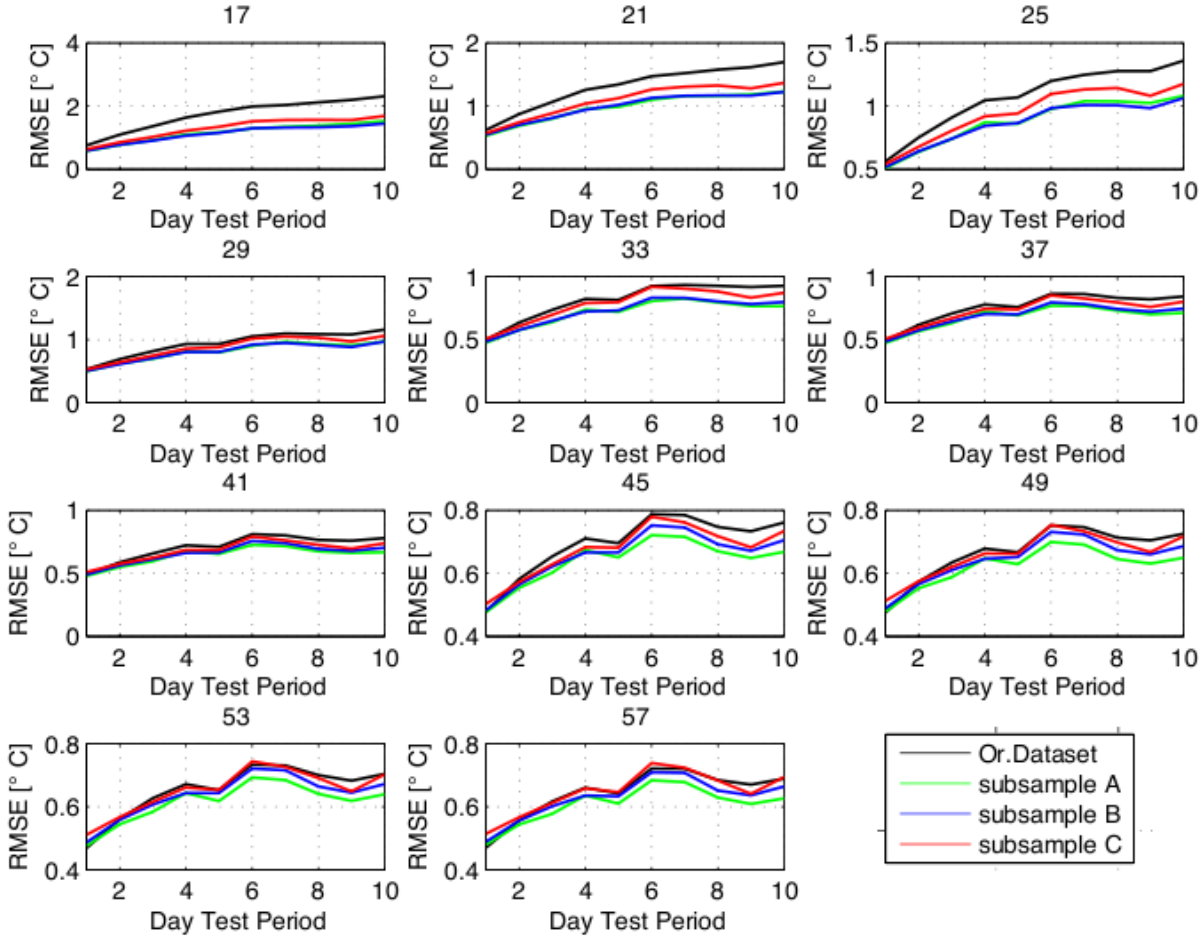

Figure 3.36: *Domain average RMSE comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C*

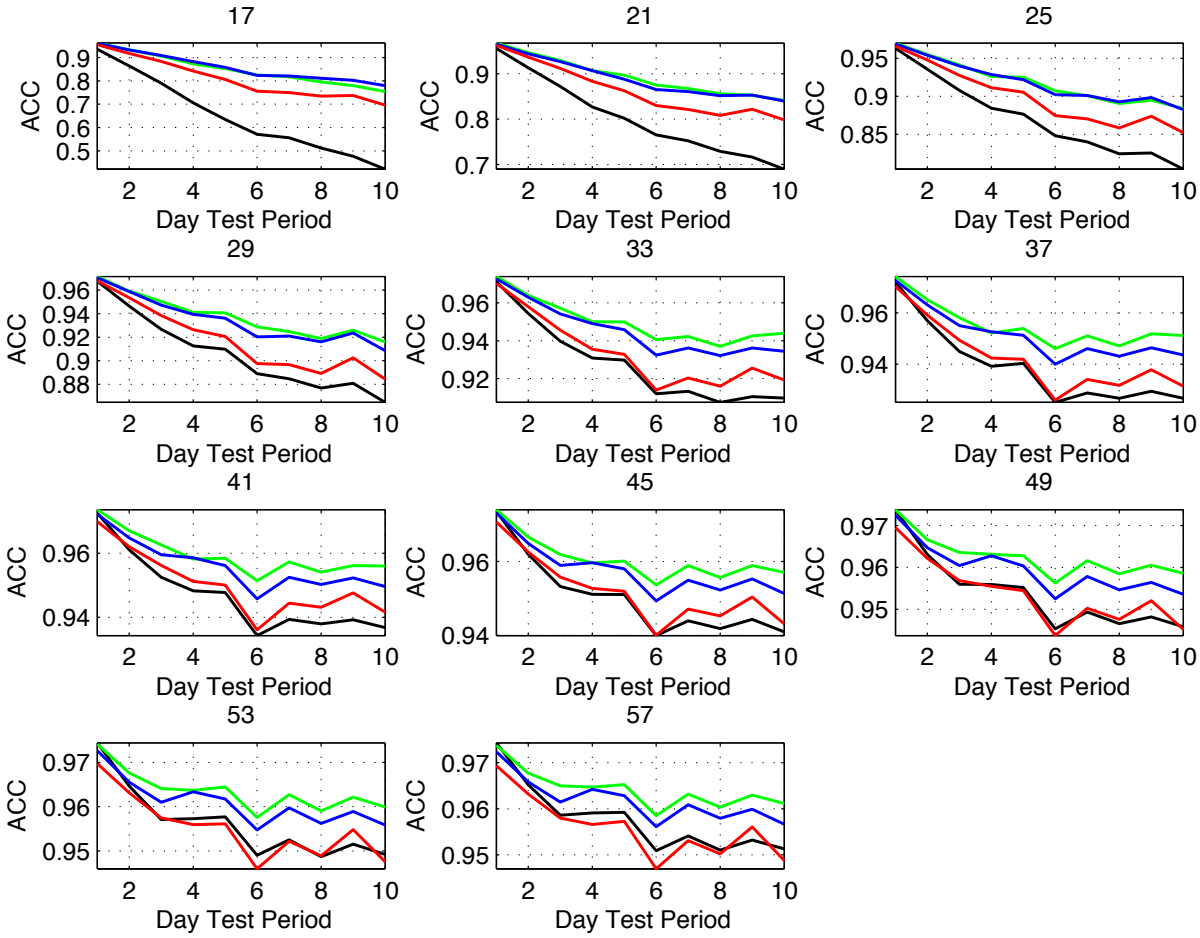

Figure 3.37: *ACC comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C*

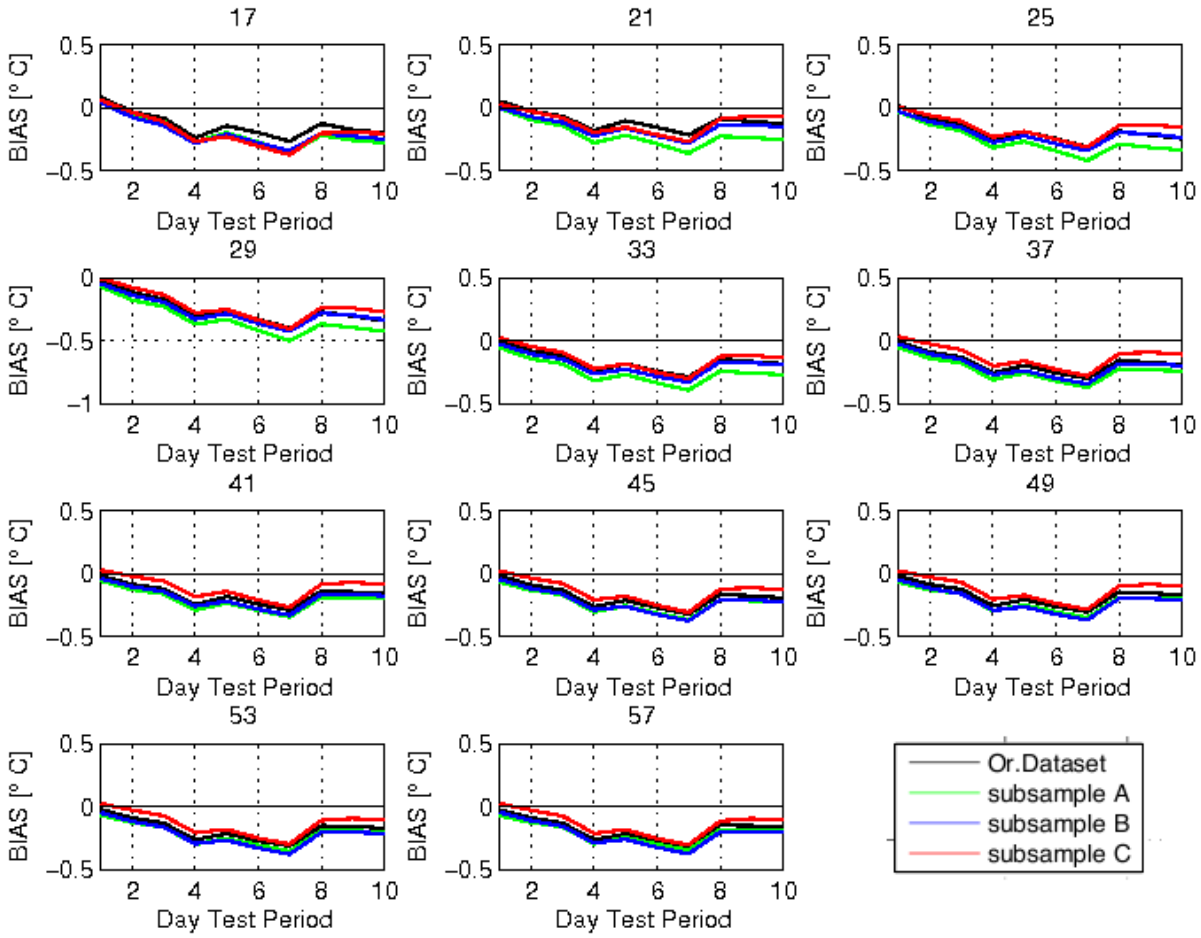

Figure 3.38: *BIAS comparisons for the same test period(1-10 June 2008), with different training period(indicated over each subplot). Black line is the SE from the original dataset, green line for the subsample A, blue line subsample B and red line subsample C*

3.4 Summary

In this chapter, we have shown the results obtained with the SuperEnsemble techniques. In the 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 identify best/worst participating models.

In the first experiment, we have seen that the lack of spread in Multi-Physics Ensemble (MPE1) leads the algorithm fail. In the second experiment, we showed that Multi-Model forced with different atmospheric forcing, MPMM1 experiment, can outperform all the participating model, but some unreliable features, due to overfitting problems, are reproduced in the SE SST horizontal maps. In the third experiment, the Multi-Model Multi-Physics approach, MPMM2 experiment, we have studied the sensitivity and limits of this procedure. A careful examination of the results showed that robust optimal weights are difficult to calculate giving short training period samples, while too long training period achieved overfitted predictions. An other key point is the choice of the dataset components: as long as the individual components are able to make a positive contribution to a relevant aspect of the prediction, the multi-model will benefit from this additional information. We have shown that better performances are reached reducing the ensemble size, leaving out the worst participating models. In this scenario, one model is superior to all the other component, so poorer model could not add information (well dispersed dataset and overconfident dataset). Hence, the key of the success of multi-model concept lies in combining independent and skilful models. Furthermore, if the quality assessment detects a single model that is always worst than the others, it should be excluded. For long training period, all the sub-samples achieve very similar performances, meaning that there is an upper limit for the predictability, over that overfitting problems arise. An other consideration when assessing the training period size, is related to the field variance. The time mean during the training period of the truth estimator would be no more "unbiased" for longer training period, being affected by the seasonal cycle effects.

An area of future work should 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

We have shown that the results of the superensemble technique proposed by Krishnamurti *et al.* (1999), can outperfom the individual ensemble members involved, when suitable conditions on free parameters are chosen. However some problems, due to the overfitting of observations during the regression procedure and ill-conditioning on the covariance matrix, arise. The SE maps have shown some unreliable features: they are very noisy fields, which neglect the spatial correlation of the variable under study, and where some sub-basin scale structures, depicted in the observations, disappeared. Kharin & Zwiers (2002) suggest 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. Either 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.

Hence, our actual study is focused on the improvement of the superensemble estimate using appropriate filtering techniques: the Empirical Orthogonal Functions (EOFs). 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 ans 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.

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 brief description of how to obtain EOFs, giving the linking between their development and application in our case. Once the anomaly data matrix (b_{ij} see) is determined : The EOFs are obained as the solution of the classical eigenvalues problem

$$\Gamma\vec{x} = \lambda^2\vec{x}; \tag{4.1}$$

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: Throughout EOFs/PCs analysis, the initial field b becomes:

$$b = \sum_{j=1}^{p} \lambda_k a_k u_k' \tag{4.2}$$

The dimensionality of the initial data can be reduced by truncating the sum at an 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 typical value is 90% of variance). Hence, we chose a set of n leading EOFs that can explain at least that desired amount of variance (in our cases the number could span from 7 to 9).

The orthogonality property provides a complete basis where the time-varying field can be separated into spatial patterns and associated time indices. This completely

characterises conventional EOFs. Orthogonality constitutes, however, a strong constraint that puts limits to the physical interpretability of individual EOFs, since in general, physical patterns tend to be non-orthogonal (Simmons *et al.*, 1983) . Furthermore, when the eigenvalues are not distinct (some may have multiplicity greater than one) the eigenspectrum degenerates. The numerical techniques employed cannot distinguish them and the choice of the right truncation order can be problematic. The most common situation in which malfunctioning can occur, is when the data represent the local variances (A Navarra, 2010). In these cases, EOFs will try to fit globally the domain under consideration, with as few modes as possible, generating first EOFs (low order modes) with very large structures. Here, the typical patterns in the analysed field may appear in secondary (higher order) EOFs. If their real relation is localized, the EOF, due to new reference system, spreads it creating artificial non-local relations and the eigenspectrum degeneration. Thus, the choice of a wrong truncation order, due to the reached amount of the desired explained variance, may cut off some important phenomena.

As first guess we applied the EOF analysis only to the weights mask, in order

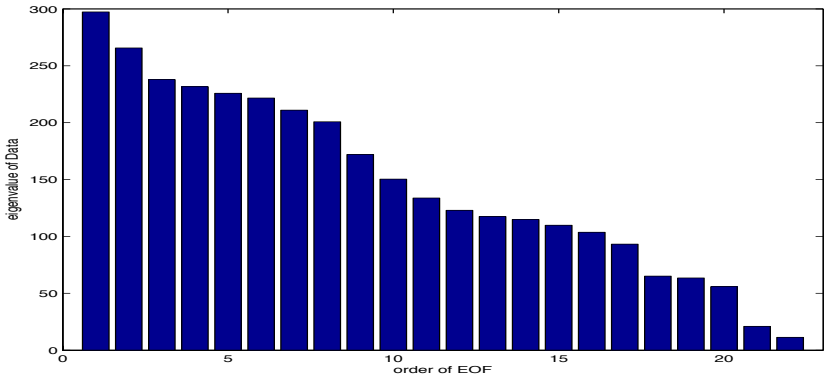

Figure 4.1: *Eigenvalues spectrum for the weights mask in the BHM dataset*

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 (depicted in figure 4.1). In this situation, we need to choose a high truncation order, 15 upon a rank matrix of 22, to

explain at least the 85% of variance. In this example 15 EOFs were not sufficient to correctly identify noise from signal, since the 16-17th eigenvalues were pretty close to the chosen threshold. Even if several *"rule of thumbs"* are available, it is better to realize that the choice on the truncation strategy, must be driven by *ad hoc* considerations.

In order to remove the degeneracy of the power spectrum we decided to apply the EOFs analysis to the multi-model anomaly field and project the observation anomaly field on the EOF too. Step back to equation 3.8, we can substitute in Γ its eigenvalues and solve the following system

$$\vec{x}\lambda\vec{a} = \vec{x}\vec{\phi} \tag{4.3}$$

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

$$\vec{a} = (\vec{x}\lambda)^{-1}\vec{x}\vec{\phi} \tag{4.4}$$

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

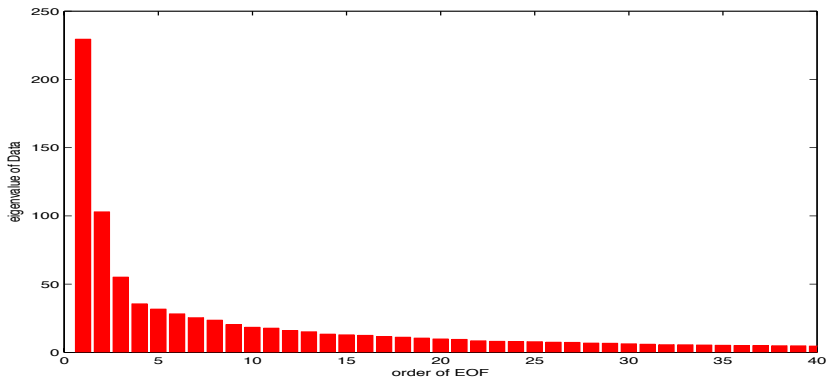

Figure 4.2: *Eigenvalues spectrum for the weights mask in the BHM dataset*

principal component analysis was performed in order to reduce the spatial dimension to a number smaller than the original number of grid points. The results for this example suggest that, with a very small number of EOFs we can explain most of the

principal modes of variability of our system.In fact, the first EOF can explain alone more than the 30% of variance. Since we decided to retain only the components that can explain the 85% of the variance, in our case it corresponds to take only the first seven-nine EOFs (according to training period length involved) of the anomaly fields, see figure 4.2 and we considered the remaining part of the spectrum a noise.

4.3 MPE1 EOFs-based SuperEnsemble

We performed the EOFs analysis on the first dataset, the perturbed physics ensemble experiment MPE 1. EOFs Analysis was able to reduce the degeneracy of the covariance matrix due to the high correlation of the member collected. The new computation of superensemble shows better performance and physical reliability compared to the classical approach (not shown). Nevertheless, SE is not the best performer of the ensemble. The diagram bars (not shown) indicates that SE estimates are not the best performer for the test period under study.

4.4 MPMM1 EOFs-based SuperEnsemble

The dataset has been described in section 2.2.1. Weights have been obtained by the regression of the projections of true estimator (SST from SYS3a2) and MPMM1 ensemble members, retaining the first nine EOFs, able to explain 87% of the system variance. We were able to reduce the noise in the SST maps (fig.4.3), and until the 20th day of test period, the EOFs-based SE gives the best results in terms of $RMSE_c$ (fig.4.4). At the end of test period, SE estimates are affected by a large bias (depicted in fig. 4.5) and it becomes the worst ensemble member both in $RMSE_c$. Figure 4.6 shows the performance of EOFs-based SE estimates in term of MSESS. Again the SE better performs better than all the other ensemble members. However, it must be enhanced that in this way we have lost again some sub-basin scale features. Further investigations are needed on the balance between filtering magnitude and the time scales of the dynamical processes involved.

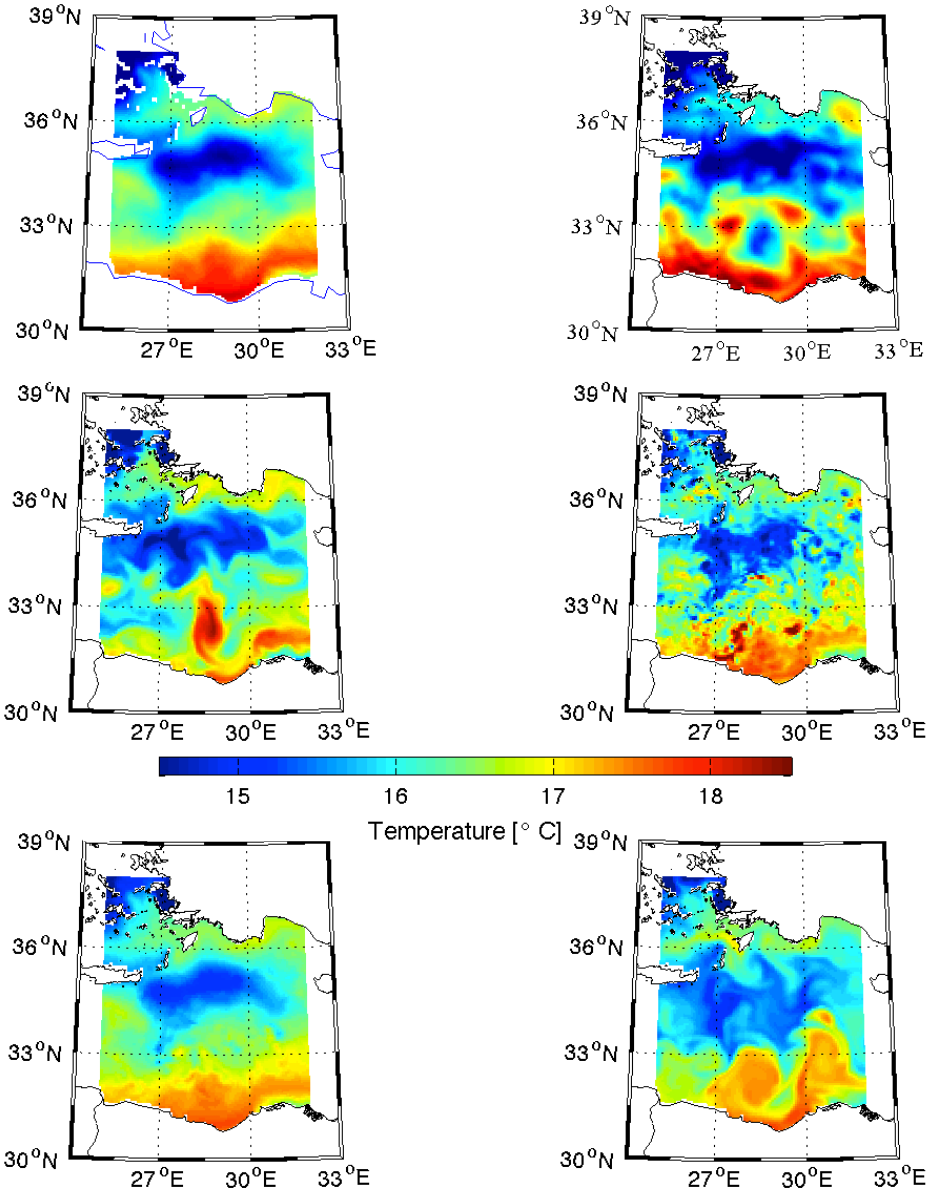

Figure 4.3: *An example of a SE estimate of SST, on day three over Rhode Gyre region: Observation mean (top left), projection of ,SYS3a2 analysis on EOFs (top right), SYS3a2 analysis (center left), Krishnamurti SE estimate(center right), EOFs-based SE (bottom left), and best ensemble member(bottom right).*

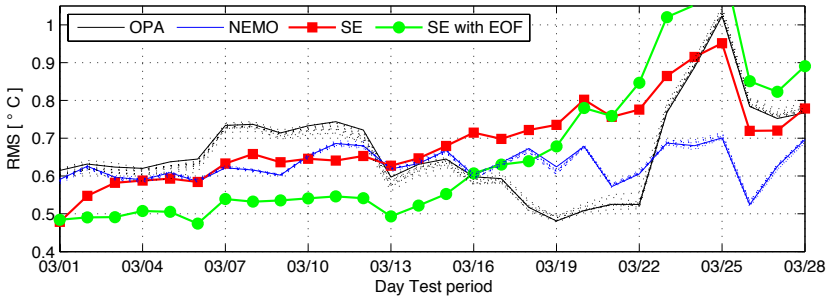

Figure 4.4: *Domain averaged RMSE between the OPA members(black), NEMO members (blue), SuperEnsemble prediction(red) and SE prediction EOFs-based(green) during the test period*

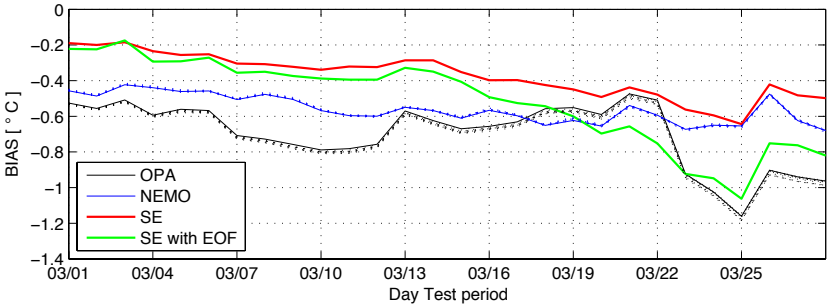

Figure 4.5: *BIAS, OPA members(black), NEMO members (blue), SuperEnsemble prediction(red) and SE prediction EOFs-based(green) during the test period*

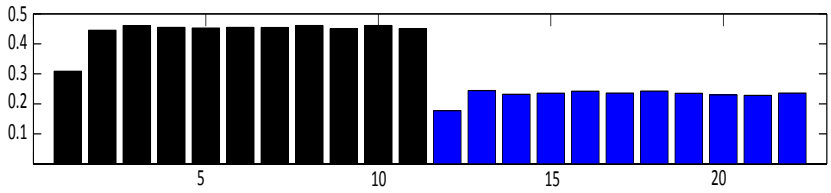

Figure 4.6: *MSESS for MPMM1 ensemble for the SST estimates over Rhode Gyre region during the test period(1-31 March 2008). Black bars represent gain respect OPA predictions, and blue bars the gain respect NEMO. Each bar is the MSESS as evaluated by3.14 with reference the MSE of each member.*

4.5 MPMM2 EOFs-based SuperEnsemble

In this section, we present the results of the multimodel Multi-physics experiment MPMM2, a collection of analyses from different research institutes and four perturbed-physics members, covering the whole year 2008 and described in 2.2.2. We performed again all the 2532 experiments with the same set up of the classical Krishnamurti approach, but in this case, we reached different conclusions.

EOFs-based SE estimates do not depict noisy SST distribution, see fig.4.7, but their performances got worst increasing the lead time of the test period (fig.4.8). These new maps seem to be affected by a bias. In fact the projection of the satellite SST on the EOFs are affected by the same bias of the time mean of the observations. A comprehensive comparison of the two methodologies employed is shown in figure 4.9. Right panels show (respectively from top to bottom) $RMSE_c$, BIAS and ACC for the experiments trained with the Krishnamurti method. The abscissae axes is the time, the ticks stand for the month of year 2008. Each bar is result for the fifth day of the test period, and obtained with the minimum training period,14 days. While on the left side are depicted, from top to bottom, $RMSE_c$, BIAS and ACC for the EOFs-based SE experiments. We can note the high correlation between BIAS and $RMSE_c$ for the EOFs-based SE estimates. This means that the EOFs-based SE estimate are highly influenced by the choice of the unbiased estimator. For example during August time, when the bias reach is minimum, $RMSE_c$ is smaller for EOFs-based SE estimate than the classical SE estimate. This inference is supported by ACC plots. This coefficient considered a measure of potential performance, ignores the bias by construnction. Figure 4.10 shows the same scores depicted in fig 4.9, but obtained with a training period of 30 days. In this latter case, we have a bigger bias, and we have a bigger $RMSE_c$. Another aspect of the quality of SE in term of RMSE is its bias. ACC plots(bottom panels in figures 4.9 4.10) clearly show a worsening of performance of EOFs-based SE estimates in terms of ACC. It is clear from those results that the new method does not bring the desired improvements.

A careful analysis of the weights mask shows smaller values for the coefficients than in the classical approach(see fig.4.11). Referring to Krishnamurti *et al.* (2003) we try to reproduce a similar set up, using seven members (the subsample A, over-

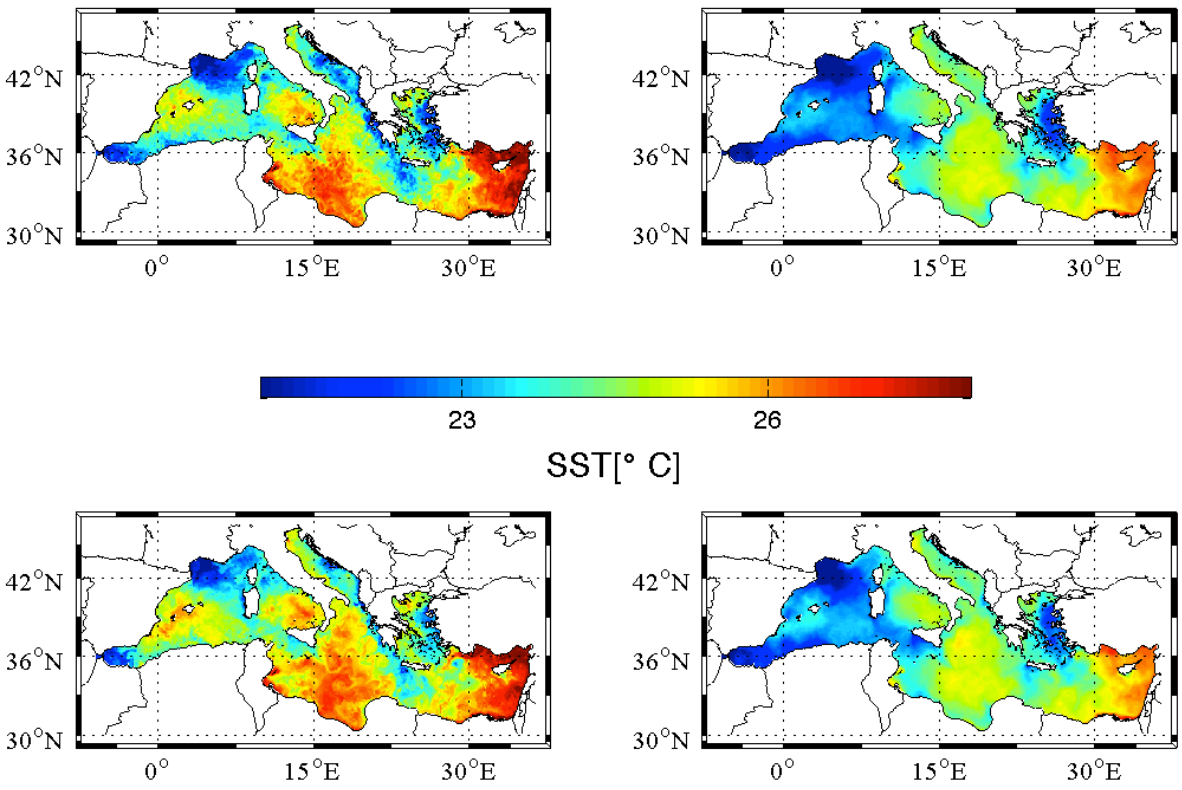

Figure 4.7: *First day of test period over the Mediterranean area, valid on the* 1^{st} *of July 2008, trained for 55 days: SE prediction (top right), EOFs-based SE estimates (top left), SST Satellite derived (bottom right) and projection of satellite SST on the EOFs (bottom right)*

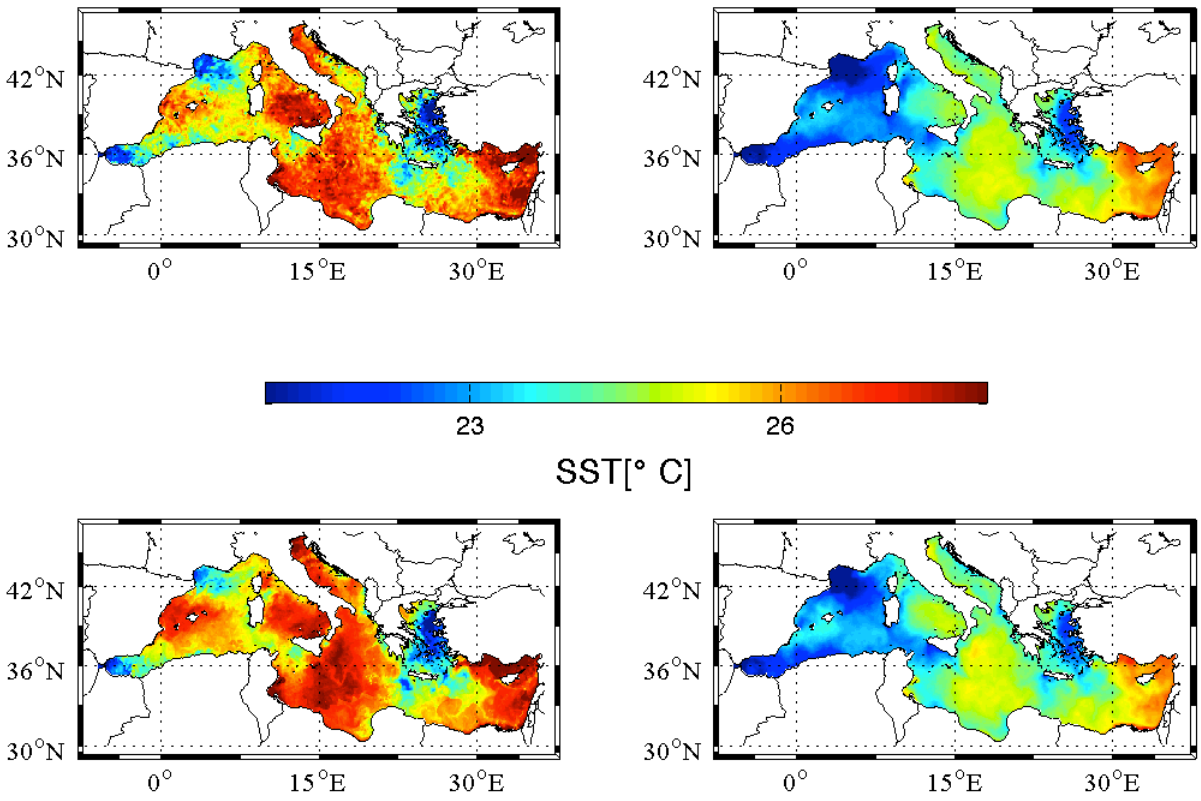

Figure 4.8: *10th day of test period over the Mediterranean area, valid on the* 10^{th} *of July 2008 and trained for 55 days: SE prediction (top right), EOFs-based SE estimates (top left), SST Satellite derived (bottom right) and projection of satellite SST on the EOFs (bottom right)*

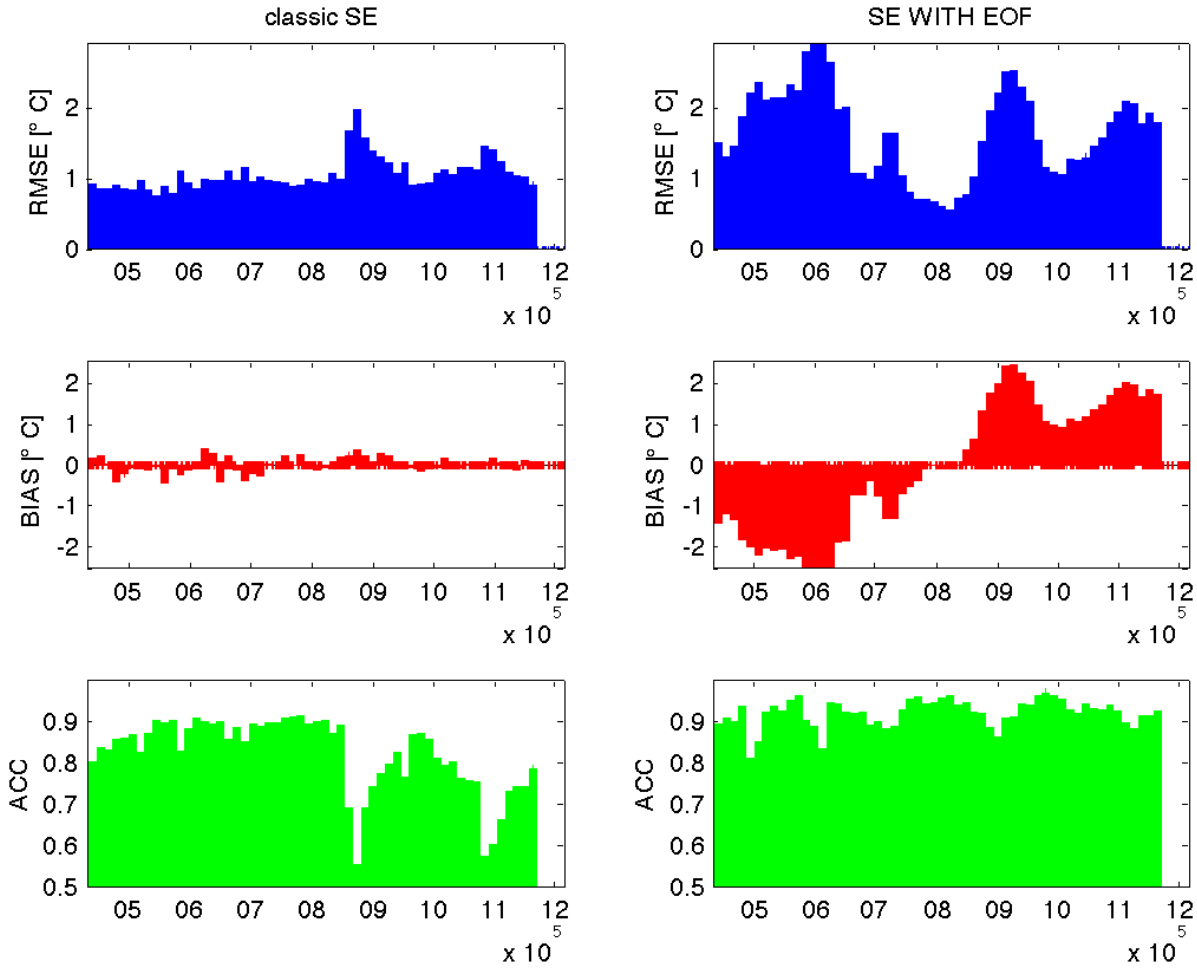

Figure 4.9: *Performances of SE obtained by 10 days of training period, during year 2008, left panels the classical approach, right side the EOFs-based SE approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.*

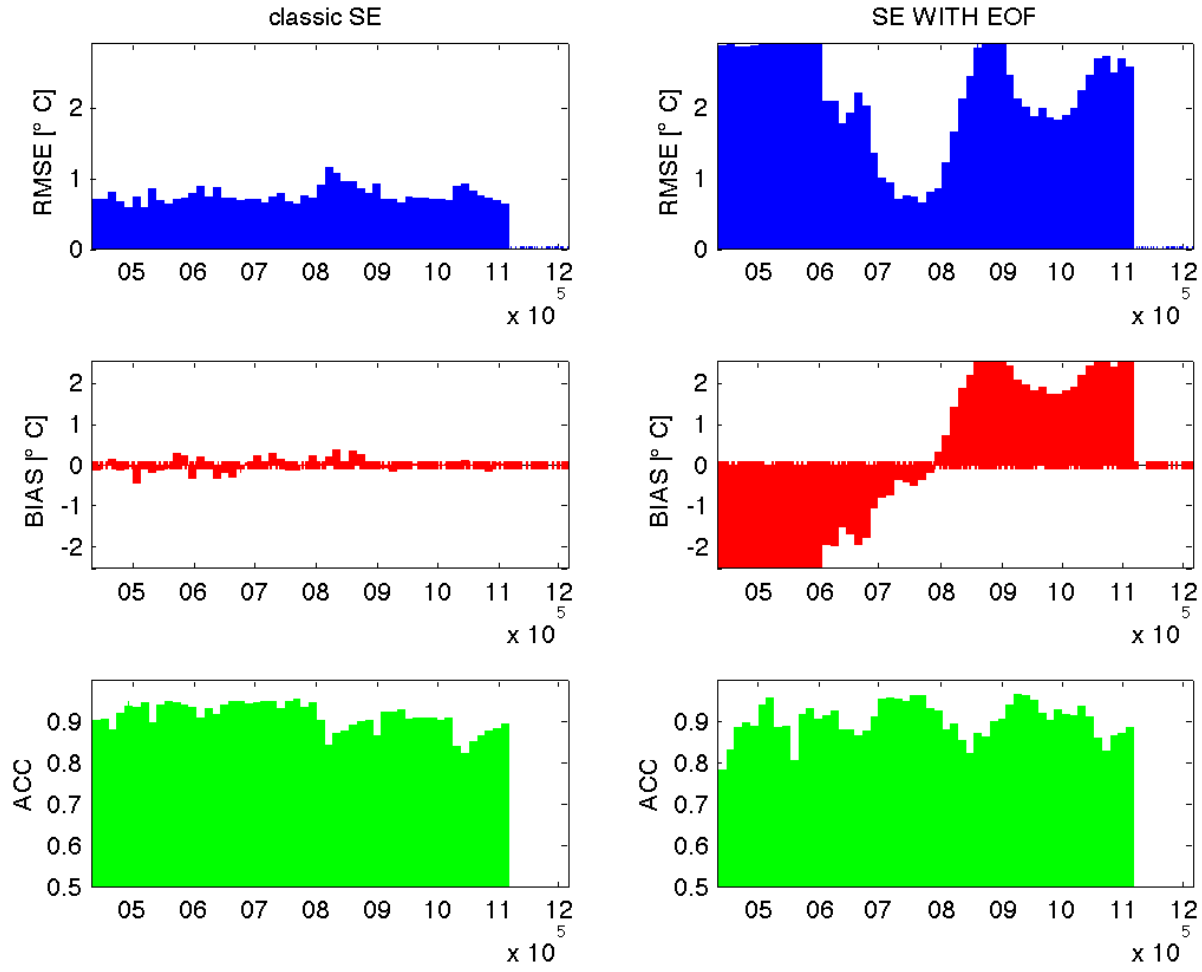

Figure 4.10: *Performances of SE obtained by 30 days of training period during year 2008, left panels the classical approach, right side the EOFs-based SE approach. Blue bars stand for RMSE, red bars for BIAS and green bars for the ACC.*

confident 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 4.11) we got larger 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:

$$|w^m_{eof}\left(F_{mn}-\overline{F_m}\right)| < |w^m\left(F_{mn}-\overline{F_m}\right)| \tag{4.5}$$

It is clear from that inequality (4.5) that the bias taken in to account in the term $\left(F_{mn}-\overline{F_m}\right)$ is under estimated with fractional weights. EOFs analysis let us obtain weight values closer to that papers (see fig.4.12), but in our experiments, especially when trained for long period, the time mean of observations,$\bar{O}$ in no more *unbiased* since it takes in account of the seasonal cycle signal, and it influenced by compensation effects of small scale circulations. Hence SE fails due to a wrong BIAS reduction since it will be affected by cold bias during spring-summer time and warm bias during fall-winter time (well depicted in figures 4.9 and 4.10). The key point is the kind of field we want to reproduce, since we are considering a faster varying field than Krishnamurti case (geopotential height 850hPa), EOFs-based SE evaluated with this kind of "unbiased estimator" was not a good way to perform the regression. An other consideration is related to the truncation order, probably we choose a wrong one. In the new reference system defined by EOFs, the remaining part of the power spectrum can describe some small scale features, but these structures being of the same order of the noise, are filtered. We should choose a new truth estimator, which can take in to account of seasonal cycle effects and that can be considered *unbiased* or we have to enlarge the desired explained variance, in this way small scale features are treated as signal and can be reproduced in the EOFs-based SE estimates.

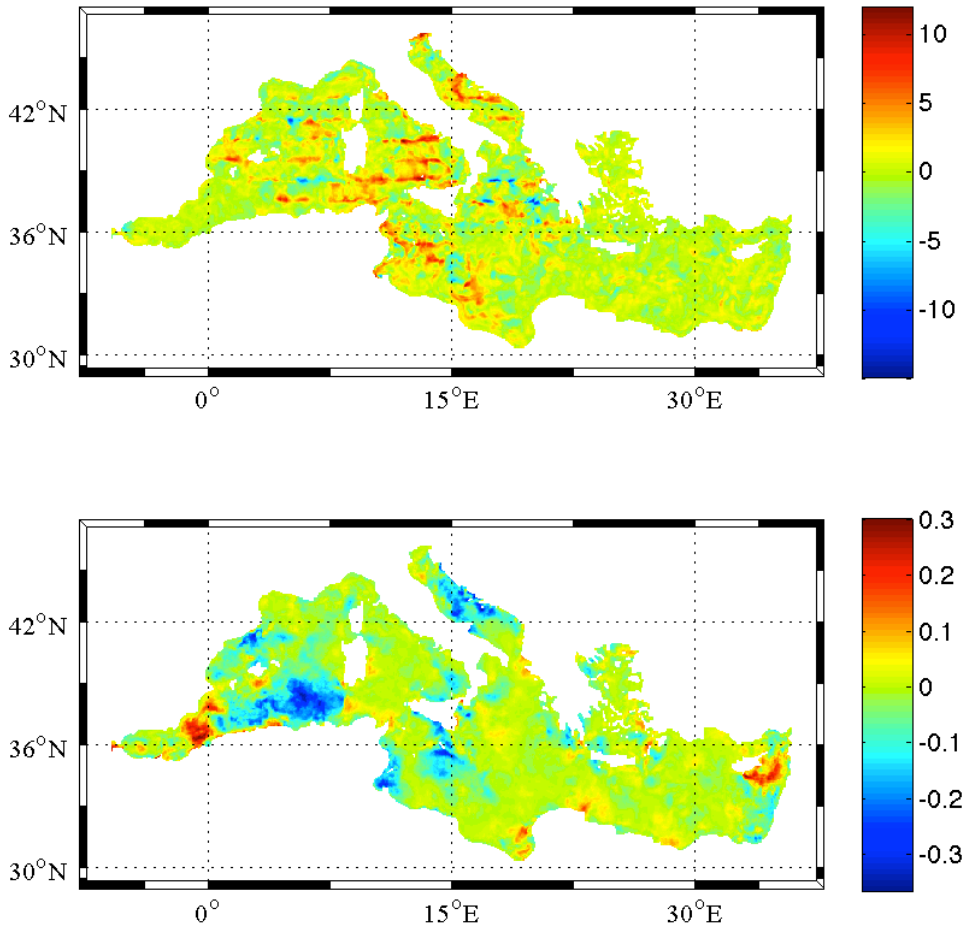

Figure 4.11: *Geographical distribution of the regression coefficients for the first model member, as the simple anomaly field (top) and for the filtered anomaly field (bottom) after 100 days of training period, note the different colorbars*

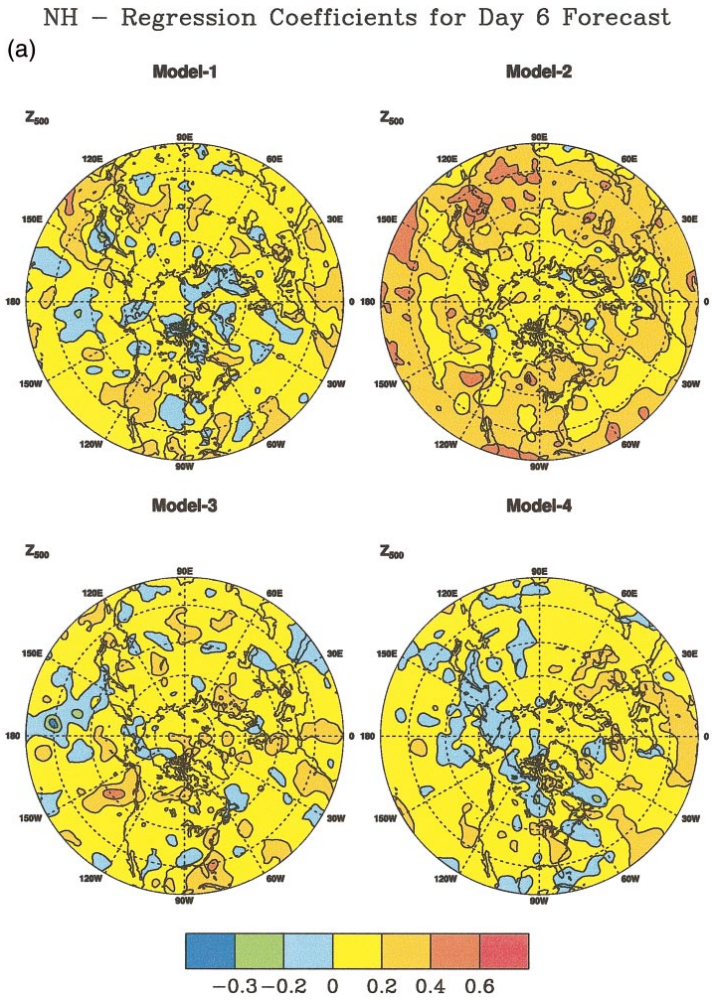

FIG. 5. Geographical distribution of statistical weights for different member models: (a) Northern Hemisphere and (b) Southern Hemisphere. Color scale of the fractional weights is shown at the bottom.

Figure 4.12: *Statistical coefficient fromKrishnamurti* et al. *(2003)*

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 processed 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 superenseble (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 physics approach (MPE1) is not a good method to generate the dataset, since lack of spread causes a degeneracy in covariance matrix, which can not be easily inverted. MPMM1 ensemble, generated by MFS BHM winds 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. A third ensemble, MPMM2 ensemble, has been used to study the sensitivity on the free parameters that determine the SE: training period length and ensemble dataset quality. Sensitivity studies show that longer training period let the unbiased estimator be a bad estimate of the field we want to reproduce (seasonal cycle effects). 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.

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 than 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.

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. The employment of EOFs didn't give the desired improvements. 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). The key

point is the kind of field we want to reproduce, since we are considering a faster varying field, EOFs-based SE evaluated with this kind of "unbiased estimator" was not a good way to perform the regression. We should choose a new truth estimator, which can take in to account of seasonal cycle effects. Another consideration is related to the truncation order. In the new reference system defined by EOFs, the remaining part of the power spectrum can describe some small scale features, but these structures being of the same order of the noise, are filtered. Probably increasing the desired explained variance, small scale features are treated as signal and can be reproduced in the EOFs-based SE estimates.

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:

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

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.

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

- 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 & Philander (1981), the background viscosity is $\nu_b = 1^{cm^2}/_s$,and diffusivity is set as $\kappa_b = 0.1^{cm^2}/_s$. The ajdustable parameters are set as: $\nu_0 = 100^{cm^2}/_s$, $n = 2$ and $\alpha = 5$. For the convection case ($Ri < 0$), a maximum value of $1 \cdot 10^{6 cm^2}/_s$ 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 :

$$\tau_u^{ups} = \begin{cases} \tau_{i+1} & if\ u_{i+1/2} < 0 \\ \tau_i & if\ u_{i+1/2} \geq 0 \end{cases}$$
$$\tau_u^{tvd} = \tau_u^{ups} + c_u \left(\tau_u^{cen2} - \tau_u^{ups}\right);$$

  - $0 \leq c_u \leq 1$ 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 :

$$\tau_u^{muscl} = \begin{cases} \tau_i + \frac{1}{2}\left(1 - \frac{u_{i+1/2}\triangle t}{e_{1u}}\right) & \widetilde{\partial_i \tau} \qquad if\ u_{i+1/2} \geq 0 \\ \tau_{i+1/2} + \frac{1}{2}\left(1 + \frac{u_{i+1/2}\triangle t}{e_{1u}}\right) & \widetilde{\partial_{i+1/2} \tau} \qquad if\ u_{i+1/2} < 0 \end{cases}$$

– $\widetilde{\partial_i \tau}$ 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 :

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

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

– 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 :

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

It preserves simmetry 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 $1/24$ °(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).

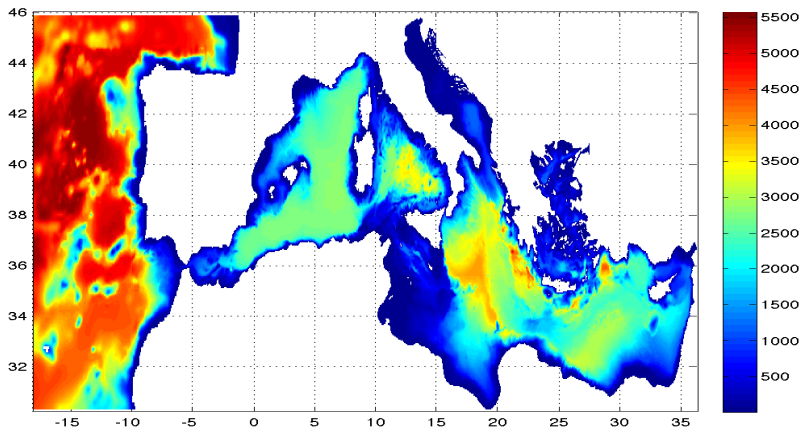

Figure B.1: *New model bathymetry and domain for the MFS2491 set up*

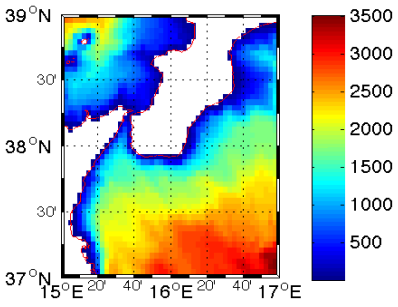

Figure B.2: *new Bathymetry near Messina Strait*

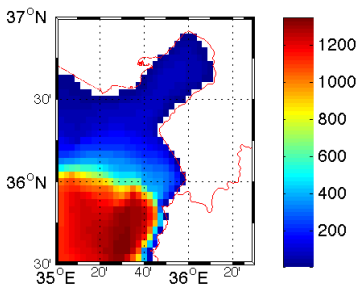

Figure B.3: *new Bathymetry near Iskenderum Gulf*

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

distribuition follow the equations (see **?**) :

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

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

In the current configuration the values are

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

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.

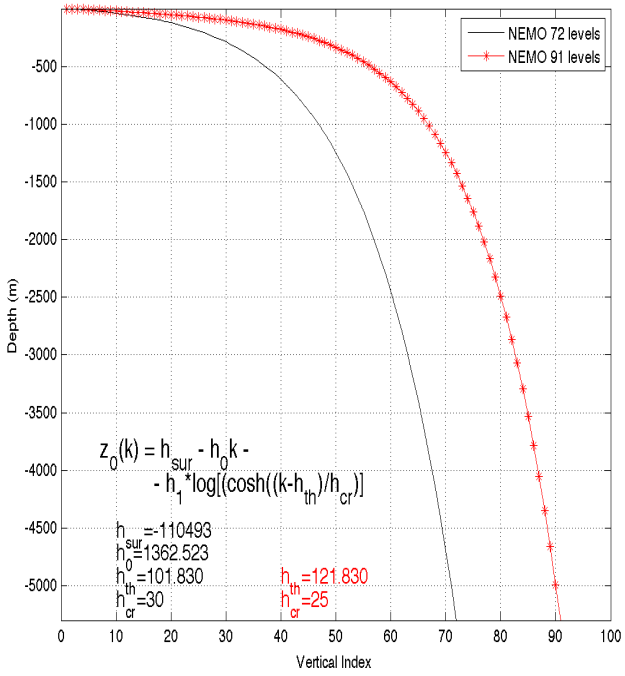

Figure B.4: *Vertical level distribution versus the depth of water column as computed by NEMO-OPA, new one 91 levels in red, in black the old distribution with 72 levels*

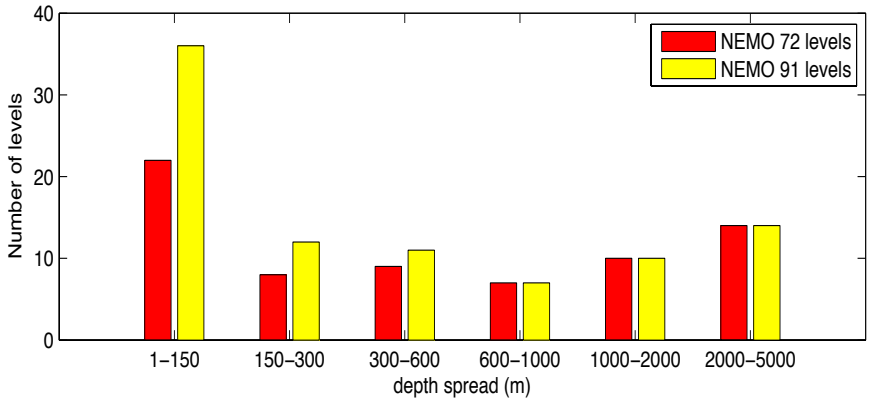

Figure B.5: *Histogram of relative frequency number of level for each interval depth, in old version in red, while the new in yellow.*

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.

Berliner, L. Mark, & Kim, Yongku. 2008. Bayesian design and analysis for superensemble-based climate forecasting. *J. climate*, **21**(9), 1891–1910.

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.

Dobricic, Pinardi, N, Adani, M, Tonani, M, Fratianni, C, Bonazzi, A, & Fernandez, V. 2007. Daily oceanographic analyses by mediterranean forecasting system at the basin scale. *Ocean science*, **3**, 149–157.

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.

Estubier, A., & Lévy, M. 2000. Quel schéma numérique pour le transport d'organismes biologiques par la circulation océanique. *Note techniquesdu pôle de modélisation, institut pierre-simon laplace.*

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.

Feddersen, 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, Lefaivre, 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, Cocke, 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 multi-model 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.

Marullo, Nardelli, B Buongiorno, Guarracino, M, & Santoleri, R. 2007. Observing the mediterranean sea from space: 21 years of pathfinder-avhrr sea surface temperatures (1985 to 2005): re-analysis and validation. *Ocean science*, 299–310.

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.

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.

Pacanowski, R. C., & Philander, S. G. H. 1981. Parameterization of vertical mixing in numerical models of tropical oceans.

Palmer, T. N., & Anderson, D. L. T. 1994. The prospects for seasonal forecasting,a review paper. *Quarterly journal of the royal meteorological society*, **120**(518), 755–793.

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.

Pettenuzzo, Daniele, Dobricic, Srdjan, Oddo, Paolo, & Pinardi, Nadia. 2010. Enhanced air-sea physics parametrization and assimilation of sst: a combined approach. *Geophysical research abstracts*, **12**.

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.*

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.

Robinson. 1966. An investigation into the wind as the cause of the equatorial undercurrent. *Journal of marine research*, **24**, 179–204.

Robinson, A.R, Leslie, W.G, Theocharis, A, & Lascaratos, A. 2001. Encyclopedia of ocean sciences. Jan, 1689–1705.

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.*

Simmons, A, Wallace, J, & Branstator, G. 1983. Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. atmos. sci.*, **40**(6), 1363–1392.

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. 2008. 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. 2011. 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.

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