Assimilation of sea surface temperature and sea surface height in a two-way nested primitive equation model of the Ligurian Sea

Alexander Barth

Departement of Astrophysics, Geophysics and Oceanography

University of Liege

A thesis submitted for the degree of

Docteur en Sciences

Committee members

Prof. Jean-Paul Donnay, president (University of Liege)
Prof. Jean-Marie Beckers (University of Liege)
Prof. Eric Delhez (University of Liege)
Prof. Robert H. Weisberg (University of South Florida)
Prof. Joaquin Tintoré (IMEDEA, CSIC-UIB)
Dr. Pierre Brasseur (LEGI/CNRS, Equipe MEOM)
Dr. Pierre De Mey (Dynamique Océanique, LEGOS)
Dr. Michel Rixen (NATO/SACLANT Undersea Research Centre)

Acknowledgements

I would like to thank my supervisor Prof. Jean-Marie Beckers who arouses my interest for oceanography and modelling. His expertise, understanding, and advises were essential for the realisation of this work. He guided me through the past four years while always giving to me the necessary freedom for developing new ideas.

I especially want to thank Dr. Michel Rixen for the enthusiasm and insight given to me during the development of this thesis. I am also indebted to Dr. Marilaure Grégoire for her patience and for the many words of encouragement.

I am grateful to Dr. Pierre Brasseur. He gave me the opportunity to work at the MEOM/LEGI in Grenoble. I had there various stimulating discussions, which widely influenced this work.

I am also indebted to Prof. Joaquin Tintoré, coordinator of the SOFT project, and to Dr. Alberto Alvarez. The collaboration with them and their inspiring ideas have been very beneficial to me.

To the rest of my thesis committee, Prof. Jean-Paul Donnay, Prof. Robert Weisberg, Dr. Pierre De Mey and Prof. Eric Delhez, I want also express my thanks for reviewing carefully this manuscript and for all of your insightful feedback.

I would like to thank especially Aida Alvera Azcárate for all of the support and encouragement. Without her support she has given me all along the stages of this thesis, this work would not have been possible.

I want also to acknowledge my colleagues, Caroline Raick, Abderaman Elkafazi, Luc Vandenbulcke and Zied Ben Bouallègue who have worked with me along the way. I am grateful for their help, suggestions and support.

I am particularly indebted to Dr. Miltos Markatos, Mireille Dondlinger, Huguette Croisier and Jonathan Margerit for kindly reading parts of this thesis and for providing me with many helpful comments.

I owe a special debt of gratitude to my family for their patience and encouragement. They have continuously supported me during all levels of my education.

This work was carried out with the support of the Fund for Research Training in Industry and Agriculture (FRIA, Belgium). The National Fund of Scientific Research (FNRS, Belgium) is acknowledged for the financing of a supercomputer. The work was realised in framework of the Satellite-based Ocean ForecasTing project (SOFT, EVK3-CT-2000-00028) of the European Union.

Abstract

A coarse grid primitive equation model of 1/4° resolution is implemented covering the whole Mediterranean Sea. Within this grid a 1/20° resolution model of the Liguro-Provençal basin and the northern part of the Tyrrhenian Sea is embedded. A third fine resolution model of 1/60° is nested in the latter one and simulates the dynamics of the Ligurian Sea (Barth *et al.*, 2003). Comparisons between one-way and two-way nesting in representing the Northern Current (NC) are made.

This system of nested models is coupled with a simplified Kalman-filter based assimilation method. The state vector for the data assimilation is composed by the temperature, salinity and elevation of the three models. The forecast error is estimated by an ensemble run by perturbing initial conditions and atmospheric forcings. The leading empirical orthogonal functions (EOF) of this ensemble are taken as the error covariance of the model forecast. This error covariance is assumed to be constant in time. Sea surface temperature (SST) and sea surface height (SSH) are assimilated in this system.

Data assimilation is also used to couple the hydrodynamic model with a statistical predictor of SST in the Ligurian Sea. The forecast improvement of this hybrid modelling system is shown and applications to operational models are highlighted.

Contents

Chapter 1

Introduction

Data assimilation is a generic technique for combining the dynamical knowledge of a system with other information in a statistically optimal way. Numerical models of environmental systems must generally represent a broad range of scales and processes. They are subject to uncertainties resulting from unresolved processes and approximative parameterisations. Data assimilation aims to reduce these uncertainties by correcting the model using other information *e.g.* satellite or *in situ* observations.

In order to model the dynamics of an environmental system, a complete description of its state at an initial time is necessary. An initial condition for the ocean is often obtained from climatology, which at best can only give a reliable estimation for large features. Direct observations give a much better estimation of the ocean state, but they are, especially in oceanography, costly and require a demanding infrastructure.

The dynamics of an environmental system are influenced by the evolution of adjacent systems. The ocean circulation, for example, is to a large extent dependent on the atmospheric heat, freshwater and momentum fluxes. It is a difficult task to obtain accurate information on these adjacent systems. They also generally involve substantial modelling efforts. The forcings are thus subjected to uncertainties and the impact of error in the forcing fields should be taken into account.

Besides the model itself, initial and boundary conditions are therefore further error sources. But once the system is closed, there is no possibility in the frame of the system dynamical formulation to impose other constrains. For example, if the surface heat flux is imposed to an ocean model, the observed sea surface temperature (SST) cannot be prescribed as an additional boundary condition. For a perfect model with the perfect initial and boundary conditions, the SST would be redundant and useless information. Imposing both, SST and heat flux, would result in an ill-posed and over-constrained problem. However, models are far from perfect and all additional information can be used to improve the model. In order to take more information into account than the necessary minimum, the model must be placed into a statistical context making allowance for the errors in the model dynamics, initial state and forcing terms. Data assimilation consists thus in finding the most likely state of the system in accordance with the additional

information available.

The two main theoretical frames of data assimilation are statistical estimation (Jazwinski, 1970; Gelb, 1974) and optimal control (Stengel, 1986; Chua and Bennett, 2001). Depending on the data assimilation algorithm, assimilation methods can also be divided into two different classes: sequential (nudging, optimal interpolation, 3D-Var, Kalman filter) and non-sequential assimilation methods (4D-Var, representer method, Kalman smoother). In each theoretical frame, both sequential and non-sequential scheme are derived.

The sequential assimilation methods consist in two successive steps: the forecast and the analysis. During the forecast step, the ocean state at the next time instant when observations are available, is predicted based on the ocean state from the last analysis. This forecast step is the same for all sequential methods. In the analysis step, the model forecast is combined with the observations. For non-sequential methods, the assimilation algorithm is more complex. Besides the forward integration, the information of the observations is also integration backward in time.

Nudging, optimal interpolation, Kalman filter and Kalman smoother are assimilation methods derived in the frame of statistical estimation. In nudging, only the error variances are considered. The analysis step is integrated in the governing equations of the model. A supplementary source or sink term is added to the prognostic equation of the observed variable. This term represents either a source or a sink if the variable is under- or overestimated respectively. This technique was applied, for example, to assimilate altimetry (*e.g.* Blayo *et al.*, 1996, in the North Atlantic), sea surface temperature and *in situ* temperature and salinity (*e.g.* Rixen, 1999, in the Alboran Sea). The main problem with this technique is that unobserved variables must adjust themselves through the model dynamics.

Optimal interpolation (Gandin, 1965; Daley, 1991) updates unobserved variables through physical assumptions and error statistics. Optimal interpolation methods are often tailored to a specific data type. Altimetry and sea surface temperature are the main observations used to correct general circulation models because remote sensing missions provide these variables with high spatial and temporal coverage. A central question in data assimilation is therefore how these surface observations can be connected to the unobserved variables at depth. For altimetry assimilation, the Cooper and Haines (1996) scheme displaces the complete water column such that the resulting dynamical height (based on a reference level with no pressure change) is in accordance with the remote sensed altimetry. This scheme has been applied to the global ocean (*e.g.* Fox *et al.*, 2000b; Haines, 2002) and works well for stratified regions. For weakly stratified seas the displacement of the Cooper and Haines (1996) scheme can become be too large.

The SOFA scheme (System for Ocean Forecasting and Analysis, De Mey and Benkiran, 2002) is based on an empirical orthogonal function analysis of the vertical temperature, salinity and surface elevation variability. These EOFs represent the covariance between

these variables and they are used to project the surface information to the depth. The SOFA scheme is more flexible than the Cooper and Haines (1996) scheme since it can take different regional and seasonal situations into account and it is not limited to altimetry observations.

These two optimal interpolation methods, or variants of these schemes, are often used in operational ocean models because they are numerically inexpensive and robust. The FOAM system (Forecasting Ocean Assimilation Model, Bell *et al.*, 2000; Bell, 2003) and the HYCOM/MICOM (Miami Isopycnic Coordinate Ocean Model/HYbrid Coordinate Ocean Model, Bleck, 2002) operational model are based on the Cooper and Haines (1996) scheme. The V1 and V2 classes of the MERCATOR models (Bahurel *et al.*, 2001, 2004), and the Mediterranean Forecasting System (Pinardi *et al.*, 2003; Demirov *et al.*, 2003) use the SOFA scheme.

For Kalman filter methods, the model itself (or the linearised model) is used to derive the relationship between the variables, the so-called error covariance. The forecast of this covariance is by far the most expensive step of the Kalman filter. A large number of simplified Kalman filters applicable to ocean models have been proposed to reduce the computational cost of this error covariance forecast.
A considerable simplification of the error covariance forecast can be done by variance-only error models (Daley, 1991, 1992). Only the model error covariance is propagated by the model dynamics. The correlation between the model variables is assumed constant and still need to be specified based on physical and statistical assumptions.

Instead of using the model with complete physics, the error forecast can be done with a model based on simplified physics. This approach reduces the cost associated to the error forecast of the Kalman filter (Dee *et al.*, 1985; Dee, 1990; Daley, 1992). However, it is not obvious if the simplified model reproduces the same error grow than the complete model.

The model can also be simplified by reducing the horizontal resolution of the model. The error covariance is thus computed on a coarser grid (Fukumori and Malanotte-Rizzoli, 1995; Cohn and Todling, 1996). Small scale errors not represented on this coarse grid are thus not corrected by this method. Miller and Cane (1989) applied a similar approach by using a low-order spectral model for the error forecast.

A consequent simplification of the original Kalman filter is the use of a steady-state error model (Fukumori *et al.*, 1993) and an asymptotically constant Kalman gain matrix. In fact, under some assumptions, the model error covariance and thus the Kalman gain converge to a constant matrix.

Other reductions deal with the explicit computation of non-null elements of matrices representing the linearised model (Parrish and Cohn, 1985; Jiang and Ghil, 1993). These studies have been carried out with a linear shallow water model. The explicit computation of the transfer matrix is only feasible with small models.

The linearised model can also be approximated by its projection on a given subspace. This subspace can either be the dominant eigenvectors of the linearised model (Cohn and Parrish, 1991; Cohn and Todling, 1996) or any other set of basis functions such as the EOFs of the model time variability (Cane *et al.*, 1996).

Another category of simplified Kalman filters works with the full model but approximates the error covariance matrix. An early attempt of Parrish and Cohn (1985) approximated the error covariance matrix by a banded matrix. This choice is motivated by the narrow stencils of ocean models. However, the positive definiteness of the error covariance is no longer guaranteed. The assimilation using a covariance matrix with (false) negative eigenvalues takes the model state away from the observations.

The representation of the error covariance matrix in terms of its leading eigenvectors and eigenvalues (Cohn and Todling, 1996) is the basis of the SEEK filter (Singular Evolutive Extended Kalman filter, Pham *et al.*, 1998; Brasseur *et al.*, 1999), the RRSQRT filter (Reduced Rank Square Root filter, Verlaan and Heemink, 1997) and the ESSE method (Error Subspace Statistical Estimation, Lermusiaux, 1997). The formulation of the error covariance in terms of its leading eigenvectors is the most accurate reduced rank representation of this error covariance matrix (Horn and Johnson, 1985, 1991). However, this approximated error covariance systematically underestimates the total error variance since a part of the error variance is lost in the truncation of the eigenvector decomposition (Heemink *et al.*, 2001). As in the Extended Kalman filter, the eigenvectors of the error covariance matrix are propagated by the tangent linear model.

The model forecast uncertainty can also be represented by an ensemble of likely model states (Miller *et al.*, 1994; Evensen, 1994, 2003). In the ensemble Kalman filter (EnKF), the model error covariance is obtained by a stochastic ensemble run. This Monte Carlo approach allows us to make a direct link between the error sources and the resulting imperfections of the model forecast. However, as in any Monte Carlo approach, the convergence is slow and high ensemble sizes are needed. It is also possible to use optimal and breed perturbations for assimilation to improve the convergence (Ehrendorfer and Errico, 1995; Toth and Kalnay, 1993).

The comparisons between the ensemble Kalman filter and Kalman filters based on an eigenvector decomposition of the error covariance have lead to interesting methods combining both approaches. Pham (2001) showed that an ensemble can also be built by the eigenvectors of the covariance matrix. The resulting ensemble represents more accurately the error covariance than the ensemble of the same size obtained by random perturbations. The error forecast of the so-called SEIK filter (Singular Evolutive Interpolated Kalman filter), a variant of the SEEK filter, is based on an ensemble and it can thus be carried out with the full nonlinear model instead of the tangent linear model.

Heemink *et al.* (2001) proposed two modifications of the RRSQRT filter, the POEnKF (Partially Orthogonal Ensemble Kalman Filter) and the COFFEE algorithm (Com-

plementary orthogonal subspace filter for efficient ensembles). The error space of the POEnKF is composed by the leading eigenvectors of the error covariance matrix and by randomly chosen ensemble members. They are propagated by the tangent linear model and the full nonlinear model respectively. During the analysis, the error directions of the ensemble already accounted by the leading eigenvectors are rejected. The COFFEE algorithm is similar except that the ensemble of states is directly sampled in the space perpendicular to the leading eigenvectors.

Due to the sequential design of the Kalman filter algorithm, all these Kalman filter techniques take only into account past observations. The Kalman smoother (*e.g.* Gelb, 1974) is an extension of the Kalman filter since it provides the best estimate using past and future observations of a hindcast experiment. The ensemble smoother (van Leeuwen and Evensen, 1996) estimates the time-space covariance by an ensemble of model trajectories. The time dimension is thus fully integrated in the estimation state vector. The ensemble Kalman smoother (Evensen and van Leeuwen, 2000) uses the EnKF as a first guess. At the analysis, the information of the current observations is propagated backward in time. An approximated smoother based on the classical Rauch-Tung-Striebel algorithm (*e.g.* Gelb, 1974) is the partitioned Kalman smoother (Fukumori, 2002). The model state is divided into several partition elements. The smoothing problem is then solved on each partition independently. The partitioning can be, for example, the decomposition into different regional sub-domains or the splitting of the baroclinic and barotropic modes (Fukumori *et al.*, 1999).

Variational assimilation methods are derived in the frame of optimal control. A set of unknown input parameters, the control variables, are adjusted in such a way that the model output comes as close as possible to the observations. In 3D-Var (*e.g.* Courtier *et al.*, 1998; Lorenc *et al.*, 2000), the model input is the model state and the model output is the model counterpart of the observations at the same instant. 3D-Var does not involve therefore the model dynamics and necessitates the adjoint[1] of the observational model relating the model state and the observations. It is a sequential assimilation method and it leads to the same result than optimal interpolation if the observational model is linear.

The variables to optimise can also be the uncertain initial conditions, the external forcings or model parameters as in the 4D-Var method (Dimet and Talagrand, 1986; Talagrand and Courtier, 1987). Here, the adjoint of the model dynamics must be derived. 4D-Var is able to take all observations into account, not only the past observations as the sequential methods. The main drawback of the 4D-Var method is the fact that all error sources must be control variables of the optimisation process. Therefore, it is not feasible to take the error introduced at each time step into account (Vidard *et al.*, 2003). The model itself is thus assumed to be perfect. This approach has been applied in oceanography to various physical and biological problems (Fennel *et al.*, 2001; Wenzel *et al.*, 2001; Spitz *et al.*, 2001; Stammer *et al.*, 2002). The representer method (Bennett, 1992; Chua and Bennett, 2001) is related to 4D-Var approach and has not this limitation

[1]The adjoint of a model can be defined as the transpose of its tangent linear model.

but it is derived in a linear frame. The 4D-Var method, the Kalman smoother and the representer method result in the same solution for linear models under the same hypotheses.

In this work, data assimilation is applied to a system of nested models. Model nesting consists in refining locally the regular model grid by a constant integer factor. It allows achieving a higher resolution locally in a domain of interest without carrying the high resolution over the entire domain. For telescopically nested models, the grid is successively refined several times and the resolution can thus be increased gradually. This reduces the problem at the interface between the grids due to an abrupt change in resolution. Traditionally (*e.g.* Fox and Maskell, 1995) two types of nesting are distinguished: (i) one-way or passive nesting (*e.g.* Robinson and Sellschopp, 2002; Zavatarelli and Pinardi, 2003; Garnier *et al.*, 2004) and (ii) two-way or interactive nesting (*e.g.* Oey and Chen, 1992; Fox and Maskell, 1996; Ginis *et al.*, 1998; Blayo and Debreu, 1999; Barth *et al.*, 2003). In one-way and two-way nesting the boundary conditions for the high resolution grid are interpolated from the coarse model values. The high resolution model depends therefore on the dynamics of the coarse resolution model. In one-way nesting, the coarse resolution model is not influenced by the high resolution model. This has the advantage that the models can be run successively, starting with the coarse resolution model. However, after long-term integration, both models may become inconsistent. The two-way nesting avoids this problem. After the boundary conditions are imposed on the fine model grid and both models are integrated one step forward, the values of the coarse model grid in the overlapping domain are replaced by the average value of the fine model grid. The underlying idea of this feedback is that the forecast of the fine model grid is more accurate than the result of the coarse model grid. The two-way nesting approaches differ in the way how the boundary conditions are interpolated, how the feedback is computed and by some possible special treatments at the boundaries.

Model nesting was first applied by Spall and Robinson (1989). This technique is now extensively used in operational oceanography (*e.g.* Robinson and Sellschopp, 2002; Zavatarelli and Pinardi, 2003) since it responds to the end-user demand of high resolution forecasts of *e.g.* regions critical for human activity.

The present work deals with data assimilation in combination with a nested model. In the case of one-way nested models, the fine resolution model is generally run after the coarse resolution model. Therefore the observations are generally assimilated into each model independently as it has been done for example in the FOAM system (*e.g.* Bell *et al.*, 2000; Martin *et al.*, 2002) or in ADRICOSM system (*e.g.* Pinardi *et al.*, 2004; Zavatarelli and Pinardi, 2003). Also in two-way nested models, each model has been treated separately for the assimilation of observations (*e.g.* Onken *et al.*, 2003).

A novel approach that consists in adopting a state vector spanning all nested models is applied in this work. This approach allows to take into account the correlation between the different models and to keep a perfect consistency between the models after assimilation. Also some implementation issues encountered when the assimilation is applied to nested

are discussed.

1.1 Observations

In this section we will introduce the two types of observations used in the assimilation experiments: sea surface temperature and sea surface height. Since there are several ways to define these variables, it is useful to recall these definitions. It is also important to take the measurement procedure and the resolved scales into account to know how the model variables can be related to the observations.

1.1.1 Sea surface temperature

The temperature structure of the upper ocean is quite complex and depends on various processes acting at the surface and on the first meters of the water column. The air-sea interface is affected by processes responsible for the heat transfer between the ocean and the atmosphere: the net longwave radiation, the heat convection between the air and the sea and the heating or cooling due to moisture transfer. Within the ocean, the heat is redistributed by molecular diffusivity and turbulent diffusivity. The heating related to the light absorption is another process modifying the temperature of the water column. All these processes acting at different depths and depth ranges complicate the definition of the SST and the intercomparison with different instruments and models. According to Donlon (2002) the temperature at different depths can be defined depending on the dominant processes.

The interface SST, $\mathrm{SST_{int}}$, is the temperature at exactly the interface between the atmosphere and the ocean. This temperature governs the surface heat transfers between the ocean and the atmosphere. However, this temperature cannot be measured directly by current technology.

The skin SST, $\mathrm{SST_{skin}}$, is the temperature of the thin surface layer where the molecular diffusion and heat conductivity dominate. The thickness of the layer influenced by these skin effects and solar heating is less than 500 μm. The temperature gradient within this layer is given by the atmospheric heat fluxes. The skin temperature is measured by radiometers operating in the spectral region of 3.7 to 12 μm (thermal infrared). The Advanced Very High Resolution Radiometer (AVHRR) and Along-Track Scanning Radiometer (ATSR) are sensors measuring in the thermal infrared (Wick, 2002). The signal of these sensors penetrates approximately 10 μm into the water column.

The sub-skin SST, $\mathrm{SST_{sub\text{-}skin}}$, represents the temperature at the base of the skin layer. Beyond the skin layer, the temperature is affected by turbulent mixing and insolation. Sensors working with microwave measure the sub-skin SST. For example, the Scanning Multichannel Microwave Radiometer (SMMR) and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) measure in this frequency band (Wick, 2002). The depth of the sub-skin SST is of the order of 1 mm.

The bulk SST or SST_{depth}, is the temperature of the water column under the skin layer at a depth completely immersed in the turbulent ocean. This temperature is observed by *in situ* buoys, CTD and XBT measurements.

The constant temperature layer SST, SST_{CTL}, is the sea temperature at a depth where the temperature is not significantly affected by the solar radiation. The diurnal variations are less than 0.2 °C. The effective position of this constant temperature layer can therefore vary from one location to another.

SST of hydrodynamic models resolving the insolation diurnal cycle corresponds to the bulk SST. Since the diurnal cycle has little impact on large time and space scales, models representing large scales are forced by daily averaged atmospheric forcings. In this case, the model SST is related to the constant temperature layer SST.

Figure 1.1 shows the approximate depths of each of these SST definitions. Since the solar radiation is an important process, the night heat distribution in the upper ocean is different from the day temperature profile. Due to this time and space variation, the SST should be given as a temperature measurement at a specified depth and ideally at a specified time of day.

1.1.2 Altimetry

Altimetry measurements are realised by a radar onboard of a satellite sending a high frequent signal (more than 1700 pulses per second) towards the ocean surface. These pulses are reflected by the ocean surface. The satellite measures the time between the emission of the pulse and the reception of the echo (figure 1.2). The distance between the satellite and the ocean surface is determined by this time delay and the speed of the light in the atmosphere. The air density and the content of water vapour in the troposphere and the density of free electrons in the ionosphere affect the speed of the signal and are taken into account.

The altitude of the satellites is known from the ellipsoid of reference. The sea surface height (SSH) is defined by the height of the ocean surface computed from this ellipsoid of reference. Thus, the SSH is obtained from the difference between the distance separating the ocean surface and satellite and the altitude of the satellite. The satellite provides us the SSH along the ground track of the satellite. The orbit of the satellite is chosen in such a way that the ground tracks of the satellite are repeated. For example, the repeat cycle of the Topex/Poseidon and the Jason-1 satellites is 9.91 days.

Besides the already mentioned tropospheric and ionospheric effects other corrections are applied to the altimetric signal. These corrections include the electromagnetic bias function of the sea state and the solid earth tides. Other corrections are carried out to eliminate the signal of high frequent ocean processes that are only weakly coupled to the ocean meso- and large scale dynamics.

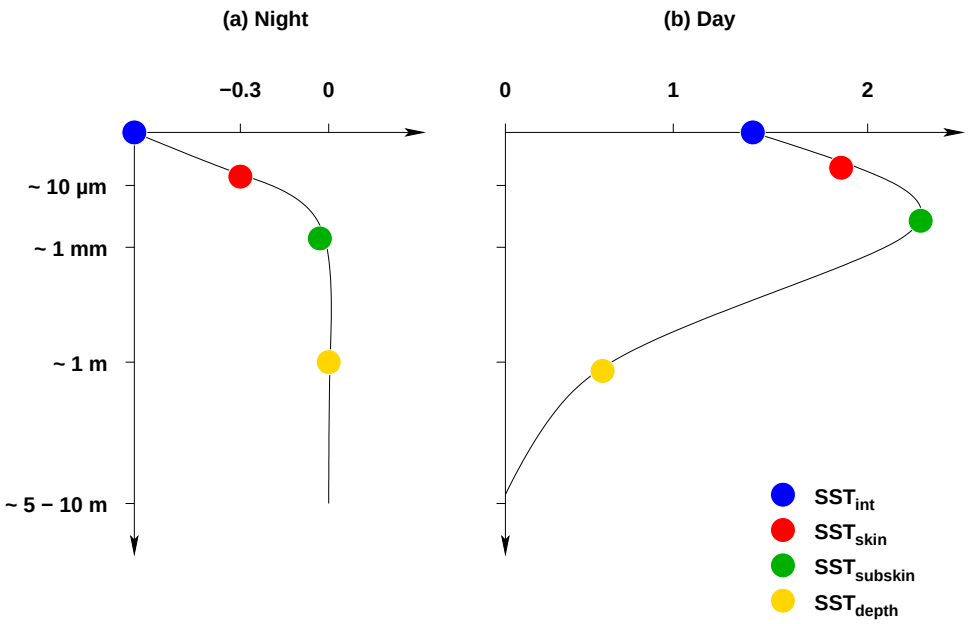

Figure 1.1: The difference between the pre-dawn temperature at 1-5 m depths and the temperature at various depths during night (a) and during day (b). Adapted from Donlon (2002)

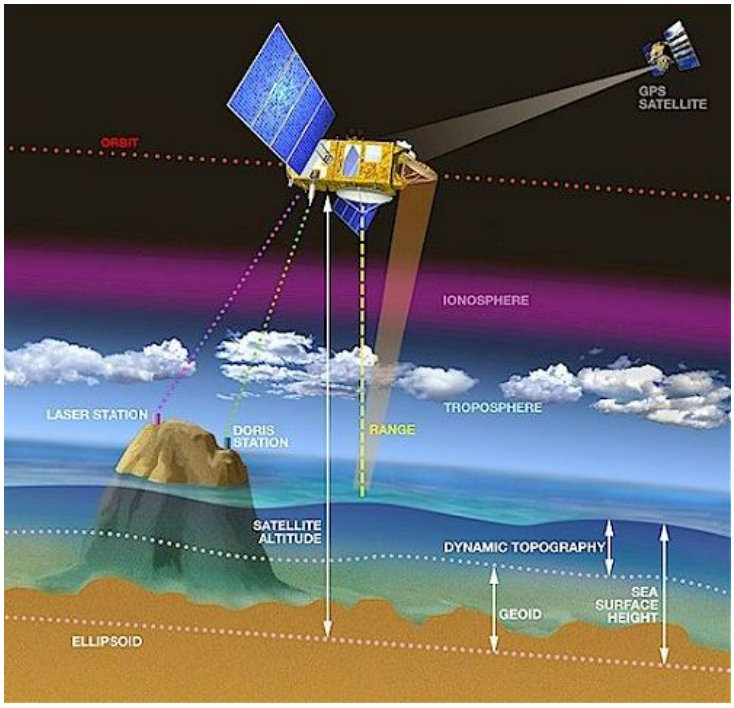

Figure 1.2: Schematic view of altimetry remote sensing and the relationship between ellipsoid of reference, geoid, SSH and dynamic topography (from `http://www.jason.oceanobs.com/`.

Like any membrane of a barometer is deformed by the atmospheric pressure, also the ocean surface is lowered or raised due to the load of the atmospheric air column. This phenomenon is called the inverse barometer effect. The resulting change of the sea level is directly related to the atmospheric pressure. This signal can be easily removed in order to isolate signal of meso- and large scale ocean processes. Often ocean models do not take into account the atmospheric pressure at the surface. This choice is due to the fact that the atmospheric pressure is compensated by the change of the surface elevation and the pressure at depth is not modified. Furthermore, regional models would require a boundary condition resolving the barotropic response of the ocean due to the atmospheric pressure.

The largest dynamical variation of the ocean surface is due to tides. Tides in the open ocean have amplitudes of the order of up to 50 cm with considerable geographic variation. These tides are removed in order investigate the signal at larger time scales. Tidal models such as the model of Cartwright and Ray (1990) and of Le Provost *et al.* (1998) are used for this purpose. Near the coast the tidal signal is sometimes not completely removed due to the coarse resolution of the tidal models. The error associated to the altimetry with tidal correction is therefore higher near the coast.

The SSH obtained after the application of the mentioned correction, is further divided into two very distinct contributions: first, the SSH to which the ocean at rest would conform and the second part of the signal is consequently the SSH due to the ocean circulation.

The former component of the SSH signal is called the geoid. The geoid is a gravitational equipotential surface and the gravity force is always perpendicular to it. The structure of the geoid is due to the mass distribution in the Earth. The geoid amplitudes are of the order of 100 m (Lemoine *et al.*, 1998). Estimations of the geoid exist but the structures at small scale are relatively unknown. The typical accuracy is of 10 to 20 cm RMS for a spherical harmonic development of the order of 20, i.e. wavelengths larger than 2000 km (Le Traon *et al.*, 2001).

The deviation of the SSH from the geoid is called the dynamic topography. It is affected by a series of dynamic ocean processes such as surface gravity waves, Rossby and Kelvin waves, frontal systems, eddies and currents. A broad range of ocean phenomena can therefore be detected by altimetry. The fraction of the oceanic signal is of the order of 10 cm and much smaller than the spatial variation of the geoid.

Ocean circulation models do not take of course the inhomogeneous mass distribution of the Earth into account. With a constant gravitational field, the model geoid is the surface with a depth equal to zero. Indeed, this surface corresponds also in the model to an ocean at rest without forcings. The model elevation represents therefore only the dynamic topography.

In principle, it would be sufficient to subtract the geoid from the SSH, to obtain the

dynamic topography. But due to the low precision of the geoid another approach is generally adopted. Sea level anomalies are computed by subtracting a time mean from the SSH measurements. A mean dynamic topography (MDT) can be determined by averaging the model results of the corresponding time period. Several alternative ways for determining the MDT exist and they are based on the Levitus climatology (Levitus and Boyer, 1994; Levitus *et al.*, 1994), *in situ* measurements of the dynamic topography (Uchida *et al.*, 1998) or from drifters (Rio and Hernandez, 2004). In order to avoid the geoid problem, the altimetry measurements are referenced from the MDT. The sea surface anomalies therefore do not contain the signal of permanent ocean currents.

1.2 Basic concepts

1.2.1 The state vector and the model

The state vector regroups all variables necessary to describe the state of the system. Typically all prognostic variables at all model's grid points can be found in the state vector. The model allows us to forecast the state vector at a future time t_{i+1}, if the state vector is known at a time t_i. Formally, the model is in this view a function mapping the state vector $\mathbf{x}_i$ at time t_i to the state vector $\mathbf{x}_{i+1}$ at time t_{i+1}.

$$\mathbf{x}_{i+1} = M_i(\mathbf{x}_i) \tag{1.1}$$

Even if the time lag $t_{i+1} - t_i$ involves a large number of model time steps on the discretised time axis, the model equations can always be formulated by equation (1.1). The function M_i already includes the boundary conditions. In some notations, they are expressed as an additional term of equation (1.1).

Some difficulties arise when the numerical scheme of the model involves different time instants. The "leapfrog" scheme is for example based on two successive past time steps. To strictly maintain the formalism, the state vector should be composed of all prognostic variables at two model time steps. In practice, however, this approach is not very interesting. It is preferable to start the model with a first order time scheme and then to continue with the second order leapfrog scheme. This last approach was used for example by Testut *et al.* (2003).

For various reasons discussed later, the model forecast $\mathbf{x}_i$ differs in general from the true model state $\mathbf{x}_i^t$ by an unknown error $\boldsymbol{\zeta}_i$.

$$\boldsymbol{\zeta}_i = \mathbf{x}_i^t - \mathbf{x}_i \tag{1.2}$$

By comparing the model to the observations and climatology, we can assess the skill of the model and evaluate a typical magnitude of this error. Data assimilation methods can only be applied if we know what a probable error is and what not. In the most general form, we can conceptualise the model error by a random vector following a probability

density function, or pdf, $p_{\boldsymbol{\varepsilon}_i}$.

Most assimilation schemes assume that the error is Gaussian distributed and has zero mean. This implies that our model is unbiased. The pdf is then completely characterised by the error covariance matrix $\mathbf{P}_i$.

$$\mathbf{P}_i = E\left[\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^T\right] \tag{1.3}$$

where E is the expectation operator. Like all covariance matrices, $\mathbf{P}_i$ is symmetric and positive definite.

Example

Let's consider the case of an ocean system governed by the primitive equations with a free surface. If data assimilation is used in order to provide the best estimation of the ocean state (data assimilation can also be implemented for parameter optimisation), the state vector would contain the surface elevation, the horizontal velocity, temperature, salinity, and the parameters of the turbulence closure scheme involved in a transport equation. The state vector is composed of all these variables at all grid points corresponding to a sea point. The number of elements in the state vector n is therefore of the order of 10^5 to 10^7. With the increasing computing capacity, the resolution of the numerical models is constantly enhanced and thus the size of the state vector increases.

The dimension of the state vector is a critical parameter in data assimilation. Assimilation schemes such as the Kalman filter (Kalman, 1960) are known for a long time but the important size of the state vector makes the direct implementation unpractical. The Kalman filter, for example, requires the storage of n^2 numbers and at least n^2 floating-point operations or n model integrations instead of one model integration for a simulation without the Kalman filter.

State vector augmentation

In some implementations, diagnostic variables are also added to the state vector. For example, the surface elevation of a rigid lid model can be part of the state vector in order to simplify the assimilation of SSH. Other variables related to the dynamics of the system such as atmospheric forcing or even scalar parameters of the model can also be present in the state vector if they have to be optimised by assimilation.

For each added variable and parameter, equation (1.1) must be extended by an evolution equation. For parameter optimisation, the "dynamical" evolution of vector holding the parameters $\mathbf{p}_i$ can simply be the persistence (Kivman, 2003):

$$\mathbf{p}_{i+1} = \mathbf{p}_i \tag{1.4}$$

According to the model dynamics, the parameters are supposed to be constant and valid for all situations encountered along the model evolution. The Kalman filter intents to provide the best estimation of the parameters taking into account all past observations.

In such a sequential assimilation algorithm, the parameter will change with time the more and more observations are available as the model runs forwards. If the hypotheses of the Kalman filter are verified, the parameter will tend to the optimal value taking into account all observations. This optimised value can be used for a new model integration. Non-sequential algorithms such as 4D-Var and Kalman smoother naturally provide the best parameter values of the complete time interval.

As Kivman (2003) has pointed out, even if the initial model is linear, the augmented model can be nonlinear. For instance, parameters in biological models and in turbulence closure schemes have in general a highly nonlinear impact on the model result since they can change completely the dynamical regime of the system. Therefore, parameter optimisations are very challenging problems in data assimilation (Fennel *et al.*, 2001; Wenzel *et al.*, 2001; Spitz *et al.*, 2001).

Furthermore, the model parameters cannot take arbitrary values. For example, most parameters must be positive. Consequently, any probability distribution expressing uncertainties in the parameter space is essentially non-Gaussian.

This technique is called state vector augmentation. These examples for state vector augmentation described here should illustrate that the concept of the state vector is very flexible and depends on the data assimilation problem. With state vector augmentation the problem is rewritten in order that its formal expression is simpler. But it is just a rearrangement of the unknowns of the problem to solve. For example, a parameter optimising 4D-Var scheme with state vector augmentation would yield exactly the same practical problem as the formulation based on separated system state vector and the parameters.

1.2.2 The observations and observational model

Observations are relevant to the studied system if they can be related to the true state of a system by:

$$\mathbf{y}_i^o = H_i(\mathbf{x}_i^t) + \boldsymbol{\varepsilon}_i \tag{1.5}$$

where the observational model H_i makes the link between model and observations. Often this function simply performs an interpolation to extract the model result at the observation locations (figure 1.3). In other cases, H_i can be a complex function computing the sea surface height of a rigid lid model or the brightness temperature from sea surface temperature and salinity. The error term $\boldsymbol{\varepsilon}_i$ takes the instrumental error into account and the fact that some processes are not resolved by the model. The latter error is called the representativeness error or the error of unresolved scales.

An example of the representativeness error is the difference between the skin temperature measured by satellites and the model surface temperature. The model surface layer resolution is generally of the order of one meter and does not resolve the skin layer,

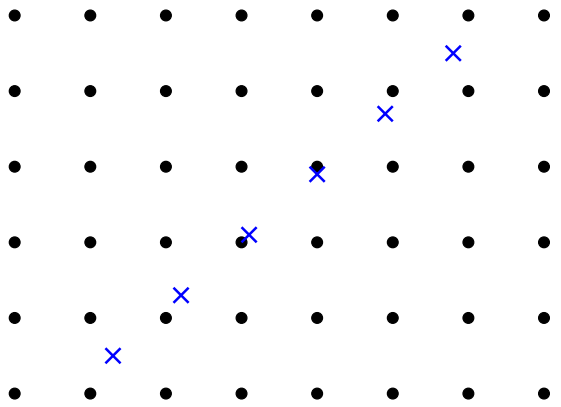

Figure 1.3: The operator $H_i(\cdot)$ interpolates the model results to the grid of the observations. This figure shows schematically the position of altimetry measurements (crosses) and the elevation grid-points of the model (dots). In the simplest form, $H_i(\cdot)$ performs a bilinear interpolation. The model results and the measurements, however, do not represent the elevation of a single point but a mean over a certain range. This can also be taken into account in the measurement operator.

whose depth is less than 500 μm.

The representativeness error is generally much larger than the instrumental error. This part of the signal in the observations should not be injected into the model since the model is not able to handle this information correctly. However, one should bear in mind that this error part does not correspond to a problem in the observations but it rather represents an inadequacy of the dynamical or the observational model.

The error term $\boldsymbol{\varepsilon}_i$ is treated as a random vector following the probability distribution $p_{\boldsymbol{\varepsilon}_i}$. In many assimilation schemes, the errors are supposed to be unbiased and the mean of the random vector $\boldsymbol{\varepsilon}_i$ is zero. A Gaussian distribution is widely used to model this pdf:

$$p_{\boldsymbol{\varepsilon}_i}(\boldsymbol{\varepsilon}) = (2\pi)^{-m/2}(\det \mathbf{R}_i)^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^T \mathbf{R}_i^{-1}\boldsymbol{\varepsilon}\right) \tag{1.6}$$

where $\mathbf{R}_i$ is the error covariance of the observations. Van Leeuwen (2003) suggested that the pdf might decrease too rapidly and does not take "outlier" observations into account. In certain circumstances, a Lorenz pdf might be more appropriate. For all assimilation methods presented here, the observational error is assumed to be uncorrelated in time and uncorrelated to the errors affecting the model.

Chapter 2

The GHER model

2.1 Primitive equations

The GHER (GeoHydrodynamics and Environment Research) model (Beckers, 1991) is used for the present telescopically nesting implementation. The free surface, hydrostatic, primitive equations under the classical Boussinesq and *beta*-plane approximations are solved. The equations are expressed in a Cartesian coordinate system where x, y and z are the longitude (positive to the East), latitude (positive to the North) and the height (zero at surface and negative in the water).

$$\nabla \cdot \mathbf{v} = 0 \tag{2.1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{u} + f \mathbf{e}_z \wedge \mathbf{u} = -\nabla_h q + \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial \mathbf{u}}{\partial z}\right) \tag{2.2}$$

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial T}{\partial z}\right) + \frac{1}{c_p \rho_0}\frac{\partial I}{\partial z} \tag{2.3}$$

$$\frac{\partial S}{\partial t} + \mathbf{v} \cdot \nabla S = \frac{\partial}{\partial z}\left(\tilde{\lambda}\frac{\partial S}{\partial z}\right) \tag{2.4}$$

$$\frac{\partial k}{\partial t} + \mathbf{v} \cdot \nabla k = \tilde{\nu}\left\|\frac{\partial \mathbf{u}}{\partial z}\right\|^2 - \tilde{\lambda}\frac{\partial b}{\partial z} - \varepsilon + \frac{\partial}{\partial z}\left(\tilde{\nu}\frac{\partial k}{\partial z}\right) \tag{2.5}$$

where f is the Coriolis frequency, q the generalised pressure, b the buoyancy, T the temperature, c_p the heat capacity of sea water, ρ_0 the reference density, S the salinity, $\tilde{\nu}$ and $\tilde{\lambda}$ are the eddy viscosity and diffusivity, k the turbulent kinetic energy and ε is the dissipation of turbulent kinetic energy. The velocity $\mathbf{v}$ is decomposed into its horizontal $\mathbf{u}$ and vertical w component:

$$\mathbf{v} = \mathbf{u} + w\mathbf{e}_z \tag{2.6}$$

The operators ∇ and ∇_h are defined as:

$$\nabla = \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z} \tag{2.7}$$

$$\nabla_h = \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} \tag{2.8}$$

$$\tag{2.9}$$

If one subtracts from the pressure p the hydrostatic pressure due to a constant density ρ_0, one obtains the generalised pressure (apart from a constant factor ρ_0):

$$q = \frac{p}{\rho_0} + gz \tag{2.10}$$

where g is gravitational acceleration. The buoyancy b is given by the state equation $\rho(T, S)$:

$$b = -\frac{\rho(T,S) - \rho_0}{\rho_0} \tag{2.11}$$

Under the hydrostatic approximation, the generalised pressure and the buoyancy are related by:

$$\frac{\partial q}{\partial z} = b \tag{2.12}$$

The vertical turbulence is parameterised by a k turbulent kinetic energy closure scheme (Nihoul *et al.*, 1989). The evolution equation for the dissipation of turbulent kinetic energy ε is replaced by an algebraic relationship. The turbulence closure scheme is based on the Brunt-Väisälä or stability frequency N, the Prandtl frequency M, the Richardson number R_i and the Richardson flux number R_f defined by:

$$N^2 = \frac{\partial b}{\partial z} \tag{2.13}$$

$$M^2 = \left\| \frac{\partial \mathbf{u}}{\partial z} \right\|^2 \tag{2.14}$$

$$\tilde{R}_i = \frac{\gamma}{2} \frac{N^2}{M^2} \tag{2.15}$$

$$R_f = \frac{\tilde{\lambda}^b N^2}{\tilde{\nu} M^2} \tag{2.16}$$

The dissipation of turbulent kinetic energy ε, the turbulent viscosity $\tilde{\nu}$, the turbulent diffusivity $\tilde{\lambda}$ and the mixing length l_m are parameterised by:

$$\varepsilon = \alpha_k \frac{k^2}{16} \tag{2.17}$$

$$\tilde{\nu} = \frac{1}{2} \alpha_k^{1/4} k^{1/2} l_m \tag{2.18}$$

$$l_m = (1 - R_f)\, l_n(z) \tag{2.19}$$

$$\tilde{\lambda} = \Psi^b \tilde{\nu} \tag{2.20}$$

$$\Psi^b = \gamma \sqrt{1 - R_f} \tag{2.21}$$

The parameters α_k and γ are dimensionless empirical quantities of the order of 1 and $1.1 - 1.4$ respectively. The length $l_n(z)$ represents the mixing length at neutral

stratification. It is assumed to depend only on geometry of the basin (Beckers, 1991).

From (2.16), (2.15), (2.20) and (2.21), one can show that the Richardson number R_i and the Richardson flux number R_f are related by the following expression:

$$1 - R_f = \left(\tilde{R}_i + \sqrt{\tilde{R}_i^2 + 1} \right)^{-2} \tag{2.22}$$

2.2 Discretisation

The domain is discretised horizontally along parallels and meridional lines. In the vertical, the model uses a double-sigma coordinate (Beckers, 1991). This coordinate transform is a particular feature of the GHER model and it is based on the observation that the most abrupt change in the bathymetry is generally located at the shelf break. It divides the model domain in an upper and a lower region at a given depth h_l (figure 2.1). In each region, a traditional sigma transform is then performed. The depth h_l is related to the mean depth of the shelf break. Unlike a normal sigma coordinate (Haney, 1991), the double σ coordinate can represent an abrupt shelf break if it is located at the depth h_l.

Scalar variables (elevation, temperature, salinity and turbulent kinetic energy) are discretised at the centre of each grid box and at the interface of each box we define the normal velocity component (a C-grid in Arakawa nomenclature). For computational efficiency, the baroclinic mode is integrated with a much larger time step than the barotropic mode. All models of the nesting system are integrated with the same time step, i.e. 4 s for the barotropic mode and 80 s for the baroclinic mode. The numerical scheme conserves volume, heat and salt. In addition, the advection scheme is a monotone TVD (Total Variation Diminishing) scheme preserving fronts and gradients. Other details of the GHER model can be found in Beckers *et al.* (2002).

2.3 Atmospheric fluxes

2.3.1 The momentum flux

The winds at the air-sea interface drag the surface water along its direction. This wind stress τ gives the momentum flux between ocean and atmosphere and it is parameterised by:

$$\tau = C_D \rho_a \|\mathbf{u}_a\| \mathbf{u}_a \tag{2.23}$$

ρ_a is the air density and $\mathbf{u}_a$ the wind vector at the reference level. The drag coefficient C_D is parameterised by the scheme of Kondo (1975). The momentum flux is a vector with the same direction of the wind vector. Therefore, a positive momentum flux is a momentum transfer from the atmosphere to the ocean (figure 2.2).

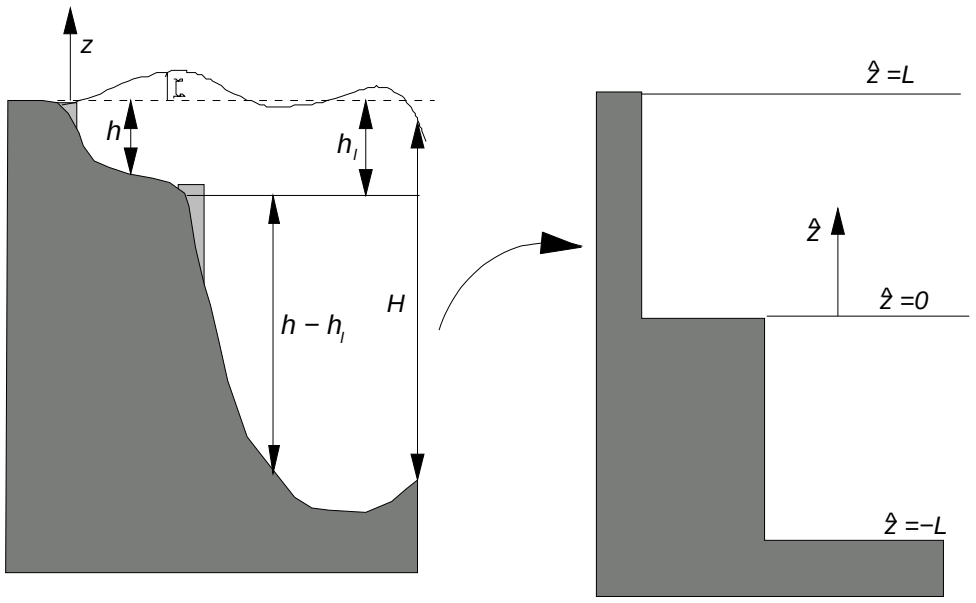

Figure 2.1: The double-sigma coordinate of the GHER model. ζ and h are the surface elevation and the ocean floor depth computed from the reference surface $z = 0$ respectively. h_l corresponds to the average depth of the shelf break. The physical domain (left) is transformed into the modelling domain (right) with the coordinate $\hat{z}$.

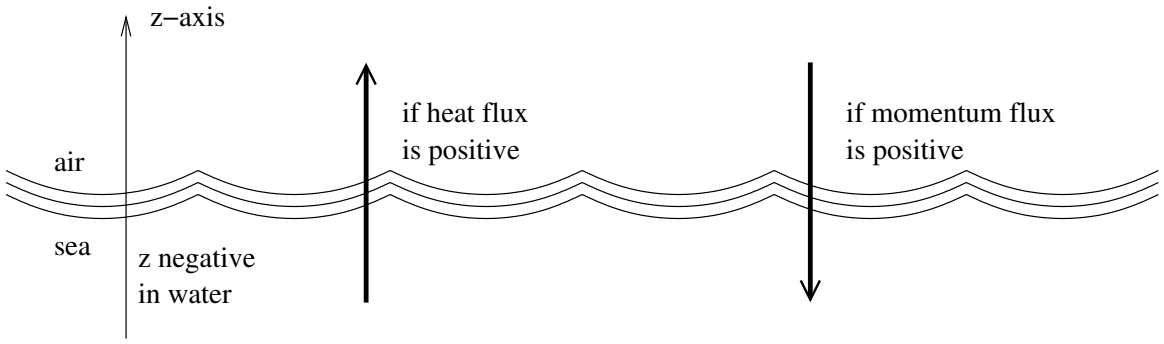

Figure 2.2: Sign convention for the atmospheric fluxes. The heat fluxes (shortwave radiation, net longwave radiation, latent and sensible heat flux) are positive if heat is transferred from the ocean to the atmosphere. For the momentum flux the opposite convention is adopted such that the momentum flux vector has the same direction than the wind vector.

2.3.2 Heat flux

The original GHER model was forced by a non-interactive total surface heat flux, a momentum flux and a surface fresh water flux. The total surface heat flux is the sum of the latent heat flux, the sensible heat flux, the net shortwave (or solar) radiation and the net longwave (or back) radiation. A relaxation toward the climatologic sea surface temperature and salinity introduces an interactive correction on the surface heat flux and the fresh water flux. The atmospheric heat forcings have a great impact on the SST forecast skill. The model was substantially improved by refining the implementation of the atmospheric fluxes, as explained in the following sections.

2.3.3 The surface shortwave radiation

We compared the 1D vertical GHER model based on the same k-turbulence parameterisation as the 3D model with the results of the GOTM model (Burchard, 2002). The GOTM model is an advanced turbulence model and in several studies, it has been validated against observations (*e.g.* Burchard *et al.*, 2002). It uses a k-ϵ turbulence scheme and treats the shortwave radiation as a volume heat source.

Without the temperature relaxation the sea surface temperature of the 1D GHER model reaches unrealistic high temperatures in summer (figure 2.3). The enhanced stratification prevents the mixing with the lower water layer and the surface heating is reinforced.

In fact, by implementing the solar radiation in GHER model also as a volume heat source the vertical structure of the temperature was significantly improved and the results were much closer to the results of the GOTM model but differences remain.

If we want to distribute the solar energy into the water column we have to make assumptions about the light absorption. By considering only two visible frequencies, the radiation flux as a function of depth is given by the following equation ($z = 0$ at the surface and negative in water):

$$I(z) = |Q_s| \left(A\, e^{g_1 z} + (1 - A)\, e^{g_2 z}\right) \tag{2.24}$$

where Q_s is the light intensity at the surface, $A = 0.58$ is the fraction long-wave solar energy and $g_1 = 0.35\ m^{-1}$ and $g_2 = 23.0\ m^{-1}$ are the absorption coefficients of the short-wave and long-wave solar energy respectively. This distribution of the light intensity corresponds to the water of type I according to the classification of Jerlov (1968) and is the same used in the GOTM model. Also the absorption profile proposed by Ivanoff (1977) was tested. Lacroix and Grégoire (2002) use this absorption formulation for the 1D GHER in the Ligurian Sea. But the differences between both absorption profiles and the resulting temperature distribution are small.

The insolation represents the main energy input into the ocean. Solar energy Q_s entering into the water arrives at the sea surface either directly or by diffusion. The

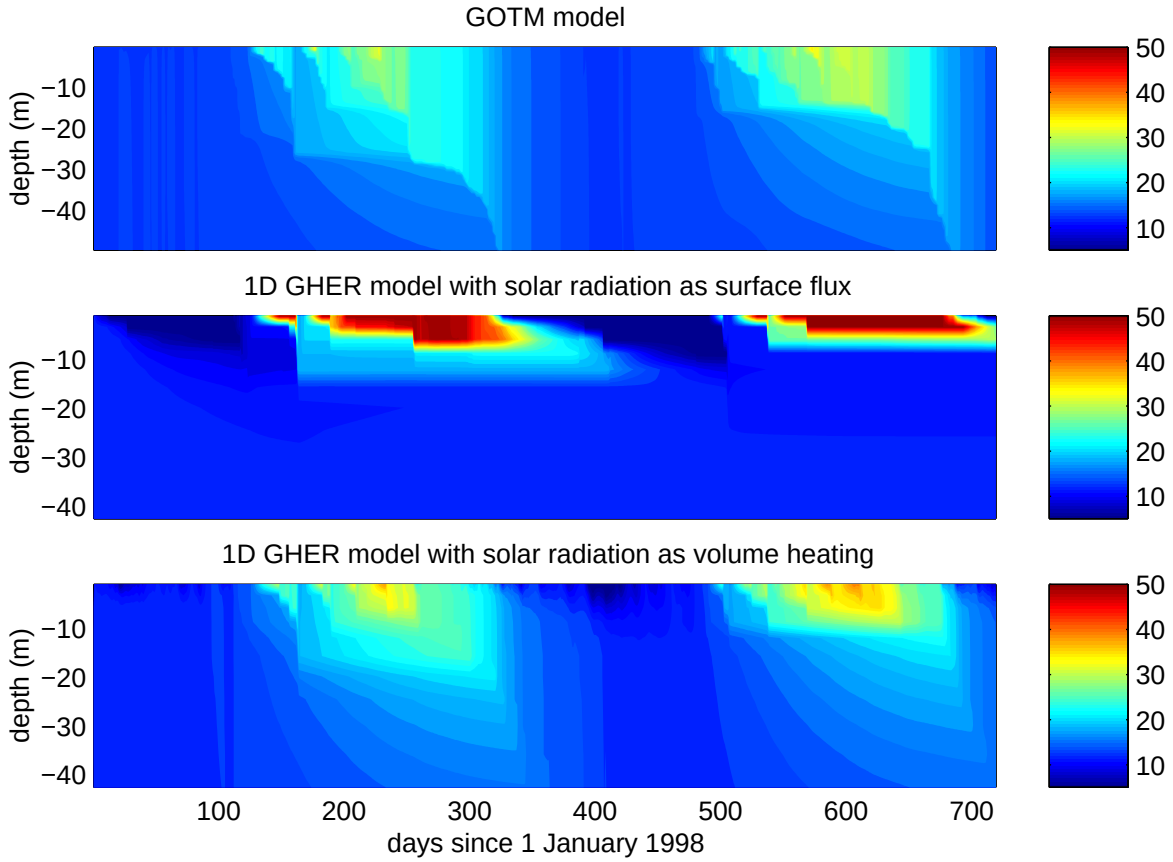

Figure 2.3: Temperature profiles in a point of the Ligurian Sea with 42°50' N latitude and 7°30' W longitude. The day 0 is the 1 January 1998. The panel (a) shows the temperature of the GOTM model, (b) the temperature of the 1D-GHER model with the shortwave radiation as a surface flux (colour bar is saturated since the temperature reaches unrealistic values) and (c) the temperature of the 1D-GHER model with the shortwave radiation as a volume source.

two components are noted Q_{dir} and Q_{diff} respectively. The radiance at the top of the atmosphere Q_0 is based on the solar constant and the zenith angle of the sun z. According to our convention Q_0 is negative (figure 2.2).

The width of the air mass crossed by the light is proportional to $\sec(z)$. If the fraction of energy transmitted by the atmosphere at normal incidence is τ, then the fraction of energy reaching the ocean surface for the sun at z is $\tau^{\sec(z)}$. The half of the energy which is not directly transmitted Q_{dir} nor absorbed due to ozone $A_a Q_0$, reaches the ocean surface by diffusion (Rosati and Miyakoda, 1988).

$$Q_{\text{dir}} = Q_0 \tau^{\sec(z)} \tag{2.25}$$

$$Q_{\text{diff}} = \frac{1}{2}\left((1 - A_a)Q_0 - Q_{\text{dir}}\right) \tag{2.26}$$

$$Q_s = (1 - 0.62C + 0.0018\beta)(1 - \alpha)(Q_{\text{dir}} + Q_{\text{diff}}) \tag{2.27}$$

where the empirical relation of Reed (1977) has been used for taking into account the fractional cloud coverage C, the ocean surface albedo α and the solar noon altitude β (in degrees).

2.3.4 Net longwave radiation Q_b

The net longwave radiation is the sum of upward longwave radiation of the ocean and the downward longwave radiation of the atmosphere (Barnier *et al.*, 1995). The upward and downward radiation are proportional to the forth power of the sea surface temperature T_s and air temperature. Not only the air at the surface but also the whole atmosphere emits longwave energy. The upward radiation is in general greater than the downward radiation. The net longwave radiation, also called "back radiation" represents thus a heat loss for the ocean. The back radiation Q_b is parameterised by the scheme of Clark *et al.* (1974):

$$Q_b = \epsilon\sigma T_s^4 (1 - 0.8C^2)(0.39 - 0.05\sqrt{e_a}) + 4\varepsilon\sigma T_s^3 (T_s - T_a) \tag{2.28}$$

where ϵ is the emissivity of the ocean, σ the Stefan-Boltzman constant, e_a the atmospheric vapour pressure in hPa and T_a is the air temperature at reference level. The first term of this parameterisation takes also the greenhouse effect due to cloud and water vapour into account (Ronday, 2000)

2.3.5 Latent Q_L and sensible heat flux Q_H

The latent heat flux is due to a difference in the water vapour content of the air at the ocean surface and at the reference level. This gradient induces an evaporation or condensation. To this mass transfer corresponds a heat exchange, which is equal to the rate of vaporisation times the latent heat of evaporation L (Ronday, 2000; Zaker, 2003).

The sensible heat flux is due to the temperature difference between the air at the ocean surface and the air at the reference level. The sensible heat is the heat exchanged

by conduction and is proportional to this temperature gradient, the heat conductivity of the ocean surface and the specific heat of air at constant pressure. For the air temperature at the ocean surface, the sea surface is taken assuming a local equilibrium.

The latent heat flux and the sensible heat flux are parameterised by classical bulk turbulent transfer formulas (Rosati and Miyakoda, 1988; Castellari *et al.*, 1998):

$$Q_L = C_L L \rho_a \|\mathbf{u}_a\| (q_s - q_a) \tag{2.29}$$
$$Q_H = C_H c_{p_a} \rho_a \|\mathbf{u}_a\| (T_s - T_a) \tag{2.30}$$

where ρ_a is the air density, $\mathbf{u}_a$ is the wind vector, q_s is the specific humidity of saturated air at T_s, q_a is specific humidity of air and c_{p_a} the heat capacity of air at constant pressure. For an air pressure p_a expressed in hPa, q_a is obtained by the air temperature T_a and the relative humidity r and q_s is obtained from the sea surface temperature T_s by:

$$q_a = r e_{\text{sat}}(T_a) \frac{0.622}{p_a} \tag{2.31}$$
$$q_s = e_{\text{sat}}(T_s) \frac{0.622}{p_a} \tag{2.32}$$

Expressions (2.29) and (2.30) are well established bulk parameterisations for the latent and sensible heat flux. Matter of discussions are however the exchange coefficient C_H (Stanton number) and C_E (Dalton Number). Numerous parameterisations are proposed in the literature and their impact on the Mediterranean Sea heat budget is discussed by Castellari *et al.* (1998). The GOTM meteorological forcings routines implemented into the GHER model use the schemes of Kondo (1975) for the turbulent coefficients.

Chapter 3

Nesting

3.1 Introduction

A system of nested models is applied to the Mediterranean Sea with two successive zooms of the Liguro-Provençal basin and of the Ligurian Sea. Two strong and variable currents, the Western Corsican Current and the Eastern Corsican Current (hereafter WCC and ECC) enter the domain of the Ligurian Sea 3.1. Both advect Modified Atlantic Water (MAW) at the surface and the ECC also transports the denser Levantine Intermediate Water (LIW). The variability of these currents has been studied by *e.g.* Astraldi and Gasparini (1992) and Sammari *et al.* (1995) and shows a seasonal cycle and a dependence on local atmospheric forcings. These currents join and give birth to the Northern Current (hereafter NC) following the French and Spanish coast. NC and WCC describe a cyclonic circulation along the Liguro-Provençal front. Especially during the winter, a high mesoscale activity associated with meanders in the NC, eddy formation or displacements in the Ligurian-Provençal front can be observed (Sammari *et al.*, 1995).

Surface cooling and mixing can transform MAW into Winter Intermediate Water (WIW) that reaches the buoyancy equilibrium between the MAW and the LIW. This process can be compared to the dense water formation in the Gulf of Lions, but WIW formation occurs for less severe weather conditions and only involves the MAW layer (Gasparini *et al.*, 1999).

The spatial resolution of oceanographic models plays an important role in representing and simulating ocean processes. But most of the time this parameter is constrained rather by computing capacity than by physical arguments. So it is desirable to concentrate the resolution efforts in a particular domain of interest or in a crucial region where important processes for a larger domain occur. Modelling an oceanographic system at variable resolution can thus be achieved by nesting a fine grid model in a coarse grid model. Traditionally (*e.g.* Fox and Maskell, 1995) we distinguish two types of nesting: (i) one-way or passive nesting and (ii) two-way or interactive nesting. In the former case, the fine grid model's boundary conditions are interpolated from the coarse grid model, but the dynamics of the fine grid model do not influence the evolution of the coarse grid model. The two-way nesting approach allows the fine grid model to give a feedback to

the coarse grid model after receiving the boundary conditions from the coarse grid model. For multiple or telescopically nested models, the grid refining procedure is repeated several times.

In oceanography, the one-way nesting technique was first applied by Spall and Robinson (1989) and is now a widespread and operational method for studying ocean system in a subdomain (*e.g.* Robinson and Sellschopp, 2002; Zavatarelli and Pinardi, 2003). For one-way nesting systems, the coarse grid model and the fine grid model can be implemented independently. This approach is therefore well suited for operational forecasts. Inconsistencies between the fine and coarse model evolution, especially after long term integrations, however make the application of the boundary conditions a delicate task. Two-way nested models avoid the divergence between fine and coarse grid model and can reduce the problem of inconsistent boundary conditions.

The interactive nesting technique was tested in idealised experiments (*e.g.* Spall and Holland, 1991; Fox and Maskell, 1995; Ginis *et al.*, 1998; Blayo and Debreu, 1999) for different resolution ratios between the fine grid and the coarse grid model. It was shown that propagating features such as a barotropic modon, baroclinic vortex, a meandering front or a Kelvin wave along the equator could enter and exit the fine model. Ginis *et al.* (1998) pointed out that the solution in the high-resolution model is similar to the solution obtained by a high-resolution model covering the whole domain. Fine grid models following a propagating feature were also implemented (Rowley and Ginis, 1999; Blayo and Debreu, 1999). Realistic ocean systems were studied e.g. the Norwegian Coastal Current (Oey and Chen, 1992), the Iceland-Faeroes front (Fox and Maskell, 1996), the tropical Pacific Ocean (Ginis *et al.*, 1998) and the Tunisia-Sardinia-Sicily region (Onken *et al.*, 2003)

The Ligurian Sea was chosen for the present study since the thermohaline structure of the NC is determined in this area. The NC is one of the major currents in the western Mediterranean Sea and surrounds the area of the Gulf of Lions where Western Mediterranean Deep Water formation occurs in winter. Thus, the NC plays an important role for the preconditioning of the surface water in the Gulf of Lions. The NC may also affect the WIW formation in the Ligurian Sea. Since eddies of WIW sink to a depth of about 200-400 m, they may in return also interact with the NC.

Due to the permanent frontal system, the primary production in the Ligurian Sea is exceptionally high for offshore water in the Mediterranean Sea. Cetacean species are also more abundant in the Ligurian Sea than in other seas of the Western Mediterranean Basin. This has lead to the instauration of Ligurian Sea Cetacean Sanctuary, which aims to protect this valuable marine environment. Primary producers are particularly sensitive to the hydrodynamical processes that also indirectly have an impact on higher species.

Besides the physical and biological particularities, the Ligurian Sea is an interesting test area since, to our knowledge, no high-resolution study at 1' has been realised in the Ligurian Sea. This resolution is achieved by two successive grid refinements of a coarse

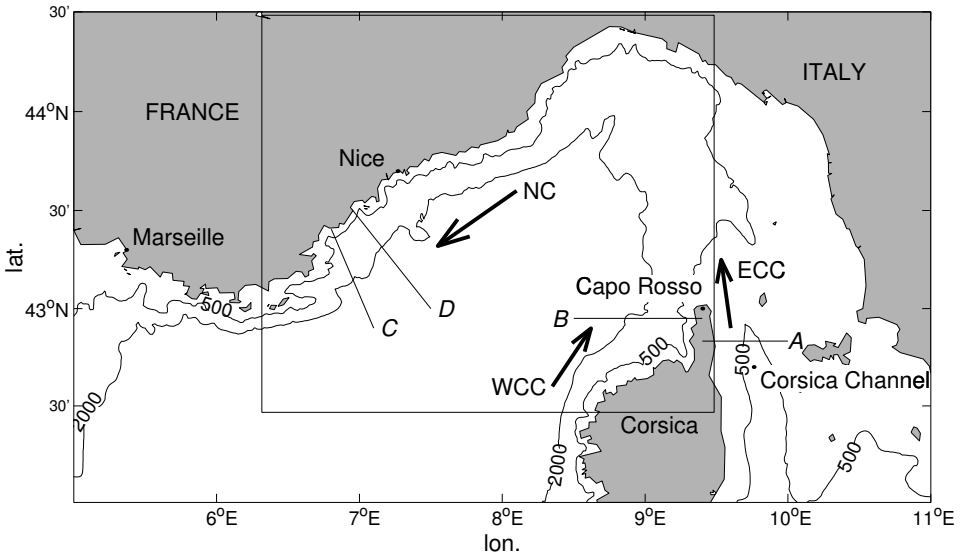

Figure 3.1: The Ligurian Sea with the three major currents: ECC, WCC and NC. The results of the model will be illustrated in sections A, B, C and D. The solid line represents the interface between the fine and intermediate grid resolution models. The 500 m and 2000 m isobaths are also shown. While the bathymetry in the Ligurian Sea is rather steep at the French coast, there is a continental shelf in the western part near the Italian coast. The Corsica Channel separates this plateau from the Corsica Island.

resolution model of the Mediterranean Sea. The present investigation aims to demonstrate that telescopically and interactively nested models applied to a realistic basin are a robust and powerful tool for simulating small-scale ocean processes induced by larger scale currents and fronts. Because of the strong winds the mesoscale activity in the Ligurian Sea is generally highest during the winter (Albérola *et al.*, 1995). In order to investigate the mesoscale flow features, a wintertime situation was thus chosen for the model run.

3.2 Nesting procedure

Although the grid was refined twice, the nesting procedure is explained for clarity in the case of a single nesting. For multiple nesting, subsequent description of the grid configuration, of the boundary condition and the feedback holds for every pair of coarse and embedded finer grid models.

The grid refinement ratio r is supposed to be odd and equal in the both horizontal directions. In this way, each grid value of the overlapping region of the coarse grid coincides with a value of the fine grid (figure 3.2). In the vertical, the grid is not refined. To each double-sigma layer of the coarse grid corresponds one layer in the fine grid at the same depth when the bathymetry is consistent at the boundary. Furthermore, the interpolation of the boundary conditions and the feedback procedure are simplified by using a land mask of the fine grid that is equal to the land mask of the overlapping

coarse grid for the first $r+1$ fine grid boxes counted normal to the boundary. The remaining land mask can be chosen to follow as closely as possible the real coastline and the bathymetry.

The nesting procedure between the coarse and fine grid models can be summarised as follows. For each variable, we:

1. interpolate the fine grid boundary conditions for that variable from the coarse grid model and impose the boundary conditions to the fine grid model in the so-called "dynamic interface",
2. integrate the coarse and fine grid models one time step ahead,
3. average the values of the fine grid model lying on and inside the "feedback interface" and replace the corresponding values of the coarse grid model. The feedback is thus performed over the entire volume of the nest delimited by the feedback interface.

Figure 3.2 shows the position of the dynamic and feedback interface for the different variables, as will be explained in the subsequent sections. For multiple nesting, the feedback starts with the finest embedded model and its average field replaces the values of the next coarsest model.

3.2.1 Boundary conditions

The integration of the fine grid model requires boundary conditions for horizontal velocity, temperature, salinity and turbulent kinetic energy. The interpolation of these boundary conditions is performed for each vertical level independently.

Normal velocity

The velocity component normal to the boundary has the greatest impact on the flow inside the fine grid model. Fine grids and coarse grids are staggered in a way that the interpolation of the normal velocity requires only an interpolation tangent to the boundary. The boundary conditions of the normal velocity are extracted and imposed in column C of figure 3.2.
We note $\mathbf{x}^c_{\mathbf{v}_n}$ the vector formed by the N normal velocity components of the coarse grid model of a given boundary segment and of a given layer and $\mathbf{x}^f_{\mathbf{v}_n}$ the corresponding rN unknown velocity components of the fine grid. The volume conservation constraint can be formulated by a linear relationship :

$$\mathbf{x}^c_{\mathbf{v}_n} = \mathbf{R}\,\mathbf{x}^f_{\mathbf{v}_n} \tag{3.1}$$

where $\mathbf{R}$ is a $N \times rN$ matrix given by:

$$\mathbf{R}_{ij} = \frac{\Delta z^f_j}{r\Delta z^c_i} \qquad \text{if } r_i \in \{j, j+1 \ldots j+r-1\} \tag{3.2}$$

and if the ith coarse grid box is a sea box

$$= 0 \qquad \text{elsewhere} \tag{3.3}$$

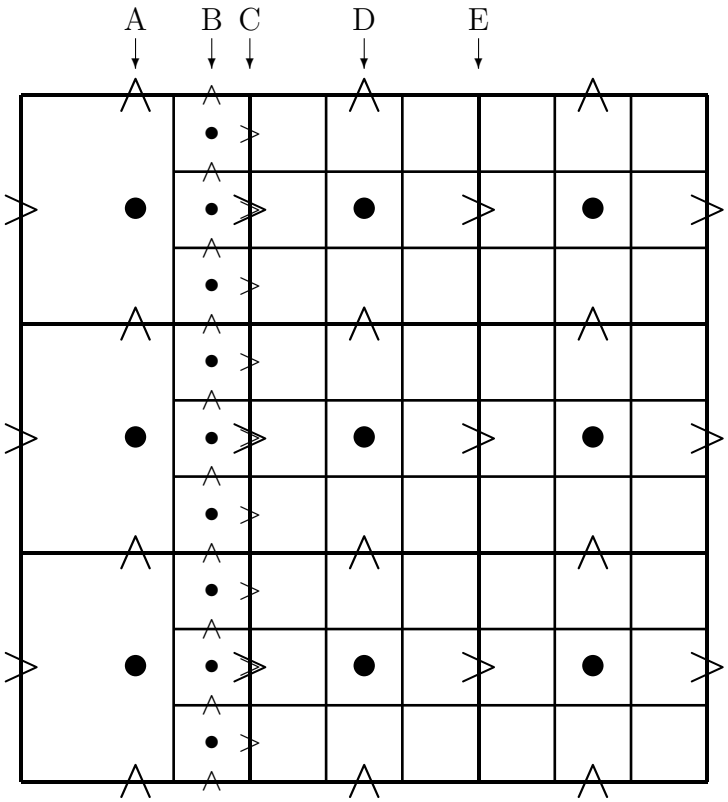

Figure 3.2: The relative position of the coarse (heavy lines) and fine grid (fine lines). The dots (•) show the position of scalar variables, > the zonal velocity and ∧ the meridional velocity component. The large symbols are associated to the coarse grid and the small symbols to the fine grid. For clarity, only the position of the variables imposed by boundary conditions are showed for the fine grid. The boundary conditions of the scalars and the tangent (to the nesting boundary) velocities interpolated from columns A and D are imposed in column B. The normal velocity component is imposed in column C. The average values of the scalars and the tangent velocities are injected in the coarse grid model, starting with column D. For the normal velocity, the feedback begins with column E.

Δz_j^f and Δz_i^c are the vertical grid spacing of the fine and coarse mesh respectively. The boundary conditions of the normal velocity for the fine grid should be smooth, but in any case the volume conservation must be satisfied. $\mathbf{x}^f_{\mathbf{v}_n}$ is obtained by minimising the cost function $J(\mathbf{x}^f_{\mathbf{v}_n})$, penalising abrupt variations under the constrain of conservation (3.1):

$$J(\mathbf{x}^f_{\mathbf{v}_n}) = (\mathbf{L}\mathbf{x}^f_{\mathbf{v}_n})^T(\mathbf{L}\mathbf{x}^f_{\mathbf{v}_n}) \tag{3.4}$$

where $\mathbf{L} \in I\!R^{(N-2)\times N}$ is the discrete second derivative ($\delta_{i,j} = 1$ if $i = j$ and 0 otherwise).

$$\mathbf{L}_{ij} = \delta_{i,j-1} - 2\delta_{i,j} + \delta_{i,j+1} \tag{3.5}$$

An application of a minimisation using Lagrangian multipliers would require the inverse of the singular matrix $\mathbf{L}^T\mathbf{L}$. The singularity is due to the fact that the cost function does not

depend on a constant shift and a constant slope of $\mathbf{x}^f_{\mathbf{v}_n}$. It is therefore preferable to identify the vector space constrained by the volume conservation and the unconstrained null space of $\mathbf{R}$. The basis vectors of the former space are columns of the matrix $\mathbf{V}_c \in I\!R^{(rN\times N)}$ and the basis vector of the latter are the columns of $\mathbf{V}_0 \in I\!R^{(rN\times(r-1)N)}$. One can then decompose $\mathbf{x}^c_{\mathbf{v}_n}$ into:

$$\mathbf{x}^f_{\mathbf{v}_n} = \mathbf{V}_c\mathbf{a}_c + \mathbf{V}_0\mathbf{a}_0 \tag{3.6}$$

The components in the constrained space are directly given by (3.1):

$$\mathbf{a}_c = (\mathbf{R}\mathbf{V}_c)^{-1}\mathbf{x}^c_{\mathbf{v}_n} \tag{3.7}$$

By minimising the square of the second derivative of the boundary condition, one obtains then the remaining components:

$$\mathbf{a}_0 = -\left(\mathbf{V}_0^T\mathbf{L}^T\mathbf{L}\mathbf{V}_0\right)^{-1}\mathbf{V}_0^T\mathbf{L}^T\mathbf{L}\mathbf{V}_c(\mathbf{R}\mathbf{V}_c)^{-1} \tag{3.8}$$

Finally, the boundary conditions applied to the fine grid mesh yield:

$$\mathbf{x}^f_{\mathbf{v}_n} = \left(\mathbf{I} - \mathbf{V}_0\left(\mathbf{V}_0^T\mathbf{L}^T\mathbf{L}\mathbf{V}_0\right)^{-1}\mathbf{V}_0^T\mathbf{L}^T\mathbf{L}\right)\mathbf{V}_c(\mathbf{R}\mathbf{V}_c)^{-1}\mathbf{x}^c_{\mathbf{v}_n} \tag{3.9}$$

For an efficient implementation of the interpolation procedure, the interpolation coefficients are computed once and stored.

This interpolation method can be generalised in order to relax the constraint that the land mask of the fine grid and coarse grid must be equal at the boundary. By a suitable definition of the reduction operator $\mathbf{R}$, any fine grid land mask can be treated.

Tangent velocity

The velocity tangent to the boundary only plays a role in the horizontal mixing and advection of momentum. These terms are generally small compared to the other forces. The tangent velocity component at the fine grid boundary (column B) is therefore obtained by a simple bilinear interpolation using the four nearest tangent velocity components of the coarse grid (columns A and D). No volume conservation aspects must be taken into account for this velocity component.

Scalars

In a first step, temperature, salinity and turbulent kinetic energy are linearly interpolated normally to the boundary (from columns A and D to column B). Then, these values are interpolated tangently to the boundary using a similar interpolation formula as the normal velocity. For the scalars we impose a constraint independent of the sigma-layer thickness.

$$\mathbf{R}_{ij} = \frac{1}{r} \quad \text{if } r_i \in \{j, j+1 \dots j+r-1\} \text{ and if the } i\text{th coarse grid box is a sea box} \tag{3.10}$$

$$= 0 \quad \text{elsewhere} \tag{3.11}$$

This procedure is obviously not conservative. The design of a stable and flux conservative interpolation is not trivial, since a smooth flux boundary condition does not guarantee a resulting smooth scalar field.

3.2.2 Feedback

After each time step the mean values of the fine grid mesh replace the values of the coarse grid model of the overlapping region. The velocity components are averaged over the interfaces of the corresponding coarse grid box. For the scalar variables (elevation, temperature, salinity and turbulent kinetic energy), the mean values over the whole coarse grid box are taken.
For stability reasons, some authors (*e.g.* Ginis *et al.*, 1998; Fox and Maskell, 1995) separate the feedback interface from the dynamic interface. Here no such problems occur and the feedback interface was chosen as close as allowed by the grid configuration to the dynamic interface.
In the present case, it would be sufficient to apply the feedback only to the first coarse grid boxes inside the overlapping region, since the feedback is performed every time step for both barotropic and baroclinic modes, and for a one-time step forecast every grid box depends, at the first order, only on its direct neighbourhood.

3.2.3 Sponge layer

It is natural to attribute to the coarse resolution model a higher diffusion than the diffusion of the fine grid model. At the boundary and over a distance of two coarse grid boxes, the diffusion in the fine grid model is raised linearly to reach the diffusion of the coarse resolution model. The utility of a sponge layer for embedded models is justified by the following reasons:

- Small scale features (compared to the grid spacing of the coarse model) moving outward the fine grid model cannot properly be resolved by the coarse grid model. These features are thus damped by the sponge layer.

- The behaviour of the coarse grid model is necessarily different than that of the fine grid model. For example, the advection speed of a perturbation depends on the grid spacing (Miyakoda, 2002). Thus, boundary conditions for the fine grid model are never perfect and totally consistent with the fine model dynamics. The sponge layer regularises the reaction of the fine grid model to imperfect boundary conditions.

- The hydrostatic inviscid primitive equations with open boundaries are mathematically ill-posed (Oliger and Sundstrom, 1978; Browning *et al.*, 1990). For one-way

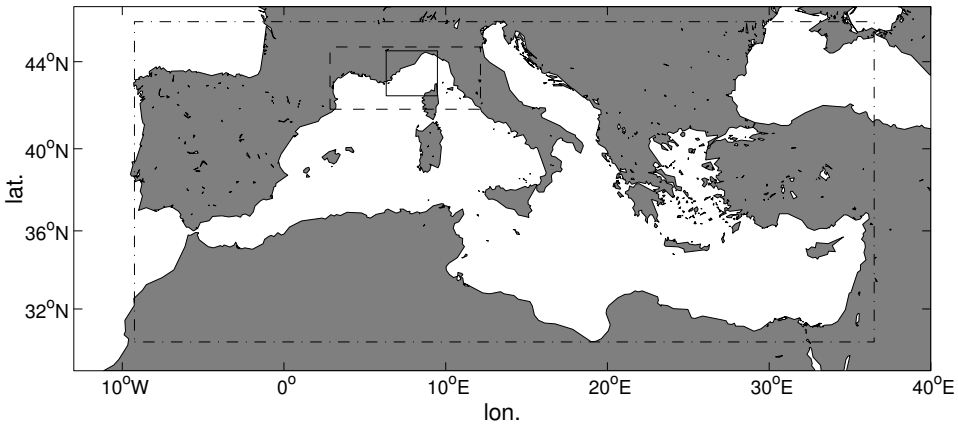

Figure 3.3: Domain of the three models. The Mediterranean Sea model is delimited by the dashed line. The dotted line represents the boundary of the Liguro-Provençal model. The Ligurian Sea model lies inside the solid line rectangle.

nesting, diffusion (or some other special treatment) is thus necessary at the boundaries. But *a priori* it is not clear if this problem occurs to the same extent in a two-way nested model.

More sophisticated solutions have been proposed to regularise the behaviour of the fine grid model at the boundary, *e.g.* the radiative-nesting Boundary Conditions of Chen (1991), a flow relaxation scheme (Oey and Chen, 1992), a Newtonian-type damping scheme (Kurihara and Bender, 1980). But most of these methods alter the volume transport inside the fine grid, and are therefore not considered here.

3.3 Implementation

Numerous idealised nesting experiments (*e.g.* Spall and Holland, 1991) show that for a fixed resolution of the fine grid model, acceptable results are obtained for a 3 and 5 times coarser large scale model. With higher nesting ratios, substantial degradation is observed. To achieve a high resolution at regional scales (1') in the Mediterranean Sea, a multiple nesting strategy was thus adopted: a fine resolution (1') regional model of the Ligurian Sea (6°19'E to 9°29'E, 42°28'N to 44°29'N) embedded in an intermediate resolution (3') model of the Liguro-Provençal sub-basin (2°51'E to 12°9', 41°51'N to 44°39'N) nested in a coarse resolution (15') general circulation model implemented for the whole Mediterranean Sea (figure 3.3). The main parameters of the three models are given in table (3.1).

During winter, the first internal Rossby Radius of deformation is of the order of $4-7$ km in the Ligurian Sea (computed from the Mediterranean Oceanic Data Base (MODB) Climatology (Brasseur *et al.*, 1996; Rixen *et al.*, 2001b)) and the associated wavelength is of order 25 - 44 km. Eddies are thus well resolved in the intermediate resolution and

	Mediterranean Sea	Liguro-Provençal Basin	Ligurian Sea
resolution	15'	3'	1'
Δx (m)	21930	4386	1462
Δy (m)	27829	5566	1855
Eddy viscosity (m^2/s)	200	8	0.9
Eddy diffusivity (m^2/s)	60	2.4	0.27

Table 3.1: The main parameters of the Mediterranean Sea, the Liguro-Provençal and the Ligurian Sea model. Δx and Δy are the grid spacing in zonal and meridional direction. The β-planes of the three grids are tangent to the Earth at a latitude of 38°.

fine resolution models.

The model bathymetry is computed from the Smith and Sandwell (1997) bathymetry. The depth of a coarser mesh box is set equal to the mean depth of the corresponding finer grid mesh. Due to the steep bathymetry near the western Corsican coast, the WCC appears to be very sensitive to the difference of the bathymetry of the nesting boundary. Unphysical results may be obtained when a sigma level of the coarse grid is connected to a sigma level of different depth. Near the coast this difference can reach several hundred meters. For simplicity we impose that the bathymetry in the $r + 1$ first grid points is equal to the bathymetry of the coarse grid model. In principle, this problem could also be solved with a vertical interpolation and sometimes an extrapolation of the boundary condition and the feedback.

The nesting system is forced by National Centers for Environmental Prediction (NCEP, http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html) reanalysis atmospheric fluxes. The coarse grid model encompasses some grid points of the Atlantic Ocean. Temperature and salinity at the Atlantic boundary are relaxed to their monthly climatology mean from the MODB and a mean flow is imposed. The freshwater discharge of the Rhône (climatology mean computed from Tusseau and Mouchel (1994)) and the Arno (Rinaldi, personal communication) are also included in the intermediate grid size model.

The coarse grid model's initial conditions are also obtained from the MODB Climatology after a 10-year spin-up of the coarse grid model. The initial conditions of the fine grid and intermediate grid are then interpolated. The whole nesting system is spun up during three months using perpetual January 1998 forcings.

The implementation of the nesting system was tested in an idealised, conservative configuration. Without atmospheric and riverine forcings and with the strait of Gibraltar "closed", the nesting system conserves the total volume. As pointed out earlier, temperature and salinity are not conserved by the boundary conditions used is the present implementation. But the losses and gains of these scalars appear to be very small compared to their spatial standard deviation.

3.4 Comparison between one-way and two-way nesting

The LIW is formed in the Eastern Mediterranean and has travelled a long path before reaching the Ligurian Sea. In the present implementation, the LIW vein crosses the two nested models and is thus particularly sensitive to the nesting strategy used. In order to elucidate the gain of a high-resolution two-way nested model, a one-way nesting experiment was carried out. The parameters, initial conditions and the interpolation of the boundary conditions of the three models are identical to the two-way nesting experiment, but now each model does not depend on the finer resolution model.

Figure 3.4 shows temperature sections of the NC. The mean and the standard deviation of the temperature during the three winter months for the Ligurian Sea model in the two-way and one-way configuration and the Liguro-Provençal basin model in the one-way case are depicted. The extension and the temperature of the LIW vein is in all three cases smaller and colder than measured by Albérola *et al.* (1995). The clearest signal of the LIW is found for the two-way nested high-resolution model and the simulated properties of this vein are close to the observed values. In the one-way high resolution model of the Ligurian Sea, the LIW core is still present, but the mean temperature is too cold and the variability of this current is also lower than in the two-way nested model. With a three times coarser grid, the model was unable to simulate the LIW vein in the Ligurian Sea as can be seen on figure 3.4, and the LIW flows through the Liguro-Provençal model in a rather unstructured and unsteady manner. The fine grid was thus necessary to maintain the temperature gradient high between the LIW and the offshore water masses and to confine the NC to the coast.

Furthermore, when a two-way nesting strategy is used instead of a one-way nesting, the results in the coarse grid are also improved. In the Gulf of Lions (covered by the $1/20^\circ$ grid model) the structure of the NC is also better represented in the two-way nested system than in the one-way nested models. The feedback from the finer resolution model thus also improves the simulation at larger scale, not only in the region of feedback but also in remote places.

3.5 Results

We present here the results of a three-month integration, starting 1 January 1998. The simulated transports and water properties of the three major currents (ECC, WCC and NC) are averaged over the integration period and are compared to results of surveys carried out in this region. The salinity and temperature gradient of the Liguro-Provençal front and the scale of mesoscale structures in the Ligurian Sea and the Gulf of Lions are also validated in the subsequent sections. The overall characteristics of the Ligurian Sea are well reproduced in the model. The signals of the ECC, WCC and the NC can clearly be seen in the surface velocity (figure 3.5). The transitions from 1' to the 3' model and from the 3' to the 15' model (not shown) are smooth for all variables.

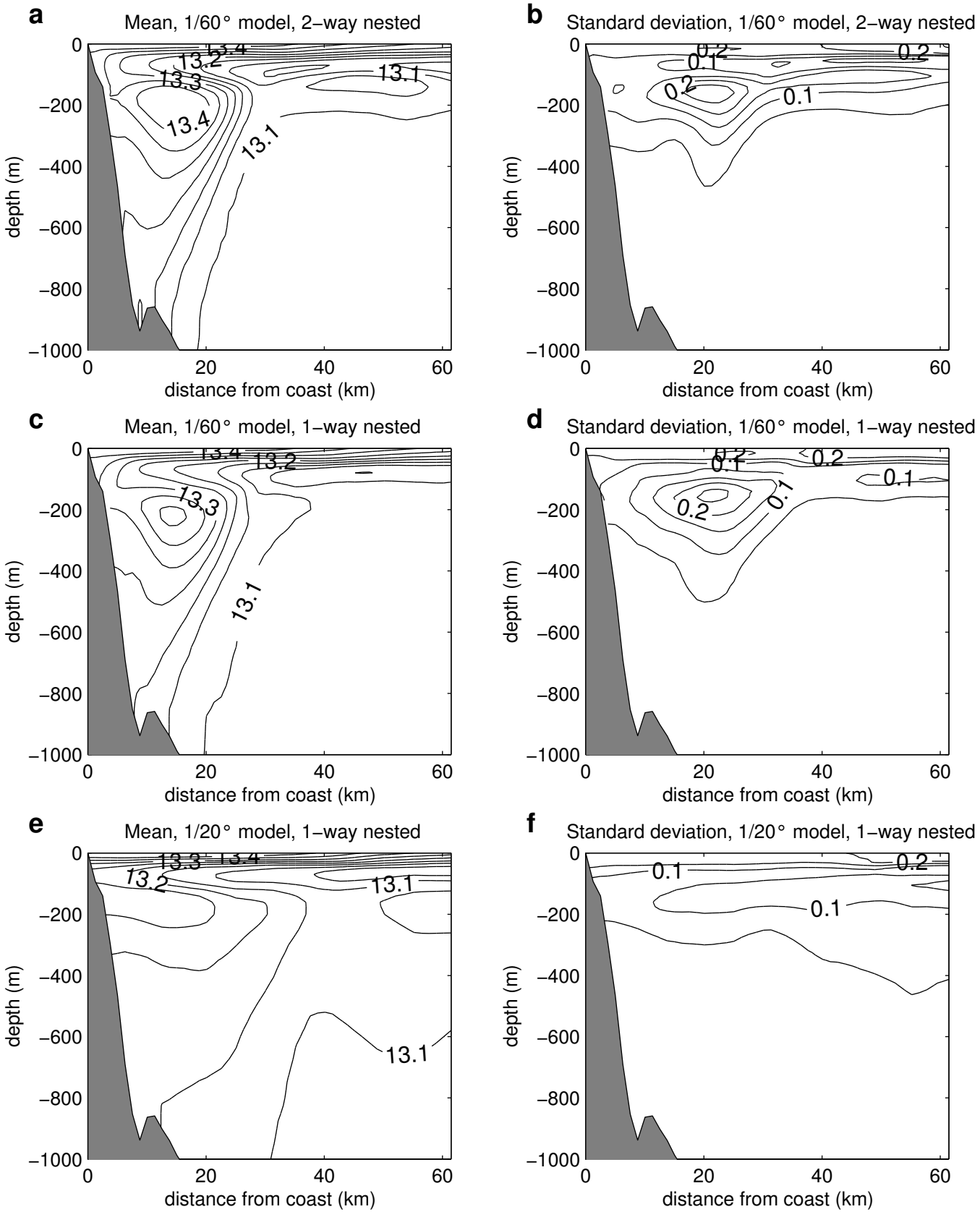

Figure 3.4: Mean temperature and standard deviation (in °C) at section C of figure 3.1 for the two-way (top) and the one-way nested model at 1/60° resolution (centre). The two bottom figures are the results of the 1/20° model for the one-way nesting strategy.

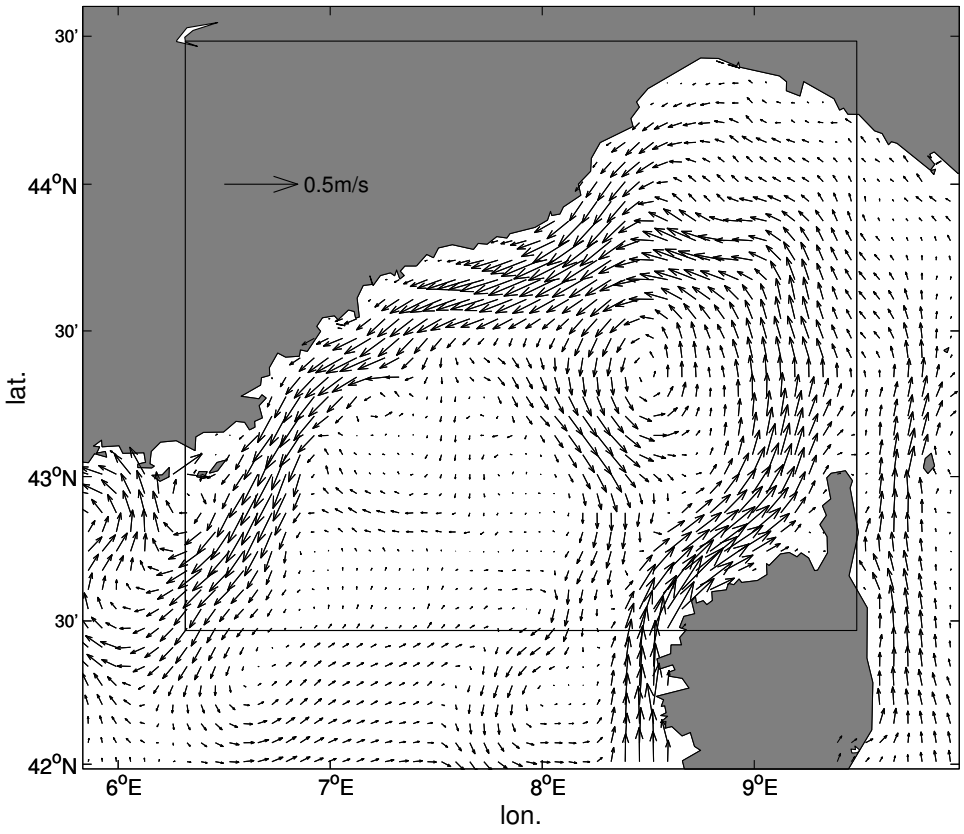

Figure 3.5: Mean velocity in winter at the surface in the Ligurian Sea. The solid line represents the interface between the fine and intermediate grid resolution models.

3.5.1 The Eastern Corsican Current

The ECC flows to the north through the Corsican Strait up to a depth of 450 m and with a width of 35 km. Its intensity increases in spring and weakens in fall (Astraldi *et al.*, 1999). The mean vertical structure of the simulated ECC at the strait of Corsica is shown in figure 3.6. The depth and the width are very close to the estimated values. A section in the salinity field shows that the MAW and the LIW are well present in the model. The order of magnitude of the velocity in the strait (0.2 m/s) reproduced by the model is in good agreement with the direct measurements (0.1 - 0.3 m/s). The simulated flows of MAW (from the surface to 200 m) and of LIW through the strait are 0.59 Sv (1 Sv = $10^6 m^3/s$) and 0.38 Sv respectively. These transports are comparable to annual mean transport based on measurements: 0.56 Sv for the surface flow and 0.14 Sv for the intermediate flow (Astraldi and Gasparini, 1992; Astraldi *et al.*, 1999).

3.5.2 The Western Corsican Current

The WCC follows the western Corsican coast in its northward flow. It extends 30 km offshore and its depth reaches 600 m. The WCC transports about 1.15 Sv MAW into the Ligurian Sea. The transport is at a maximum during spring. Typical velocities of the

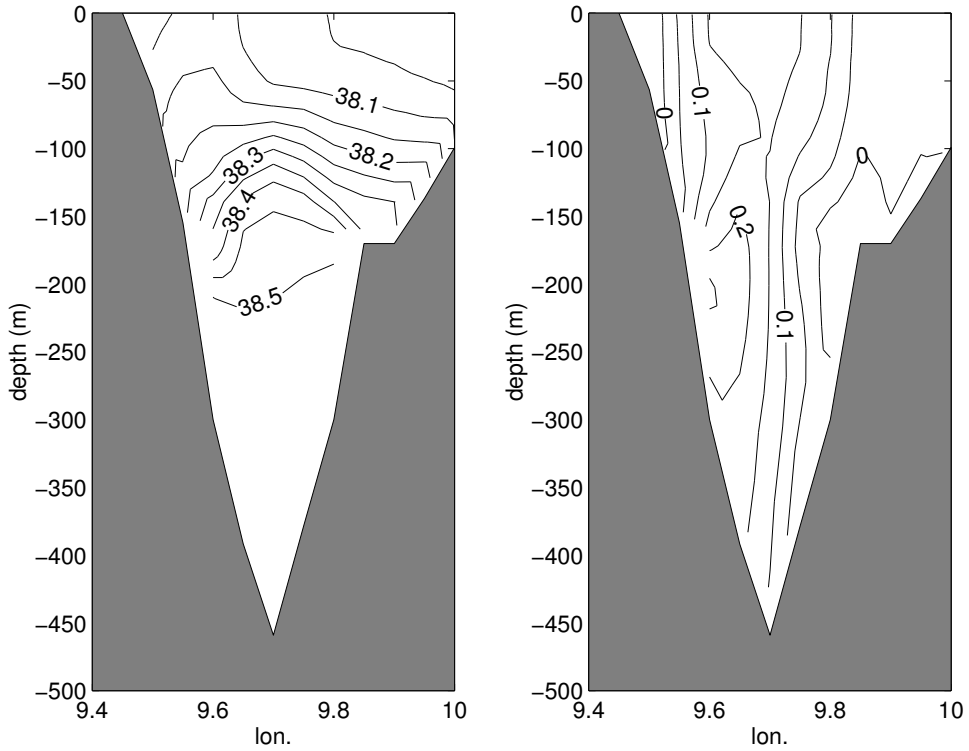

Figure 3.6: Cross-section in the strait of Corsica (section A of figure 3.1) of the mean salinity and the mean zonal velocity (in m/s) of the model. The two water masses MAW and LIW transported by the ECC can be seen in the salinity section.

WCC range around 0.09 m/s (+/- 0.06 m/s standard deviation) during winter (Astraldi and Gasparini, 1992; Sammari *et al.*, 1995).

This current flows through all three models of the nesting system. Figure 3.7 shows the N-S component of the WCC west of the Capo Rosso and the salinity at the same location. The depth (800 m) and the width (32 km) of the WCC are comparable to the depth obtained by measurements. An unphysical countercurrent next to the WCC can be found in the intermediate resolution model. This countercurrent is characterised by a maximum velocity of 0.2 m/s. Mean transport of the WCC during the model integration near the Capo Rosso was 1.22 Sv and is in good agreement with values of Astraldi and Gasparini (1992). However, a high variability of the WCC in position and strength was observed in the model. The steep bathymetry and the nesting boundaries transecting the topographic slope make the modelling of the WCC a delicate task.

3.5.3 The Northern Current

The NC is one of the major veins of the Western Mediterranean Sea. It takes birth in the Ligurian Sea and follows the continental slope until the Catalan Sea. Its width ranges from 20 km in winter to 30 km in summer. From late January to mid-March, the 0.1 m/s isotach is steep and deeper than 200 m. The transport of the upper 200 m ranges

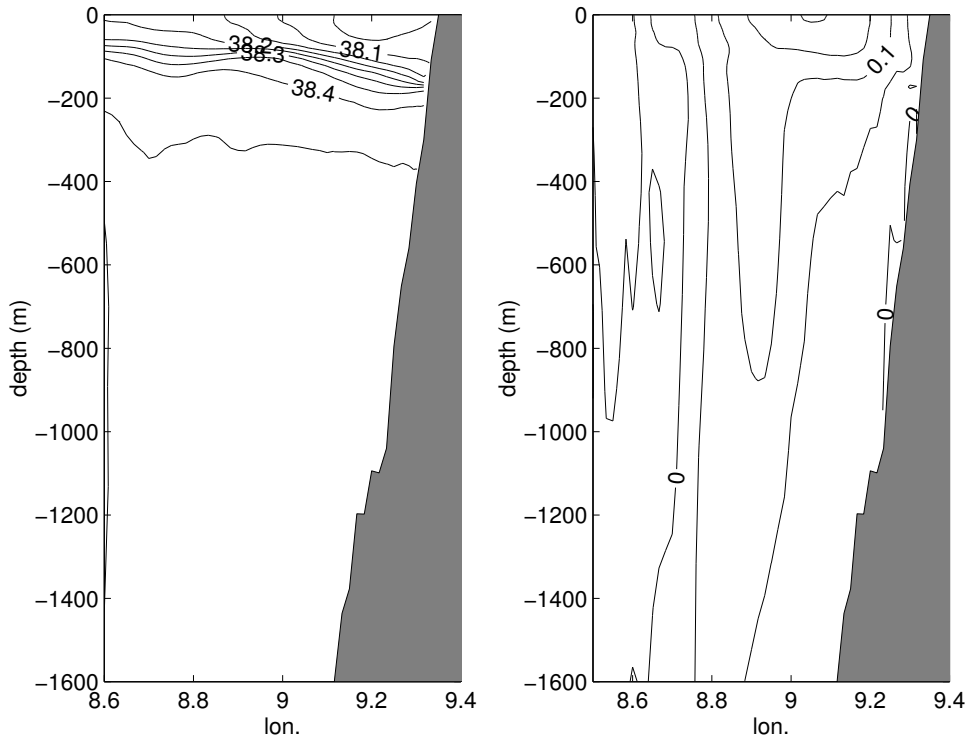

Figure 3.7: Cross-section of the WCC (section B of figure 3.1) of the mean salinity and the mean zonal velocity (in m/s) of the model.

from 0.9 to 1.5 Sv. Below this level (between 200 and 700 m), only a weak transport of 0.2 to 0.4 Sv is measured. Generally, MAW can be found at the surface in the NC lying above a core of LIW located relatively close to the coast at 20-25 km and at depths of 400-500 m. Cooling and homogenisation of the MAW processes occur in winter and a relatively thin layer of WIW (T = 13.0 - 12.9 °C, S = 38.12-38.20) between 100 and 200 m is formed in late February (Sammari *et al.*, 1995; Albérola *et al.*, 1995).

In figure 3.4a and 3.5.3 cross-sections of the temperature and the normal velocity are shown. The velocity of NC (0.23 m/s) is lower than the estimation of the maximum velocity based on geostrophy (0.6 m/s, Sammari *et al.* (1995)) and direct measurements (0.5 m/s, Albérola *et al.* (1995)). The transports simulated by the model give satisfactory results at the surface (0.95 Sv for the first 200 m), while the transport of 0.12 Sv between 200 m and 700 m is slightly lower than measurements by Sammari *et al.* (1995). The signature of the different water masses is also well present in the model. The temperature section (figure 3.5.3) very clearly displays the core of LIW advected by the ECC that joins the NC. At a depth of 100 m, a temperature minima characterising the presence of WIW is successfully reproduced by the model. Above the WIW, a layer of less modified and advected MAW can be observed in the model.

During wintertime the NC exhibits a high mesoscale activity. This observed variability was associated to meanders of the NC due to baroclinic instabilities. The meanders have

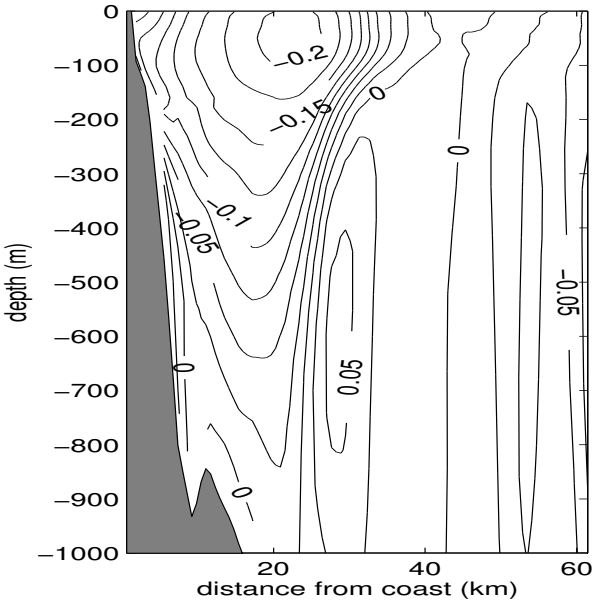

Figure 3.8: Mean velocity (in m/s and normal to the cross-section) of NC (section C of figure 3.1) in function of depth and distance from the coast.

a typical wavelength of 30-60 km and propagate with a phase velocity of 10 km/day downstream. The fluctuations perpendicular to the coast are of a few tens of km and can thus affect the whole width of the current (Albérola *et al.*, 1995; Sammari *et al.*, 1995). The offshore displacement of the NC can be easily seen on figure 3.4b showing the standard deviation of the temperature. Since the temperature of LIW is very different from the surrounding water masses, the highest variability is located next to the core of LIW.

Figure 3.9 shows a horizontal section of temperature at 200 m. Around the longitude 7°30' and latitude 43°20' we see the development of a meandering instability. Its wavelength is about 55 km. The phase velocity is more difficult to determine because different perturbations of the front with different phase velocity are overlaid. The phase velocities reproduced in the model are about 20 km/day.

The unstable path of the NC also favours the exchange between the LIW and the much colder and saltier offshore water masses. Filaments and eddies of LIW leave the main flow path of the NC and maintain their characteristic thermohaline properties for sometimes up to 20 days before they diffuse or return back to the main flow of the NC. The standard deviation of the temperature at a section of 200 m depth shows that the outbreak of LIW filaments and eddies often occurs in a region around 7°30' longitude and 43°20' latitude (Figure 3.10). This might be due to the topographic feature at this location, which perturbs the NC following the continental slope. The topographic feature is characterised by the 2000 m isobath on figure 3.1. This plateau located in the continental slope extends about 15 km offshore.

Figure 3.11 shows the transport of LIW crossing the section in the NC during the model integration. The high variability is due to interaction of the LIW vein with the offshore water masses. The sharp minimum at day 24 is due to a breaking of the LIW vein by

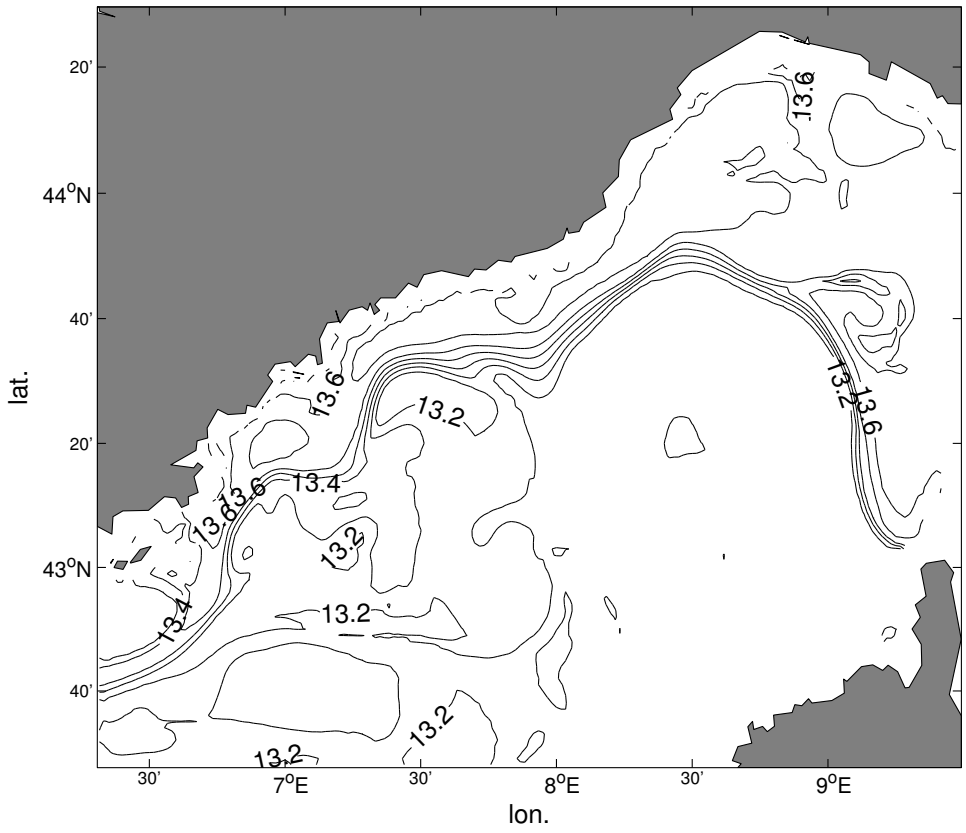

Figure 3.9: Temperature (in °C) at 200 m for 20 April 1998 and meandering structures of the LIW vein.

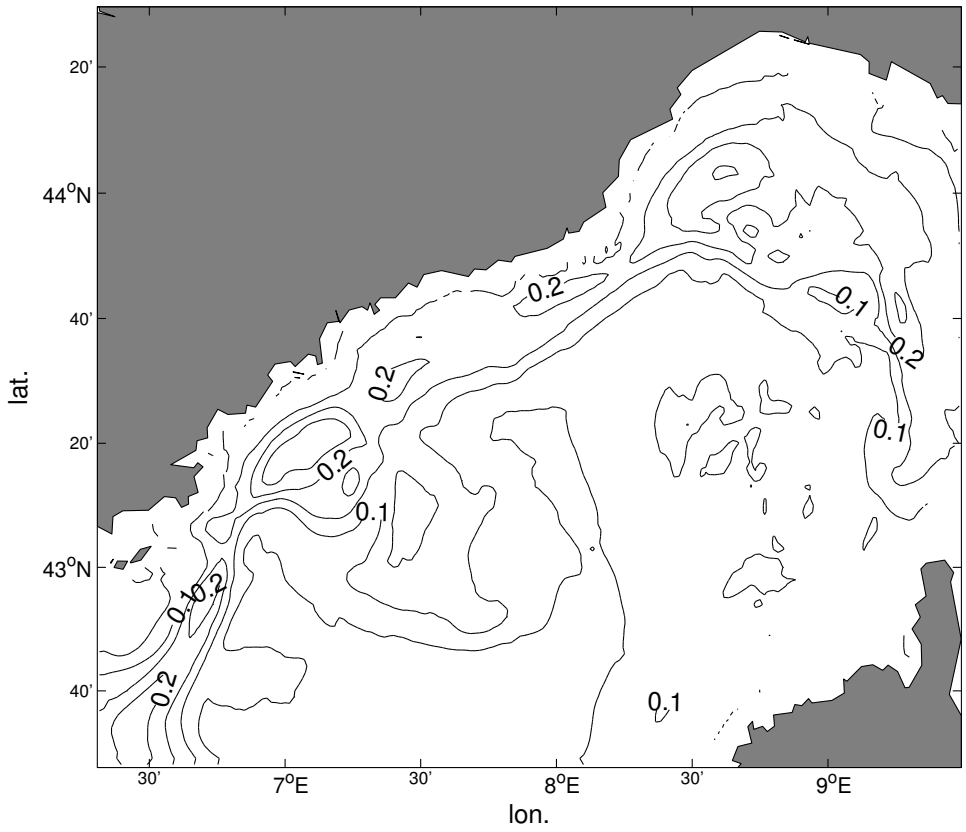

Figure 3.10: Standard deviation of temperature (in °C) at 200 m. The high variability of the NC at 43°10'N and 7°30'E might be due to topographic interactions.

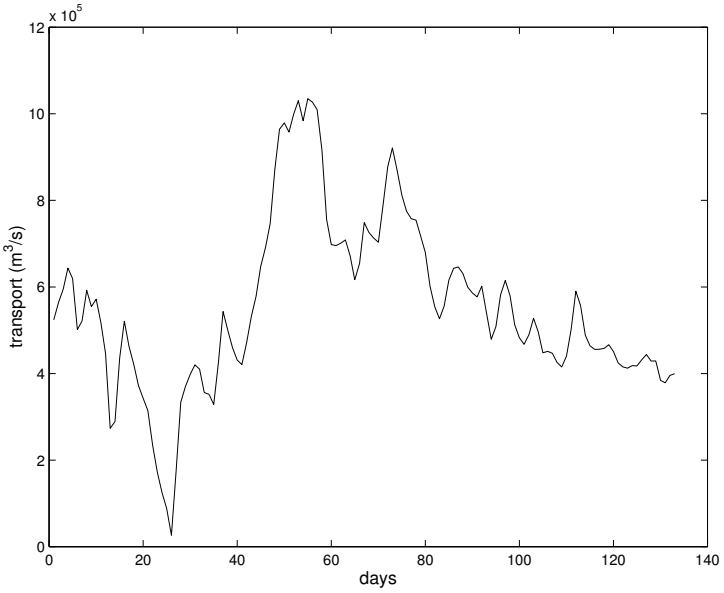

Figure 3.11: The transport of LIW (in m^3/s) trough section D of 3.1 as a function of time (days). Water masses warmer than 13.4 °C and below 100 m were identified as LIW.

offshore water. After this blockade is released, the LIW transport reaches its highest value observed during the model integration around day 50.

In the section where the transport of LIW was computed, another interesting process occurs. Surface cooling has produced a water mass called winter intermediate water (WIW), with a temperature of 12.9°C -13°C and a salinity of 38.3-38.4 (figure 3.12). These rotating lenses of WIW (also called weddies) have been found in the Catalan Sea and in the Ligurian Sea (Gasparini *et al.*, 1999). The characteristics of the simulated WIW lens agree with the observed range of temperature 12.5-13°C and salinity 38.1 - 38.3 (Salat and Font, 1987). Particular to the present situation is its strong interaction with the vein of LIW broken into two branches. This caused a decrease of the LIW transport, as can be seen on figure 3.11 at day 64. During the three winter months the formation of eight WIW lenses has been observed, most of them in January.

3.5.4 The Liguro-Provençal front

The Liguro-Provençal front is formed by a high gradient in temperature and salinity separating coastal waters from those located offshore. Coastal waters are warmer (more than 14°C) and less saline (about 38) than the interior waters, with temperatures below 13.8°C and salinities of 38.1-38.2 (data collected in early spring at the Gulf of Genoa (Gasparini *et al.*, 1999)). Other studies (e.g. Albérola *et al.*, 1995) indicate that the vertical mixing occurring in winter leads to the homogenisation of the surface layer temperature in both the coastal and the interior waters. The winter surface temperature is about 13°C, and these data are collected in a section of 55 km from the coast to

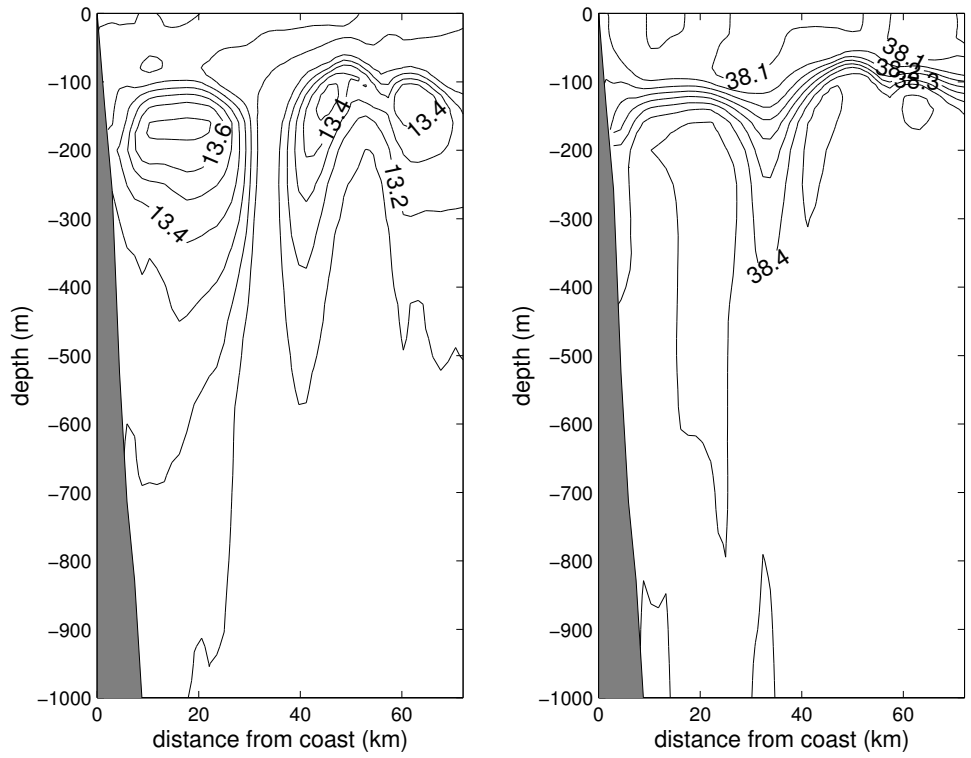

Figure 3.12: Temperature (in °C) and salinity at section D of figure 3.1.

the interior, starting near Nice. The model simulates well the Liguro-Provençal front. Gradients are consistent, with a difference of about 0.3°C and 0.2 between coastal and interior waters. Coastal waters show temperature and salinity values of about 13.3°C and 38 , and in the interior the values are of about 13°C and 38.2 respectively during winter. In early spring (March), the salinity rises by about 0.1 over the whole Ligurian Sea, but for temperature, similar values to those of Gasparini *et al.* (1999) are obtained. The temperature of the coastal zones ranges between 13.9 and 14.5°C and between 13.6 and 13.7 °C in the interior of the basin (not shown). The Liguro-Provençal front can also be seen in remote sensed sea surface temperature (SST). Figure 3.13 shows a composite AVHRR (Advanced Very High Resolution Radiometers) SST for the 23 March 1998 and the corresponding model result. The model gives a quite satisfactory result, except near the Italian coast where the temperature is about 0.5 °C too low. Also the temperature at the surface is very smooth. This is due to the coarse resolution of the atmospheric heat flux.

3.6 Conclusions

A complex nesting system was implemented at three different resolutions in the Mediterranean Sea. Two successive zooms are performed to study small-scale processes of the Ligurian Sea. At a reasonable cost it was thus possible to set up a model of 1' resolution of

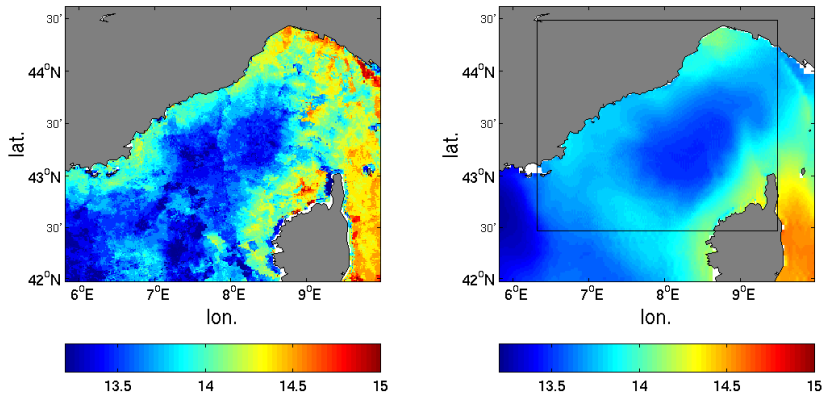

Figure 3.13: Remotely sensed SST (AVHRR) and model SST (in °C) for 23 March 1998

the Ligurian Sea by avoiding the open boundary problem. The two-way nesting approach appears to be an efficient and robust method for simulating oceanographic systems at variable resolutions. The Northern Current is better represented in the two-way nested model than in a coarser resolution grid and in the one-way nested model. The improvement was not only observed in the region covered by the high-resolution model, but also downstream in the coarse resolution model. The interaction between small-scale processes and larger scale currents and fronts are taken into account: meanders and eddies induced by instabilities of basin-wide currents and fronts are well observed in the intermediate and fine grid models. No particular problems were observed at the nesting boundaries. Due to the feedback and the sponge layer, the transition between the finer and the coarser grid model is smooth and confirms the robustness of the adopted nesting procedure.

The nesting system was able to simulate the major characteristics of the Ligurian Sea and the Liguro-Provençal basin. The three coastal currents influencing this area are present in the model at position, width and depth comparable to those obtained from different surveys. The transports of the Western Corsican Current (WCC), the Eastern Corsican Current (ECC) and the Northern Current (NC) are also consistent with values computed from observations. Especially the transports of the WCC and the ECC are in very good agreement with previous measurements (Sammari *et al.*, 1995; Astraldi and Gasparini, 1992; Astraldi *et al.*, 1999). However, a countercurrent next to the WCC is formed. This countercurrent is, to our knowledge, not observed and could be an artefact of the interpolation of the boundary condition. This current seems to weaken the NC and should be further analysed.

Chapter 4

Sequential assimilation methods

In the present review of sequential assimilation methods we limit ourselves to two classes of assimilation techniques: optimal interpolation and assimilation filters such as the Kalman filter. Given a certain number of observations, the central problem in data assimilation is how the model results and especially how the unobserved model variables should be corrected. Optimal interpolation and filtering methods are based on two different approaches to estimate the relation between the observed variables and the unobserved variables.

Optimal interpolation is based at the same time on a statistical optimality criterion and on physical assumptions concerning the relationship between the surface elevation, temperature, salinity and velocity. Dynamical balances between these variables such as geostrophy or a constant hydrostatic pressure at a given depth are used to prescribe the covariance between these variables. Data assimilation via optimal interpolation can therefore ensure that the model state is dynamically balanced.

For filter methods, the dynamical model itself is used to determine the relationship between variables. Different dynamical regimes can therefore implicitly be taken into account. For example, during a deep water formation, the surface properties are highly correlated to the subsurface properties, but a strong summer stratification almost decouples the surface layer from the lower layers. These changes in the dynamical regimes must be prescribed explicitly for optimal interpolation schemes, for example by parameterising the covariance differently during each season (De Mey and Benkiran, 2002).

Filtering methods such as the Kalman filter can however learn from the optimal interpolation schemes in order to validate the physical structures and the dynamical coherence of the relationship between the variables.

4.1 Optimal interpolation

In this section we will describe first the general frame of optimal interpolation and how the optimal interpolation schemes merge model results with observations. A

justification of the results will be given later on when the filtering methods are derived. Optimal interpolation can in fact be viewed as a simplified Kalman filter. But optimal interpolation does not require that the model is placed in a statistical frame. Therefore it is preferable to introduce it independently of the statistical description of the model.

Given the model forecast $\mathbf{x}^f$ and the observations $\mathbf{y}^o$, and their respective error covariances $\mathbf{P}^f$ and $\mathbf{R}$, we want to find the optimal combination of the model and the observations. For the optimal interpolation approach, these two error covariances are inputs to the assimilation problem.

We also assume that observations and model are related linearly by means of the observation operator $\mathbf{H}$. This is generally the case for most types of observations and models. Different ways are possible to define the optimal melting of observations and model results. They are discussed in section 4.3.2. If the errors follow a Gaussian pdf and if the observation operator is linear, they all lead to the same result. The model state $\mathbf{x}^a$ taking the model forecast $\mathbf{x}^f$ and the observations $\mathbf{y}^o$ into account, is given under these assumptions by:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{4.1}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{4.2}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f \tag{4.3}$$

The matrix $\mathbf{P}^a$ is the error covariance matrix of the *a posteriori* state $\mathbf{x}^a$. The main difficulty in optimal interpolation is the specification of the *a priori* error covariance matrix $\mathbf{P}^f$. If one is only interested in the *a posteriori* state, it is sufficient to know the matrix $\mathbf{H}\mathbf{P}^f$. This matrix represents the covariance between the complete state vector $\mathbf{x}^f$ and the observed part $\mathbf{H}\mathbf{x}^f$:

$$\mathbf{H}\mathbf{P}^f = E\left[\left(\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t\right)\left(\mathbf{x}^f - \mathbf{x}^t\right)^T\right] \tag{4.4}$$

This covariance tells us how an error, detected in the observed part of the state vector, should affect the unobserved variables.

4.1.1 Lowering and Lifting scheme

The Topex/Poseidon mission provided to the oceanographic community sea level measurements with valuable information on the ocean dynamics at mesoscale and at large scale. Cooper and Haines (1996) proposed an optimal interpolation scheme for altimetry assimilation. By ignoring the horizontal covariance in (4.1) and (4.2), the update for each single elevation point is given by:

$$\zeta^a = \zeta^f + \frac{{\sigma^f}^2}{{\sigma^f}^2 + {\sigma^o}^2}(\zeta^f - \zeta^o) \tag{4.5}$$

where ζ^o, ζ^f and ζ^a are the observed surface elevation, the *a priori* and the *a posteriori* elevation respectively. Especially for altimetry, one has to pay attention that the observed

elevation corresponds to the model elevation in terms of reference level and resolved processes as tides, inverse barometer effects and the seasonal steric cycle (Fox *et al.*, 2000a).

The change in temperature and salinity is based on the assumption that the horizontal velocity is zero at a given reference level H. This implies that the pressure is constant at this level and, under the hydrostatic approximation, the surface elevation, temperature and salinity are related in the following way:

$$\rho_0 g\zeta = -g\int_H^0 \rho(T,S)dz + p_b \tag{4.6}$$

where p_b is the constant pressure at the reference level and ρ_0 is the density of sea water at the surface. An elevation increment is therefore related to a density increment by:

$$\rho_0 g\Delta\zeta = -g\int_H^0 \Delta\rho dz \tag{4.7}$$

The whole vertical temperature and salinity profiles are lifted or lowered such that equation (4.7) is satisfied. Water is removed or added at the surface and at the bottom. The added top or bottom water has the same properties than the water already present at this location. Obviously, this method conserves the temperature and salinity properties. However, in weakly stratified regions, *e.g.* near the poles, this approach can not be applied since the displacement becomes very large even for small sea level increments.

4.1.2 System for Ocean Forecasting and Analysis (SOFA)

The major problem in data assimilation is to specify a realistic error covariance matrix that can be implemented in an efficient way. In particular, the error covariance should not need to be formed explicitly since it is a $n \times n$ matrix with $n \sim 10^5 - 10^7$. De Mey and Benkiran (2002) introduced a simplification operator, which projects the error covariance on a low dimensional error space. The columns of the $n \times r$ matrix $\mathbf{L}$ form an orthonormal basis of this error space and the projection operator[1] is therefore $\mathbf{L}^T$. The error covariance of the forecast in this error space is noted by $\tilde{\mathbf{P}}^f$:

$$\mathbf{P}^f = \mathbf{L}\tilde{\mathbf{P}}^f\mathbf{L}^T \tag{4.8}$$

The Kalman gain $\mathbf{K}$ and the Kalman gain expressed in the reduced state space $\tilde{\mathbf{K}}$ are then given by:

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f(\mathbf{HL})^T\left[(\mathbf{HL})\tilde{\mathbf{P}}^f(\mathbf{HL})^T + \tilde{\mathbf{R}}\right]^{-1} \tag{4.9}$$

$$\mathbf{K} = \mathbf{L}\tilde{\mathbf{K}} \tag{4.10}$$

[1] In De Mey and Benkiran (2002) the projection operator is noted $\mathbf{S}$ and the error space is defined by the columns of $\mathbf{S}^T$. A different notation is chosen here for coherence with other chapters.

The error space reduction decreases the error covariance at the observation locations $\mathbf{HP}^f\mathbf{H}$.

$$\mathbf{HP}^f\mathbf{H} = E\left[\left(\mathbf{Hx}^f - \mathbf{Hx}^t\right)\left(\mathbf{Hx}^f - \mathbf{Hx}^t\right)^T\right] \tag{4.11}$$

This loss of variance must be compensated by increasing the observation error covariance $\tilde{\mathbf{R}}$ such that:

$$\mathbf{HP}^f\mathbf{H} + \mathbf{R} = (\mathbf{HL})\tilde{\mathbf{P}}^f(\mathbf{HL})^T + \tilde{\mathbf{R}} \tag{4.12}$$

The additional observation error can be interpreted as a representativeness error, since an error perpendicular to the error space formed by $\mathbf{L}$ cannot be corrected and represented in this given error space. In practice, $\tilde{\mathbf{R}}$ is specified directly without estimating explicitly the contribution of the perpendicular space. Therefore, many authors do not make a difference between $\mathbf{R}$ and $\tilde{\mathbf{R}}$.

Two choices were considered by De Mey and Benkiran (2002) for the error reduction: 3D EOFs and 1D vertical EOFs. The 3D EOFs are intended to be the eigenvectors of the error covariance matrix $\mathbf{P}^f$. Therefore they should be computed from an ensemble of states representing the uncertainty associated to the model results. For instant, the difference between the observations or model output and the climatology, or the model result minus the time mean are possible choices (Pham *et al.*, 1998; Brasseur *et al.*, 1999). The EOFs can also be obtained from an ensemble of model states by perturbing uncertain inputs to the model, *e.g.* the position of an incoming current or even the bathymetry for a 2D wave forecast (De Mey *et al.*, 2004). The matrix $\tilde{\mathbf{P}}^f$ is consequently diagonal and holds the eigenvalue of the corresponding EOFs.

In another approach, the vertical and horizontal covariances are handled differently. The matrix $\mathbf{L}$ is formed by a series of $n' \times r$ blocks involving only the n' grid points of the same water column.

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}^{(1)} & & \mathbf{0} \\ & \mathbf{L}^{(1)} & \\ \mathbf{0} & & \ddots \end{pmatrix} \tag{4.13}$$

The blocks $\mathbf{L}^{(j)}$ are composed of the vertical EOFs which can vary in space. The matrix $\tilde{\mathbf{P}}^f$ is no longer diagonal. But we assume that the kth vertical EOF amplitude is only correlated with the kth vertical EOF amplitude of the other water column. This implies that $\tilde{\mathbf{P}}^f$ can be rearranged in r blocks along its diagonal. Each of these blocks describes the horizontal covariance of a given EOF amplitude. This covariance $\tilde{\mathbf{P}}^{f,i}$ is modelled by specifying the variance $\mathbf{D}^i$ and the correlation $\mathbf{C}^i$ separately.

$$\tilde{\mathbf{P}}^{f,i} = {\mathbf{D}^i}^{1/2}\mathbf{C}^i{\mathbf{D}^i}^{1/2} \tag{4.14}$$

where $\mathbf{D}^i$ is a diagonal matrix and the diagonal elements of $\mathbf{C}^i$ are all equal to 1 and the other elements of the correlation matrix are between -1 and 1. $\mathbf{D}^i$ is specified on the basis of the eigenvalue of the vertical EOFs and the horizontal correlation $\mathbf{C}^i$ is parameterised

by a given function of the distance between two grid points.

The SOFA analysis scheme was implemented *e.g.* in the Azores-Madeira region (Gavart *et al.*, 1999) and in North Atlantic (Faucher *et al.*, 2002). On an operationally basis, SOFA has been applied to the Mediterranean Sea in frame of the MFS projects (De Mey and Benkiran, 2002) and to the North and Tropical Atlantic, the Mediterranean Sea and to the Global Ocean within the MERCATOR project (Bahurel *et al.*, 2001, 2004).

4.2 Statistical description

Optimal interpolation methods require the specification of the model state's error covariance. The choice of the error covariance for those schemes is based partially on physical assumptions. The magnitude of a probable error can also be estimated by identifying the error sources of the model and by studying the propagation of the error through the model. In the following section, the model is placed in a broader statistical frame, which takes the uncertainties affecting the model results into account.

4.2.1 Probabilistic forecast

If the system state is perfectly known at an initial time t_0, a perfect model would allow us to forecast the state of the system at any future state by performing a series of successive forecasts.

$$\mathbf{x}_0 = \mathbf{x}^i \tag{4.15}$$
$$\mathbf{x}_{i+1} = M_i(\mathbf{x}_i) \tag{4.16}$$

It is clear that in real world, neither the initial state vector nor the model are perfectly known. The number and the quality of observations are also limited and therefore errors on the estimated initial state always remain. The uncertainties of the model stem from two different sources: the boundary conditions and the discretised model equation including the parameterisation of unresolved processes. The true state $\mathbf{x}_i^t$ is therefore governed by the following system introducing two unknown error terms: $\boldsymbol{\zeta}^i$ and $\boldsymbol{\eta}_i$

$$\mathbf{x}_0^t = \mathbf{x}^i + \boldsymbol{\zeta}^i \tag{4.17}$$
$$\mathbf{x}_{i+1}^t = M_i(\mathbf{x}_i^t) + \boldsymbol{\eta}_i \tag{4.18}$$

The estimated initial state vector differs from the true initial state by the error $\boldsymbol{\zeta}^i$, which is as unknown and inaccessible as the true initial state. But our experience can tell us what a probable error is and what is not. For instance, a small error is likelier than a large error. An error that would produce a dynamical inconsistent true initial state is also improbable. It is therefore interesting to describe the error by a probabilistic density function (pdf). The uncertainties lead us to the representation of the error of the initial condition as a random vector following a given pdf.

The estimated initial state is then a realisation of this random vector. It should be noted that the description of the true initial state by a random variable reflects only our limited capability to determine the true state of the studied deterministic system.

The approximate dynamical model is another source of error. At each integration step, the error $\boldsymbol{\eta}_i$ is introduced. This error leads in most cases to a progressive error growth of the forecast. As before, we assume that we know what a likely error is and what is not. The unknown error $\boldsymbol{\eta}_i$ is therefore treated as a random vector following a known distribution.

If these errors are small, the straightforward integration of equation (4.16) by neglecting the errors, can give us a likely future state of the system for short forecast time lead. For many applications, however, it is not sufficient to provide only the forecast of a system. An estimation of its reliability is also necessary in some cases, in particular for data assimilation. Therefore we seek the pdf of the forecast $\mathbf{x}_i$.

The pdf of the state vector at the initial time is the pdf of the initial condition noted $p_{\mathbf{x}_i}(x)$. For a moment, the model error is neglected, *i.e.* the model is considered as perfect. The probability that a forecast $\mathbf{x}_{i+1}$ has a value within the small volume $d\mathbf{x}$ centred in $\mathbf{x}$, is the probability that the system was at time t_i near the state $\mathbf{x}' = M_i^{-1}(\mathbf{x})$ within the corresponding volume $d\mathbf{x}'$.

$$p_{\mathbf{x}_0}(\mathbf{x}) = p_{\mathbf{x}_i}(\mathbf{x}) \tag{4.19}$$

$$p_{\mathbf{x}_{i+1}}(\mathbf{x}) = \left|\frac{\partial M_i(\mathbf{x}')}{\partial \mathbf{x}'}\right|^{-1} p_{\mathbf{x}_i}(\mathbf{x}') \qquad \text{for } \mathbf{x} = M_i(\mathbf{x}') \tag{4.20}$$

where $\left|\frac{\partial M_i}{\partial \mathbf{x}'}\right|$ is the Jacobian of the function M_i. It can be interpreted as the ratio of the volumes $d\mathbf{x}'$ and $d\mathbf{x}$.

This relation implies the inverse of M_i. For a general function, this inverse might not exist or is not unique. If two different states can produce the same forecast, then the probability of this forecast is simply the sum of the probabilities assuming the system was either in one state or in the other. The probability of a forecast for which no previous state exists is of course zero. But for a model M_i derived from a dynamical equation, the inverse exists and can be obtained by inversing the time step. In practice, this is not possible for stability reasons. Therefore we will introduce an alternative formulation, the Fokker-Plank equation, based only on the forward model.

So far we have neglected the model error. The ocean is a dissipative system and forced by the atmosphere. Therefore, it will in most cases "forget" the error of the initial condition after the integration over several characteristic time lengths. The model error (including the forcing errors) will be then the dominant error source (Navarra, 2002). We suppose the random vector $\boldsymbol{\eta}_i$ follows the distribution $p_{\boldsymbol{\eta}_i}(\mathbf{x})$ and that it is neither time correlated nor correlated with the error of the initial condition. The error $\boldsymbol{\eta}_i$ introduce a spread of the probability and can be computed by (Pham, 2001):

$$p_{\mathbf{x}_0}(\mathbf{x}) = p_{\mathbf{x}_i}(\mathbf{x}) \tag{4.21}$$

$$p_{\mathbf{x}_{i+1}}(\mathbf{x}) = \int_{I\!R^n} \left[\left| \frac{\partial M_i(\mathbf{x}')}{\partial \mathbf{x}'} \right|^{-1} p_{\mathbf{x}_i}(\mathbf{x}') \right]_{x'=M_i^{-1}(\mathbf{y})} p_{\boldsymbol{\eta}_i}(\mathbf{x}-\mathbf{y}) d\mathbf{y} \tag{4.22}$$

This equation gives a complete description of the probabilistic, nonlinear time evolution of the system. Its solution provides us the probability of finding the system in a certain state at any time t_i. If the dynamical model is unbiased, the last equation yields the Fokker-Plank equation in the limit of a continuous time and for a Gaussian distributed model error. The continuous time model of the so-called Markov process can be written as follows:

$$d\mathbf{x}(t) = g(\mathbf{x}, t)dt + d\boldsymbol{\eta} \tag{4.23}$$

where $d\boldsymbol{\eta}$ is a random term describing the model uncertainties. The characteristics of the model error are more difficult to define in the continuous-time case. The error integrated over the time interval Δt has a variance of $\Delta t\mathbf{Q}$. The normalised error $\mathbf{Q}^{-1/2}d\boldsymbol{\eta}$ is also called a Wiener process. The pdf of $\mathbf{x}(t)$ is the solution of the Fokker-Plank equation:

$$\frac{\partial p_{\mathbf{x}}}{\partial t} + \sum_{\alpha=1}^{n} \frac{\partial (g_\alpha p_{\mathbf{x}})}{\partial \mathbf{x}_\alpha} = \frac{1}{2} \sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \frac{\partial^2 p_{\mathbf{x}}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta} \tag{4.24}$$

An original proof of the Fokker-Plank equation derived from equation (4.22) can be found in appendix A. A detailed derivation can also be found in Jazwinski (1970). However, it is impracticable for large models such as ocean models to solve (4.22) or the Fokker-Planck equation directly. The pdf $p_{\mathbf{x}_{i+1}}$ is an n-dimensional function. If $p_{\mathbf{x}_{i+1}}$ is discretised on a grid of for example 100 grid points in each dimension, the pdf would be composed of 100^n scalars where n is about 10^6 for most ocean models. But fortunately we do not need to know the probability of all states. An important part of the state space represents very unlikely ocean states. For data assimilation, the exact probabilities of these states are irrelevant. Therefore, the Monte Carlo method and the Lyapunov Equations only explore the small part of the state space constituted by likely states.

4.2.2 Ensemble forecasts, the Monte Carlo method

The ensemble method for probabilistic forecasts in the context of data assimilation was first applied in oceanography by Evensen (1994) and in meteorology by Houtekamer and Mitchell (1998). A review of this method can be found in Evensen (2003). It consists in the generation of a large number N of perturbations of the two errors sources: the initial conditions and the model error. The perturbations ${\boldsymbol{\zeta}^i}^{(k)}$ and ${\boldsymbol{\eta}_i}^{(k)}$ for $k = 1...N$ are simply realisations of the random vectors $\boldsymbol{\zeta}^i$ and $\boldsymbol{\eta}_i$. An ensemble of future ocean states is then obtained by integrating the model for each member separately.

$$\mathbf{x}_0^{(k)} = \mathbf{x}^i + {\boldsymbol{\zeta}^i}^{(k)} \tag{4.25}$$

$$\mathbf{x}_{i+1}^{(k)} = M_i(\mathbf{x}_i^t) + {\boldsymbol{\eta}_i}^{(k)} \tag{4.26}$$

Each member of this ensemble can be integrated independently of the others. Ensemble forecasts can therefore be implemented very efficiently on parallel clusters.

In principle, the pdf of the future ocean state could be computed by the fraction of ensemble members given in a certain volume of the state space. But for data assimilation this step is not necessary since the analysis can be expressed directly in terms of the ensemble members.

The advantage of ensemble forecasts is their relatively simple concept. Often it is also easier to define a likely perturbation than a pdf. In particular, the error term $\boldsymbol{\eta}_i$ can be implicitly defined by perturbing, for each ensemble member, uncertain forcings and parameters. Each model can be integrated with perturbed atmospheric forcings and a different set of empirical constants of the parameterised processes. But the perturbation of each member must reflect the real uncertainty associated to the forcings terms or parameters (Houtekamer and Mitchell, 1998).

Some members can also be obtained by a completely different model. Such ensembles are called poor man's ensemble (Ziehmann, 2000) or super-ensembles when the different model results are merged into an improved deterministic forecast (Krishnamurti *et al.*, 1999). The assimilation of independent data can also *a posteriori* also help to determine which combination of perturbation produces the best forecast. This is interesting for calibration and validation of the model and its parameterisations.

The main drawback however is that large ensembles are needed to obtain accurate error statistics. The statistical moments of the pdf, in particular the error covariance which plays a central role in data assimilation, are affected by a relative error of the order of $N^{-1/2}$ where N is the ensemble size. This problem can be reduced by choosing the perturbation with care (Houtekamer, 1995; Evensen, 2004).

4.2.3 The Lyapunov equation

Ocean models are generally nonlinear, but the linear case is important since it leads to some fundamental properties of error propagation. Some properties are also applicable to weakly nonlinear models. In particular, growing and decaying error modes can be determined for the linear models. Pham *et al.* (1998) has shown that this is important for the stability of an assimilation scheme.

Linear models

The true model dynamics can be written in the linear context as:

$$\mathbf{x}_0^t = \mathbf{x}^i + \boldsymbol{\zeta}^i \tag{4.27}$$

$$\mathbf{x}_{i+1}^t = \mathbf{M}_i \mathbf{x}_i^t + \mathbf{f}_i + \boldsymbol{\eta}_i \tag{4.28}$$

The vector $\mathbf{f}_i$ represent all terms of the linear model independent of the system state. For linear ocean models, this vector represents the external forcings such as the winds. The model forecast error is defined by the difference between the true state $\mathbf{x}_i^t$ and the predicted state $\mathbf{x}_i$.

$$\boldsymbol{\zeta}_i = \mathbf{x}_i^t - \mathbf{x}_i \tag{4.29}$$

This error of the model state is governed by the following stochastic equation:

$$\begin{aligned} \boldsymbol{\zeta}_0 &= \boldsymbol{\zeta}^i & (4.30)\\ \boldsymbol{\zeta}_{i+1} &= \mathbf{M}_i\boldsymbol{\zeta}_i + \boldsymbol{\eta}_i & (4.31) \end{aligned}$$

For a linear model, the evolution of each statistical moment of $\boldsymbol{\zeta}_i$ is independent of each other. We also assume that the initial condition and the model are unbiased and their error covariance are written as $\mathbf{P}^i$ and $\mathbf{Q}_i$ respectively.

$$\begin{aligned} E(\boldsymbol{\zeta}^i) &= 0 \qquad & \mathbf{P}^i &= E(\boldsymbol{\zeta}^i {\boldsymbol{\zeta}^i}^T)\\ E(\boldsymbol{\eta}_i) &= 0 \qquad & \mathbf{Q}_i &= E(\boldsymbol{\eta}_i {\boldsymbol{\eta}_i}^T) \end{aligned}$$

By taking the expectation of equation (4.28) one can show that the mean of the random vector $\mathbf{x}_{i+1}$ is the central forecast:

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{x}^i & (4.32)\\ \mathbf{x}_{i+1} &= \mathbf{M}_i\mathbf{x}_i & (4.33) \end{aligned}$$

and the error covariance $\mathbf{P}_i$ is given by:

$$\begin{aligned} \mathbf{P}_0 &= \mathbf{P}^i & (4.34)\\ \mathbf{P}_{i+1} &= \mathbf{M}_i\mathbf{P}_i\mathbf{M}_i^T + \mathbf{Q}_i & (4.35) \end{aligned}$$

This equation is called the Lyapunov equation. It describes how the error variance and covariance are propagated through a linear model. It is interesting to note that the Lyapunov equation does not make any assumption concerning the pdf of the errors $\boldsymbol{\zeta}^i$ and $\boldsymbol{\eta}_i$ except that their mean is zero.

But if these errors are Gaussian distributed, then also the model state vector at any time will follow a Gaussian distribution. This is due to the fact, that the linear stochastic model implies multiplications with a constant matrix and sums of two random vectors. In fact, the product of a constant matrix and Gaussian random vector and the sum of two Gaussian distributed variables can always be described by a Gaussian pdf.

Nonlinear models

The propagation of this error through a nonlinear model is obtained by subtracting the evolution of the true state (4.18) from the central forecast (4.16).

$$\boldsymbol{\zeta}_{i+1} = M_i(\mathbf{x}_i + \boldsymbol{\zeta}_i) - M_i(\mathbf{x}_i) + \boldsymbol{\eta}_i \tag{4.36}$$

If the error is supposed to be small, then the error propagation can be expanded by a Taylor series. The αth component of the model state's error $\boldsymbol{\zeta}_{i+1}$ is obtained by:

$$\boldsymbol{\zeta}_{i+1\alpha} = \frac{\partial M_{i\alpha}}{\partial x_\beta}\boldsymbol{\zeta}_{i\beta} + \frac{\partial^2 M_{i\alpha}}{\partial x_\beta \partial x_\gamma}\boldsymbol{\zeta}_{i\beta}\boldsymbol{\zeta}_{i\gamma} + \mathcal{O}(\boldsymbol{\zeta}^3) + \boldsymbol{\eta}_{i\alpha} \tag{4.37}$$

The greek indexes denote the components of vectors of matrices. For clarity, the summation sign for repeated indexes has been omitted. We neglect here the third order contribution in the error of the model state. As it was shown before, the error $\boldsymbol{\zeta}_i$ can be expressed as a probability density function. From equation (4.37) the error can also be quantified in terms of its statistical moments. The two first statistical moments are:

$$E\left[\boldsymbol{\zeta}_{i+1\alpha}\right] = \frac{\partial M_{i\alpha}}{\partial x_\beta}E\left[\boldsymbol{\zeta}_{i\beta}\right] + \frac{\partial^2 M_{i\alpha}}{\partial x_\beta \partial x_\gamma}E\left[\boldsymbol{\zeta}_{i\beta}\boldsymbol{\zeta}_{i\gamma}\right] + E\left[\mathcal{O}(\boldsymbol{\zeta}^3)\right] \tag{4.38}$$

$$E\left[\boldsymbol{\zeta}_{i+1\alpha}\boldsymbol{\zeta}_{i+1\beta}\right] = \frac{\partial M_{i\alpha}}{\partial x_\gamma}\frac{\partial M_{i\beta}}{\partial x_\delta}E\left[\boldsymbol{\zeta}_{i\gamma}\boldsymbol{\zeta}_{i\delta}\right] + E\left[\mathcal{O}(\boldsymbol{\zeta}^3)\right] + \mathbf{Q}_{i\alpha\beta} \tag{4.39}$$

The first moment is the mean error or the bias and the second moment is related to the covariance by:

$$\mathbf{P}_{i\alpha\beta} = E\left[\boldsymbol{\zeta}_{i\alpha}\boldsymbol{\zeta}_{i\beta}\right] - E\left[\boldsymbol{\zeta}_{i\alpha}\right]E\left[\boldsymbol{\zeta}_{i\beta}\right] \tag{4.40}$$

Some important properties of linear models are no longer true if the model is nonlinear. First of all, even if the initial condition and the model are unbiased, there is no guarantee that the nonlinear forecast will also be unbiased. This effect can be shown with a simple example. The bottom stress τ is often parameterised by a quadratic function of the velocity u:

$$\tau = au^2 \tag{4.41}$$

We suppose that this parameterisation is correct and that the velocity estimation is unbiased. A velocity error $\boldsymbol{\zeta}_u$, lead to the following error in the bottom stress:

$$\boldsymbol{\zeta}_\tau = a(u + \boldsymbol{\zeta}_u)^2 - \tau = 2au\boldsymbol{\zeta}_u + a\boldsymbol{\zeta}_u^2 \tag{4.42}$$

The mean error in the bottom stress is therefore positive:

$$E\left[\boldsymbol{\zeta}_\tau\right] = aE\left[\boldsymbol{\zeta}_u^2\right] > 0 \tag{4.43}$$

Positive velocity errors have a larger impact on the bottom stress than negative errors. Both errors do not compensate each other and therefore a positive bias remains.

This approach is very similar to the turbulence closure problem. The errors here correspond to the subgrid scale processes of the closure problem. Therefore, the model parameterisation takes the bias due to unresolved subgrid scale features (here the bottom boundary layer), already into account. The errors resolved on the model grid however, have also an effect on the expected mean flow.

The other characteristic of the probabilistic nonlinear models forecast is that the statistical moments are no longer independent of each other (*e.g.* Evensen, 1994). The forecast of a statistical moment may depend on all higher moments. It has been shown for instance, that the mean error depends on the covariance. In the extended Kalman filter, all moments higher than the covariance are neglected. More sophisticated closure schemes for the statistical moments exist (Fleming, 1971a,b; Leith, 1971; Leith and Kraichnan, 1972; Leith, 1974) but their application to large ocean models is questionable since the estimation of the error covariance is already very difficult and can only be done in a approximative way.

The coupling of the statistical moments has also the consequence that an initial Gaussian distributed error may become distorted. The pdf of the forecast error can have a more complex distribution than a simple Gaussian one. For example, the error can follow for a multimodal pdf. Such a phenomenon occurs for instance in the case of flow instabilities leading to two (or more) likely but distinct ocean states.

For a weakly nonlinear ocean model, the second order derivative of the model is neglected in equations (4.38) and (4.39). This is the approach chosen to derive the extended Kalman filter.

As in the linear case, the pdf of the model state's error can be approximated at any time by a Gaussian function if the errors $\boldsymbol{\zeta}^i$ and $\boldsymbol{\eta}_i$ are Gaussian random vectors. Then, the pdf of the model state's error is determined only by its mean and covariance. The mean of this pdf is the central forecast $\mathbf{x}_i$ and the error covariance is obtained by the tangent linear model $\mathbf{M}_i$:

$$\mathbf{x}_{i+1} = M_i(\mathbf{x}_i) \tag{4.44}$$

$$\mathbf{P}_{i+1} = \mathbf{M}_i \mathbf{P}_i \mathbf{M}_i^T + \mathbf{Q}_i \tag{4.45}$$

The tangent linear model $\mathbf{M}_i$ is computed at the state $\mathbf{x}_i$.

$$\mathbf{M}_i = \left[\frac{\partial M_i}{\partial \mathbf{x}}\right]_{\mathbf{x}_i} \tag{4.46}$$

The error forecast of the extended Kalman filter can be schematised as in figure 4.1. Covariances are always positive defined matrices. The hypersurface in the state space of

all states with the same probability is thus an ellipsoid for a Gaussian pdf. The shape of this ellipsoid depends on the covariance matrix. The principal axes, for example, are the eigenvectors of the covariance matrix. Figure 4.1 shows the ellipsoids representing the error covariances $\mathbf{P}_i$ and $\mathbf{P}_{i+1}$ of the states $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$. The vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are the eigenvectors of $\mathbf{P}_i$ multiplied by the square root of the corresponding eigenvalues.

$$\mathbf{P}_i = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T \tag{4.47}$$

The vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ have therefore the magnitude of a typical error.

If the true system is not in the state $\mathbf{x}_i$, but in the state $\mathbf{x}_i + \mathbf{u}_1$, then at time t_{i+1}, the true system will be in the following state:

$$M_i(\mathbf{x}_i + \mathbf{u}_1) \sim \mathbf{x}_{i+1} + \mathbf{M}_i\mathbf{u}_1 \tag{4.48}$$

provided that the model is perfect ($\mathbf{Q}_i = 0$). In the same way, the state $\mathbf{x}_i + \mathbf{u}_2$ maps on the state $\mathbf{x}_{i+1} + \mathbf{M}_i\mathbf{u}_2$ at time t_{i+1}.

Every point $\mathbf{x} = \mathbf{x}_i + \boldsymbol{\zeta}_i$ on the ellipsoid centred at $\mathbf{x}_i$ satisfies the following condition:

$$\boldsymbol{\zeta}_i^T\mathbf{P}_i^{-1}\boldsymbol{\zeta}_i = 1 \tag{4.49}$$

The error $\boldsymbol{\zeta}_i$ at time t_i has grown or shrunken to $\boldsymbol{\zeta}_{i+1} = \mathbf{M}_i\boldsymbol{\zeta}_i$ at time t_{i+1}. The hypersurface representing the errors at t_{i+1} can be obtained by substituting $\boldsymbol{\zeta}_i$ by $\mathbf{M}_i^{-1}\boldsymbol{\zeta}_{i+1}$ in equation (4.49):

$$\boldsymbol{\zeta}_{i+1}^T\left(\mathbf{M}_i^{-1}\right)^T\mathbf{P}_i^{-1}\mathbf{M}_i^{-1}\boldsymbol{\zeta}_{i+1} = 1 \tag{4.50}$$

$$\boldsymbol{\zeta}_{i+1}^T\left(\mathbf{M}_i\mathbf{P}_i\mathbf{M}_i^T\right)^{-1}\boldsymbol{\zeta}_{i+1}^T = 1 \tag{4.51}$$

The ellipsoid of the error at time t_{i+1} is therefore represented by the following covariance matrix:

$$\mathbf{P}_{i+1} = \mathbf{M}_i\mathbf{P}_i\mathbf{M}_i^T \tag{4.52}$$

In fact, this is the forecast equation of the error covariance matrix (4.45) for a perfect model.

In summary, the linear theory for error forecast can also be applied to weakly nonlinear models. The mean of the Gaussian pdf is prediced by the full nonlinear model and the error covariance by the tangent linear model. The criterion "weakly nonlinear" means that the nonlinearities of the model for a typical error ε must be small. For instance, the following approximation must be possible:

$$M_i(\mathbf{x} + \varepsilon) \sim M_i(\mathbf{x}) + \mathbf{M}_i\varepsilon \tag{4.53}$$

This means that perturbations around the central forecast with the magnitude of a typical error can be forecast by the tangent linear model. This requirement can fail either if model is too nonlinear or if the errors are too large.

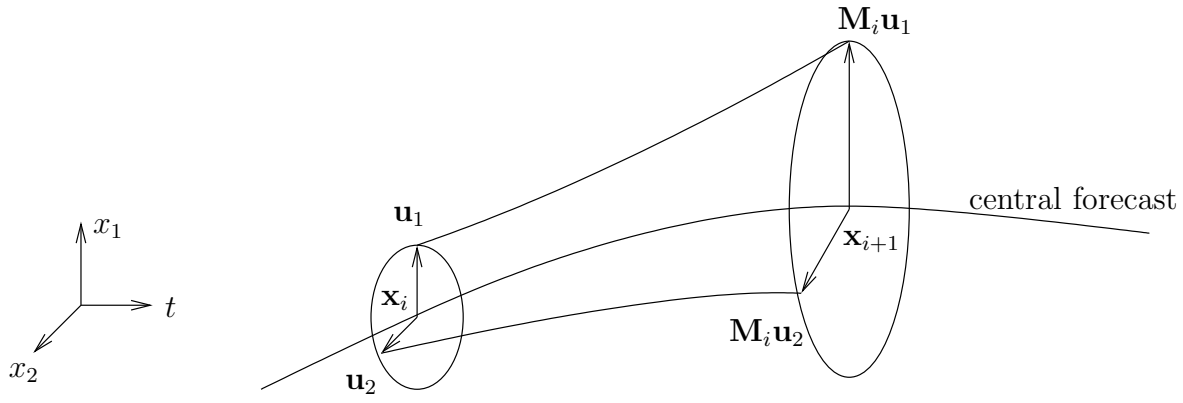

Figure 4.1: Forecast of the error covariance with the tangent linear model

4.2.4 The Lyapunov equation with a Reduced Rank error covariance

Error forecast with a square root representation

The error covariance matrices can also be expressed in terms of the following products:

$$\mathbf{P}^i = \mathbf{S}^i {\mathbf{S}^i}^T \tag{4.54}$$
$$\mathbf{Q}_i = \mathbf{S}_{\mathbf{Q}_i} {\mathbf{S}_{\mathbf{Q}}}_i^T \tag{4.55}$$
$$\mathbf{P}_i = \mathbf{S}_i {\mathbf{S}_i}^T \tag{4.56}$$

The representation of the error covariance in terms of its square root matrices has the advantage that the underlying error covariance is by definition symmetric and positive definite (Verlaan, 1998). But the main advantage is that with the square root representation one can easily derive a reduced rank error forecast requiring much less model integration than the Lyapunov equation (4.45).

The link between ensemble members and error covariance can be highlighted with this formulation. The columns of $\mathbf{S}^i$ can be interpreted as perturbations of the initial condition. An ensemble generated by this perturbations would in fact have a covariance of $\mathbf{P}^i$. A more detailed description of the relation between an ensemble and the square root representation is given in section 4.2.5. It is possible to express the Lyapunov equation using the square root matrices:

$$\mathbf{S}_0 = \mathbf{S}^i \tag{4.57}$$
$$\mathbf{S}'_{i+1} = \begin{pmatrix} \mathbf{M}_i\mathbf{S}_i & \mathbf{S}_{\mathbf{Q}_i} \end{pmatrix} \tag{4.58}$$
$$\mathbf{S}_{i+1} = \mathbf{S}'_{i+1}\mathbf{U}_i \tag{4.59}$$

In the operation (4.58) all rows of $\mathbf{M}_i\mathbf{S}_i$ and $\mathbf{S}_{\mathbf{Q}_i}$ are concatenated to form an $n \times 2n$ matrix. But only n of these errors vector are linearly independent. The $2n \times n$ matrix $\mathbf{U}_i$ (with $\mathbf{U}_i^T\mathbf{U}_i = \mathbf{I}$) is chosen in a way that the covariance is not modified:

$$\mathbf{S}_{i+1}\mathbf{S}_{i+1}^T = \mathbf{S}'_{i+1}{\mathbf{S}'_{i+1}}^T \tag{4.60}$$

This is the forecast step of the SQRT filter. It involves as many model integrations as linearly independent error vectors in $\mathbf{S}_i$. In the general case, it implies therefore n model forecasts. The unacceptable numerical load can be decreased by approaching the covariance matrix $\mathbf{P}_i$ by a reduced rank matrix.

$$\mathbf{P}_i = \mathbf{S}_i{\mathbf{S}_i}^T + \boldsymbol{\xi} \tag{4.61}$$

where $\mathbf{S}_i$ is now an $n \times r$ matrix with $r << n$. The error $\boldsymbol{\xi}$ is minimum under the spectral[2] and Frobenius norm[3] if the columns of $\mathbf{S}_i$ are defined by the r dominant eigenvectors and eigenvalues of $\mathbf{P}_i$ (Horn and Johnson, 1985, 1991). This error space reduction is performed during the step (4.59). The covariance matrix is then limited only to the r dominant eigenvectors and only r model integrations must be performed. In this procedure the covariance matrix $\mathbf{P}_i$ do not need to be formed explicitly. All operations are done on the square root matrices $\mathbf{S}^i$, $\mathbf{S}_{\mathbf{Q}_i}$ and $\mathbf{S}_i$. Under the reduced rank approximation, these $n \times r$ matrices are much smaller matrices than the $n \times n$ covariance matrices $\mathbf{P}^i$, $\mathbf{Q}_n$ and $\mathbf{P}_i$. This error forecast step is part of the Reduced Rank SQRT filter (RRSQRT filter).

Error forecast with a subspace representation

The traditional SEEK (Singular Evolutive Extended Kalman) filter, as introduced by Pham *et al.* (1998), and the ESSE (Error Subspace Statistical Estimation) approach (Lermusiaux, 1997; Lermusiaux and Robinson, 1999) are based on an eigenvector decomposition of the initial error covariance matrix $\mathbf{P}^i$.

$$\mathbf{P}^i = \mathbf{L}^i\tilde{\mathbf{P}}^i{\mathbf{L}^i}^T \tag{4.62}$$

where $\tilde{\mathbf{P}}^i$ is a diagonal $n \times n$ matrix containing the eigenvalues of $\mathbf{P}^i$ on its diagonal and the columns of $\mathbf{L}$ are the corresponding eigenvectors of the initial covariance matrix. We first consider the general case where all eigenvalues of $\mathbf{P}^i$ have arbitrary values. In a second step, we will assume that some of the eigenvalues are negligible. This hypothesis leads to a practical formulation of the error covariance forecast.

For a perfect model ($\mathbf{Q}_i = 0$), the Lyapunov equation yields:

$$\mathbf{L}_0 = \mathbf{L}^i \tag{4.63}$$
$$\mathbf{L}_{i+1} = \mathbf{M}_i\mathbf{L}_i \tag{4.64}$$

[2] The spectral norm $\|\mathbf{A}\|_2$ of a real matrix $\mathbf{A}$ is the square root of the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$.

[3] The Frobenius norm $\|\mathbf{A}\|_F$ of a $m \times n$ matrix is the square root of the sum of the absolute squares of its elements:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2}$$

The tangent linear model is applied to the columns of $\mathbf{L}_i$. It should be noted that they are not necessarily orthonormal after the application of the linear model $\mathbf{M}_i$. The error covariance at any time t_i is then given by

$$\mathbf{P}_i = \mathbf{L}_i \tilde{\mathbf{P}}^i {\mathbf{L}_i}^T \tag{4.65}$$

For a model with dynamical noise, the error covariance matrix $\mathbf{Q}_i$ must be projected onto the error directions given by the columns of $\mathbf{L}_i$ in order to obtain a formulation of the error covariance similar to equation (4.65). If all columns are linearly independent, one can show that:

$$\mathbf{Q}_i = \mathbf{L}_i \left(\mathbf{L}_i^T \mathbf{L}_i\right)^{-1} \mathbf{L}_i^T \mathbf{Q}_i \mathbf{L}_i \left(\mathbf{L}_i^T \mathbf{L}_i\right)^{-1} \mathbf{L}_i^T \tag{4.66}$$

The projection operator takes this complex expression since the columns are not orthonormal. If this is the case, the projection operator will take the more familiar form of $\mathbf{L}_i \mathbf{L}_i^T$, which has to be applied to the left and to the right of the model error covariance $\mathbf{Q}_i$.

The Lyapunov equation in the presence of model errors is then:

$$\begin{aligned}
\mathbf{L}_0 &= \mathbf{L}^i && (4.67)\\
\tilde{\mathbf{P}}_0 &= \tilde{\mathbf{P}}^i && (4.68)\\
\mathbf{L}_{i+1} &= \mathbf{M}_i \mathbf{L}_i && (4.69)\\
\tilde{\mathbf{P}}_{i+1} &= \tilde{\mathbf{P}}_i + \left(\mathbf{L}_i^T \mathbf{L}_i\right)^{-1} \mathbf{L}_i^T \mathbf{Q}_i \mathbf{L}_i \left(\mathbf{L}_i^T \mathbf{L}_i\right)^{-1} && (4.70)
\end{aligned}$$

However, this equation involves n integrations of the tangent linear model which is not feasible for oceanographic or atmospheric models. The interest of the SEEK error forecast step is the possibility of using a low rank approximation of the initial error covariance matrix:

$$\mathbf{P}^i \sim \mathbf{L}^i \tilde{\mathbf{P}}^i {\mathbf{L}^i}^T \tag{4.71}$$

where $\tilde{\mathbf{P}}^i$ is a diagonal $r \times r$ matrix containing only the r largest eigenvalues of $\mathbf{P}^i$ on its diagonal. The columns of $\mathbf{L}$ are the corresponding eigenvectors. The same approximation has been made for the RRSQRT error forecast. Finally, the equations (4.67)-(4.70) are solved for this reduced error covariance matrix $\tilde{\mathbf{P}}_i$ and the $n \times r$ matrix $\mathbf{L}_i$.

But SEEK and the RRSQRT error forecast handle the model error in a different way. Instead of extending the error space by the error modes introduced by the dynamical model, the SEEK "lengthens" the error vectors in order to take into account the model error. It assumes therefore that the model error lies in the space formed by the r columns of $\mathbf{L}_i$. In this case equation (4.66) also holds for $\mathbf{L}_i$ containing only r error modes. Model errors outside of the subspace are ignored. Since the error space evolves, different parts of the model error covariance $\mathbf{Q}_i$ are rejected.

Pham *et al.* (1998) proposed also a simple parameterisation of $\mathbf{Q}_i$ in terms of a "forgetting factor" ρ. This approach assumes that the model error covariance is a fraction of the state vector error covariance:

$$\mathbf{Q}_i = \frac{1-\rho}{\rho}\mathbf{L}_i\tilde{\mathbf{P}}_i\mathbf{L}_i^T \tag{4.72}$$

This parameterisation is consistent with the assumption that the model error lies in the subspace spanned by the r columns of $\mathbf{L}_i$. In the context of data assimilation, an adaptative and space dependent extension for the forgetting factor has been proposed by Testut *et al.* (2003). In order to extent the error space, Brasseur *et al.* (1999) have used an adaptive mechanism to add error modes based on the observation residuals.

The error forecast of the SEEK and the RRSQRT filters are based on the tangent linear model. The derivation of an exact tangent linear model of ocean models is an extremely complex task. Often the tangent linear model is approximated by the gradient:

$$[\mathbf{M}_i\mathbf{x}]_{\mathbf{x}_i} = \frac{1}{\varepsilon}\left(M_i(\mathbf{x}_i+\varepsilon\mathbf{x}) - M_i(\mathbf{x}_i)\right) \tag{4.73}$$

In theory ε should be as small as possible, but, in practice, better results are obtained if the vector $\varepsilon\mathbf{x}$ is of the order of magnitude of a typical error as it has been done in Brasseur *et al.* (1999). The specification of the dynamical error covariance matrix $\mathbf{Q}$ is a difficult task for the SEEK and the RRSQRT filter. This covariance contains for example the effect of a heat flux error at time t_i on the currents at time t_{i+1}. Such indirect repercussions are difficult to specify since they depend on the dynamics. The "forgetting factor" approximation is therefore often used

Error forecast with a fixed subspace

Another approach to forecast the model's error covariance consists in projecting the Lyapunov equation into a fixed low dimensional space. The matrix formed by the basis vector of this subspace is noted $\mathbf{E}$. The empirical orthogonal functions of the model time variability were used by Cane *et al.* (1996) to specify this low dimensional subspace. The error covariances $\mathbf{P}_i$ and $\mathbf{Q}_i$ and the tangent linear model are projected in this subspace by:

$$\begin{aligned}
\tilde{\mathbf{P}}_i &= \mathbf{E}\mathbf{P}_i\mathbf{E}^T && (4.74)\\
\tilde{\mathbf{M}}_i &= \mathbf{E}\mathbf{M}_i\mathbf{E}^T && (4.75)\\
\tilde{\mathbf{Q}}_i &= \mathbf{E}\mathbf{Q}_i\mathbf{E}^T && (4.76)\\
& && (4.77)
\end{aligned}$$

If a vector $\mathbf{x}$ belongs to the error subspace spanned by the columns of $\mathbf{E}$, then we assume that its forecast $\mathbf{M}_i\mathbf{x}$ still lies in this error space. This is a strong assumption for ocean models where nonlinear interactions couple the different scales and modes. The time evolution of the error covariance can then be expressed as:

$$\tilde{\mathbf{P}}_{i+1} = \tilde{\mathbf{M}}_i \tilde{\mathbf{P}}_i \tilde{\mathbf{M}}_i^T + \tilde{\mathbf{Q}}_i \tag{4.78}$$

The main difference between this approach and the RRSQRT and SEEK error forecast is that the error subspace remains constant. However, the error covariance within this space is allowed to change with time.

This approach reduces greatly the cost compared to the error propagation of the SEEK or the RRSQRT filter if the reduced model $\tilde{\mathbf{M}}_i$ can be maintained constant over a certain time interval at least.

4.2.5 Relationship between the ensemble and reduced rank representation

An ensemble is an approximation of the underlying probability density function. But when we take into account only the two first moments of the pdf, *i.e.* the mean and the covariance (and most analysis schemes do so), the ensemble can be treated as a reduced rank approximation of the error covariance.

On the other hand, we will show that a reduced rank error covariance expressed in terms of its square root or eigenvectors/eigenvalues can also be transformed into an ensemble with the same covariance and a given mean. All this transformations imply matrix manipulations of the order of the ensemble size or the rank of the covariance matrix and one operation with a cost linear in n. The transition from one representation to another is therefore easy to perform.

It is important to realise the direct link between the different representations in order to take advantage of the different methods based either on the square root, eigendecomposition or ensemble representation. The ensemble representation allows stochastic forecasts and, to some extend, nonlinear observation operators and analysis of non-Gaussian variables by anamorphosis transform (Bertino *et al.*, 2002). But an important ensemble size is required. The convergence of the method as a function of the ensemble size N is of the order of $N^{-1/2}$. A square root factorisation or an eigendecomposition of the covariance matrix converge substantially faster since it is based on a more clever choice of the error subspace than simply random perturbations. But the framework of these methods assumes linear or linearised models and observation operators.

From an ensemble to the square root of the error covariance

Pham (2001) derived these relationships and applied them in the SEIK filter which combines the ensemble forecast with SEEK analysis.

The covariance $\mathbf{P}$ of an ensemble of states $\mathbf{x}^{(k)}$ with $k = 1, \ldots, N$ is given by:

$$\mathbf{P} = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{x}^{(k)} - \overline{\mathbf{x}})(\mathbf{x}^{(k)} - \overline{\mathbf{x}})^T \tag{4.79}$$

where $\overline{\mathbf{x}}$ is the ensemble mean. We look now for a matrix square root $\mathbf{S}$ such that:

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T \tag{4.80}$$

It is convenient to introduce the $n \times N$ matrix $\mathbf{X}$ whose columns are the ensemble states:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} & \dots & \mathbf{x}^{(N)} \end{pmatrix} \tag{4.81}$$

Applied to this matrix $\mathbf{X}$, the operator $\mathbf{B}$ subtracts the ensemble mean:

$$\mathbf{B} = \mathbf{I}_{N\times N} - \frac{1}{N}\mathbf{1}_{N\times N} \tag{4.82}$$

where $\mathbf{I}_{N\times N}$ is the $N \times N$ identity matrix and $\mathbf{1}_{N\times N}$ the $N \times N$ matrix with all elements equal to 1.

With these notations it is straightforward to find a square root representation of the covariance, since:

$$\mathbf{P} = \frac{1}{N-1}\mathbf{X}\mathbf{B}\mathbf{B}^T\mathbf{X}^T \tag{4.83}$$

A possible form of the matrix $\mathbf{S}$ is therefore:

$$\mathbf{S} = \frac{1}{\sqrt{N-1}}\mathbf{X}\mathbf{B} \tag{4.84}$$

The columns of $\mathbf{S}$ are simply the difference between each ensemble member and the ensemble mean scaled by $\frac{1}{\sqrt{N-1}}$. But all the columns of $\mathbf{S}$ are not linearly independent since the sum of all columns is zero:

$$\mathbf{S}\mathbf{1}_{N\times 1} = 0 \tag{4.85}$$

The rank of the covariance matrix obtained from N ensemble members is therefore $N-1$ at most. Consequently, it should be possible to find an $n \times N-1$ matrix $\mathbf{S}$ satisfying equation (4.80). Equation (4.85) also implies that we can express the column of $\mathbf{S}$ as a linear combination of the other columns of $\mathbf{S}$. The N-th column of $\mathbf{S}$ is the opposite of the sum of all previous ones. The error space can hence be described by the $N-1$ first columns of $\mathbf{S}$.

The last column of $\mathbf{B}$ is also just the opposite of the sum of all previous ones. Therefore, we split the $\mathbf{B}$ matrix into two parts:

$$\mathbf{B} = \begin{pmatrix} \mathbf{T} & -\mathbf{T}\mathbf{1}_{N-1\times 1} \end{pmatrix} = \mathbf{T}\begin{pmatrix} \mathbf{I}_{N\times N} & -\mathbf{1}_{N\times 1} \end{pmatrix} \tag{4.86}$$

where $\mathbf{T}$ is the $N \times N-1$ matrix holding the $N-1$ first columns of $\mathbf{B}$. The matrix $\mathbf{T}$, as introduced by Pham (2001) has the following form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_{N-1\times N-1} \\ \mathbf{0}_{1\times N-1} \end{pmatrix} - \frac{1}{N}\mathbf{1}_{N\times N-1} \tag{4.87}$$

Using this $\mathbf{T}$ matrix, the error covariance $\mathbf{P}$ can be expressed as:

$$\begin{aligned} \mathbf{P} &= \frac{1}{N-1}\mathbf{X}\mathbf{T}\begin{pmatrix} \mathbf{I}_{N\times N} & -\mathbf{1}_{N\times 1} \end{pmatrix}\begin{pmatrix} \mathbf{I}_{N\times N} \\ -\mathbf{1}_{1\times N} \end{pmatrix}\mathbf{T}^T\mathbf{X}^T & (4.88)\\ &= \frac{1}{N-1}\mathbf{X}\mathbf{T}\left(\mathbf{I}_{N-1\times N-1} + \mathbf{1}_{N-1\times N-1}\right)\mathbf{T}^T\mathbf{X}^T & (4.89) \end{aligned}$$

A slightly different expression is proposed by (Pham, 2001):

$$\mathbf{P} = \frac{1}{N-1}\mathbf{X}\mathbf{T}\left(\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{X}^T \tag{4.90}$$

But the derivations presented here show us that the inverse in this last equation can be computed analytically. It seems that this property was not previously recognised. The analytical inverse allows us to go further and to extend the results of Pham (2001):

$$\left(\mathbf{T}^T\mathbf{T}\right)^{-1} = \mathbf{I}_{N-1\times N-1} + \mathbf{1}_{N-1\times N-1} \tag{4.91}$$

In order to obtain an expression for $\mathbf{S}$, we only need to factorise the rhs of (4.91) into its square root matrix. This can also be done analytically by making the following *ansatz*:

$$\mathbf{I}_{N-1\times N-1} + \mathbf{1}_{N-1\times N-1} = \left(\mathbf{I}_{N-1\times N-1} + \alpha\mathbf{1}_{N-1\times N-1}\right)\left(\mathbf{I}_{N-1\times N-1} + \alpha\mathbf{1}_{N-1\times N-1}\right)^T \tag{4.92}$$

By solving a second order equation in α, one can find the values of α

$$\alpha = \frac{-1 \pm \sqrt{N}}{N-1} \tag{4.93}$$

Either the positive or the negative signs can be taken here. By combining equation (4.89) and (4.92), one gets the final expression of $\mathbf{S}$:

$$\mathbf{S} = \frac{1}{\sqrt{N-1}}\mathbf{X}\mathbf{T}\left(\mathbf{I}_{N-1\times N-1} + \alpha\mathbf{1}_{N-1\times N-1}\right) \tag{4.94}$$

Finally, one can show that the columns of $\mathbf{S}$ can be obtained directly from the $N-1$ first ensemble members:

$$\mathbf{S}_k = \frac{1}{\sqrt{N-1}}\left(\mathbf{x}^{(k)} - \mathbf{x}_m\right) \qquad \text{with } k = 1, \ldots, N-1 \tag{4.95}$$

where the vector $\mathbf{x}_m$ is obtained from the ensemble mean $\overline{\mathbf{x}}$ and the Nth ensemble member $\mathbf{x}^{(N)}$ by:

$$\mathbf{x}_m = (1-\alpha)\overline{\mathbf{x}} + \alpha\mathbf{x}^{(N)} \tag{4.96}$$

The number of operations of the approach based on (4.95) and (4.96) is only of the order of nN. The direct solution (Pham, 2001) of the equation (4.90) to obtain a reduced rank

representation requires $\mathcal{O}(nN^2)$ operations.

This matrix $\mathbf{S}$ meets our initial requirement, *i.e.* it represents the same covariance than the ensemble. Furthermore, it has the minimum number of columns for a maximum rank ensemble. The naïve way of obtaining an $\mathbf{S}$ matrix with only $N-1$ columns consisting in simply ignoring the last column of $\mathbf{S}$ in (4.84) would lead to the correct error space but to an underestimated covariance. The last term in expression (4.94) can therefore be interpreted as a compensation factor necessary to obtain the right variance.

From the square root of the covariance to an ensemble

Knowing the ensemble mean and the square root of the covariance $\mathbf{S}$, one can construct an ensemble with the same covariance. Since the columns of $\mathbf{S}$ are related to the ensemble perturbations, we can try the following ensemble:

$$\mathbf{x}^{(k)} = \overline{\mathbf{x}} + \sqrt{N-1}\mathbf{S}_k \qquad \text{with } k = 1, \ldots, N-1 \tag{4.97}$$

The ensemble has indeed the right covariance around the given ensemble mean $\overline{\mathbf{x}}$, but the mean state of all $\mathbf{x}^{(k)}$ with $k = 1, \ldots, N-1$ is not necessary equal to $\overline{\mathbf{x}}$. By means of equation (4.97) we have only constructed $N-1$ ensemble members. If the Nth ensemble member was chosen in such way that the ensemble had the right mean, the covariance would be altered.

Using the same notation as in the previous section, we intend to construct an ensemble in a similar way as equation (4.97) where the ensemble perturbations are linear combinations of the columns of $\mathbf{S}$ (Pham, 2001).

$$\mathbf{X} = \overline{\mathbf{x}}\mathbf{1}_{N\times N} + \sqrt{N-1}\mathbf{S}\boldsymbol{\Omega}^T \tag{4.98}$$

where $\boldsymbol{\Omega}$ is an $N \times N-1$ matrix which holds the coefficients of the linear combinations of the columns of $\mathbf{S}$. We choose this matrix such that the ensemble has the right mean and covariance.

The covariance of this ensemble is given by:

$$\mathbf{P} = \mathbf{S}\boldsymbol{\Omega}^T\boldsymbol{\Omega}\mathbf{S}^T \tag{4.99}$$

The columns of $\boldsymbol{\Omega}$ must thus be normalised and orthogonal to each other such that $\boldsymbol{\Omega}^T\boldsymbol{\Omega} = \mathbf{I}$. The mean of this ensemble can be obtained by multiplying equation (4.98) to the right by $\frac{1}{N}\mathbf{1}_{N\times 1}$:

$$\frac{1}{N}\mathbf{X}\mathbf{1}_{N\times 1} = \overline{\mathbf{x}} + \frac{\sqrt{N-1}}{N}\mathbf{S}\boldsymbol{\Omega}^T\mathbf{1}_{N\times 1} \tag{4.100}$$

The columns of $\boldsymbol{\Omega}$ must therefore also be orthogonal to the vector $\mathbf{1}_{N\times 1}$. This is equivalent to require that the columns of $\boldsymbol{\Omega}$ have a zero mean. Hoteit *et al.* (2002) proposed a method for generating a random $\boldsymbol{\Omega}$ with the required properties. The generation of an ensemble by (4.98) is also called second order exact sampling since the mean and the

reduced rank error covariance are exactly represented by the ensemble.

The unscented transform is also a second order exact sampling technique transforming $\mathbf{S}^f$ into an ensemble (Julier and Uhlmann, 1997; Verlaan and Heemink, 2001). The ensemble contains $2r+1$ member and they are formed by:

$$\begin{aligned} \mathbf{x}^{(0)} &= \overline{\mathbf{x}} && (4.101)\\ \mathbf{x}^{(k)} &= \overline{\mathbf{x}} + \sqrt{r+\kappa}\mathbf{S}_k \qquad \text{with } k = 1,\ldots,r && (4.102)\\ \mathbf{x}^{(k+r)} &= \overline{\mathbf{x}} - \sqrt{r+\kappa}\mathbf{S}_k \qquad \text{with } k = 1,\ldots,r && (4.103) \end{aligned}$$

where κ is an arbitrary parameter chosen in such a way that the ensemble forecast yields relevant error modes. The following weight is attributed to each ensemble member:

$$\begin{aligned} w^{(0)} &= \frac{\kappa}{r+\kappa} && (4.104)\\ w^{(k)} &= \frac{1}{2}\frac{1}{r+\kappa} \qquad \text{with } k = 1,\ldots,2r && (4.105) \end{aligned}$$

The statistics of this ensemble obtained by the unscented transform are based on those weights. The ensemble mean and covariance are thus weighted sums and one can show that:

$$\begin{aligned} \overline{\mathbf{x}} &= \sum_{k=0}^{2r} w^{(k)} \left(\mathbf{x}^{(k)} - \overline{\mathbf{x}}\right) && (4.106)\\ \mathbf{P} = \mathbf{S}\mathbf{S}^T &= \sum_{k=0}^{2r} w^{(k)} \left(\mathbf{x}^{(k)} - \overline{\mathbf{x}}\right)\left(\mathbf{x}^{(k)} - \overline{\mathbf{x}}\right)^T && (4.107) \end{aligned}$$

This method generates $2r+1$ ensemble members. For an ensemble forecast, each member must be integrated by the model. The ensemble size should therefore be as small as possible. The method of Pham (2001) which produces $r+1$ ensemble members is thus preferred.

Another method to create an ensemble is the Monte Carlo approach. It consists in generating the ensemble by:

$$\mathbf{x}^{(k)} = \overline{\mathbf{x}} + \mathbf{S}\mathbf{v}^{(k)} \tag{4.108}$$

where the $N-1\times 1$ vectors $\mathbf{v}^{(k)}$ are drawn randomly from the standard Gaussian distribution. The accuracy of the Monte Carlo sampling augments with increasing the ensemble size. The difference between the ensemble mean and covariance and the $\overline{\mathbf{x}}$ and $\mathbf{S}\mathbf{S}^T$ is of the order of $N^{-1/2}$. In general large ensemble sizes are needed to obtain a representative ensemble.

Relationship between square root representation and eigenvector decomposition

The connection between a covariance matrix factorised in its square root matrix and its eigenvector and eigenvalues is much easier to prove than the relationship with the ensemble. We require that:

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T = \mathbf{L}\tilde{\mathbf{P}}\mathbf{L}^T \tag{4.109}$$

where $\mathbf{S}$ and $\mathbf{L}$ are $n \times r$ matrices and $\tilde{\mathbf{P}}$ is an $r \times r$ matrix. By a Cholesky decomposition we factorise $\tilde{\mathbf{P}}$ into $\tilde{\mathbf{P}} = \mathbf{C}\mathbf{C}^T$. This decomposition is trivial when $\tilde{\mathbf{P}}$ is diagonal. The corresponding square root representation of the $\mathbf{L}\tilde{\mathbf{P}}\mathbf{L}^T$ is given by:

$$\mathbf{S} = \mathbf{L}\mathbf{C} \tag{4.110}$$

We can obtain an eigenvector decomposition of $\mathbf{P}$ from a square root representation of the error covariance by carrying out the following singular value decomposition:

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{4.111}$$

where $\boldsymbol{\Sigma}$ is a diagonal $r \times r$ matrix containing the singular values of $\mathbf{S}$. $\mathbf{U}$ and $\mathbf{V}$ are respectively the left and right singular vectors of $\mathbf{S}$. An eigenvector decomposition can therefore be obtained by:

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T \tag{4.112}$$

All representations of the error covariance $\mathbf{P}$ have in common that they express $\mathbf{P}$ as a reduced rank matrix. It is therefore not surprising that a relationship between those representations exists. The rank of the error covariance matrix is actually a critical parameter for data assimilation since the error associated to the directions perpendicular to the error space is assumed to be zero. The prescribed rank should thus be as large as possible.

4.3 Analysis

In the previous section, we have explained how a probabilistic forecast can be carried out. Such a forecast allows us to determine the most likely ocean state and to assess the confidence that we can have in this forecast. A probabilistic forecast can also quantify the probability that a certain event occurs such as a "red tide" or an accumulation of wastewater in a coastal region. Probabilistic forecasts are also increasingly desired by a wide range of forecast users, especially in numerical weather prediction (*e.g.* Fritsch *et al.*, 1998).

The pdf of the model forecast is also necessary when the model results are combined with observations. The observations can reduce the uncertainties of the model and limit, to some extent, the error growth of an ocean model. The vector $\mathbf{y}_i^o$ contains all observations at time t_i. Sometimes, the time index is dropped for simplicity if all quantities are taken

at the same instance.

The number of observations m is generally much less than the dimension of the state space n. The enhanced resolution of ocean models and their increasing complexity makes the growth of the state space more rapid that the increased number of observations. Recent progresses of the observation systems have also increased the number of observations to assimilate into ocean models. However oceanographic measurements still remain costly and are limited in number.

The pdf $p_{\mathbf{x}_i}(\mathbf{x}|\mathbf{y}_i^o)$ of the ocean state given the observations $\mathbf{y}_i^o$ at the time t_i, can be inferred from the Bayes formula:

$$p_{\mathbf{x}_i}(\mathbf{x}|\mathbf{y}_i^o) = \frac{p_{\mathbf{y}_i^o}(\mathbf{y}_i^o|\mathbf{x})p_{\mathbf{x}_i}(\mathbf{x})}{\int_{I\!R^n} p_{\mathbf{y}_i^o}(\mathbf{y}_i^o|\mathbf{x})p_{\mathbf{x}_i}(\mathbf{x})d\mathbf{x}} \tag{4.113}$$

In words, the probability that the system is in state $\mathbf{x}$ given the observations $\mathbf{y}^o$ is proportional to the probability that the system state is $\mathbf{x}$ without knowing the observations times the probability that the observations are $\mathbf{y}^o$ provided the ocean is in state $\mathbf{x}$. The denominator is a normalisation constant.

Often, the error of the observations is supposed to be Gaussian distributed and its covariance is $\mathbf{R}_i$:

$$p_{\mathbf{y}_i^o}(\mathbf{y}) = (2\pi)^{-m/2}(\det \mathbf{R}_i)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - H_i(\mathbf{x}_i^t))^T \mathbf{R}_i^{-1}(\mathbf{y} - H_i(\mathbf{x}_i^t))\right) \tag{4.114}$$

For the Bayes formula, we are not bothered with the true state of the system $\mathbf{x}_i^t$ since it requires only the probability of the observations given an ocean state $\mathbf{x}$.

For systems with a few degrees of freedom, equations (4.22) and (4.113) form already a probabilistic forecast system capable to deal with nonlinearities and arbitrarily distributed errors and to assimilate observations.

$$p_{\mathbf{x}_0^a}(\mathbf{x}) = p_{\mathbf{x}_i}(\mathbf{x}) \tag{4.115}$$

$$p_{\mathbf{x}_{i+1}^f}(\mathbf{x}) = \int_{I\!R^n} \left[\left|\frac{\partial M_i(\mathbf{x}')}{\partial \mathbf{x}'}\right|^{-1} p_{\mathbf{x}_i^a}(\mathbf{x}')\right]_{x'=M_i^{-1}(\mathbf{y})} p_{\boldsymbol{\eta}_i}(\mathbf{x}-\mathbf{y})d\mathbf{y} \tag{4.116}$$

$$p_{\mathbf{x}_{i+1}^a}(\mathbf{x}) = \frac{p_{\mathbf{y}_{i+1}^o}(\mathbf{y}_{i+1}^o|\mathbf{x})p_{\mathbf{x}_{i+1}^f}(\mathbf{x})}{\int_{I\!R^n} p_{\mathbf{y}_{i+1}^o}(\mathbf{y}_{i+1}^o|\mathbf{x})p_{\mathbf{x}_{i+1}^f}(\mathbf{x})d\mathbf{x}} \tag{4.117}$$

The probability density function is forecast until a set of observations becomes available. The prior pdf is then modified according to the Bayes formula. This *a posteriori* pdf is integrated in time until the next observations instant and so on. This method is not practicable for large ocean system, but it can help to assess the performance of simplified assimilation methods. They are often tested on small systems such as the Lorenz model

(Lorenz, 1963). An exact solution for the nonlinear assimilation problem can be obtained from equations (4.115)-(4.117).

4.3.1 Particle filter and Sequential Importance Resampling

In practice, the probability distribution is not represented as an n-dimensional function. The Monte Carlo method describes the pdf by an ensemble of states. Initially, each ensemble member is equally likely, but when the first observations come into play, it must be realised that some ensemble members are closer to the observations than others. According to the Bayes formula, for each ensemble member *a posteriori* probability $w_i^{(k)}$ can be obtained (van Leeuwen, 2003):

$$w_i^{(k)} = \frac{p_{\mathbf{y}_i^o}(\mathbf{y}_i^o|\mathbf{x}_i^{f(k)})w_{i-1}^{(k)}}{\sum_{k=1}^{N} p_{\mathbf{y}_i^o}(\mathbf{y}_i^o|\mathbf{x}_i^{f(k)})w_{i-1}^{(k)}} \tag{4.118}$$

The weights are not modified during the prediction step and initially all weights are set to $1/N$. This method is called the Particle Filter. It has the merit of handling correctly any nonlinearity. The particle filter is directly drawn from the Bayes formula without any additional hypothesis. But obviously a certain number of ensemble members will have very low weight and so very large ensembles are needed. This method closely resembles to a simple trial and error approach where the model error is guessed. But the important number of uncertain variables (error of initial conditions and dynamical error) makes this method unfeasible for large systems.

It is natural that the members with a negligible weight should be excluded and no longer integrated in time. But a mechanism is needed to create new ensemble members close to the true state of the system in order that the number of ensemble member remains constant.

A possible solution is Sequential Importance Resampling. This technique creates a new ensemble of equally probable members representing the same pdf. In its simplest form, a sequence of N integer numbers between 1 and N are chosen randomly. The probability of the number k is the weight $w^{(k)}$.

The new ensemble will contain as many identical copies of the ensemble member k as its number has been drawn. If the first ensemble member is for example two times more probable than the second member, there will be in average twice as much copies of this first ensemble member than of the second. The identical copies will progressively diverge due to the stochastic noise introduced when integrating the ensemble forward in time. Variants of the importance resampling procedure can be found in Doucet *et al.* (2001) and van Leeuwen (2003).

The Guided Sequential Importance Resampling (GSIR) uses observations that are not yet assimilated, to decimate irrelevant members at an earlier stage. All members which are not consistent with future observations with a several times augmented error bar are rejected.

The ensemble is therefore guided towards the observations. Van Leeuwen (2003) applied the Guided Sequential Importance Resampling to a 5-layer quasi-geostrophic ocean model of the Agulhas Current. He found that the ensemble mean of 32 member ensemble assimilating SSH by the GSIR method produced similar results than a 1024 member ensemble assimilating SSH by the SIR method.

4.3.2 What is the "best" estimation of the state vector?

It has been shown that the pdf of the system state can be represented efficiently either by an ensemble or by a Gaussian pdf with a given mean and a reduced rank covariance. Once the pdf of the system state vector taking into account the observations is known, the filtering problem can be considered as solved.

For some applications a single estimate of the ocean state is needed and we want to provide the "best" and "most accurate" estimation of the ocean state. However, the choice of the optimality criterion is not unique. Different ways to define an optimality criterion exist:

- Maximum likelihood: The best estimate of the ocean is the state vector with the highest probability. If the pdf is known as a function, then one needs to find its global maximum. For an ensemble representation, the determination of the most probable state is a non-trivial task since the pdf estimated by an ensemble is a sum of Dirac's delta functions. A possible solution could be based on the cumulative density function which is a sum of Heaviside step functions. The most probable estimate can then be the maximum gradient of a smoothed version of the cumulative density function. The high dimensionality of this problem makes such an approach inefficient and assimilation methods based on ensemble forecasts do not use in practice this approach.

  But since the exact shape of the pdf is in practice poorly known, the assumption that the errors follow a Gaussian distribution is often used. The application of the maximum likelihood criterion is straight forward. The pdfs of the *a priori* state $p_{\mathbf{x}}(\mathbf{x})$ and of the observations $p_{\mathbf{y}^o}(\mathbf{y}^o|\mathbf{x})$ giving the state $\mathbf{x}$ are written as:

$$p_{\mathbf{x}}(\mathbf{x}) = C_1 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^f)^T \mathbf{P}^{f^{-1}}(\mathbf{x}-\mathbf{x}^f)\right) \tag{4.119}$$

$$p_{\mathbf{y}^o}(\mathbf{y}^o|\mathbf{x}) = C_2 \exp\left(-\frac{1}{2}(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right) \tag{4.120}$$

  where C_1 and C_2 are normalisation constants such that the integrated probability is one. From the Bayes rule, the *a posteriori* pdf is then:

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) = C_3 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^f)^T \mathbf{P}^{f^{-1}}(\mathbf{x}-\mathbf{x}^f) - \frac{1}{2}(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right) \tag{4.121}$$

In the maximum likelihood approach, the "best" state of the system is the most probable one given the observations. The optimal state $\mathbf{x}^a$ minimises the following cost function:

$$J(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}^f\right)^T \mathbf{P}^{f^{-1}} \left(\mathbf{x} - \mathbf{x}^f\right) + \left(\mathbf{H}\mathbf{x} - \mathbf{y}^o\right)^T \mathbf{R}^{-1} \left(\mathbf{H}\mathbf{x} - \mathbf{y}^o\right) \tag{4.122}$$

This last equation is the classical least square error fit with the inverses of the error covariances, which weight the model forecast and the observations.

- The error variance diminishing state: From the *a posteriori* pdf one can also define the state with the lowest error variance as the optimal state.

  Let $\mathbf{x}$ be an estimation of the unknown true state $\mathbf{x}^t$ treated as a random vector following the pdf obtained by the Bayes formula. The expected error variance of $\mathbf{x}$ is given by:

$$\begin{aligned} E\left[\|\mathbf{x} - \mathbf{x}^t\|^2\right] &= E\left[\|\mathbf{x} - E\left[\mathbf{x}^t\right] - \left(\mathbf{x}^t - E\left[\mathbf{x}^t\right]\right)\|^2\right] & (4.123)\\ &= \|\mathbf{x} - E\left[\mathbf{x}^t\right]\|^2 + E\left[\|\mathbf{x}^t - E\left[\mathbf{x}^t\right]\|^2\right] & (4.124) \end{aligned}$$

  The norm $\|\cdot\|$ implies a metric with a physical meaning. A more detailed discussion of the role of this norm is given in chapter 6. This equation shows that the error variance of any estimation is always greater or equal to the last term of equation (4.124). This minimum is obtained if the estimation is the expected mean value of the random vector following the pdf of equation (4.121).

  In the variance diminishing approach, the "best" state is the mean of the *a posteriori* pdf $p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o)$.

$$\mathbf{x}^a = \int_{\mathbb{R}^n} \mathbf{x}\, p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} \tag{4.125}$$

- The unbiased linear combination of model and observations with the least error variance. This estimator is also called the Best Linear Unbiased Estimator (BLUE) introduced by Talagrand (1997). This criterion is also used for the previous estimator. However, the BLUE estimator assumes a linear combination of observations and model forecast, previously it was a consequence. The BLUE estimator is defined independently of the Bayes formula and it makes no assumptions concerning the distribution of the errors.

  All unbiased linear combination of the model forecast $\mathbf{x}^f$ and observations $\mathbf{y}^o$ can be written in the following way:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{4.126}$$

In fact, any other linear combination would be biased. The "best" *a posteriori* state in this approach is the estimator with the lowest total error variance:

$$J(\mathbf{K}) = E\left[(\mathbf{x}^a - \mathbf{x}^t)^T \mathbf{W} (\mathbf{x}^a - \mathbf{x}^t)\right] \tag{4.127}$$

This expression represents the mean squared "length" of the error vector in the state space. The diagonal matrix $\mathbf{W}$ is the error norm in the state space. Even if the *a priori* estimate $\mathbf{x}^f$ and $\mathbf{y}^o$ are non-Gaussian, the cost function $J(\mathbf{K})$ implies only their covariances:

$$J(\mathbf{K}) = \mathrm{tr}\left(\mathbf{W}\left((\mathbf{I} - \mathbf{KH})\,\mathbf{P}^f\,(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T\right)\right) \tag{4.128}$$

The result of this minimisation is independent of the chosen norm.

Although these three approaches make different assumptions regarding the definition of the best estimation, they lead to the same solution under the assumption of Gaussian distributed errors and linear observation operators. The best *a posteriori* state $\mathbf{x}^a$ and its error covariance $\mathbf{P}^a$ are given by:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{4.129}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{4.130}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f \tag{4.131}$$

However, the maximum likelihood and the variance diminishing approach can be in theory easily extended to arbitrarily distributed errors and nonlinear observation operators unlike the BLUE approach. A linear combination of observations and model forecast is difficult to justify in the presence of a nonlinear observation operator.

The differences between the maximum likelihood and variance diminishing approaches are illustrated in figure (4.2) where the extreme case of a bimodal pdf is shown. In this example the errors of *a priori* state and the observations are Gaussian distributed. The square of the system state variable is measured. The error variance of the *a priori* state ($x^f = 0.1$ and $P^f = 1$) is high compared to the error variance of the observations ($y^o = 1$ and $R = 0.1$). The *a posteriori* pdf is bimodal since the values near zero can be excluded by the observations but the sign of the state variable cannot be determined neither by the observations nor by the *a priori* state. The variance diminishing estimate (dot) has a low probability compared to the maximum likelihood estimate (asterisk). But in average an estimation based on the latter approach has a higher error than the estimation based on the error variance criterion.

This discussion may seem somewhat academic since in most cases the observation operator is linear and it is also very difficult to obtain accurate estimations of multimodal

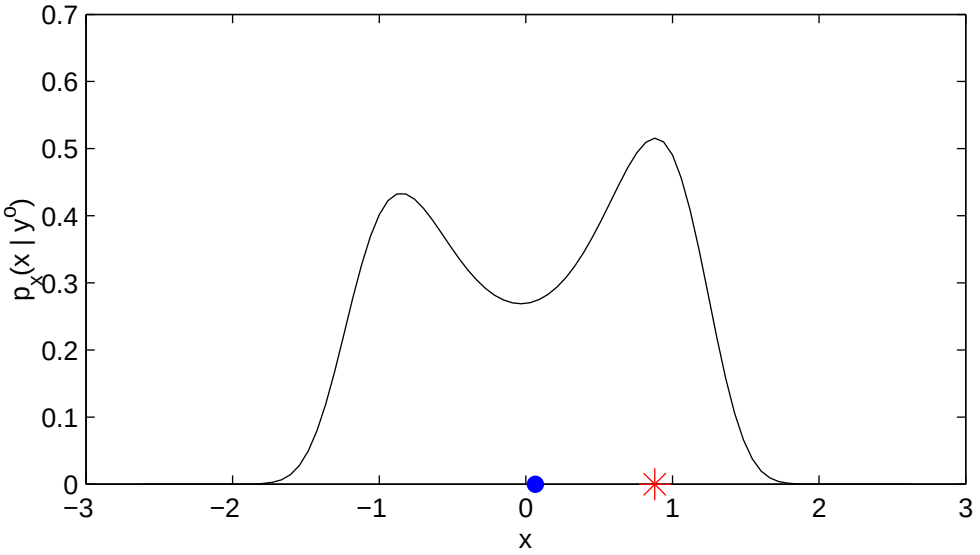

Figure 4.2: The pdf of an *a posteriori* state. The dot denotes the variance diminishing state and the asterisk the maximum likelihood.

pdfs of realistic ocean systems. However, also in the context of purely observation oriented techniques such as the computation of a climatology, the multimodal character of the ocean has been recognised (Levitus, 2002). Should a climatology provide the mean state of the ocean even if it corresponds, in some dynamically active regions, to an unlikely state or should it provide the most likely regime, in which the ocean can be found? In the latter approach the climatology mean RMS error would be higher. In some cases, the application can give an answer to this question. It is preferable to choose a dynamical consistent state of the system for a model initial condition provided by the maximum likelihood approach. On the other hand, the state with the lowest mean RMS error might be more appropriate for determining outliers from a set of observations.

The 4D-Var method and the Bayesian variant of the Kalman Smoother (Evensen and van Leeuwen, 2000) are two assimilation methods which are not explained here. However, they can be treated in the present framework if the vector $\mathbf{x}$ contains the state vector of the system of all observation times. Both approaches give the same *a posteriori* evolution of a linear system. In theory, they can also handle nonlinear dynamical models without approximations, but the 4D-Var method seeks the most likely evolution of the system and the Kalman Smoother the variance diminishing system trajectory.

4.3.3 The Kalman Filter

If the prior estimation of the ocean state and the observation error are normal distributed, then the *a posteriori* pdf of the ocean state is also normal distributed and can completely be described by its mean $\mathbf{x}^a$ and covariance $\mathbf{P}^a$. We also suppose that the observation operator is linear and given by matrix $\mathbf{H}$. According to the Kalman filter update, $\mathbf{x}^a$ and $\mathbf{P}^a$ are given by:

$$\mathbf{x}^a = \mathbf{P}^a \left(\mathbf{P}^{f\,-1}\, \mathbf{x}^f + \mathbf{H}^T \mathbf{R}^{-1}\, \mathbf{y}^o \right) \tag{4.132}$$

$$\mathbf{P}^{a\,-1} = \mathbf{P}^{f\,-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \tag{4.133}$$

The derivation of the Kalman filter analysis in the Bayesian approach can be found in appendix D.

The matrix inversion $\mathbf{P}^{f\,-1}$ in the state space is very difficult to perform. The Sherman-Morrison-Woodbury formula (B.1) allows us to compute the analysis and its covariance by a matrix inversion in observation space. The dimension of this space is generally much less than the dimension of the state space.

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K} \left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f \right) \tag{4.134}$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R} \right)^{-1} \tag{4.135}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}\mathbf{P}^f \tag{4.136}$$

The difference between observations $\mathbf{y}^o$ and its model counterpart $\mathbf{H}\mathbf{x}^f$ is called the innovation vector. This vector introduces into the model the information contained in the observations, in order to improve the ocean state estimation. The innovation vector is "amplified" or "reduced" by the matrix $\mathbf{K}$ called the Kalman gain. The Kalman gain depends on the error covariances of the model forecast and the observations.

From equation (4.132) and (4.133) one can show that the Kalman gain can also be expressed in the following way:

$$\mathbf{K} = \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1} \tag{4.137}$$

This expression is particularly interesting for filtering algorithms computing the *a posteriori* covariance $\mathbf{P}^a$ before the state $\mathbf{x}^a$.

The main assumption of the Kalman filter update is the Gaussian distribution of the errors when the Kalman filter is derived from a Bayesian approach. But even for a non-Gaussian distribution, acceptable results can be expected if the distribution has only one maximum and if it is sufficiently symmetric. If this is not the case, unrealistic or at least very improbable results can be obtained by the analysis, for example negative values of a concentration of a given biological species or an unstable water column. An *ad hoc* adjustment is then necessary (Brankart *et al.*, 2003). Another approach rendering these unphysical states impossible is the application of a so-called anamorphosis transform (Bertino *et al.*, 2002, 2003) which transforms these problematic variables by a nonlinear function into Gaussian distributed variables. The Kalman filter is then applied to these latter variables before they are transformed back into the physical variables.

Another hypothesis is that the model results can be related linearly to the observations. In oceanography, the model variables are, for the most part, directly observed and this

hypothesis is not very restrictive.

Here we have derived the analysis of the Kalman filter in order to give the best estimate provided the model forecast and the observations and their respective error covariances at the given time t_i. But the Kalman filter is more than this. In fact, the sequential algorithm formed by equation (4.32)-(4.35) and (4.134)-(4.136) gives the best estimate of the ocean taking *all* past observation into account (Gelb, 1974). Its solution after assimilating the last observation is thus equal to the result of the non-sequential methods (Kalman Smoother, 4D-Var and representer method), which use all observations.

The Kalman filter depends on the error covariance matrix of the model forecast $\mathbf{P}^f$. However, this matrix can never be formed explicitly for large system. Several alternative ways to express the model uncertainties, and implicitly the error covariance, have been proposed in the previous sections, namely the ensemble representation and the eigenvector decomposition. The Kalman filter has to be adapted for each of these formulations. The *a posteriori* state should be computed directly from the ensemble or the eigenvector decomposition of $\mathbf{P}^f$ and the *a posteriori* error covariance $\mathbf{P}^a$ should be expressed as an ensemble or in terms of an eigenvector decomposition rather than a huge covariance matrix.

4.3.4 Stochastic Kalman Filter Analysis

The way the uncertainties of the observations is taken into account is not unique. The stochastic Kalman filter analyses are based on an ensemble of observations. In the other hand, deterministic Kalman filter analyses do not need perturbed observations.

The Ensemble Kalman filter analysis

Here we describe the traditional ensemble Kalman filter as proposed by Evensen (1994). Several variants and modifications of the analysis scheme were proposed in the literature. Some of them are discussed later on.

An ensemble contains the information of all statistical moments of the pdf. The ensemble Kalman Filter (EnKF) analysis takes only the ensemble mean $\overline{\mathbf{x}^f}$ and the ensemble covariance $\mathbf{P}_e^f$ into account:

$$\overline{\mathbf{x}^f} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}^{f(k)} \tag{4.138}$$

$$\mathbf{P}_e^f = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{x}^{f(k)} - \overline{\mathbf{x}^f})(\mathbf{x}^{f(k)} - \overline{\mathbf{x}^f})^T \tag{4.139}$$

An ensemble of perturbed observations must also be created in order to take into account their uncertainties (Burgers *et al.*, 1998). The perturbation must follow the error statistics given by the Gaussian pdf (4.114).

$$\mathbf{y}^{o(k)} = \mathbf{y}^{o} + \boldsymbol{\varepsilon}^{(k)} \tag{4.140}$$

The ensemble covariance $\mathbf{R}_e$ of the observation perturbations is used as an approximation of the error covariance matrix $\mathbf{R}$.

$$\mathbf{R}_e = \frac{1}{N-1} \sum_{k=1}^{N} \boldsymbol{\varepsilon}^{(k)} \boldsymbol{\varepsilon}^{(k)T} \tag{4.141}$$

Each ensemble member is updated by the following analysis equation:

$$\mathbf{x}^{a(k)} = \mathbf{x}^{f(k)} + \mathbf{K}\left(\mathbf{y}^{o(k)} - \mathbf{H}\mathbf{x}^{f(k)}\right) \tag{4.142}$$

where the Kalman gain has been computed by the ensemble covariance of $\mathbf{x}^{f(k)}$ and the perturbed observations $\mathbf{y}^{o(k)}$ (Evensen, 2003).

$$\mathbf{K} = \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T + \mathbf{R}_e\right)^{-1} \tag{4.143}$$

The use of $\mathbf{R}_e$ instead of the *a priori* more accurate $\mathbf{R}$ for the Kalman gain is not without consequences.

The perturbations of the observations are necessary for the ensemble to have in average the covariance given by equation (4.136). Otherwise the spread of the ensemble will be underestimated (Burgers *et al.*, 1998). The need of degrading the observations by random noise is however an undesirable characteristic of the EnKF and several variants have been proposed to avoid this.

The covariance matrix $\mathbf{P}_e^f$ and $\mathbf{R}_e$ must never be formed explicitly. By definition, these covariances can be written as the following products according to equation (4.84):

$$\mathbf{P}_e^f = \mathbf{S}^f \mathbf{S}^{fT} \tag{4.144}$$
$$\mathbf{R} = \mathbf{E}\mathbf{E}^T \tag{4.145}$$

Since the covariance $\mathbf{R}_e$ is the ensemble covariance of the perturbations, the matrix to inverse in equation (4.143) is the sum of two matrices with ranks of $N-1$ at most. Their sum will therefore have a maximum rank of $2N-2$. If the number of observations is higher than $2N-2$ the matrix to be inverted $\mathbf{H}\mathbf{P}_e^f\mathbf{H} + \mathbf{R}_e$ is singular. Generally the ensemble size is of the order of 10 to 100 and the number of scalar observations in oceanography is often of the order of 10^4 to 10^5. So what went wrong ? The reduced rank approximation has lead us to a situation where the model and the observational constrains have become incompatible. The matrix to invert is nothing more than the covariance matrix of the innovation vector:

$$E\left[\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right)\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right)^T\right] = \mathbf{H}\mathbf{S}^f(\mathbf{H}\mathbf{S}^f)^T + \mathbf{E}\mathbf{E}^T \quad (4.146)$$

$$= \begin{pmatrix} \mathbf{H}\mathbf{S}^f & \mathbf{E} \end{pmatrix} \begin{pmatrix} (\mathbf{H}\mathbf{S}^f)^T \\ \mathbf{E}^T \end{pmatrix} \quad (4.147)$$

where we have already used the low rank approximation of the covariance matrices. This expression shows that we have made implicitly the assumption that the innovation vector lies in a $2N-2$ dimensional subspace. If this is really the case, then the singular matrix inversion can be avoided by an inversion in this $2N-2$ dimensional space. But for independent observations this is generally not true.

Traditionally, the pseudoinverse[4] of the innovation covariance matrix is computed by performing the following eigenvalue value decomposition:

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R} = \mathbf{H}\mathbf{S}^f\mathbf{H}\mathbf{S}^{fT} + \mathbf{E}\mathbf{E}^T = \mathbf{Z}\mathbf{\Gamma}\mathbf{Z}^T \quad (4.148)$$

where $\mathbf{Z}$ is an orthogonal matrix and $\mathbf{\Gamma}$ is diagonal. Once this eigenvector decomposition is performed, the pseudoinverse of this matrix is simply given by:

$$\left(\mathbf{H}\mathbf{S}^f\mathbf{H}\mathbf{S}^{fT} + \mathbf{E}\mathbf{E}^T\right)^+ = \mathbf{Z}\mathbf{\Gamma}^+\mathbf{Z}^T \quad (4.149)$$

If the rank of the innovation covariance matrix is not full, the role of the pseudoinverse is to filter out the components of the innovations outside the space defined by the columns of $\mathbf{HS}$ and $\mathbf{E}$. This makes the filtered innovation consistent with the reduced order covariance matrix (4.147)

Matrices $\mathbf{Z}$ and $\mathbf{\Gamma}$ in (4.149) are generally computed by a singular value decomposition. Its cost is of the order of m^3 where m is the number of observations. This step might be prohibitive for large observation vectors (Nerger *et al.*, 2004a). From equation (4.147) it can be seen that the singular value decomposition can also be done in a $2N-2$ dimensional space using the matrix properties derived in appendix C. This optimisation seems not to be recognised by all authors.

The EnKF is extensively discussed in the literature and several problems were pointed out:

[4] The pseudoinverse $\mathbf{A}^+$ of a $m \times n$ real matrix $\mathbf{A}$ is the unique matrix which satisfies the following criteria:

- $(\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+$ (That is, $(\mathbf{A}\mathbf{A}^+)$ is symmetric).
- $(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A}$
- $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
- $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$

If $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the singular value decomposition of $\mathbf{A}$, then $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^T$. For a diagonal matrix such as $\mathbf{\Sigma}$, we get the pseudoinverse by inverting each non-zero element on the diagonal.

- The problems of "in breeding" or "filter divergence" were identified by Houtekamer and Mitchell (1998) and analysed by van Leeuwen (1999) and by Houtekamer and Mitchell (1999). Even if the ensemble based covariances $\mathbf{P}_e^f$ and $\mathbf{R}_e$ are unbiased estimations of the true error covariances, there is no guarantee that after the analysis, which is a nonlinear function of these covariances, the *a posteriori* ensemble will be unbiased (Whitaker and Hamill, 2002). Van Leeuwen (1999) has showed that the sampling error of $\mathbf{P}_e^f$ leads to a systematic underestimation of the *a posteriori* error covariance. The sampling error of the observation error covariance tends also to reduce the variance of the *a posteriori* ensemble.

  The sampling error of the observation error covariance matrix can be avoided by using a deterministic analysis scheme (see following section). Whitaker and Hamill (2002) show the beneficial impact of the Ensemble Square Root Filter on the variance of the *a posteriori* ensemble. Filter divergence might still occur with this analysis method due to the sampling error of the *a priori* ensemble.

  A very simple method applied by Anderson and Anderson (1999) consists in increasing the ensemble spread by an empirical determined factor. The deviation from the ensemble mean of each member is augmented by this inflation factor. The forgetting factor of Pham (2001) does the same, but it is used in a different context. Pham (2001) introduced the forgetting factor as a simple parameterisation of the model error.

  The double ensemble Kalman filter proposed by Houtekamer and Mitchell (1998) uses two parallel ensemble Kalman filter. But the Kalman gain for the analysis is obtained from the covariances based on the other ensemble. Van Leeuwen (1999) has discussed how this method compensates biases associated with the nonlinearities of the Kalman gain. The sampling error of the *a priori* ensemble tends now to overestimate the ensemble spread.

- Kepert (2004) has showed that the ensemble Kalman filter algorithm described by equations (4.140), (4.142), and (4.143) results in an *a posteriori* ensemble with a rank lower than $N-1$ if more than two measurement perturbations have components in the space perpendicular to the space formed by the column of $\mathbf{HS}^f$. For random perturbations this is generally the case.

  This can be interpreted intuitively since with the ensemble error covariance $\mathbf{R}_e$, one assumes that the observational error lies only in a subspace of $N-1$ dimensions at most. Along the directions outside this space, the observations are assumed to be exact. The assimilation of an observation vector without an error along a given direction will reduce the variance of the ensemble in this direction to zero after the analysis. In particular, if the model error direction at observation location $\mathbf{HS}^f$ lies outside the space formed by the columns of $\mathbf{E}$, this error mode will no longer be present in the ensemble after the analysis.

Kepert (2004) suggested taking a full rank measurement error covariance to cope with this rank loss. Evensen (2004) replied that this rank issue could be avoided by sampling the measurement perturbation inside the space of $\mathbf{HS}^f$. However, general error observation covariance matrix cannot be decomposed into a covariance matrix lying inside of the space defined by $\mathbf{HS}^f$ and into the perpendicular space. The new algorithm needs therefore a new hypothesis whose effect is unclear.

In any circumstances, an ensemble based representation of $\mathbf{R}$ is inaccurate and large sampling errors would occur for weakly or not correlated observations. The use of the full $\mathbf{R}$ instead of its ensemble representation solves the problem of sampling error associated to the observations and also reduces "the filter divergence" problem. The only advantage of the ensemble based representation of $\mathbf{R}_e$ is the fact that it is very efficient if the singular vector decomposition of the innovation covariance matrix is done in the $N-2$ dimensional space formed by the perturbations.

The stochastic ESSE filter

Lermusiaux and Robinson (1999) propose two analysis schemes of the Error Subspace Statistical Estimation (ESSE) approach. Here we describe the variant using perturbed observations for the ensemble update (Lermusiaux, 1997; Lermusiaux and Robinson, 1999). From equation (4.134), it can be shown that the *a posteriori* error covariance is given by:

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}^f\,(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \tag{4.150}$$

The first term takes into account the uncertainty due to the *a priori* state and the second is due to the uncertainties of the observations. If the *a priori* ensemble perturbations $\mathbf{S}^f$ are updated in the following way,

$$\mathbf{S}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{S}^f \tag{4.151}$$

the resulting ensemble would have too low spread since the uncertainties due to the observations are neglected. This is the reason why Lermusiaux and Robinson (1999) introduce an additional stochastic term in the ensemble perturbation update as in the Ensemble Kalman filter:

$$\mathbf{S}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{S}^f + \mathbf{KE} \tag{4.152}$$

where the columns of $\mathbf{E}$ are again random perturbations drawn from a Gaussian pdf with zero mean and covariance given by $\mathbf{R}$ multiplied by $\frac{1}{\sqrt{N-1}}$ as in the EnKF. A similar update equation can be obtained by subtracting the ensemble mean from (4.142). But here the full observation error covariance is used for the Kalman gain $\mathbf{K}$. The ESSE scheme does not work directly with the ensemble perturbation but with the singular vector decomposition of these perturbations.

$$\mathbf{S}^{f,a} = \mathbf{U}^{f,a}\mathbf{\Sigma}^{f,a}\mathbf{V}^{f,a^T} \tag{4.153}$$

From (4.152) the analysis update of the right and left singular vectors and for the singular values can be derived. The *a priori* state is obtained by the standard Kalman filer analysis (4.134)-(4.136) without perturbed observations.

4.3.5 Deterministic Kalman filter analysis

If unperturbed observations are used during the analysis step (4.142), the resulting ensemble would have a too low variance, compared to the Kalman filter in the case of Gaussian statistics (Burgers *et al.*, 1998). But the need of perturbed observations is not a desirable feature since the analysis is not fed with the true observations but only with observations degraded by an artificial noise. For the assimilation of perturbed observations, generally larger ensemble sizes are needed (Nerger *et al.*, 2004a,b). The next class of Kalman filter analysis are deterministic in the sense that no Monte Carlo approach for the observation uncertainty is needed.

The SESAM and the Ensemble Transform Kalman filter analysis

Practical methods for forecasting the model error use an ensemble approach or a leading error mode representation of the error space. In most methods presented here the forecast error covariance can be expressed in terms of its square root matrices (Tippett *et al.*, 2003):

$$\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{fT} \tag{4.154}$$

The *a posteriori* error covariance matrix $\mathbf{P}^a$ can also be expressed in terms of square roots:

$$\begin{aligned} \mathbf{P}^a &= \mathbf{S}^a \mathbf{S}^{aT} && (4.155)\\ &= \mathbf{S}^f \left(\mathbf{I} - (\mathbf{H}\mathbf{S}^f)^T \left((\mathbf{H}\mathbf{S}^f)^T (\mathbf{H}\mathbf{S}^f) + \mathbf{R} \right)^{-1} \mathbf{H}\mathbf{S}^f \right) \mathbf{S}^{fT} && (4.156) \end{aligned}$$

The square root of the inner matrix can be computed by an eigenvector-decomposition:

$$(\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{S}^f) = \hat{\mathbf{U}}^T \hat{\mathbf{\Lambda}} \hat{\mathbf{U}} \tag{4.157}$$

or by a singular value decomposition of $\mathbf{R}^{-1/2}(\mathbf{H}\mathbf{S}^f)$. From equation (4.157), the Kalman gain can then be written as:

$$\begin{aligned} \mathbf{K} &= \mathbf{S}^f (\mathbf{H}\mathbf{S}^f)^T \left((\mathbf{H}\mathbf{S}^f)(\mathbf{H}\mathbf{S}^f)^T + \mathbf{R} \right)^{-1} && (4.158)\\ &= \mathbf{S}^f \left(\mathbf{I} + (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{S}^f) \right)^{-1} (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (4.159)\\ &= \mathbf{S}^f \hat{\mathbf{U}} (\mathbf{I} + \hat{\mathbf{\Lambda}})^{-1} \hat{\mathbf{U}}^T (\mathbf{H}\mathbf{S}^f)^T \mathbf{R}^{-1} && (4.160) \end{aligned}$$

For equation (4.159), the Sherman-Morrison-Woodbury formula (B.2) was used to transform the inverse in the observation space to an inverse in the error subspace. If we insert the Kalman gain (4.160) in equation (4.134), we obtain the expression of the *a posteriori* state:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{S}^f\hat{\mathbf{U}}(\mathbf{I}+\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{U}}^T(\mathbf{H}\mathbf{S}^f)^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{4.161}$$

This equation shows that the correction introduced by the analysis is a linear combination of the columns of $\mathbf{S}^f$. This property is common to all assimilation methods based on a reduced rank error covariance of the model state. This approximation assumes that the model is perfect along the directions not included in the error space defined by the error covariance. It is therefore not surprising that the correction term can only lie inside the space formed by the columns of $\mathbf{S}^f$.

There is an interesting link between the eigenvalues of $\hat{\mathbf{\Lambda}}$ and the controllability of a system. The magnitude of the eigenvalue in $\hat{\mathbf{\Lambda}}$ indicates which model error can be reduced with the given observations and which can not. For instance, if the eigenvector $\mathbf{u}_0$ of the rhs of (4.157) has a zero eigenvalue then:

$$(\mathbf{H}\mathbf{S}^f\mathbf{u}_0)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}^f\mathbf{u}_0) = 0 \tag{4.162}$$

Since the norm of the vector $\mathbf{R}^{-1/2}\mathbf{H}\mathbf{S}^f\mathbf{u}_0$ is zero, the vector itself must be zero:

$$\mathbf{H}\mathbf{S}^f\mathbf{u}_0 = 0 \tag{4.163}$$

The model error $\mathbf{S}^f\mathbf{u}_0$ has no signature in the observed part of the model state vector. It is therefore not possible for the observations to give any information reducing this error. The model is not controllable along this direction in the error space by the given observations. One can also show that the analysis increment $\mathbf{x}^f - \mathbf{x}^a$ has no contribution along the unobservable error direction $\mathbf{S}^f\mathbf{u}_0$. Although the error space describes only the model and it is independent of the observations, it should for efficiency reasons contain only error modes that can be constrained by the observations. The magnitude of the eigenvalues $\hat{\mathbf{\Lambda}}$ can help to identify uncontrollable error modes. On the other hand, it can also help to define the observation strategy (the operator $\mathbf{H}$) to reduce the error along a given error mode corresponding to a specific event, *e.g.* an eddy detaching form a current.

Finally the *a posteriori* error covariance can also easily be obtained by

$$\mathbf{P}^a = \mathbf{S}^a\mathbf{S}^{aT} \tag{4.164}$$

$$= \mathbf{S}^f\hat{\mathbf{U}}(\mathbf{I}+\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{U}}^T\mathbf{S}^{fT} \tag{4.165}$$

The error subspace after the analysis is then given by

$$\mathbf{S}^a = \mathbf{S}^f\hat{\mathbf{U}}(\mathbf{I}+\hat{\mathbf{\Lambda}})^{-1/2} \tag{4.166}$$

In this algorithm the error covariance matrices $\mathbf{P}^f$ and $\mathbf{P}^a$ must never be formed explicitly. All operations involve only the square root matrices $\mathbf{S}^f$ and $\mathbf{S}^a$.

This method is also used for the SEEK filter as it is implemented in the SESAM (An integrated System of Sequential Assimilation Modules) package (Testut *et al.*, 2002).

This approach is also identical to the analysis step of Ensemble Transform Kalman Filter (Bishop *et al.*, 2001). For the assimilation experiments realised in the present work, the analysis is based on equations (4.161) and (4.166).

It is important to note that the matrix $\mathbf{S}^a$ is not uniquely defined. The application of any $m \times m$ orthogonal matrix to the right of (4.166) would yield another *a posteriori* error space describing the same error covariance.

$$\mathbf{S}^a = \mathbf{S}^f \hat{\mathbf{U}} (\mathbf{I} + \hat{\mathbf{\Lambda}})^{-1/2} \mathbf{U}' \tag{4.167}$$

The model error forecast based on the tangent linear model is also invariant with respect to an m by m orthogonal matrix $\mathbf{U}'$.

The need of a tangent linear model for the error modes is avoided by transforming $\mathbf{S}^a$ into an ensemble (*e.g.* Pham, 2001). But for nonlinear systems the forecast ensemble mean and covariance will depend on $\mathbf{U}'$.

Numerous filtering algorithms based on the decomposition (4.154) exist and they are reviewed by Tippett *et al.* (2003). They are all formally equivalent and give the same $\mathbf{S}^a$ up to an arbitrary rotation matrix. For nonlinear systems they can produce in practice different results, but *a priori* it is not possible to determine which method produces the best results.

While the error covariance is not modified by the application of $\mathbf{U}'$, higher moments of the ensemble are modified. These higher moments reflect *e.g.* that unstable water columns are improbable and negative concentrations even impossible. Ott *et al.* (2004) have proposed an interesting method for determining an $m \times m$ orthogonal matrix such that the *a posteriori* ensemble is as close as possible to the initial ensemble and still satisfying equation (4.155).

The Ensemble Adjustment Kalman filter

Anderson (2001) has proposed the Ensemble Adjustment Kalman filter in order to use a Monte Carlo method for the ensemble forecast but a deterministic analysis scheme for the update. The *a priori* error space $\mathbf{S}^f$ is adjusted by an operator in such way that $\mathbf{S}^a = \mathbf{A}\mathbf{S}^f$ has the right covariance. This operator is obtained by rotating the error space into the space formed by the eigenvector of $\mathbf{P}^f$. This eigenvectors can be obtained directly by a singular value decomposition of the matrix $\mathbf{S}^f$.

According to Anderson (2001), the *a posteriori* state $\mathbf{x}^a$ is obtained by equation (4.132). The eigenvector decomposition of $\mathbf{P}^f$ is then used to compute its inverse. However the inverse should be treated as pseudo-inverse since the ensemble size is much less than the state space dimension. Alternatively, the *a posteriori* state can directly be computed by (4.137). No pseudo-inverse is needed in this case.

Reduced Rank Square root Kalman filter (RRSQRT filter)

If the observations are not correlated, then each observation can be assimilated serially. The square root filter, introduced in oceanography by Verlaan and Heemink (1997), computes the Kalman gain of a given observation directly by equation (4.135). The inversion involves only scalar quantities because it is performed in the observation space. The *a posteriori* error space can be obtained by:

$$\mathbf{S}^a = \mathbf{S}^f \left(\mathbf{I} - \beta(\mathbf{H}\mathbf{S}^f)^T(\mathbf{H}\mathbf{S}^f)\right) \tag{4.168}$$

where the scalar β is chosen in order that the *a posteriori* error covariance $\mathbf{S}^a\mathbf{S}^{aT}$ has the covariance predicted by the Kalman filter theory. Therefore, β has the following value:

$$\beta = (\mathbf{H}\mathbf{S}^f)^T(\mathbf{H}\mathbf{S}^f) + \mathbf{R} + \sqrt{((\mathbf{H}\mathbf{S}^f)^T(\mathbf{H}\mathbf{S}^f) + \mathbf{R})\,\mathbf{R}} \tag{4.169}$$

This expression involves essentially only scalar quantities. For correlated observations, the observations have to be transformed first into a space where the observation error covariance matrix is diagonal. The multiplication of the observation vector and the observation operator by $\mathbf{R}^{-1/2}$ transforms the error covariance matrix into the identity matrix. Each of those new observations can then be assimilated sequentially.

Whitaker and Hamill (2002) proposed a similar algorithm for assimilating uncorrelated observations serially called the Ensemble Square root Kalman filter. Tippett *et al.* (2003) showed that this method and the scheme derived by Whitaker and Hamill (2002) yield the same *a posteriori* error covariance than the update of equation (4.168).

In this method, it is particularly easy to assess the impact of each observation on the *a posteriori* state and covariance. Hamill and Snyder (2000) used this property to limit the correction of an observation to a given influence circle. In this approach unrealistic long range correlation introduced by the reduced rank approximation are suppressed.

The SEEK filter analysis

Instead of expressing the covariance matrix in terms of its square root matrices, the covariance can also be projected into a subspace:

$$\mathbf{P}^f = \mathbf{L}\tilde{\mathbf{P}}^f\mathbf{L}^T \tag{4.170}$$

The subspace is spanned by the columns of $\mathbf{L}$. A full rank covariance matrix cannot be projected into a low dimension subspace without losing small scale structures in the initial covariance matrix. However, it can be shown that the subspace defined by the dominant eigenvectors of $\mathbf{P}^f$ is the closest approximation, given a fixed subspace dimension of $\mathbf{P}^f$ in terms of the spectral and Frobenius norm (Horn and Johnson, 1985, 1991).

There is however a clear link between the square root decomposition (4.154) and the eigenvector decomposition (4.170). The crucial point of this factorisation is the

assumption that the covariance matrix can be approximated by a low rank matrix.

An important aspect here is that the *a posteriori* error covariance matrix $\mathbf{P}^a$ can be expressed in the same subspace without any other approximation (Pham *et al.*, 1998):

$$\mathbf{P}^a = \mathbf{L}\tilde{\mathbf{P}}^a\mathbf{L}^T \tag{4.171}$$

The analysis step can therefore be computed in the subspace defined by $\mathbf{L}$ and then transformed back into the physical state space.

The base of the initial error space is generally orthonormal. But the columns of $\mathbf{L}$ are generally no longer perpendicular after the application of the tangent linear model to the vectors of this basis. A decomposition into the eigenvectors of $\mathbf{P}^f$ is generally not needed for the analysis. The space spanned by the $\mathbf{L}$ should contain the most dominant eigenvectors of $\mathbf{P}^f$, but they are not necessary equal to them.

The analysis of the SEEK filter, as proposed by Pham *et al.* (1998), used traditionally an eigenvector decomposition (4.170). The *a posteriori* error covariance is computed by projecting equation (4.133) onto the subspace defined by the columns of $\mathbf{L}$.

$$\tilde{\mathbf{P}}^{a\,-1} = \tilde{\mathbf{P}}^{f\,-1} + (\mathbf{HL})^T\mathbf{R}^{-1}(\mathbf{HL}) \tag{4.172}$$

Except for the inverse of the observation error covariance, the inverses have to be computed in the error subspace. The resulting *a posteriori* error covariance expressed in the error subspace $\tilde{\mathbf{P}}^a$ and the columns of $\mathbf{L}$ are not necessary the eigenvalues and eigenvectors of the *a posteriori* error covariance $\mathbf{P}^a$.

The *a posteriori* state can be obtained directly from (4.134) by using the Kalman gain given by equation (4.137):

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{L}\tilde{\mathbf{P}}^a(\mathbf{HL})^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right) \tag{4.173}$$

The SEIK (The Singular Evolutive Interpolated Kalman) filter (Pham, 2001) relies on the same analysis scheme but it is based on an ensemble forecast in order to predict the error covariance due to the uncertain initial condition. The model error covariance $\mathbf{Q}_i$ is projected onto this error space and introduced as in the SEEK filter. Numerous variants of the SEIK filter use the same analysis scheme but different error forecast methods have been proposed (Hoteit *et al.*, 2002).

The ESSE analysis scheme

Besides the stochastic ESSE analysis scheme, there exists also a deterministic variant of the ESSE analysis (Lermusiaux, 1997; Lermusiaux and Robinson, 1999).

Starting from the eigenvector decomposition (4.170), the Kalman gain has the following form:

$$\mathbf{K} = \mathbf{L}\tilde{\mathbf{P}}^f(\mathbf{HL})^T \left[(\mathbf{HL})\tilde{\mathbf{P}}^f(\mathbf{HL})^T + \mathbf{R}\right]^{-1} \tag{4.174}$$

As mentioned before, if the Kalman gain is applied to an innovation vector $\mathbf{y}^o - \mathbf{H}\mathbf{x}^f$, the resulting correction lies inside the error subspace. The amplitudes of these corrections along the error subspace base can be obtained by applying the matrix $\tilde{\mathbf{K}}$ to the innovation vector if $\tilde{\mathbf{K}}$ is defined by:

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f(\mathbf{HL})^T \left[(\mathbf{HL})\tilde{\mathbf{P}}^f(\mathbf{HL})^T + \mathbf{R}\right]^{-1} \tag{4.175}$$

$$\mathbf{K} = \mathbf{L}\tilde{\mathbf{K}} \tag{4.176}$$

$\tilde{\mathbf{K}}$ can be viewed therefore as the Kalman gain expressed in the error subspace, which also allows the computation of the *a posteriori* error covariance. The updated equation of the error covariance (4.136) is projected into the error subspace formed by the columns of $\mathbf{L}$.

$$\tilde{\mathbf{P}}^a = \tilde{\mathbf{P}}^f - \tilde{\mathbf{K}}(\mathbf{HL})\tilde{\mathbf{P}}^f \tag{4.177}$$

As for the SEEK filter, the resulting *a posteriori* error covariance expressed in the error subspace $\tilde{\mathbf{P}}^a$ is generally not diagonal. By an eigenvector decomposition of $\tilde{\mathbf{P}}^a$, one can obtain the necessary rotation in the error subspace such that $\mathbf{P}^a$ is decomposed into its eigenvectors and eigenvalues. The advantage of expressing $\mathbf{P}^a$ this way, is that error directions with too small variance can be neglected. The ESSE scheme can also increase the error subspace dimension by means of a procedure called the "adaptive learning". Based on the observation residuals, the missing directions in the error space are "guessed" by some empirical assumptions (Lermusiaux and Robinson, 1999)

4.4 Summary of reduced rank Kalman filters

	Model error covariance			Observations	Analysis
	Initial condition	Forcings	Error forecast		
SEEK (Pham *et al.*, 1998)	Eigenvectors of $\mathbf{P}_i$	Projection of $\mathbf{Q}_i$ onto the actual error space	Tangent linear model	Unperturbed observations	no further approximations
SEIK (Pham, 2001)	Eigenvectors of $\mathbf{P}^i$	Projection of $\mathbf{Q}_i$ onto the actual error space	Error vector is transformed into an ensemble. The nonlinear model is applied to each ensemble member	Unperturbed observations	no further approximations
RRSQRT (Verlaan and Heemink, 1997)	Eigenvectors of $\mathbf{P}^i$	Eigenvectors of $\mathbf{Q}_i$	The tangent linear model is applied to the dominant eigenvectors of $\mathbf{P}^i$. The error space is extended by the eigenvectors of $\mathbf{Q}_i$. The dimension of the error space dimension is limited by the leading eigenvectors of the merged space.	Unperturbed observations	Observations are assimilated serially
EnKF (Evensen, 1994)	Random perturbation drawn from $N(0, \mathbf{P}^i)$	Random perturbation drawn from $N(0, \mathbf{Q}_i)$	The nonlinear model starts for each ensemble member from the perturbed initial conditions and is driven by the stochastic forcings.	Random perturbation drawn from $N(0, \mathbf{R}_i)$ are added to the observations	The Kalman gain is computed from the model forecast and observation ensemble error covariances. Each member in the ensemble is updated separately by the analysis.
EnKF variant proposed by Kepert (2004)	As EnKF	As EnKF	As EnKF	As EnKF	The observation error covariance is not approximated by an ensemble. The original error covariance $\mathbf{R}$ is used for the analysis.

	Model error covariance			Observations	Analysis
	Initial condition	Forecings	Error forecast		
Deterministic ESSE (Lermusiaux and Robinson, 1999)	Eigenvectors of $\mathbf{P}_i$	Random linear combinations of the dominant eigenvectors of $\mathbf{Q}_i$.	An ensemble is constructed by random linear combinations of the dominant eigenvector of the model error covariance. Each ensemble member is forecast by the nonlinear model.	Unperturbed observations	The analysis is performed in the error subspace.
Stochastic ESSE (Lermusiaux and Robinson, 1999)	Eigenvectors of $\mathbf{P}_i$	Random linear combinations of the dominant eigenvectors of $\mathbf{Q}_i$.	As the deterministic ESSE. Alternatively the ensemble is constructed directly from the ensemble of perturbations of the previous analysis.	Perturbed observations	The *a posteriori* state vector is determined by assimilation of the unperturbed observations. An ensemble of perturbation representing $\mathbf{P}_i^a$ is obtained by the perturbed observations.
ETKS (Bishop *et al.*, 2001)	As EnKF	As EnKF	As EnKF	Unperturbed observations	The ensemble mean and the ensemble perturbations are updated separately. The square root of the model error covariance is estimated from the ensemble run. The analysis is computed in the error space.
EAKS (Anderson, 2001)	As EnKF	As EnKF	As EnKF	Unperturbed observations	The ensemble mean and the ensemble perturbations are updated separately. The former is updated directly from the Kalman filter analysis. An operator is applied to the latter such that the *a posteriori* ensemble covariance is in accordance with the Kalman filter theory.

4.5 Description of the assimilation scheme implemented to the system of nested models

An assimilation strategy based on the combination of the SEEK filter, RRSQRT filter, ESSE and EnKF has been adopted. Each of these methods presents advantages and drawbacks. But the different steps of these algorithms can be exchanged since all are based on the Kalman filter and on the reduced rank approximation of the error covariance matrix $\mathbf{P}^f$.

4.5.1 Estimation of error covariance of the model forecast

The error covariance $\mathbf{P}^f$ is estimated as in the EnKF and the ESSE. The specification of the model error is particularly straightforward for these methods. Once the variables causing the main errors are identified, an ensemble of these variables reflecting the uncertainties attributed to them is created. The ensemble of model forecast obtained with the perturbed variables represents thus the error of the model forecast. In this stochastic forecast, the full nonlinear model is used.

4.5.2 Dominant error modes

The error space is obtained from an ensemble simulation and contains thus redundant error modes. Like in the ESSE, the error space is reduced to the leading EOFs. The EOF analysis provides a set of orthonormal basis functions of the error space ordered by their importance *i.e.* , the variance of the ensemble along the basis function.

4.5.3 Analysis

The analysis schemes of the SEEK and the RRSQRT filters are particularly efficient since they do not require observation perturbations. They provide the exact solution of the Kalman filter analysis to which the EnKF analysis converges for an infinitely large ensemble size. In particular, when the observation error covariance is diagonal, the analysis requires only a number of operations linear in m.

The representation of the error covariance in terms of square root matrices is borrowed from the RRSQRT filter. The error modes have the same physical units as the state vector. The square root matrices are thus easier to interpret since their magnitude represents "typical" errors. However, as in the SEEK filter, all observations are assimilated simultaneously.

4.5.4 Simplification of the error forecast

From a computational point of view, the most expensive part of the Kalman filter assimilation schemes is the forecast of the error covariance, since it implies r or $r + 1$ model

integrations, where r is the rank of the error covariance matrix. A single two-month integration of the Ligurian Sea nesting system takes about 2.5 days CPU time[5]. Even small ensembles of $\mathcal{O}(10)$ are not feasible for the complete nesting. Thus the error space describing the forecast error is only computed once and finally it is maintained constant over the different assimilation cycles. More details on generation of the ensemble are given in section 6.1.

[5]The complete nesting system is distributed on six 400 MHz processors of a SGI Origin 3800.

Chapter 5

Programming aspects

In this chapter, some programming aspects of the developed assimilation package and the model are given. It should give an overview of their capabilities and limits. This chapter is not necessary of the understanding for the remaining work. Readers not interested in the computing aspects can skip this chapter. But in data assimilation, programming aspects can be important for practical implementation since the computational costs of the assimilation methods are often a limiting factor. These limitations must be taken into account in the development of an assimilation package.
aspect

Furthermore, there is a growing demand for relocatable model implementations. In many situations such as the dispersion of oil or another pollutant, marine construction or questions related to marine transport and safety, models need to be rapidly implemented. The Harvard Ocean Prediction System is an example of a relocatable model implementation (Robinson and Sellschopp, 2002). This kind of applications has also consequences for the development of a modelling and data assimilation system.

5.1 The GHER model

5.1.1 Relocatable model implementation

The model implementation consists in a series of steps with little or no intervention of the modeller. It is therefore possible to automate an important part of the model implementation and thus to accelerate the model set-up. The implementation procedure relies on a database constituted by the MEDAR/Medatlas and MODB Climatology (Rixen *et al.*, 2001b; Brasseur *et al.*, 1996), a 1/60° bathymetry of the Mediterranean Sea and a 1/4° typical winter initial conditions obtained from a 10 years model spin-up (Beckers *et al.*, 2002). From these data sets, a nested model implementation with an arbitrary number of nesting levels at any location of the Mediterranean Sea can be built in an automated way. The only parameter that has to be specified is the number of nesting levels, the position, size and resolution of each nested model grid.

The model bathymetry

The model bathymetry is obtained from 1/60° bathymetry of the Mediterranean Sea. If the model resolution is lower than 1/60°, bathymetry is averaged over all grid-points of the 1/60° bathymetry falling within the model grid cell. If the model resolution is higher than 1/60°, the bathymetry is interpolated linearly (this functionality was added by L. Vandenbulcke, personal communication). For models with a sigma coordinate, the depth of the water column must be greater than a given threshold and smaller depths should be avoided for numerical stability. The GHER model uses a double-sigma coordinate and consequently two ranges of depth are not allowed. The depth must be greater than 30 m and it should not be between 170 and 300 m for the vertical discretisation adopted in the Mediterranean Sea (Beckers *et al.*, 2002). For each grid point with a depth outside the range of stability, the bathymetry is put to the closest permitted value. It is also possible to adjust the bathymetry at specific locations in order to avoid, for example, disconnected sea points.

The climatology

The monthly MEDAR or MODB temperature and salinity climatology is automatically interpolated to the model grids of the nesting system. These fields can be used to correct the surface heat and freshwater flux by relaxing the model SST and SSS towards the climatological values. It is also possible to nudge the temperature and salinity fields in depth towards the climatology. This is useful for example at open boundaries.

The initial conditions

As stated earlier, typical initial conditions for a 1 January are also interpolated to the model grid. In order to obtain initial conditions for a given date, the model needs to be spun-up with realistic forcings and possibly with data assimilation. The spin-up can also start from climatological temperature and salinity. Utilities exist for computing the surface elevation (based on a no-motion-level assumption) and for the calculation the geostrophic velocity.

Remaining implementation steps

Of course, the automation has to stop somewhere. The atmospheric forcings are not included in this implementation procedure since they are not part of the database. The size of a database containing the atmospheric fields of several years would be too important. Nevertheless, a specialised ftp-script was written to download and to convert either the NCEP Reanalysis forcings from the NCEP site (http://www.cdc.noaa.gov/ncep_reanalysis/) or the ECMWF forcings fields from the MFSTEP server (ftp://data.bo.ingv.it).

Model parameters such as the time step and diffusion coefficient have to be specified. Otherwise it would make the modeller redundant.

5.1.2 The Ligurian Sea model implementation

The oceanographic aspects of the Ligurian Sea model implementation were explained in chapter 3. Figure 5.1 shows the two successive grid refinements carried out in this model setup. The model resolution is first refined by a factor of 5 and then by a factor of 3.

Data flow

Figure 5.2 shows the interdependence of the three models of the nesting system. Only the models of two successive nesting levels communicate with each other. They exchange boundary conditions and feedback.

The dynamical equations are integrated in a synchronous way with the same time step. For each dynamical equation, the exchanges shown in figure 5.2 are realised. First, the necessary boundary conditions for a given variable are interpolated from the results of the next coarser nesting level. The Mediterranean Sea model provides the boundary conditions for the Liguro-Provançal Basin and then the Liguro-Provençal Basin model provides the boundary conditions for the Ligurian Sea model. The dynamical equation is then integrated one time step forward. Each nested model average its results to the resolution of its parent model. The highest resolution model begins with this feedback procedure. This averaged field of the Ligurian Sea model is sent back to the Liguro-Provençal Basin model. Model results of the Liguro-Provençal Basin in the overlapping domain are replaced by these values. Then, the Liguro-Provençal Basin model sends the average values to the Mediterranean Sea model. The feedback cascades thus information from the highest resolution grid to the coarsest resolution grid. The values of the Ligurian Sea model are thus found also in the Mediterranean Sea model.

The model integrates by mode-splitting the barotropic mode with a much smaller time step than the baroclinic modes. The boundary conditions and feedback are also exchanged between the barotropic time steps. During the integration of the barotropic flow, the data exchange is therefore particularly high.

Model parallelisation

The GHER model can run in parallel on different CPUs (Central Processing Unit) using a domain decomposition. The model domain is divided statically into several subdomains with the same resolution and with a similar number of sea points. The exchange between the different domains is realised by the message-passing library PVM (Parallel Virtual Machine). But this form of parallelisation was not used in the present work.

Each model of the nesting system is a separate executable. The transfer of the boundary conditions and the feedback is also realised by PVM (figure 5.3). The nesting is also a form of model parallelisation since all models of the nesting system run on different processors.

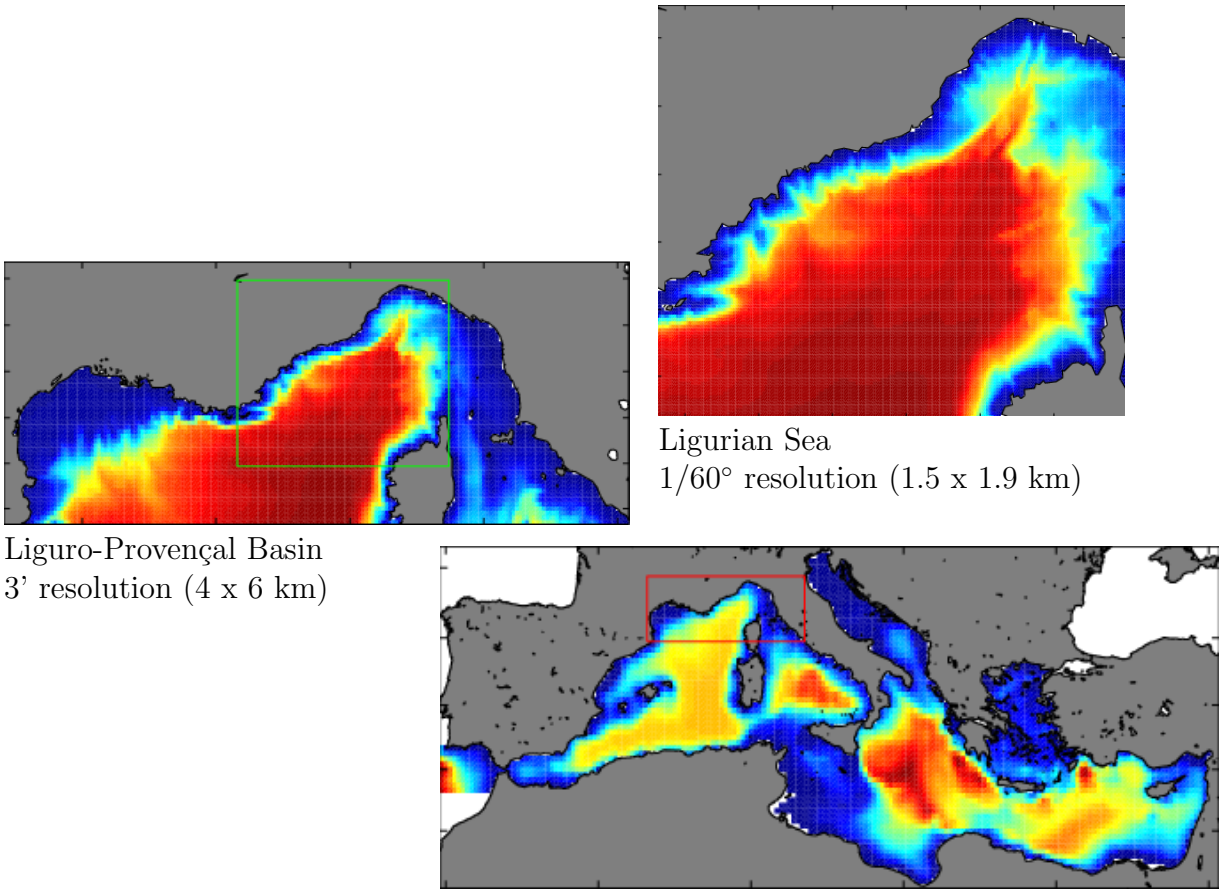

Figure 5.1: The three domains of the Ligurian Sea implementation. The colour indicates the depth.

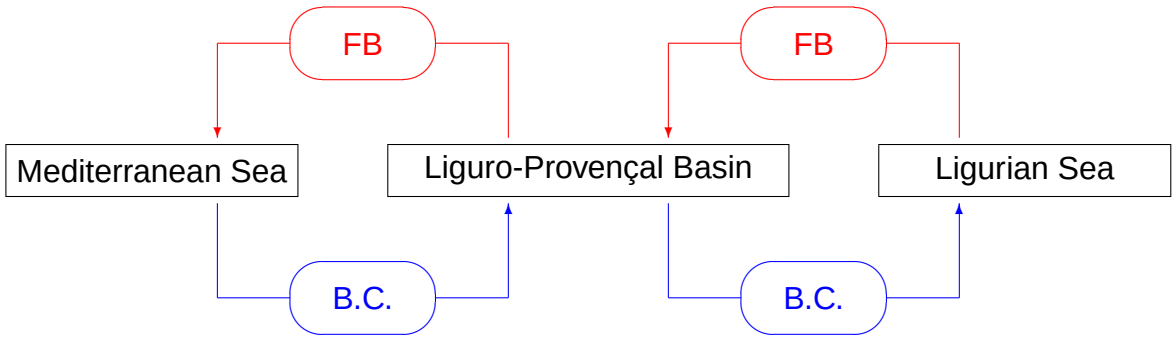

Figure 5.2: The data flow between the three models. "B.C." stands for boundary conditions and "FB" for feedback.

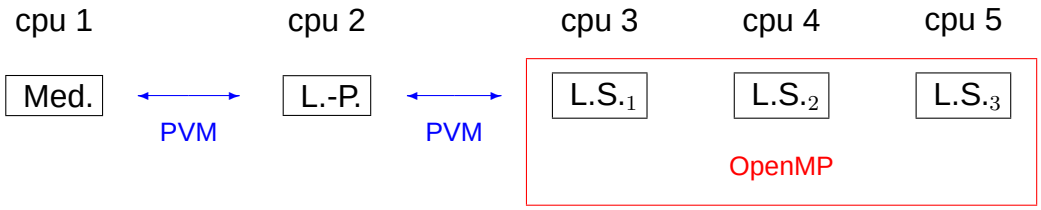

Figure 5.3: Parallelisation in the system of nested models. Such approaches are also known as dual-level parallelisation (Luong *et al.*, 2004).

Model	Nb. of Sea points
Mediterranean Sea (Med.)	120374
Liguro-Provençal Basin (L.-P.)	151596
Ligurian Sea (L. S.)	477364

Table 5.1: Comparison of the number of sea points between the different models

Load balance of the nested models is of major concern since the grids of the nesting system have a different spatial resolution and a different number of grid points (table 5.1). The nesting mode is presently only to a limited extent compatible with the domain decomposition since only the coarsest level of the nesting system can be parallelised by the domain decomposition. But in a typical nesting implementation the fine grid model has the most grid points. An efficient nesting set-up would require that the fine resolution grid model is further parallelised in order to balance the execution speed of each model.

To overcome this problem, the model code was parallelised with OpenMP in a way that it can be used at all levels of the nested models. The main drawback of OpenMP is the need for a shared memory computer-architecture. OpenMP code cannot be ported, for example, on clusters, which have gained popularity due to their low cost.

On the other hand, OpenMP parallelisation requires less programming effort than a message passing parallelisation and it can be done incrementally: starting from a low level, the parallel region can progressively be extended to the entire program.

Each model of the nesting system can run with a different number of CPUs with minimal modifications. The Ligurian Sea model, which has the highest number of sea points, is run therefore with three or four CPUs as shown in figure 5.3. Depending on the CPU time availability, the Mediterranean Sea model can be run on two CPUs, the Liguro-Provençal Basin model also on two CPUs and the Ligurian Sea model on six CPUs.

To change the number of processors, the PVM domain decomposition would require a complete re-implementation since each domain has its specialised executable. Of course, a multiple data/single program (MDSP) approach allowing a straight forward change

of the number of CPUs is also possible with a message passing library such as PVM or MPI. Such strategies are easier to realise with dynamic allocation, which is possible in Fortran 90.

The OpenMP parallelisation is done in such a way that a team of threads running on different CPUs is created at the beginning of the main time loop. This parallel region is only closed at the end of the model run. The work of the loops is then shared among the team of threads. For nested loops, the outermost loop is parallelised, which is in most cases the vertical or the zonal direction. A rearrangement of some variables is necessary for rendering each loop-iteration independent of the others. The parallelisation was validated by verifying that the model produces the same results up to the last digit for different number of CPUs.

Table 5.2 shows the performance of the OpenMP parallelisation for different numbers of CPUs without model nesting. The model covers the Mediterranean Sea with a unique grid of 1/4° resolution. The barotropic time step is 30 s and the baroclinic time step is 1800 s. The model simulation is done for one month.

The speed-up of a program is determined by how much faster it is on p processors than on one processor. Therefore, the speed-up $S(p)$ of a program on p processors is the ratio of the execution time obtained with one processor $t(1)$ and the execution time with p processors $t(p)$:

$$S(p) = \frac{t(1)}{t(p)} \tag{5.1}$$

Generally, the speed-up of a program on p CPUs is lesser or equal than p. The speed-up can be greater than this limit due to the effect of the cache architecture, for example. But in real programs, the communication overhead between the processes slows down the execution time and the speed-up is less than the number of CPUs.

The efficiency $E(p)$ of a parallel code running on p processors is the ratio between the speed-up and the number of processors p:

$$E(p) = \frac{S(p)}{p} \tag{5.2}$$

The efficiency measures how close is the speed-up to the theoretical limit.

The efficiency of the parallelisation obtained here is acceptable for a low number of CPUs. For higher number of CPUs the acceleration of the program becomes smaller and smaller. This is due to the parallelisation overhead. At several stages, there are implicit or explicit barriers in the code. This barrier can only be passed if all threads have reached this barrier. These barriers ensure that all the work before the barrier has been completed. Some future optimisation can be done to verify the necessity of each barrier.

Number of CPUs	Execution time (s)	Speed-up	Efficiency
1	8306.89	1.00	1.00
2	4978.02	1.66	0.83
4	3272.34	2.53	0.63
8	2146.16	3.87	0.48

Table 5.2: The execution time, seed-up and efficiency of the Mediterranean Model with different number of CPUs.

As mentioned earlier, the parallel region has been extended to the entire time loop. Only a single fork[1] of processes at the beginning of the time loop is done. Only parts of the model code without computations such as input and output have not been parallelised. The fraction of unparallelised code is therefore very low and significant improvement due parallelisation of these parts is not expected.

5.1.3 Algerian Current

By presenting only a single nesting implementation, one cannot claim that a model is really relocatable. The nesting system has therefore been applied to another domain. The Algerian Sea was chosen for the nesting implementation. The Algerian Current is formed by Modified Atlantic Water (MAW). This water mass enters the Mediterranean Sea by the Strait of Gibraltar and follows the Moroccan, Algerian and Tunisian coast. This meandering current is dynamically very active and often eddies are detached from the Algerian Current. The dynamics of this region is therefore very different from the dynamics of the Ligurian Sea.

The coarse model of the Mediterranean Sea has, as before, a resolution of 1/4°. The grid of the Algerian Current is refined by a factor of 5. The domain of this model is shown in figure 5.4. The time necessary to implement a model can be a critical factor. The model implementation, starting with the choice of the model domain to getting the first results, has been done in two days. Figure 5.5 shows the SST of the model after an integration of one month starting from the 1 January 2000. Some discrepancies are apparent when comparing the model with the observed SST. In particular, the North-South gradient is less pronounced in the model than in the observed SST. In the model water from Algerian Current reaches the Balearic Island. However, this has also been observed in simulations of the Mediterranean DieCAST model (Fernández *et al.*, 2003).

The implementation time here does not include the necessary calibration and validation of the model. This is probably the most important and time-consuming task of the model setup.

[1] Generation of child threads which are exact copies of the parent process.

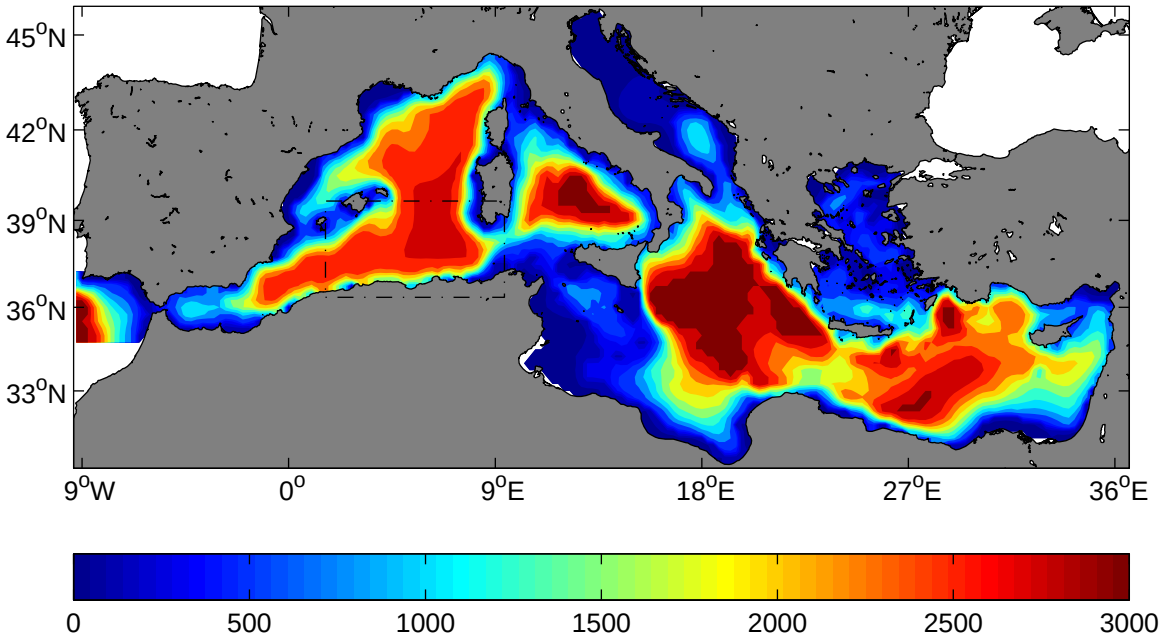

Figure 5.4: The Algerian Current model nested in the Mediterranean Sea model.

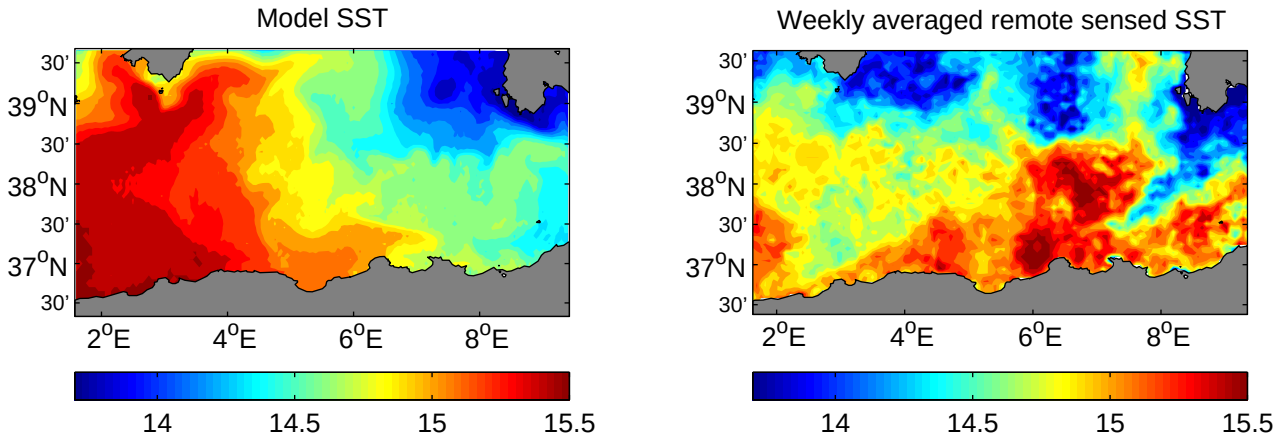

Figure 5.5: The model SST of the 6 February 2000 is shown on the left and the weekly averaged remote sensed SST of the 3 February 2000 on the right.

5.2 Data assimilation

5.2.1 Coding aspects

The developed software is presented in the form of seven Fortran 90 modules. A specific task such as binary input/output, reading the entry of the configuration file, high-level matrix operation and the implementation of different analysis schemes, is assigned to each module. The computation in data assimilation involves mainly linear matrix algebra. Fortran is well suited for these computations since the matrices are primitive data structures. Fortran 90 allows the definition of data types and operators, and the assimilation package makes extensive use of this possibility. A new data type is defined for sparse matrices. This structure contains only the non-zero elements of the matrix and their indexes. This sparse matrix type is primarily intended for the observation operator.

Operators

The definition and overloading of operators makes use of data types particularly interesting. For example, the matrix product was implemented with the operator "`.x.`". The left and right operant can be full or sparse matrices. A specialised multiplication subroutine corresponding to the data types is associated to the operator "`.x.`" by the compiler. The product of the two matrices $\mathbf{A}$ and $\mathbf{B}$ can then be written as:

```
C = A.x.B
```

Other operators involving the product of a diagonal matrix and a full matrix (the operator `.dx.`) or the transpose of a matrix and a matrix (the operator `.tx.`) are also defined. For performance reasons, it would be very inefficient to compute all products by a generic multiplication routine. The multiplication itself is done by the BLAS library (Basic Linear Algebra Subprograms). Most compilers allow us to inline specific subroutines. If all multiplication functions are inlined, then the program strongly resembles to a Fortran 77 code with direct call to the BLAS library. Therefore, the use of operators for the matrix multiplication does not reduce necessary the performance.

The primary goal of this approach is to obtain a program code close to the underlying mathematical formulas. For example, let us consider the following mathematical expression. It computes the amplitude of the correction along the error modes and is related to equation (4.161).

$$A = \mathbf{U}\mathbf{D}\mathbf{U}^T \left(\mathbf{H}\mathbf{S}^f\right)^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f) \tag{5.3}$$

where $\mathbf{D}$ and $\mathbf{R}$ are diagonal matrices. Its diagonal elements are stored in the vectors `D` and `R` respectively. This expression can be implemented with following the Fortran 90 code:

```
Hxf = H.x.xf
HSf = H.x.Sf
A   = U.x.(D.dx.(U.tx.(HSf.tx.((yo-Hxf)/R))))
```

We believe that the resulting code is more readable than the code with direct calls to the BLAS library. The code is thus also easier to maintain and to extend.

The order of the multiplication can be controlled by brackets. The order does not change the result but it has implications for the number of operations performed. Sometimes, brackets are even necessary while they are optional in the mathematical expression. Any user-defined operator has a lower precedence than a built-in operator. In particular, the addition has a higher precedence than our multiplication operator.

5.2.2 Interface to the assimilation package

The assimilation package was developed so that the analysis can be performed in two different ways: either by subroutine calls from the model directly to the assimilation package or by separate standalone programs reading all necessary data from the disk. Both approaches have advantages and limitations. The first method is the most efficient one since the access to the disk is minimised. This approach is based on a single executable running from the beginning to the end of the experiment.

The standalone program for the assimilation is very useful for testing the implementation and for repeating an analysis step with different parameters. The same configuration file can be used in both approaches and the standalone program can be used for a consistency check without the run of the complete hydrodynamic model. The sequential character of the Kalman filter algorithm also allows combining the model and the assimilation with shell scripts. This approach is for example used by the SESAM package (Testut *et al.*, 2002). The coupling of the model and the assimilation scheme can be done very easily.

5.2.3 Binary data format

Large quantities of data are more efficiently stored in a binary data format than as ASCII text. Currently, the GHER and a subset of the NetCDF format are supported. The data stored and read through the assimilation program is limited to 1D, 2D and 3D arrays of real numbers. The matrix can contain missing values ("holes"). These points are generally attributed to land points or, for example, missing observations due to clouds.

Matrices A, where the elements are a linear combination of the indices, can also be efficiently represented:

$$A(i,j,k) = a_0 + a_1 i + a_2 j + a_3 k \tag{5.4}$$

Only the coefficients a_0, a_1, a_2 and a_3 are stored. These files are called degenerated files. For example, the longitude and latitude of each grid point can often be expressed in this way.

For the GHER format, each file represents an array of real numbers. If the filename ends with `.gz`, then the file is automatically decompressed. Simple Fortran 90-style and Matlab-style extraction can also by performed. A coma-separated list of indices or

ranges of indices in parenthesis can be appended to the filename, if only a subsection of the matrix should be loaded.

For example, if the file `GHER_file.TEM` is a $10 \times 10 \times 10$ matrix, the "file":

`GHER_file.TEM(:,:,6)` is $10 \times 10 \times 1$ matrix containing all elements with the 3rd indices equal to 6.

`GHER_file.TEM(:,end,:)` is $10 \times 1 \times 10$ matrix containing all elements with the 2nd indices equal to 10.

`GHER_file.TEM(1:,:end,1:end)` is $10 \times 10 \times 10$ matrix equal to the original matrix

But no arithmetic with the indices (for example `GHER_file.TEM(:,end-1,:)`) are allowed. If data extraction is used with degenerated matrices, the four coefficients are changed accordingly to the subsection chosen.
A variable in a NetCDF file can be loaded by specifying a "filename" of the following form:

```
NetCDF_filename#NetCDF_variable
```

If the NetCDF filename end with `.gz`, then the file is uncompressed as with the GHER file format. The data extraction follows also the same rules as above.

The subroutines related to the binary output/input are grouped in one Fortran 90 module. Most new utility programs are based on this module. All new added functionality in this module is therefore beneficial for a large number of individual programs. The support of compressed data and the data extraction facility have given all these programs an additional flexibility.

5.2.4 The initialisation file

The assimilation package is independent of the model and the implementation. At compile time, no problem-specific parameters have to be given. The whole assimilation procedure is driven by a single initialisation containing all necessary information about the model, the observations and the error covariances.

The initialisation file can specify parameters of different types: integer number, floating number or a character string (mostly filenames). Also vectors of these types can be specified. It is beyond the scope of this chapter to describe the complete syntax of the initialisation file, but the following two examples should clarify how it works:

```
runtype = 2
Model.variables = ['ETA','TEM','SAL']
```

When the program search for example the key "`runtype`", it gets the integer 2. A query for the key `Model.variables` returns the vector containing the three character strings `'ETA'`,`'TEM'` and `'SAL'`.

Sometimes a sequence of keys is attributed to the same values:

```
Obs001.path      = '/home/user/Obs/'
Obs002.path      = '/home/user/Obs/'
Obs003.path      = '/home/user/Obs/'
```

In this case one can use wildcards and write the following:

```
Obs*.path      = '/home/user/Obs/'
```

The meanings of the wildcards are the same as for filename generation of the Burne Shell.

5.2.5 Configuration

The initialisation file of the assimilation module is composed mainly of four sections: configuration of the model (model state vector, position of the individual variables, error space of the model and the partitioning of the state vector for local assimilation), observations to assimilate (observations, their position, and their error), possible diagnostics of the analysis and miscellaneous flags.

The model

The following example contains the definition of the multivariate state vector. The key `Model.variables` is a vector of character strings attributing to each variable a user chosen name. The keys `Model.gridX`, `Model.gridY`, `Model.gridZ` and `Model.mask` are vectors of filenames. The files in `Model.gridX` and `Model.gridY` are degenerated and give the longitude and latitude of each variable. The files in `Model.gridZ` can be plain files and contain the depth. The key `Model.mask` is used to determine the sea-land mask of each variable. The exclusion value (or missing value in NetCDF terminology) marks a land point and all other values represent a sea point. All files assembled into a state vector should have physical values where the mask assumes a sea point.

The string in `Model.path` in prepended to each filename. Example:

```
Model.variables = ['ETA'              ,'TEM'    ,'SAL']
Model.gridX     = ['ligur.X(:,:,end)','ligur.X','ligur.X']
Model.gridY     = ['ligur.Y(:,:,end)','ligur.Y','ligur.Y']
Model.gridZ     = ['ligur.Z(:,:,end)','ligur.Z','ligur.Z']
Model.mask      = ['ligur.Z(:,:,end)','ligur.Z','ligur.Z']
Model.path      = '/home/user/Data/'
```

For nested grids the variables of the same nested grid must be grouped and the groups must be ordered according to the resolution, starting with the highest resolution one. A `Model.gridnum` is associated to each model grid: one for the highest resolution one, two for the next highest resolution one, and so on.

```
Model.variables = ['TEM'    ,'SAL'    ,'TEM',   'SAL']
Model.gridX     = ['ligur.X','ligur.X','med.X','med.X']
Model.gridY     = ['ligur.Y','ligur.Y','med.Y','med.Y']
Model.gridZ     = ['ligur.Z','ligur.Z','med.Z','med.Z']
Model.mask      = ['ligur.Z','ligur.Z','med.Z','med.Z']
Model.gridnum   = [        1,        1,      2,      2]
Model.path      = '/home/user/Data/'
```

The error space of the model $\mathbf{P}^f$ is given either by specifying the columns of the matrix $\mathbf{S}^f$ directly or by an ensemble of states. In the latter case $\mathbf{S}^f$ is obtained by the matrix $\mathbf{E}^f$ containing the N members given by

$$\mathbf{S}^f = \frac{1}{\sqrt{N-1}}\left(\mathbf{E}^f - \mathbf{E}^f \mathbf{1}_{n\times N}\right) \tag{5.5}$$

It is not necessary for the columns of $\mathbf{S}^f$ to be orthogonal. Sometimes it is useful to change the error variance to modify the amplitude of the correction. The error variance can either by changed by a constant factor α^2:

$$\mathbf{P}^f \rightarrow \alpha^2 \mathbf{P}^f \tag{5.6}$$

or by a diagonal matrix $\mathbf{D}$:

$$\mathbf{P}^f \rightarrow \mathbf{D}\mathbf{P}^f\mathbf{D} \tag{5.7}$$

The later approach allows us to change the error variance by a space dependent factor. For example, the error variance can be decreased progressively to zero at depth in order to avoid the assimilation to modify the variables at depth. However, it is important to check that the new error covariance is in accordance with the basic equilibrium relations of the ocean. For example, if the temperature and salinity variance is modified and the elevation is not changed, then it is possible that the new error covariance no longer satisfies the dynamical height relationship (6.2). Pressure gradients at depth can then introduce important velocity changes at depth.

Covariance localisation makes assimilation feasible with a low dimensional error space. This procedure suppresses (unrealistic) long-range correlations and increases the effective rank of the covariance matrix. The covariance localisation is realised as in Testut *et al.* (2002). The state vector is partitioned into several zones. These zones are specified by a vector of the size of the state vector formed only by integral values. All elements in the state vector with the same value in this partition vector belong to the same zone. The analysis is done for each zone independently, taking only into account observations nearby. This is implemented by changing the error variance of the observations. For each

zone located around the point $\mathbf{x}_c$, the weight of the observations $\mathbf{R}(\mathbf{x})^{-1}$ is multiplied by a Gaussian function:

$$\mathbf{R}'(\mathbf{x})^{-1} = \exp\left(-\frac{\mathrm{dist}_h(\mathbf{x},\mathbf{x}_c)^2}{l_c(\mathbf{x})^2}\right)\mathbf{R}(\mathbf{x})^{-1} \tag{5.8}$$

Here $\mathbf{x}$ and $\mathbf{x}_c$ denote two points in the three dimensional ocean, $\mathrm{dist}_h(\mathbf{x},\mathbf{x}_c)$ is the horizontal distance between these points and l_c is the length scale defining which observations should influence the analysis in a given zone. This length scale can depend on the position $\mathbf{x}$ and is related to the horizontal correlation length. Since in a system of nested models, the scale of the smallest processes resolved is different in each model, the length l_c is smaller in the high-resolution model than in the coarse resolution model. The problem with this local analysis is that true long-range correlations are also filtered out. Long-range errors can therefore only be corrected by a sufficiently dense observational network.

The analysis in each zone is independent of the other. Therefore, a parallelisation of the loop running over the zones is straightforward. As the GHER model, the local assimilation scheme has been parallelised with PVM and OpenMP. Both parallelisation approaches can be used at the same time. A parallelisation based on both, PVM and OpenMP, is necessary to distribute the work of the analysis over all CPUs of the nesting system.

The observations

The observations to assimilate are ordered chronically and a time index starting with 1 is attributed to them. Observations at the same time are specified simultaneously. The first observation vector $\mathbf{y}^o$ constituted by SST and SSH can be defined for example by:

```
Obs001.date      = '06/07/2000'
Obs001.time      = '00:00:00'
Obs001.variables = ['TEM','ETA']
Obs001.value     = ['med.SST','med.SSH']
Obs001.mask      = ['med.SST','med.SSH']
Obs001.rmse      = ['med.SST.RMSE','med.SSH.RMSE']
Obs001.gridX     = ['med.SST.X','med.SSH.X']
Obs001.gridY     = ['med.SST.Y','med.SSH.Y']
Obs001.gridZ     = ['med.SST.Z','med.SSH.Z']
```

The nearest model time step for the assimilation of the observations is computed from the two first keys. The key `ObsXXX.variables` determines which model variables are observed. The filenames of the observations themselves are specified in `ObsXXX.value`. All points flagged as missing points (for example land and cloud points for the SST) in the files `ObsXXX.mask` are not assimilated. These two last keys have generally the same value since the mask is included in the observations. With the key `ObsXXX.rmse`, the root mean square error of the observations (or the square root of the error variance) is specified. The error variance can vary in space, but the observations are supposed to be uncorrelated.

Observations				Model				
var. index	i-index	j-index	k-index	var. index	i-index	j-index	k-index	Interpolation coefficient

Table 5.3: Structure of the sparse matrix

The three last keys define the grid on which the observations are given. For each scalar observation the longitude, latitude and depth are specified. The observations can be distributed in an irregular grid. For observations related to two dimensional variables such as the surface elevation, a valid file for the depth is necessary, but its values are not used.

The observation operator performs (possibly) a shift of the model state vector by the offset $\mathbf{x}'$ followed by the multiplication of the matrix $\mathbf{H}$. The observation operator $H(\mathbf{x})$ can therefore be written as:

$$H(\mathbf{x}) = \mathbf{H}(\mathbf{x} - \mathbf{x}') \tag{5.9}$$

In most cases, the vector $\mathbf{x}'$ is zero. But when the sea level anomalies are assimilated, the elements of $\mathbf{x}'$ corresponding to the elevation can contain, for example, the mean dynamic topography (MDT). There is no need to interpolate the MDT to the satellite tracks.

The matrix $\mathbf{H}$ is constructed with the information about the position of the observations. If possible, a trilinear interpolation is done (temperature measurements at depth, for example); otherwise, the observation operator is based on a bilinear interpolation (SST for instance).

The implementation was done keeping the model nesting in mind. Several variables can represent the same physical point due to the domain overlapping. This ambiguity is solved by taking in this case the model grid with the highest resolution. The highest resolution model is supposed to be more accurate than the model with coarser resolution.

Internally the observation operator is stored as a sparse matrix holding the interpolation coefficients and the indexes of the non-zero elements of this matrix. It is also possible to specify directly the sparse matrix $\mathbf{H}$. This feature has been used to assimilate the amplitudes of the SST EOFs (see chapter 7). Only the non-zero elements of $\mathbf{H}$ are given in a $nz \times 9$ matrix, where nz are the number of non-zero elements. Each line of this matrix contains 9 numbers with the structure given in table 5.3. The first four integer values are related to the observations. The index of the variable is the position where the observed variable appears in `ObsXXX.value` and i,j,k-indexes are the three spatial indexes of a single scalar observation. The integers in column 5 to 8 are related to the model state vector. Again, the index of the variable is the position where the observed variable

appears in `Model.variables` and i,j,k-indexes are the three spatial indexes of a single scalar model state. If one of the model indexes is -1 the corresponding observation is treated "out of grid" and the associated weight will be zero.
The column 9 is the value of the non-zero element of $\mathbf{H}$. In the case of a linear interpolation, it takes values between 0 and 1. Utilities for creating the observation operator, based on a linear interpolation or a spatial low-pass filter and for multiplying sparse matrices in these formats, have been developed.

The diagnostics

After the assimilation, the *a posteriori* state $\mathbf{x}^a$ is returned to the GHER model. Several optional diagnostics can be saved for future analysis of the assimilation experiment. These diagnostics include:

- the model forecast, $\mathbf{x}^f$
- the model forecast at observation locations, $H(\mathbf{x}^f)$,
- the error space of the model forecast, $\mathbf{S}^f$,
- the standard deviation of the forecast error, $\mathrm{diag}(\mathbf{P}^f)^{\frac{1}{2}}$
- the standard deviation of the forecast error at observation locations, $\mathrm{diag}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T)^{\frac{1}{2}}$

All these diagnostics can also be computed for the *a posteriori* state $\mathbf{x}^a$ and the *a posteriori* error covariance $\mathbf{P}^a$. Furthermore the analysis increment $\mathbf{x}^a - \mathbf{x}^f$ and the analysis increment at observation locations $\mathbf{H}\mathbf{x}^a - \mathbf{H}\mathbf{x}^f$ can be stored since they are useful to detect systematic errors. Since the analysis increment is a linear combination of the columns of $\mathbf{S}^f$, these coefficients can also be diagnosed. When the covariance localisation is applied, these coefficients are different for each zone. If the analysis has produced an unrealistic state, it is thus possible to determine which error mode is dominant in the analysis increment.

Also the error of the *a priori* state $\mathbf{H}\mathbf{x}^f - \mathbf{y}^o$ and the *a posteriori* state $\mathbf{H}\mathbf{x}^a - \mathbf{y}^o$ compared to the observations can be stored directly.

All these diagnostics are highly redundant and it would be useless to export them all. But the variety of possible diagnostics allows us to save to the disk exactly the needed information. Furthermore, scalar diagnostics such as RMS errors and bias for each assimilated variables are computed, as well the value of the χ^2 testing the consistency of the error covariances (Bennett, 1992):

$$J = \left(H(\mathbf{x}^f) - \mathbf{y}^o\right)^T \left(\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T\right)^{-1} \left(H(\mathbf{x}^f) - \mathbf{y}^o\right) \tag{5.10}$$

If the error covariance $\mathbf{R}$ and $\mathbf{P}^f$ are correctly specified, J follows a χ^2_m distribution. The degrees of freedom of this distribution is equal to the number of scalar measurements m. In average, J should be equal to m.

However, from the analysis equation (4.134) we can see that the analysis will not change if the model and observations error covariance are both multiplied by the same constant factor. But this constant factor will change the mean value of J. Therefore, for any $\mathbf{P}^f$ and $\mathbf{R}$ one can find a multiplication factor for this covariances so that the average of J is m but without changing the analysis.

Furthermore, it is not clear how this quantity should be interpreted when the error variance of the observations was artificially inflated to take into account the correlation of the observations (see section 6.7).

5.2.6 Alternative analysis schemes

Besides the global analysis and local analysis, the bias aware analysis scheme of Dee and Silva (1998) is also implemented (both in global and local mode). But the error space $\mathbf{S}^a$ is not computed in this approach. This has been tested for the assimilation of the sea level anomaly. An error in the mean dynamic topography used for referencing the sea level anomaly would in fact introduce a constant bias in the assimilation. But the bias aware assimilation scheme did not improve significantly the ability of the model in predicting future observations.

Some physical constraints cannot be imposed by error covariances based on Gaussian statistics. An example is stable water columns. Unstable water columns are possible but they are rare and related to extreme atmospheric events. Gaussian statistics are symmetric and therefore they suppose that positive corrections of the vertical density gradient are as likely as negative corrections. It is not uncommon that the analysis renders the water column unstable (Brasseur, personal communication). An analysis scheme enabling us to maintaining the water columns stable by an anamorphosis transform is under development. An anamorphosis transform consists in applying a nonlinear invertible function f to the model state vector $\mathbf{x}$ (Bertino *et al.*, 2002):

$$\mathbf{x}' = f(\mathbf{x}) \tag{5.11}$$

The pdf of this transformed state vector $p_{\mathbf{x}'}$ is related to the pdf of the original state vector $p_{\mathbf{x}}$ by:

$$p_{\mathbf{x}'}(\mathbf{x}') = \left|\frac{\partial f}{\partial \mathbf{x}}\right|_{x=f^{-1}(\mathbf{x}')} p_{\mathbf{x}}(f^{-1}(\mathbf{x}')) \tag{5.12}$$

If the pdf $p_{\mathbf{x}}$ is specified by an ensemble of state $\mathbf{x}^{(k)}$, $k = 1, \ldots, N$, then an ensemble of realisations of $\mathbf{x}'$ can be obtained by applying the nonlinear function to all states of $\mathbf{x}^{(k)}$:

$$\mathbf{x}'^{(k)} = f(\mathbf{x}^{(k)}) \tag{5.13}$$

The interest of the method lies in the fact that a non-Gaussian distributed state vector can in principle be transformed into a Gaussian state vector.

The most analysis schemes assume Gaussian distributed variables. If the ensemble of model forecast $\mathbf{x}^f$ does not have this property, then the anamorphosis transform can be used to obtain a Gaussian distributed ensemble. The assimilation is performed on the transformed variables. In general, the observation operator becomes a nonlinear function. The *a posteriori* ensemble is finally transformed back by applying f^{-1} in order to get the usual quantities which can be predicted by the model.

In practice, the anamorphosis transform is only applied to a subset of problematic variables, such as the concentration of biological variables, which cannot become negative. The distribution of the biological variables resembles more to a log-normal distribution and by an appropriate anamorphosis transform, for example the logarithm of the concentrations, these variables become Gaussian distributed variables. After the assimilation, the variables are transformed back (by the exponential function for example) and by definition all concentrations will be positive.

This approach of the anamorphosis transform has been implemented to control the stability of the water columns. The difficulty resides in the fact that the stability of the water column at a given depth depends on temperature and salinity above and under this depth. The anamorphosis transform is therefore a non-local function and acts on different variables simultaneously. Furthermore, the SST is often assimilated and the observation operator involves the inverse anamorphosis transform. The observation operator is therefore nonlinear, non-local and multivariate.

The choice of the anamorphosis transform is not unique. The anamorphosis transform implemented in the assimilation package is based on a linearised state equation:

$$b = \alpha(T - T_0) - \beta(S - S_0) \tag{5.14}$$

where T_0, S_0, b, α and β are respectively the reference temperature, salinity, buoyancy, thermal expansion coefficient and the saline contraction coefficient. The vertical gradient of the buoyancy is the square of the Brunt-Väisälä frequency N^2.

$$N^2 = \frac{\partial b}{\partial z} = \alpha \frac{\partial T}{\partial z} - \beta \frac{\partial S}{\partial z} \tag{5.15}$$

For stable water columns, N^2 is positive and for unstable water columns it is negative. N^2 is thus the parameter that we want to control by the anamorphosis transform. It would not be a judicious choice to eliminate completely the possibility of unstable water columns since they are real phenomena's responsible for deep or intermediate water formation. But the probability is very low and anamorphosis transform function can be chosen in such a way that unstable water columns are possible but occurring only with a given probability.

The anamorphosis transform proposed here changes T and S into the variables X and Y by:

$$X = f\left(\alpha\frac{\partial T}{\partial z} - \beta\frac{\partial S}{\partial z}\right) \tag{5.16}$$

$$Y = \alpha\frac{\partial T}{\partial z} + \beta\frac{\partial S}{\partial z} \tag{5.17}$$

$$\tag{5.18}$$

The bottom temperature and salinity are not transformed. The function f is a nonlinear function acting on N^2. The expression of this function still has to be determined. Different expressions for the second variable Y are of course possible. Here Y has been chosen so that a change in Y does not affect at all the Brunt-Väisälä frequency. If f is invertible, the whole transformation is also invertible. Temperature and salinity profiles are obtained by integrating their gradients starting from the bottom.
A user-defined function can be specified for f. Further research is necessary to find a good choice for the function f. This function can be determined by the cumulative density function F_{N^2} of N^2. The value of the cumulative density function F_N^2 in t is the probability that the square of the Brunt-Väisälä frequency is less than t:

$$F_{N^2}(t) = \Pr(N^2 < t) \tag{5.19}$$

The cumulative density function F_X of the transformed variable $X = f(N^2)$ can be related to F_{N^2} since:

$$F_X(x) = \Pr(X < x) = \Pr(N^2 < f^{-1}(x)) = F_{N^2}(f^{-1}(x)) \tag{5.20}$$

The function f is invertible, therefore it must be either monotonously increasing or monotonously decreasing. In equation (5.20) we have assumed the former case. The cumulative density function F_N^2 can be determined from the data and F_X should be a Gaussian cumulative density function and is thus also known. The anamorphosis transform function f is therefore:

$$f(t) = F_X^{-1}(F_{N^2}(t)) \tag{5.21}$$

To our knowledge, this method for determining the anamorphosis function is a new approach. Previously, only simple analytical functions such as the logarithm have been applied. Here, the function f is obtained from the statistics of the variable. A simple and efficient way to express the cumulative density functions F_X and F_{N^2} is by means of a piecewise linear function. The inverse of a piecewise linear function is obtained by permuting the independent and dependent values of the vertices. A function of a function is determined by a simple linear interpolation.

The main difficulty of the anamorphosis method is to find a suitable transformation. It is not clear if it can be assumed that this function is constant in time and over the whole domain. Furthermore, as it has been stated earlier, the observation operator becomes nonlinear if the transformed variables are observed. In some particular cases, this can be

avoided. If the anamorphosis transform involves only the observed variables individually, the same transformation can also be applied to the observations. A nonlinear observation operator is not mainly a problem for the implementation, but the standard Kalman analysis is derived for a linear observation operator. The ensemble Kalman filter analysis (4.142) can be extended to nonlinear operator if the matrices $\mathbf{HP}^f\mathbf{H}^T$ and $\mathbf{P}^f\mathbf{H}$ are estimated directly from the ensemble (Nerger, 2004). However, for a highly nonlinear observation operator, there is no guarantee that the analysis will produce an *a posteriori* state closer to the observations.

Chapter 6

Assimilation of SST and SSH

SST and SSH are assimilated in the system of nested models described in the previous chapters. The assimilation scheme is based on the equations (4.161) and (4.166). This chapter explains how this assimilation scheme is implemented to the nested models. In particular, the choice of the model state vector, the observations, the model error covariances and observation error covariance are described. Finally the results of the assimilation experiment are discussed.

6.1 Ensemble generation

Errors on the initial conditions and the atmospheric forcings are considered as the major error sources affecting the model forecast. The error space for the data assimilation has been defined consequently. An ensemble of 200 members has been created by perturbing the initial conditions and atmospheric forcings. For computational efficiency, only the coarse and intermediate resolution models were used.

6.1.1 Initial conditions

The initial temperature and salinity of the coarse resolution model are perturbed by a pseudo-random field with a horizontal correlation length of 100 km and a vertical correlation length of 20 m. A description of the random field generation can be found in Evensen (2003). At the surface, the standard deviations of the temperature and salinity perturbations are 0.5 °C and 0.1 respectively. Since the variability of the ocean is higher at the surface than at depth, the standard deviation of the perturbation is supposed to be a function of depth. A function of Gaussian shape with an inflection point at the depth of $h_{\mathrm{IP}} = 170$ m has been chosen.

$$\sigma(z) = \sigma(0) \exp\left(-\frac{z^2}{h_{\mathrm{IP}}^2}\right) \tag{6.1}$$

The depth of 170 m is also the transition depth of the two sigma regions of the coordinate transform. By comparing the model with CTD profiles measured in the Ligurian Sea (Sirena Cruises carried out by the Saclant Centre, La Spezia, Italy in August 2000), the model errors are indeed small, about 0.1 °C below that depth. In any circumstances,

even with a higher error at this depth, the surface data assimilated in the present work would not allow us to correct it. In this sense, the depth h_{IP} and the vertical correlation length also reflect our *a priori* idea of the depth to which a surface measurement can have a significant impact.

The perturbation on elevation is computed from the temperature and salinity perturbation. At a depth of $h_0 = 700$ m, horizontal variations of the hydrostatic pressure are small. This hypothesis was validated on several model results at different times. It implies a balance between temperature, salinity and surface elevation. The elevation perturbation $\Delta\zeta$ is therefore computed from the temperature and salinity perturbations ΔT and ΔS by the following linearised relation (Haines, 2002):

$$\Delta\zeta = \int_{-h_0}^{0} \alpha\Delta T - \beta\Delta S \; dz \tag{6.2}$$

This relationship is based on the assumption that the hydrostatic pressure does not change at depth h_0. The parameters α and β are the mean values of the thermal expansion coefficient and the saline contraction coefficient respectively:

$$\alpha = -\frac{1}{\rho}\left(\frac{\partial\rho}{\partial T}\right)_{p,S} \tag{6.3}$$

$$\beta = \frac{1}{\rho}\left(\frac{\partial\rho}{\partial S}\right)_{p,T} \tag{6.4}$$

Elevation, temperature and salinity perturbations allow us to compute the perturbation on the hydrostatic pressure by a linearised state equation. The perturbation of the horizontal velocity is supposed to be in geostrophic balance with the hydrostatic pressure perturbation. While the geostrophic balance relation is well respected in the interior of a basin, the relationship between pressure and velocity is more complex near the coast. For instance the geometry of the coast, nonlinear and non-stationary effects cannot be neglected. Therefore, at the coast, the velocity field is not perturbed. A coefficient $c(\mathbf{x})$ ranging from 0 at the coast to 1 in the interior of the basin ensures a smooth transition.

$$\Delta p_h(z) = g\rho_0\Delta\zeta - g\rho_0\int_z^0 \alpha\Delta T - \beta\Delta S \; dz \tag{6.5}$$

$$\Delta\mathbf{u} = \frac{c(\mathbf{x})}{f}\,\nabla\left(\Delta p_h\right) \wedge \mathbf{e}_z \tag{6.6}$$

The coefficient $c(\mathbf{x})$ decreases over a length scale of 50 km. Important perturbations of the velocity near the coast and inconsistent with the coast line geometry might also produce an "adjustment shock" and should be avoided also for the sake of the numerical stability.

The turbulent kinetic energy is not perturbed. It depends on a highly nonlinear way on the stratification and the shear and it is not clear what a consistent perturbation could

be. Furthermore the turbulent kinetic energy adjusts itself very rapidly to the density structure and the velocity profile of the model.

Perturbations of the coarse grid model initial conditions are finally interpolated to the Provençal basin model grid. The length scale of the perturbation is sufficiently resolved on coarse grids and the perturbations are comparable to those that one would obtain by generating the perturbation directly on the grid of the Provençal basin. But the latter approach would need some special treatment at the boundary between the coarse and intermediate resolution model to ensure a smooth transition between the two models initial conditions.

6.1.2 Atmospheric forcings

The atmospheric forcings are a non- negligible source of uncertainties in ocean forecasting. The dissipation in ocean systems will progressively let the system "forget" its initial conditions (*e.g.* Navarra, 2002). The uncertainties due to the initial conditions will therefore tend to zero after a sufficient long time integration. The errors due to the atmospheric forcings are however constantly injected into the model and will therefore have a significant impact on the longer term.

The air temperature at 2 m, the surface wind and cloud coverage were perturbed with a random field of 3 °C, 5 m/s and 30% standard deviation respectively. The two first standard deviations are also those chosen by Brusdal *et al.* (2003). The cloud coverage standard deviation is inspired by the temporal variability of the cloud coverage in summer. The perturbed cloud coverage is brought back in the permitted range of 0% and 100%. The spatial correlation is 1000 km and the time correlation is seven days.

6.1.3 Model integration

The Provençal Basin Model nested in the Mediterranean Sea Model has been integrated for 2 weeks starting on 5 July 2000. Each member starts from different initial conditions and is forced by perturbed atmospherics fields, but otherwise the conditions of each model integration are identical. The final ensemble contains therefore the cumulative spreading effects of unknown initial conditions and atmospheric forcings. Before this ensemble is used for the assimilation experiments, its statistics such as variance and covariance are analysed. The feedback between the Provençal Basin Model and the Mediterranean Sea Model produces naturally an ensemble where the values of the intermediate resolution model are consistent with those of the coarse resolution model and where the transition between the two models is smooth and dynamically meaningful.

6.1.4 Analysis of the ensemble

The spatial structure of the ensemble correlation is studied by choosing a surface point in the Gulf of Genova at 44°20' N and 8° 48 E. Figure 6.1 shows the horizontal and vertical correlation between the temperature at this location and the temperature of the other

grid points of the Provençal Basin Model. The correlation clearly reflects the path of the Northern Current. Since this current transports at the surface the same water mass (Modified Atlantic Water, (Sammari *et al.*, 1995)), the temperature is strongly correlated along its track. This horizontal structure of the correlation is an imprint of the dynamical properties of the region under study since this anisotropic distribution of the correlation was neither present in the perturbation of the initial conditions nor in the atmospheric forcings.

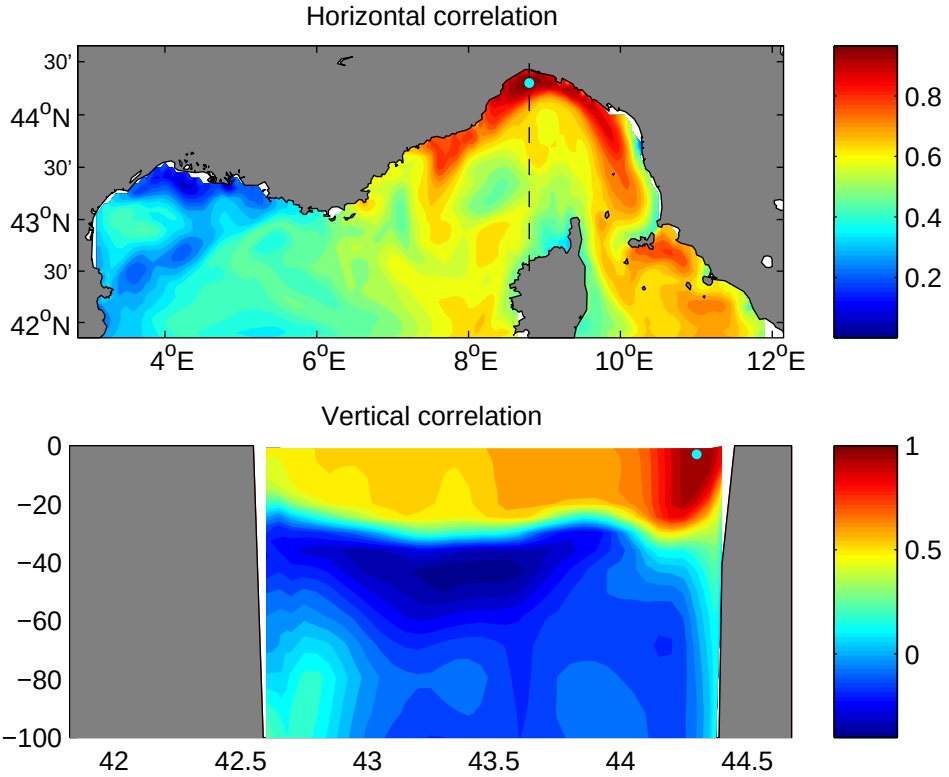

Figure 6.1: Horizontal and vertical correlation between the surface temperature at 44°20' N and 8° 48 E with the temperature at other positions in the Provencal basin.

6.2 Multivariate covariance

In some optimal interpolation and 3D-Var assimilation schemes, the error covariances are parametrised for example in terms of variance and correlation length (De Mey and Benkiran, 2002). The covariances between different variables are the most difficult to estimate in this context. A stochastic ensemble of model states or the time-variability of the model constitutes a very efficient and simple approach to estimate the multivariate error covariances. Examples exist in literature where an ensemble is used to calibrate variables of a parameterised error covariance (Etienne and Dombrowsky, 2003).

The upper panel of figure 6.2 shows the covariance of the elevation at the marked location in the Provençal Basin with the SST in the Mediterranean Sea. Near this point

the covariance is positive. This translates our physical understanding that if the surface is lifted, the water column must become lighter in order to avoid pressure and velocity changes at the reference level of no motion.

The error of these covariances due to the limited ensemble size can be estimated as described in Hamill *et al.* (2001). If c is the covariance between two random variables x_1 and x_2 and $\hat{c}$ is the estimated covariance based on n ensemble members, then the standard deviation of this estimator is for large n given by:

$$\mathrm{std}(\hat{c}) \sim \sqrt{\frac{1}{n}\left(1+\mathrm{corr}(x_1,x_2)^2\right)\mathrm{var}(x_1)\mathrm{var}(x_2)} \tag{6.7}$$

The lower panel of figure 6.2 illustrates this error estimation of the covariance between elevation and temperature. Uncertain error covariances can degrade the unobserved variables after assimilation. This is in average the case when the relative error of the covariance is greater than 1. The relative error is the ratio between the error of the covariance and the absolute value of the covariance (Daley, 1991; Hamill *et al.*, 2001).

Figure 6.3 shows the relative error for covariance between the elevation at the chosen location and the temperature. In red are the regions where the threshold is exceeded and the assimilation of a single elevation point in the Provençal-Basin will degrade in average the result. But where the covariances are small, only a small degradation would occur. In general, the estimation of small correlation requires a large ensemble size in order to be sufficiently accurate. From equation (6.7) one can show that the needed ensemble size n to accurately estimate a correlation useful for data assimilation must be greater than:

$$n > \frac{1+\mathrm{corr}(x_1,x_2)^2}{\mathrm{corr}(x_1,x_2)^2} \tag{6.8}$$

In general, the correlation between two variables is expected to decrease with their distance. Therefore it is not surprising that the regions where the assimilation of a single elevation measurement in the Provençal-Basin is likely to degrade the results are found far away from the observation point. This has consequences for the implementation of the assimilation and will be discussed in more details in section 6.3.3.

The finite ensemble size is only one of the error sources affecting the covariance estimation. The model inadequacy introduces also an error in the covariances. It is in reality a problem of the stochastic perturbation when uncertainties on the model parameterisation, for example, are not taken into account. Multiphysics and multimodel ensemble have been studied in meteorology for this purpose (Buizza *et al.*, 1999; Stensrud *et al.*, 2000; Houtekamer *et al.*, 1996).

Empirical orthogonal functions

The elevation, temperature and salinity of the coarse and intermediate model resolution grid have been grouped into a multivariate and multigrid state vector. The ensemble of these state vectors has been produced by a stochastic forecast. In this approach it is

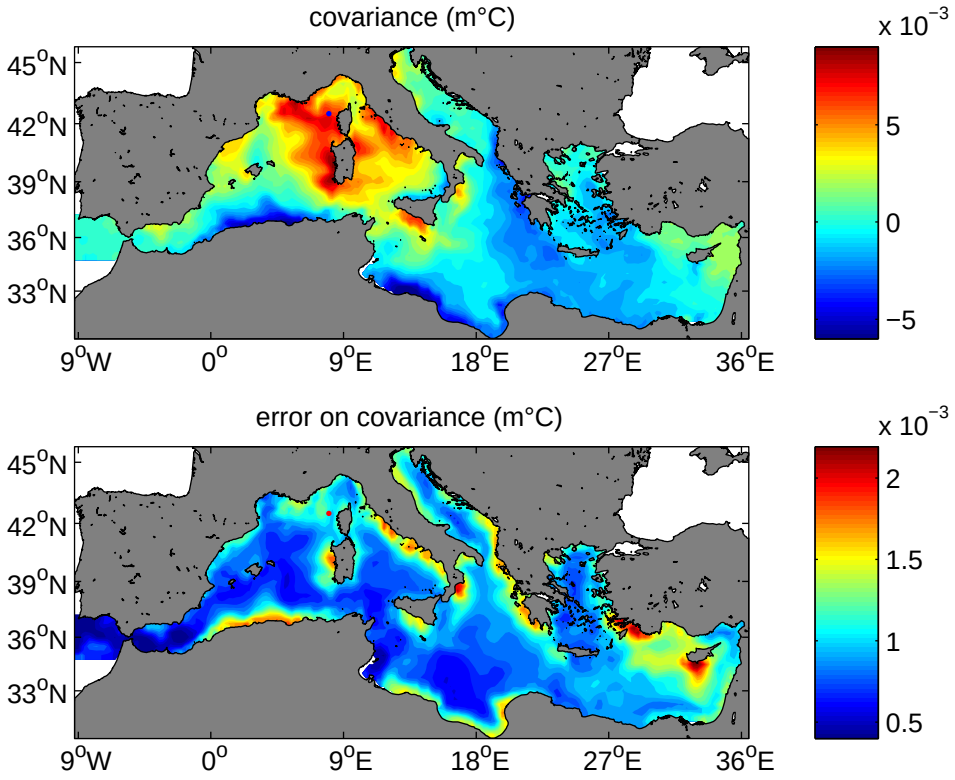

Figure 6.2: Covariance between the elevation of a point in the Provençal Basin at 42°30' N and 8° E and the sea surface temperature in the Mediterranean Sea in the upper panel. The lower panel shows the associated error due to the finite ensemble size by using the estimated variances and correlations in equation (6.7).

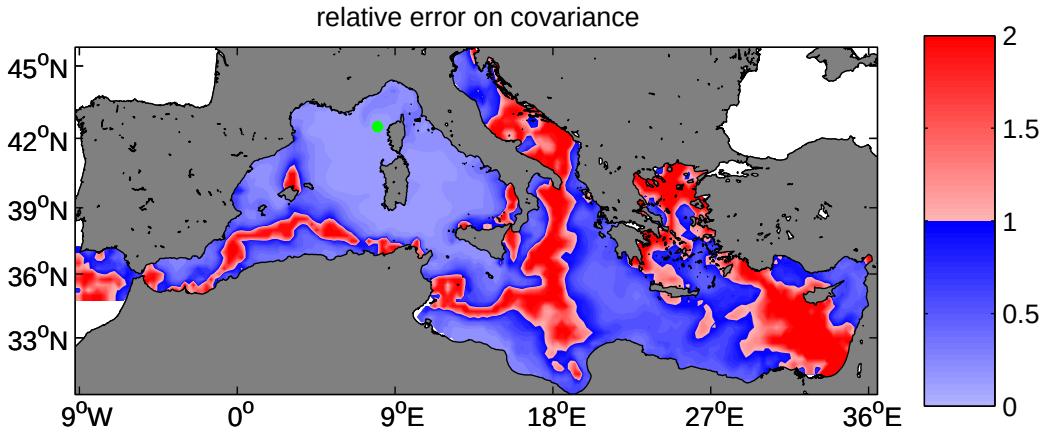

Figure 6.3: Relative error of the covariance between the elevation of a point in the Provençal Basin at 42°30' N and 8° E and the sea surface temperature in the Mediterranean Sea. The colour bar is limited to 2.

inevitable that some members will resemble to each other. By assuming that the ensemble statistic is sufficiently Gaussian distributed, an EOF decomposition of the ensemble can quantify the redundancy in the ensemble. The EOF analysis seeks the most important and recurrent differences between the ensemble members and the ensemble mean. This implies naturally that we need a norm in the state space to compare the importance of an error in the elevation, temperature and salinity. The sum of barotropic potential energy PE and available potential energy APE (Lorenz, 1955) was used for this purpose.

$$PE + APE = \frac{1}{2}\rho_0 g \int_{\mathcal{S}} \Delta\zeta^2 ds + \frac{1}{2}\rho_0 g^2 \int_{\mathcal{V}} \frac{\alpha^2 \Delta T^2 + \beta^2 \Delta S^2}{N^2} dv \tag{6.9}$$

where $\Delta\zeta$, ΔT and ΔS are respectively the difference in elevation, temperature and salinity between an ensemble member and the ensemble mean. The first integral is performed over the sea surface $\mathcal{S}$ and the second of the entire volume $\mathcal{V}$. The Brunt-Väisälä frequency N is computed from the ensemble mean and spatially filtered at sub-basin scale. The parameters α, β and ρ_0 are constants and represent typical values of the thermal expansion coefficient, the saline contraction coefficient and density respectively.

The two parts of this norm are related to the wave processes. The PE represents the energy that could be released into barotropic waves by transforming the surface elevation into a plan surface. The APE accounts for the energy that could be released by an adiabatic transformation of the density field into horizontal homogenous distributed density (here at subbasin scale, see figure 6.4). This kind of transformation generally produces internal waves.

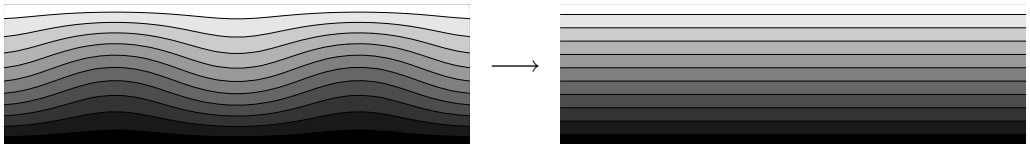

Figure 6.4: Adiabatic transformation of the density field. The lines represent the isopycnals. The horizontal homogeneous density field is the state with the minimum potential energy that can be reached without mixing the water masses.

The APE and PE of the free model run were computed. Generally, the APE is three to four orders of magnitude larger than the PE. The norm as expressed by equation (6.9) takes therefore mainly the baroclinic variability into account. For the computation of the EOFs this means that only the surface elevation variability originating from baroclinic processes is contained in the dominant EOFs. In ocean circulation model, the simulated processes are indeed focused on baroclinic variability. Rigid lid models even exclude the barotropic mode. Also, the observations assimilated in ocean circulation models often do not resolve the high frequent barotropic variability. The low weight attributed to the elevation is thus justified for ocean circulation models. However, if barotropic wave processes are studied, the error space should contain also purely barotropic error modes. A possible way for doing so is to scale PE and APE by their standard deviation.

The impact of temperature and salinity on the norm becomes important where the Brunt-Väisälä frequency is small. This can be interpreted in physical terms: where N^2 is small, a slight error in temperature or salinity will have an important effect on the dynamics since it corresponds to a large displacement of the isopycnals.

The norm expressed as surface and as volume integrals has also the desirable property that its value is independent of the chosen spatial discretisation. In our implementation, the discretisation is neither regular in the vertical nor in the horizontal due to the grid refinement. The norm makes allowance for the different surfaces and volumes of the grid cells. Furthermore, it is not trivial to combine into a consistent norm two- and three-dimensional variables. Indeed, the impact of one surface point affecting the whole water column has to be compared to the impact of the temperature and the salinity of a single grid box. The norm (6.9) scales the 2D elevation field and the 3D temperature and salinity field in a consistent way.

The grid refinement introduces another subtlety. The state vector contains the grid points in the Provençal Basin twice: once at coarse resolution for the Mediterranean Sea model and once at intermediate resolution for the Provençal Basin model. In order to avoid counting twice the model error in these regions, the weights of the Provençal Basin grid points at coarse resolution are put to zero during the EOF calculation.

Figure 6.5 shows the Mediterranean Sea components of the first three EOFs obtained by the ensemble. The elevation of the ensemble shows a high variability especially at the coast along the basin-wide currents. Since the geostrophic transport of these currents is related to the slope of the elevation, the elevation variability can be viewed as variability in transport. The conservation of the water volume imposes a high correlation of the elevation along the path of the current. The constraint of orthogonality between the EOF modes is responsible for the fact that different branches of the coastal current have successive non-zero values.

The first three temperature EOFs show very large scale structures at the surface, having some resemblance with the first Fourier modes. In fact, if a covariance matrix $P(\mathbf{x}, \mathbf{y}) = f(\mathbf{x} - \mathbf{y})$ is translation invariant, one can easily show that the eigenfunctions of the covariance matrix are the Fourier modes:

$$\int_V P(\mathbf{x}, \mathbf{y}) \exp(i\ \mathbf{k} \cdot \mathbf{y})\, d\mathbf{y} = \exp(i\ \mathbf{k} \cdot \mathbf{x}) \int_V f(\mathbf{t}) \exp(i\ \mathbf{k} \cdot \mathbf{t})\, d\mathbf{t} \tag{6.10}$$

To some extent the error covariance in temperature is similar for different positions in the Mediterranean Sea. This is due to the fact that the surface temperature is largely influenced by the air temperature at 2 m and by the cloud coverage. The prescribed error covariance of the atmospheric forcings is indeed translation invariant.

In the contrary, salinity was not directly perturbed by the atmospheric forcings (since the freshwater flux was not perturbed). The EOF modes contain therefore the indirect dynamical response to the atmospheric forcings and the remaining perturbation of the

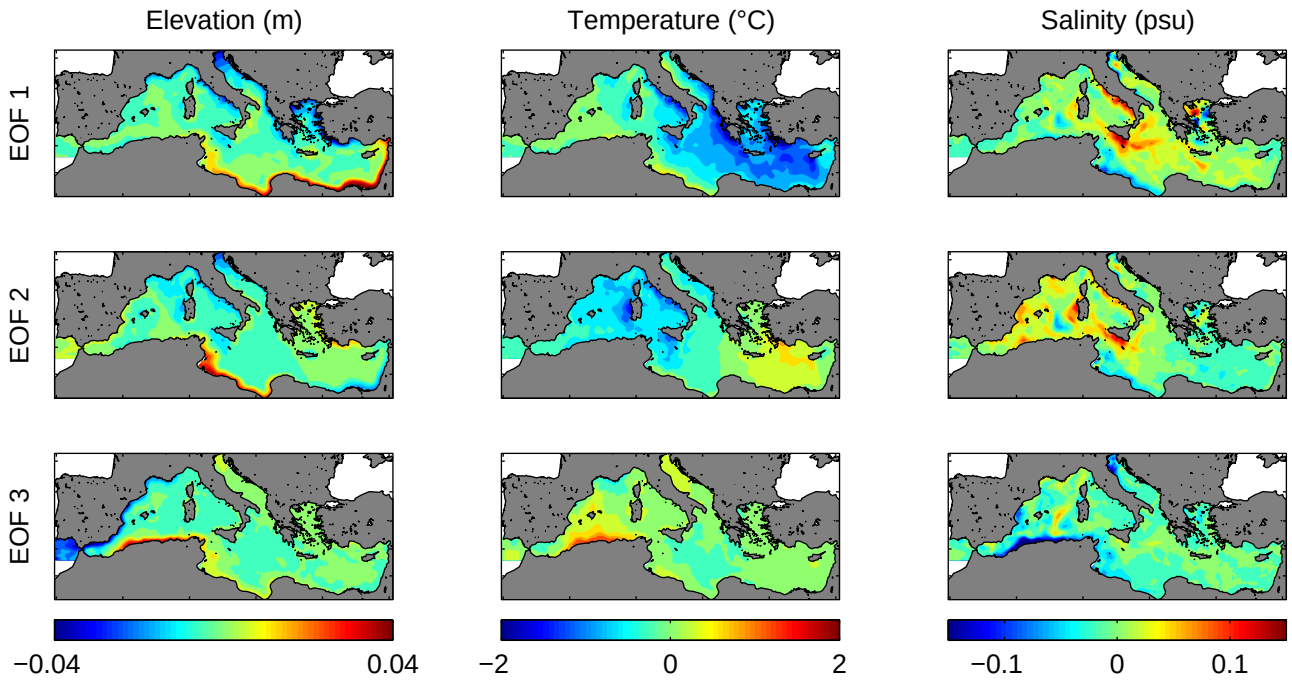

Figure 6.5: The 3 leading multivariate EOFs of the Mediterranean Sea model showing the elevation, surface temperature and surface salinity. The normalised EOFs were multiplied by the square root of the eigenvalues such that the resulting field is a dimensional quantity. EOF components on the Provençal model grid are not shown.

initial conditions.

It was verified that the multivariate EOF modes are close to the following balance:

$$\Delta\zeta = \int_{-h_0}^{0} \alpha\Delta T - \beta\Delta S \, dz \tag{6.11}$$

A positive temperature and a negative salinity anomaly are therefore associated to a positive elevation anomaly. It is important to validate physically the EOFs modes when they are used as basis functions of the error subspace in a data assimilation context. The *a posteriori* state will be a linear combination of the EOFs modes. Unphysical EOFs would leave the system in an unbalanced state after assimilation of observations.

6.3 Implementation

6.3.1 State vector

The state vector for the assimilation is composed of the elevation, temperature and salinity in all nested grids. Since the coarse grid contains grid points at the same geographic locations than next finer nesting level, the state vector contains some redundancy. These elements are linearly related to corresponding elements of the next

finer resolution nesting level by the feedback procedure of the nesting method. It is not necessary to exclude the redundant grid points, as long as their relationship is correctly represented by the error covariance matrix.

Novel in the choice of the state vector is that these three variables from all nested model grids are assembled into one multigrid state vector. This implementation allows us to take into account the correlation of the variables across the nested model grids in order to avoid, for example, artificial gradients at the nesting boundary after an assimilation cycle.

The velocity is not present in the state vector since the error covariance between surface elevation, temperature and salinity were prescribed by the geostrophic balance (based on a linearised state equation) and no velocity observations are used. This assumption was also made by *e.g.* De Mey and Benkiran (2002), Haines (2002) and Brankart *et al.* (2003). One can show that the analysis can be made with a state vector containing only elevation, temperature and salinity if the analysis increment of the horizontal velocity is obtained by the geostrophic balance. Near the coast, the velocity of the model was not corrected by a geostrophic current because of the nonlinear interactions with the coastline geometry.

The turbulent kinetic energy is discarded because of its highly nonlinear dependence on the other variables. Also, the fact that the eddy kinetic energy is always positive cannot be taken into account by an analysis scheme based on Gaussian statistics. Furthermore the kinetic energy adjusts itself rapidly to the local shear and stratification. However, studies (*e.g.* Korres *et al.*, 2004) exist where the potential benefit for analysing also the variables of the turbulence closure scheme was stressed.

6.3.2 Error space

The 50 leading modes obtained from the multivariate and multigrid EOF decomposition have been used as a basis of the background error space. The components of this error space basis for the Ligurian Sea have been linearly interpolated from the EOFs of the Provençal basin. As a consequence, the error modes are no longer guaranteed to be orthogonal but this property is not necessary for the analysis scheme.

The covariance represented by the 50 EOFs is less accurate than the covariance obtained from the 200 ensemble members due to the truncation of the EOFs series. However, the covariance matrix based on the EOFs is by construction the closest (in spectral and Frobenius norm) to the initial ensemble covariance among all possible reduced rank covariance matrices. But a crucial consequence of the reduced rank representation is the fact that only the model error components lying inside this space can be corrected.

6.3.3 Covariance localisation

For representing accurately small correlations, a large number of ensemble members and a high dimensional error space are needed. Small error covariances obtained from a limited number of EOFs are therefore unreliable and they are likely to degrade the

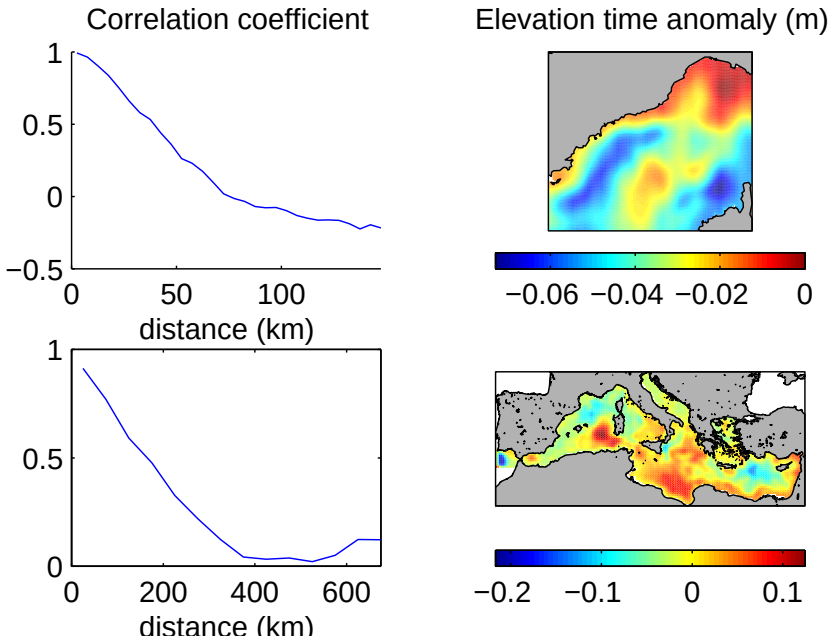

Figure 6.6: Correlation coefficient of the elevation time anomaly (model snapshot minus a two-year average of the model) as a function of distance. The upper and lower part shows the correlation coefficient of the Ligurian Sea and Mediterranean Sea of a model snapshot.

analysis. It is generally assumed that the covariance between two points decreases with the horizontal distance. The assimilation scheme is implemented in such a way that horizontal correlations beyond a certain horizontal distance h_c are neglected. Thus, observations outside this radius do not have an impact on the water column inside this circle. Therefore, the assimilation is done water column per water column as in Brankart *et al.* (2003) using only the subset of observations lying inside the influence radius. The continuity of the analysis is ensured by decreasing the weight $\mathbf{R}^{-1}$ of observations outside the influence circle progressively to zero.

This method enhances also the flexibility of the analysis to reach any state. The EOFs amplitudes of the analysis increment are no longer constant over the whole domain but can vary in space. A similar approach has also been implemented in the EnKF (Evensen, 2003). The problem with this approximation is that errors with scale greater than h_c can only be corrected with a sufficiently dense observational network.

The correlation distance h_c has been determined by the correlation length of the spatial structures present in the elevation time anomaly. A correlation distance of 47 km has been chosen for the Ligurian Sea. This length increases gradually to 200 km on the Mediterranean Sea model grid (figure 6.6).

Coarse grid model	Fine grid model	Interaction
assimilation	assimilation	no interaction
assimilation	no assimilation	no interaction
assimilation	no assimilation	interpolation from CGM to FGM
no assimilation	assimilation	no interaction
no assimilation	assimilation	averaging from FGM to CGM

Table 6.1: Different possibilities of data assimilation and interactions in a system formed by a coarse grid model (CGM) and a nested fine grid model (FGM).

6.4 Benefit of a unique multigrid state vector

The traditional approach of data assimilation with nested models is to use separate state vectors for the different models (*e.g.* Bell *et al.*, 2000; Martin *et al.*, 2002; Pinardi *et al.*, 2004; Onken *et al.*, 2003). For each model, one can decide whether or not to assimilate observations and if one model does not assimilate observations, it can be influenced or not by the model with assimilation. Five different scenarios are thus possible for data assimilation in a system composed of a coarse and an embedded fine resolution model. These possibilities are summarised in table 6.1. All these options can be applied in the frame of one-way or two-way nested models. Different approaches are currently tested (Vandenbulcke *et al.*, 2003, 2004) and these experiments are part of the MFSTEP research program.

The first and the third scenarios of table 6.1 are attractive since they allow us to assimilate observations of any location within the nesting system. For the two last methods only observations covering the fine grid model can be assimilated. The nested models are sensitive to the boundary conditions and it should be avoided to impose boundary conditions inconsistent with the dynamics of the nested models. This obvious problem of the second approach can be reduced by the third approach.

Let us consider the case where the assimilation is done in coarse and nested model separately. A point located on the coarse grid near the nesting boundary modifies the results of the coarse grid but it is not taken into account in the assimilation of the nested model. Obviously, this leads to inconsistencies along the nesting boundaries. To illustrate this problem, SST and SSH of the 27 July 2000 have been assimilated into the Ligurian Sea and Provençal Basin model forecast separately. Figure 6.7 shows the SST difference between the *a posteriori* states of both models in the Ligurian Sea. Far from the boundaries, both analyses lead to the same result but at the boundaries the difference reach 0.7 °C. This test was repeated for other days and the highest differences were generally observed at the corner of the Ligurian Sea domain. For geometrical reasons, the number of surface observations influencing this part of the domain is only a quarter of the number of observations affecting the *a posteriori* state of the Provençal basin model while at the boundaries far from the corner, this ratio of influencing observations is one half. The

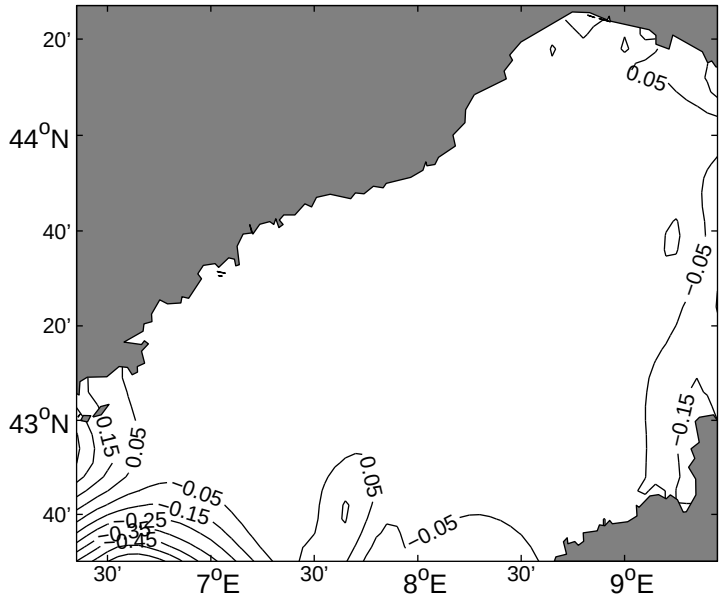

Figure 6.7: SST difference between the *a posteriori* state compute from a single state vector and separated state vector. Isoline interval is 0.1 °C. Maximum and minimum values are 0.3 °C and −0.7 °C respectively.

values at the boundary in figure 6.7 represent thus the discontinuity produced by the assimilation between the Ligurian Sea and the Provençal Basin. The other variables, observed or not, present similar discontinuities. It is not obvious how this inconsistency between the parent and nested models can be reduced. A possible *ad hoc* solution could be to combine linearly both fields in the overlapping domain where combination coefficients depend on the distance to the nesting boundary.

In the third method of table 6.1, the fine grid model is restarted from the interpolated *a posteriori* field of the coarse grid model. Apparently, there are no problems of inconsistencies between the coarse and the fine grid model. However, the small-scale information of the fine grid model is lost after the assimilation. This is an undesired characteristic especially for short assimilation cycles. Alternatively, only the analysis increment is interpolated and added to the fine grid state. However, this is only feasible for two-way nested models.

6.5 Observations

The model dynamics were constrained by two data types: sea surface temperature (SST) and altimetry (SSH). These two observations were assimilated each week in the system of nested models.

6.5.1 Sea surface temperature

The SST data sets used are weekly means of daily composites of NOAA-scenes processed by the remote sensing data center of the DLR (Deutschen Zentrum für Luft- und Raumfahrt). Several satellite swaths have been combined by taking the maximum temperature. Also daytime images have been used, which reduces the information for the subsurface layers. The sea surface temperature was assimilated at 14:00 GMT. The model sea surface temperature reaches its maximum on this diurnal cycle at this time.

The initial resolution of the SST is 1 km. Despite that the load of the analysis step is linear with respect to the number of observations, there is no interest to assimilate this high resolution temperature in the coarse resolution model. The resolution was degraded to 10 km where the high-resolution information in the SST is not expected to have a significant impact on the Ligurian Sea model. The SST data is therefore split into a 1 km resolution SST covering mainly the Provençal-Basin and a 10 km resolution SST of the rest of the Mediterranean Sea.

6.5.2 Sea surface height

The assimilated SSH are tracks from Topex/Poseidon and the ERS-2 mission (European Remote-Sensing). From the 31 August 2000 also tracks of the GFO mission (Geosat Follow-on) were assimilated. These data were obtained from the Naval Research Laboratory Stennis Space Center website. All data falling within the time window of one week were assimilated at the same time. The mean dynamic topography is obtained by a two-years free model run starting the 1 January 1998. The along track resolution is about 7 km. Often spatial scales shorter than 40 km are removed (*e.g.* Larnicol *et al.*, 2002). This step was not done here since the analysis step should filter out the information that is not in agreement with the model error statistics. The difference between the model and the observations contains therefore an important noise part and small scale variations that cannot be represented on the coarse model grid of the Mediterranean Sea.

The part of the signal with variations smaller than the model resolution can be quantified by a singular value decomposition of the observation matrix $\mathbf{H}$. This operator linearly interpolates the model elevation to the observations points. In the overlapping domains, the surface elevation with the highest resolution is taken. The singular decomposition of $\mathbf{H}$ can be written in the following way:

$$\mathbf{H} = \mathbf{U}_{\mathrm{H}} \mathbf{\Sigma}_{\mathrm{H}} \mathbf{V}_{\mathrm{H}}^{T} \tag{6.12}$$

with $\mathbf{U}_{\mathrm{H}}^{T}\mathbf{U}_{\mathrm{H}} = \mathbf{I}$, $\mathbf{V}_{\mathrm{H}}^{T}\mathbf{V}_{\mathrm{H}} = \mathbf{I}$ and where $\mathbf{\Sigma}_{\mathrm{H}}$ is a diagonal matrix containing only non-zero elements on its diagonal. The matrix product $\mathbf{U}_{\mathrm{H}}\mathbf{U}_{\mathrm{H}}^{T}$ is not equal to the identity matrix since the eigenvectors with a zero eigenvalue have been rejected. The mean square difference between the SSH observations $\mathbf{y}^{o}$ and the model counterpart $\mathbf{Hx}$ can be divided into two parts: the RMS error in the space formed by the columns of $\mathbf{U}_{\mathrm{H}}$ and the RMS error of the space orthogonal to the space formed by $\mathbf{U}_{\mathrm{H}}$.

$$\begin{aligned}\|\mathbf{y}^o - \mathbf{H}\mathbf{x}\|^2 &= \|\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o + \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x}\|^2 \\ &= \|\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o\|^2 + \|\mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x}\|^2 + 2(\mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x})^T(\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o) \\ &= \|\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o\|^2 + \|\mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x}\|^2 \qquad (6.13)\end{aligned}$$

This is the Pythagoras' theorem for orthogonal vector spaces. The two vectors $\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o$ and $\mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x}$ are orthogonal since:

$$(\mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{H}\mathbf{x})^T(\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o) = (\mathbf{U}_\mathrm{H}^T\mathbf{y}^o - \mathbf{\Sigma}_\mathrm{H}\mathbf{V}_\mathrm{H}^T\mathbf{x})^T\mathbf{U}_\mathrm{H}^T(\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o) = 0 \quad (6.14)$$

The first part of the error in equation (6.13) is only due to the scales that cannot be represented on the model grid. It is not possible that the model can get closer to the observations than this amount. The residual $\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o$ contains scales smaller than 25 km (the resolution of the Mediterranean Sea model) and can be treated as observational noise. Figure 6.8 shows the minimum RMS error for the different bins of SSH tracks. This minimum error should be considered when interpreting the RMS error between model and observations.

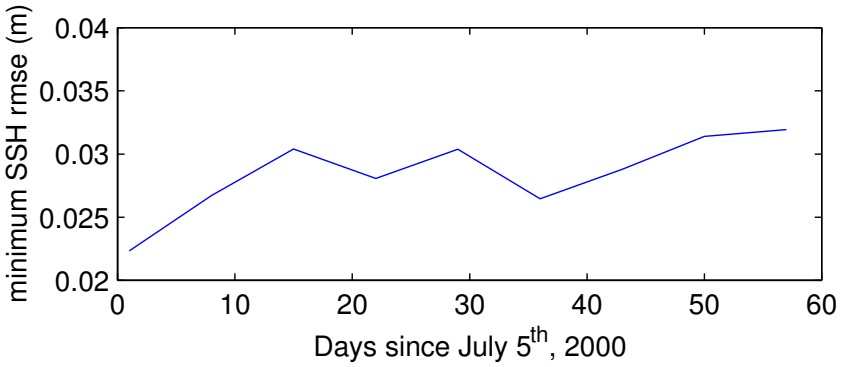

Figure 6.8: The minimum SSH RMS error $\|\mathbf{y}^o - \mathbf{U}_\mathrm{H}\mathbf{U}_\mathrm{H}^T\mathbf{y}^o\|^2$ between model and observations due to geometric constrains.

6.5.3 Observation operator

The model counterpart of this two SST data sets and the SSH tracks are obtained by bilinear interpolation of the models surface temperature layer and elevation respectively. Since the data assimilation scheme has been implemented to work on the top of the system of nested models, it can handle the fact that the interpolation may involve data from different grids. In the overlapping domains the information on the finest resolution grid is chosen.

6.6 Treatment of the observation error covariance matrix

The observation error is the sum of two different contributions: the instrumental error and the representativeness error. The latter is generally larger than the former. While it can be assumed that the instrumental error is not correlated in space, this is not true for the representativeness error. The correlation length scale of the representativeness error is related to the size of the unresolved processes.

For oceanic systems, the number of observations m is often larger than 10^4. If the correlation of the observations is taken into account, a matrix of the size $m \times m$ must be inverted. Depending on the specific formulation, this is either the observation error covariance matrix $\mathbf{R}$ or the sum $\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}$. This last matrix represents the covariance of the difference between the observation and their model counterpart $(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f)$. The inversion of such matrices is prohibitive since they are ill-conditioned and since the inversion is computationally intensive. The conditioning of the matrix $\mathbf{R}$ is related to the correlation length scale. For larger correlation lengths, the condition number[1] increases. The complete inversion of $\mathbf{R}$ requires $\mathcal{O}(m^3)$ operations. However, in the assimilation algorithm only products between $\mathbf{R}^{-1}$ and a vector must be computed. These products can be computed by solving the linear system for $\mathbf{y}$:

$$\mathbf{R}^{-1}\mathbf{b} = \mathbf{y} \Rightarrow \mathbf{R}\mathbf{y} = \mathbf{b} \tag{6.15}$$

The performance and accuracy of a linear solver depends on the condition number of the matrix $\mathbf{R}$. If the *a posteriori* error space $\mathbf{P}^a$ has to be computed, then r linear systems have to be solved, where r is the dimension of the error space.

6.6.1 Impact of correlated observations

Often the error correlations of observations are ignored during the analysis. In order to find alternative ways to take the spatial correlation of the observations into account, it is interesting to determine the impact of correlated observations and the consequences of neglecting this correlation.

The simplest system for assessing the impact of the correlated observations is constituted by a scalar state x^f with an error variance of P^f. The observation vector $\mathbf{y}^o$ contains two correlated direct measurements of the state x^f. The error covariance matrix $\mathbf{R}$ of $\mathbf{y}^o$ and the observation operator $\mathbf{H}$ have the following form:

[1] The condition number C is the ratio of the largest to smallest singular value in the singular value decomposition of a matrix. The base-10 logarithm of C is an estimate of how many digits are lost in solving a linear system with that matrix. In other words, it estimates worst-case loss of precision. A system is said to be singular if the condition number is infinite, and ill-conditioned if it is too large, where "too large" means roughly that $\log(C)$ is greater than the precision of matrix entries.

$$\mathbf{R} = \sigma_o^2 \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \tag{6.16}$$

$$\mathbf{H} = \begin{pmatrix} 1 & 1 \end{pmatrix} \tag{6.17}$$

The parameter c is the correlation between the observations y_1^o and y_2^o and σ_o^2 their error variance. The analysed state of this system is:

$$x^a = x^f + P^f \mathbf{H}^T \left(\mathbf{H} P^f \mathbf{H}^T + \mathbf{R}\right)^{-1} \left(\mathbf{y}^o - \mathbf{H} x^f\right) \tag{6.18}$$

$$= x^f + \frac{P^f}{\sigma_o^2(1+c) + 2P^f}(y_1^o + y_2^o) \tag{6.19}$$

This simple example shows that the correlation decreases the weight attributed to the observations.

6.6.2 Alternative schemes

Different ways can be considered for assimilating correlated observations with a diagonal observation error covariance matrix. Either the error covariance matrix $\mathbf{R}$ is modified or the observations.

From equation (6.19), we can see that a diagonal error covariance with an error variance inflated by $1 + c$ would yield the same analysis. In the case of m perfectly correlated observations, one can also show that the correlation can be ignored if the error variances is multiplied by m. In practice, the observations are not perfectly correlated. The correlation is described by a correlation function decreasing over at a given length scale. In a simplified view, the observations separated by a distance shorter than this correlation length can be considered as strongly correlated and the other observation pairs can be considered as uncorrelated. The number of observations r of a two-dimensional data set correlated to a single observation can be roughly estimated by:

$$r \sim \frac{L_x}{\Delta x} \frac{L_y}{\Delta y} \tag{6.20}$$

where L_x and L_y are the correlation lengths along the two dimensions of the data set and Δx and Δy the resolution. In summary, the correlation between the observations has the effect to reduce the weight of the observations during the analysis. If an existing correlation is neglected (for simplicity of the implementation for example), then the weight of the observations has to be decreased explicitly. This can be done by inflating the error variance of the observations by the average number of highly correlated and redundant observations. For a 2D dimensional data set this number can be estimated by equation (6.20).

If the two observations assimilated in the simple example described in section 6.6.1 are perfectly correlated ($c = 1$), then the assimilation of a single observation with the same error variance would give the same results. In this case, the second observation y_2^o does not provide any new information. This suggests that correlated observations can also be handled by excluding the redundant information from the observation vector. For a correlation length L, only observations separated by a distance of the order of L provide additional constraints to the model that can be considered as independent. The observations can for example be subsampled. In a 2D data set, this means that only every $r_x = \frac{L_x}{\Delta x}$ observations in the x-direction and only every $r_y = \frac{L_y}{\Delta y}$ observations in the y-direction are retained for the analysis.

In order to take all data into account, the observations can be binned into the so-called super-observations (*e.g.* Smith and Reynolds, 2003; Caires and Sterl, 2003). Every data bin represents then the mean of the observations within a rectangle of the size $L_x \times L_y$. Also, means based on distance depending weights can be considered.

Formally, the subsampling and the binning consists in replacing the observations $\mathbf{y}^o$ by the transformed observations $\mathbf{C}\mathbf{y}^o$. The operator $\mathbf{C}$ performs either the supsampling or the data binning. With the common notations, the analysis for these transformed observations can be written as:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T\mathbf{C}^T\left(\mathbf{C}\mathbf{H}\mathbf{P}^f\mathbf{H}^T\mathbf{C}^T + \mathbf{C}\mathbf{R}\mathbf{C}^T\right)^{-1}\left(\mathbf{C}\mathbf{y}^o - \mathbf{C}\mathbf{H}\mathbf{x}^f\right) \tag{6.21}$$

The observation operator $\mathbf{H}$ and the observation error covariance matrix $\mathbf{R}$ have been changed into $\mathbf{CH}$ and $\mathbf{CRC}^T$ respectively. This transformed observation error covariance is assumed to be diagonal.

6.6.3 Comparison of the different schemes

Different methods for assimilating correlated observations as uncorrelated ones have been proposed in the previous section. The impacts of these simplifications are tested in an idealised experiment. The model state $\mathbf{x}$ is a two dimensional field of the variables x and y and of a 2×2 units size. This field is discretised on a $n_x \times n_y$ grid ($n_x = n_y = 32$). The true state is shown in figure 6.9a. This true field is supposed to be unknown and is used for the assessment of the different schemes. A forecast is shown in figure 6.9b. It has been obtained by adding a random error to the true field. This error follows a Gaussian probability density function (pdf) with zero mean and the covariance between two points (x_1, y_1) and (x_2, y_2) is given by $P(x_1, y_1, x_1, y_2)$.

$$P(x_1, y_1, x_1, y_2) = 0.01 \exp\left(-\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{0.3^2}\right) \tag{6.22}$$

The complete true field is observed. The observations are thus also fields of size $n_x \times n_y$. But the observations are degraded by an error following a Gaussian pdf with zero mean and the covariance between two points (x_1, y_1) and (x_2, y_2) is given by $R(x_1, y_1, x_1, y_2)$.

$$R(x_1, y_1, x_1, y_2) = \sigma^{o2} \exp\left(-\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{(4\Delta x)^2}\right) \tag{6.23}$$

where $\sigma^{o2} = 0.01$ is the error variance and $\Delta x = 1/16$ is the resolution of the field. The correlation of the observations is thus the same in x and y directions and the correlation length scale L is of the order of $4\Delta x$ and the redundancies in both directions are $r_x = r_y = 4$. A realisation of the observations is shown in figure 6.9c as well as the corresponding *a posteriori* state (figure 6.9d) obtained with the error covariance (6.23).

The following experiment with different observation error covariance matrixes $\hat{\mathbf{R}}$ and/or observation vectors $\hat{\mathbf{y}}^o$ were carried out. For simplicity, correlation length and the size of the domain have been chosen such that r_x, r_y, $\frac{n_x}{r_x}$ and $\frac{n_y}{r_y}$ are integer values.

Non-diagonal R: The observation error covariance matrix is the true non-diagonal matrix given by equation (6.23) and all observations are assimilated. The inverse of $\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}$ is approximated by the pseudoinverse.

$$\hat{\mathbf{y}}^o = \mathbf{y}^o \tag{6.24}$$
$$\mathbf{C} = \mathbf{I} \tag{6.25}$$
$$\hat{\mathbf{R}} = \mathbf{R} \tag{6.26}$$

Since the observations are not changed, the transformation operator $\mathbf{C}$ is the identity matrix. It is defined in order to treat all experiments in a uniform way.

Without covariances: All observations are assimilated and the error covariance between two observations is neglected. The error variance is equal to the true error variance.

$$\hat{\mathbf{y}}^o = \mathbf{y}^o \tag{6.27}$$
$$\mathbf{C} = \mathbf{I} \tag{6.28}$$
$$\hat{\mathbf{R}} = \sigma^{o2}\mathbf{I} \tag{6.29}$$

Inflated R: All observations are assimilated and the observation error covariance is diagonal. But the error variance of the observations is multiplied by $r_x r_y$.

$$\hat{\mathbf{y}}^o = \mathbf{y}^o \tag{6.30}$$
$$\mathbf{C} = \mathbf{I} \tag{6.31}$$
$$\hat{\mathbf{R}} = r_x r_y \sigma^{o2}\mathbf{I} \tag{6.32}$$

Subsampled observations: Only every r_x observation in the x-direction and every r_y observation in the y direction are assimilated. Since the subsampling operator $\mathbf{C}$ acts on a two dimensional fields of size $n_x \times n_y$ and it returns a subsampled field of size $\frac{n_x}{r_x} \times \frac{n_y}{r_y}$, its elements can be referred by four indexes.

$$\begin{aligned} \mathbf{C}_{i_1 j_1 i_2 j_2} &= 1 \quad \text{if } r_x(i_1-1) = (i_2-1) \text{ and } r_y(j_1-1) = (j_2-1) \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{6.33}$$

The assimilated observations and their assumed error covariance are thus in this experiment given by the following expression.

$$\hat{\mathbf{y}}^o = \mathbf{C}\mathbf{y}^o \tag{6.34}$$
$$\hat{\mathbf{R}} = \sigma^{o2}\mathbf{I} \tag{6.35}$$

The error covariance matrix is assumed diagonal and its diagonal elements are equal to the error variance of the true observation error covariance matrix.

Binned observations: The mean observations of a $r_x \times r_y$ rectangle are assimilated. The elements of the $\frac{n_x}{r_x} \times \frac{n_y}{r_y} \times n_x \times n_y$ binning operator can be written as[2]:

$$\begin{aligned} \mathbf{C}_{i_1 j_1 i_2 j_2} &= \frac{1}{r_x r_y} \quad \text{if } i_1 - 1 = \text{floor}\left(\frac{i_2-1}{r_x}\right) \text{ and } j_1 - 1 = \text{floor}\left(\frac{j_2-1}{r_y}\right) \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{6.36}$$

The binned observations $\hat{\mathbf{y}}^o$ are assimilated as the subsampled observations with a diagonal observation error covariance matrix.

$$\hat{\mathbf{y}}^o = \mathbf{C}\mathbf{y}^o \tag{6.37}$$
$$\hat{\mathbf{H}} = \mathbf{C}\mathbf{H} \tag{6.38}$$
$$\hat{\mathbf{R}} = \sigma^{o2}\mathbf{I} \tag{6.39}$$

Binned observations with correct $\hat{\mathbf{R}}$: This is a variant of the previous experiment. Instead of a diagonal observation covariance matrix, the true observation covariance matrix of the binned observations is used.

$$\hat{\mathbf{R}} = E\left[\left(\hat{\mathbf{y}}^o - \mathbf{C}\mathbf{H}\mathbf{x}^t\right)\left(\hat{\mathbf{y}}^o - \mathbf{C}\mathbf{H}\mathbf{x}^t\right)^T\right] = \mathbf{C}\mathbf{R}\mathbf{C}^T \tag{6.40}$$

This matrix is generally non-diagonal but it is smaller and better conditioned than the original error covariance matrix $\mathbf{R}$. The inversion of $\hat{\mathbf{R}}$ is thus computationally less intensive than the inversion of $\mathbf{R}$. The assimilated observations and the binning operator are the same than in the previous experiment.

[2] The function floor(x) returns the largest integer that is less than x.

For all these experiments, the *a posteriori* state $\mathbf{x}^a$, the gain $\mathbf{K}$ and the *a posteriori* error covariance $\mathbf{P}^a$ are computed by the standard Kalman filter analysis based on the modified observations and observation error covariance.

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\hat{\mathbf{y}}^o - \hat{\mathbf{H}}\mathbf{x}^f\right) \tag{6.41}$$

$$\mathbf{K} = \mathbf{P}^f\hat{\mathbf{H}}^T\left(\hat{\mathbf{H}}\mathbf{P}^f\hat{\mathbf{H}}^T + \hat{\mathbf{R}}\right)^{-1} \tag{6.42}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\hat{\mathbf{H}}\mathbf{P}^f \tag{6.43}$$

where $\hat{\mathbf{H}} = \mathbf{CH}$. However, the analysis is only optimal if the matrix $\hat{\mathbf{R}}$ is the true error covariance of the observations. This is only the case for the experiments with "non-diagonal $\mathbf{R}$" and "Binned observations with correct $\hat{\mathbf{R}}$". Also the error covariance $\mathbf{P}^a$ is well estimated only for these two experiments since its expression is derived for the correct error covariances and not for the approximated error covariance used in the experiments.

The real error covariance of the *a posteriori* state obtained by equation (6.41) can be derived. By subtracting the true state from equation (6.41), one obtains the following expression for the error of the *a posteriori* state:

$$\mathbf{x}^a - \mathbf{x}^t = \left(\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}\right)\left(\mathbf{x}^f - \mathbf{x}^t\right) + \mathbf{K}\left(\mathbf{C}\mathbf{y}^o - \hat{\mathbf{H}}\mathbf{x}^t\right) \tag{6.44}$$

The covariance of this error is given by:

$$\mathbf{P}_r^a = E\left[\left(\mathbf{x}^a - \mathbf{x}^t\right)\left(\mathbf{x}^a - \mathbf{x}^t\right)^T\right] \tag{6.45}$$

$$= \left(\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}\right)\mathbf{P}^f\left(\mathbf{I} - \mathbf{K}\hat{\mathbf{H}}\right)^T + \mathbf{K}\mathbf{C}\mathbf{R}\mathbf{C}^T\mathbf{K}^T \tag{6.46}$$

Note that the last term in this equation depends on the true error covariance $\mathbf{CRC}^T$ and not on the approximated error covariance $\hat{\mathbf{R}}$. This expression for the *a posteriori* error covariance holds for any gain matrix $\mathbf{K}$, whereas the equation (6.43) is strictly speaking only valid for the observation error covariance corresponding to the observations. But assimilation schemes for ocean model implement error space updates based on equation (6.43). Therefore, it is also interesting to study possible the discrepancies between the error covariance obtained from equation (6.43) and from equation (6.46).

6.6.4 Results

The *a posteriori* state of all experiments has been computed for a particular realisation of $\mathbf{x}^f$ and $\mathbf{y}^o$. The difference between the *a posteriori* state with non-diagonal $\mathbf{R}$ and the truth is shown in figure 6.10a. It represents the smallest error that can be expected since the scheme is optimal. Figures 6.10b-6.10f show the difference between the *a posteriori* state of the other experiments and this optimal *a posteriori* state. Visually, the analysis schemes with an inflated $\mathbf{R}$, with subsampled observations and with binned observations

Experiment	$\mathrm{tr}\,(\mathbf{P}_r^a)$	$\mathrm{tr}\,(\mathbf{P}^a)$
non-diagonal $\mathbf{R}$	5.01	5.01
without covariances	8.65	0.73
inflated $\mathbf{R}$	5.57	4.30
sub sampled observations	5.84	4.47
binned observations	5.70	4.71
binned observations with correct $\mathbf{R}$	5.19	5.19

Table 6.2: This table summarises the accuracy of the analysis of the different experiments given by $\mathrm{tr}\,(\mathbf{P}_r^a)$. The last column $\mathrm{tr}\,(\mathbf{P}^a)$ shows the error estimated based on the assumption that the used observation error covariance is the right one.

perform all similar (figures 6.10f-6.10e). They have all a higher error in the central part of the field. The *a posteriori* state assimilating binned observations with a correct error covariance matrix is very close to the optimal analysis (figure 6.10f).

Table 6.2 shows the trace of the error covariance $\mathbf{P}_r^a$ and $\mathbf{P}^a$. The trace of the error covariance $\mathbf{P}_r^a$ is directly related to the RMS error that one would obtain if the experiments are repeated many times with different realisations of $\mathbf{x}^f$ and $\mathbf{y}^o$.

$$\mathrm{tr}\,(\mathbf{P}_r^a) = \mathrm{tr}\left(E\left[\left(\mathbf{x}^a-\mathbf{x}^t\right)\left(\mathbf{x}^a-\mathbf{x}^t\right)^T\right]\right) \tag{6.47}$$

$$= E\left[\sum_{\alpha=1}^{n}\left(\mathbf{x}_\alpha^a-\mathbf{x}_\alpha^t\right)^2\right] \tag{6.48}$$

where α is index for the state vector's components. It is not surprising that the best result is obtained when the true observation error covariance is used. Among the analysis schemes with a diagonal error covariance, the method consisting in inflating the error variance gives in average the most accurate *a posteriori* state. Simply ignoring the error covariance gives the worst result since the weight attributed to the observations is too high. The approach based on subsampling gives the second worst result. This is attributed to the fact that the rejected data contained also useful information. The approach with binned observations takes all observations into account and leads therefore to better results than the subsampling. Even better results than the inflation of the error covariance are obtained when binned observations are assimilated with the corresponding error covariance $\mathbf{CRC}^T$. This error covariance is not diagonal but it is smaller and better conditioned than the initial error covariance matrix $\mathbf{R}$. For an assimilation scheme capable of assimilating correlated observations, this last option is thus an interesting alternative to the inversion of a poorly conditioned $m \times m$ matrix.

As expected, this total error variance $\mathrm{tr}\,(\mathbf{P}^a)$ is equal to the error variance $\mathrm{tr}\,(\mathbf{P}_r^a)$ of the *a posteriori* state for the first and the last experiments. However, the error variances of

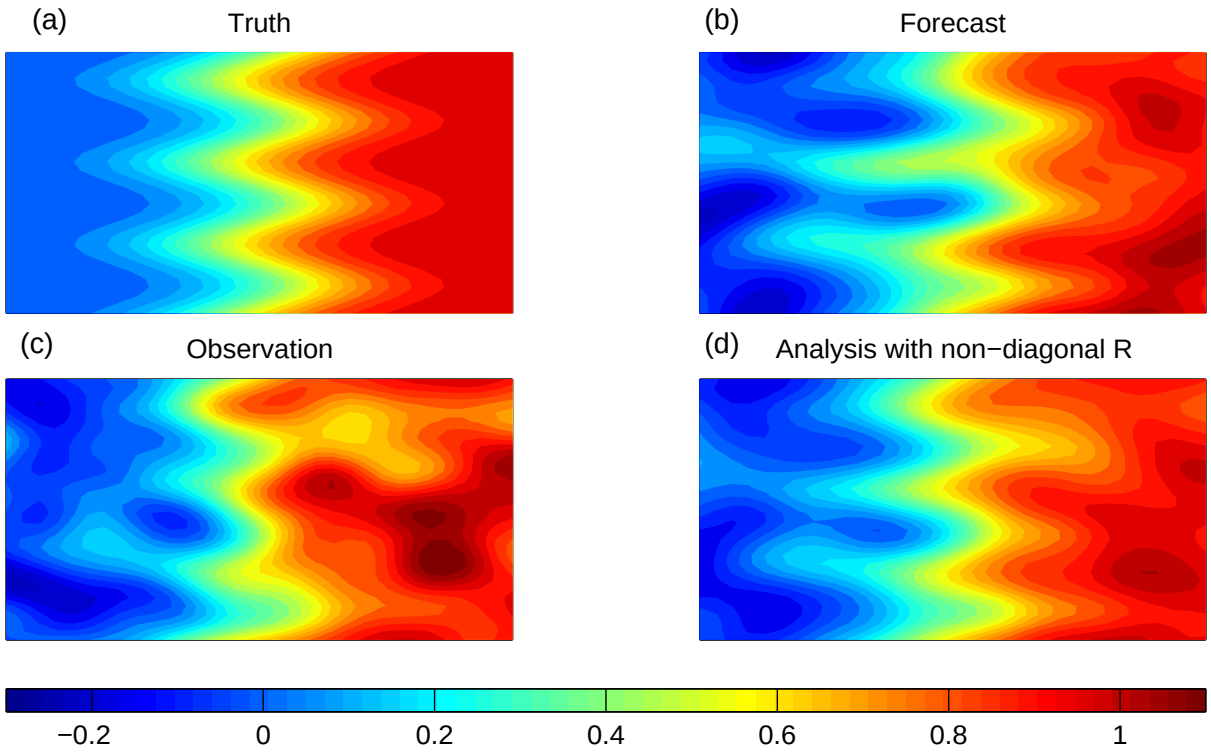

Figure 6.9: True, *a priori* , observed and *a posteriori* field of an experiment assimilation of synthetic data. The forecast and the observations are obtained by adding noise to the true field. The analysis combines forecast and observations based on their respective error covariances.

the experiments with diagonal $\mathbf{R}$ are generally underestimated. The discrepancy between the estimated error variance and the true error variance is the highest for the experiment ignoring the correlation.

For the assimilation method used in the present work the *a posteriori* error covariance is not used since the $\mathbf{P}^f$ is prescribed to be constant. For assimilation schemes propagating the error covariance from one assimilation cycle to the next, a correct *a posteriori* error covariance $\mathbf{P}^a$ is necessary. Further research is necessary to understand why these error variances are underestimated. A possible solution might be the fine tuning of the inflation factor or the subsampling and binning rate such that the *a posteriori* error variance is well represented.

6.7 Calibration

The success of a data assimilation experiment depends greatly on the specification of the error covariances of the model forecast and the observations. The forecast error covariance has been specified by the dominant error modes of an ensemble as it has been explained in section 6.1. The present assimilation implementation can handle only uncorrelated errors in observations.

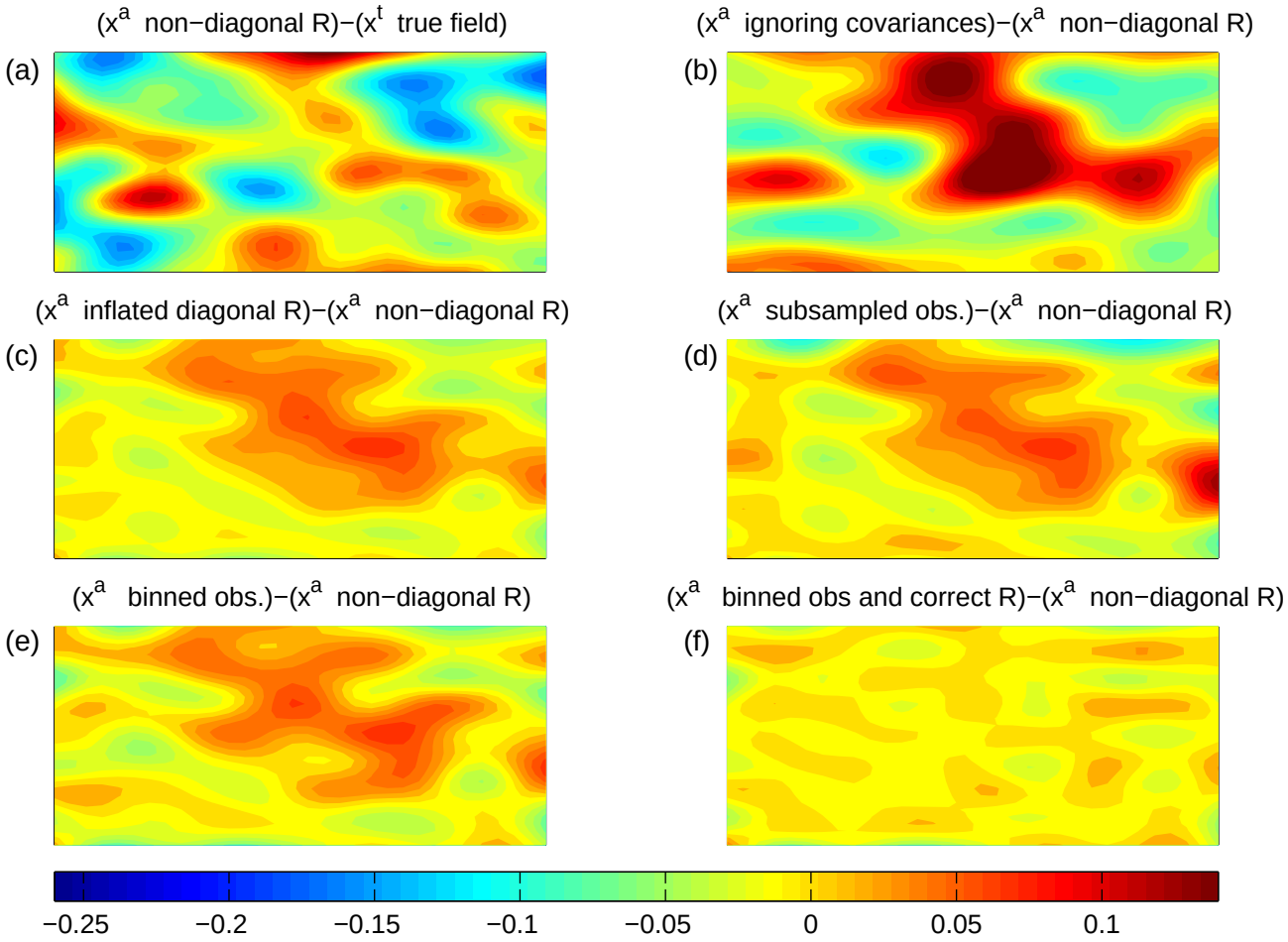

Figure 6.10: In panel (a), we see the difference between the optimal *a posteriori* and the true state. The other panels show how close the *a posteriori* states of the simplified schemes are to this optimal estimation.

The error variances of the observations have an important impact on the analysis. If the error variance is underestimated, the analysis deteriorates the unobserved variables. In the case of overestimation, the information in the observations is not sufficiently transferred into the model.

The assumption of uncorrelated errors might be verified for the instrumental error, but this is not the case for the representative error part. If the space correlation of the observations is not taken into account explicitly, the error covariance of the observations should be increased:

$$\mathbf{R}' = r \, \mathrm{diag}(\mathbf{R}) \tag{6.49}$$

where $1/r$ is the fraction of non-redundant scalar observations. For a 2D observation data set with a horizontal resolution of Δx and Δy, we can estimate this fraction by the horizontal correlation length L_x and L_y:

$$r \sim \frac{L_x}{\Delta x} \frac{L_y}{\Delta y} \tag{6.50}$$

A rationale for this parameterisation of the observation error covariance was given in section 6.6. The correlation in the observations can also be reduced by subsampling or binning the observations. But in this approach, the spatial resolution of the observations is degraded and their information content about the position of a front or an eddy for instance, is reduced.

6.7.1 Calibration of the 10 km SST and SSH

The error covariances of the 10 km resolution SST and the SSH were jointly calibrated. These data were assimilated in the Mediterranean Sea model with different observation error covariances without the nested models. The best error covariances are assumed to be those who produce the smallest RMS error between the SST and SSH observations and the one-week forecast. The observations error covariances are supposed to be diagonal and constant in space. Two variables have been calibrated, noted R_{SST} and R_{SSH} and two diagnostics, *i.e.* the mean forecast errors of the SST and SSH are computed. Table 6.3 summarises several experiments carried out with different values of R_{SST} and R_{SSH}.

	RMS SST	RMS SSH
Free model run	1.147 °C	8.60 cm
$\sqrt{R_{\mathrm{SST}}} = 1.25$ °C and $\sqrt{R_{\mathrm{SSH}}} = 3.75$ cm	0.913 °C	8.12 cm
$\sqrt{R_{\mathrm{SST}}} = 1.875$ °C and $\sqrt{R_{\mathrm{SSH}}} = 3.75$ cm	0.921 °C	8.10 cm
$\sqrt{R_{\mathrm{SST}}} = 2.5$ °C and $\sqrt{R_{\mathrm{SSH}}} = 3.75$ cm	0.937 °C	8.03 cm
$\sqrt{R_{\mathrm{SST}}} = 1.875$ °C and $\sqrt{R_{\mathrm{SSH}}} = 2.5$ cm	0.934 °C	8.13 cm
$\sqrt{R_{\mathrm{SST}}} = 1.875$ °C and $\sqrt{R_{\mathrm{SSH}}} = 5$ cm	0.918 °C	8.18 cm

Table 6.3: Different experiments for the calibration of R_{SST} and R_{SSH}. The row in grey marks the values adopted for the SST and SSH assimilation experiment.

The interaction between these two parameters is quite complex. A decrease of R_{SST} improves the SST forecast but it degrades the SSH forecast. For $\sqrt{R_{\mathrm{SST}}} = 1.875$ °C, the SSH forecast is optimal for $\sqrt{R_{\mathrm{SSH}}} = 3.75$ cm. Smaller and greater values for altimetry error covariance reduce the accuracy of the SSH forecast. On the other hand the best SST forecast has been obtained with the lowest SSH constrain. A reasonable compromise between SST and SSH forecast is obtained for the values $\sqrt{R_{\mathrm{SST}}} = 1.875$ °C and $\sqrt{R_{\mathrm{SSH}}} = 3.75$ cm. These error covariances are used in the following assimilation experiments.

The error covariances suggested by these experiments depend on various parameters, notably the data and their post-processing and the model error covariance. Therefore, it is expected that these error covariances of the observations cannot be applied without caution to other cases.

6.7.2 Calibration of the 1 km SST

The assimilation of the SST will modify the vertical structure according to the error covariances. Error on these covariances can degrade the vertical structure of the water column. This is mainly a problem of the error covariance between the SST and the sub-surface properties. But once the best effort has been made to quantify these covariances, we can reduce the problem of extrapolation based on uncertain error covariances by choosing carefully the error variance of the SST observations.

In order to assess the impact of the SST assimilation on the vertical temperature and salinity structure, the analysis was compared to temperature and salinity profiles of the Sirena 2000 cruise carried out by the Saclant Centre. The 1 km SST of the 24 August 2000 has been assimilated in the Ligurian Sea model with different error variances and the mean RMS error with the 33 temperature and salinity profiles within a centred time window of 1 week. The model background state of the 24 August 2000 has been obtained from a preliminary model integration with SST and SSH assimilation.

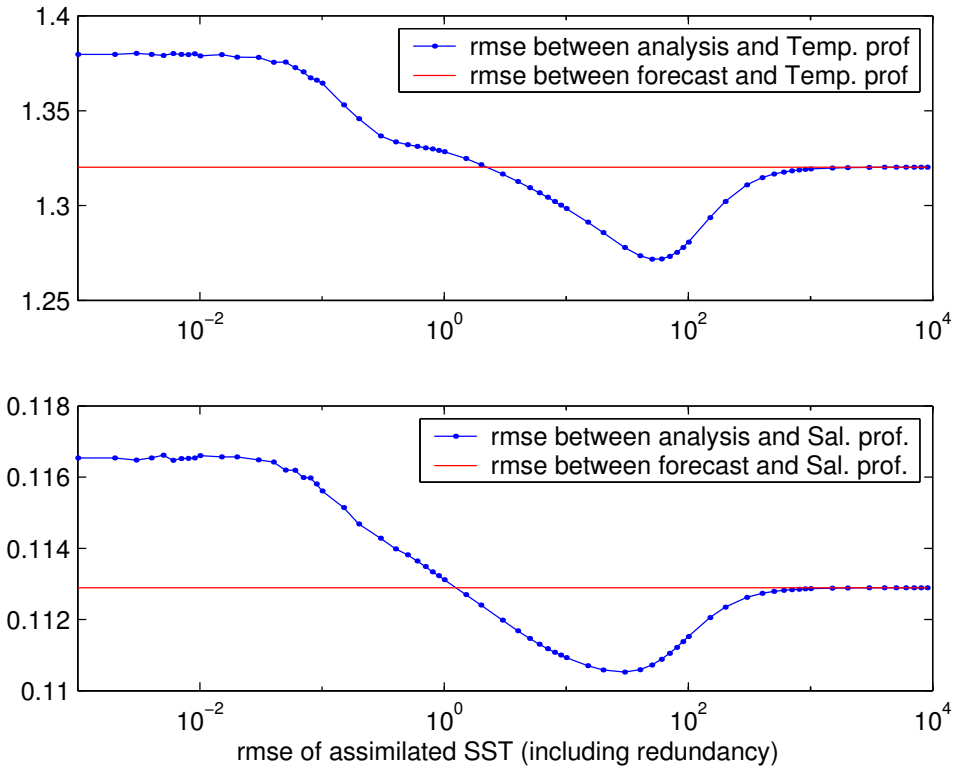

Figure 6.11: The RMS difference between the background and the 33 temperature (upper panel) and salinity profiles (lower panel) are shown as a continuous horizontal line. On each panel, the dotted line is the RMS difference between the analysis for different SST error variances and the measured profile.

For very high error variances, there is of course no difference between the background and the analysis. In the other extreme of error variance range, the assimilation degrades the vertical structure both in temperature and salinity. The error in density is minimum for $\sqrt{R'_{\mathrm{SST}}} = 70$ °C. This error variance includes redundancy in the data. If the error variance of SST observations (with explicit correlation) is assumed to be 1 °C^2, one obtains from equation (6.50) with $\Delta x = \Delta y = 1$ km and $L_x = L_y$, a correlation length of 70 km:

$$L_x = \Delta x \frac{\sqrt{R'_{\mathrm{SST}}}}{\sqrt{R_{\mathrm{SST}}}} = 70 \text{ km} \tag{6.51}$$

In the Ligurian Sea the correlation length of the SST is about 80 km. This correlation length represents an upper bound for the error correlation length since obviously only a part of the SST can be treated as observation error. The correlation length is therefore consistent with the optimised error variance including the redundancy.

6.8 Results

Figure 6.12 shows the SST of the 31 August 2000 obtained from the free model run (*i.e.* the model without assimilation) and the model run with assimilation as well as the observed SST. The SST of the model run without assimilation is too warm in the Ligurian Sea and too cold in the Eastern part of the Gulf of Lions compared to the observations. In particular, the water of Rhone plume is too cold and spreads over an extended area.

In the contrary, the forecast SST is closer to the observations. The mean temperatures of the Ligurian Sea and the part of the Gulf of Lions and the Tyrrhenian Sea match already well the observations. This is due to the beneficial impact of the previous assimilation cycles, which prevent the model to drift too far away from the observations. The *a posteriori* state combines the model forecast and the observed SST. The assimilation has changed the temperature distribution in the Ligurian Sea: warmer water masses are found near the coast and colder waters in the interior basin. This is indeed the typical temperature distribution in the Ligurian Sea. This temperature structure is also present in the free model run but it is about 1 °C warmer than the observed SST.

The RMS errors between the results of a model integration without assimilation and the 1 km SST of the Provençal Basin, the 10 km SST of the Mediterranean Sea and the altimetry tracks are shown in figure 6.13. The mean RMS error between observed SST and model SST without assimilation is 1.45 °C for the Provençal Basin and 1.14 °C for the rest of the Mediterranean Sea. The SST is very sensitive to the atmospheric forcings and the implementation of these forcings. The reasonable performance of the free model for reproducing the observed SST is attributed to the ECMWF forcings and the bulk formula parameterisation of the heat fluxes. Tests realised with NCEP forcings and non-interactive heat fluxes showed poorer results.

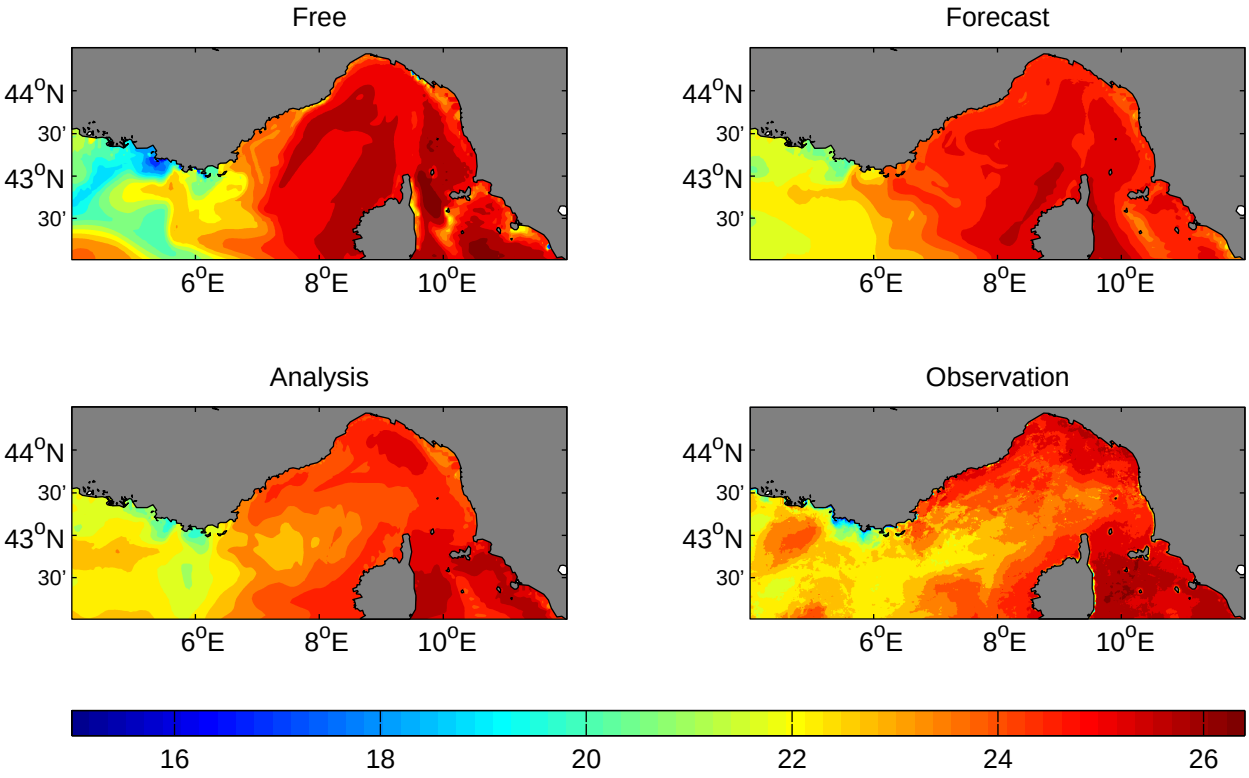

Figure 6.12: Sea surface temperature of the 31 August 2000 (9th assimilation cycle). The model results without assimilation (free), the model *a priori* state (forecast), the model *a posteriori* state and the assimilated observations are shown.

For the model run with assimilation, two estimations at each observation time exist: the forecast and the analysis. At each analysis step the model comes closer to the observations. Obviously, only the RMS error of the forecast is a comparison with independent data. This is not the case for the RMS error of the analysis. For all observations, the RMS error of the model forecast is in average lower than the RMS error of the free model run. This shows the beneficial impact of the previous assimilation cycles.

The SST forecast is always better than the SST of the free model except in the Provençal Basin for 24 August 2000 (8th assimilation cycle). This is related to a strong surface warming event in this region present in the observed SST and in the atmospheric fluxes. But the warming in the atmospheric fluxes is delayed compared to the observed SST. Even after the observed SST has reached its maximum, the atmospheric fluxes continue to warm the sea surface. The previous SST assimilation has caught up the effect of the too weak warming. The model with data assimilation overshoots the observed SST when the heating in the ECMWF forcing begins. In the contrary, the free model run, which is colder than the observed SST during the first half of August, comes fairly close to the observed SST after the atmospheric heating event. A delay in the atmospheric forcings is a time-correlated model error and in most assimilation approaches time-correlation is neglected.

For interpreting the errors of the sea surface height, the fact that the observations were not filtered should be taken into account. A part of the SSH signal variance is noise and increased the RMS error between the model and observations. A "perfect" model forecast would not produce a zero RMS error. The standard deviation of the noise with spatial scales smaller than the model grid size (figure 6.8) is a lower bound of the RMS error. Instead of filtering the data, we prefer to assimilate the raw data and to filter the observations by the assimilation procedure itself. But this implies that the RMS error contains also the variance of the observation errors, for which figure 6.8 provides an estimation.

During the first month, the SSH predictions of the model with data assimilation are similar to the free model run. The SSH forecast of the 3 August 2000 is even worse than the predicted sea surface elevation of the free model. But after this date, the improvement of the forecast due to the assimilation of the SSH is clear. The overall average RMS of the SSH forecast with assimilation is also better than the SSH forecast of the model run without assimilation. Further validation of the model can be found in Alvera-Azcárate *et al.* (2004).

6.9 Sirena Cruise

The model was also compared to the observations of the Sirena Cruise. This cruise was carried out by the Saclant Centre (La Spezia, Italy) starting on the 22 August 2000 until the 30 August 2000. 51 CTD cast were taken in the Ligurian Sea at the positions shown in figure 6.14. At some locations two casts separated by several hours were realised.

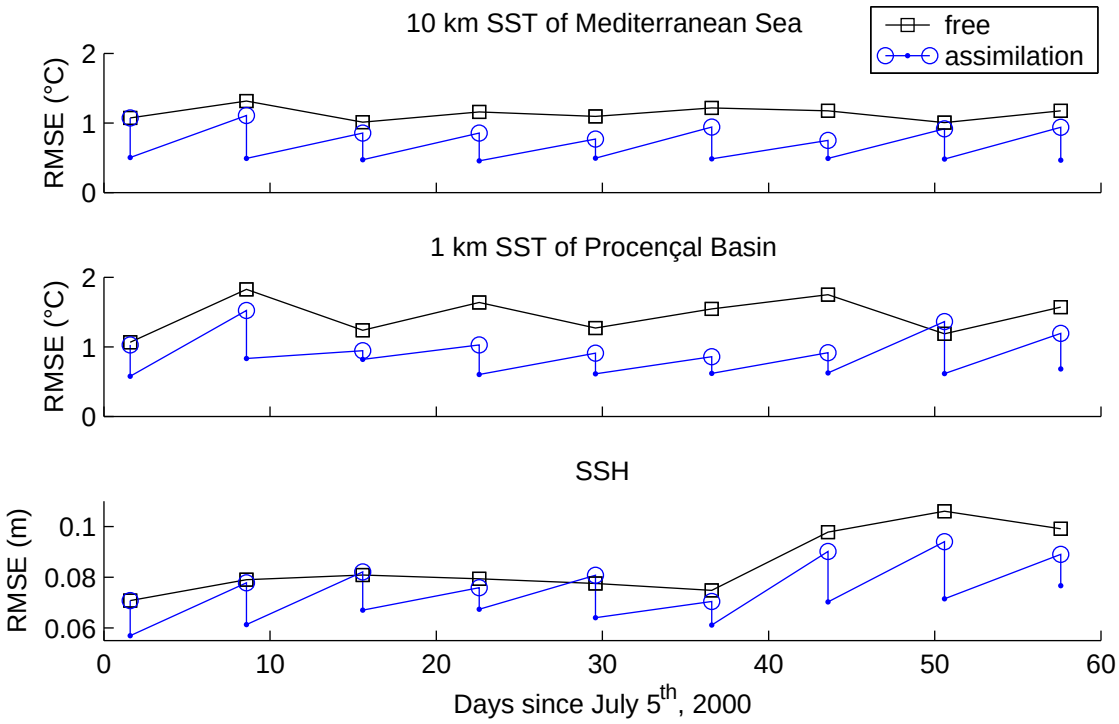

Figure 6.13: RMS error between the model results and the observations. "Free" is the model run without assimilation and "Assimilation" is the model integration with SST and SSH assimilation. The circles are the *a priori* estimations (or forecast) and the dots are the *a posteriori* estimations (or analysis).

The model temperatures of the corresponding day were spatially interpolated on the location of the profiles. Figure 6.14 shows the RMS error integrated over the depth of the model run with and without assimilation. The assimilation degrades the vertical structure of the model. Especially in the interior of the basin, the model with assimilation produces poorer results.

Figure 6.15 shows the RMS error of the free model run and the model simulations with SST and SSH assimilation. The error is mostly located at the thermocline. In fact, the thermocline of the model (with and without assimilation) is too deep in comparison with the observations. Since the thermocline represents a temperature jump of the order of 5 to 9 °C, a shift in the thermocline depth produces a large RMS error. In the free model, the mixed layer is too deep and the assimilation makes the mixed layer even deeper. This effect is even more critical in the interior of the Ligurian Sea where the overestimation of the thermocline depth is the highest.

These problems seem to be related to the covariance between the SST and SSH and the variables at depth. In the initial ensemble, the vertical correlation length of the temperature was prescribed to 20 m which is an estimation of the mixed layer depth during summer. Even after the integration of two weeks of each ensemble member this vertical correlation length is still dominant. However, in the Ligurian Sea during summer 2000 the mixed layer depth is as shallow as 10 m in some areas. The assimilation projects the surface temperature to a depth of 20 m and degrades the representation of the thermocline. Future research is necessary to have a better representation of the initial ensemble of members.

6.10 Surface velocity

Besides the statistical measures assessing the impact for the assimilation, it is also interesting to look at the impact on the oceanographic structures. Figure 6.16 shows the surface currents of the high resolution model with and without assimilation. The salinity is also displayed in order to locate the position of the fronts. The Northern Current is the main feature in the Ligurian Sea. Its baroclinic instabilities are the principal source of mesoscale activity. In the model run without assimilation, the Northern Current is detached from the coast at 43°40'N and 7°30'E. According to observations from Albérola *et al.* (1995) and Sammari *et al.* (1995), the Northern Current follows in this region the French coastline. This unrealistic detachment is accompanied with a long and straight front clearly visible in the salinity field. When the Northern Current of the free model reaches the southern boundary of the model, it gets back to the coast.

In the model run with assimilation, the path of the Northern Current is much better represented. It is confined to the coast as in the observations. Also meandering structures are reproduced by the model. The observed meanders have a typical wavelength of 30-60 km (Sammari *et al.*, 1995). The size of meanders reproduced by the model is in

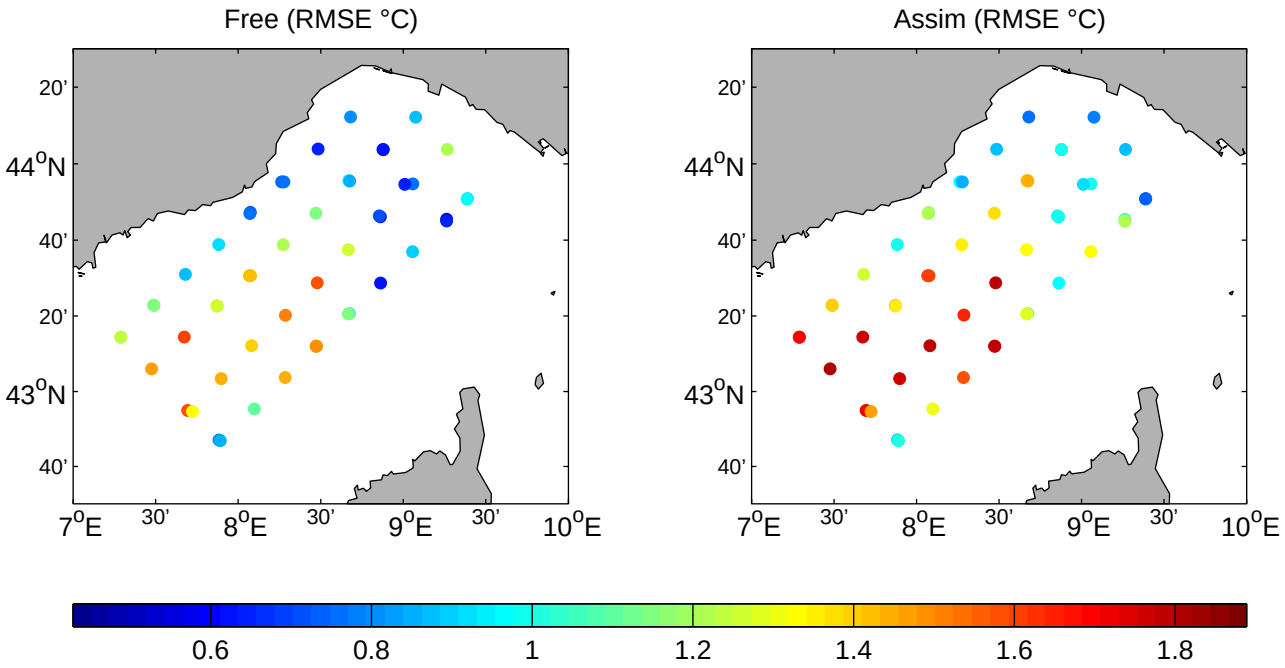

Figure 6.14: Horizontal distribution of the RMS error of the free model run and the model run with assimilation compared to the profiles of the Sirena Cruise. The RMS error is computed by summing over the depth.

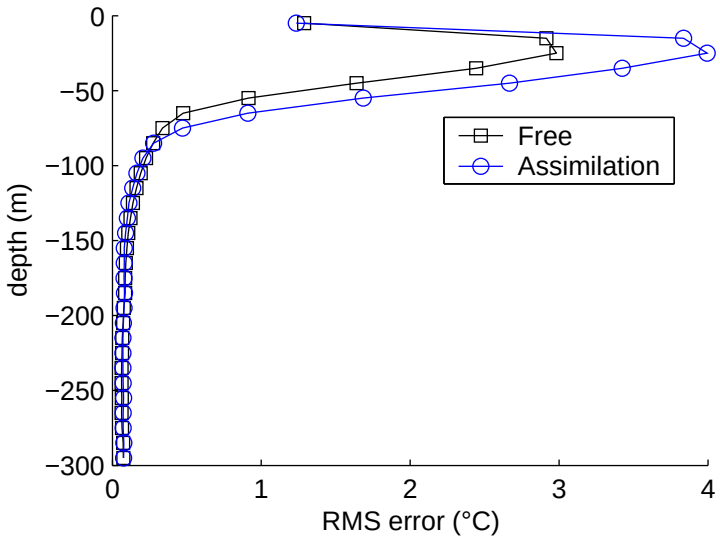

Figure 6.15: Vertical distribution of the RMS error of the free model run and the model run with assimilation compared to the profiles of the Sirena Cruise. The RMS error is computed by summing over all stations and over 10 m depth intervals.

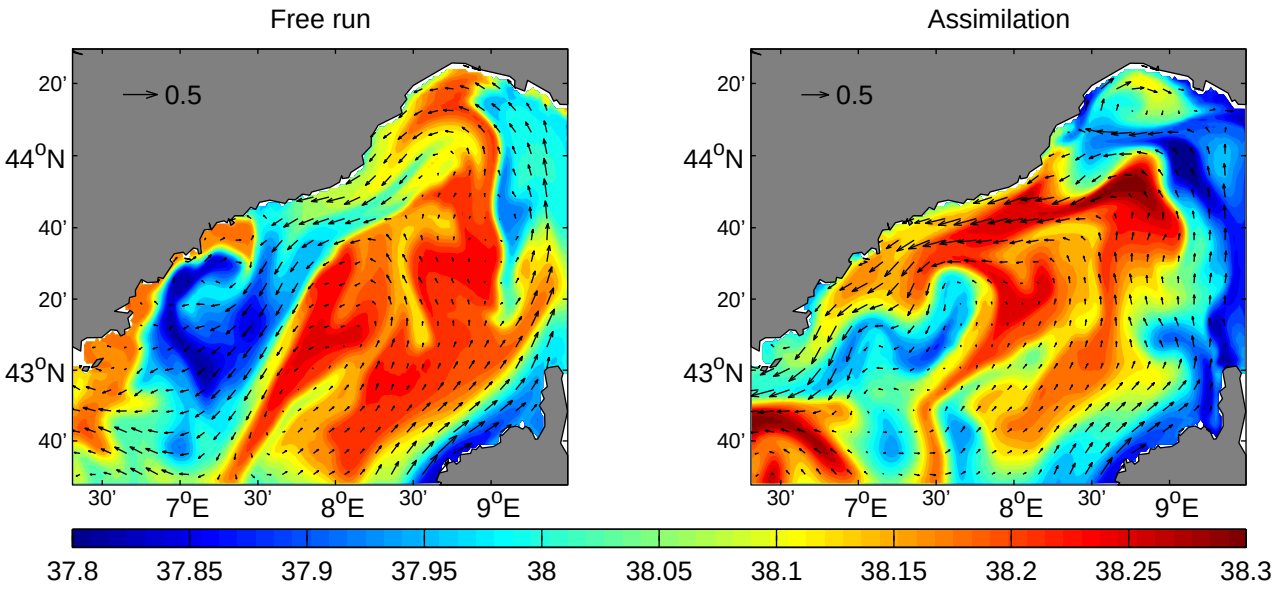

Figure 6.16: Surface velocity and salinity in the Ligurian Sea of the 25 July 2000.

good agreement with these observations. Velocity and salinity are not assimilated and the improvements visible on the figure 6.16 are thus the response of the model to SST and SSH assimilation. The assimilation has thus also a beneficial impact on the overall representation of the ocean processes in the Ligurian Sea.

6.11 Conclusions

The present chapter showed the application of a Kalman filter based assimilation scheme to a system of two-way nested models. Implementation aspects specific to the two-way nested model were highlighted and discussed.

Since the nested models are strongly coupled through the dynamics, the state vector for the data assimilation encompasses all nested models. The feedback between the nested models relates linearly the variables of the coarse model grid in the overlapping domain to the variables of the fine model grid. The redundancy in the state vector is not a problem as long as the covariance between the redundant variables and the others is correctly specified. This is the case when the covariance is estimated from model results either directly by an ensemble of model states or by the empirical orthogonal functions (EOFs) of this ensemble.

EOF decomposition needs the introduction of a norm in the state space. For variables with different units and discretised on a non-regular grid, the Euclidian norm is unacceptable since the result would depend on the units and discretisation chosen. A norm based on the sum of potential energy and available potential energy was first applied to the EOF computation. Model nesting has also an implication on the choice of the norm since the variables in the overlapping domain have to be counted only one time.

A novel method for choosing the error variance of the observations is also presented. This method resembles closely to the cross-validation technique. Unassimilated vertical temperature and salinity profiles are used to assess the impact of the assimilation of sea surface temperature (SST) at 1 km resolution. It has been shown that when the correlation in this data set is neglected, the weight given to the observations is by far too high. But the correlation in the observations can be included by inflating the weight of the error variance by the proportion of redundant data. The method for determining the error variance of observations includes this inflation factor.

The improvement of the SST and sea surface height (SSH) forecast due to the assimilation is also shown. For all weeks except one, the model with SST assimilation provides better SST forecasts in the Provençal Basin than the free model run. The forecast of the surface elevation is also improved in average. The comparison of forecasts and observations not yet assimilated helps to assess the beneficial impact of the previous assimilation cycles.

While the vertical structure is degraded by the assimilation, the circulation in the Ligurian Sea is better represented in the model run driven by the SST and SSH assimilation. The assimilation has corrected the path of the NC and displaced the Liguro-Provençal frontal system such that the model corresponds to the observed oceanographic characteristics of this zone.

Chapter 7

Hybrid modelling system

Data assimilation is traditionally used to combine model dynamics and observations in a statistically optimal way. Assimilation of observations improves therefore hindcasts and nowcasts of the ocean state than otherwise obtained by the model alone. The observational constraints are necessary to reduce uncertainties and imperfections of the ocean model. Due to the obvious lack of future observations, the model forecast cannot be controlled by observations and the predictive skill degrades as the forecast time lag increases. The error growth is not only caused by the chaotic nature of the system but also by the biases and drifts of the model. The later part can be reduced by considering different models with different imperfections. Data assimilation provides the statistical frame for merging different model results.

In this chapter, we will first introduce the statistical SOFT predictor (Álvarez *et al.*, 2004). The applied methodology as well as the accuracy of its forecast will be briefly discussed. Then the SST predicted by this SOFT system will be assimilated into the GHER ocean model. The skill of this hybrid modelling system composed of the GHER primitive equation model and the statistical predictor will be assessed. Finally, we will discuss the results and make conclusions in the perspective of operational forecasts.

7.1 Statistical Model

The system of nested models described in the chapter 2 and3 was coupled with the statistical model of Álvarez *et al.* (2004). This statistical model is applicable to ocean properties, which are measured at a high temporal and spatial coverage. Satellites sensors meet these coverage requirements. The statistical predictors, also called satellite-based ocean forecasting (SOFT) systems, have been applied to remote sensed surface properties such as sea surface temperature (SST) and sea level anomaly.

This model is based on an EOF decomposition and on a genetic algorithm. The EOF decomposition reduces significantly the degrees of freedom in the system and therefore the number of variables to predict.

7.2 EOF decomposition

The $m \times n$ data matrix $\mathbf{D}$ contains a time series of spatial fields. The first index i is the space index ($1 \leq i \leq m$) containing the two horizontal dimensions and the second index j is the time index ($1 \leq j \leq n$). The EOFs are not directly compute from this data matrix but from the anomalies $\mathbf{X}$ based on a given mean $\mathbf{M}$.

$$\mathbf{X} = \mathbf{D} - \mathbf{M} \tag{7.1}$$

where $\mathbf{X}$ and $\mathbf{M}$ are $m \times n$ matrices. The mean can be for example the time mean. In this case all columns of $\mathbf{M}$ are the same and are equal to the time average of the data:

$$M_{ij} = \frac{1}{n} \sum_{k=1}^{n} D_{ik} \qquad \forall j \in \{1, \ldots, n\} \tag{7.2}$$

The matrix $\mathbf{M}$ can also represent the space mean:

$$M_{ij} = \frac{1}{m} \sum_{k=1}^{m} D_{kj} \qquad \forall i \in \{1, \ldots, m\} \tag{7.3}$$

Another values of the mean $\mathbf{M}$, such as the time and space mean or a seasonal climatologic mean, can also be taken depending on the kind of variability which one wants to analyse by the EOF decomposition.

The anomalies $\mathbf{X}$ are decomposed into spatial and temporal variability patterns by a singular value decomposition:

$$\mathbf{X} = \mathbf{U}_\mathrm{X} \mathbf{\Sigma}_\mathrm{X} \mathbf{V}_\mathrm{X}^T \tag{7.4}$$

where $\mathbf{U}_\mathrm{X}$ is a $m \times m$ orthogonal matrix containing in its columns the spatial EOFs. The columns of the $n \times n$ orthogonal matrix $\mathbf{V}_\mathrm{X}$ are the temporal EOFs. The diagonal elements of the $m \times n$ matrix $\mathbf{\Sigma}_\mathrm{X}$ are the ordered singular values starting with the most important singular value. Other elements of $\mathbf{\Sigma}_\mathrm{X}$ are equal to zero.

The spatial EOFs are the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$. This matrix is, up to a factor, the spatial covariance matrix based on a time series of states if the time mean (7.2) was subtracted.

$$\mathbf{X}\mathbf{X}^T \mathbf{U}_\mathrm{X} = \mathbf{\Sigma}_\mathrm{X} \mathbf{\Sigma}_\mathrm{X}^T \mathbf{U}_\mathrm{X} \tag{7.5}$$

The data at a given time can be viewed as a linear combination of the spatial EOFs, which represent spatial patterns of variability:

$$D_{ij} = M_{ij} + \sum_{k=1}^{m} A_{kj} U_{\mathrm{X}ik} \tag{7.6}$$

where A_{kj} is the amplitude of the kth EOF at time j. The EOF amplitudes can be obtained directly by projecting the data at a given time on the spatial EOFs.

$$\mathbf{A} = \mathbf{\Sigma}_\mathrm{X}\mathbf{V}_\mathrm{X}^T = \mathbf{U}_\mathrm{X}^T\mathbf{X} \tag{7.7}$$

The time series formed by the EOF amplitudes are orthogonal vectors:

$$\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}_\mathrm{X}\mathbf{\Sigma}_\mathrm{X}^T \tag{7.8}$$

If the time mean has been subtracted to compute the anomalies $\mathbf{X}$, then the mean of each EOF amplitude time series is zero. In this case the product $\mathbf{A}\mathbf{A}^T$ is the time covariance between the different EOF amplitudes and we see that the amplitudes for different EOFs are not correlated in time. An EOF analysis can therefore help to identify different uncorrelated physical processes and to split the evolution of a system into these uncorrelated modes. It is particularly efficient when the dynamics of the system can be related to "standing waves".

However, the physical interpretation of the EOF reaches its limit since different processes at similar frequency cannot be separated. An EOF pattern can thus contain the signal of two processes that are dynamically not related. A high correlation does not imply a cause-effect-relationship. The dynamics of nonlinear systems such as the baroclinic ocean cannot in general be separated into modes due to the interaction between the different scales. The orthogonality constraint between the EOFs makes also the interpretation of the spatial pattern difficult.

In practice, the EOF analysis is limited to a small number r of the eigenvectors corresponding to the leading eigenvalues:

$$D_{ij} = M_{ij} + \sum_{k=1}^{r} A_{kj}U_{\mathrm{X}ik} + \varepsilon_{ij} \tag{7.9}$$

where ε_{ij} is the truncation error introduced by limiting the number of EOFs. One can show that no other of r dimensional set of basis functions can give a total truncation error $\|\varepsilon\|^2 = \mathrm{tr}(\varepsilon^T\varepsilon)$ less than the error obtained by the EOFs (see appendix E).

Other basis functions used in signal analysis are for example Fourier modes and wavelets. In oceanography, the irregular coastline makes however the application of spatial Fourier modes a difficult task. Wavelets have a compact support and the wavelet coefficients in the interior of the ocean are not affected by the boundaries. However, near the coast a special treatment is required (Alvera-Azcárate *et al.*, 2004).

For all characteristics of the spatial EOF a similar property can be derived for the temporal EOF since there is a perfect symmetry between temporal and spatial EOFs. In general, it is sufficient to apply the results to the transposed data matrix $\mathbf{D}^T$. The time average becomes a space average and vice versa.

Álvarez *et al.* (2004) carried out a truncated spatial EOF decomposition of remote sensed SST of the Ligurian Sea. The amplitudes of the EOFs were filtered by a singular spectrum analysis in order to extract the deterministic part of the signal. For a deterministic time

series $a_j, j = 1, \ldots, n$, the Takens' theorem (Takens, 1981) proves the existence of a smooth map $\gamma : \mathbb{R}^m \rightarrow \mathbb{R}$ relating the elements of the time series by:

$$a_{k+1} = \gamma(a_k, a_{k-1}, a_{k-2}, \ldots, a_{k-m}) \tag{7.10}$$

where m is the embedding dimension. Álvarez *et al.* (2004) determined this function γ for each EOF amplitude time series. The genetic algorithm approximating γ is called DARWIN (Álvarez *et al.*, 2000; Álvarez *et al.*, 2001) and is based on the past evolution of the studied system. The smooth map can also be approximated by neuronal networks (Rixen *et al.*, 2001a).

The EOF analysis of weekly SST from 1 March 1993, to 4 October 1999 serves as training set in order to obtain a statistical predictor of the SST in the Ligurian Sea. Once the function γ fits sufficiently well the evolution of this training data set, γ can be used to predict the EOF amplitudes. From the predicted amplitudes, the complete SST scene is reconstructed by using the spatial EOFs.

7.3 Application of the SOFT predictor

Operational models run basically in two modes: the hindcast mode and the forecast mode. They are illustrated in figure 7.1a. In the former mode, the model assimilates the available data to obtain a model state as accurate as possible at the end of the hindcast. When the model has reached the present time instant, the run turns into the forecast mode. In this phase, no observations constrain the model. Eventually climatologic data is assimilated in order to control the model integration. For example, the model SST is nudged towards the climatologic SST to correct the atmospheric heat fluxes. In the MERCATOR system, climatologic data are also assimilated in depth when the difference between model and climatology is too large (Bahurel *et al.*, 2004).

The climatology has some natural drawbacks. It does not contain extreme or rare events. If the model predicts such an event, the assimilation of the climatology will attenuate it. Real and unreal events will be affected. The SOFT predictor can be integrated in such operational models as shown on figure 7.1b. The model forecast can be constrained by predicted ocean properties. Here we focus on the prediction of SST. One question is if the assimilation of predicted SST gives more skilful forecast than the assimilation of climatologic SST.

7.4 Assimilation Experiments

In order to assess the impact of the assimilation of predicted SST, five different experiments were carried out. All the experiments start the 5 July 2000 and simulate the dynamics for 60 days. They use the same initial conditions and the atmospheric forcings.

`Free:` The system of nested models is run without data assimilation but with realistic initial conditions and ECMWF atmospheric forcings.

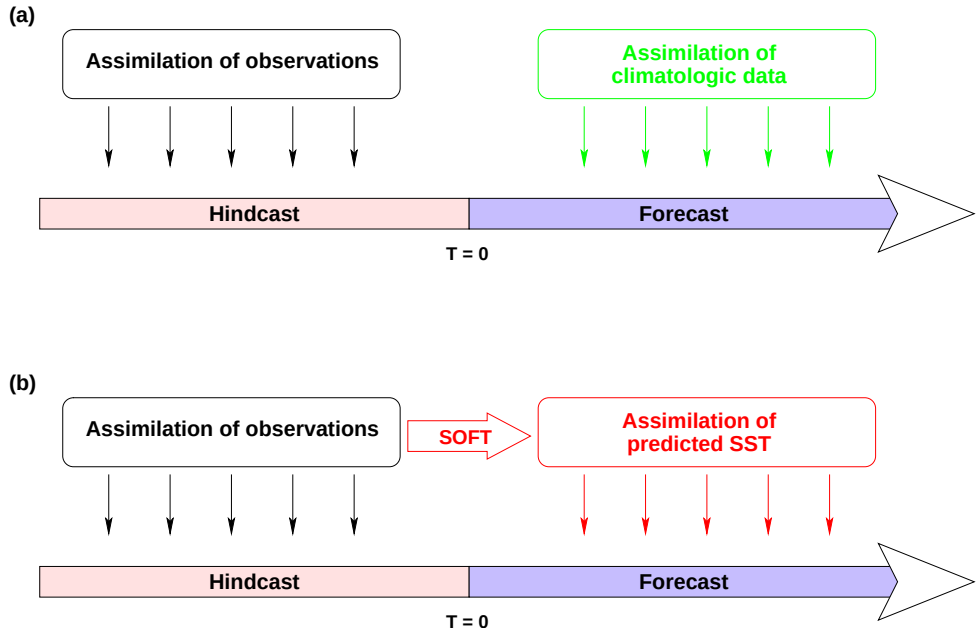

Figure 7.1: Application framework for the SOFT predictor. Instead of performing an unconstrained model forecast or a model forecast with assimilation of climatologic data (panel a), the forecast of the SOFT system can be assimilated into the operational model during the forecast (panel b).

`AssimObs`: The hydrodynamic model is constrained by weekly remote sensed SST from DLR. This model run intents to validate and to assess the capabilities of the system of nested models.

`AssimPred`: In this experiment, the hydrodynamic model assimilates the SST from the statistical predictor of Álvarez *et al.* (2004). The objective is to study model behaviour when it is constrained by the result of another model and to quantify the expected benefit of assimilating predicted SST in comparison with no assimilation and with the assimilation of observed SST. The 60-days model run only assimilates predicted SST but each SST prediction used the observed SST of the previous weeks. The model does not make a 60 days forecast without any influence of observations within these 60 days. But the model results of the 14 days following an analysis step do not depend on any observations within these two weeks.

`AssimAmpl`: The predicted SST is a linear combination of the dominant spatial EOFs. This implies that the rejected EOF components always have zero amplitude. The SST of the model, however, can contain components of these rejected EOFs. When assimilating the reconstructed EOF field (`AssimPred`) the variance of these components is reduced by the assimilation of the predicted SST that contains no variance at all in this space.
The approach in this experiment `AssimAmpl` consists in leaving these components as "unobserved" by assimilating directly the predicted EOF amplitudes and not the complete SST. The observation model makes therefore the following operation on

the state vector $\mathbf{x}^f$:

$$\mathbf{a}^f = \mathbf{U}_{\mathrm{X}}^T \left(\mathbf{H}\mathbf{x}^f - \mathbf{y}_m^{\mathrm{SST}} \right) \tag{7.11}$$

where the vector $\mathbf{a}^f$ contains the EOF amplitudes predicted by the hydrodynamic model and $\mathbf{y}_m^{\mathrm{SST}}$ is the annual mean SST in the Ligurian Sea used to compute the spatial EOFs $\mathbf{U}_{\mathrm{X}}$. The matrix $\mathbf{H}$ is the observation operator extracting the SST from the model state vector. This operator is the same than the one used in the two previous experiments.

`AssimClim`**:** In this experiment, the interpolated SST from the MEDAR climatology (Rixen *et al.*, 2001b) is assimilated into the model. For consistency with the other experiment only the SST of the Ligurian Sea is used.

All three assimilation experiments use the same error covariance matrix described in 6.1. The covariance localisation with the same parameters than in 6.3.3 was also used for the experiments `AssimObs` and `AssimPred`. The covariance localisation technique is based on the distance between the model grid point and each single scalar observation. However, EOF amplitudes are not local observations since they involve the temperature of all surface grid points of the domain. For the assimilation of the amplitudes, the original global assimilation scheme (4.161) was used.

7.5 The predicted SST compared to climatology and persistence

The SST predictor produced SST forecast of the Ligurian Sea from 11 October 1999, to 28 August 2000. The overall skill assessment of the method compared to persistence of the observations during this period was made by Álvarez *et al.* (2004). But it is useful to take a closer look to the performance of the predictor in the time interval of the data assimilation experiments since Álvarez *et al.* (2004) showed that the forecast skill of the predictor depends on the seasons. Figure 7.2 shows the RMS error of the predicted SST, the SST observed one week ago (persistence) and the climatologic SST compared with the observations. The monthly MEDAR climatology is interpolated linearly to obtain weekly SST of the Ligurian Sea.

During this time period the predicted SST is substantially better than the persistence or the climatology. The RMS error of the predictor is also better than the RMS error of the hydrodynamic model (figure 7.3), which justifies the assimilation of the predicted SST. The climatology is too cold, especially at the end of the studied period. Interannual variations have an important impact on the SST. These variations ignored by a monthly climatologic cause the large discrepancy between the climatology and the observations.

7.6 Results

The RMS error as a function of time between the observed SST and the five experiments are shown in figures 7.3 and 7.4. At each observation time, two estimates of the ocean state exists for the model run with assimilation: the model forecast ($\mathbf{x}^f$ in the standard notation) and the analysis ($\mathbf{x}^a$). The time averaged RMS error between the observed SST and both estimates for the four assimilation experiments and the free model integration are summarised in table 7.1.

	Forecast	**Analysis**
`Free`	1.111	-
`AssimObs`	1.031	0.580
`AssimPred`	1.037	0.732
`AssimAmpl`	1.063	0.883
`AssimClim`	1.096	1.479

Table 7.1: The time averaged SST RMS error of the different experiments

The RMS error between the observed SST and the analysis of the model run with observed SST assimilation, `AssimObs`, is obviously not a comparison with independent data. But all other RMS errors of table 7.1 are in fact skill assessments based on independent data since the model forecast at a given time has not yet been influenced by the observations of this instant and the analysis of the experiments `AssimPred` and `AssimAmpl` assimilate the predicted SST of the analysis time. This predicted SST is also independent of the observed SST at the same time.

The model without assimilation predicts the SST with an average RMS error of 1.111 °C. The improvement on the forecast of the experiment `AssimObs` shows the beneficial impact of the previous assimilation cycles. The standard deviation of the error of this experiment is in average 0.08 °C lower that the error of the free model run.

The forecast and the analysis of `AssimPred` are also better in average than the free model. The RMS of the forecast is reduced by 0.07 °C and the RMS of the analysis even by 0.4 °C.

The best results are obtained of course when assimilating the observations. But these are not available in a real operational forecast. The error reduction of the forecast in `AssimPred` is however comparable to the improvement of the forecast of experiment `AssimObs`. The RMS error of the analysis in `AssimObs` is the lowest since the observations have been assimilated. The improvement of the analysis in this experiment can be achieved in real operational forecast since experiment `AssimPred` does not use the observed SST. The difference between the forecast of `AssimObs` and the analysis of `AssimPred` highlights therefore the benefit of the statistical predictors to constrain ocean models.

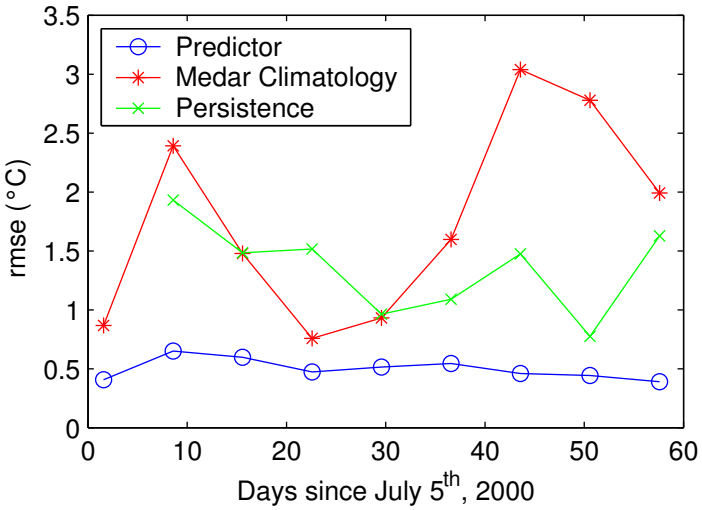

Figure 7.2: The RMS error based on the observed SST of the SOFT system, persistence and climatology.

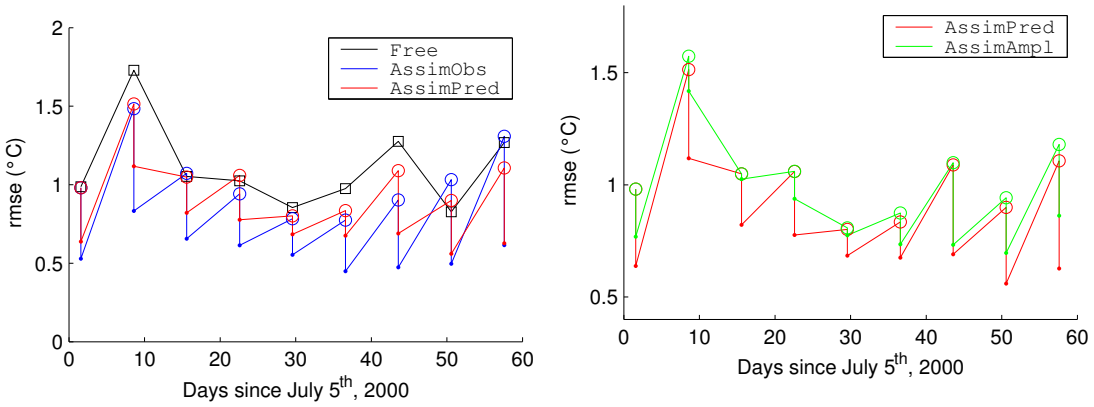

Figure 7.3: RMS error between the observations and the SST of the free model run, model run with assimilation of observed SST and of predicted SST. The dots correspond to the RMS error of the analysis and the circles to the forecast.

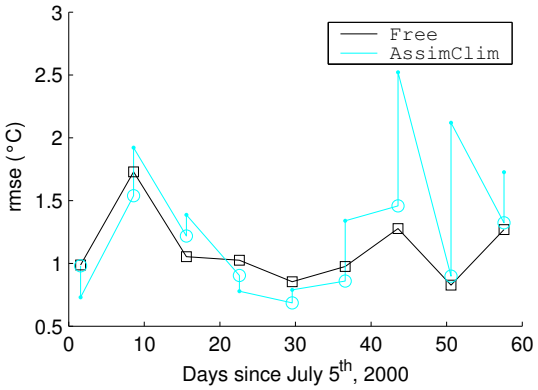

Figure 7.4: RMS error between the observations and the SST of the free model run and the model run with assimilation of the MEDAR SST. The dots correspond to the RMS error of the analysis and the circles of the forecast.

The `AssimAmpl` experiment produced similar results than the `AssimPred` experiment. However, the improvement of the `AssimAmpl` experiment is not so important than the assimilation of the reconstructed SST. This difference suggests that the errors of the rejected EOF amplitudes are not negligible. In particular, the variance of the model SST within this space is too high and the concentration of the SST variance in the space formed by the dominant EOFs improves the result. The assimilation of EOF amplitudes is also technically more difficult since it implies non-local observations.

Large differences between the climatologic SST and the observed SST exist. Thus, the failure of the experiment `AssimClim` is not a surprise. Most of the time, the model forecast is closer to the observations than the climatology. The assimilation of the climatologic SST takes therefore the model away from the observations. This is in particular the case for the strong warming event at the end of the assimilation period. This warming event is absent in the climatology but it is present in the atmospheric heat forcings. Therefore the free model is able to predict this event and the assimilation of the climatologic SST degrade substantially the model result during this period.

7.7 Discussion

The SOFT predictor relies on a given number of past observations, the embedding dimension. For SST prediction in the Ligurian Sea, Álvarez *et al.* (2004) used an embedding dimension of 8. After assimilating 8 weekly SST images of the Ligurian Sea should not the model forecast be better than the forecast of the statistical predictor? If the hypotheses of the Kalman filter are verified, this should be the case for a reasonable good ocean model, since the Kalman filter should provide the best estimate of the ocean state taken into account the dynamics and all previous observations. But in ocean modelling these hypotheses are not exactly satisfied. For instance, the dynamics are nonlinear, model and observations are sometimes biased, the model error is correlated in time and

the error covariances specified in the assimilation scheme are only approximations.

In particular, the reduced rank approximation of the error covariance matrix can only take into account model error within a certain error subspace. Any error outside of this space is ignored. A part of the signal in the observations is therefore systematically rejected.

The manifold of the model trajectory is generally different from the manifold of the true dynamics. This can be highlighted by comparing the model climatology and the climatology based on observations. Often the model error covariances are not well specified and the model error bias and its time correlation are neglected. This leads to the situation where the model falls back on its manifold after assimilation (Judd and Smith, 2001). In this context, it makes sense to use different models with different dynamics biases and drifts. If the models are independent, one can suppose that by merging the different model results, the problems associated to the biases and drifts are reduced.

The SOFT system forecasts the SST using an algebraic function of the m past SST scenes of the Ligurian Sea. The operations (the four elementary operations $+, -, *, /$ are implemented) of this function and their order are determined by genetic algorithm. This approach is very different from ocean modelling based on differential equation such as the primitive equations. Therefore, they can be considered as independent. The models are independent with respect to their dynamics. However, they are influenced by the same observations in this case. After taking into account m SST images of the Ligurian Sea, the forecast of the primitive equation model with SST assimilation and the forecast of the statistical predictor are affected by different model errors which can be reduced by merging them via data assimilation.

Of course, this procedure is not limited to two models. The Kalman filter analysis can be extended to any number of independent models. An option that has not be considered here is to use the analysis as initial state of the statistical predictor. In this case, the model coupling would be bi-directional. But it is not clear how the statistical model behaves when it is fed with the SST of the analysis. The statistical model was "trained" with observed SST with different scales and noise characteristics than the SST of the analysis.

7.8 Conclusions

The main conclusion of this work is that the forecasts of a statistical predictor like the SOFT system developed by Álvarez *et al.* (2004) can be used to improve the forecast skill of models based on primitive equations. The SST obtained by the model with assimilation of the predicted SST is closer to the observations than the SST forecast by the model without assimilation. The accuracy of the SST forecast was improved by 0.3 °C when the forecast of the run `AssimObs` is compared to the analysis of the

experiment `AssimPred`. This corresponds to a 30% reduction of the average RMS error.

The SST is a key parameter for the heat exchange between ocean and atmosphere. In our experiments, the assimilation of predicted SST appears to correct also the errors and biases in the atmospheric heat fluxes. The assimilation of the results obtained from statistical predictors might have an application for operational forecast. Currently, the operational forecasts beyond the present time instant are unconstrained model runs. But the systematic errors due to biases can be detected and therefore corrected by comparing the forecasts with other model. By assimilating the forecasts of specialised statistical predictors, a model can still be constrained even in forecast-mode, and this reduces the error growth due to model's imperfections.

Another way to constrain a forecast model run is to assimilate climatologic SST. It is a common practice in ocean modelling to nudge weakly the model towards climatologic SST to reduce the long-term impact of biased heat fluxes. Since the SST is a highly variable parameter, the assimilation of climatologic SST does not improve the capacity of the model to forecast SST. In the present case, the model without assimilation gives a better SST forecast than the interpolated climatologic SST. For short-range forecasts, the SST predicted by the SOFT system is therefore a better choice.

To our knowledge, it is also the first time that a data assimilation method is used to combine the results of two different models. Traditionally, data assimilation merges observations with model dynamics. Here we show that in the frame of data assimilation, it is also possible to combine the results of different models of very different natures, *i.e.* statistical model and a primitive equation model. The improvements on the prediction skill of this hybrid system are encouraging, but further research is necessary to assess and to control the impact of the assimilation on the unobserved model state vector (see also the model validation part of chapter 6).

Chapter 8

Summary and conclusions

The model nesting has been implemented successfully in a realistic case with two successive grid refinements. A coarse resolution model of 1/4° covering the Mediterranean Sea has been implemented. In this model, a Liguro-Provençal model with a grid refined by a factor of 5 in both horizontal dimensions has been nested. The resolution of this model has been again refined by a factor of 3 in the Ligurian Sea. The coupling between the nested models is bidirectional, *i.e.* boundary conditions of the nested models are obtained from the next coarser resolution model and the variables in the overlapping domains are replaced by the average field of the nested child model.

A novel interpolation scheme for boundary conditions has been applied to scalar fields and to the normal velocity is developed. This interpolation scheme conserves the volume transport and produces a boundary condition, which is as "smooth" as possible. The interpolated velocity does not present any discontinuity and minimises the square of the second derivative integrated along the nesting boundary. It was verified that the whole system of nested models conserves the water volume. However, temperature and salinity are imposed to the fine grid model as Dirichlet boundary condition. The total heat and salt are thus not conserved by the nesting procedure.

The benefit of the two-way nesting compared to the one-way nesting has been highlighted. The temperature structure of the Northern Current, the main dynamical feature in the Ligurian Sea, has been studied for both nesting strategies. The Northern Current is better represented in the two-way nesting than in the one-way nesting in terms of mean position and variability. It is interesting to note that also in the Gulf of Lions, which is located outside of the 1/60° resolution domain of the Ligurian Sea, the representation of the Northern Current is better in the two-way nesting case than for the one-way nesting approach. The improvement due the high resolution is thus propagated through the feedback between the nested models along the downstream path of the Northern Current.

It is known that an abrupt change in the horizontal resolution of the model can lead to wave reflection since the wave speed in the model depends on the resolution (Miyakoda, 2002). Structures with a wavelength of two times of the horizontal grid spacing are found in the nested model at the boundary with the parent model. These structures

cannot be represented on the coarse resolution model grid and they are thus blocked at the boundary. This observation was also made in other two-way nesting model implementations (Sylvain Cailleau, personal communication). The horizontal diffusion coefficient was raised in the nested model near the boundary to reach progressively the diffusion of the coarse grid model. This layer at the boundary with higher diffusion could efficiently filter this small scale noise.

We have also studied the response of the vertical heat transfer of the 1D vertical GHER model. This model has been forced by passive heat and momentum fluxes without any relaxation towards climatological values. The results have been compared to a simulation of the GOTM model realised under the same conditions. When the solar heat flux is modelled as a surface flux, a by far too strong stratification was observed. This stratification inhibits the turbulent heat transfer to the depth and consequently the thermal stratification is reinforced. The SST relaxation towards the climatology might be the reason why this problem was not identified earlier. The treatment of the insolation as a volume source has significantly improved the temperature structure of the GHER model. This has also revealed the complex connection between turbulence scheme and the modelling of the insolation. The vertical light intensity profiles from Jerlov (1968) and Ivanoff (1977) have been tested. The results are similar and the profile of Jerlov (1968) was adopted.

Since SST is an assimilated variable, we are interested in improving the forecast skill of this variable. Independently of the question how the insolation should be modelled, the other heat fluxes can be treated in two different ways. Their values can be imposed passively and independently of the model SST since the atmospheric models compute also these fluxes based on the SST estimation of the atmospheric model. The problem of this approach is that the ocean model SST and atmospheric model SST might be inconsistent. Alternatively, these heat fluxes can be computed interactively using the model SST according the bulk formula parameterisation. This latter approach has been implemented. The formulation has also had a beneficial impact on the realism of the SST predictions. The diurnal cycle of the forcings, in particular of the insolation, is now resolved.

Several sequential assimilation schemes based on optimal interpolation and on the Kalman filter have been reviewed. Due to the similarity of these approaches, different steps of the methods can be exchanged. A simplified approach based on the SEEK, RRSQRT, EnKF and the ESSE has been derived under the constraint that the method should be feasible for a large ocean model.

This assimilation approach has been implemented in a problem and model independent way. Multigrid models are supported as well as arbitrary linear observation operators (applicable, *e.g.*, to the assimilation of SST, SSH, temperature and salinity profiles) and state vector partitioning for local assimilation. The local assimilation is parallelised with PVM and OpenMP. This package is freely available on the GHER webpage `http://modb.oce.ulg.ac.be` or from the author.

This assimilation strategy has been applied to a system composed of three nested models. Sea surface temperature (SST) and sea surface height (SSH) are assimilated into this system of nested models. The covariance of the observation error cannot be taken into account by the current assimilation scheme. Otherwise, a $m \times m$ matrix has to be inverted which is often ill-conditioned. We have examined the effect of ignoring these error covariances. Several simplified assimilation methods have been proposed. They transform the error covariance and the observations in order that the observations can be treated as uncorrelated ones. The accuracy of the different simplified and suboptimal assimilation schemes has been assessed. The method inflating artificially the error variance has lead to the best results among all schemes using a diagonal error covariance matrix. It was also found that the assimilation of binned observations with a transformed observation error covariance has leaded the results, which are the closest to the optimal analysis. However, this transformed observation error covariance matrix is also non-diagonal but it is smaller and better conditioned than the original observation error covariance matrix.

The state vector for the data assimilation compass all nested models. This novel approach ensures that after the analysis all models will be in a consistent state if the covariance between the model states is correctly specified. Traditionally, data assimilation is done in each model independently. The problem with this approach is that observations near the nesting boundary but located on the coarse model grid correct the state of the coarse model but cannot influence the nested fine grid model. This leads to inconsistent boundary conditions. The unique state vector approach takes into account the correlation between the different models. Since the error space is constructed from an ensemble of model states, the relationship between the models states due to the feedback is taken into account.

In general, suboptimal assimilation schemes are derived in a way that they are as close as possible to the unapproximated assimilation scheme. This implies thus an error norm in the state space in order to quantify this "as close as possible" whereas the optimal assimilation schemes (*e.g.* the original Kalman filter) do not depend on any norm. An error norm has been introduced based on the potential energy and the available potential energy. This error norm combines consistently the two-dimensional elevation field and the three-dimensional temperature and salinity fields and it does not depend on the discretisation of these fields. Furthermore, this norm is based on the physical variables and not on statistical quantities, which may change from one particular experiment to another.

A calibration method for the error variance of observations similar to the cross-validation technique has been proposed. This approach has been used to determine the error variance of the high resolution 1 km SST. The error variances of the 10 km SST and SSH have been calibrated by performing several tests with different observation error covariances and by examining the skill in predicting the SST and SSH with a lead time of one week. With these optimised values, a two month simulation with SST and SSH assimilation was carried out and the results were compared to assimilated and independent obser-

vations. In addition, the realism and the validity of the ocean processes have been verified.

We have shown that the assimilation has a beneficial in the SST and SSH forecast. Most of the time, the SST and SSH forecasts of the model with assimilation are more accurate than the predictions of the free model. The path of the Northern Current is also better represented in the model with assimilation than without. However, the assimilation degrades the vertical structure of the temperature when the results are compared to the CTD profiles of the Sirena cruise. In particular, in the interior of the basin where the thermocline is very shallow (about 10 m), the SST assimilation deepens the mixed layer. Possible remedies of this problem are still under study. For example, the ensemble perturbations of the initial conditions can depend on the climatologic mixed layer depth of this zone.

A novel application of data assimilation is also proposed. It can be used to combine different models in order to give an improved estimate of the ocean state. In the present case, this concept has been applied to the nested hydrodynamics model of the Ligurian Sea and the statistical SOFT predictor (Álvarez *et al.*, 2004). The model coupling has been realised simply by assimilating the prediction of the SOFT system into the forecasts of the hydrodynamic model. This approach is interesting in a forecast mode, when observations are not available for correcting the model biases and drifts. It has been showed that the forecast skill of the model could indeed be improved by assimilating the SST predicted by the SOFT system.

Operational models often use climatological SST to reduce the impact of erroneous heat fluxes. Therefore a comparison with assimilation of climatologic SST and predicted SST was done. The experiments with hybrid assimilation system revealed that this heat flux correction can be done more accurately by using the SST from SOFT system. The SST forecast by the SOFT system is numerically inexpensive and by no means comparable to a hydrodynamic model. Furthermore, the variables that can be predicted by the SOFT system are already routinely assimilated into the operational models. Therefore, the SOFT predictor can be integrated easily in existing operational systems.

Appendix A

Derivation of the Fokker-Plank equation

From equation (4.22) we can derive the Fokker-Plank equation for a zero mean, normal distributed model error in the limit of a time continuous model.

$$p_{\boldsymbol{\eta}_i}(\mathbf{x}) = (2\pi\Delta t)^{-n/2}(\det \mathbf{Q})^{-1/2} \exp\left(-\frac{1}{2\Delta t}\mathbf{x}^T\mathbf{Q}^{-1}\mathbf{x}\right) \tag{A.1}$$

$$M(\mathbf{x}) = \mathbf{x} + \Delta t\, g(\mathbf{x}) + \mathcal{O}(\Delta t^2) \tag{A.2}$$

Δt is the time interval between two successive forecast steps of equation (4.22). We are interested in the limit for $\Delta t \longrightarrow 0$. At the first order of Δt we have the following relations:

$$M^{-1}(\mathbf{y}) = \mathbf{y} - \Delta t\, g(\mathbf{y}) + \mathcal{O}(\Delta t^2) \tag{A.3}$$

$$p_{\mathbf{x}_i}(\mathbf{y} - \Delta t\, g(\mathbf{y})) = p_{\mathbf{x}_i}(\mathbf{y}) + \Delta t \sum_{\alpha=1}^{n} g_\alpha(\mathbf{y}) \frac{\partial p_{\mathbf{x}_i}}{\partial \mathbf{x}_\alpha} + \mathcal{O}(\Delta t^2) \tag{A.4}$$

$$\left|\frac{\partial M_i(\mathbf{x}')}{\partial \mathbf{x}'}\right| = 1 + \Delta t \sum_{\alpha=1}^{n} \frac{\partial g_\alpha}{\partial \mathbf{x}_\alpha} + \mathcal{O}(\Delta t^2) \tag{A.5}$$

From equation (4.22) one obtains:

$$p_{\mathbf{x}_{i+1}}(\mathbf{x}) = -\Delta t \sum_{\alpha=1}^{n} \frac{\partial (g_\alpha p_{\mathbf{x}_i})}{\partial \mathbf{x}_\alpha} + \int_{I\!R^n} p_{\mathbf{x}_i}(\mathbf{y}) p_{\boldsymbol{\eta}_i}(\mathbf{x} - \mathbf{y}) d\mathbf{y} + \mathcal{O}(\Delta t^2) \tag{A.6}$$

At order zero, the last integral is the pdf of $\mathbf{x}_n$ since $p_{\boldsymbol{\eta}_i}$ tends to the Delta Dirac function as $\Delta t \longrightarrow 0$. The first order in Δt is obtained by differentiating the integral with respect to Δt.

$$\frac{\partial}{\partial \Delta t}\int_{I\!R^n} p_{\mathbf{x}_i}(\mathbf{y}) p_{\boldsymbol{\eta}_i}(\mathbf{x}-\mathbf{y})\, d\mathbf{y} = \int_{I\!R^n} p_{\mathbf{x}_i}(\mathbf{y}) \frac{\partial p_{\boldsymbol{\eta}_i}}{\partial \Delta t}(\mathbf{x}-\mathbf{y})\, d\mathbf{y} \quad (A.7)$$

$$= \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \int_{I\!R^n} p_{\mathbf{x}_i}(\mathbf{y}) \frac{\partial^2 p_{\boldsymbol{\eta}_i}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta}(\mathbf{x}-\mathbf{y})\, d\mathbf{y} \quad (A.8)$$

$$= \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \int_{I\!R^n} \frac{\partial^2 p_{\mathbf{x}_n}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta}(\mathbf{y}) p_{\boldsymbol{\eta}_i}(\mathbf{x}-\mathbf{y})\, d\mathbf{y} \quad (A.9)$$

$$\rightarrow \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \frac{\partial^2 p_{\mathbf{x}_n}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta}(\mathbf{x}) \qquad \text{if } \Delta t \rightarrow 0 \quad (A.10)$$

where we have used the particular expression of the model error. In fact, one can show that:

$$\frac{\partial p_{\boldsymbol{\eta}_i}(\mathbf{x})}{\partial \Delta t} = -\frac{n}{2\Delta t}\, p_{\boldsymbol{\eta}_i}(\mathbf{x}) + \frac{\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x}}{2\Delta t^2}\, p_{\boldsymbol{\eta}_i}(\mathbf{x}) \quad (A.11)$$

$$= \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \frac{\partial^2 p_{\boldsymbol{\eta}_i}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta} \quad (A.12)$$

Finally, we get the expression of the Fokker-Plank equation, for $\Delta t \rightarrow 0$.

$$\frac{\partial p_{\mathbf{x}}}{\partial t} + \sum_{\alpha=1}^{n} \frac{\partial (g_\alpha p_{\mathbf{x}})}{\partial \mathbf{x}_\alpha} = \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Q}_{\alpha,\beta} \frac{\partial^2 p_{\boldsymbol{\eta}_i}}{\partial \mathbf{x}_\alpha \partial \mathbf{x}_\beta} \quad (A.13)$$

The time index i has been dropped, since the pdf $p_{\mathbf{x}}$ is a continuous function of time.

Appendix B

Some matrix identities

For all symmetric matrices $\mathbf{A}$ and $\mathbf{C}$ and for all matrices $\mathbf{B}$ of conforming dimension, we have:

$$\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{B}^T \left(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T - \mathbf{A}\right)^{-1} \mathbf{B}\mathbf{C}^{-1} = \left(\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \tag{B.1}$$

This is a special case of the Sherman-Morrison-Woodbury formula. From this last equation, one can show that:

$$\left(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\right)^{-1} \mathbf{B}^T\mathbf{C}^{-1} = \mathbf{A}\mathbf{B}^T \left(\mathbf{C} + \mathbf{B}\mathbf{A}\mathbf{B}^T\right)^{-1} \tag{B.2}$$

Instead of performing the inverse in space of matrix $\mathbf{A}$ the inverse is done in space of matrix $\mathbf{C}$.

Appendix C

Eigenvector decomposition of covariance matrices

Covariance matrices are symmetric and positive definite. If $\mathbf{P}$ is a $n \times n$ covariance matrix, then it can be expressed in terms of a $n \times m$ matrix $\mathbf{S}$ by:

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T \tag{C.1}$$

This decomposition is obtained by construction of an error covariance estimated by an ensemble for example. In this case m is much smaller than n. We show here that it is possible to obtain the eigenvectors and eigenvalues of $\mathbf{P}$ without forming $\mathbf{P}$ explicitly.

The singular value decomposition of the matrix $\mathbf{S}$ is given by:

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{C.2}$$

where $\mathbf{\Sigma}$ is a $r \times r$ matrix containing the non-zero singular values of $\mathbf{S}$. Its size r is thus the rank of the matrix $\mathbf{S}$. The columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal ($\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$). The singular value decomposition of $\mathbf{S}$ is directly related to the eigenvalue decomposition of $\mathbf{P}$ since:

$$\begin{aligned} \mathbf{P} &= \mathbf{S}\mathbf{S}^T & \text{(C.3)} \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T & \text{(C.4)} \\ &= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T & \text{(C.5)} \end{aligned}$$

The matrices $\mathbf{\Sigma}^2$ and $\mathbf{U}$ contain thus the non-zero eigenvalues and the corresponding eigenvectors of $\mathbf{P}$. The matrix $\mathbf{V}$ is composed of the eigenvectors of the $m \times m$ matrix $\mathbf{S}^T\mathbf{S}$ since:

$$\begin{aligned} \mathbf{S}^T\mathbf{S} &= \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T & \text{(C.6)} \\ &= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T & \text{(C.7)} \end{aligned}$$

An efficient way to compute the eigenvectors of the model error covariance $\mathbf{P} = \mathbf{S}\mathbf{S}^T$ is thus to form the matrix $\mathbf{S}^T\mathbf{S}$. An eigenvalue decomposition of this $m \times m$ matrix gives $\mathbf{V}$ and $\mathbf{\Sigma}^2$:

$$\mathbf{S}^T\mathbf{S} \;=\; \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T \tag{C.8}$$

The eigenvectors $\mathbf{U}$ of the covariance $\mathbf{P}$ with non-zero eigenvalues are finally obtained by:

$$\mathbf{U} = \mathbf{S}\mathbf{V}\mathbf{\Sigma}^{-1} \tag{C.9}$$

It is important to note that the eigenvectors of $\mathbf{P}$ represented in terms of its square roots can be computed without forming this matrix explicitly.

Appendix D

Derivation of the Kalman filter analysis from a Bayesian approach

The Kalman filter analysis can be derived from the Bayes formula assuming Gaussian error statistics for the *a priori* model state $\mathbf{x}^f$ and for the observations $\mathbf{y}^o$. Gaussian pdf can be entirely described by their mean and their covariance. The pdf of the *a priori* model state $\mathbf{x}^f$ and of the observations $\mathbf{y}^o$ given the model is in the state $\mathbf{x}$ can therefore be written as:

$$p_{\mathbf{x}}(\mathbf{x}) = C_1 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^f)^T \mathbf{P}^{f\,-1}(\mathbf{x}-\mathbf{x}^f)\right) \quad \text{(D.1)}$$

$$p_{\mathbf{y}^o}(\mathbf{y}^o|\mathbf{x}) = C_2 \exp\left(-\frac{1}{2}(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right) \quad \text{(D.2)}$$

where C_1 and C_2 are normalisation constant such that the integrated probability is one. From the Bayes rule, the *a posteriori* pdf is then:

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) = C_3 \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^f)^T \mathbf{P}^{f\,-1}(\mathbf{x}-\mathbf{x}^f) - \frac{1}{2}(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right) \quad \text{(D.3)}$$

From this expression it is not obvious what the mean and the covariance of this pdf is. But since it is a quadratic expression in $\mathbf{x}$, we already now that the *a posteriori* pdf is Gaussian. Therefore we seek its mean $\mathbf{x}^a$ and covariance $\mathbf{P}^a$, such that:

$$(\mathbf{x}-\mathbf{x}^f)^T \mathbf{P}^{f\,-1}(\mathbf{x}-\mathbf{x}^f) + (\mathbf{y}^o-\mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x}) = (\mathbf{x}-\mathbf{x}^a)^T \mathbf{P}^{a\,-1}(\mathbf{x}-\mathbf{x}^a) + C_4 \quad \text{(D.4)}$$

The additional constant C_4 has no importance here, since it can be absorbed into the normalisation constant of the pdf. By requiring that the 1st and the 2nd derivative of the rhs and lhs of the last equation are equal, one obtains x^a and P^a:

$$\mathbf{x}^a = \mathbf{P}^a\left(\mathbf{P}^{f\,-1}\,\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\,\mathbf{y}^o\right) \quad \text{(D.5)}$$

$$\mathbf{P}^{a\,-1} = \mathbf{P}^{f\,-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \quad \text{(D.6)}$$

Equations (D.5) and (D.6) describe the Kalman filter update. It must be noted that alternative formulation of these equation exists. They are derived in section 4.3.3.

Appendix E

EOF as an optimal function basis for representing a series of fields

For data analysis purposes geophysical fields are often projected onto a basis of functions such as Fourier modes or wavelets. A common technique to separate the noise from the deterministic signal or to limit the number of variables to interpret is to truncate the expansion to the most relevant basis functions. EOFs are widespread in ocean and atmospheric data analysis since they constitute the optimal function basis for approximating a series of fields by a truncated linear combination of basis functions.
The $m \times n$ matrix $\mathbf{D}$ contains a time series of spatial fields. The first index i is the space index and the second index j is the time index. We want to describe the variability of this field in a r dimensional subspace $\mathcal{S}_1$. The columns of the $m \times r$ matrix $\mathbf{U}_1$ are a orthonormal basis of this subspace. In matrix notation, the truncated expansion can be written as:

$$\mathbf{D} = \mathbf{U}_1\mathbf{A}_1 + \varepsilon \tag{E.1}$$

The $r \times n$ matrix $\mathbf{A}_1$ contains the amplitudes of each basis vector and $m \times n$ matrix ε represent the error due to the fact that the dimension of the subspace is less than the dimension of the initial data space. Our goal is to find the vector basis $\mathbf{U}_1$ and the amplitudes $\mathbf{A}_1$ such that the total error $\|\varepsilon\|^2$ is minimum. The total error is defined as:

$$\|\varepsilon\|^2 = \mathrm{tr}(\varepsilon^T\varepsilon) \tag{E.2}$$

By taking into account that the columns of $\mathbf{U}_1$ form a orthogonal basis ($\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}$), this error can be written as:

$$\|\varepsilon\|^2 = \|\mathbf{U}_1^T\mathbf{D} - \mathbf{A}_1\|^2 + \|\mathbf{I} - \mathbf{U}_1\mathbf{U}_1^T\mathbf{D}\|^2 \tag{E.3}$$

For a given basis $\mathbf{U}_1$, the optimal amplitudes are therefore:

$$\mathbf{A}_1 = \mathbf{U}_1^T\mathbf{D} \tag{E.4}$$

Consequently the error ε is orthogonal to the subspace $\mathcal{S}_1$

$$\varepsilon^T\mathbf{U}_1 = 0 \tag{E.5}$$

The error ε lies therefore in the $n-r$ dimensional subspace $\mathcal{S}_0$ perpendicular to the subspace $\mathcal{S}_1$. If the columns of the $n \times (n-r)$ matrix $\mathbf{U}_0$ is an orthonormal basis of this subspace $\mathcal{S}_1$, then:

$$\varepsilon = \mathbf{U}_0 \mathbf{A}_0 \tag{E.6}$$

The entire signal rejected by the truncated series of basis vectors lies in the subspace $\mathcal{S}_0$. Both spaces are complementary and if the space $\mathcal{S}_0$ known then also the space $\mathcal{S}_1$ is known and vice versa. It appears to be simpler to derive first the basis function of the space $\mathcal{S}_0$. The total error can be expressed as:

$$\|\varepsilon\|^2 = \|\mathbf{U}_0^T \mathbf{D}\|^2 \tag{E.7}$$

The optimal basis can be found by minimising the following cost function $J(\mathbf{U}_0)$

$$J(\mathbf{U}_0) = \|\varepsilon\|^2 = \mathrm{tr}(\mathbf{U}_0^T \mathbf{D}\mathbf{D}^T \mathbf{U}_0) \tag{E.8}$$

under the constrain that the columns of $\mathbf{U}$ are orthonormal:

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \tag{E.9}$$

There are r^2 scalar constrains on the matrix $\mathbf{U}$. We introduce this constrain by the Langrangian multiplier technique. The new cost function is:

$$J(\mathbf{U}_0) = \mathrm{tr}(\mathbf{U}_0^T \mathbf{D}\mathbf{D}^T \mathbf{U}_0) + \mathrm{tr}(\left[\mathbf{I} - \mathbf{U}_0^T \mathbf{U}_0\right] \mathbf{\Lambda}) \tag{E.10}$$

where the $r \times r$ matrix $\mathbf{\Lambda}$ holds the Langrangian multipliers. The cost function reaches its minimum (or maximum) if a small perturbation of $\mathbf{U}_0$ does not modify the value of the cost function in the first order of the perturbation.

$$\delta J = J(\mathbf{U}_0 + \delta\mathbf{U}_0) - J(\mathbf{U}_0) = 2\mathrm{tr}(\delta\mathbf{U}_0^T \left[\mathbf{D}\mathbf{D}^T \mathbf{U}_0 - \mathbf{U}_0 \mathbf{\Lambda})\right]) = 0 \tag{E.11}$$

This must be true of any perturbation, thus

$$\mathbf{D}\mathbf{D}^T \mathbf{U}_0 = \mathbf{U}_0 \mathbf{\Lambda} \tag{E.12}$$

This expression shows that the basis functions must be a linear combination of $n-r$ eigenvectors of the matrix $\mathbf{D}\mathbf{D}^T$. From equation (E.8) one can show that the total truncation error $\|\varepsilon\|^2$ is the sum of the corresponding eigenvalues. This error is minimum if the columns of $\mathbf{U}_0$ are the eigenvectors with the smallest eigenvalue. Since the spaces $\mathcal{S}_0$ and $\mathcal{S}_1$ are perpendicular to each other, from a basis of one space we can determine a basis of the perpendicular space. The optimal basis $\mathbf{U}_1$ is therefore formed by all other eigenvectors of $\mathbf{D}\mathbf{D}^T$, *i.e.* the eigenvectors corresponding to the largest eigenvalues.

Appendix F

Notations

Physical variables

Symbol	Description	Ensemble
z	depth, zero at the reference surface and negative in water	$\mathbb{R}$
ζ	surface elevation computed from the reference surface $z = 0$	$\mathbb{R}$
h	ocean floor depth computed from the reference surface $z = 0$	$\mathbb{R}$
$\mathbf{v}$	velocity	$\mathbb{R}^3$
$\mathbf{u}$	horizontal velocity	$\mathbb{R}^3$
w	vertical velocity	$\mathbb{R}$
f	Coriolis frequency	$\mathbb{R}$
T	temperature	$\mathbb{R}$
S	salinity	$\mathbb{R}$
k	turbulent kinetic energy	$\mathbb{R}$
q	generalised pressure	$\mathbb{R}$
p	pressure	$\mathbb{R}$
p_b	bottom pressure	$\mathbb{R}$
ρ_0	reference density	$\mathbb{R}$
g	gravity constant	$\mathbb{R}$
b	buoyancy	$\mathbb{R}$
N	Brunt-Väisälä frequency or stability frequency	$\mathbb{R}$
M	Prandtl frequency	$\mathbb{R}$

Basic concepts of data assimilation

Symbol	Description	Ensemble
n	dimension of the model state space	$\mathbb{N}$
m	number of observation at a given time	$\mathbb{N}$
i	time index	$\mathbb{N}$
t_i	time of the ith instant	$\mathbb{R}$
$\mathbf{x}_i$	model state vector	$\mathbb{R}^n$
$\mathbf{x}_i^t$	true model state vector	$\mathbb{R}^n$
$\mathbf{x}_i^f$	*a priori* model state vector (f: forecast)	$\mathbb{R}^n$
$\mathbf{x}_i^a$	*a posteriori* model state vector (a: analysis)	$\mathbb{R}^n$
$\boldsymbol{\zeta}_i^f$	error of the *a priori* model state vector	$\mathbb{R}^n$
$\boldsymbol{\zeta}_i^a$	error of the *a posteriori* model state vector	$\mathbb{R}^n$
$\mathbf{P}_i^f$	error covariance of the *a priori* model state vector	$\mathbb{R}^{n\times n}$
$\mathbf{P}_i^a$	error covariance of the *a posteriori* model state vector	$\mathbb{R}^{n\times n}$
$M_i(\cdot)$	dynamical model	$\mathbb{R}^n \rightarrow \mathbb{R}^n$
$\mathbf{M}_i$	linear dynamical model or tangent linear model of M_i	$\mathbb{R}^{n\times n}$
$\mathbf{f}_i$	External forecings of the linear model	$\mathbb{R}^n$
$\boldsymbol{\eta}_i$	error of the model dynamics	$\mathbb{R}^n$
$\mathbf{Q}_i$	error covariance of the model dynamics	$\mathbb{R}^{n\times n}$
$\mathbf{y}_i^o$	observations	$\mathbb{R}^m$
$H_i(\cdot)$	observation operator	$\mathbb{R}^n \rightarrow \mathbb{R}^m$
$\mathbf{H}_i$	linear observation operator or tangent linear operator of H_i	$\mathbb{R}^{m\times n}$
$\boldsymbol{\varepsilon}_i$	error of the observations	$\mathbb{R}^m$
$\mathbf{R}_i$	error covariance of the observations	$\mathbb{R}^{m\times m}$
$\mathbf{x}^i$	initial condition	$\mathbb{R}^n$
$\boldsymbol{\eta}^i$	error of the initial condition	$\mathbb{R}^n$
$\mathbf{P}^i$	error covariance of the initial condition	$\mathbb{R}^{n\times n}$

Symbol	Description	Ensemble
$p_{\mathbf{x}_i}(\mathbf{x}), p_{\mathbf{x}_i^f}(\mathbf{x})$	*a priori* probability density function of $\mathbf{x}_i$	
$p_{\boldsymbol{\eta}_i}(\mathbf{x})$	probability density function of the error of the model dynamics	
$p_{\mathbf{y}_i^o}(\mathbf{y})$	probability density function of the observations	
$p_{\mathbf{y}_i^o}(\mathbf{y}\|\mathbf{x})$	probability density function of the observations provided the model is in state $\mathbf{x}$	
$p_{\mathbf{x}_i}(\mathbf{x}\|\mathbf{y}_i^o), p_{\mathbf{x}_i^a}(\mathbf{x})$	*a posteriori* probability density function provided the observation $\mathbf{y}_i^o$	
$\mathbf{K}_i$	Kalman gain	$\mathbb{R}^{n\times m}$
$J(\mathbf{x}), J(\mathbf{K})$	cost function	
$\mathbf{W}$	diagonal matrix defining the scalar product of two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ by $\mathbf{x}_1^T\mathbf{W}\mathbf{x}_2$	$\mathbb{R}^{n\times n}$

Stochastic Kalman Filter

Symbol	Description	Ensemble
N	ensemble size	$\mathbb{N}$
k	ensemble index, $k = 1, \ldots, N$	$\mathbb{N}$
$\mathbf{x}_i^{f(k)}, \mathbf{x}_i^{a(k)}$	ensemble of state vectors representing the *a priori* and *a posteriori* states	$\mathbb{R}^n$
$\mathbf{P}_{e_i}^f, \mathbf{P}_{e_i}^a$	ensemble covariance of the *a priori* and *a posteriori* state vectors	$\mathbb{R}^{n\times n}$
$\boldsymbol{\eta}_i^{(k)}$	ensemble of error of the model dynamics	$\mathbb{R}^n$
$\boldsymbol{\zeta}^{i(k)}$	ensemble of error of the initial condition	$\mathbb{R}^n$
$\boldsymbol{\varepsilon}_i^{(k)}$	ensemble of measurment errors	$\mathbb{R}^m$
$\mathbf{E}$	matrix whose columns are the measurment errors multiplied by $\frac{1}{\sqrt{N-1}}$	$\mathbb{R}^{m\times N}$
$\mathbf{y}_i^{o(k)}$	ensemble of observations	$\mathbb{R}^m$
$\mathbf{R}_{e_i}$	ensemble covariance of the measurment errors	$\mathbb{R}^{m\times m}$
$N(\mathbf{x}, \mathbf{P})$	Gaussian pdf with mean $\mathbf{x}$ and covariance $\mathbf{P}$	

Deterministic Kalman filters

Symbol	Description	Ensemble
$\mathbf{S}_i^f$	square root matrices of *a priori* error covariance matrix , $\mathbf{P}_i^f = \mathbf{S}_i^f \mathbf{S}_i^{f^T}$.	$I\!R^{n\times r}$
$\mathbf{U}_i^f$	right singular vectors of $\mathbf{S}_i^f$, $\mathbf{S}_i^f = \mathbf{U}_i^f \mathbf{\Sigma}_i^f \mathbf{V}_i^{f^T}$ (or eigenvector of the *a priori* error covariance matrix, $\mathbf{P}_i^f = \mathbf{U}_i^f \mathbf{\Sigma}_i^{f^2} \mathbf{U}_i^{f^T}$)	$I\!R^{n\times r}$
$\mathbf{\Sigma}_i^f$	diagonal matrix containing the singular values of $\mathbf{S}_i^f$ (or the square root of the eigenvalues of $\mathbf{P}_i^f$)	$I\!R^{r\times r}$
$\mathbf{V}_i^f$	right singular vectors of $\mathbf{S}_i^f$	$I\!R^{r\times r}$
$\tilde{\mathbf{P}}_i^f$	error covariance of the *a priori* state vector expressed in the error subspace, $\mathbf{P}_i^f = \mathbf{L}_i \tilde{\mathbf{P}}_i^f \mathbf{L}_i^T$	$I\!R^{r\times r}$
$\mathbf{S}_i^a, \mathbf{U}_i^a, \mathbf{\Sigma}_i^a, \mathbf{V}_i^a, \tilde{\mathbf{P}}_i^a$	idem but for the *a posteriori* state	
$\mathbf{L}_i$	matrix whose columns define the error subspace. The columns are not necessary orthogonal.	$I\!R^{n\times r}$
$\tilde{\mathbf{Q}}_i$	error covariance of the model dynamics error expressed in the error subspace, $\mathbf{Q}_i = \mathbf{L}_i \tilde{\mathbf{Q}}_i \mathbf{L}_i^T$	$I\!R^{r\times r}$
$\tilde{\mathbf{K}}_i$	Kalman gain expressed in the error subspace, $\mathbf{K}_i = \mathbf{L}_i \tilde{\mathbf{K}}_i$	$I\!R^{r\times m}$
$\tilde{\mathbf{R}}_i$	error covariance of the observations including the rank reduction truncation error of $\mathbf{P}^f$	$I\!R^{m\times m}$
ρ	forgetting factor	$[0, 1]$

List of Figures

List of Tables

Bibliography

Albérola, C., C. Millot, and J. Font, 1995: On the seasonal and mesoscale variabilities of the Northern Current during the PRIMO-0 experiment in the western Mediterranean Sea. *Oceanologica Acta*, **18**, 163–192.

Álvarez, A., C. Lopez, M. Riera, E. Hernandez-Garcia, and J. Tintoré, 2000: Forecasting the SST space-time variability of the Alboran Sea with genetic algorithms. *Geophysical Research Letters*, **27**, 2709–2712.

Álvarez, A., A. Orfila, and J. Tintoré, 2001: DARWIN: An evolutionary program for nonlinear modelling of chaotic time series. *Computer Physics Communications*, **136**, 334–349.

Álvarez, A., A. Orfila, and J. Tintoré, 2004: Real-time forecasting at weekly timescales of the SST and SLA of the Ligurian Sea with a satellite-based ocean forecasting (SOFT) system. *Journal of Geophysical Research*, **109**, C03023.

Alvera-Azcárate, A., A. Barth, Z. B. Bouallègue, L. Vandenbulcke, M. Rixen, and J.-M. Beckers, 2004: Forecast Verification of a 3D model of the Ligurian Sea. The use of Discrete Wavelet Transforms in the skill assessment of spatial forecasts. *Journal of Marine Systems*, in preparation.

Anderson, J. L., 2001: An Ensemble Adjustment Filter for Data Assimilation. *Monthly Weather Review*, **129**, 2884–2903.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127**, 2741–2758.

Astraldi, M., S. Balopoulos, J. Candela, J. Font, M. Gacic, G. P. Gasparini, B. Manca, A. Theocharis, and J. Tintoré, 1999: The role of straits and channels in understanding the characteristics of Mediterranean circulation. *Progress in Oceanography*, **44**, 65–108.

Astraldi, M. and G. P. Gasparini, 1992: The seasonal characteristics of the circulation in the north Mediterranean basin and their relationship with the atmospheric-climatic conditions. *Journal of Geophysical Research*, **97**, 9531–9540.

Bahurel, P., E. Dombrowsky, J.-M. Lellouche, and the Mercator Project Team, 2004: Mercator Ocean Monitoring and Forecasting System, near-real-time assimilation of

satellite and in situ data in different operational ocean models. *Marine Environmental Monitoring and Predictions*, 37th Liège Colloquium, oral presentation.

Bahurel, P., P. D. Mey, T. D. Prada, E. Dombrowsky, P. Josse, C. L. Provost, P.-Y. L. Traon, A. Piacentini, and L. Siefridt, 2001: *AVISO Altimetry Newsletter*, CNES, Toulouse, France, chapter MERCATOR, forecasting global ocean. 8, 14–16.

Barnier, B., L. Siefridt, and P. Marchesiello, 1995: Thermal forcing for global ocean circulation model using a 3-year climatology of ECMWF analyses. *Journal of Marine Systems*, **6**, 363–380.

Barth, A., A. Alvera-Azcarate, M. Rixen, and J.-M. Beckers, 2003: Two-way nested model of mesoscale circulation features in the Ligurian Sea. *Progress In Oceanography*, accepted.

Beckers, J.-M., 1991: Application of a 3D model to the Western Mediterranean. *Journal of Marine Systems*, **1**, 315–332.

Beckers, J.-M., M. Rixen, P. Brasseur, J.-M. Brankart, A. Elmoussaoui, M. Crépon, C. Herbaut, F. Martel, F. Van den Berghe, L. Mortier, A. Lascaratos, P. Drakopoulos, P. Korres, N. Pinardi, E. Masetti, S. Castellari, P. Carini, J. Tintore, A. Alvarez, S. Monserrat, D. Parrilla, R. Vautard, and S. Speich, 2002: Model intercomparison in the Mediterranean. The MedMEx simulations of the seasonal cycle. *Journal of Marine Systems*, **33–34**, 215–251.

Bell, M., R. M. Forbes, and A. Hines, 2000: Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *Journal of Marine Systems*, **25**, 1–22.

Bell, M. J., 2003: Conservation of Potential Vorticity on Lorentz Grids. *Monthly Weather Review*, **131**, 1498–1501.

Bennett, A. F., 1992: *Inverse methods in physical oceanography*. Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press.

Bertino, L., G. Evensen, and H. Wackernagel, 2002: Combining geostatistics and Kalman Filtering for data assimilation in an estuarine system. *Inverse Methods*, **18**, 1–23.

Bertino, L., A. Hollard, G. Evensen, and H. Wackernagel, 2003: An ensemble Kalman filter for non-gaussian variables. *Geophysical Research Abstracts*, EGS-AGU-EUG Joint Assembly, Nice, France, volume 5, oral presentation.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129**, 420–436.

Blayo, E. and L. Debreu, 1999: Adaptive mesh refinement for finite difference ocean model: some first experiments. *Journal of Physical Oceanography*, **29**, 1239–1250.

Blayo, E., J. Verron, J. M. Molines, and L. Testard, 1996: Monitoring of the Gulf Stream path using Geosat and Topex/Poseidon altimetric data assimilated into a model of ocean circulation. *Journal of Marine Systems*, **8**, 73–89.

Bleck, R., 2002: An Oceanic Circulation Model Framed in Hybrid Isopycnic-Cartesian Coordinates. *Ocean Modelling*, **1**, 55–88.

Brankart, J. M., C. E. Testut, P. Brasseur, and J. Verron, 2003: Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean modes: application to a 1993-96 hindcast of the North Atlantic Ocean circulation. *Journal of Geophysical Research*, **108 (C3)**.

Brasseur, P., J. Ballabrera, and J. Verron, 1999: Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. *Journal of Marine Systems*, **22**, 269–294.

Brasseur, P., J.-M. Beckers, T. J. of Meteorology (AMS), and R. Schoenauen, 1996: Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of an historical data set. *Deep–Sea Research*, **43**, 159–192.

Browning, G. L., W. R. Holland, H.-O. Kreiss, and S. J. Worley, 1990: An accurate hyperbolic system for approximately hydrostatic and incompressible oceanographic flows. *Dynamics of Atmosphere and Oceans*, **14**, 303–332.

Brusdal, K., J. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. J. van Leeuwen, E. Dombrowsky, and J. Verron, 2003: An evaluation of ensemble based assimilation methods with a layered OGCM. *Journal of Marine Systems*, **40–41**, 253–289.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF EPS. *Quarterly Journal of the Royal Meteorological Society*, **125**, 2887–2908.

Burchard, H., 2002: *Applied turbulence modelling in marine waters*, volume 100 of *Lecture Notes in Earth Sciences*. Springer, Berlin, Heidelberg, New York, 215 pp.

Burchard, H., K. Bolding, T. P. Rippeth, A. Stips, J. H. Simpson, and J. Sündermann, 2002: Microstructure of turbulence in the Northern North Sea: A comparative study of observations and model simulations. *Journal of Sea Research*, **47**, 223–238.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, **126**, 1719–1724.

Caires, S. and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research*, **108**, 3098.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, 1996: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *Journal of Geophysical Research*, **101**, 22599–22617.

Cartwright, D. E. and R. D. Ray, 1990: Oceanic tides from Geosat altimetry. *Journal of Geophysical Research*, **95**, 3069–3090.

Castellari, S., N. Pinardi, and K. Leaman, 1998: A model study of air-sea interactions in the Mediterranean Sea. *Journal of Marine Systems*, **18**, 89–114.

Chen, C., 1991: A Nested Grid, Nonhydrostatic, Elatic Model Using a terrain following coordinate transormation: The radiative-nesting boundary conditions. *Monthly Weather Review*, **119**, 2851–2869.

Chua, B. S. and A. F. Bennett, 2001: An inverse ocean modeling system. *Ocean Modelling*, **3**, 137–165.

Clark, N. E., R. M. Eber, J. A. Renner, and J. F. T. Saur, 1974: Heat exchange between ocean and atmosphere in the eastern north pacific for 1961-71. Technical report, NOAA, U. S. Dept. Commerce, Wash., D. C.

Cohn, S. E. and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Monthly Weather Review*, **119**, 1757–1785.

Cohn, S. E. and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *Journal of the Meteorological Society of Japan*, **74**, 63–75.

Cooper, M. and K. Haines, 1996: Data assimilation with water property conservation. *Journal of Geophysical Research*, **10**, 1059–1077.

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, **124**, 1783–1808.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, New York, 457pp.

— 1992: Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Monthly Weather Review*, **120**, 627–643.

De Mey, P., F. Auclair, G. Jordà, J. Lamouroux, F. Lyard, S. Mangiarotti, P. Marsaleix, B. Mourre, I. Pairaud, C. Estournel, and C. Le Provost, 2004: Data assimilation in regional and shelf seas. *Marine Environmental Monitoring and Predictions*, Liège Colloquium, oral presentation.

De Mey, P. and M. Benkiran, 2002: *Ocean Forecasting, Conceptual basis and Applications*, Springer-Verlag, Berlin Heidelberg New York, chapter A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. 281–306.

Dee, D. P., 1990: Simplified adaptive Kalman filtering for large-scale geophysical models. *Realization and Modelling in System Theory*, M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, eds., Birkhäuser, volume 1 of *Proceedings of the International Symposium MTNS-89*, 567–574.

Dee, D. P., S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Transactions on Automatic Control*, **AC-30**, 1057–1065.

Dee, D. P. and A. Silva, 1998: Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, **124**, 269–295.

Demirov, E., N. Pinardi, C. Fratianni, M. Tonani, L. Giacomelli, and P. Mey, 2003: Assimilation scheme of the Mediterranean Forecasting System: operational implementation. *Annales Geophysicae*, 189–204.

Dimet, F.-X. L. and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.

Donlon, C. J., 2002: GHRSST-PP Data Poduct Specifivations v2.0. Technical report, GODEA Hight Resolution Sea Surface Temperature Pilot Project, GHRSST-PP Reference Document GHRSST/10, http://ghrsst-pp.org.

Doucet, A., N. de Freitas, and N. J. Gordon, eds., 2001: *Sequential Monte Carlo Methods in Practice*. Springer, 581 pp.

Ehrendorfer, M. and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *Journal of the Atmospheric Sciences*, **52**, 3475–3500.

Etienne, H. and E. Dombrowsky, 2003: Estimation of the optimal interpolation parameters in a quasi-geostrophic model of the Northeast Atlantic using ensemble methods. *Journal of Marine Systems*, **40–41**, 317–339.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10143–10162.

— 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

— 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, submitted.

Evensen, G. and P. J. van Leeuwen, 2000: An Ensemble Kalman Smoother for Nonlinear Dynamics. *Monthly Weather Review*, **128**, 1852–1867.

Faucher, P., M. Gavart, and P. D. Mey, 2002: Isopycnal empirical orthogonal functions (EOFs) in the North and tropical Atlantic and their use in estimation problems. *Journal of Geophysical Research*, **107**, 21–1.

Fennel, K., M. Losch, J. Schroter, and M. Wenzel, 2001: Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *Journal of Marine Systems*, **28**, 45–63.

Fernández, V., D. E. Dietrich, R. L. Haney, and J. Tintoré, 2003: A Numerical Study of the Mediterranean Sea Internal Variability: Sensitivity to the Model Surface Boundary Conditions. *EOS Transactions AGU*, Ocean Sciences Meeting, Suppl., volume 84(52).

Fleming, R. J., 1971a: On stochastic dynamic prediction. Part I: The energetics of unvertainty and the question of closure. *Monthly Weather Review*, **99**, 851–872.

— 1971b: On stochastic dynamic prediction. Part II: Predictability and utility. *Monthly Weather Review*, **99**, 927–938.

Fox, A. D., K. Haines, B. A. de Cuevas, and D. J. Webb, 2000a: Altimeter assimilation in the OCCAM global model: Part I. a twin experiment. *Journal of Marine Systems*, **26**, 303–322.

— 2000b: Altimeter assimilation in the OCCAM global model: Part II. TOPEX/POSEIDON and ERS-1 assimilation. *Journal of Marine Systems*, **26**, 323–347.

Fox, A. D. and S. J. Maskell, 1995: Two-Way Interactive Nesting of Primitive Equation Ocean Models with Topography. *Journal of Physical Oceanography*, **25**, 2977–2996.

— 1996: A nested primitive equation model of the Iceland-Faeroe front. *Journal of Geophysical Research*, **101**, 18259–18278.

Fritsch, J. M., R. A. H. Jr., R. Adler, H. Bluestein, L. Bosart, J. Brown, F. Carr, C. Davis, R. H. Johnson, N. Junker, Y.-H. Kuo, S. Rutledge, J. Smith, Z. Toth, J. W. Wilson, E. Zipser, , and D. Zrnic, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U. S. Weather Research Program. *Bulletin of the American Meteorological Society*, **79**, 285–299.

Fukumori, I., 2002: A partitioned Kalman filter and smoother. *Monthly Weather Review*, **130**, 1370–1383.

Fukumori, I., J. Benveniste, C. Wunsch, and D. B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. *Journal Physical Oceanography*, **23**, 1831–1855.

Fukumori, I. and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with one idealized Gulf Stream model. *Journal of Geophysical Research*, **100**, 6777–6793.

Fukumori, I., R. Raghunath, L.-L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/Poseidon altimeter data into a global ocean circulation model: How good are the results? *Journal of Geophysical Research*, **104(C11)**, 25647–25666.

Gandin, L. S., 1965: *Objective analysis of meteorological fields.*. Israel Program for Scientific Translation, Jerusalem, 242 pp.

Garnier, V., G. Andre, and P. Garreau, 2004: Monitoring of the Coastal Circulation in the Northern Part of the Western Mediterranean Sea using nested models. *Marine Environmental Monitoring and Predictions*, Liège Colloquium.

Gasparini, G. P., G. Zodiatis, M. Astraldi, C. Galli, and S. Sparnocchia, 1999: Winter intermediate water lenses in the Ligurian Sea. *Journal of Marine Systems*, **20**, 319–332.

Gavart, M., P. De Mey, and G. Caniaux, 1999: Assimilation of satellite altimeter data in a primitive-equation model of the Azores-Madeira region. *Dynamics of Atmospheres and Oceans*, **29**, 217–254.

Gelb, A., 1974: *Applied optimal estimation*. MIT Press, Cambridge, MA, 374 pp.

Ginis, I., R. A. Richardson, and L. M. Rothstein, 1998: Design of a Multiply Nested Primitive Equation Ocean Model. *Monthly Weather Review*, **126**, 1054–1079.

Haines, K., 2002: *Ocean Forecasting, Conceptual basis and Applications*, Springer-Verlag, Berlin Heidelberg New York, chapter Assimilation of Satellite Altimetry in Ocean Models. 472.

Hamill, T. M. and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. *Monthly Weather Review*, **128**, 2905–2919.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthy Weather Review*, **129**, 2776–2790.

Haney, R. L., 1991: On the pressure gradient force over steep topography in sigma coordiante ocean models. *Journal of Physical Oceanography*, **21**, 610–619.

Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance Reduced Ensemble Kalman Filtering. *Monthly Weather Review*, **129**, 1718–1728.

Horn, R. A. and C. R. Johnson, 1985: *Matrix Analysis*. Cambridge University Press, 561 pp.

— 1991: *Topics in Matrix Analysis*. Cambridge University Press, 607 pp.

Hoteit, I., D.-T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine Systems*, **36**, 101–127.

Houtekamer, P. L., 1995: The construction of optimal perturbations. *Monthly Weather Review*, **123**, 2888–2898.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Monthly Weather Reviev*, **124**, 1225–1242.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811.

— 1999: Reply. *Monthly Weather Reviev*, **127**, 1378–1379.

Ivanoff, A., 1977: *Modelling and Prediction of the Upper Layers of the Ocean*, Pergamon, School of Marine and Atmospheric Science, University of Miami, Florida, chapter Oceanic absorption of solar energy. 4771.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic, San Diego, California.

Jerlov, N. G., 1968: *Optical oceanography*. Elsevier Publishing Co., New York, 194 pp.

Jiang, S. and M. Ghil, 1993: Dynamical properties of error statistics in a shallow-water model. *Journal of Physical Oceanography*, **23**, 2541–2566.

Judd, K. and L. A. Smith, 2001: Indistinguishable states II: Imperfect model scenarios. *Physica D*, **2-4**, 125–141.

Julier, S. J. and J. K. Uhlmann, 1997: A new extension of the Kalman Filter to nonlinear systems. *Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, Florida.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, **82**, 35–45.

Kepert, J. D., 2004: Comment on "The Ensemble Kalman Filter: Theoretical formulation and practical implementation". *Ocean Dynamics*, submitted.

Kivman, G. A., 2003: Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, **10**, 253–259.

Kondo, J., 1975: Air-sea bulk transfer coefficients in diabatic conditions. *Boundary–Layer Meteoroplogy*, 91–112.

Korres, G., I. Hoteit, and G. Triantafyllou, 2004: Data Assimilation into a Princeton Ocean Model of the Mediterranean Sea using the SEEK Filter. *Marine Environmental Monitoring and Predictions*, 37th Liège Colloquium, oral presentation.

Krishnamurti, T. N., C. M. Kishtawal, T. La Row, D. Bachiochi, Z. Zhang, E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.

Kurihara, Y. and M. A. Bender, 1980: Use of a movable nested-mesh model for tracking a small vortex. *Monthly Weather Review*, **108**, 1792–1809.

Lacroix, G. and M. Grégoire, 2002: Revisited ecosystem model (MODECOGeL) of the Ligurian Sea: seasonal and interannual variability due to atmospheric forcing. *Journal of Marine Systems*, **37**, 229–258.

Larnicol, G., N. Ayoub, and P. Y. Le Traon, 2002: Major changes in Mediterranean Sea level variability from seven years of TOPEX/Poseidon and ERS-1/2 data. *Journal of Marine Systems*, **33–34**, 63–89.

Le Provost, C., F. Lyard, J. M. Molines, M. L. Genco, and F. Rabilloud, 1998: A hydrodynamic ocean tide model improved by assimilating a satellite altimeter-derived data set. *Journal of Geophysical Research*, **103**, 5513–5529.

Le Traon, P.-Y., P. Schaeffer, S. Guinehut, M.-H. Rio, F. Hernandez, G. Larnicol, and J.-M. Lemoine, 2001: Mean Ocean Dynamic Topography from GOCE and Altimetry. *First International GOCE Workshop*, ESTEC, ESA, Noordwijk, The Netherlands, 135–139.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *Journal of the Atmospheric Sciences*, **28**, 145–161.

— 1974: Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, **102**, 409–418.

Leith, C. E. and R. H. Kraichnan, 1972: Predictability of turbulent flows. *Journal of the Atmospheric Sciences*, **29**, 1041–1058.

Lemoine, F. G., S. C. Kenyon, J. K. Factor, R. G. Trimmer, N. K. Pavlis, D. S. Chinn, C. M. Cox, S. M. Klosko, S. B. Luthcke, M. H. Torrence, Y. M. Wang, R. G. Williamson, E. C. Pavlis, R. H. Rapp, and T. R. Olson, 1998: *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96*. NASA Goddard Space Flight Center, Greenbelt, Maryland, 20771 USA, TP-1998-206861.

Lermusiaux, P. F. J., 1997: *Error Subspace Data Assimilation Methods for Ocean Field Estimation: Theory, Validation and Appliciation*. Ph.D. thesis, Harvard University, Cambridge, Mass.

Lermusiaux, P. F. J. and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part 1: Theory and schemes. *Monthly Weather Review*, **127**, 1385–1407.

Levitus, S., 2002: Building global ocean profile-plankton databases for scientific research. *Tracer Methods in Geophysical Fluid Dynamics*, Liège Colloquium.

Levitus, S. and T. P. Boyer, 1994: *World Ocean Atlas 1994 Volume 4: Temperature*. NOAA Atlas NESDIS 4, Washington, D. C., 117 pp.

Levitus, S., R. Burgett, and T. P. Boyer, 1994: *World Ocean Atlas 1994 Volume 3: Salinity*. NOAA Atlas NESDIS 3, Washington, D. C., 99 pp.

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, 2000: The Met Office global 3-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2991–3012.

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167.

— 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20**, 130–141.

Luong, P. V., C. P. Breshears, and L. N. Ly, 2004: Application of Multiblock Grid and Dual-Level Parallelism in Coastal Ocean Circulation Modeling. *Journal of Scientific Computing*, **20**, 257–277.

Martin, M. J., M. J. Bell, and A. Hines, 2002: Estimation of three-dimensional error covariance statistics for an ocean assimilation system. Ocean Applications Technical Note No. 30, Met Office, Bracknell.

Miller, R. N. and M. A. Cane, 1989: A Kalman Filter Analysis of Sea Level Height in the Tropical Pacific. *Journal of Physical Oceanography*, **19**, 773–790.

Miller, R. N., M. Ghil, and F. Gauthier, 1994: Data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, **51**, 1037–1056.

Miyakoda, K., 2002: *Forecasting: Conceptual Basis and Applications*, Springer-Verlag, chapter Strategy for Regional Seasonal Forecasts. 179–197.

Navarra, A., 2002: *Ocean Forecasting, Conceptual basis and applications*, Springer-Verlag, Berlin Heidelberg New York, chapter Ensembles, Forecasts and Predictability. 131–148.

Nerger, L., 2004: *Parallel Filter Algorithms for Data Assimilation in Oceanography*. Ph.D. thesis, Universität Bremen.

Nerger, L., W. Hiller, and J. Schröter, 2004a: A comparison of error subspace Kalman filters, part 1: Filter algorithms. *Monthly Weather Review*, submitted.

— 2004b: A comparison of error subspace Kalman filters, part 2: Evaluation of numerical experiments. *Monthly Weather Review*, submitted.

Nihoul, J. C. J., E. Deleersnijder, and S. Djenidi, 1989: Modelling the general circulation of shelf seas by 3D $k - \epsilon$ models. *Earth-Science Reviews*, **26**, 163–189.

Oey, L.-Y. and P. Chen, 1992: A Nested-Grid Ocean Model: With Application to the Simulation of Meanders and Eddies in the Norwegian Coastal Current. *Journal of Geophysical Research*, **97**, 20063–20086.

Oliger, J. and A. Sundstrom, 1978: Theoretical and practical aspects of some initial and boundary value problems in fluid dynamics. *SIAM Journal of Applied Mathematics*, **35**, 419.

Onken, R., A. R. Robinson, P. F. J. Lermusiaux, P. J. H. Jr., and L. A. Anderson, 2003: Data-driven simulations of synoptic circulation and transports in the Tunisia-Sardinia-Sicily region. *Journal of Geophysical Research*, **108 (C9)**, 8123.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus*, submitted.

Parrish, D. F. and S. E. Cohn, 1985: A Kalman filter for a two-dimensional shallow-water model: Formulation and preliminary experiments, Office Note 304, NOAA/NWS/NMC, 64 pp.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**, 1194–1207.

Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, **16**, 323–340.

Pinardi, N., I. Allen, E. Demirov, P. Mey, G. Korres, A. Lascaratos, P. Y. Traon, C. Maillard, G. Manzella, and C. Tziavos, 2003: The Mediterranean ocean forecasting system: first phase of implementation (1998-2001). *Annales Geophysicae*, **21**, 3–20.

Pinardi, N., A. Grezio, P. Oddo, and the ADRICOSM parters, 2004: Adriatic Sea integrated costal areas and river basin management system pilot project (ADRICOSM). *37th Congress Proceedings*, CIESM, Barcelona, Spain, volume 37.

Reed, R. K., 1977: On estimating insolation over the ocean. *Journal of Physical Oceanography*, **17**, 854–871.

Rio, M.-H. and F. Hernandez, 2004: A Mean Dynamic Topography computed over the world ocean from altimetry, in-situ measurements and a geoid model. *Journal of Geophysical Research*, accepted.

Rixen, M., 1999: *Data analysis and assimilation in the Alboran Sea*. Ph.D. thesis, University of Liège, 198 pp.

Rixen, M., A. Barth, and J.-M. Beckers, 2001a: A new hydrid forecasting system using neural network forecasted satellite imagery. *The use of data assimilation in coupled hydrodynamic, ecological and bio-geo-chemical models of the ocean*, University of Liège, 33rd International Liège Colloquium on Ocean Hydrodynamics, oral presentation.

Rixen, M., J.-M. Beckers, J.-M. Brankart, and P. Brasseur, 2001b: A numerically efficient data analysis method with error map generation. *Ocean Modelling*, **2**, 45–60.

Robinson, A. and J. Sellschopp, 2002: *Forecasting: Conceptual Basis and Applications*, Springer-Verlag, chapter Rapid Assessment of the Coastal Ocean Environment. 199–230.

Ronday, F., 2000: Dynamique des basses couches de l'atmosphére et interactions air-mers. Lecture at the University of Liege.

Rosati, A. and K. Miyakoda, 1988: A general circulation model for upper ocean simulation. *Journal of Physical Oceanography*, **18**, 1601–1626.

Rowley, C. and I. Ginis, 1999: Implementation of a mesh movement scheme in a multiply nested ocean model and its application to air-sea interaction studies. *Monthly Weather Review*, **127**, 1879–1896.

Salat, J. and J. Font, 1987: Water mass structure near and offshore the Catalan coast during the winters of 1982 and 1983. *Annales Geophysicae*, **1B**, 49–54.

Sammari, C., C. Millot, and L. Prieur, 1995: Aspects of the seasonal and mesoscale variabilities of the Northern Current in the Western Mediterranean Sea inferred from the PROLIG-2 and PROS-6 experiments. *Deep-Sea Research I*, **42,6**, 893–917.

Smith, T. M. and R. W. Reynolds, 2003: Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854-1997). *Journal of Climate*, **16**, 1495–1510.

Smith, W. H. F. and D. T. Sandwell, 1997: Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. *Science*, **277**, 1956–1962.

Spall, M. A. and W. R. Holland, 1991: A nested primitive equation model of oceanic application. *Journal of Physical Oceanography*, **21**, 205–220.

Spall, M. A. and A. R. Robinson, 1989: A new open-ocean hybrid coordinate primitive equation model. *Mathematics and Computers in Simulations*, **31**, 241–269.

Spitz, Y. H., J. R. Moisan, and M. R. Abbott, 2001: Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). *Deep Sea Research Part II: Topical Studies in Oceanography*, **48**, 1733–1768.

Stammer, D., C. Wunsch, R. Giering, C. Eckerts, P. Heimbach, J. Marortzke, A. Adcroft, C. N. Hill, and J. Marshall, 2002: The global ocean circulation during 1992-1997, estimated from ocean observations and a general circulation model. *Journal of Geophysical Research*, **107(C9)**, 3118.

Stengel, R. F., 1986: *Optimal Control and Estimation*. Wiley, New York.

Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, **128**, 2077–2107.

Takens, F., 1981: *Lecture Notes Math.*, Springer-Verlag, New York, volume 898, chapter Detecting strange attractors in fluid turbulence, in Dynamical Systems and Turbulence. 366–369.

Talagrand, O., 1997: Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, **75**, 191–209.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equations: Theory. *Quarterly Journal of the Royal Meteorological Society*, **113**, 1311–1328.

Testut, C.-E., J.-M. Brankart, and L. Parent, 2002: An integrated system of sequential assimilation modules. SESAM 3.1 reference manual. Technical report, MEOM-LEGI.

Testut, C.-E., P. Brasseur, J.-M. Brankart, and J. Verron, 2003: Assimilation of sea-surface temperature and altimetric observations during 1992-1993 into an eddy permitting primitive equation model of the North Atlantic Ocean. *Journal of Marine Systems*, **40-41**, 291–316.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square-Root Filters. *Monthly Weather Review*, **131**, 1485–1490.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, **74**, 2317–2330.

Tusseau, M. H. and J. M. Mouchel, 1994: Nitrogen inputs to the Gulf of Lions via the Rhône river. *Water Pollution Research Reports, Proceedings of the EROS 2000 Workshop, Hamburg*, J. M. Martin and H. Barth, eds., commission of the European Communities.

Uchida, H., S. Imawaki, and J.-H. Hu, 1998: Comparison of Kuroshio surface velocities derived from satellite altimetry and drifting buoy data. *Journal of Oceanography*, **54**, 115–122.

van Leeuwen, P. J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique". *Monthly Weather Review*, **127**, 1374–1377.

— 2003: A Variance-Minimizing Filter for Large-Scale Applications. *Monthly Weather Review*, **131**, 2071–2084.

van Leeuwen, P. J. and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, **124**, 2898–2913.

Vandenbulcke, L., A. Barth, A. Alvera-Azcarate, Z. Ben Bouallegue, M. Rixen, and J. Beckers, 2004: A nested-grid model with data assimilation in the Gulf of Lions. *Geophysical Research Abstracts*, EGU General Assembly, Nice.

Vandenbulcke, L., A. Barth, Z. B. Bouallegue, M. Rixen, and J. Beckers, 2003: Demonstration of the two-way nesting in the Gulf of Lions. *Multivariate estimation tools for the basin and regional scales - Workpackage 7*, MFSTEP Workpackage 7, Toulouse.

Verlaan, M., 1998: *Efficient Kalman Filtering Algorithms for Hydrodynamics Models*. Ph.D. thesis, Technische Universiteit Delft.

Verlaan, M. and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square filters. *Stochastic Hydrology and Hydraulics*, **11**, 349–368.

— 2001: Nonlinearity in data assimilation applications: A practical method for analysis. *Monthly Weather Review*, **129**, 1578–1589.

Vidard, A., F.-X. Le Dimet, and A. Piacentini, 2003: Optimal determination of nudging coefficient. *Tellus*, **55A**, 1–15.

Wenzel, M., J. Schröter, and D. Olbers, 2001: The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. *Progress in Oceanography*, **48**, 73–119.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed obsevations. *Monthly Weather Review*, **130**, 1913–1924.

Wick, G. A., 2002: Infrared and microwave remote sensing of sea surface temperature. Seminar at the University of Colorado at Boulder "Remote Sensing Seminar" graduate course.

Zaker, N. H., 2003: Computation and Modeling of the Air-Sea Heat and Momentum Fluxes. Lecture at the International Centre for Theoretical Physics.

Zavatarelli, M. and N. Pinardi, 2003: The Adriatic Sea modelling system: a nested approach. *Annales Geophysicae*, **21**, 345–364.

Ziehmann, C., 2000: Comparison of a single model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52A**, 280–299.