UNIVERSITÉ CATHOLIQUE DE LOUVAIN

FACULTÉ DES SCIENCES
DÉPARTEMENT DE PHYSIQUE

Model state and parameter estimation by assimilation of temperature profiles with an ensemble Kalman filter in North Sea applications

Doctoral dissertation presented by

Stéphanie Ponsar

in fulfillment of the requirements for the degree of Doctor in Sciences

Prof. Eric Deleersnijder (Supervisor)	Université catholique de Louvain
Dr Anouk de Brauwere	Université catholique de Louvain
Prof. Arnold Heemink	Delft University of Technology
Prof. Vincent Legat	Université catholique de Louvain
Dr Patrick Luyten	Institut Royal des Sciences Naturelles de Belgique
Dr Georges Pichot	Institut Royal des Sciences Naturelles de Belgique
Prof. Jean-François Remacle (Chair)	Université catholique de Louvain

LOUVAIN-LA-NEUVE
AUGUST 2012

A l'issue de ce travail, je souhaiterais remercier toutes les personnes qui, par leurs encouragements ou leur participation, ont contribué à son aboutissement. Ma reconnaissance s'adresse en particulier au Professeur Eric Deleersnijder qui a accepté de promouvoir cette thèse. Ik ben ook Doctor Patrick Luyten dankbaar voor zijn ideeën die de basis vormen van dit werk en voor zijn talrijke verbeteringen die de kwaliteit van deze tekst ten goede komen. Merci aussi au Docteur Georges Pichot qui a encouragé et soutenu ce projet.

Je tiens également à remercier le Président du jury, le Professeur Jean-François Remacle, et le Professeur Grégoire Winckelmans pour ses remarques sur le manuscript. Een hartelijk woordje van dank aan Professor Arnold Heemink voor onze verhelderende discussies. Merci aussi au Docteur Anouk de Brauwere pour ses commentaires, et au Professeur Vincent Legat qui m'ont fait l'honneur de juger ce texte.

Les Docteurs Laurent Bertino, Isabel Andreu-Burillo, Alexander Barth et Pierre De Mey m'ont permis de profiter de leur expérience en assimilation de données. Je les en remercie.

Merci à tous mes collègues de l'Unité de Gestion des Modèles Mathématiques de la Mer du Nord (UGMM), et en particulier à Geneviève, Géraldine, Katrijn, Yolande, Dries, Fabrice, Frank et Sébastien.

J'exprime toute ma reconnaissance à Monsieur José Ozer pour sa disponibilité et nos nombreuses discussions. J'adresse aussi un très grand merci aux Docteurs Valérie Dulière et Alain Norro pour leur aide et leurs conseils avisés.

Les simulations ont été réalisées au Centre Européen pour les Prévisions Météorologiques à Moyenne Échelle (ECMWF). Je remercie Monsieur Dominique Lucas pour son aide lors de l'installation du programme ainsi que Madame Liliane Frappez de l'Institut Royal Météorologique de Belgique. Merci aussi à Monsieur Serge Scory et à l'équipe informatique de l'UGMM, Lucien et Pierre-Alain, qui m'ont donné l'accès à des précieuses ressources de calcul.

Je souhaite aussi exprimer toute ma gratitude à mes parents ainsi qu'à Vincent pour leur soutien indéfectible durant ce long parcours.

Contents

Chapter 1

Introduction

1.1 Data assimilation: what and why

Nowadays, the prevision of storm-surges, of current profiles in the neighborhood of harbors, the prediction of the drift of oil slicks and of the morphological development of the coast rely on accurate modelling of coastal waters. However, there are several sources of model error: interaction processes at the surface and at the bottom are often badly constrained, the bathymetry, forcings, initial and boundary conditions are never known exactly. Therefore, it is necessary to adjust models in order to provide a good picture of the physical processes.

On the other hand observations are usually sparse in time or in space. Satellites provide only surface data with a very high spatial resolution in cloud free conditions whereas buoys measurements have a very high temporal resolution but are limited to a few fixed locations.

As the true state of the ocean is neither fully captured by the model nor by the observations, one way to improve the model representation of the sea state is to combine model and data, so that all the available information is used to this purpose. The technique for blending the dynamical model and the observations has been widely used and is known as data assimilation. This approach to data assimilation focuses mainly on model state estimation. In a sequential assimilation method, each assimilation step can be interpreted as a re-initialization of the model. Another motivation for data assimilation in oceanic models is the better understanding of the dynamics. This is closely related to model parameter estimation which is also allowed by the method. An optimal set of values can be derived for the badly constrained parameters which will afterwards be used to provide the forecasts.

At least two issues are related to these aims [Ghil and Malanotte-Rizzoli, 1991] [36]. The first one is how to determine variables not directly observed from observations. Answers are linked to the dynamical coupling between the variables. The second issue concerns the propagation of information by the dynamical model through the fluid flow:

- how the information provided locally can be transferred to ocean areas far away from the data dense location by dynamically connected to it,
- how to use the information available at one place, for example at the surface, to infer the state of the other parts, for example at depth.

The answers to these questions are related to the determination of temporal and spatial scales adequate for the assimilation of local data.

1.2 A brief history of data assimilation

There are mainly two classes of data assimilation methods: variational and sequential [Bouttier and Courtier, 2002] [9]. Variational methods minimize a cost function measuring the quadratic difference between the observations and the model solution over a given time interval. The model state is adjusted at all times by all the observations available over the estimation period. The more widely applied variational methods are 3D-VAR and 4D-VAR. The main difficulty with variational methods stems from the fact that the cost function can have multiple minima and that convergence is not guaranteed. In low data coverage areas, the global minimum might even be not physical and it can be hard to determine the proper minimum [Robinson et al., 1998] [84]. However, Vermeulen and Heemink, [2006] [102] propose an approach based on model reduction for variational data assimilation that does not require the computation of the adjoint of the original model and that is less sensitive to local minima.

In sequential methods, the model is integrated in time and when the model time reaches the instant at which observations are available the state predicted by the model is corrected with the observations. The integration of the model is then restarted from the updated state. The operation consisting of correcting the background estimate at a given time with new observations is called the analysis. Sequential assimilation is therefore an alternate sequence of analyses performed at observation times and of integrations of the model between successive analyses [Talagrand, 1997] [93]. Sequential data assimilation is illustrated on Figure 1.1. An initial state containing errors is integrated in time by the model operator (blue). At the time where observations are available (green), the model state is updated taking into account an estimate of the model errors as well as of the observations' errors. If the combination between the model state and the observations is properly performed, the resulting model state is closer to the true state represented in red.

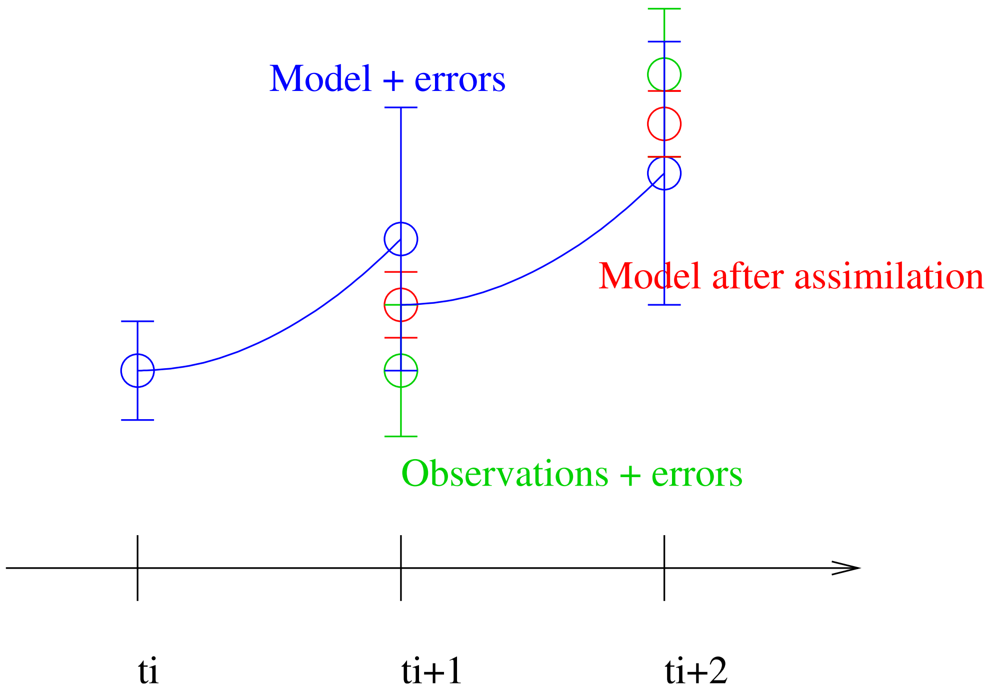

Figure 1.1: Sequential data assimilation. The model state and time integration are represented in blue, the observations in green and the true state in red.

The great advantage of sequential methods is that they provide explicit error estimates of the solution issued and that the derivation of a tangent linear operator of the model and backward integrations in time are not necessary [Ghil and Malanotte-Rizzoli, 1991] [36]. Their main drawback comes from the sequential character of the assimilation where each piece of observation influences the estimated state of the flow only at later times [Talagrand, 1997] [93].

Particle filters, also known as sequential Monte Carlo methods are sophisticated estimation techniques. As they do not need to assume gaussianity, they are more general than the Kalman filtering. The particles evolve in time with the numerical model and their assigned weights are updated each time new measurements are available. The filter solution is the weighted average of the particle ensemble [Hoteit et al., 2008] [44]. With sufficient samples (particles) they approach the Bayesian optimal state so that they can be more accurate than techniques based on the Kalman filter. However, while they are already current practice in the statistical community, their use in realistic data assimilation applications is still sparse because of the computing time they require.

Most of sequential assimilation methods are based on the Kalman filter [Kalman, 1960] [49] analysis step:

$$\psi^a = \psi^f + K(d - H\psi^f)$$

where ψ^a is the analysed (after assimilation) model state vector, ψ^f is the forecast state vector, M is the dynamical model operator, H is an operator mapping the model field to the observations' space, d is the observations' vector and K is the Kalman gain.

In the Kalman filter the computation of the Kalman gain requires that the error covariance matrix of the model state be stored and propagated in time which is computationally non feasible for realistic simulations in which the state vectors have large dimensions.

A classical method that is applied in operational forecasts is the optimal interpolation. It is an algebraic simplification of the computation of the Kalman weights K so that the analysis equation becomes a list of scalar analysis equations. While the Kalman filter allows a time evolution of the error statistics, the optimal interpolation neglects it. As a result the updating procedure is not consistent with the dynamics of the system which is critically important in coastal areas.

Another disadvantage of the Kalman filter is that the equations provide an optimal estimate only for linear models so that in a non linear context its solution is sub-optimal. A simple approximate method for non linear systems is the extended Kalman filter. It is a generalised Kalman filter for non linear systems in which the error covariance matrices are propagated in time from a tangent linear model (jacobian of the model operator M) and the non-linear observations' operator H is linearized around the model forecast. These linearizations can result in unbounded linear instabilities sot that the method can only be used for weakly non linear systems while it does not work properly for strongly non linear models.

Another efficient scheme allowing a time varying Kalman filter step is based on a reduced rank approximation of the error covariance matrix. In this case the error covariance matrix is approximated by a matrix of lower rank including only the most significant eigenvectors of the matrix. This reduced rank square root filter has been applied in several coastal area modelling applications, e.g. [Heemink et al., 1997] [41], [Verlaan and Heemink, 1997] [99], [Cañizares et al., 1998] [16], [Madsen and Cañizares, 1999] [66].

Other approaches based on the Kalman filter have been developed by Pham et al., [1998] [81] to reduce the prohibitive computational burden of the Kalman filter. They are the SEEK (Singular Evolutive Extended Kalman filter) and the SEIK (Singular Evolutive Interpolated Kalman filter) filters. Both filters use a low-rank approximation of the error covariance matrices. They differ by the forecast error in time: the SEEK filter uses the tangent linear model while the SEIK filter uses an ensemble technique. If a low rank approximation of the ensemble Kalman filter is used, the ensemble Kalman filter and the SEIK filters are equivalent [Nerger et al., 2007] [75]. These two filters have been successfully applied for realistic ocean simulations, e.g. [Brasseur et al., 1999] [11], [Brusdal et al., 2003] [14].

Provided that sufficient computer resources are available, ensemble methods constitute a good means to represent the uncertainties on forecasts taking into account the particularities of coastal dynamics [Mourre et al., 2004] [73]. The ensemble Kalman filter (EnKF) was initially developed for the global ocean by Evensen, [1994] [28]. It combines Monte Carlo methods with the traditional Kalman filter equations and provides a good approximation to the optimal Kalman filter solution [Annan et al., 2004] [4]. The error statistics of the system are propagated in time with the fully non linear model equations. Another scheme based on ensemble method is the ensemble smoother [van Leeuwen, 2001] [98]. In this case the ensemble represents the probability density function integrated over the whole assimilation period and all the data are processed in one step so that they influence the solution before and after their time.

The choice of the ensemble Kalman filter for this study has been mainly guided because it is adequate for non linear systems. Moreover, the proposed algorithm is suitable for parallel computation and can be relatively easily implemented. Then, although the EnKF is generally applied for model state estimation, parameter estimation can readily be included in the same framework by state space augmentation [Annan and Hargreaves, 2004 [4]; Anderson, 2001 [2]]. Numerous studies dealing with the EnKF may be found in ocean science and in atmospheric modelling. However, two major differences characterize data assimilation for the atmosphere and ocean [Ghil and Malanotte-Rizzoli, 1991] [36]. First of all, the oceanographic data set currently available is considerably smaller than the atmospheric one. Then, the second difference concerns the motivation. Data assimilation in meteorological models has been driven largely by the need for forecasts. Though this motivation is also present in physical oceanography, another pressing aspect is the optimization of the use of data sets expected in the near future to deepen and broaden our understanding of ocean circulation on coastal, regional, basin and global scales.

Amongst the applications in numerical weather forecasts, we report the studies that have had a large influence on the ocean implementations. Houtekamer and Mitchell focus on the assimilation of atmospheric data with the purpose of operational meteorological applications. To eliminate the need to estimate small correlations associated with observations located at remote distance of the assimilation point, they propose to localise the analysis. In a first study they propose to use a cut-off radius [Houtekamer and Mitchell, 1998] [45]. Then they have further applied the EnKF for operational atmospheric data assimilation with a particular focus on the localization of the analysis and on the sampling of model errors [Houtekamer and Mitchell, 2001] [46]. In this latter study [Houtekamer and Mitchell, 2001] [46], they multiply element by element the ensemble based covariance estimates with a distance dependent correlation function varying from 1 at the observation location to 0 at a specified radial distance. Hamill et al., [2001] [38] explore further the impact of distance-dependent filtering and show how the noise in the estimates of covariances affects the accuracy of the analyses. Houtekamer and Mitchell, [2005] [47] have further examined the importance of the model error representation by using a model error parameterized by adding noise consistent in structure with 3D-VAR background error covariances. A further step to real-life applications has been made by Zhang et al., [2004] [107] who have assessed the EnKF for the assimilation of more realistic radar observations at convective scale, by taking into account the variations in the quality and availability of these observations. Issues related to the application of a local ensemble square root filter for operational numerical weather prediction are investigated in Szunyogh et al. [2005] [92] using simulated observations. Finally, Meng and Zhang, [2007] [71] have shown how to use an EnKF for meso-scale and regional scale modelling of convective phenomena through observing system simulation experiments. A last atmospheric application worth mentioning is the forecasting of tropospheric ozone concentrations [Eben, 2005] [24].

In ocean modelling, the EnKF was primarily applied to the global ocean. One of its first realistic oceanography applications has been the assimilation of altimeter data in a quasi-geostrophic model of the Agulhas current [Evensen and van Leeuwen, 1996] [29]. Then, Keppenne and Rienecker [2002] [51] have implemented a parallel ensemble Kalman filter in an isopycnal ocean general circulation model. They have further discussed the impact of the assimilation of simulated equatorial temperature field on model currents [Keppenne and Rienecker 2003] [52]. Their study as well as that of Leeuwenburgh, [2005] [57] have shown that the time evolution of ensemble-based error covariances could capture seasonal and inter-annual changes in the mean state. As a result, by assimilation of satellite temperature, salinity and surface height measurements, the EnKF can also be applied to provide analyzed ocean states that will be used as initial conditions for seasonal forecasts [Leeuwenburgh, 2007] [58]. An EnKF is also implemented in the National Centers for Environmental Prediction operational global data assimilation system [Whitaker et al., 2008] [106]. The EnKF is also abundantly applied for the assimilation of ecological parameters. Its application to a 1-D ecological model has shown that it is able to handle non linear instabilities during the spring bloom [Eknes and Evensen, 2002] [27]. It has further been applied for the assimilation of ocean colour data in a biochemical model of the North Atlantic [Natvik and Evensen, 2003] [74], in a 1-D coupled hydro-ecological model of the Ligurian Sea [Lenartz et al., 2007] [56] and in a North Atlantic configuration of the coupled ocean ecosystem model HYCOM-NORWECOM [Simon and Bertino, 2009] [87]. Further applications are to be found in the assimilation of ice thickness data in a coupled ice-ocean model [Lisaeter et al., 2007] [59].

1.3 Scope and objectives

Coastal zones are complex systems, their dynamics is influenced by winds, tides and exchanges between deep seas and estuaries at open boundaries. Moreover their topography is abrupt and the geometry of land-sea boundaries is usually complex. The challenges of data assimilation in coastal zones stem from non linearities and the wide range of spatial and temporal scales of the physical processes such as tides and temperature and salinity fronts. A given region can be dominated by one type of regime and a neighboring one by a completely different type of phenomenon [Robinson et al., 1998] [84]. Therefore a large number of in situ or space borne measurements is necessary for assimilation in coastal zones [Echevin and De Mey, 2000] [25].

Up to now the wide variety of existing assimilation methods is reflected in coastal applications. A Kalman based method is used for operational forecasts of coastal water levels in the Netherlands [Verlaan et al., 2005] [101]. Using an optimal interpolation scheme, Oke et al., [2002] [76] assimilate surface currents in a high resolution model of the Oregon coast. Breivik and Saetra, [2001] [12] use a quasi-steady ensemble Kalman filter to assimilate high frequency radar currents in a nested model of the Norwegian Sea. However, their simulations are not able to correct the hydrodynamics of the area probably because of the absence of data assimilation in the outer model nested on their coastal zone. Surface winds in the German Bight are successfully estimated from the assimilation of surface current observations by Barth et al., [2010][6] with an ensemble smoother technique.

Although the ensemble Kalman filter has been widely used in many areas, its applications in coastal tidal zones remain up to now sparse, particularly in the North Sea. It has been applied for investigating the error representation in the assimilation of sea level measurements in a European shelf model [Mourre et al., 2004] [73]. Based on twin experiments of a hypothetical bay region, it has been assessed for data assimilation in a coastal area by Madsen and Cañizares, [1999] [66]. Kalman based methods including a reduced rank Kalman filter, an ensemble Kalman filter and an ensemble square root filter were also applied and compared for coastal data assimilation in three idealized regimes by Chen et al., [2009] [18]. However, for realistic North Sea simulations, the results reported here are pioneering work.

This study is the first application of the ensemble Kalman filter in 3-D simulations of the North Sea. Its aim is to provide a detailed assessment of the role that the ensemble Kalman filter can play in model state estimation as well as in combined model state and parameter estimation in a coastal area model.

Simulations dealing with the assimilation of temperature profiles have been performed for a 1-D case study and in a 3-D North Sea implementation of the COHERENS regional model developed by Luyten et al., [2005] [64], [2011] [65]. The 1-D and 3-D data assimilation experiments are the mirror of one another. At first stage, we directly started with the 3-D simulations. But soon, a better understanding of the functioning of the filter appeared necessary, so we moved to simplified 1-D simulations. They allowed to perform sensitivity tests to the values of model error sampling parameters and to better assess the impact of the filter on the model results. As a by-product, they made it possible to assess to what point the assimilation scheme could compensate for missing dynamical processes. However, as the physics of the 3-D simulations is much more complex than in the 1-D test case, the impact of the model error sampling parameters and of the filter on the model results is also investigated in the 3-D case.

The choice to assimilate temperature profiles is guided by the need to refine the modelling of the seasonal thermocline and because the sea temperature is one of the most important parameters in particular for ecological modelling [Holt and James, 1999] [42]. The thermocline constitutes a barrier between the chlorophyll present in the euphotic zone and the nutrients located in the lower layer of the water column. Then, as most ecological processes are exponential functions of the temperature, even small errors in its estimated magnitude play a large role in the misfit between observations and forecasts.

In the 1-D vertical test case, observations from the 'North Sea Project' database are assimilated while in the 3-D simulations data are generated from a higher horizontal resolution model run. Two algorithms of the ensemble Kalman filter already implemented in the COHERENS model are tested: a square root algorithm where observations are considered as perfect and a low rank square root algorithm allowing an ensemble representation of the observations' errors [Evensen, 2004] [32].

A series of numerical experiments have been performed to determine an adequate set of parameters for ensemble generation. They include the number of ensemble members, the vertical and horizontal standard deviations of the ensemble as well as the spatial and temporal decorrelation scales.

Moreover, as two dynamical regimes (well mixed and thermally stratified in summer) coexist in the North Sea, the conducted simulations allow one to examine the spatial variability of the model parameter estimation [Aksoy et al., 2006] [1]. The model parameters selected for estimation are primarily related to the representation of the temperature of the upper mixed layer, they are the optical attenuation and surface heat exchange coefficients.

1.4 Overview

The next chapter is dedicated to the description of the ensemble Kalman filter method and to the algorithms proposed by Evensen. In chapter 3, a brief description of the hydrodynamical component of the COHERENS model and of the numerical methods is given, with a particular focus on the parameterizations adopted for the simulations reported in this study. The issues related to the modelling of the North Sea temperature are also reviewed and all details concerning the simulations are provided. The discussion of the results is divided into four chapters. The EnKF is primarily applied for model state estimation. Results of simulations aiming at model state estimation are presented in chapter 4 (1-D test case) and chapter 5 (3-D simulations). Those dealing with the combined estimation of model state and parameters are given in chapter 6 and 7. Conclusions and perspectives are presented in the last chapter.

1.5 Supporting publications

This work is partly based on the content of the following refereed publications:

- Stephanie Ponsar and Patrick Luyten: Data assimilation with the EnKF in a 1-D numerical model of a North Sea station, *Ocean Dynamics*, **59**, 983-996, 2009 - doi: 10.1007/s10236-009-0224-3.

- Stephanie Ponsar, Patrick Luyten and Jose Ozer: Combined model state and parameter estimation with an ensemble Kalman filter in a North Sea station 1-D numerical model, *Ocean Dynamics*, **61**, 1869-1886, 2011 - doi: 10.1007/s10236-011-0477-5.

Chapter 2

Data assimilation with the ensemble Kalman filter

The data assimilation method is described in this chapter. At first we review how the combination of model and observations allows to refine model forecasts. The second section is dedicated to the traditional Kalman filter. Then, we describe the ensemble Kalman filter method. The fourth section deals with the generation of the ensemble which is the basis for model error representation. In the next section, we focus on the two square root algorithms that are effectively used in this study. Two applications of data assimilation are highlighted, they consist of model state estimation and combined model state and parameters estimation. Finally, the specific features of the implementation considered here are mentioned.

2.1 Model error sources and applications of data assimilation

Model error is present in all models and limits the accuracy of the predictions of the ocean state. By model error is meant the generic inaccuracies inherent in the physical and numerical model representation of the true state of the ocean. Indeed, a dynamical model always provides a simplification of the truth. There are many reasons for this. Physical processes may be incorrectly represented by the model formulation, the fluxes at the surface and bottom of the ocean are often badly constrained and model parameters may have approximate values. Then, the forcing fields and boundary conditions are never known exactly and errors may also exist on initial conditions and on the measures of bathymetry. Furthermore, a given model formulation with a structured grid may not capture features that are important to the coastal system.

Observations contain important information about the true state of the ocean. However, measurements may be contaminated by errors arising from different sources such as instrumental noise. Real observations may be sparse in space and time, and can only provide information at the scales they represent. Compared to the open ocean, in coastal areas an observation is able to correct the model only in a very limited neighborhood.

As both model and observations contain errors, a complete and accurate picture of the ocean can be provided by properly balancing information from dynamical model and observations [Ghil and Malanotte-Rizzoli, 1991] [36]. This can be achieved by means of data assimilation which aims at deriving an optimal from model forecasts, observations and given error statistics [Natvik and Evensen, 2003] [74].

The model uncertainties enumerated above stem from different error sources: errors in the specification of the forcings, initial and open boundary conditions and imperfections in the model formulation [Ehrendorfer and Tribbia, 1997] [26]. Therefore there are several applications to data assimilation: the estimation of errors on model forecasts, the refinement of the representation of the dynamical processes and the estimation of model parameters [Robinson and Lermusiaux, 2000] [85]. Coastal model calibration by trial and error comparisons with data is cumbersome, parameter estimation via data assimilation is thus an important issue for accurate coastal simulations [Robinson et al., 1998] [84].

2.2 The Kalman filter

2.2.1 Model state vector and observations vector

The model state vector ψ contains the n model state variables:

$$\psi = (x_1, \cdots, x_n)^T, \tag{2.1}$$

while the observations vector d contains m measurements:

$$d = (d_1, \cdots, d_m)^T. \tag{2.2}$$

In practical applications, observations are not available for all the model variables and the dimension of the model state space is different from that of the observations' space. In order to convert the information contained in the model state vector to the observations vector, an observations' operator H is necessary:

$$d = H\psi. \tag{2.3}$$

2.2.2 Error covariances to represent uncertainties

Given a forecast field ψ^f, there is one and only one vector of errors that separates it from the true state ψ^{tr}, so that the model error is defined as follows [Bouttier and Courtier, 2002] [9]:

$$\epsilon^f = \psi^f - \psi^{tr},$$

and the observation errors are defined as:

$$\epsilon^0 = d - H[\psi^{tr}].$$

As the true state is unknown, the model and the observation errors are unknown but it is assumed that they are statistically unbiased:

$$E[(\psi^f - \psi^{tr})] = 0,$$

$$E[(d - H[\psi^{tr}])] = 0,$$

and that the model errors and the observations errors are uncorrelated:

$$E[(\psi^f - \psi^{tr})(d - H[\psi^{tr}])^T] = 0,$$

$E[.]$ denotes the expected value (statistical average) [Bouttier and Courtier, 2002] [9]. The hypothesis of unbiased errors is a difficult one in practice because there are often significant biases in the forecast model (e.g. a model drift) and in the observations. If the biases are known, they can be subtracted from the forecast and observations values, [Bouttier and Courtier, 2002] [9]. On the other hand, the hypothesis of uncorrelated model errors and measurement errors could be dealt with in the same way if specifications of the correlation matrix could be given, which is quite difficult. As a result, it is usually assumed that the causes of errors in the model forecast and in the observations are completely independent, [Bouttier and Courtier, 2002] [9].

The model error covariances are defined as:

$$P^f = E[(\psi^f - \psi^{tr})(\psi^f - \psi^{tr})^T],$$

they represent the gap between the true state of the ocean and the state given by the model.

2.2.3 Description of the Kalman filter

When a stochastic description of the model and measurements is available in the following form:

$$\psi^{tr}(t_{i+1}) = M_i[\psi^{tr}(t_i)] + q(t_i), \tag{2.4}$$

$$d(t_i) = H_i[\psi^{tr}(t_i)] + v(t_i), \tag{2.5}$$

with M_i the transition operator from time t_i to time t_{i+1}, $q(t_i)$ a white Gaussian noise of covariance matrix Q_i and $v(t_i)$ a white Gaussian noise of covariance matrix R_i, it is possible to incorporate measurements into the model to obtain an optimal estimate of the model state [Jazwinski, 1970] [48].

The Kalman filter, [Kalman, 1960] [49] is a method for solving the sequential data assimilation problem consisting of estimating $\psi^{tr}(t_{i+1})$ given observations up to time t_{i+1}, starting from an estimate $\psi^a(t_i)$ at time t_i and corresponding error covariance matrix $P^a(t_i)$. This is done in two steps: the forecast and analysis steps.

At the forecast step, the integration of the model is restarted from the updated state:

$$\psi^f(t_{i+1}) = M_i[\psi^a(t_i)], \tag{2.6}$$

and the error covariance matrix is propagated with the model dynamics:

$$P^f(t_{i+1}) = M_i P^a(t_i) M_i^T + Q(t_i). \tag{2.7}$$

At the analysis step, the model forecast is updated with the observations available at time t_{i+1}. Given a vector of model forecast values ψ^f of error covariance matrix P^f, a vector of observations d of error covariance matrix R, a model integration operator M_i, and a transform H which maps the field of model variables to their observational equivalents, then the estimate ψ^a can be written as:

$$\psi^a(t_{i+1}) = \psi^f(t_{i+1}) + K_{i+1}[d(t_{i+1}) - H_{i+1}[\psi^f(t_{i+1}])], \tag{2.8}$$

where K_{i+1} is the Kalman gain given by:

$$K_{i+1} = P^f(t_{i+1}) H_{i+1}^T [H_{i+1} P^f(t_{i+1}) H_{i+1}^T + R(t_{i+1})]^{-1}, \tag{2.9}$$

and the term $d(t_{i+1} - H_{i+1}\psi^f(t_{i+1})$ is referred to as the innovation vector. The analysis error covariance matrix is calculated as:

$$P^a(t_{i+1}) = (I - K_{i+1} H_{i+1}) P^f(t_{i+1}), \tag{2.10}$$

which states that the uncertainty of the analysed state is smaller or equal to that of the forecast state. The estimate ψ^a is optimal if the operators M_i and H are linear.

The forecast equations can also be thought of as a predictor equations while the analysis equations can be thought of as a corrector equations so that the final estimation algorithm looks like a predictor-corrector algorithm [Welch and Bishop, 2006] [104].

2.3 Description of the ensemble Kalman filter

The updating scheme (analysis step) of the Kalman filter requires to know the error covariance matrix of the model forecast, P^f. Given the size of this matrix in practical applications, the computation of its elements is prohibitive. Moreover, as the true state of the system is unknown, the exact computation of the error covariance matrix is just not possible. Provided that sufficient computer resources are available, ensemble methods provide a number of model instances and constitute a good means to estimate the error statistics [Mourre et al., 2004] [73].

The ensemble Kalman filter (EnKF) is a sequential data assimilation method developed by Evensen [1994, 2003, 2004] [28], [31], [32]. It combines the traditional Kalman filter with Monte Carlo methods to generate an ensemble of states. A first guess of the model state is introduced on input, then an ensemble of model states is generated from random perturbations of the badly constrained variables and/or parameters of the initial state. The ensemble members are simultaneously integrated forward in time until measurements are available [Eknes and Evensen, 2002] [27]. At that stage, the analysis step is applied to all the members of the ensemble and the mean of the ensemble is also updated through a multivariate analysis computed from the ensemble statistics. Thus, a change in one of the model variables will influence the other variables. As a result the analysis step explicitly constructs an analysed ensemble to serve as initial state for the next forecast steps [Snyder and Zhang, 2003] [90].

The symbols used for the description of the EnKF are listed in Table 8.1 in Appendix A. Let A be the matrix representing an ensemble of N model state vectors of dimension n:

$$A = (\psi_1, \psi_2, \ldots, \psi_N), \tag{2.11}$$

the ensemble mean is stored in each column of the matrix $\bar{A}$ which can be defined as:

$$\bar{A} = \frac{1}{N} \sum_{j=1}^{N} \psi_j = A 1_N, \tag{2.12}$$

where 1_N is a square matrix of dimension $N \times N$ where each element is equal to $\frac{1}{N}$. A matrix of ensemble perturbations with respect to the ensemble mean is defined as:

$$A' = A - \bar{A} = A(I - 1_N), \tag{2.13}$$

where I is the identity matrix. The model error covariance matrix is estimated from the ensemble of perturbations with respect to the ensemble mean:

$$P_e = \frac{A'(A')^T}{N-1}. \tag{2.14}$$

The matrix of error covariances of the measurements, R, can be prescribed from the knowledge of the accuracy of the measurements. Given a vector of measurements d of size m (the number of measurements), an ensemble of N vectors of perturbed observations can be defined as:

$$d_j = d + \epsilon_j, \ j = 1, \ldots, N, \tag{2.15}$$

from which an ensemble representation of the measurement error covariance matrix can be built in the form of a matrix D:

$$D = (d_1, \ldots, d_N), \tag{2.16}$$

and the mean of the ensemble of observations is stored in each column of the matrix $\bar{D}$ which can be defined as:

$$\bar{D} = \frac{1}{N} \sum_{j=1}^{N} d_j = D 1_N, \tag{2.17}$$

while the ensemble of perturbations of the measurements can be stored in the matrix E:

$$E = (\epsilon_1, \ldots, \epsilon_N). \tag{2.18}$$

Then, the error covariance matrix of the observations estimated from an ensemble of measurements has the following low-rank representation:

$$R_e = \frac{EE^T}{N-1}. \tag{2.19}$$

The forecast and analysis steps of the EnKF are those of the Kalman filter given from equations (2.6) - (2.9) where the analysis step is applied to each ensemble member. The dimensions of the matrices involved in the above description are given in Table 8.1 in Appendix A.

Although based on the optimal Kalman filter equations, a couple of approximations are inherent to the method. First of all, as the number of ensemble members N is small with respect to the dimension of the model state vector n, the ensemble of model state errors represented by the matrix A' will never be able to provide an exact representation of all possible model error covariances. As a result this may lead to an excessive decrease of the ensemble variance and possibly to filter divergence. This issue can be dealt with by localizing the analysis step. Another limitation stems from the sampling error when stochastic perturbations of the measurements are introduced. This can be avoided through square root implementation of the analysis scheme. Another approximation inherent to all Kalman filters is that the optimality of the analysed state relies on the assumption that the error statistics are Gaussian and on the linearity of the operators H and M [Gelb, 1974] [35]. If the model is non linear, the time evolution of the error distribution will not be Gaussian and the ensemble mean will no longer be the maximum likelihood estimate [Evensen and van Leeuwen, 2000] [30], but an estimate of the state that minimizes the root mean square forecast error [Burgers et al., 1998] [15].

However, while a linear equation is applied when computing the analysed state (equation (2.8)) [Haugen and Evensen, 2002] [40], the strength of the ensemble Kalman filter is that the error covariance matrix is propagated in time via the fully non linear model equations [Annan et al., 2005] [5] .

2.4 Generation of the ensemble

The specification of model error is very important for the data assimilation process to work properly. In the ensemble Kalman filter the model error is represented by an ensemble of model states which must be chosen adequately to represent the error statistics of the model. An ensemble of model states is generated by adding perturbations to a best-guess estimate, and then the members are integrated over a time interval covering a characteristic time scale of the dynamical system. This is performed to ensure that the system is in dynamical equilibrium, that multivariate correlations between the variables of the state vector have developed and that correlations between the model stochastic terms and the model variables are present. Each individual member of the ensemble is integrated as a stochastic differential equation of the type of equation (2.4) i.e. forced with a random noise component which represents the model error.

There are several ways to generate perturbations of the system: perturbations of the forcing fields of the system, perturbations of model state vector, perturbations of the model parameters [Zheng and Zhu, 2008] [108]. In this study, the last two approaches are investigated. When perturbing the model state vector, the model parameters are kept fixed with predefined values. In a second step the perturbation of the model state vector is combined with the perturbation of model parameters.

The ensemble Kalman filter implemented in the COHERENS model allows to generate horizontally and vertically correlated model error terms. The generation of these correlations is described below and follows the one given by Evensen, [2003] [31].

2.4.1 Generation of horizontal model error terms

A horizontal distribution of random model error terms with specific mean and variance is generated through the use of their Fourier transform discretised on a $N \times M$ grid:

$$q_h(x_n, y_m, :, :) = \sum_l \sum_p \hat{q}_h(\kappa_l, \lambda_p, :, :)\Delta k e^{i(\kappa_l x_n + \lambda_p y_m)} \tag{2.20}$$

with $x_n = n\Delta x$, $y_m = m\Delta y$, $\kappa_l = \frac{2\pi l}{N\Delta x}$, $\lambda_p = \frac{2\pi p}{M\Delta y}$ and $\Delta k = \Delta\kappa\Delta\lambda = \frac{(2\pi)^2}{NM\Delta x\Delta y}$. If we define their Fourier transform as:

$$\hat{q_h}(\kappa_l, \lambda_p, :, :) = \frac{c}{\sqrt{\Delta k}} e^{-(\kappa_l^2+\lambda_p^2)/\sigma^2} e^{2\pi i\phi_{l,p}} \tag{2.21}$$

where $\phi_{l,p} \in [0, 1]$ is a pseudo-random number introducing a phase shift. To ensure that the pseudo-random fields are real, the condition :

$$\hat{q_h}(\kappa_l, \lambda_p, :, :)^* = \hat{q_h}(\kappa_{-l}, \lambda_{-p}, :, :) \tag{2.22}$$

must be satisfied, where $*$ denotes complex conjugate. As a result the expression of the horizontal model error terms (equation 2.20) becomes:

$$q_h(x_n, y_m, :, :) = \Delta k \sum_{l,p} \frac{c}{\sqrt{\Delta k}} e^{-(\kappa_l^2+\lambda_p^2)/\sigma^2} e^{2\pi i\phi_{l,p}} e^{i(\kappa_l x_n+\lambda_p y_m)}. \tag{2.23}$$

The above equation (2.23) contains two unknowns c and σ which determine the specific covariance of the distribution of horizontal model error terms. Two equations for these unknowns can be derived as follows. The variance of the distribution at a location (x,y) is given by:

$$E[q_h(x, y, :, :)q_h(x, y, :, :)] = \Delta k c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2}, \tag{2.24}$$

by requiring that the variance be equal to 1, this equation becomes:

$$\Delta k c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2} = 1, \tag{2.25}$$

and provides an expression for c. Then, by requiring that two fields be correlated at a distance μ (i.e. that their correlation decreases by a factor e at a distance μ), a second equation can be derived:

$$E[q(0)q(\mu)] = \Delta k c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2} \cos(\kappa_l \mu) = e^{-1}, \tag{2.26}$$

in which the expression for c can be replaced so that this equation is a non linear scalar equation for σ which can be solved using the algorithm given by Brent, [1973] [13].

2.4.2 Vertical correlation

In the context of coastal dynamics, a specific vertical correlation between the values of the fields at different water depths may exist. In the well mixed areas, the stirring caused by tides and winds ensures a vertical mixing of the whole water column. By contrast, in deeper areas, when the mixing is weaker and when the solar heating increases in summer, a robust two-layer structure develops. Such correlations can be included in the ensemble of states representing error terms.

In the 1-D case, w, a sequence of pseudo-random numbers with mean equal to 0 and variance equal to 1 is generated following the algorithm given by L'Ecuyer and Côté [1991] [55]. A vertical spatial correlation between the model error terms is expressed in terms of the vertical level of the model grid k:

$$q_v(:,:,k,:) = \alpha_v q_v(:,:,k-1,:) + \sqrt{1-\alpha_v^2}\gamma w(k,:) \tag{2.27}$$

where the multiplicative factor γ modulates the magnitude of the pseudo-random numbers. In the case of 3-D simulations, the random characteristic is introduced in the horizontal error terms and the vertical correlation is imposed as:

$$q_v(:,:,k,:) = \alpha_v q_v(:,:,k-1,:) + \sqrt{1-\alpha_v^2}\gamma q_h(:,:,k,:). \tag{2.28}$$

where γ modulates the magnitude of the horizontal error terms.

The factor α_v is related to the vertical spacing and to a vertical decorrelation length scale ξ. When excluding the random sequence of equation (2.27), this equation becomes a differential equation of the following form:

$$\frac{\partial q_v}{\partial z} = \frac{-1}{\xi} q_v \tag{2.29}$$

which states that q is damped by a factor e on a length scale $z = \xi$ called the vertical decorrelation length. An approximation of the solution of this equation is given by:

$$q_v(:,:,k,:) = (1 - \frac{\Delta z}{\xi})q_v(:,:,k-1,:) \tag{2.30}$$

where Δz is the vertical grid spacing. So that α_v can be expressed as:

$$\alpha_v = 1 - \frac{\Delta z}{\xi}. \tag{2.31}$$

Two limit cases clarify the role of ξ: if the sequence of stochastic terms is vertically uncorrelated, the vertical decorrelation scale is of the same order of magnitude as the vertical grid spacing, $\xi = \Delta z$ ($\alpha_v = 0$), so that the model error is totally random and equal to w. If the vertical decorrelation scale ξ tends to infinity ($\alpha_v = 1$), there is no random term in the sequence of model noise which becomes deterministic.

The magnitude of the vertical error terms q_v is modulated by a multiplicative factor η representing the magnitude of the vertical model error so that:

$$q(t) = \eta q_v(:,:,k,:). \tag{2.32}$$

2.4.3 Specific features of the ensemble generation in COHERENS

At the sampling step, the ensemble members contain only the deviations with respect to the ensemble mean. At each sampling time step small pseudo-random perturbations are added to each ensemble member. This allows to progressively generate Gaussian perturbations with a sufficiently large standard deviation of the ensemble while not shaking too strongly the model dynamics by the addition of large error terms at once. As soon as data are assimilated the standard deviation of the ensemble decreases. After the beginning of the assimilation process small random perturbations are still added to the ensemble members at the sampling stage in order to prevent the under-estimation of the model error.

2.5 Square root algorithms for the analysis scheme of the EnKF

Although the Kalman filter provides the optimal estimate of a state given the measurements, the traditional equations are not a good choice from a numerical point of view. Square root algorithms provide an algebraically equivalent but numerically more robust alternative to the traditional Kalman equations [Verlaan, 1998] [100]. They avoid numerical problems by using the square root of the error covariance matrix. If the matrix P is hermitian (symmetric with real elements) and positive semi-definite, then it can be decomposed as $P = LL^*$ where L is a lower triangular matrix and L^* denotes the conjugate transpose of L. By construction, the error covariance matrix remains implicitly symmetrical and positive semi-definite. All computations involved by this algorithm deal with the factor L directly and the error covariance matrix P is never computed explicitly. The matrix elements L_{kl} have a much smaller range of eigenvalues and square root algorithms are therefore numerically better conditioned than the original Kalman filter algorithm. Most square root algorithms have the same predictor-corrector structure as the original Kalman filter [Verlaan, 1998] [100].

A description of square root filters may be found in Tippett et al., [2003] [95]. The study presented here is based on two square root algorithms for the EnKF analysis step proposed by Evensen, [2004] [32]. These algorithms have been chosen for their ease of implementation and for their computational efficiency. They do not add a further computational cost with respect to the traditional ensemble Kalman filter and provide better results [Evensen, 2004] [32]. Both algorithms use the dominant singular vectors of the ensemble.

The analysis equation can be expressed in terms of the ensemble matrices as:

$$A^a = A^f + P^f H^T (HP^f H^T + R)^{-1}(D - HA^f). \tag{2.33}$$

Let's define the matrix S holding the model ensemble perturbations in the measurements' space:

$$S = HA', \tag{2.34}$$

$\in R^{m \times N}$ and of rank $N - 1$, and the matrix C:

$$C = SS^T + (N - 1)R, \tag{2.35}$$

and let's introduce them in the above equation which becomes:

$$A^a = A^f [I + S^T C^{-1}(D - HA^f)], \tag{2.36}$$

and define the matrix X as:

$$X = [I + S^T C^{-1}(D - HA^f)], \tag{2.37}$$

so that:

$$A^a = A^f X. \tag{2.38}$$

In square root algorithms, the ensemble perturbations and the ensemble mean are updated separately. The analysis step for the update of the model error covariance matrix can be written as:

$$P^a = P^f - P^f H^T (HP^f H^T + R)^{-1} HP^f, \tag{2.39}$$

when using the ensemble representation for the model error covariance matrix, P, this equation expresses the analysis step for the ensemble perturbations in terms of the matrices S and C:

$$A^{a'} A^{a'T} = A'(I - S^T C^{-1} S) A'^T. \tag{2.40}$$

To ensure that the ensemble is the best possible representation of the error covariance matrix for a given number of ensemble members, the mean of the ensemble perturbations matrix A' must be zero, so that an eventual mean must be subtracted. Removing the mean of the ensemble leaves the maximum possible rank of A' to be $N - 1$ [Evensen, 2004] [32]. Then the analysed ensemble mean is computed from the Kalman filter analysis equation as:

$$\bar{A}^a = \bar{A}^f + A' S^T C^{-1} (\bar{D} - H\bar{A}^f), \tag{2.41}$$

or

$$\bar{\psi}^a = \bar{\psi}^f + A' S^T C^{-1} (\bar{d} - H\bar{\psi}^f). \tag{2.42}$$

2.5.1 Role and issues related to an ensemble representation of the observations' errors

The success of the analysis step in the EnKF depends on the diversity of the ensemble members represented by the ensemble variance. Houtekamer and Mitchell, [1998] [45] have found that the ensemble members tend to converge when observations are considered as perfect ($R = 0$). This problem arises when the same unperturbed observations are used to update all the ensemble members at the analysis stage. To prevent this, they proposed to perform the analysis update with a perturbed set of observations that are different for each ensemble member. Burgers et al., [1998] [15] provided a theoretical justification for adding random perturbations to the observations in EnKF applications. They pointed out that if the observations are not treated as random variables and hence if the same observations and the same gain are used to update each member of the ensemble, analysis error covariances will be systematically too low, and may lead to filter divergence.

For large ensembles, this problem can be alleviated by assimilating an ensemble of perturbed observations whose statistics reflect the observations errors [Whitaker and Hamill, 2002] [105]. However, as ensemble sizes are finite, the noise added to the observations might induce unwanted observation error covariances, hence reducing the accuracy of the analysis [Whitaker and Hamill, 2002] [105], [Anderson, 2001] [2].

When observations error and model error covariance matrices are both represented by an ensemble, the matrix C can be rank deficient when the number of observations is greater or equal to the ensemble size (m $\geq$ N) [Kepert, 2004] [50]. The proof given by Kepert, [2004] [50] is the following.

As R is symmetric and positive semi-definite, there exists an orthonormal matrix $F \in R^{m\times m}$ such that:

$$\tilde{R} = FRF^T$$

where $\tilde{R}$ is diagonal.

Then, D which is the matrix of the perturbed observations can be transformed by F so that:

$$\tilde{D} = FD,$$

and the observations' operator for $\tilde{D}$ is FH and the transformed observations error covariance matrix is $\tilde{R}$. Precisely $m - N + 1$ of the diagonal elements of R are equal to zero because R is of rank $N - 1$. As a result, when an ensemble representation of the observations' error is considered, $m - N + 1$ of the transformed observations are perfect.

Lorenc, [2003] [61] shows that assimilating one perfect observation in the EnKF removes one degree of freedom form the ensemble so that, in this case, the rank of the analysis ensemble becomes equal to$N - (m - N + 1)$ $= 2N - m - 1$. Lorenc's argument indicates that this is valid only if the ensemble has rank > 1. Hence, a complete collapse of the analysis ensemble occurs when $2N - m - 1 \leq 1$. As a result, rank deficiency in the analysis ensemble occurs when $2N - m - 1 <$ N or for $N \leq m$.

This issue is related to the fact that, as it is nearly diagonal, the observations' error covariance is not adequate for a Monte Carlo representation which may lead to a substantial systematic bias in its eigenvalues. This problem is less critical for the model error covariance matrix which has a broader spectrum [Kepert, 2004] [50].

In the framework of this study, we will examine and compare the performance of two square root algorithms for the analysis: the square root algorithm without perturbation of the measurements and the low rank square root with an ensemble representation of the measurements' errors.

2.5.2 Square root algorithm without ensemble for measurements error

Because of the issues related to the perturbation of the observations, Evensen, [2004] [32] proposes a square root algorithm for the analysis scheme without perturbations of measurements. This algorithm provides a significant improvement of the results because of the removal of sampling errors introduced by the perturbations of measurements. Its derivation starts by forming C as defined in equation (2.35) with $R = 0$ so that C is of full rank and C^{-1} exists. Therefore the eigenvalue decomposition of $C = Z\Lambda Z^T$ can be computed, and:

$$C^{-1} = Z\Lambda^{-1}Z^T \tag{2.43}$$

where all matrices are of dimension $m \times m$ and Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues of A.

The equation (2.40) of the analysis step for the ensemble perturbations can be written as follows:

$$\begin{aligned} A^{a'}A^{a'T} &= A'(I - S^T Z\Lambda^{-1}Z^T S)A'^T & (2.44)\\ &= A'(I - X_2^T X_2)A'^T & (2.45) \end{aligned}$$

where $X_2 = \Lambda^{-\frac{1}{2}}Z^T S \in R^{m\times N}$ is a rectangular matrix whose singular value decomposition is:

$$X_2 = U_2\Sigma_2 V_2^T, \tag{2.46}$$

where $U_2 \in R^{m\times m}$, $\Sigma_2 \in R^{m\times N}$ and $V_2 \in R^{N\times N}$. U_2 and V_2 are orthogonal matrices, and Σ_2 contains $p = min(m, N)$ singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ on the diagonal. When the singular value decomposition of X_2 is introduced in equation (2.45), this equation takes the "square root" form:

$$\begin{aligned} A^{a'}A^{a'T} &= A'(I - [U_2\Sigma_2V_2^T]^T[U_2\Sigma_2V_2^T])A'^T & (2.47)\\ &= (A'V_2\sqrt{I - \Sigma_2^T\Sigma_2})(A'V_2\sqrt{I - \Sigma_2^T\Sigma_2})^T. & (2.48) \end{aligned}$$

Therefore, the analysis step for the ensemble perturbations is given by:

$$A^{a'} = A'V_2\sqrt{I - \Sigma_2^T\Sigma_2}, \tag{2.49}$$

and for the ensemble mean it is:

$$\bar{A}^a = \bar{A}^f + A'S^T(Z\Lambda^{-1}Z^T)(\bar{D} - H\bar{A}^f). \tag{2.50}$$

The whole analysed ensemble is obtained as:

$$A^a = \bar{A}^a + A'^a. \tag{2.51}$$

2.5.3 Low rank square root analysis scheme with ensemble for the representation of measurements' error

In this section, we follow the derivation of the low rank square root algorithm given by Evensen, [2004] [32] where a low rank representation is used for the measurements' error covariance matrix R. This results in a computationally stable and efficient algorithm.

When $R \neq 0$, C is a matrix of dimension $m \times m$. If the number of available measurements is low, $m < N$, then if C is of full rank its inverse can be computed. However, when the number of measurements is large, $m \geq N$, the ensemble representation for the measurements' error covariance matrix may lead to a loss of rank of the matrix C. In this case the pseudo-inverse $C^{\dagger}$ can be computed and is obtained from a singular value decomposition.

Let's define the following eigenvalue factorization of the matrix C:

$$C = Z \Lambda Z^T, \tag{2.52}$$

the factorization for its pseudo-inverse is given by:

$$C^{\dagger} = Z \Lambda^{\dagger} Z^T, \tag{2.53}$$

where Λ is a diagonal matrix of rank $p = rank(C)$ and:

$$diag(\Lambda^{\dagger}) = (\lambda_1^{-1}, \ldots, \lambda_p^{-1}, 0, \ldots, 0) \tag{2.54}$$

Let's assume that $rank(S) = N - 1$, which is the case if the ensemble is properly chosen and if the measurement operator is of full rank. The singular value decomposition of S is:

$$S = U_0 \Sigma_0 V_0^T, \tag{2.55}$$

with $U_0 \in R^{m \times m}$, $\Sigma_0 \in R^{m \times N}$ and $V_0^T \in R^{N \times N}$. The subspace of S is defined by the $N-1$ first singular vectors of S contained in U_0. The pseudo-inverse of S is defined as:

$$S^{\dagger} = V_0 \Sigma_0^{\dagger} U_0^T, \tag{2.56}$$

where $\Sigma_0^{\dagger} \in R^{N \times m}$ is a diagonal matrix such that:

$$diag(\Sigma_0^{\dagger}) = (\sigma_1^{-1}, \sigma_2^{-1}, \ldots \sigma_{N-1}^{-1}, 0). \tag{2.57}$$

When the expressions of S and $S^{\dagger}$ are introduced in the expression of C, equation (2.35) becomes:

$$C = SS^T + (SS^{\dagger})(N-1)R(SS^{\dagger})^T, \tag{2.58}$$

the low rank expression of R is: $(N-1)R = EE^T$ so that the expression of C is given by:

$$\begin{aligned} C &= SS^T + (SS^\dagger)EE^T(SS^\dagger)^T, && (2.59) \\ &= SS^T + \hat{E}\hat{E}^T, && (2.60) \end{aligned}$$

where

$$\hat{E} = (SS^\dagger)E, \tag{2.61}$$

is the projection of E on the $N-1$ first singular vectors contained in U_0. Let's replace S and $S^\dagger$ by their singular value decompositions so that the expression of C becomes:

$$\begin{aligned} C &= U_0\Sigma_0[I + \Sigma_0^\dagger U_0^T EE^T U_0 \Sigma_0^{\dagger T}]\Sigma_0^T U_0^T && (2.62) \\ &= U_0\Sigma_0[I + X_0X_0^T]\Sigma_0^T U_0^T && (2.63) \end{aligned}$$

where

$$X_0 = \Sigma_0^\dagger U_0^T E \tag{2.64}$$

is of dimension $N \times N$ and of rank $N-1$. Its singular value decomposition is:

$$X_0 = U_1\Sigma_1 V_1^T \tag{2.65}$$

where all the matrices are of dimension $N \times N$. If we replace X_0 in the expression of C given at equation (2.63), we get:

$$\begin{aligned} C &= U_0\Sigma_0[I + (U_1\Sigma_1V_1^T)(U_1\Sigma_1V_1^T)^T]\Sigma_0^T U_0^T && (2.66) \\ &= U_0\Sigma_0 U_1[I + \Sigma_1^2]U_1^T\Sigma_0^T U_0^T. && (2.67) \end{aligned}$$

In the case where C is singular, its pseudo-inverse is expressed as:

$$\begin{aligned} C^\dagger &= (U_0\Sigma_0^{\dagger T}U_1)(I + \Sigma_1^2)^{-1}(U_0\Sigma_0^{\dagger T}U_1)^T && (2.68) \\ &= X_1(I + \Sigma_1^2)^{-1}X_1^T && (2.69) \end{aligned}$$

where $X_1 \in R^{m\times N}$ is of rank (N-1) and is defined as:

$$X_1 = U_0\Sigma_0^{\dagger T}U_1. \tag{2.70}$$

Let's take the the singular value decomposition of S defined in equation (2.55), the definition of C given by equation (2.35) and let's replace its pseudo-inverse defined at equation (2.69) in equation (2.40), we get the following "square root" expression for the analysis of the ensemble perturbations:

$$\begin{aligned} A^{a'}(A^{a'})^T &= A'[I - S^TC^\dagger S]A'^T && (2.71) \\ &= A'[I - S^TX_1(I+\Sigma_1^2)^{-1}X_1^TS]A'^T && (2.72) \\ &= A'[I - ((I+\Sigma_1^2)^{-\frac{1}{2}}X_1^TS)^T((I+\Sigma_1^2)^{-\frac{1}{2}}X_1^TS)]A'^T && (2.73) \end{aligned}$$

if X_2 is defined as:

$$
\begin{aligned}
X_2 &= (I + \Sigma_1^2)^{-\frac{1}{2}} X_1^T S &(2.74)\\
&= (I + \Sigma_1^2)^{-\frac{1}{2}} (U_0 \Sigma_0^{\dagger T} U_1)^T (U_0 \Sigma_0 V_0^T) &(2.75)
\end{aligned}
$$

the above equation becomes:

$$A^{a'}(A^{a'})^T = A'(I - X_2^T X_2)A'^T. \tag{2.76}$$

As the singular value decomposition of X_2 is given by:

$$X_2 = U_2 \Sigma_2 V_2^T,$$

so that the analysis step for the ensemble perturbations is given by:

$$A^{a'} = A' V_2 \sqrt{I - \Sigma_2^T \Sigma_2} \tag{2.77}$$

and for the ensemble mean it is:

$$\bar{A}^a = \bar{A}^f + A' S^T X_1 (I + \Sigma_1^2)^{-1} X_1^T (\bar{D} - H\bar{A}^f), \tag{2.78}$$

and we end-up with the same final equations as for the square root algorithm. The whole analysed ensemble is also obtained as:

$$A^a = \bar{A}^a + A'^a. \tag{2.79}$$

2.6 Combined parameter and state estimation

In this section, we describe the approach that is used for combined model state and parameter estimation. We define the estimation state vector and how it is modified for combined model state and parameter estimation.

2.6.1 Estimation state vector

The model state vector represents all the variables of the model. However, in practice, it is impossible to estimate all the correlations between the different state variables and an estimation state vector of smaller dimensions is used [Magri, 2002] [67]. The estimation vector contains all the variables that are updated at the analysis step. The variables not included will be adjusted by the model dynamics. Starting from a given observations' vector (the temperature profiles), we have to identify the state variables that can not be included in the estimation vector because they can not be linked to the observed variables using a Kalman filtering approach, for instance in case of strongly non linear relationships between the observed and estimated variables [Raick, 2006] [83].

2.6.2 Combined model state and parameter estimation

Generally speaking, the objective of parameter estimation is to improve the estimates of a set of poorly known model parameters in order to get model forecasts that are closer to the observations. The underlying assumption is that all errors in the model state representation are associated to uncertainties in the selected set of parameters and that the model representation as well as the initial and boundary conditions are exactly represented. This is equivalent to assuming that the model contains no more errors as soon as the parameters have been estimated and is of course far from true. So, the parameter estimation problem should be considered as a combined model state and parameter estimation problem. In this case, an improved estimate of the model state and of the model parameters are sought simultaneously [Evensen, 2009] [33].

2.6.3 State vector augmentation

At first sight the Kalman filter equations do not appear to be related to the problem of parameter estimation as it is merely the model state that is estimated [Annan et al., 2005] [5]. However, the (ensemble) Kalman filter may be used for combined model state and parameter estimation by means of state space augmentation [Derber, 1989] [23]. The technique consists of considering the model parameters as part of the model state so that they are estimated simultaneously with the rest of the model state. The analysed augmented state vector therefore contains not only the updated model variables but also the newly estimated parameter values [Annan et al., 2005] [5]. This method can be applied for the estimation of parameters slowly varying in time with noise and therefore acting like states [Vermeulen and Heemink, 2006] [102]. However, in this case, the estimate will not converge but will always fluctuate. For parameters not affected in time by noise, the estimation might converge to a certain value; but in this case, the Kalman filter is biased [Ljung, 1979] [60] and tends to diverge from optimality [Jazwinski, 1970] [48].

2.7 Specific features of the EnKF implementation used in this study

In this section, we review the specific features of the ensemble Kalman filter implemented in the COHERENS model. They consist of a parallel implementation as well as of a local and multivariate analysis.

2.7.1 Localization of the analysis

Potential problems with ensemble methods arise because of the limited size of the ensemble [van Leeuwen, 1999] [97]. Though augmenting the size of the ensemble reduces the problem, the ensemble Kalman filter could generate unwanted correlations between physically unrelated fields [Hamill et al., 2001] [38]. Moreover, observations located at a long distance of the point being analysed are expected to have a small positive impact on this point. The influence of spatially remote observations at one point of the domain can be limited by localizing its impact to a subset of grid points closely related to the observations [Hamill et al., 2001] [38]. Localization can be essential to the performance of small size ensembles [Anderson, 2007] [3]. In many cases it significantly reduces the impact of an ensemble of limited size and allows to apply the ensemble Kalman filter for high dimensional models. There are two ways for imposing localization, either the local updating or the covariance localization by using a Schur (elementwise) product of the model error covariances and a correlation function with local support [Hamill et al., 2001] [38].

In the implementation of the ensemble Kalman filter on which the 3-D simulations are based, the spatial extent of the model error covariance is controlled through horizontal and vertical decorrelation lengths. Then, the analysis update is computed grid point by grid point. Hence the Kalman weights for the analysis update are different at each grid point and the analysed state results from the combination of different ensemble member values at each grid point. In this case, it is necessary to limit the analysis update to the grid points closely related to the observations. Different representations of the localization radius exist, from a cut-off to smoothly varying functions. As the horizontal extent of the ensemble members is controlled by the horizontal decorrelation length between the model error terms, a horizontal smoothing is already incorporated in the computation of the Kalman weights and the use of a cut-off radius is adequate. Moreover, because of the use of a cut-off radius, the Kalman filter update at each grid point relies on data located within a fixed distance. A further advantage of the use of a cut-off radius is that it greatly reduces the size of the matrices that must be inverted in order to compute the Kalman gain [Houtekamer and Mitchell, 1998] [45].

2.7.2 Multivariate analysis

To ensure that the analysed model state remains physically and dynamically consistent after the assimilation of observations of one model variable, the other model variables need to be also updated, this can be achieved by using a multivariate analysis scheme [Brasseur et al., 1999] [11].

A multivariate analysis consists of correcting all the variables of the estimation state vector from multivariate covariances between these variables. It is difficult to know all the covariances between all model state variables. Wrong multivariate error covariances could prevent the convergence of the model state to the "true" state of the system. On the other hand, if a consistent representation of these error covariances is available, a multivariate analysis of the variables of the estimation state vector will provide a better model state [Magri, 2002] [67].

In particular, currents and temperature are dynamically correlated and the impact of the assimilation of temperature profiles is propagated to the temperature, salinity and (horizontal) currents.

Moreover, a multivariate update is justified as the assumption of Gaussian error statistics for the model state variables seems to be reasonable [Brasseur et al.] [10].

2.7.3 Implementation in a parallel dynamical model

In order to perform simulations for the North Sea domain with a good horizontal spatial resolution, the ensemble Kalman filter is implemented in the parallel version of the COHERENS model with the same strategy as the one applied in Keppenne and Rienecker, [2002] [51]. The domain decomposition is based on the physical domain so that each processor deals with a piece of the state vector of all the ensemble members and processes the observations that are local to the region for which it is responsible. The main advantage of this parallelization strategy is that it avoids the costly ensemble transfers across processing elements that would be required at the analysis time if the integration of the ensemble was parallelized by letting each ensemble member run on a different processor.

An ensemble is associated to each variable and is represented as a matrix containing the mean and the perturbations around the mean. At each sampling time step the model error is estimated by adding to the variable related to the assimilated data (in this study temperature profiles) a random noise whose magnitude is modulated by the vertical and horizontal standard deviations. The time integration is applied to each member of the ensemble.

At the analysis step, the mean value of the ensemble is computed. The model variable corresponding to the observations is interpolated on the observations' grid and the innovations are calculated. Depending on the size of the assimilation radius, the data neighboring points that have to be analysed are identified. At these points the Kalman gain weights are computed following the chosen square root analysis algorithm; the ensemble mean and perturbations for temperature, salinity and horizontal currents are updated. Then, the model time integration is restarted.

Chapter 3

Simulations of the North Sea temperature with the hydrodynamical component of the COHERENS model

This chapter deals firstly with the description of the model (physical and numerical aspects). Then, the features of the North Sea temperature are reviewed and the simulations are described in the last section. The focus of this study is on how a given data assimilation scheme allows to improve model forecasts rather than on the model itself. So, a short description of the hydrodynamical component of the COHERENS (Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas) model is given. It is based on the COHERENS User Documentation, [Luyten, 2011] [65]. The hydrodynamical component of the COHERENS model consists of the equations of momentum, continuity, temperature and salinity, written in cartesian or spherical coordinates. Conservative finite differences are used to discretise the mathematical model in space on an Arakawa C-grid. Equations are solved with a mode splitting technique.

3.1 Physical model

3.1.1 Equations and hypotheses

3-D mode equations

The 1-D case study uses a Cartesian system of coordinates (x, y, z) whereas the 3-D simulations require a spherical coordinates system (λ, ϕ, z) allowing to take into account a latitudinal dependence of the Coriolis frequency; λ and ϕ are respectively the longitude and latitude. The z-axis is directed upwards along the vertical and is chosen so that the surface $z = 0$ corresponds to the mean sea level. Hence the equations of the free surface and sea bottom take the following form:

$$z = \zeta(x, y, t) \quad or \quad z = \zeta(\lambda, \phi, t) \text{ at the free surface} \tag{3.1}$$

$$z = -h(x, y) \quad or \quad z = -h(\lambda, \phi) \text{ at the bottom} \tag{3.2}$$

where ζ is the sea surface elevation and h is the mean water depth so that the total water depth is given by $H = h + \zeta$.

Coordinate units are meters in the Cartesian system and degrees in the spherical case.

The hydrodynamical component of the COHERENS model consists of the equations of momentum, continuity, temperature and salinity. They are derived under the following assumptions:

- the Boussinesq approximation which states that the density is constant except for the Earth's gravity force,
- the assumption of vertical hydrostatic equilibrium that reduces the vertical component of the momentum equations to the balance of the vertical pressure gradient and gravity force.

The horizontal component of the Earth's rotation vector is set to zero, this assumption becomes invalid for non hydrostatic water masses or near the equator. The horizontal sub-grid scale processes not resolved by the model can be parameterized by means of horizontal diffusion terms in the momentum, heat and salt equations. Momentum and scalar diffusion coefficients can be taken as constant or computed according to the Smagorinsky, [1963] [89] parameterizations. In the framework of this study, because of the relatively high horizontal resolution ($\sim$ 8km) for the 3-D North Sea simulations, it is assumed that they can be neglected. However, the process of horizontal diffusion arising from the combination of horizontal advection with vertical diffusion is resolved. In a Cartesian system of coordinates, the three-dimensional equations are expressed as follows:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \frac{\partial}{\partial z}(\nu_T\frac{\partial u}{\partial z}) \tag{3.3}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + fu = -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + \frac{\partial}{\partial z}(\nu_T\frac{\partial v}{\partial z}) \tag{3.4}$$

$$\frac{\partial p}{\partial z} = -\rho g = -\rho_0(g - b) \tag{3.5}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{3.6}$$

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z} = \frac{1}{\rho_0 C_p}\frac{\partial I}{\partial z} + \frac{\partial}{\partial z}(\lambda_T \frac{\partial T}{\partial z}) \tag{3.7}$$

$$\frac{\partial S}{\partial t} + u\frac{\partial S}{\partial x} + v\frac{\partial S}{\partial y} + w\frac{\partial S}{\partial z} = \frac{\partial}{\partial z}(\lambda_T \frac{\partial S}{\partial z}) \tag{3.8}$$

where (u, v, w) are the components of the velocity, T denotes the potential temperature, S the salinity, $f = 2\Omega \sin\phi$ the Coriolis frequency, $\Omega = 2\pi/86164\ rad/s$ the rotation frequency of the Earth, ϕ the latitude, g the acceleration of gravity, p the pressure, ν_T and λ_T the vertical eddy viscosity and diffusivity coefficients, ρ the density, ρ_0 a reference density, C_p the specific heat of seawater at constant pressure, I the solar irradiance and b the buoyancy defined as:

$$b = -g(\frac{\rho - \rho_0}{\rho_0}).$$

T is not the in situ temperature but the potential temperature defined as the temperature of a fluid parcel which is moved adiabatically to a certain level usually taken at the surface. The reason is that equation (3.7) is derived from the equation of conservation of heat which contains an extra term stemming from compressibility. This term vanishes if T is interpreted as the potential temperature.

The density is computed from the equation of state proposed by McDougall et al., [2003] [69]. It is expressed in terms of the potential temperature. This formulation, given in Appendix B in Table 8.2, is chosen because it is more accurate and computationally more efficient than the International Equation of State [McDougall et al., 2003] [69].

In the horizontal components of the momentum equations, the horizontal pressure gradient terms are expressed as a function of the surface elevation ζ, of the atmospheric pressure and of the vertically integrated buoyancy.

2-D mode equations

The 3-D continuity and momentum equations presented in the above section need to be supplemented by additional 2-D equations for the depth-integrated velocity and surface elevation. The numerical solution of 3-D equations is constrained by the CFL limit which poses a severe limit on the time step used in the numerical discretisations. This can be overcome by solving the simpler 2-D equations with a smaller 2-D time step and inserting the results

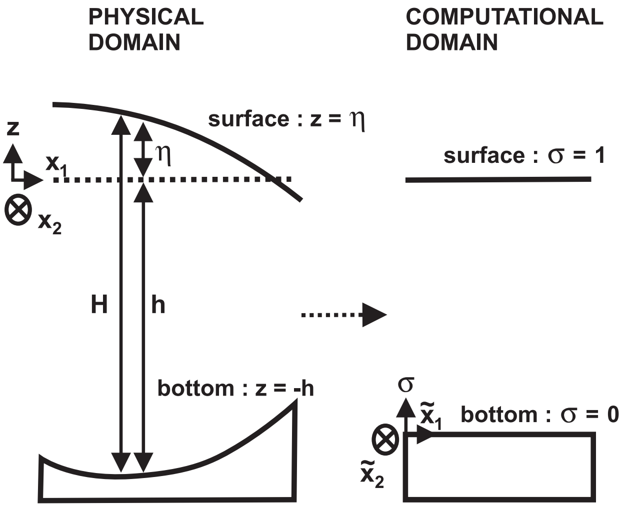

Figure 3.1: The σ-coordinate transform in the vertical - From: Deleersnijder and Ruddick, [1992] [20], page 494.

into the 3-D equations which can afterwards be solved with a larger 3-D time step.

3.1.2 Vertical coordinates system

In the vertical, the numerical solutions of the model equations are greatly simplified when a σ-coordinate system is defined, with σ varying between 0 at the bottom and 1 at the surface [Freeman et al., 1972] [34], [Owen, 1980][79], [Deleersnijder, 1989] [19]. The formulation of this coordinates' change is inspired by the one initially suggested for atmospheric modelling by [Phillips, 1957] [82]:

$$\sigma = \frac{z+h}{\zeta+h} = \frac{z+h}{H}, \tag{3.9}$$

where ζ is the sea surface elevation, h is the mean water depth and $H = h+\zeta$ is the total water depth. It is illustrated on Figure 3.1.

The σ-coordinate system is particularly suitable for coastal regions, where bathymetric effects are important [Kourafalou et al., 1996] [54]. The advantages of the σ-coordinates consist of a better resolution in shallower areas and a simpler expression of the boundary conditions at the surface and at the bottom. However they have well-known disadvantages: areas with steep bathymetry gradients are difficult to represent, and large numerical errors can be produced by discretisation of the baroclinic pressure gradient. Further insight on the issues related to the use of σ-coordinates can be found in Deleersnijder and Ruddick, [1992] [20].

3.1.3 Turbulence closure

A wide range of turbulence parameterisations exists in the literature and is implemented in COHERENS, from simple algebraic formulations up to second order closure schemes with additional transport equations for turbulence quantities [Luyten et al., 1996] [62].

Turbulent kinetic energy

The vertical exchange processes are represented through the eddy viscosity and diffusivity coefficients ν_T and λ_T. These coefficients are expressed as a function of the turbulent kinetic energy k and its dissipation rate ϵ. The model implementation used in this study is a one-equation model in which the turbulent kinetic energy k is obtained by solving the following equation:

$$\frac{\partial k}{\partial t} = \frac{\partial}{\partial z}(\nu_T \frac{\partial k}{\partial z}) + \nu_T M^2 - \lambda_T N^2 - \epsilon$$

where N is the Brunt-Väisälä (buoyancy) frequency defined by $N^2 = \frac{\partial b}{\partial z}$ and M is the Prandtl (shear) frequency defined by $M^2 = \left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2$. The eddy coefficients are expressed as follows:

$$\nu_T = S_u \frac{k^2}{\epsilon}, \tag{3.10}$$

$$\lambda_T = S_b \frac{k^2}{\epsilon}, \tag{3.11}$$

where S_u and S_b are referred to as stability functions.

The turbulent dissipation ϵ is determined according to:

$$\epsilon = c_\epsilon \frac{k^{3/2}}{l}$$

where c_ϵ is a constant and l is a prescribed turbulent mixing length given in the next paragraph. The parameter c_ϵ is determined for an equilibrium neutral flow near a wall where l is equal to κ (von Karman constant) times the distance to the wall, this gives $c_\epsilon = 0.166$.

The stability functions, S_u and S_b, are expressed as functions of α_N:

$$\alpha_N = \frac{k^2}{\epsilon^2} N^2,$$

they are derived in Luyten et al., [1996] [62]:

$$S_u = \frac{0.108 + 0.0229\alpha_N}{1 + 0.471\alpha_N + 0.0275\alpha_N^2},$$

$$S_b = \frac{0.177}{1 + 0.403\alpha_N}.$$

They have no explicit dependency on the current shear so that the stability of the scheme is improved [Deleersnijder and Luyten, 1994] [21].

Mixing length formulation

In the case of a one-equation model, a mixing length formulation is needed. The basic requirement is that l reduces to the traditional bottom and surface layer length scales i.e. l_1 near the bottom and to l_2 near the surface, with l_1 and l_2 defined as:

$$l_1 = \kappa(\sigma H + z_{0b}),$$

$$l_2 = \kappa(H - \sigma H + z_{0s}),$$

where κ is the von Karman constant and is equal to 0.4, z_{0b} and z_{0s} are the bottom and surface roughness lengths. These expressions indicate that, close to the surface or bottom boundary, the turbulent mixing length is a linear function of the distance to that boundary.

The asymptotic mixing length formulation proposed by Blackadar, [1962] [7] is used, it has the following form:

$$\frac{1}{l} = \frac{1}{l_1} + \frac{1}{l_2} + \frac{1}{l_a},$$

so that $l \to l_a$ sufficiently far from the boundaries. A commonly used value for l_a is the one proposed by Mellor and Yamada, [1974] [70] who defined l_a as the ratio of the first order to the zero order moment of the vertical profile of the turbulent velocity scale $k^{1/2}$:

$$l_a = \frac{\alpha_1 \int_{-h}^{\zeta} (\zeta - z) k^{1/2} dz}{\int_{-h}^{\zeta} k^{1/2} dz} \tag{3.12}$$

where ζ is the surface elevation, h is the mean water depth, and the constant α_1 is set to 0.2. The Blackadar formulation depends explicitly on turbulent kinetic energy via equation (3.12).

Background mixing scheme using limiting conditions

The assumptions of the theory presented above are valid in the case of a nearly-isotropic, fully developed turbulence under homogeneous or weakly stratified conditions but becomes invalid for strongly stratified flows for two main reasons. Firstly, theoretical and laboratory experiments have shown that vertical turbulence excursions are impeded by stable stratification. In this case the turbulence becomes anisotropic and the assumption of nearly-isotropic turbulence can no longer be maintained. Then, even when the turbulence decays because of an increasing stable stratification, additional turbulence is generated by the shear and breaking of internal waves which are not resolved by the model. In the absence of a comprehensive theory of both effects, simple parameterisations have been designed and implemented in COHERENS. They consist of adding background mixing coefficients to equations (3.10) and (3.11). The method used in the framework of this study is based on limiting conditions for the dissipation rate of the turbulent kinetic energy of the form $\epsilon \sim N$:

$$\epsilon > \epsilon_{lim} = \alpha_{max}^{-1/2} k N,$$

and a second limiting condition needs to be imposed for the turbulent kinetic energy:

$$k > k_{lim},$$

where N is the buoyancy frequency, $k_{lim} = 10^{-6}\ m^2.s^{-2}$ and $\alpha_{max} = 4.72$.

These limitations of vertical overturns by stable stratification yield the following limiting or background values for the eddy coefficients of the form $\nu_T \sim N^{-1}$, $\lambda_T \sim N^{-1}$:

$$\nu_T > \nu_{T-lim} = \nu_{0l}\frac{k}{N},$$
$$\lambda_T > \lambda_{T-lim} = \lambda_{0l}\frac{k}{N}.$$

Following the formulation by Hossain and Rodi, [1982] [43], $\nu_{0l} = 0.123$ and $\lambda_{0l} = 0.132$.

Besides these physical limiting conditions, the following numerical lower bounds are imposed:

$$k_{min} = 10^{-14} m^2.s^{-2}, \epsilon_{min} = 10^{-12} m^2.s^{-3}, l_{min} = 1.7 \times 10^{-17} m.$$

The reason for taking into account these lower limits is to prevent that unrealistically large eddy coefficients are created by rounding errors in the intermittent zone. With these values and taking $S_u \sim S_b \sim 0.1$ the values of ν_T and λ_T are of the order of 10^{-17} which is much smaller than the corresponding laminar viscosity and diffusivity coefficients.

3.1.4 Solar heat absorption

The downward radiation is expressed as the sum of an infrared and a short-wave component so that:

$$I(z) = Q_{rad}[Be^{-z/\lambda_1} + (1 - B)e^{-z/\lambda_2}], \tag{3.13}$$

where B represents the infrared fraction absorbed in the upper (1-2) meters of the sea surface and $(1-B)$ is the fraction of blue-green light absorbed over larger depths ($B = 0.54$). Q_{rad} is the radiation incident on the sea surface (the reflectance by albedo is taken into account). λ_1 is the attenuation depth of the infrared radiation which is taken as a constant equal to 1 m so that the infrared radiation is almost completely absorbed in the surface layer. λ_2 is the penetration depth of solar light or, equivalently in terms of the model implementation, the inverse of the optical attenuation coefficient, $\lambda_2 = 1/K_{opt}$, where K_{opt} is given in m^{-1}. Its value depends on the optical properties of the water masses.

3.1.5 Surface boundary conditions

For temperature

The surface heat flux is given by:

$$\rho_0 C_p \lambda_T \frac{\partial T}{\partial z} = Q_s, \tag{3.14}$$

where Q_s is the downwards heat flux at the surface, C_p is the specific heat of seawater at constant pressure and ρ_0 a reference density. The net total heat flux coming into the water column is composed of a term $-Q_{nsol}$ of all non solar contributions plus the radiative flux Q_{rad}. This last term does not contribute to the surface heat flux because the solar radiation is absorbed within the water column so that $Q_s = -Q_{nsol}$ and only $-Q_{nsol}$ contributes to the surface heat flux.

The upwards directed non solar heat flux at the surface is decomposed into three terms:

$$Q_{nsol} = Q_{la} + Q_{se} + Q_{lw}, \tag{3.15}$$

The first one, Q_{la}, is the latent heat flux released by the evaporation. The second one, Q_{se}, is the sensible heat flux generated by the (turbulent) transport of heat across the air-sea interface. The last one, Q_{lw}, is the net long-wave thermal radiation emitted at the sea surface.

The first two terms are related to the turbulent fluxes of humidity and temperature. They are calculated as follows:

$$Q_{la} = \rho_a L_v C_e W_{10}(q_s - q_a) \tag{3.16}$$

$$Q_{se} = \rho_a C_{pa} C_h W_{10}(T_s - T_a) \tag{3.17}$$

where W_{10} is the magnitude of the wind speed at 10 metres height defined as $W_{10} = (U_{10}^2 + V_{10}^2)^{1/2}$, with U_{10} and V_{10} the horizontal components of the wind velocity. T_s, q_s and T_a, q_a are the sea surface and air temperature and humidity, ρ_a is the air density and C_{pa} is the specific heat of air at constant pressure:

$$C_{pa} = 1004.6\,(1 + 0.8375\,q_a)\; J.kg^{-1}.(°C)^{-1}, \tag{3.18}$$

and L_v is the latent heat of vaporization given as a function of the sea surface temperature:

$$L_v = 2.5008 \times 10^6 - 2300\,T_s\; J.kg^{-1} \tag{3.19}$$

The sea surface and air humidities q_s and q_a are calculated using:

$$q = \frac{0.62\,e}{P_{a0} - 0.38\,e} \tag{3.20}$$

where $P_{a0} = 1013.25\; mb$ is a reference atmospheric pressure and e is the vapor pressure obtained in mb from the empirical relation given by Gill, [1982] [37]:

$$\log_{10} e = \log_{10} RH + \frac{0.7859 + 0.03477\,T}{1 + 0.00412\,T}$$

where RH is the relative humidity comprised between 0 and 1 and T represents either sea surface temperature or air temperature at a reference height.

The surface fluxes of heat involve two dimensionless parameters, the latent and sensible heat exchange coefficients, C_e, C_h for which neutral and stratified formulations are used [Kondo, 1975] [53]. In the absence of temperature stratification, the neutral values C_{en} and C_{hn} are given by:

$$C_{en} = 10^{-3}[a_e + b_e W_{10}^{p_e} + c_e(W_{10} - 8)^2], \tag{3.21}$$

$$C_{hn} = 10^{-3}[a_h + b_h W_{10}^{p_h} + c_h(W_{10} - 8)^2], \tag{3.22}$$

where W_{10} is the magnitude of the wind speed at 10 metres height, and a_e, a_h, b_e, b_h, c_e, c_h, p_e and p_h are functions of wind speed, their values are given in Table 8.4 in Appendix C.

The stratified formulation is obtained by multiplying neutral values by a factor $f(D)$ depending on the ratio of the sea-air temperature difference and wind speed:

$$D_0 = (T_s - T_a)/W_{10}^2, \tag{3.23}$$

$$D = D_0 \frac{|D_0|}{|D_0| + 0.01}. \tag{3.24}$$

In case of stable conditions ($T_s < T_a$), one has:

$$f(D) = 0.1 + 0.03D + 0.9e^{4.8D}, \quad if - 3.3 < D < 0, \tag{3.25}$$
$$f(D) = 0, \qquad if D \leq -3.3,$$

while for unstable conditions ($T_s > T_a$):

$$f(D) = (1.0 + 0.63D^{1/2}). \tag{3.26}$$

The stratified formulation is given by:

$$C_{es} = C_{en} f(D), \tag{3.27}$$

$$C_{hs} = C_{hn} f(D). \tag{3.28}$$

where C_{es} and C_{hs} represent the stratified coefficients.

Figure 3.2 compares the neutral (black line) to the stratified formulations (colors) of the sensible heat exchange coefficient. In the neutral formulation, the minimum value corresponds to a wind speed of 2.2 $m.s^{-1}$ and the maximum value is attained for a wind speed of 18 $m.s^{-1}$. For $W_{10} < 2.2$ $m.s^{-1}$ the coefficient decreases with increasing wind speed while for $W_{10} > 2.2$ $m.s^{-1}$ it increases.

When considering the stratified formulation, in the case of light winds, the effect of the temperature stability on the value of the coefficient is important even when the sea-air temperature difference is small.

The study presented here is based on these two formulations of the surface heat exchange coefficients. The stratified formulation undergoes the influence of the assimilated data through its dependence on the sea surface temperature. It is also considered as more realistic than the neutral formulation as it incorporates the effects of the wind and sea air temperature difference. In applications for model parameters estimation, the heat exchange coefficients in their neutral formulation will be adjusted by the assimilation process and the results will be compared to the stratified formulation.

Finally, the long-wave radiation flux is parameterized following Gill, [1982] [37] as a function of the emissivity at the sea surface, Stefan's constant, the fractional cloud cover and the vapour pressure.

For salinity

The salinity change caused by the difference between the evaporation and precipitation rates is given by a surface boundary condition of the following

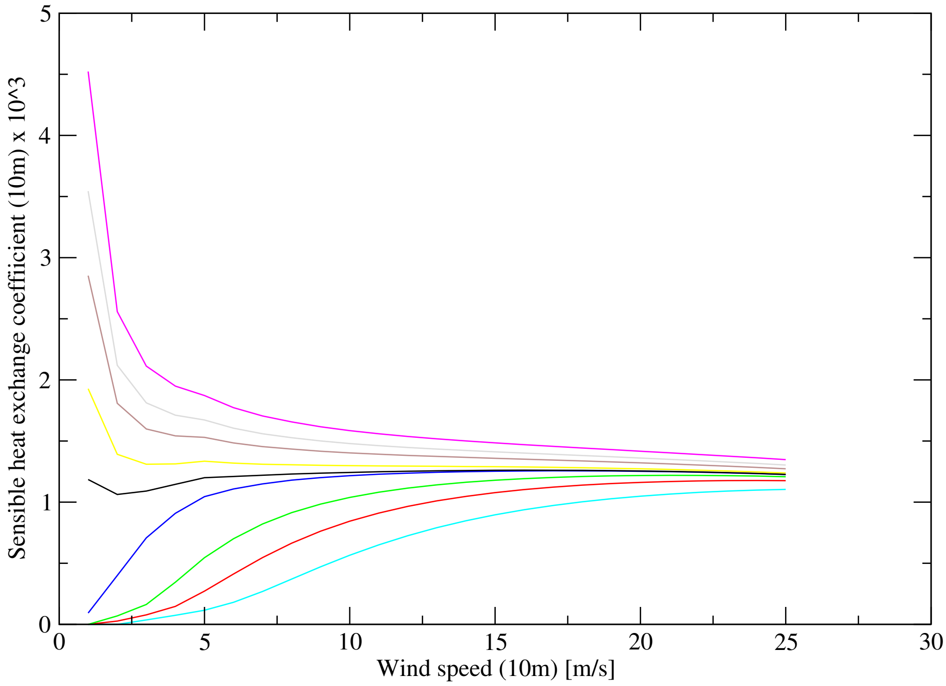

Figure 3.2: Sensible heat exchange coefficient as a function of the wind speed for the neutral formulation (black) and the stratified formulation for a sea-air temperature difference of 20°C in magenta, 10°C in grey, 5°C in brown, 1°C in yellow, −1°C in blue, −5°C in green, −10°C in red, −20°C in cyan

.

form:

$$\rho_0\lambda_T\frac{\partial S}{\partial z} = \frac{S_s(R_{evap} - R_{prec})}{1 - 0.001S_s} \tag{3.29}$$

where $R_{evap} = \frac{Q_{la}}{L_v}$ and R_{prec} are the evaporation and precipitation rates in $kg.m^{-2}.s^{-1}$ and S_s the surface salinity. The evaluation of the surface salinity flux requires the input of precipitation as additional meteorological data.

For the horizontal components of the velocity

The surface stress is specified as a function of the surface wind speed:

$$\rho_0\nu_T(\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}) = (\tau_{s1}, \tau_{s2}) = \rho_a C_{ds} W_{10}(U_{10}, V_{10}), \tag{3.30}$$

where (U_{10}, V_{10}) are the components of the wind at a reference height of 10 metres, $W_{10} = (U_{10}^2 + V_{10}^2)^{1/2}$ is the wind speed and $\rho_a = 1.2\ kg.m^{-3}$ is the air density.

The Kondo formulation of the sea surface drag coefficient C_{ds} is analogous to that of the heat exchange coefficients. In the absence of stratification, its neutral value C_{dn} is given by:

$$C_{dn} = 10^{-3}(a_d + b_d W_{10}^{p_d}) \tag{3.31}$$

where a_d, b_d and p_d are functions of wind speed given in Appendix C at Table 8.4, whereas its stratified expression is:

$$C_{ds} = C_{dn} f(D). \tag{3.32}$$

The transformed vertical velocity vanishes at the surface (σ=1).

3.1.6 Bottom boundary conditions

At the bottom, a slip boundary condition following a quadratic friction law is applied to the horizontal components of the velocity:

$$\nu_T(\frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}) = (\tau_{b1}, \tau_{b2}) = C_{db}(u_b^2 + v_b^2)^{1/2}(u_b, v_b) \tag{3.33}$$

where the bottom velocity components (u_b, v_b) are evaluated at the grid point nearest to the bottom.

In the boundary layer approximation a vertically uniform shear stress is assumed, yielding a logarithmic profile for the current. The quadratic bottom drag coefficient can then be expressed as a function of the bottom roughness length z_{0b} and the vertical grid spacing:

$$C_{db} = \left(\frac{\kappa}{ln(z_r/z_{0b})}\right)^2 \tag{3.34}$$

where z_r is a reference height taken at the grid centre of the bottom cell. The bottom roughness length z_{0b} has a constant value equal to 0.0035 m.

The transformed vertical velocity vanishes also at the bottom (σ=0). A zero flux of temperature and salinity normal to the seabed is imposed.

3.1.7 Lateral boundary conditions

The lateral boundary conditions are described in section 3 of this chapter.

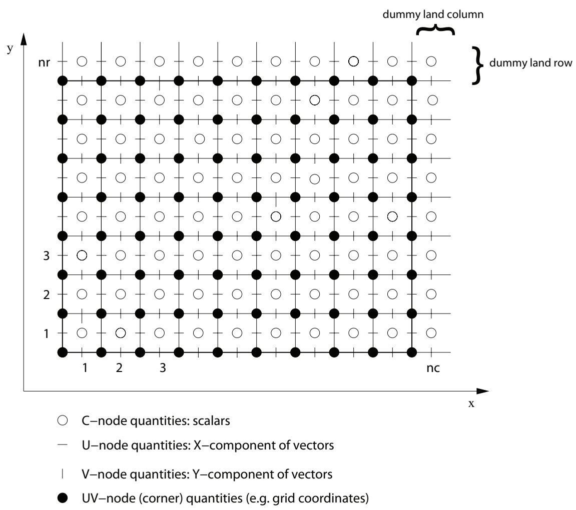

Figure 3.3: Computational grid in the horizontal - From: Luyten, [2011] [65].

3.2 Numerical methods

3.2.1 Spatial discretisation

The grid chosen is the Arakawa C-grid [Mesinger and Arakawa, 1976] [72] represented on Figure 3.3. It staggers the currents and pressure/elevation nodes to give a good representation of the gravity waves and provides simple representations of open and coastal boundaries.

3.2.2 Numerical schemes for time discretisation

The mode splitting technique proposed by Blumberg and Mellor, [1987] [8] is used for the time integration of the governing equations. The method consists of solving the simpler depth-integrated momentum and continuity equations for a two-dimensional external or barotropic mode (fast free surface gravity waves) with a time step ($\Delta\tau$) small enough to satisfy the Courant-Friedrichs-Lewy (CFL) condition. The results are then inserted into the three-dimensional momentum and scalar transport equations (slower baroclinic waves) which can be solved with a 10 to 20 times longer time step (Δt) [Kourafalou et al., 1996] [54]. Moreover, for the 3-D momentum equations a predictor-corrector method is used, satisfying the basic requirement that the currents obtained from the 3-D equations should have the same depth integral as the ones obtained from the 2-D depth-integrated equations.

All horizontal derivatives are evaluated explicitly in time. Along the vertical, diffusion is computed fully implicitly and advection is computed semi-implicitly with a central scheme that is monotonic provided that the implicity factor $\theta_a \geq 0.5$. In this way, only simple tridiagonal systems need to be solved.

A quasi-implicit method is implemented for the Coriolis term.

3.2.3 Advection schemes for the momentum and scalar equations

A TVD scheme is employed for the spatial discretisation of horizontal and vertical advection terms in conjunction with the operator splitting method. The TVD scheme has the advantage of combining the monotonicity of the upwind scheme with the second order accuracy of the Lax-Wendroff scheme. A superbee limiter function is used to reduce the numerical diffusion due to the upwind scheme in areas of low gradients and to provide sufficiently large diffusion in areas of strong gradients so that over- and under-shooting due to the non montonicity of the Lax-Wendroff schemes are suppressed.

3.3 Modelling the North Sea temperature

The specification of model error plays an important role in the performance of a data assimilation scheme. The features of coastal sea processes render the specification of error statistics needed for data assimilation much more complicated than in the open ocean [Mourre et al., 2004] [73].

In this section, we describe the issues related to the modelling of the North Sea temperature in two steps. First, general considerations are given on the North Sea temperature reported from [Otto et al., 1990] [78] and from surveys of the North Sea Project ([Charnock et al., 1994] [17]). Then, we refer to modelling issues discussed in previous studies by Holt and James, [1999] [42], Warrach, [1998] [103], and Luyten et al., [2003] [63].

A description of the particular model error with respect to the data that will be assimilated in the two test cases that are dealt with in the present study is given at the beginning of the chapters where the results of these applications are presented.

3.3.1 Variability of the North Sea temperature

The North Sea is a semi-enclosed sea receiving several large river inflows with three open sea boundaries: a narrow connection to the English Channel through the Dover Strait, a connection to the Baltic sea through the Skagerrak, and a wide northern boundary (Figure 3.4). Its bathymetry varies widely, many areas are less than 40 metres deep (the Southern and German Bights as well as the Dogger Bank) while there are deeper regions, east and west of the Dogger Bank where depths exceed 90 metres as well as along the Norwegian trench and the Skagerrak where the water depth is up to 700 metres [Charnock et al., 1994] [17].

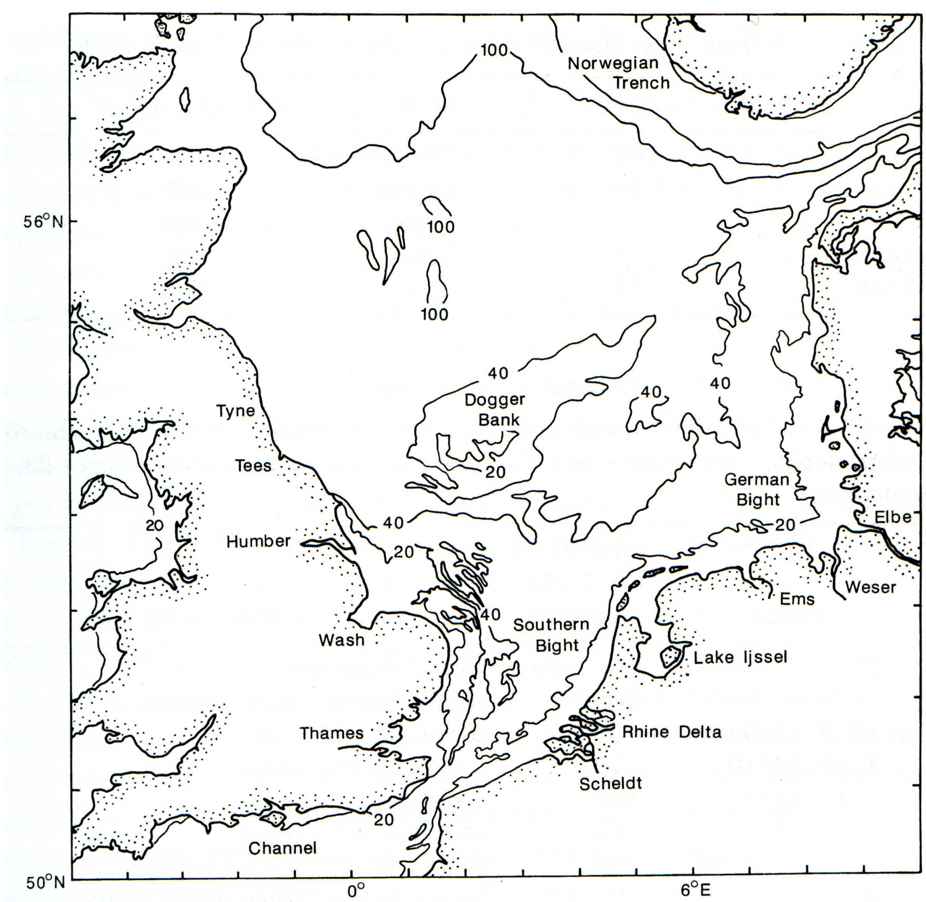

Figure 3.4: Map of the North Sea and surrounding coasts, with the 20, 40 and 100 isobaths - From: Charnock et al., [1994] [17], page 7.

The main forcings mechanisms of the North Sea dynamics (Figure 3.5) are the tides and the wind [Otto et al., 1990] [78]. Semi-diurnal tides predominate at the latitudes under consideration. Tidal energy enters the North Sea across the northern boundary from the Atlantic and through the Channel. The resulting motions dominate the kinetic energy budget and induce strong frictional stresses stirring the water column. The other main mechanical forcing is the wind, its time scales range broadly from one hour to one day. Along with atmospheric pressure changes it is responsible for driving storm surges motions [Charnock et al., 1994] [17].

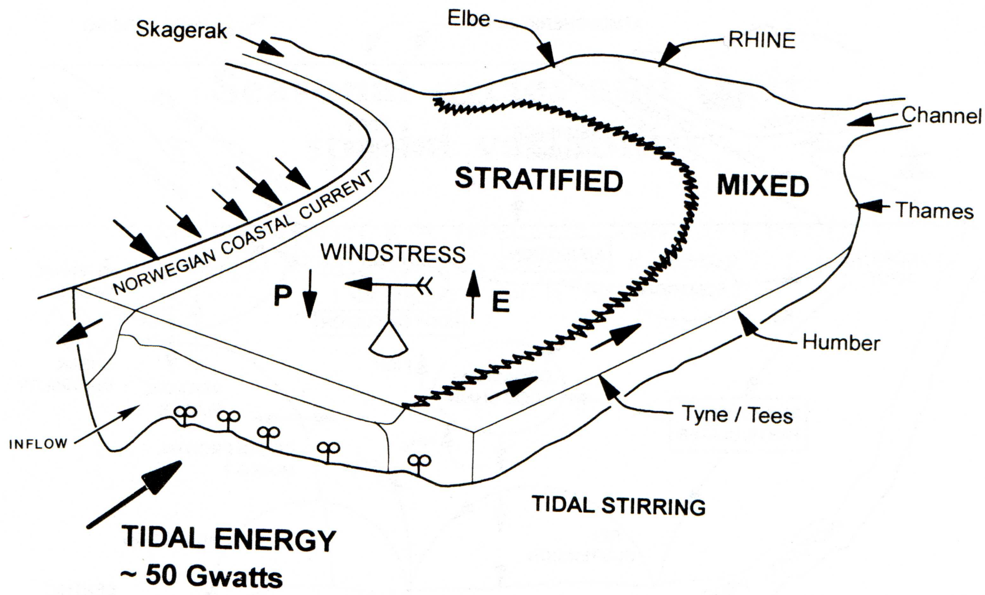

Figure 3.5: Schematic of the basic physical processes in the North Sea - From: Charnock et al., [1994] [17], page 3.

The dominant factor governing the temperature field in the North Sea is the surface seasonal heating and cooling which in the central and northern parts of the North Sea leads to a thermal stratification of the water column in summer. The combination of tides, winds and solar forcings induces three different ranges of timescales: seasonal (long range), semi-monthly as given by the spring-neap cycle (medium range) and semi-diurnal and inertial (short range) [Luyten et al., 2003][63].

Seasonal variations dominate many processes in the North Sea. 97% of the surface temperature's variance in the seasonal cycle is driven by solar forcing [Charnock et al., 1994] [17] with surface heating and cooling producing density differences leading to thermal stratification. Other important density differences are due to freshwater inputs, from rivers [Charnock et al., 1994] [17]), and from water discharge in the Norwegian fjords.

The surface temperatures of the North Sea show a yearly cycle with amplitudes ranging from 8°C in the Wadden Sea to less than 2°C at the northern entrances. The high amplitude in the south-east region is related to the reduced depth [Ospar commission, 2000] [77].

3.3.2 Issues related to the modelling of the North Sea temperature

The evolution of the North Sea temperature field is governed by different processes: surface forcings, absorption of solar heat in the water column, vertical distribution of heat by turbulent mixing, horizontal and vertical advective transport [Luyten et al., 2003] [63].

In winter, most areas of the North Sea are vertically well-mixed. In summer, in areas of small water depths like the Southern Bight and the English Channel, the tidal currents maintain a turbulent regime which ensures a vertical mixing of the whole water column. By contrast, in the central North Sea, the stirring is weaker and, in summer, as the solar heating increases, the water column stratifies with a robust two-layer structure persistent during the summer season [Charnock et al., 1994] [17]. The strength of the thermocline depends on the heat input and on the turbulent mixing generated by tides and wind so that their representation is critical to the modelling of realistic temperature profiles. The better the wind stress is represented, the better the observed depths of the isotherms are represented. Moreover, the choice of parameterization of the vertical mixing of heat and momentum has a considerable influence on the thermal stratification calculated by the model [Warrach, 1998] [103].

Contrary to the open ocean where the depth of the thermocline is determined by the surface forcing, the deepening of the thermocline in tidal shelf seas is limited from below because of a tidally mixed bottom layer [Luyten et al., 2003] [63]. The depth of the thermocline increases from May to September. In August and September it is typically 30 metres deep in the central North Sea. In these stratified waters, the tides are able to generate internal waves that propagate along the interface between the two layers. In autumn, the increasing number and severity of storms combined with seasonal cooling at the surface destroys the thermocline and mixes the surface and bottom layers [Ospar commission, 2000] [77]. A good representation of the depth of the thermocline is of critical importance to the modelling of temperature as it determines the depth to which surface forcing has an influence and is crucial in the prediction of sea surface temperature (and hence sensible and latent heat loss) [Holt and James, 1999] [42].

The well mixed and stratified regions are separated by thermal fronts resulting of a balance between tidal mixing and surface heating [Simpson and Hunter, 1974] [88].

Finally, observational studies by van Haren, [2000] [96] have revealed an important variability in the temperature field on time scales of up to a few days, mainly due to advection by wind (inertial frequencies) and tidally driven currents (semi-diurnal frequencies). During a wind event, the inertial mode may even become dominant [Luyten et al., 2003] [63]. These modes and their higher-order harmonics generate an important current shear across the thermocline that tends to enhance vertical mixing [van Haren, 2000] [96]. Therefore in shelf seas such as in the North Sea, most of the short-time variability in the temperature field can be attributed to the semi-diurnal and inertial modes and their first higher-order harmonics [Luyten et al., 2003] [63].

3.4 Description of the simulations

This section is dedicated to the description of simulations of the North Sea. Temperature profiles are assimilated with an ensemble Kalman filter. Simulations are performed in a 1-D vertical test case at the CS station and in 3-D simulations on the whole North Sea domain. Two applications of data assimilation are considered: model state estimation and combined model state and parameter estimation. A series of numerical experiments are conducted with twe two square root algorithms of the ensemble Kalman filter that are assessed.

We describe the forcings, numerical aspects and observations for the 1-D test case in the first paragraph and those for the 3-D case in the second paragraph of this section. Many features of the data assimilation experiments are common to both cases, hence they are described in the third paragraph and the specificities of the two cases are separately highlighted.

3.4.1 Simulations in the 1-D case: the North Sea CS station

Motivation for the 1-D simulations

The 1-D version of the program is applied at the CS station (Figure 3.6) located in the central part of the North Sea where a thermocline develops in summer. This 1-D version of the program has the advantage of limiting the computing time of the simulations allowing therefore the performance of many tests to determine a set of adequate parameters for model error sampling and data assimilation. Moreover, these experiments with a simplified model dynamics allowed to assess the validity of the implementation of the EnKF.

Another motivation for the use of the 1-D version of the model is to examine the ability of the EnKF to correct a "bad/very simplified" physical model and in keeping the correction. Furthermore, the study of Luyten et al., [2003] has shown that even with well parameterized physics, the 3-D model could not fully reproduce oscillations present in the observations at the level of the thermocline. The representation of these oscillations provides therefore a good test of the performance of the EnKF in coastal zones.

Parameterization of the 1-D physical model and forcings

In the 1-D version of the program only the vertical component of the equations is solved, so that all horizontal gradients are neglected except for the slope of the sea surface elevation. Vertical advection is also neglected therefore the advective effects are ignored. The momentum equations are then given by:

$$\frac{\partial u}{\partial t} - fv = -g\frac{\partial \zeta}{\partial x} + \frac{\partial}{\partial z}(\nu_T \frac{\partial u}{\partial z}), \tag{3.35}$$

$$\frac{\partial v}{\partial t} + fu = -g\frac{\partial \zeta}{\partial y} + \frac{\partial}{\partial z}(\nu_T \frac{\partial v}{\partial z}), \tag{3.36}$$

where $(u,\ v)$ are the horizontal components of the current, ζ is the surface elevation, f is the Coriolis frequency, ρ is the density and ν_T is the vertical coefficient of eddy viscosity. The surface slope term is provided by a depth averaged tidal model of the North-West European Shelf model.

The temperature field is obtained as a solution of the following equation:

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}(\lambda_T \frac{\partial T}{\partial z}) + \frac{\partial I}{\partial z}, \tag{3.37}$$

where λ_T is the vertical coefficient of eddy diffusivity.

The consequence of these hypotheses is that the physics is very simplified and the model is not fully able to account for the stratification of the temperature.

Salinity is kept constant and uniform at a value of 34.8. Given the location of the CS station in the central part of the North sea where the influence of freshwater inputs is very small, this assumption does not induce the main errors on the temperature profile for the 1-D simulations.

Simulations performed at the CS station

The 1-D model with its very simplified physics (lack of advection terms) is not adequate to properly represent the temperature of the North Sea area where the CS station is located. However, it is applied in a first stage to better understand the role of parameters for model error sampling and the impact of the assimilation on the model results.

The simulations are carried out for the year 1989 at a station located in the central part of the North sea at 55°30′N and 0°55′E and referred to as CS station after the 'North Sea Project' (Figure 3.6). The water depth is equal to 83 metres.

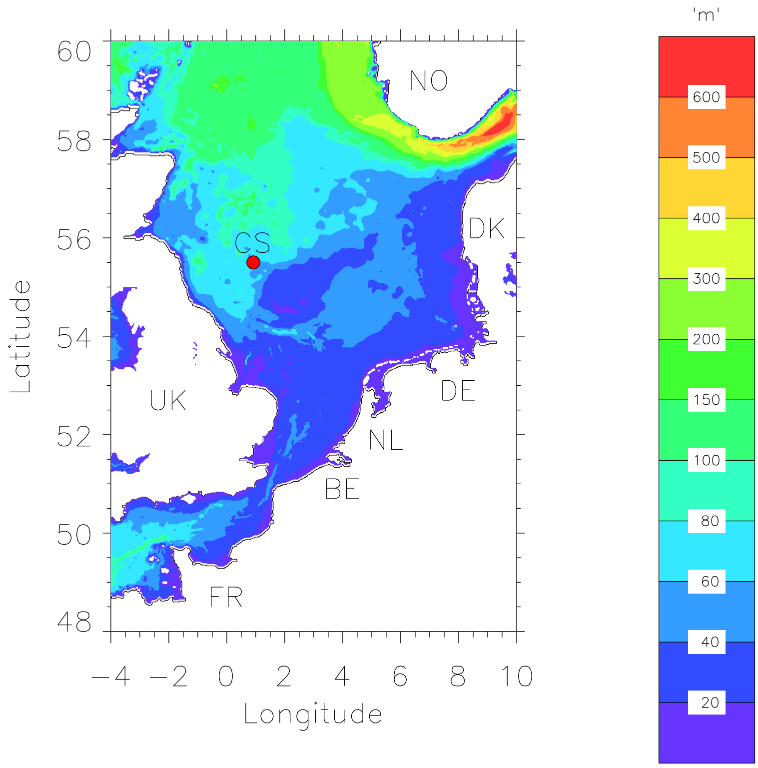

Figure 3.6: Location of the North Sea CS station (55°30′N, 0°55′E).

Wind velocity, air temperature and relative humidity are taken from the NWP model of the UK MetOffice with a 3-hour resolution. The model is run with a time step of 20 minutes and 50 σ-levels in the vertical. The temperature field is initialised using CTD data available from January 3, 1989. The model is integrated with a time step of 20 minutes and 50 σ levels in the vertical.

The model is run from January to April 1989 without ensemble generation, an initial ensemble of states is generated in May and is integrated without data assimilation till the end of July. The model error is sampled once a day. 50 ensemble members are used. Data is assimilated once a day at midnight, from August 5 to August 24, 1989.

Observations assimilated at the CS station

High resolution temperature profiles from the 'North Sea Project' data set (thermistor chain data) is assimilated for August 1989. The measurements of temperature profiles range from 10 metres to 60 metres depth with a resolution of 5 metres. Time sampling rate was approximately equal to 2 minutes. The rms error of the observations is about 0.1°C [Charnock et al., 1994]. The assimilated data have been extracted from the data set at the assimilation time step (they are not hourly or daily averaged). Contrary to 3-D simulations, no observations from an independent data set are available for validation.

3.4.2 Simulations in the 3-D case: the North Sea

Motivation for the 3-D simulations

Even with more realistic physics than in the 1-D case there are always observed features not correctly represented by the model. For example, the study by Luyten et al., [2005] [64] has shown that a horizontal resolution of 4 nautical miles was not sufficient to represent the typical North Sea internal oscillations present in the thermocline of summer stratified regions.

The 3-D runs reported in this study present the two regimes of the North Sea: well-mixed and stratified. The existence of these two regimes is one of the reasons why data assimilation in coastal zones is a challenge. One objective of this work is to determine a set of model error parameters adequate for the ensemble Kalman filter to deal with such an issue.

Model forcings in the 3-D case

The 3-D simulations were performed for september 2001 with a spatial resolution of 4 nautical miles on the North Sea domain extending from 4°W to 10°E and from 48.5°N to 60°N. The model is run with a time step of 20 seconds and 20 σ levels in the vertical. Meteorological data surface data have been supplied by the Danish Meteorological Institute from the HIRLAM model with a temporal resolution of one hour. Tidal harmonics and daily profiles of currents, temperature, salinity and inflow/outflow conditions at the boundaries of the domain are derived from simulations with the POLCOMS model (Proudman Oceanographic Laboratory) covering a larger area. River runoffs from the Elbe, Scheldt, Rhine/Meuse, Thames, Humber, Tyne/Tees are taken into account. Baroclinic inflow/outflow conditions are imposed at the eastern boundary to include the exchange of water masses with the Baltic Sea. The POLCOMS values are introduced at open boundaries via a flow relaxation scheme [Martinsen and Engedahl, 1987] [68]. The initial conditions for September are provided from conditions at the end of August resulting themselves from model runs starting on the 1st of January 2001.

Simulations performed in the North Sea domain

The 3-D simulations performed on the whole North Sea domain are based on the method of twin experiments. Twin experiments allow to assess the experimental design of the data assimilation experiments that will be used afterwards to assimilate data from "real" observations. To this aim, a least three series of simulations are necessary:

- a simulation providing the true state of the system (true state simulation) from which the data that will be assimilated are extracted,
- a simulation without data assimilation (free run), differing from the true state simulation either by the initial conditions or by the resolution of the forcings or of the model grid,
- a simulation with data assimilation (reference simulation) with the same setup as the simulation without data assimilation and in which the data from the true state simulation will be assimilated.

In this study, the data were extracted at the assimilation time step from a true state simulation generated with the same setup but with a higher horizontal spatial resolution than the free run and reference simulation. The true state simulations have a horizontal spatial resolution of one nautical mile (1.852 km) whereas the free run and reference simulation have a horizontal spatial resolution of four nautical miles.

For the free run and reference simulation, an initial ensemble of states is generated on the 1st of September and is integrated without data assimilation till the 11th of September. The model error is sampled once a day using 50 ensemble members. Eight temperature profiles are assimilated once a day at midnight from the 12th of September till the 28th of September. Their impact on the neighboring temperature field is limited by means of an assimilation cut-off radius.

The one nautical mile simulations of the North Sea show significant differences in comparison to those with a four nautical miles resolution [She et al., 2006] [86] :

- eddy structures with scales of a few kilometres, visible along the thermal fronts in the one nautical mile resolution simulations, are not resolved by the coarser grid,

- an unrealistic over-dispersion of coastal plume water is taking place within the Channel and Southern North Sea in the 4 nautical mile resolution surface salinity plots,

- large vertical displacements of the thermocline: this feature of the North Sea dynamics is induced by winds and tides, the amplitudes are much larger in the high resolution run.

Available observations for the North Sea domain

For assimilation The study presented here started in the perspective of the European Commission FP5 project ODON (Optimal Design of Observations Networks). One aim of the ODON project was to develop a tool based on data assimilation to assess existing observations' networks and to design future ones. It was chosen to focus on a network of buoy data for the following reasons. Satellites provide surface data with a very high spatial coverage and a temporal resolution around one day. However, even in cloud free conditions their error remains larger than that from CTD in situ data. Furthermore, no vertical information is available from satellite data while profile measurements can be provided either by CTD data or buoys. They have a very high temporal resolution but are limited to a fixed location and exist only at a few locations; data can be transferred in near real time and they are less expensive to operate than satellites.

The network of observations examined in this study has been designed by J. Hoyer (Danish Meteorological Institute). It is made of eight buoys extracted from a network of existing observation locations. They are located along the Belgian coast, in the German Bight and in the United Kingdom central part of the North Sea. Their location allows to represent the two dynamical regimes of the North Sea in summer. It is shown on Figure 3.7 and their coordinates are given in Table 3.1.

For validation To validate the results for Septembre 2001, an independent set of observations is only available at stations 10, 11 and 15.

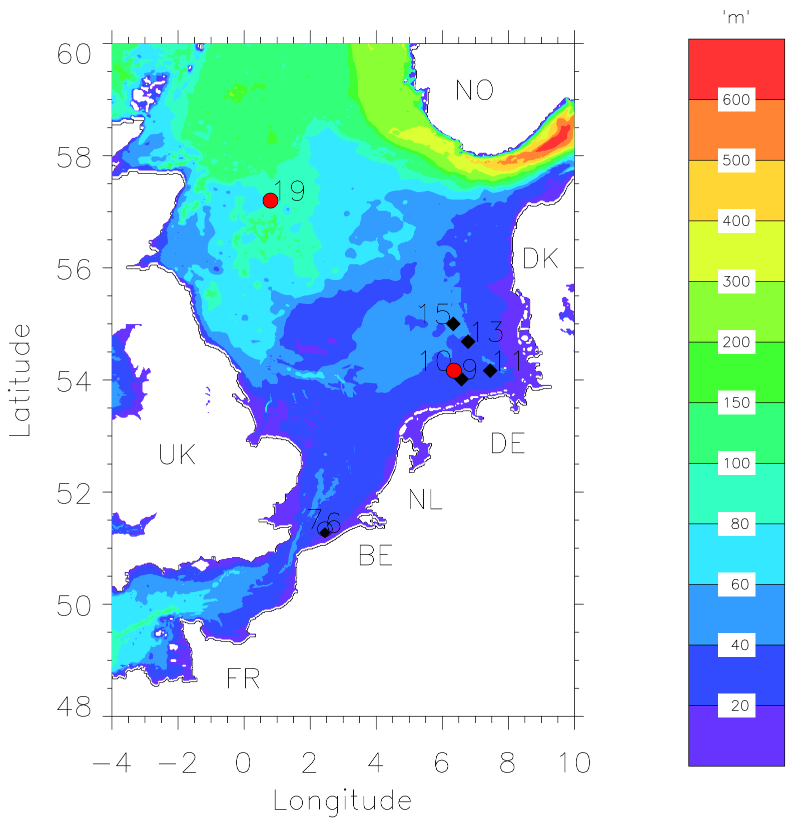

Figure 3.7: Stations at which temperature profiles will be assimilated for the 3-D study.

3.4.3 Data assimilation experiments

Simulations for model state estimation

Model state vector and estimation state vector There are six model state variables so that the model state vector is (T, S, U, V, ζ, k). The estimation state vectors contain the elements that are corrected by the data assimilation process while the variables not included are updated by the model dynamics. As temperature data are assimilated, a multivariate analysis is applied only to variables primarily related to temperature so that the estimation state vector is defined as (T, S, U, V). The kinetic turbulent energy is not included as it is a strongly non linear variable adjusting quickly to the model dynamics at the forecast time steps [Parent et al., 2003 [80];

Stat.	Lon. [°E]	Lat. [°N]	Reg.
6	2.448	51.274	M
7	2.448	51.338	M
9	6.583	54.016	M
10	6.354	54.166	M
11	7.450	54.166	M
13	6.783	54.683	M
15	6.333	55.000	M
19	0.800	57.200	S

Table 3.1: Description of the eight stations where data are assimilated in the 3-D case: station number, longitude, latitude, North Sea regime (M: well-mixed, S: stratified in summer).

Testut et al., 2003 [94]].

An ensemble of temperature is generated by adding terms of random perturbation, q, to the temperature variable. Ensembles for other model state variables are initialised to the value of the variable and then generated from the model dynamics.

1-D case: CS station Possible error sources in the model representation of the 1-D test case stem from the meteorological forcing (mainly the wind), from discrepancies between the measured water depth locations of the CTD sensor and the model bathymetry. The advection terms are omitted in the 1-D simulations which are therefore based on a very simplified or 'bad' representation of the physical processes governing the temperature field. When comparing these possible error sources, it is clear that the oversimplified physics is the main source of error in model representation. The error stemming from the meteorological wind forcing is difficult to quantify because the propagation of errors from meteorological data to the model temperatures is not straightforward [Holt and James, 1999] [42].

Simulations focusing on model state estimation are summarized in Table 8.5 given in Appendix D. A first series of experiments deals with the square root algorithms. The experiment performed without data assimilation is referred to as the 'free run' or 'CS-1', while the one with data assimilation as the 'reference simulation' or 'CS-2'. These experiments are compared to determine the benefit of data assimilation on the temperature representation.

Then, experiments 'CS-3' to 'CS-8' are performed to determine an adequate set of parameters for model error sampling by varying the vertical decorrelation length, the ensemble standard deviation and the number of ensemble members. Simulations 'CS-9' and 'CS-10' allow to test the influence of the assimilation time step.

A second series of experiments for model state estimation are performed with the low rank square root algorithm allowing an ensemble representation of the observations' error. Experiment 'CS-11' serves as reference experiment for the low rank square root algorithm. Experiments 'CS-12' to 'CS-13' are related to a quantitative assessment of the role of the vertical correlation when errors on the observations are taken into account.

3-D case: North Sea domain Table 8.6 in Appendix E summarizes the numerical experiments that are performed. Experiment 'NS-1', performed without data assimilation, is referred to as the 'free run' while experiment 'NS-2' involving data assimilation is referred to as the 'reference simulation'.

Experiments labelled 'NS-3' to 'NS-10' are designed to determine a set of adequate parameters for model error sampling. They are performed with the square root algorithm to determine the vertical and horizontal decorrelation scales, standard deviations of the ensemble, number of ensemble members and assimilation radius.

A series of simulations has also been performed with the low rank square root algorithm allowing an ensemble representation of the model error. Experiment 'NS-11' is the reference data assimilation experiment with the low rank square root algorithm. In experiments 'NS-12' to 'NS-13' we come back to the influence of a vertically correlated model error and provide a quantitative assessment of its impact in the case of the low rank square root algorithm.

Simulations for combined model state and parameter estimation

Another source of error in the modelled temperature is related to the parameterization of the physical processes. The selected parameters enter directly in the formulation of the temperature which is the assimilated data. Heat exchanges at the surface, either by radiation or by fluxes of latent and sensible heat have a large influence on the vertical exchange in the water as they change the buoyancy of the water [Otto et al., 1990] [78]. The parameters more directly involved in the surface heat fluxes are the surface heat exchange coefficients. Then, the optical attenuation has a direct influence on the penetration depth of the solar radiation in the upper mixed layer. The choice to estimate surface heat exchange coefficients is especially justified in the 3-D case as most of the stations are well mixed so that the temperature profiles are largely influenced by the surface heat fluxes.

As mentioned in the model description, the heat flux at the sea surface has a solar and a non solar component. The solar component comprises two terms. The first one corresponds to the solar radiation absorbed in the first two metres of the sea surface and can not be dealt with in this study. On the other hand the second term, K_{opt} the optical attenuation coefficient, is related to the temperature and optical properties of the upper mixed layer.

The non solar component of the heat flux comprises three terms: the latent and sensible heat components and a long wave radiation term. The formulation of the sensible and latent heat exchange fluxes depends on the latent and sensible heat exchange coefficients for which a large variety of parameterizations exists.

The neutral and stratified Kondo formulations of the surface heat exchange coefficients are used in this study. The stratified formulation of the surface heat exchange coefficients is related to the sea-air temperature difference and wind speed following equations (3.27) and (3.28). Therefore the stratified formulation is considered as more realistic than the neutral one. The combined estimation of the temperature and heat exchange coefficients will be performed on their neutral formulation and the results will be assessed by comparison to the stratified formulation.

Two cases for parameter estimation arise: constant and time-varying parameters. The optical attenuation coefficient is a function of the optical properties of the water (suspended particulate matter). These processes are not incorporated in our implementation of the model. In the 1-D case, "real" observations are assimilated so that it is expected that the information from these processes is present in the assimilated observations. So, once the assimilation has started, this parameter becomes time-varying. In the 3-D simulations, data are assimilated from 3-D twin experiments so that this parameter remains constant. Concerning the heat exchange coefficient, simulations are performed with the neutral Kondo formulation depending only on the wind speed at 10 metres height. However, the assimilated temperature data are derived from a simulation with a more realistic formulation taking into account a further dependence on sea-air temperature difference. So, in this case, a time variation of the state variable type is introduced in these parameters.

Ensemble of parameters are generated with the same technique as the temperature ensembles, by adding error terms to their default values used in the simulations of model state estimation. As the parameters considered are 2-D matrices, the error terms have a horizontal correlation. These error terms are constant. So, it is expected that their estimation converges to a certain value. At each analysis step, they are updated following the Kalman analysis step described by equation (2.8) and their ensemble spread is expected to become smaller and smaller.

The aims of these simulations are first to identify the benefit of the combined estimation of model state and parameters so that the model state is estimated in all the simulations in the same way as in 'sole' model state estimation. The second objective is to examine the spatial variability of the adjusted model parameters.

1-D case: CS station Simulations dealing with combined model state and parameter estimation are listed in Table 8.7 in appendix F. In experiment 'CPA-1' the optical attenuation coefficient is estimated while experiment 'CPA-2' deals with the estimation of the heat exchange coefficients. The simultaneous estimation of the set of selected parameters is performed in experiment 'CPA-3'.

3-D case: North Sea domain The same kind of simulations as in the 1-D test case are performed except that in the 3-D case a time correlated model error is used. They are listed in Table 8.8 in Appendix G.

Model error and estimation state vectors

The combined estimation of the model state and parameters is based on the state vector augmentation technique. Four different configurations for the estimation state vectors can be considered:

$$AD0 = \begin{pmatrix} T \\ S \\ U \\ V \end{pmatrix} \quad AD1 = \begin{pmatrix} T \\ S \\ U \\ V \\ K_{opt} \end{pmatrix} \quad AD2 = \begin{pmatrix} T \\ S \\ U \\ V \\ C_{en} \\ C_{hn} \end{pmatrix} \quad AD3 = \begin{pmatrix} T \\ S \\ U \\ V \\ K_{opt} \\ C_{en} \\ C_{hn} \end{pmatrix}$$

The $AD1$ state vector consists of the $AD0$ state vector augmented with the optical attenuation coefficient. The $AD2$ state vector comprises the $AD0$ state vector and the latent and sensible heat exchange coefficients. The $AD3$ state vector consists of the $AD0$ state vector augmented by the set of selected model parameters.

Data assimilation is applied in a twofold prospect: for model state estimation and for combined model state and parameter estimation. The $AD0$ state vector is used in experiments for model state estimation while 'AD1', 'AD2' and 'AD3' are for combined estimation.

Chapter 4

Model error sampling and state estimation with the ensemble Kalman filter in the 1-D case: the North Sea CS station

The North Sea CS station 1-D vertical test case is a standard simulation of the validation chain of the COHERENS model, the testing of the implementation of the analysis scheme in that case is a preliminary step to more realistic 3-D simulations. Temperature profiles from the 'North Sea project' data base are assimilated once a day from August 5th till August 24th with a vertical decorrelation length of 10 metres and an ensemble standard deviation of 1.0°C (product of "horizontal" and vertical standard deviation). The setup of the simulations is described with more details in the previous chapter.

This chapter is divided in five sections. In the first section, the gap between the model and the assimilated observations is examined. Then, in the next section, results obtained with the square root algorithm (without ensemble representation of the observations' error) are analysed. As they are based on the assumption of perfect observations, they allow to assess the maximum effect of the data assimilation process. So, this algorithm is used to determine a set of adequate parameters for the ensemble generation: vertical decorrelation length, multiplicative factor of the standard deviation of the ensemble, influence of the number of ensemble members. Also examined is how the assimilation process allows to take into account features badly represented by the model. The role of the assimilation time interval is also studied. Finally we also assess if the corrected features stay incorporated in the model when the assimilation of data is stopped.

Furthermore, as it permits to take into account an ensemble representation of the model error the low rank square root algorithm is applied. The third section is dedicated to the study of the results obtained with that algorithm. A quantitative analysis of the influence of a vertically correlated representation of the model error on the analysed temperature profiles is provided (root mean square error, model bias, correlation coefficient). Then, the effect of the temperature assimilation on the stratified expression of the heat exchange coefficients is examined. In the fourth section, results obtained with both algorithms are compared. The fifth section is dedicated to the conclusions. The characteristics and reference numbers of the different simulations are summarized in Table 8.5 in Appendix D.

4.1 Model error estimation in the 1-D case

In a first step, we identify the model errors that are particularly related to the 1-D configuration of the model. As already mentioned in the description of the simulations, the main source of errors for these 1-D simulations is the lack of advection terms.

The numerical experiment performed without data assimilation is referred to as 'CS-1'. Figure 4.1 presents time-depth series of the temperature modelled without data assimilation (left) and of the observations used for assimilation (right). Both are plotted with a temporal resolution of three hours.

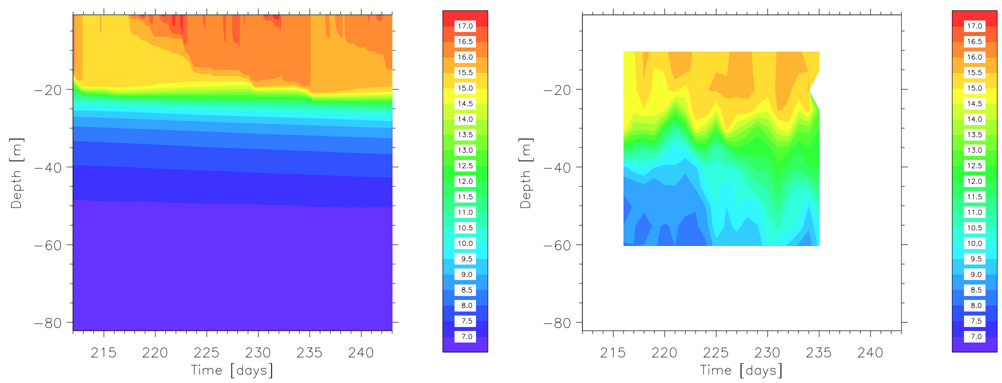

Figure 4.1: Time-depth series of the temperature modelled without data assimilation (left), 'CS-1', and of the observations (right).

The white frame around the graph of the right panel indicates that observations are missing at the surface (above 10 metres depth), and below 60 metres, and are available only from the 5th (day 216) until the 24th of August (day 235). The observed temperature profiles are composed of three zones (surface, thermocline and bottom), they show a sharp thermocline positioned at about 30 metres depth.

Figure 4.2 provides time-depth series of the difference between the observations and the model. To better highlight the error features, the differences are plotted with a temporal resolution of one hour. The left panel corresponds to the difference between the observations and the temperature modelled without data assimilation (experiment referred to as 'CS-1') while the right panel to the difference between the observations and the temperature modelled with data assimilation (experiment referred to as 'CS-2').

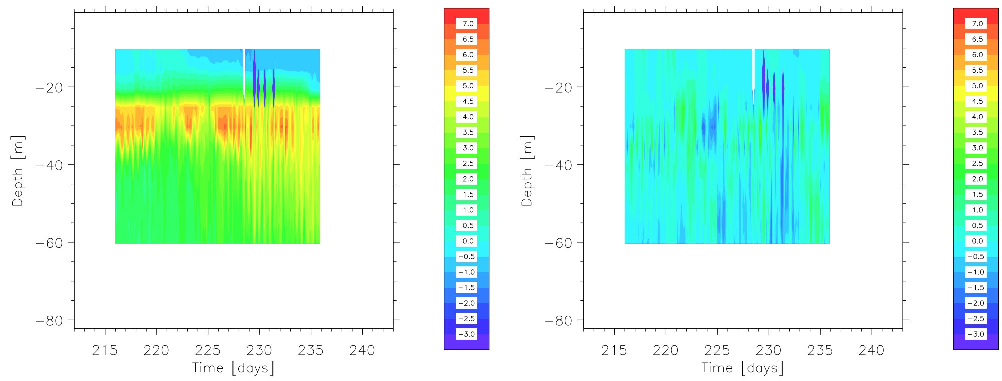

Figure 4.2: Time-depth series of the difference between the observations and the temperature modelled without data assimilation 'CS-1' (left) and of the difference between the observations and the temperature modelled with data assimilation 'CS-2' (right).

A comparison of the temperature profiles modelled without data assimilation with the observations indicates that the modelled position of the thermocline is not correctly represented, it is located at about 20 metres depth rather than at the observed 30 metres. The left panel of Figure 4.2 indicates that the largest temperature difference between the observations and the model without assimilation is located at the level of the thermocline. Because of the too high modelled position of the thermocline, the modelled temperature in that zone is over-estimated (too warm) and the temperature of the water mass located below the thermocline is under-estimated (too cold). Between 10 and 15 metres depth the temperature difference between the observations and the model is mainly negative, with values of up to −1°C. Around 20 metres, the difference between the observations and the model becomes positive. That depth corresponds to the thermocline position modelled without assimilation, in this case model values are colder than the observations. At

30 to 35 metres depth, which is the depth of the observed thermocline, the difference between the observations and the model is the largest, with values reaching 7°C. Below 40 metres the differences between the observations and the model are mainly positive with values ranging between 2°C and 4°C.

Moreover, the observations show regular upward and downward displacements of the thermocline. An analysis of North Sea temperature observations by van Haren, 2000 [96] indicates that this variability is primarily due to the advection by tidally driven currents. That study also shows that the basic harmonics in the frequency spectrum are the semidiurnal (ω_{M_2}) and inertial (f, Coriolis frequency) modes. Interaction between these basic harmonics generates an additional series of harmonics ($2\omega_{M_2} - f$, $2f$, $\omega_{M_2} + f$, ω_{M_4}, ...). The investigation of the observations used for assimilation in this study has been provided by Luyten et al., 2003 [63]. It reveals that the oscillations are mainly semi-diurnal during the period of spring tide under moderate wind conditions while the inertial mode becomes dominant during a strong wind event with a $2f$ harmonic.

The model simulation without assimilation at the CS station ('CS-1') is unable to represent the oscillations present in the thermocline of the observations. Although the tides are introduced as lateral forcings, the absence of advection terms prevents a correct propagation of the induced perturbations. The simulations by Luyten et al., [2003] [63] were performed with the 3-D version of the model and a horizontal spatial resolution of 7.3km; even with that model setup, these oscillations were underestimated. In a further study, Luyten et al., [2005] [64] show that these vertical oscillations, driven by tidal-inertial motions, are only realistically represented with a one nautical mile horizontal resolution (a four nautical mile resolution is too coarse).

4.2 Considering perfect observations: the square root algorithm

In this section, the square root algorithm is applied to compute the analysed model state. In this case, the observations are considered as perfect and their impact on the model is maximal. At first, we determine what missing features of the modelled temperature profiles are improved by the assimilation process and if the improvement is held when the assimilation is stopped. Then we seek for a set of adequate parameters to generate the ensemble and examine the role of the assimilation time step. Finally, the overall quality of the model error sampling is assessed by comparison of the standard deviation of the ensemble to the root mean square of the innovations.

4.2.1 Modelled temperature and ensemble standard deviation

This section focuses on comparisons of the temperature profiles and of the corresponding standard deviation of the ensembles obtained with and without data assimilation. The ensemble members are generated following equation (2.27) so that they are vertically correlated.

Parameters for model error sampling are the same for the simulations with and without assimilation. As these parameters are at the basis of the generation of the ensemble representing the model error, their adequacy is examined in terms of their impact on the temperature representation with data assimilation.

Temperature field

Figure 4.3 presents the time-depth series of the assimilated observations(left) and of the temperature modelled with data assimilation for August (right) referring to experiment 'CS-2'. The observations are plotted once a day at the assimilation time while the plot for the modelled temperature has a temporal resolution of 3 hours.

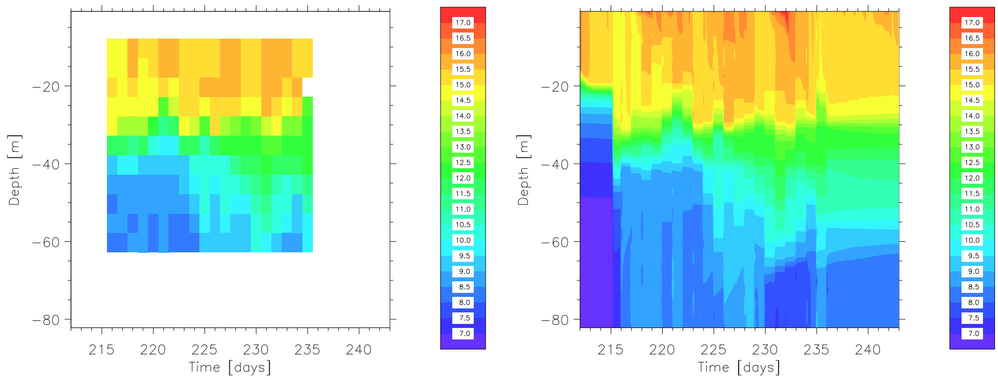

Figure 4.3: Time-depth series of the assimilated observations (left) and temperature modelled with data assimilation, experiment 'CS-2' (right).

The effect of the assimilation of temperature profiles is shown on the right side of Figure 4.3, it is compared to the assimilated observations on the left panel of Figure 4.3. The position of the thermocline is improved, it is shifted from 20 metres depth to about 30 metres depth. The temperature field located above 10 metres depth and below 60 metres depth, where no observations are available for assimilation, is also updated with the correct tendency i.e. colder above the thermocline and warmer below it.

The assimilative run presents oscillations at the level of the thermocline with a period of about one day which is the assimilation time step. In this very simplified test case, advection terms are missing. If one wishes that the assimilation process compensates for this, it should have a time step much smaller than one day. Indeed, the period of the oscillations shown on Figure 4.14 is of the order of a few hours. The influence of the assimilation time step will be further discussed in Section 6.2.4.

Figure 4.2 presents hourly differences between the observations and the temperature modelled without data assimilation (experiment 'CS-1') (left) as well as differences between the observations and the temperature modelled with data assimilation (experiment 'CS-2') (right). A comparison of the left and right panels of Figure 4.2 illustrates the effect of the assimilation process in reducing the gap between the model and the observations.

The differences between the observations and the temperature modelled with assimilation are of the same order of magnitude along the whole water column. After assimilation the remaining gap is close to zero (light blue) at many places, with maximum differences in the range of [−2.5°C, 2.5°C]. The error reduction is particularly strong in the thermocline zone where it is decreasing from 7°C to a maximum value of 2.5°C.

Correlation coefficient between the results with assimilation and the observations

The computation of the correlation coefficient between the model results after assimilation and the observations helps to assess the efficiency of the assimilation process. It is computed following equation (8.8) given in Appendix H. The correlation coefficient is computed at each analysis step and the averaging is performed on each vertical level where observations are available. It is shown on Figure 4.4.

On the first day of assimilation, the correlation coefficient is very bad (0.27) but it is quickly rising from the second day on to reach a value close to 1. Such high values for the correlation coefficient are interpreted as the fact that this 1-D trimmed down instance of the model with incomplete physics has incorporated the information contained in the observations as a substitute to the missing physical processes.

Standard deviation of the temperature ensemble

The model error representation is a critical issue when working with Kalman based methods. In ensemble Kalman filter applications it is estimated by means of an ensemble of model states. In this sub-section, we examine the

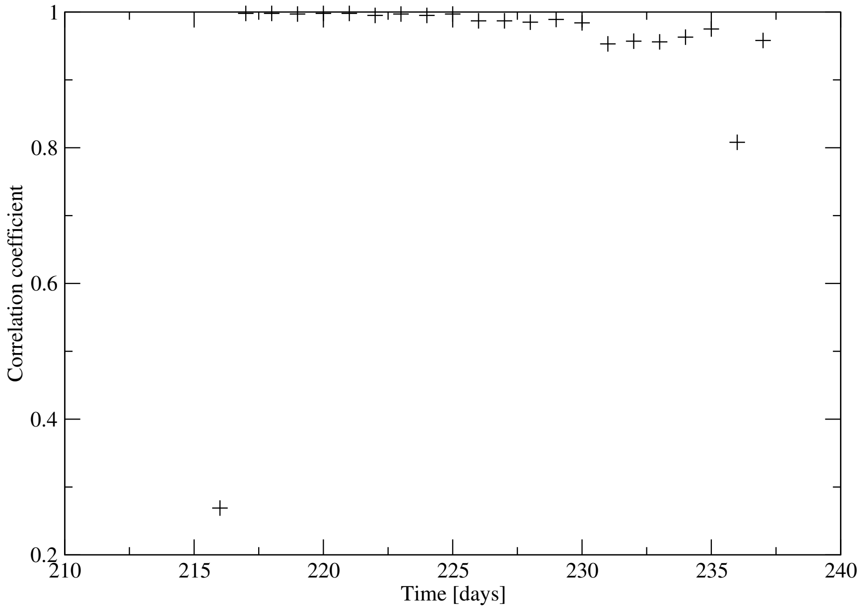

Figure 4.4: Correlation coefficient between the model 'CS-2' and observations.

temporal evolution and the vertical distribution of the standard deviation of the temperature ensemble. Figure 4.5 presents time-depth series of the standard deviation of the ensemble without (left panel) and with data assimilation (right panel).

Standard deviation of the temperature ensemble without data assimilation The left panel of Figure 4.5 corresponds to the standard deviation of the temperature ensemble without data assimilation. The generation of the ensemble starts in May. An ensemble of vertical temperature profiles is generated from a distribution of pseudo-random numbers with zero mean and variance equal to one. At each sampling time step (every 24 hours), small random perturbations are added to each ensemble member allowing to progressively generate gaussian perturbations while not shaking too strongly the model dynamics by the addition of large error terms at once. They are integrated by the model dynamics so that correlations develop between the model variables before starting the assimilation. The ensemble standard deviation reaches a value of about 1°C at the beginning of August when the assimilation process starts. The vertical correlation between the model error terms is built from bottom to surface following equation (2.27). As a result the random perturbations are added along the vertical, leading to smaller standard deviations at the bottom than at the surface.

When no data are assimilated, the modelled position of the thermocline is around 20 metres depth. During the ensemble generation, a small perturbation added to the temperature at that depth is increased by the model dynamics and results in a standard deviation of the ensemble larger than 1°C around 20 metres depth.

Standard deviation of the temperature ensemble with data assimilation As soon as data are assimilated (right panel of Figure 4.5), a sharp decrease in the standard deviation of the ensemble appears. This is observed from the first day of assimilation, day 216 (5th of August), and applies to the whole water column. This sharp decrease in the ensemble standard deviation is related to the error variance minimizing property of the Kalman filter.

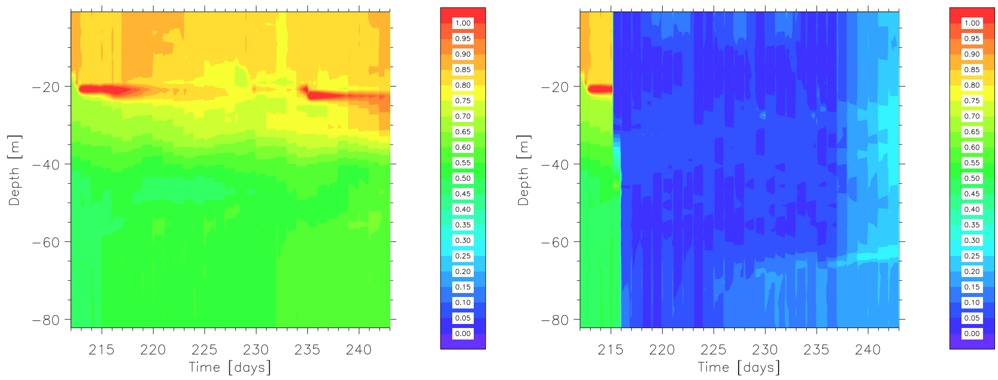

Figure 4.5: Time-depth series of the temperature standard deviation modelled without (left) and with (right) data assimilation, respectively experiments 'CS-1' and 'CS-2'.

4.2.2 Fate of model corrections when the assimilation process stops

Up to now, we have seen that the assimilation process allows to correct the wrong position of the thermocline. The next question is of course to know if this correction is kept by the model once the assimilation process is stopped. To this aim, we have performed a simulation where data are assimilated until the 25th of August. Then, the model is run without data assimilation for September. This simulation is compared to a simulation of the September temperature modelled without any assimilation step.

The September temperature modelled without data assimilation in August is presented on Figure 4.6. Its thermocline is located around 20 metres depth.

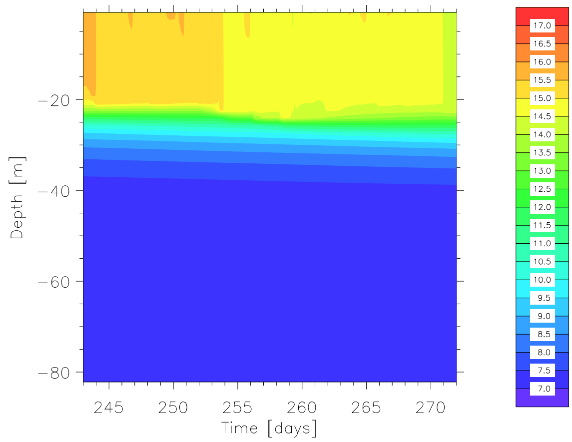

Figure 4.6: Time-depth series of the September temperature modelled without data assimilation in August.

The temperature modelled for September with preliminary assimilation in August is shown on the left panel of Figure 4.7 and partial observations are on its right panel. Its thermocline is displayed at 30 metres depth which is the adjusted position at the end of the assimilation process in August. Moreover, the temperatures of the water masses located above and below the thermocline are in good agreement with the available observations.

These features seem to indicate that a daily assimilation of temperature profiles with the ensemble Kalman filter is able to partly compensate for missing model dynamics. However, the observed position of the thermocline in September is shown on the left panel of Figure 4.7 and is located around 40 metres depth. This corresponds to the deepening of the thermocline occurring beginning of autumn.

The discrepancy between the modelled and observed thermocline depths is attributed to the fact that the dynamical processes allowing to update the thermocline depth are missing. As a result the assimilated features corrected by assimilation in August are not updated and can not be properly represented in September.

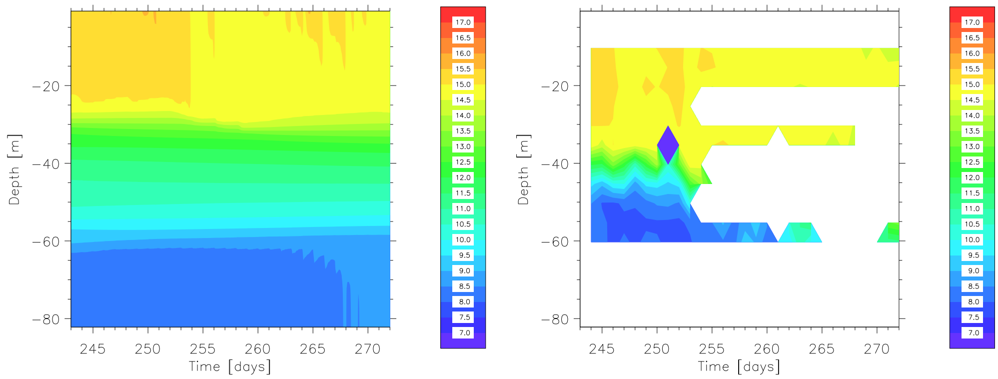

Figure 4.7: Time-depth series of the September temperature modelled with data assimilation in August (left) and of September observations (right).

4.2.3 Influence of the parameters for model error sampling

The estimate of the model error necessary to compute the Kalman gain is obtained through an ensemble of model instances. Several parameters are involved in the definition of the ensemble:

- η_T, defining the magnitude of the vertical standard deviation of the ensemble,
- ξ_T, the vertical decorrelation length of the model error,
- the number of ensemble members.

Multiplicative factor of the vertical standard deviation of the temperature ensemble

In this section, we test two values of η_T, the multiplicative factor of the vertical standard deviation of the temperature ensemble, and we examine how it affects the temperature profile and the standard deviation of the temperature ensemble.

Figure 4.8 represents the temperature profile modelled with a vertical standard deviation of 0.5°C (experiment 'CS-3') which is half the value used in the reference simulation 'CS-2'. The results for this experiment are close to those of the reference simulation 'CS-2' (right panel of Figure 4.3) except that the water mass below 60 metres depth is slightly warmer than in experiment 'CS-2'.

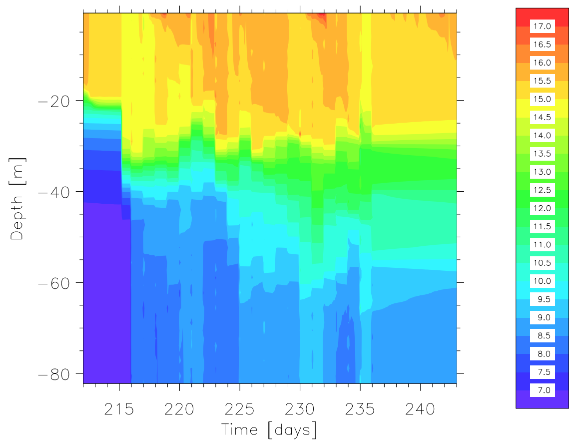

Figure 4.8: Time-depth series of the temperature modelled with a vertical deviation of 0.5°C, experiment 'CS-3'.

As expected from the respective values of the η_T multiplicative factor, figure 4.9 shows that the vertical standard deviation of the temperature ensemble for experiment 'CS-3' (left panel) has about half the value of that of the reference simulation 'CS-2' (right panel) .

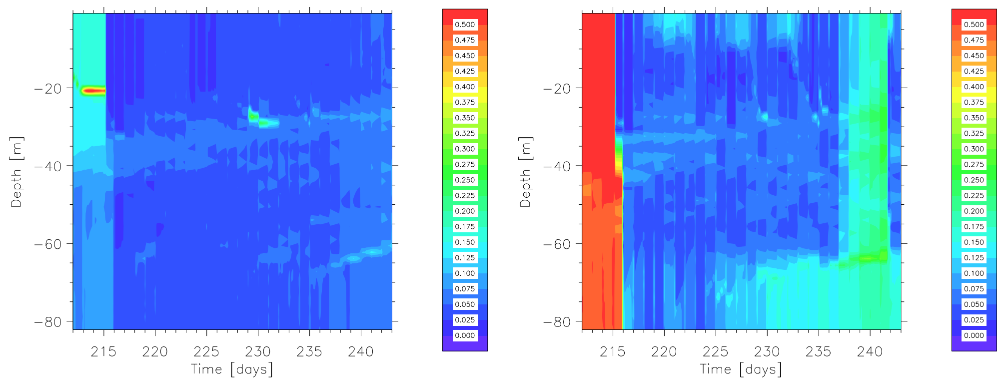

Figure 4.9: Time-depth series of the ensemble standard deviation with a vertical deviation of 0.5°C (left) and for the reference simulation (right), respectively experiments 'CS-3' and 'CS-2'.

In order to assess which value of the two multiplicative factors η_T is the most adequate, the root mean square of the innovations can be compared to the standard deviation of the ensemble. More details concerning this assessment are given in Appendix H.

To allow the assimilation process to incorporate the innovations into the model, the sampled model error should be of the same order of magnitude as the gap between the model and the observations [Haugen and Evensen, 2002] [40]. The model error is represented through the ensemble spread. To check that the gap between the model and the observations is at least contained in the ensemble spread, the standard deviation of the ensemble is compared to the root mean square of the innovations. Moreover, such a test allows to further use the assimilated data to validate the model results [De Mey et al.] [22].

The root mean square of the innovation vectors (black squares) is compared to the ensemble standard deviation (red circles) for experiment 'CS-3' (left panel) and for the reference simulation 'CS-2'. The results are plotted respectively on the left ('CS-3') and right ('CS-2') panels of Figure 4.10. In the reference simulation (right panel) these two quantities are close together so that the model error sampling is adequate. In experiment 'CS-3' (left panel), the innovations are larger than the modelled standard deviation indicating that the standard deviation of the ensemble is too low to represent the gap between the model. These results indicate therefore that a multiplicative factor for the vertical standard deviation of the ensemble, η_T, of 1°C is more adequate than a η_T of 0.5C.

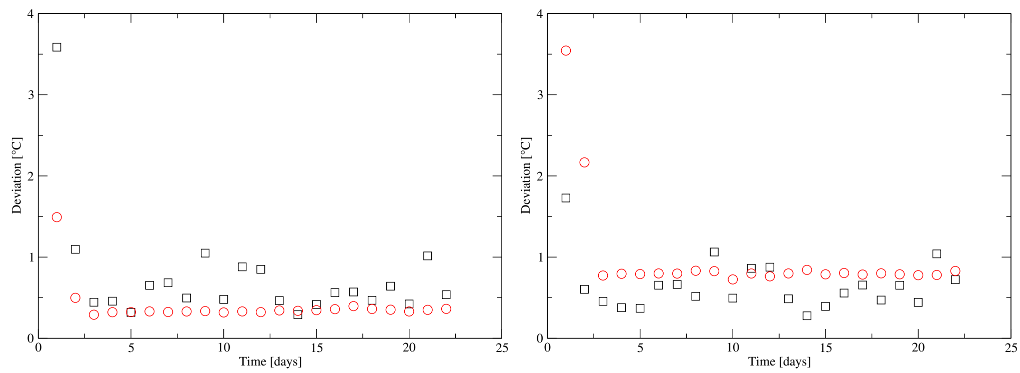

Figure 4.10: Root mean square of the innovations (black squares) and standard deviation of the modelled temperature (red circles) with a vertical deviation of 0.5°C (left) and for the reference simulation (right), respectively experiments 'CS-3' and 'CS-2'.

Vertical decorrelation length of the model error

The simulations with the square root algorithm presented up to now have been performed with a vertically correlated model error. A quantitative assessment of the benefit provided by tuning the vertical correlation of the model error will be given in a next section dealing with the low rank square root algorithm.

Here we provide a qualitative study of the influence of the vertical decorrelation length and examine therefore four cases among which two limit cases. The first limit case corresponds to numerical experiment 'CS-5' and is performed with a vertical decorrelation length of 2 metres which is of the order of magnitude of the vertical grid spacing; so $\alpha_v \sim 0$ and the sequence of model error terms is almost non-vertically correlated and made of random terms. The second limit case appears for a vertical decorrelation of 80 metres, the depth of the water column; in this case $\alpha_v \sim 1$ and the error terms are the same at all vertical levels so that the model error is vertically correlated over the whole water depth, it corresponds to experiment 'CS-6'. For the two intermediate cases, the vertical decorrelation length of the model error has been chosen by examining the observations. Observations are available between 10 and 60 metres depth. The time-depth series of the observed temperature on Figure 4.1 indicate a relatively good mixing within the upper and lower layers and suggest that a vertical decorrelation length ranging between 10 and 20 metres is a good candidate. A vertical decorrelation length of 10 metres (reference simulation 'CS-2') is sufficient to propagate to the surface the error reduction due to the data assimilated at 10 metres depth, while a vertical decorrelation length of 20 metres (double the one used for the reference simulation) might be useful for the data assimilated at 60 metres depth to influence the bottom temperature. The simulation performed with a vertical decorrelation length of 60 meters is referred to as numerical experiment 'CS-4'.

Hovmöller diagram of the modelled temperature The left panel of Figure 4.11 presents a time series of temperatures at the surface, at 30 metres and at the bottom for the four decorrelation lengths under study: 10 metres (reference simulation), 20 metres, 2 metres ($\alpha_v \sim 0$) and 80 metres ($\alpha_v \sim 1$).

No data are available at the surface and at 80 metres depth, so that the effects of the data assimilation on the model error at those two locations stem from the vertical correlation between the ensemble members. At the surface, the effect of varying the decorrelation lengths on the modelled temperature is negligible especially when the assimilation process is stopped, from day

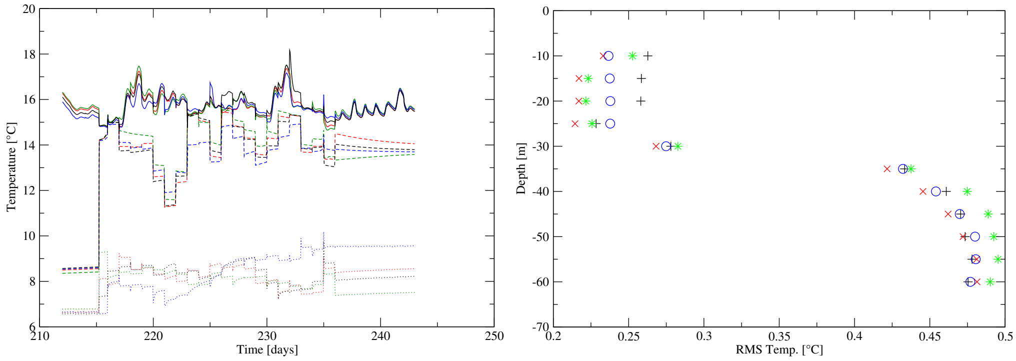

Figure 4.11: **Left panel:** Time series of the surface (straigth line), 30 metres depth (dashed line), bottom (dotted line) temperature modelled with a vertical decorrelation length of 10 metres (black), 20 metres (red), 2 metres (blue) and 80 metres (green) - **Right panel:** Root mean square error between the observations and the temperature modelled respectively with a vertical decorrelation length of 10 metres (black plusses), 20 metres (red crosses), 2 metres (blue circles) and 80 metres (green stars). Respectively experiments 'CS-2', 'CS-4', 'CS-5' and 'CS-6'.

236 (25th of August) on.

Generally speaking the effect of the assimilation of temperature profiles is to cool the water above the thermocline and to warm it below the thermocline. At 30 metres depth, a longer vertical decorrelation scale slightly reinforces this effect in the case of a temperature modelled with a decorrelation length of 20 metres (red dashed curve) warmer than for the reference simulation (black dashed curve).

At the bottom, the temperature modelled with a vertical decorrelation length of 2 metres (blue dotted curve) starts with the lowest values and increases to become larger than the bottom temperatures modelled with any other decorrelation length. Before day 225 (14th of August), it is colder than in the other cases with a difference reaching 3°C. After day 225 (14th of August) it is warmer, and the gap reaches 2°C when the assimilation process stops. As a result, during the first half of the assimilation period, the bottom temperature modelled with a decorrelation length of 2 metres is colder than in the other cases. This is not desired as we wish that the assimilation process reduces the surface temperature and increases the bottom temperature. So a vertical decorrelation length of 2 metres corresponding to $\alpha_v \sim 0$ or a model error not vertically correlated is not considered as the most adequate choice.

During the assimilation period, the 30 metres depth temperature modelled with a decorrelation length of 80 metres (green dashed curve) is essentially warmer than in the other cases while the temperature modelled with a decorrelation length of 2 metres (blue dashed curve) is mainly colder. The thermocline temperature modelled with vertical decorrelation lengths equal to 10 (black dashed curve) and 20 (red dashed curve) metres are close together during the whole assimilation period.

Root mean square errors between the observations and the modelled temperature In order to better interpret the model features related to the vertical decorrelation length, the root mean square errors between the observations and the modelled temperature have been computed for the four vertical decorrelation lengths considered. Results are presented on the right panel of Figure 4.11.

At 10, 15 and 20 metres depth, a vertical decorrelation length of 10 metres (black plusses) shows a root mean square error slightly larger than in the other cases, while below 20 metres depth it is not the case any more. Below the thermocline, the root mean square error of the simulation performed with a vertical decorrelation length of 80 metres (green stars) is always larger than for the other simulations.

From the results described above, the following remarks concerning the role of the vertical decorrelation length on the modelled temperature can be mentioned. First of all, the influence of the vertical decorrelation length is not critical to the results though these ones can be slightly improved when an adequate value is chosen.

Then, the bottom results obtained with nearly no vertical correlation (decorrelation length of 2 metres) are slightly worst than with larger decorrelation lengths. Below the thermocline, the root mean square errors between the model and the observations are larger than above it. The largest values appear for a vertical decorrelation length of 80 metres. Finally the performance obtained with decorrelation lengths of 10 metres or 20 metres are virtually equivalent.

As the CS station is located in the stratified zone of the North Sea, one would imagine a priori that a non vertically correlated model error is more adequate. However, as no data are available below 60 metres, the bottom temperature obtained in the absence of a vertical correlation is worst than with a vertically correlated model error. At the opposite, when using a too long vertical correlation in stratified waters the quality of the results decreases. A slightly improved performance is obtained with a decorrelation length of 10-20 metres, and a decorrelation length of 10 metres is finally chosen as reference value.

Number of ensemble members

Influence of the number of ensemble members on the temperature
In this section we show to which extent the temperature modelled with data assimilation is improved when the number of ensemble members is increased. For a given standard deviation of the ensemble, augmenting the number of ensemble members increases the chances to have an ensemble member very close to the observations. The numerical experiments 'CS-7' and 'CS-8' are performed respectively with 10 and 200 ensemble members. Their results are compared to the results of the reference simulation 'CS-2' performed with 50 ensemble members.

The root mean square (rms) error between the simulated temperature profiles and the observations is provided at each vertical level where observations are available. The mean is computed in time from values extracted every three hours. Results are plotted on Figure 4.12.

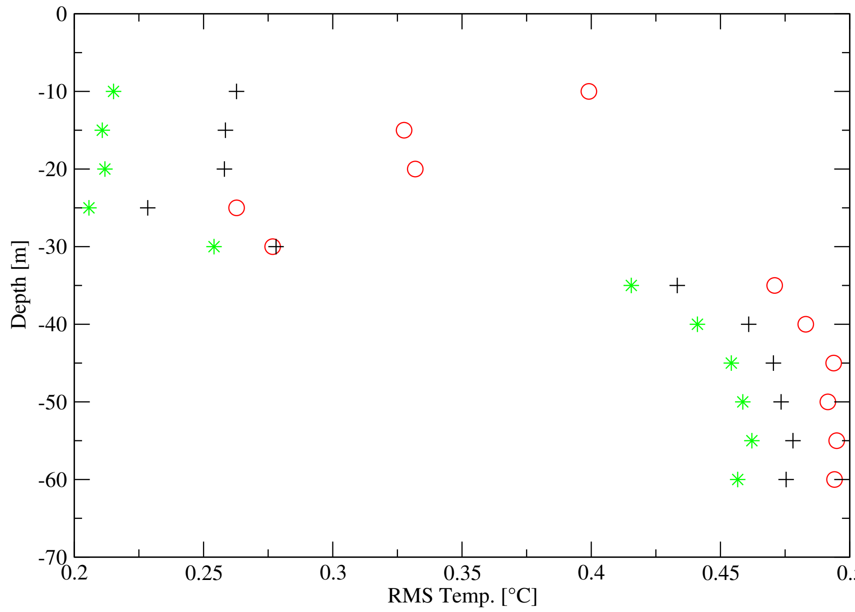

Figure 4.12: Root mean square error between the observations and the temperature modelled respectively with the following number of ensemble members: 10 (red circles) 'CS-7', 50 (black plusses) 'CS-2' and 200 (green stars) 'CS-8'.

At 10 metres depth, the benefit of augmenting the number of ensemble members is clear, with decreased root mean square errors when 200 ensemble members are used in place of 10. The rms errors of the simulations with 10 ensemble members (red circles) are nearly double of those of simulations performed with 50 (black plusses) ensemble members, with values of 0.4 °C for simulations with 10 members against 0.26 °C for simulations with 50 members. The rms errors between the simulations with 50 ensemble members and 200 ensemble members (green stars) are smaller, respectively 0.26 °C with 50 members against 0.21 °C with 200 members.

The root mean square errors of the simulations with 10 ensemble members (red circles) decrease from 10 metres depth to 25 metres depth while those of the simulations with 200 ensemble members are rather constant.

At 30 metres depth, the rms errors of the 10 and 50 ensemble members simulations are the same, around 0.28°C. This value is close to that of the simulation with 200 ensemble members (0.25°C).

Below the thermocline and down to the bottom, the rms errors of the three simulations increase continuously while remaining close together. At 60 metres depth, they reach 0.45°C in the case of 200 ensemble members and 0.49°C with 10 ensemble members.

To summarize, augmenting the number of ensemble members improves the temperature representation at all water depths. In the upper mixed layer, simulations with 50 ensemble members present clearly better results than with 10 ensemble members and are close to the simulations with 200 ensemble members. At the thermocline, the rms errors between simulations with 50 and 200 members are nearly the same. Below the thermocline, the gap between the three set of simulations is smaller than in the upper mixed layer. Therefore, the improvement in the model results from 50 to 200 ensemble members is not so significant as from 10 to 50 members. On the other hand, the augmentation of the number of ensemble members increases the duration of the simulations. As a result, 50 ensemble members are considered as performing sufficiently well without being prohibitive in computing time.

Influence of the number of ensemble members on the standard deviation of the temperature ensemble The influence of the number of ensemble members is further examined in terms of the improvement of the sampling of the model error. Here we examine how the number of ensemble members influences the representation of the standard deviation of the temperature ensemble when no data are assimilated.

Time-depth series of the standard deviation of simulations performed without data assimilation with 10 and 200 ensemble members are presented on Figure 4.13.

The standard deviation of the temperature ensemble is compared to that of the free simulation 'CS-1' performed with 50 ensemble members and shown

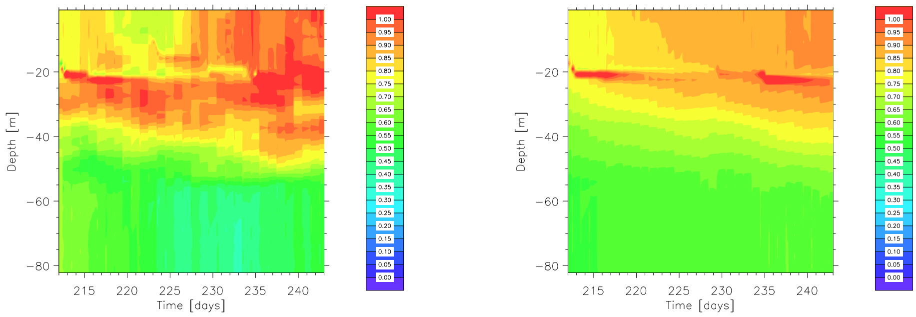

Figure 4.13: Time-depth series of the ensemble standard deviation obtained without data assimilation, using 10 and 200 ensemble members.

on the left panel of Figure 4.5. At the beginning of the assimilation period until day 226 (15th of August), the standard deviation of the upper mixed layer is smaller with 10 ensemble members than with 50 ensemble members; after day 236 it is much larger, with values around 0.95°C. The zone around the thermocline corresponding to the largest standard deviation is much thicker with 10 ensemble members than with 50 members.

Next, the standard deviation of the temperature ensemble derived from simulations with 50 ensemble members (left panel of Figure 4.13) is compared to that from simulations with 200 ensemble members (right panel of Figure 4.13). The standard deviation of the ensemble modelled with 200 ensemble members is vertically uniform in the upper mixed layer and shows four zones corresponding to values clearly defined. The zone around the thermocline presents the largest standard deviation and has roughly the same extent as with 50 ensemble members.

The standard deviation of the temperature ensemble is determined by the parameters γ_T and η_T. For given values of these parameters, the range of the standard deviation of the temperature ensemble is fixed. When the number of ensemble members taking their values in that interval is 200 in place of 10, many more possible values of temperature are represented. This resuls in smaller differences between each ensemble member and in a more uniformly represented standard deviation of the temperature. Moreover, the standard deviation of the ensemble slowly converges with the number of ensemble members [Hanea et al., 2007] [39]. A comparison of the two panels of Figure 4.13 highlights the influence of the number of ensemble members on the convergence of the standard deviation of the ensemble.

Final remark concerning the estimation of the model error sampling

In this section, we come back to the comparison of the root mean square of the innovations to the standard deviation of the ensemble for the reference simulation 'CS-2' and for experiment 'CS-3'. Experiment 'CS-3' is performed with a standard deviation of the temperature ensemble having half the value of the reference simulation (0.5°C in place of 1°C for the reference simulation). The results are presented on Figure 4.10, the root mean square of the innovations is represented by black squares and the standard deviation of the ensemble by red circles. The left side of Figure 4.10 presents the results of experiment 'CS-3' and its right side corresponds to the results of the reference simulation 'CS-2'.

For both experiments, on the first day of assimilation the standard deviation of the ensemble over-estimates the innovations and then, the difference is reduced due to the variance-minimizing property of the EnKF. On the right panel, the results of the reference simulation indicate that the standard deviation of the ensemble is of the same order of magnitude as the root mean square of the innovations. When the vertical standard deviation of the model error is reduced by half, the results (left panel) show an equivalent reduction of the ensemble standard deviation. In this latter case, the standard deviation of the ensemble has become too small with respect to the values of the innovations.

A comparison of the two panels of Figure 4.10 indicates that, given the values of other parameters for model error sampling of these 1-D simulations, a multiplicative factor of the vertical standard deviation of the ensemble, η_T, of 1°C is an adequate value to allow the assimilation scheme to incorporate the innovations.

4.2.4 Influence of the time interval between two assimilation steps - Representation of the thermocline oscillations

As mentioned previously, the two main features present in the observations and missed by the model without assimilation are the position of the thermocline and its oscillations. Up to now, the reference simulation has shown that the assimilation process is able to correct the position of the thermocline but that the oscillations are only partly represented because of a too large time interval between two assimilations. In this section, the role played by the assimilation time interval on the representation of these oscillations is investigated.

The frequency of these oscillations is around six hours. Figure 4.14 presents a time-depth series of the observed temperature plotted with a temporal resolution of one hour. The oscillations of the thermocline are clearly represented.

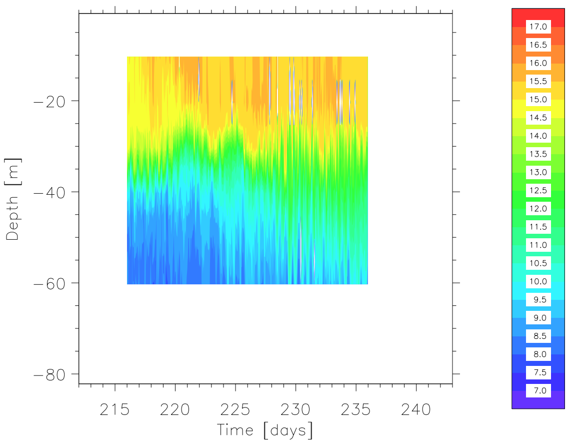

Figure 4.14: Time-depth series of the temperature observations at each hour.

Two experiments are reported in this section, they are respectively performed with an assimilation time interval of 3 hours (experiment 'CS-9') and of 6 hours (experiment 'CS-10').

The left panel of Figure 4.15 corresponds to experiment 'CS-9' which has been performed with the same filter parameters as the reference simulation except that data are assimilated every 3 hours. "Temperature anomalies" in the form of warmer and colder patches, which do not exist in the observations, are present in the modelled temperature profile. These "anomalies" are attributed to a possible difficulty of the model when assimilating data with a high frequency. An assimilation interval of 3 hours contains only 9 time integration steps of the model which could be insufficient for the model to incorporate the information and relax between two analyses.

Therefore an experiment with an assimilation time interval of 6 hours has been performed, the results are presented on the right panel of Figure 4.15 and correspond to experiment 'CS-10'. The "temperature anomalies" present in the 3-hours simulation are reduced. The oscillations of the thermocline are better represented in particular from day 217 until day 225; from day 226 until the end of the assimilation period, the agreement between the modelled oscillations and the observations is not sufficient in terms of values

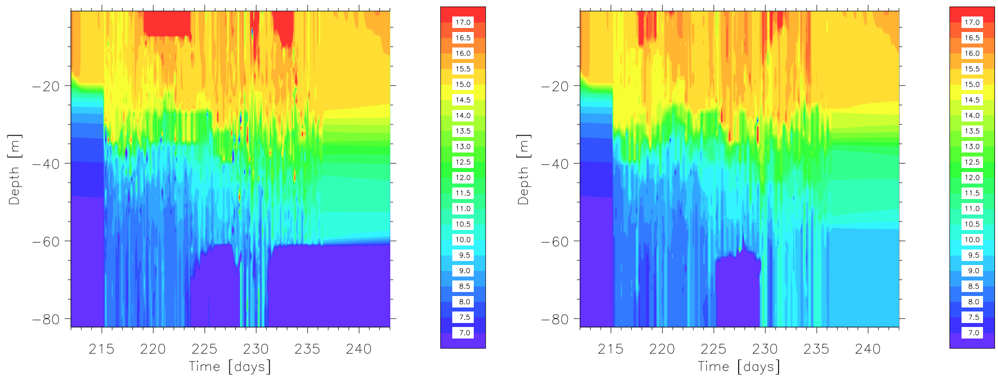

Figure 4.15: Time-depth series of the temperature modelled with data assimilation at a frequency of 3 hours (left) and 6 hours (right), respectively experiments 'CS-9' and 'CS-10'.

and amplitude of the oscillations.

However, assimilation time steps of 3 to 6 hours are too large to represent a signal with a period around 6 hours. Simulations not shown here have been performed with an assimilation time step of one hour. The spurious features reported in the above paragraph were still larger.

Finally, when applying a model with a very simplified physics, a small time interval between two assimilations may be necessary to represent important features of the observations missed by the model. However, an assimilation time interval that is too small to allow a sufficient relaxation of the model between two assimilation steps may worsen the situation. For the assimilation of temperature profiles reported in this 1-D vertical case study, an assimilation time interval of 6 hours was the smallest value consistent with the numerical integration time step. It allowed a better representation of the oscillations than a one day assimilation time interval.

4.3 Considering ensemble errors for model and observations: the low rank square root algorithm

Two square root algorithms of the ensemble Kalman filter are implemented in the COHERENS model. The first one is based on the assumption that the observations are free of error. It has been used in the previous sections of this chapter in order to assess the most adequate values for the specification of the model ensemble. The low rank square root (LRSR) algorithm examined in this section is based on an ensemble representation of the observations' error.

In this section, we present at first a quantitative analysis of the influence of the vertical correlation of the model error on the results by computing root mean square errors, model biases and correlation coefficients. Then, the modelled temperature is compared to that obtained with the square root (SR) algorithm. Finally we examine the effect of the temperature assimilation on the stratified formulation of the surface heat exchange coefficients.

4.3.1 Influence of a vertically correlated or non correlated model error on the modelled temperature

The ensemble Kalman filter has been implemented with the possibility to generate vertically correlated ensemble members. This allows to relate the model error at one vertical point to the neighboring ones according to equation (2.27). However, such a representation can be tricky to use in stratified regions as the temperature in one layer can be totally independent of that a few metres deeper. The key issues that will be dealt with in this section are:

- does a vertically correlated model error perform better than a non correlated one in a stratified region?
- if it is the case, than which value of the correlation scales are allowed?

Simulations are performed with vertically correlated and non-vertically correlated ensemble members. In this latter case, the sequence of model error terms is purely random, $\alpha_v = 0$, and the equation (2.27) reduces to:

$$q_v(k) = w$$

where w is the sequence of white noise generated from the distribution of pseudo random numbers. In the case of vertically correlated model error, the chosen vertical decorrelation length is of 10 metres. As the simulations are performed with 50 vertical σ level and as the depth of the CS station is around 80 metres, $\alpha_v = 0.84$.

Quantitative analysis of the influence of the vertical correlation of the model error on the modelled temperature

The performance of the assimilation process in the case of a vertically correlated or non-vertically correlated model error is compared through the computation of correlation coefficients, root mean square errors and model bias between the model and observations. They are computed at all vertical levels of the model. The time period for computing the mean extends from the beginning until the end of the assimilation period, i.e. from the 5th until the 24th of August.

Correlation coefficients The correlation coefficient gives a measure of the similarity between the modelled and observed fields, it is computed as:

$$C_{cor} = \frac{\frac{1}{n}\sum_{i=1}^{n}[(T_{mod}(i) - \bar{T}_{mod})(T_{obs}(i) - \bar{T}_{obs})]}{\sigma_{T_{mod}}\sigma_{T_{obs}}} \tag{4.1}$$

where n is the number of time series for model output, $\bar{T}_{mod}$ and $\bar{T}_{obs}$ are the mean of the model T_{mod} and observations T_{obs} over n points, $\sigma_{T_{mod}}$ and $\sigma_{T_{obs}}$ are their respective standard deviations.

Figure 4.16 represents the correlation coefficient between the modelled temperature profiles and the observations. First of all, the correlation coefficients of the non assimilative runs (black plusses and red crosses) are small at the surface and larger at the bottom with respective values of 0.2 and 0.6. Values for the vertically correlated and non-vertically correlated simulations are very close together in the case of the non assimilative runs.

The thermocline is the most badly represented feature of the runs without data assimilation. After assimilation (blue circles and green stars), it is the water depth where the correlation coefficient is the largest (0.85) while without assimilation its value was extremely small (less than 0.1).

At the surface the assimilation process increases the correlation coefficient by a factor of 3. At the bottom the increase is not so large but the correlation coefficient passes from around 0.6 to 0.8 which indicates a clear benefit of the assimilation process.

At all depths the largest values of the correlation coefficient are obtained for the assimilative runs with a vertically correlated model error (green stars).

Root mean square errors The correlation coefficient measures the similarity of the two fields while the magnitude of their differences is given by the root-mean square error. It is computed under the following form:

$$RMS = \frac{1}{\sqrt{n}}[\sum_{i=1}^{n}(T_{mod}(i) - T_{obs}(i))^2]^{1/2}, \tag{4.2}$$

where n is the number of assimilative time steps.

The root mean square errors between the modelled temperature profiles and the observations are presented on Figure 4.17. For the non assimilative

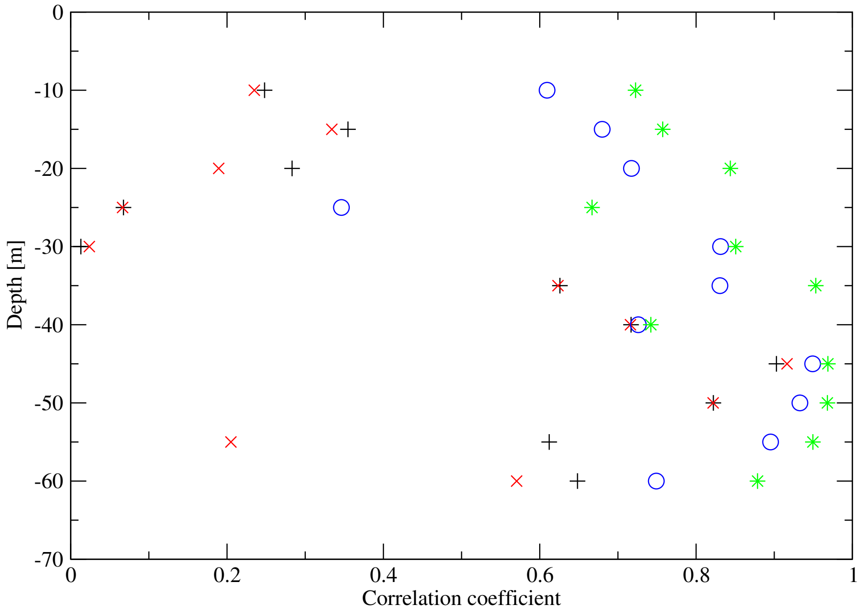

Figure 4.16: Correlation coefficient between the observed and modelled temperature without data assimilation for a non-vertically correlated 'CS-12' (black plusses), a vertically correlated 'CS-1' (red crosses) model error and with data assimilation for a vertically non correlated 'CS-13' (blue circles) and a vertically correlated 'CS-11' (green stars) model error.

runs the rms errors of the vertically correlated (red crosses) and non-vertically correlated (black plusses) model errors are very close together at all depths.

The rms errors between the runs without data assimilation and the data are the largest at the thermocline (around 5°C). After data assimilation this difference is reduced by an order of magnitude. The rms errors are clearly dominated by model bias features as can be seen from Figure 4.18 where rms values are a mirror of model biases.

In the upper mixed layer the rms errors between the modelled temperature field and the assimilated data are almost the same for the two assimilative runs (green crosses and blue circles). The assimilation process reduces the rms errors by a factor of two.

Below the thermocline, the rms errors of the non assimilative runs with respect to the observations are quite large, ranging from 2.25°C at 60 metres depth to 4°C at the thermocline. The assimilation process reduces the rms errors in the range of [0.25°C, 0.5°C].

At all examined water depths the results from the runs performed with

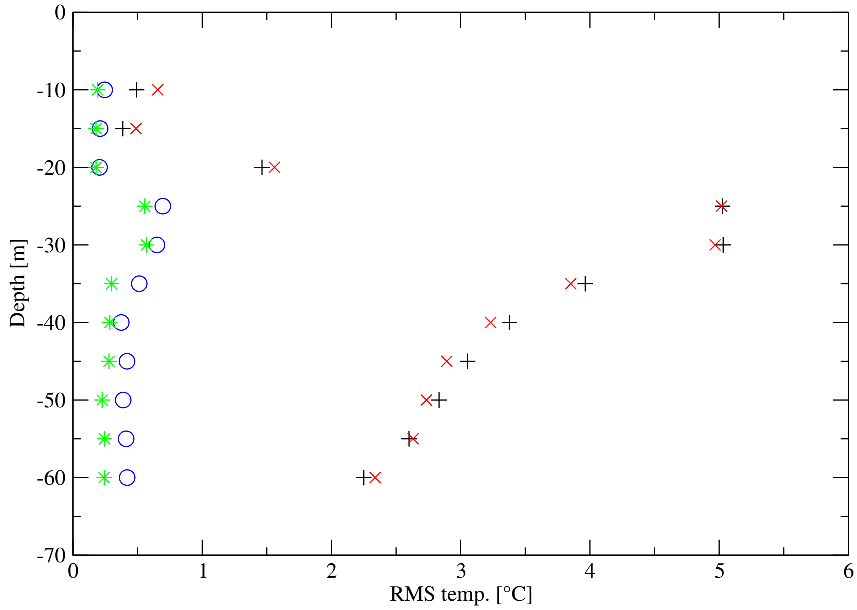

Figure 4.17: Root mean square error between the modelled and observed temperature. Four cases of modelled temperature are considered: without data assimilation for a non-vertically correlated 'CS-12' (black plussses), a vertically correlated 'CS-1' (red crosses) model error and with data assimilation for a non-vertically non correlated 'CS-13' (blue circles) and a vertically correlated 'CS-11' (green stars) model error.

vertically correlated model errors (green stars) present the smallest rms errors between the modelled temperature field and the observations.

Model bias The mean error between the model and the observations is computed as follows:

$$\bar{T}_{mod} - \bar{T}_{obs}. \tag{4.3}$$

Results are presented on Figure 4.18.

After data assimilation the bias between the modelled temperature fields and the observations is close to zero. For the non assimilative runs, the mean biases at 10 and 15 metres depth are positive. This indicates that the surface temperature is over-estimated with respect to the observations.

The model bias attains a maximum value of $-5°$C at 25 and 30 metres depth. This is partly caused by the too high representation of the thermocline without data assimilation. As a result the temperature of the upper mixed layer is too warm with respect to the observations, and below the thermocline

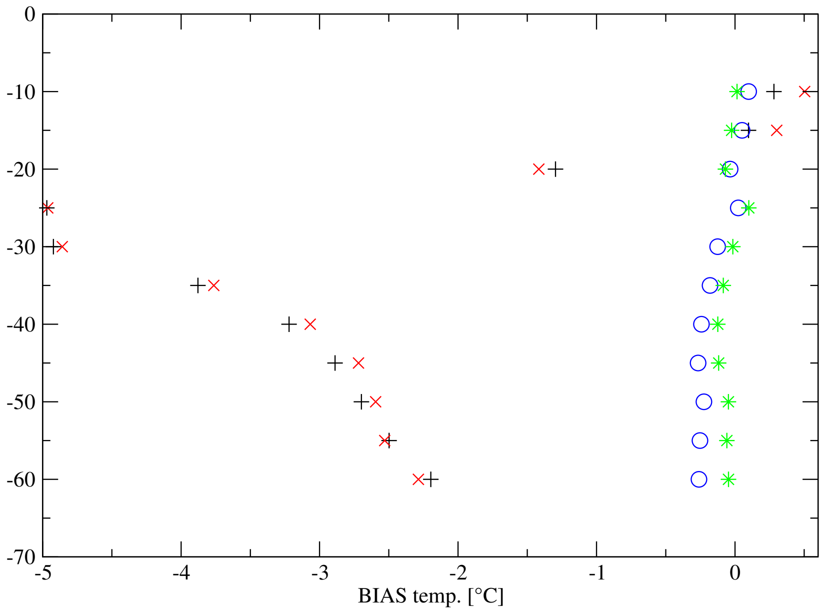

Figure 4.18: Bias between the modelled temperature and observed temperature. Four cases of modelled temperature are considered: without data assimilation for a non-vertically correlated 'CS-12' (black plusses), a vertically correlated 'CS-1' (red crosses) model error and with data assimilation for a non-vertically correlated 'CS-13' (blue circles) and a vertically correlated 'CS-11' (green stars) model error.

the temperatures are under-estimated.

Conclusion concerning the vertical correlation in the representation of the model error

All quantitative assessments of the effect of a vertical correlation in the error representation of the temperature indicate that there is a positive impact of a vertically correlated model error. However, the adequacy of such a method for a stratified temperature profile is not expected a priori. One possible reason for the benefit of a vertically correlated model error representation at a stratified station is that a vertical correlation scale of 10 metres depth is small enough not to induce unrealistic correlations between for example the upper mixed layer temperature and the thermocline.

4.3.2 Modelled temperature

This section examines the modelled temperature profiles when data are assimilated with the low rank square root algorithm allowing an ensemble representation of the observations' errors.

The numerical experiment is performed with a vertically correlated model error and the same ensemble parameters as the reference simulation: 50 ensemble members, sampling and assimilation time steps of one day, a vertical standard deviation of 1°C, and a vertical decorrelation length of 10 metres. Time-depth series of the August temperature corresponding to the experiment referred to as 'CS-11' are presented on Figure 4.19. Model outputs are plotted every three hours.

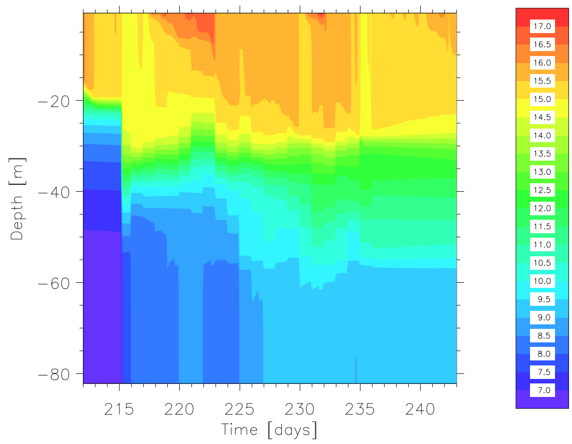

Figure 4.19: Time-depth series of temperature modelled with data assimilation using the low rank square root filter, experiment 'CS-11'.

The ensemble parameters being the same for experiments 'CS-2' and 'CS-11', in the absence of assimilation, the temperature profile modelled without assimilation with the low rank square root algorithm is the same as with the square root algorithm (free run 'CS-1)'. Therefore the observations features that are missed by the model are the same: the main one being a thermocline not correctly located (at about 20 metres depth compared to 30 metres). This wrong position of the thermocline results in an over-estimation of the temperature above the thermocline (too warm) and in an under-estimated temperature below it (too cold).

Figure 4.19 indicates that the assimilation process improves the representation of the temperature profile with the position of the thermocline shifted from 20 metres to 30 metres depth.

A quantitative assessment of the effect of the assimilation process is provided by comparison of the root mean square error between respectively the non assimilative run and the assimilative run and the observations. Results are presented on Figure 4.20. The average is performed in time. For the left panel the rms is computed using midnight values (which is the time of assimilation) while for the right panel, it is computed using values 23 hours later, one hour before the next analysis step.

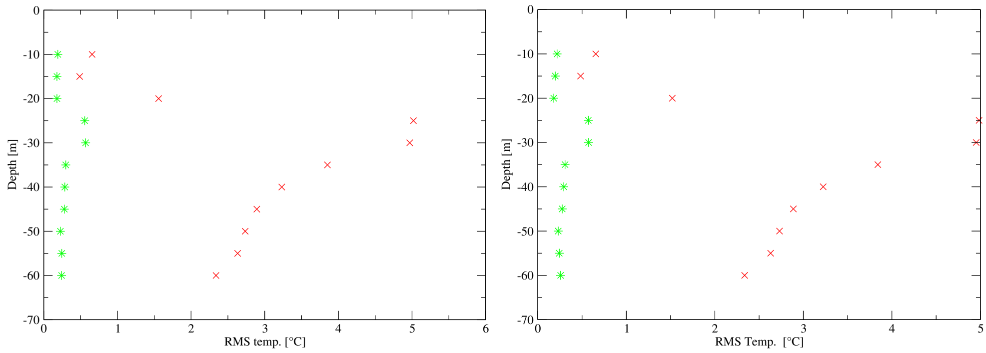

Figure 4.20: Root mean square error between the observed and modelled temperature without assimilation (red crosses) and with assimilation (green stars) at midnight (left) and at 23.00 (right).

The profiles of the root mean square errors of the non assimilative run (red crosses) show exactly the same features and have roughly the same values at midnight and twenty-three hours later. The rms errors of the assimilative run (green stars) are roughly the same at midnight and twenty-three hours later. This confirms what was already mentioned from the representation of the September temperature i.e. that when model features are corrected by the analysis, the model integrates them and these corrections are maintained for a certain period of time, even in the absence of assimilation.

4.3.3 Effect of the temperature assimilation on the stratified formulation of the surface heat exchange coefficients

In this section we examine the effect of the assimilation of temperature data on the surface heat exchange coefficients in the case of the Kondo stratified formulation which is dependent on the sea-air temperature difference. The heat exchange coefficients are not adjusted but are simply updated by the surface temperature improved representation resulting from the temperature analysis.

The sea-air temperature difference is represented on Figure 4.21. It is limited to day 236 (25th of August) as no more data are available for assimilation after that day.

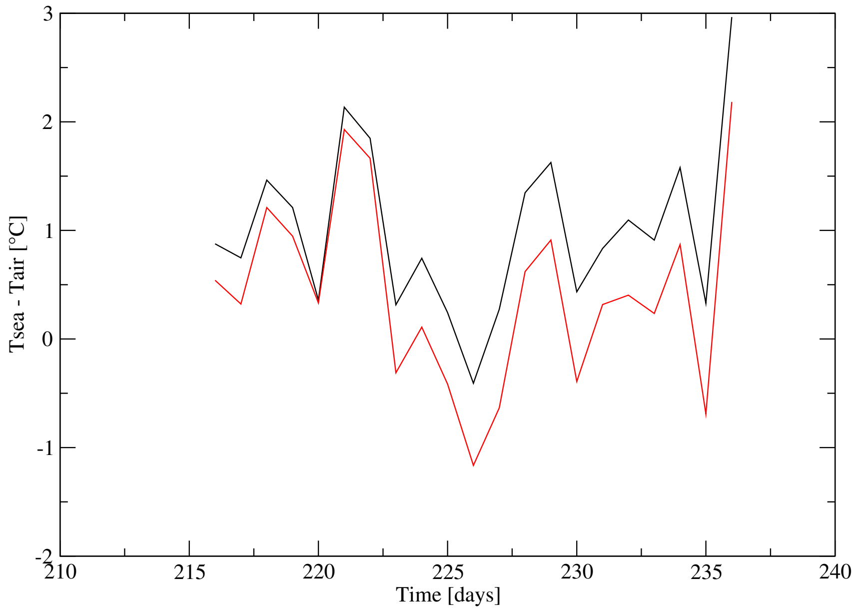

Figure 4.21: Time series of the sea air temperature difference without (black) and with (red) temperature data assimilation.

In the absence of assimilation, this difference is mainly positive (black line). The assimilation process provides a colder sea surface temperature so that it reduces the sea-air temperature difference and induces a shift of the system from unstable conditions ($T_{sea} > T_{air}$) to stable conditions ($T_{sea} < T_{air}$) at three periods:

- from day 222 (11th of August) until day 227 (16th of August),
- around day 230 (19th of August),
- around day 235 (24th of August).

In order to highlight the effects of the temperature assimilation, the differences between the stratified coefficients computed without and with temperature assimilation have been calculated. Results are presented on Figure 4.22, the left panel corresponds to the latent heat coefficients while the right panel is for the sensible heat coefficients. The vertical scale indicates that the differences correspond to ± 10 % of the default mean values of the coefficients. Most of the values are equal to zero. When non equal to zero the differences are mainly positive except markedly around day 230 (19th of August).

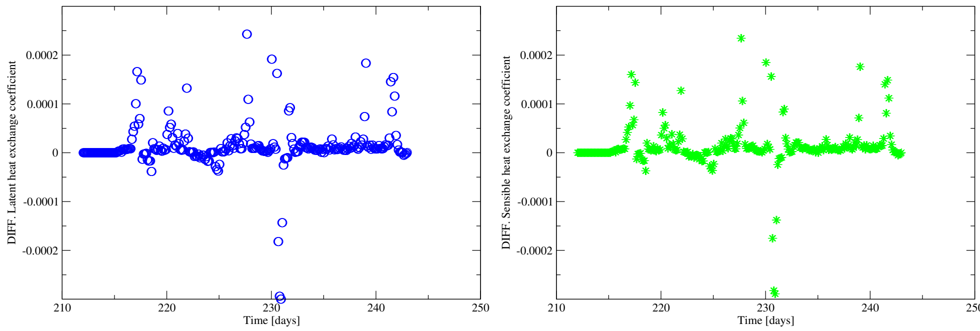

Figure 4.22: **Left:** Time series of the difference between the latent heat exchange coefficients computed without 'CS-1' and with temperature assimilation 'CS-11' (blue circles). **Right:** Time series of the difference between the sensible heat exchange coefficients computed without 'CS-1' and with temperature assimilation 'CS-11' (green stars).

The only negative difference between the values of the coefficients without and with data assimilation takes place around day 230 (19th of August) capturing the only increase of the heat exchange coefficients when data are assimilated. This occurs when the assimilation diminishes the sea surface temperature up to a level where it becomes colder than the air temperature.

When interpreting these results, it must be kept in mind that the sea-air temperature difference has a large influence on the surface heat exchange coefficients in small wind speed conditions (Figure 3.2).

4.4 Comparison of the results obtained without and with considering observations' errors

When assessing the efficiency of the assimilation process, the model results appear better if they are closer to the assimilated data. However, the assimilation scheme should not just interpolate the model closer to the data but rather smooth them according to the prescribed model error statistics [De Mey et al.] [22].

With respect to the results of the reference simulation 'CS-2' (right panel of Figure 4.3) where observations are considered as perfect, the temperature profile obtained with an ensemble representation of the observations' error 'CS-11' (Figure 4.19) are further away from the observations. However, it does not mean that the square root algorithm performs better than the low rank square root algorithm; it is just attributing a larger weight to the observations.

Depending on the quality of the observations, the most adequate of these two algorithms has to be chosen. Moreover, the low rank square root algorithm is easier to run as the constraint put on the model at the analysis step is reduced when errors on the observations are taken into account.

4.5 Conclusions for the model error sampling and state estimation in the 1-D case

Temperature profiles from the 'North Sea project' data base have been assimilated at the North Sea CS station from the 5th of August till the 24th of August. One of the motivations for these 1-D simulations is that their low cost in computing time allows to perform many numerical experiments to determine the best configuration and parameters for model error sampling.

The overall quality of the model error representation has been assessed by comparison of the root mean square of the innovations to the standard deviation of the ensemble and from a series of numerical experiments. We will retain that:

- an ensemble of 50 ensemble members provides good results (augmenting the number of members does not significantly improves the results but increases the computing time),
- a vertical decorrelation length of 10 metres for the model error representation is adequate,
- a model error sampled with a vertical standard deviation of 1°C allows the best representation of the innovations.

This set of values for the model error sampling parameters provides a first estimate for the 3-D simulations. However, it will not be directly extrapolated to 3-D simulations as their physics is more complex and a similar assessment will be performed for the simulations of the whole North Sea domain.

A vertically correlated model error is more adequate than a non vertically correlated model error when a vertical decorrelation length of 10 metres is used, even in the case of a stratified water column. A possible explanation is that a vertical decorrelation length of 10 metres is small enough not to induce unrealistic correlations between physically different water layers.

The assimilation of data allows to represent observed features that were absent from the model representation. In particular the too high position of the thermocline and too warm surface temperature are corrected to values closer to the observations. As soon as the Kalman filtering step is applied the ensemble standard deviation of the ensemble drops. Provided data are assimilated with an adequate time interval, oscillations present in the observed thermocline can be partly captured by data assimilation.

Finally, we have examined if the corrected position of the thermocline remains in the modelled temperature once the assimilation process is stopped. A simulation of the September temperature has been performed with preliminary assimilation in August. Results indicate that the position of the thermocline in September remains at the same depth as in August. This is attributed to the fact that the dynamical processes influencing the thermocline position are missing in this 1-D vertical test case.

Chapter 5

Model error sampling and state estimation with the ensemble Kalman filter for the 3-D case: the North Sea

This chapter deals with model state estimation for the more realistic 3-D simulations. The domain under study is the North Sea area comprised between 48° and 60° North in latitude and −3° and +9° in longitude. Temperature profiles generated from simulations with a higher horizontal resolution are assimilated. The ensemble generation starts on the 1st of September and data are assimilated from the 12th of September until the 28th of September 2001. A detailed description of the simulations is given in section 4 of chapter 3.

This chapter comprises five sections. In the first one, the gap between the model and the observations that will be assimilated is examined. On the basis of that primary information the second section deals with the choice of the parameters for the ensemble generation. In the next section results obtained with the square root algorithm (data are considered as perfect) are examined. The assimilative runs are considered versus the non assimilative runs and a quantitative analysis against the assimilated data is provided. The issues related to the choice of number of ensemble members and assimilation radius are discussed in detail. The fourth section is dedicated to the low rank square root algorithm where an ensemble representation of the observations' error is taken into account. This corresponds to a more realistic representation of the system. Finally the model results obtained with the two algorithms are compared to an independent set of observations available at three well-mixed stations. Conclusions are drawn in the fifth section. The simulations and their features are summarized in Table 8.6 in Appendix E.

The model outputs presented in this chapter are available along the 20 σ-levels of the model grid. However, results showing the whole domain under study focus mostly on the surface and bottom temperature at the end of the assimilation period when the maximum information available from the data has been incorporated. Time-depth series of temperature are also considered, usually for two stations, one located in the German Bight (referred to as station 10) and representing the well-mixed regime and the other one located in the summer stratified central part of the North Sea (referred to as station 19).

5.1 Model error features of the 3-D case

The model error is represented by the standard deviation of an ensemble of representations of the model. When generating these ensemble members, different parameters can be adjusted to specify as well as possible the model error. They are the standard deviation of the ensemble, vertical and horizontal decorrelation lengths and the number of ensemble members.

The choice of the values for the above mentioned ensemble parameters is governed by the discrepancy between the non assimilative model results and the data used for assimilation. The data are generated from model runs with a horizontal resolution of one nautical mile. To examine the model error of the 3-D simulations, we have calculated the root mean square difference and the bias between the temperature field modelled without data assimilation and the data at the two stations representative of the two North Sea regimes, well mixed (station 10) and stratified (station 19). The error means are computed in time.

Results at station 10 are shown on Figure 5.1, the root mean square error is presented on the left panel while the bias is on the right panel. At this stage, we focus on the difference between the non assimilative runs and the observations represented by red crosses. The bias is positive and is roughly constant along the vertical with a value around 0.25°C. It indicates a temperature profile over-estimated by the model.

Figure 5.2 presents the results at the stratified station. The error between the model and the data at station 19 contrasts with that of station 10. At the surface, there is a very good agreement between the model results without assimilation and the data that will be assimilated. The model bias indicates that the modelled temperature is under-estimated. The differences vary by a factor of four from the surface to the bottom. The temperature above the thermocline is under-estimated by about 0.15°C while below the thermocline

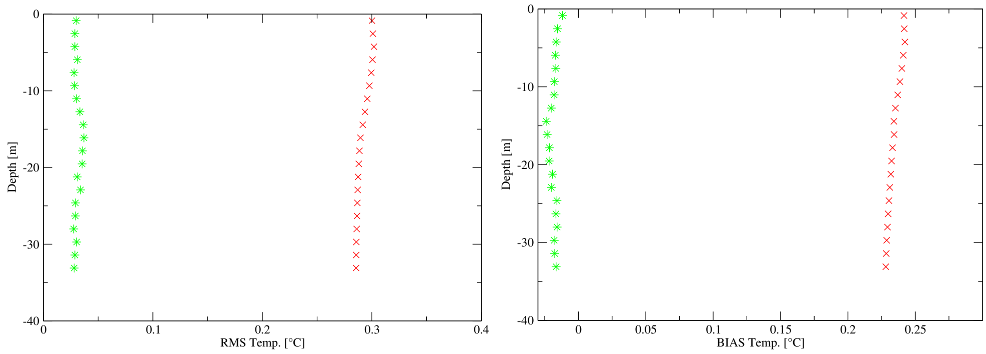

Figure 5.1: Root mean square error (left) and bias (right) between the model and the data, without (red crosses) and with (green stars) data assimilation at a well-mixed station (station 10).

the under-estimation reaches 0.6°C. The largest discrepancies between the model an the data are located around the thermocline with a maximum root mean square error of 1.5°C. They correspond to differences between the observed and modelled positions of the thermocline.

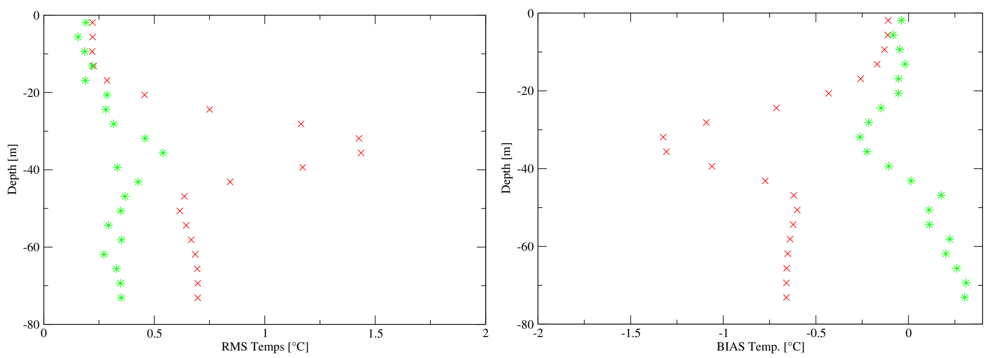

Figure 5.2: Root mean square error (left) and bias (right) between the model and the data, without (red crosses) and with (green stars) data assimilation at a summer stratified station (station 19).

These features illustrate the challenges related to the assimilation of data in coastal zones where the processes dominating one region can be totally different from those in a neighboring one. In the present study, the patterns of the temperature differences between the model and the data that will be assimilated at the well-mixed station are very different from those of the stratified zone.

	RMS [$^\circ C$]		BIAS [$^\circ C$]	
Station	Sfce	Bot.	Sfce	Bot.
Stat. 6	0.31	0.32	-0.25	-0.27
Stat. 7	0.20	0.19	-0.11	-0.12
Stat. 9	0.24	0.23	0.19	0.19
Stat. 10	0.30	0.29	0.24	0.23
Stat. 11	0.51	0.27	-0.09	0.02
Stat. 13	0.21	0.21	-0.04	-0.13
Stat. 15	0.31	0.86	0.28	0.42
Stat. 19	0.22	0.70	-0.10	-0.66

Table 5.1: Root mean square error and model bias between the model without data assimilation and the data that will be assimilated, at the surface and at the bottom.

The root mean square error and the model bias have been computed at the eight assimilation stations. Surface and bottom values are presented in Table 5.1.

The root mean square errors for the well mixed stations are of the same order of magnitude. At the surface they range from 0.2°C to 0.5°C. These errors are larger at the surface than at the bottom, except at stations 15 and 19 where they are much larger at the bottom than at the surface. At station 15, it is larger by nearly a factor of two while at station 19 it is by a factor of three.

At stations 6 and 7, in the Southern Bight, the bias is negative at the surface and at the bottom, indicating that the model under-estimates the temperature. In the German Bight, at stations 9, 10 and 15 the temperature is over-estimated at the surface as well as at the bottom while at station 11, the bias is negative at the surface (-0.09°C) and positive at the bottom (0.02°C).

The data that will be assimilated are generated from a model run with a one nautical mile horizontal resolution. In the Southern Bight, the negative bias indicates that when run with a lower horizontal resolution (four nautical miles) the model under-estimates the temperature.

In the German Bight, there are stations at which the bias is positive and other stations where it is negative, its values can have different signs at the surface and at the bottom. One possible explanation is that it could be related to slightly different positions of the temperature fronts depending on the horizontal resolution of the model.

5.2 Parameters for model error sampling

Critical issues for data assimilation in coastal areas are related to the wide range of characteristic lengths of the processes. The parameters for model error sampling (the decorrelation scales and the standard deviation of the ensemble) are defined for the whole domain on which the EnKF is applied. They are chosen by comparison of the model forecasts without data assimilation to the temperature profiles that will be assimilated.

5.2.1 Vertical decorrelation length

When generating the ensemble members, it is possible to introduce a vertical correlation between the error terms. This is done following equation (2.28). The total perturbation added to one vertical level consists of a fraction of the perturbation added just below and is augmented by the complementary fraction of random perturbation. The vertical decorrelation length corresponds to the vertical distance at which the error is damped by a factor of e.

The correlation rate is not uniform. The 3-D simulations are performed with 20 vertical σ levels. If we consider a well mixed station with a water depth of 20 metres, Δz is of about 1 meter so that $\alpha_v = 0.9$. In this case the proportion of random terms is small, the vertical error at one level comes at 90% from the level below and the vertical error tends to be uniform along the vertical. On the other hand, in the deepest parts of the domain, if we consider a station with a water depth of 80 metres, Δz is of about 4 metres so that $\alpha_v = 0.6$. In these regions, the contribution of random terms to the vertical model error is in the same proportion as the contribution of the error term from the level below.

In the case of a well-mixed regime, the diffusion processes propagate errors from the surface heat fluxes along the whole water column so that the root mean square errors between the model without assimilation and the data take the same values at the surface and at the bottom (except at stations 11 and 15). For the summer stratified station 19, the errors at the surface are related to the errors in the estimation of heat fluxes and turbulent mixing due to the wind while the bottom error is mainly due to the tidal turbulent mixing.

Figure 5.3 shows the data that will be assimilated at stations 11 (left panel) and 15 (right panel). Both stations are located in a well mixed area (German Bight), hence for these stations a vertical decorrelation length equal to the water depth could even be considered.

At station 11 (left panel of Figure 5.3), around day 255 (the 12th of September 2001), an unstable stratification of the temperature is present in

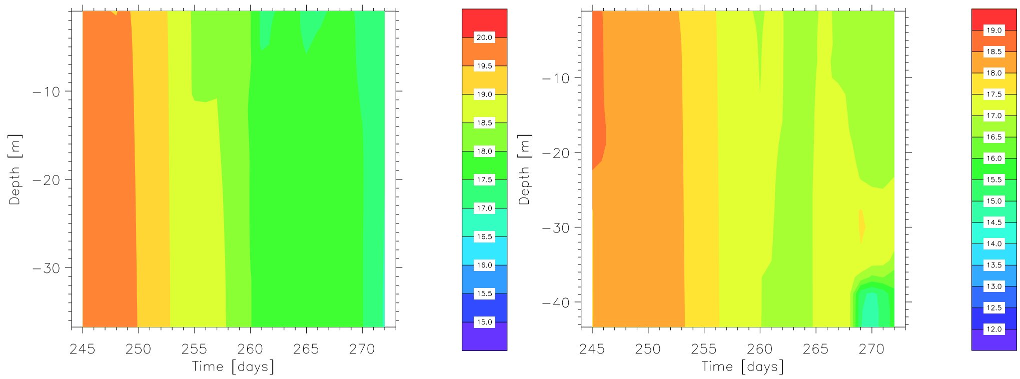

Figure 5.3: Time-depth series of the data that will be assimilated at station 11 (left) and at station 15 (right).

the first 10 metres of the water column. It could be due to the influence of salinity in the German Bight because of a variation of the freshwater input of the river Elbe.

The temperature profile that will be assimilated at station 15 (right panel of Figure 5.3) indicates that, at the beginning of September, the water column is well mixed. Though, at the end of the period, a feature of unstable stratification of the temperature develops until the end of the month. This feature is located between 25 and 35 metres depth and covers roughly 10 metres. Convection is not represented in the COHERENS model so that it is integrated in the water column by the turbulent mixing due to the wind and tides.

The implementation of the ensemble Kalman filter in the COHERENS model is based on a uniform vertical correlation length, so its determination has to take into account features of well mixed as well as of stratified stations. In the summer stratified part of the North Sea, the thermocline is located around 30 metres depth. The small unstable temperature stratifications present in profiles of well-mixed zones have characteristic lengths of about 10 metres. Therefore a vertical decorrelation length of 10 metres is a good candidate as a common value for the eight stations under consideration.

Further confirmation of this value of the vertical decorrelation length has been sought by examining the temporal evolution of the temperature profiles at station 10 and at station 19. The temperature profiles are highlighted at three time steps: on the 1st of September which is the first day of simulation, on the 12th of September corresponding to the first day of assimilation and finally on the 28th of September, the last day of assimilation.

The temperature profiles at these three time steps are shown on Figure

5.4 at stations 10 (left panel) and 19 (right panel).

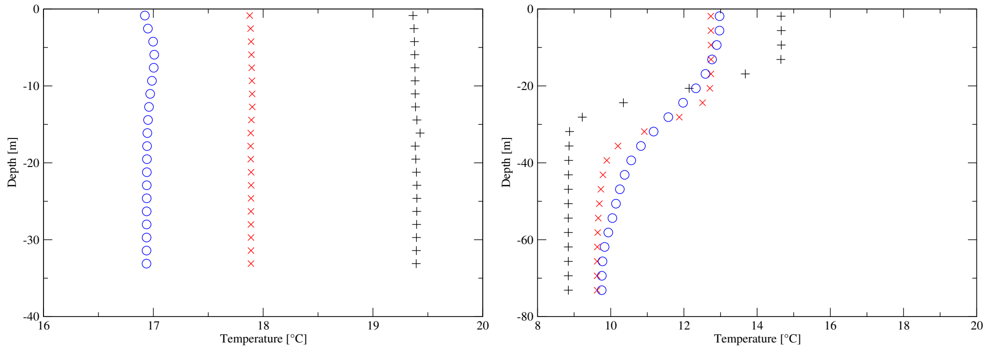

Figure 5.4: Temperature profiles at station 10 (left) and at station 19 (right) on the 1st of September (black plusses), the 12th of September (red crosses) and the 28th of September (blue circles).

At station 10 (left panel), the temperature clearly cools as time passes, starting from 19.5°C (black crosses) on the 1st of September and decreasing to 17°C on the 28th of September (blue circles). The temperature stratification at station 19 (right panel) evolves in time with a surface temperature cooling from 14°C to 12°C, and a bottom temperature which warms from 8°C to 10°C. The sharpness of the position of the thermocline varies according to the temporal evolution of the temperature profile.

5.2.2 Horizontal decorrelation length

For a given vertical level, between two fronts, the horizontal temperature at one point is correlated to a certain extent to that of its neighbors. Hence the error on the modelled temperature at one point is also correlated to that of its neighboring points. Such correlation is introduced between the ensemble members by using a horizontal correlation length. The horizontal correlation between the perturbation's terms is described by equation (2.23). The horizontal decorrelation length corresponds to the damping by a factor of e of the amplitude of this perturbation.

In stratified regions, the main variation of the bottom temperature is seasonal, so it is roughly constant in summer; on the other hand, the surface temperature is influenced by solar heating and by the wind. So, the horizontal decorrelation length has been determined by examining the modelled surface temperature at stations located close together. It is the case for stations 6 and 7 in the North Sea Southern Bight and for stations 9, 10 and 11 in the German Bight.

The left panel of Figure 5.5 presents time series of the surface temperature at stations 6 (black plusses) and 7 (red crosses) while its right panel deals with the surface temperature at stations 9 (blue circles), 10 (black plusses) and 11 (red crosses).

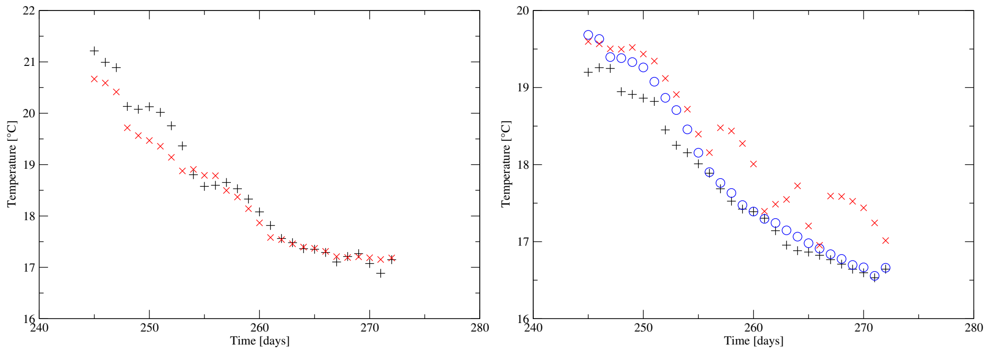

Figure 5.5: Time series of the surface temperature - **Left panel:** stations 6 (black plusses) and 7 (red crosses) - **Right panel:** stations 9 (blue circles), 10 (black plusses) and 11 (red crosses).

The distance between stations 6 and 7 is of about 7 km. The correlation between the surface temperature at those stations is quite strong, especially for the second half of the September month with superimposed temperature values.

Stations 9 and 10 are distant by 11 km in latitude and by 25 km in longitude. As in the case of stations 6 and 7, the temperature values at station 9 (blue circles) are superimposed to values at station 10 (black plusses) for the second half of the September month. These temperature values may differ by up to 0.5°C of those of station 11 which is located at 122 km from station 10. Moreover, the root mean square error between the temperature modelled without assimilation and the data at station 9 is very close to that of station 10. The surface values of the root mean square errors are respectively 0.24°C and 0.3°C while the bottom values are 0.23°C and 0.29°C.

To sum up, the surface temperature at stations distant by 10 km show a strong correlation, taking identical values in some cases. However, for stations located further away, such a correlation is no more valid. Hence a horizontal decorrelation length of 10 km has been chosen as it corresponds to the minimal horizontal correlation length for areas free of temperature fronts.

5.2.3 Standard deviation of the temperature ensemble

The determination of the multiplicative factors for the vertical standard deviation of the ensemble, η_T, and for its horizontal standard deviation, γ_T, is based on the estimated gap between the temperature obtained without assimilation and the temperature data computed with a higher horizontal spatial resolution.

Figure 5.1 presents the root mean square difference (left) and bias (right) between the temperature modelled without data assimilation and the data that will be assimilated (red crosses) at station 10. They are roughly constant along the vertical and positive. The root mean square difference is of 0.3°C.

At station 19, the root mean square difference and bias between the model without assimilation and the data are shown on Figure 5.2 (red crosses). At the surface, the root mean square difference is of 0.15°C while below the thermocline it is of the order of 0.6°C; at the thermocline it is around 1.5°C.

At the summer stratified station 19, the root mean square difference varies by one order of magnitude between different zones of the water column. Such a feature of the model error illustrates the difficulty to find a set of common parameters for model error sampling which is adequate for the whole North Sea domain and for the time period under consideration.

The assessment of the differences between the temperature profiles modelled without data assimilation and the temperature data that will be assimilated indicates root mean square discrepancies of up to 1.5°C. In a primary step, a vertical standard deviation η_T of 1°C combined with a horizontal standard deviation γ_T of 1°C is chosen. The ensemble standard deviations resulting from these values will be further examined.

5.3 Considering perfect observations: the square root algorithm

In this section, the results obtained by assimilation with the square root algorithm are examined in detail. Assimilated data are considered as perfect, this allows to assess the assimilation process when its effect on the model results is maximum.

We will first study the effect of the assimilation process by comparing the surface and bottom temperature obtained with and without data assimilation as well as time-depth series of temperature at stations 10 and 19. A quantitative assessment will also be provided through the mean model errors (bias) and root mean square errors between the assimilative runs and the assimilated data. Then, the standard deviation of the temperature ensemble without data assimilation will be compared to its standard deviation with data assimilation.

A particular issue related to 3-D data assimilation in coastal regions is the influence radius of the assimilated data. In this study we use a cut-off assimilation radius. The German Bight contains several stations where data are assimilated, the surface temperature at these locations is examined in detail. In order to determine the assimilation radius and the number of ensemble members, the root mean square error between the assimilative runs and the assimilated data has been computed using 30, 40 and 50 ensemble members with assimilation radii of 30km, 50km and 70km.

5.3.1 Comparison of temperature fields modelled without and with data assimilation

Results presented in this section correspond to experiments 'NS-1' and 'NS-2'. They are respectively referred to as the free run, performed without data assimilation, and the reference simulation, performed with the square root algorithm in which the error covariance matrix of the observations is set to zero. Both experiments use the same parameters for model error sampling. The vertical and horizontal standard deviations are of 1°C. The vertical correlation length is of 10m while the horizontal correlation scale is of 10km. 50 ensemble are used and the cut-off assimilation radius is of 50km.

The assimilated data are generated from model run with a higher horizontal resolution of one nautical mile which is considered as the proxy ocean. In a next section, these results will be compared to those obtained with the low rank square root algorithm in which the observations' error covariance matrix is sampled through the use of an ensemble.

The discussion presented here aims at assessing how the ensemble Kalman filter allows to represent features missing in the modelled temperature but present in the data.

Surface and bottom horizontal fields

The horizontal temperature fields presented on Figures 5.6 and 5.8 correspond respectively to the surface and bottom temperatures at the end of the assimilation period (28th of September 2001). The left panels correspond to the temperature fields without data assimilation while the right panels concern the assimilative runs.

The surface temperature modelled without data assimilation shows three main zones:

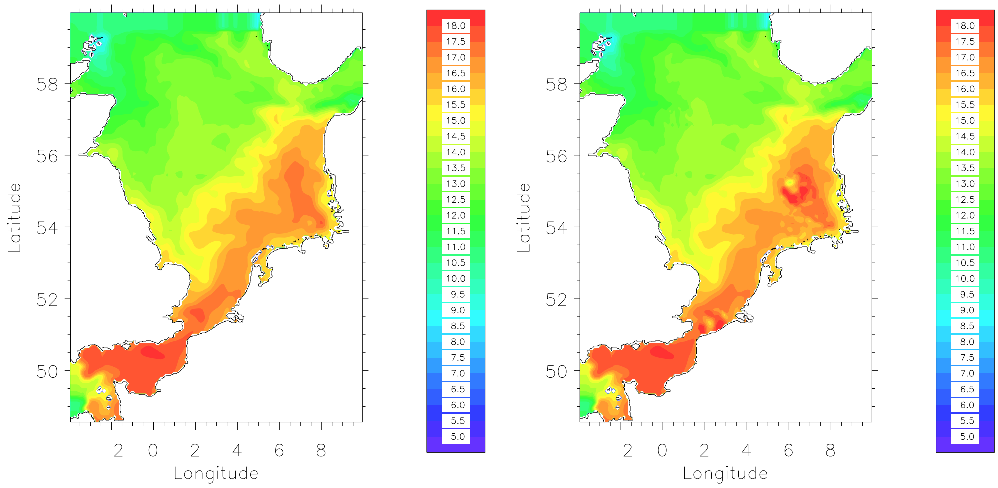

Figure 5.6: Surface temperature modelled without (left) and with (right) data assimilation on the last day of the assimilation period (September 28), respectively experiments 'NS-1' and 'NS-2'.

- a coastal region on the East side of the North Sea (English Channel, Southern Bight and German Bight) with a temperature varying from 16.5°C to 18°C,
- the North West part of the North Sea with a temperature of about 12°C (area in green),
- a transition zone in yellow, with a temperature of around 15°C.

The transition zone in yellow is the thinnest with a length of about 0.5 degrees in latitude. It also acts as a limiting factor in the determination of the maximum horizontal correlation length of the temperature field.

A comparison of the two panels of Figure 5.6 indicates that the effect of the assimilation process is to locally lower the surface temperature with a clear effect of the cut-off radius:

- in the Southern Bight (stations 6 and 7),
- in the German Bight (stations 9, 11, 13, 15),
- in the central part of the North Sea (station 19).

The model bias at stations 6 and 7 in the Southern bight of the North Sea is negative at the surface as well as at the bottom. Because of diffusion processes this was expected at stations in well mixed regime. The surface temperature at these stations is colder with assimilation (right panel) than without (left panel). The surface and bottom model biases at stations 9, 10 and 15 are positive, so that the effect of the assimilation process is a warming. At stations 11 and 13, the surface bias is negative so that the assimilation process decreases the surface temperature which is well marked by a green spot (right panel) at these stations. Concerning station 19, typical of a summer stratified regime, the bias without assimilation is negative, indicating that the assimilation process also induces a warming at that location.

Figure 5.7 provides a zoom on the surface temperature of the German Bight on the last day of the assimilation period. The left panel represents the surface temperature without data assimilation while the right panel the temperature with data assimilation. The circles have a radius of 50km and are centred on the stations at which data are assimilated.

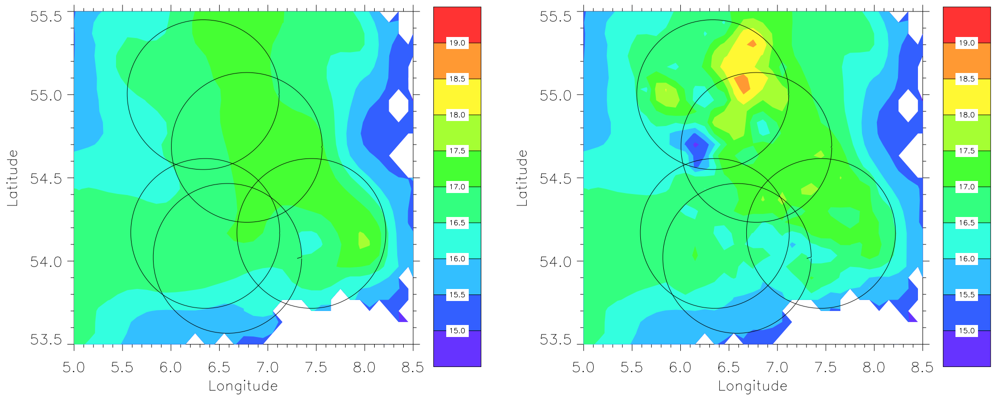

Figure 5.7: Surface temperature modelled without (left) and with (right) data assimilation in the German bight on the last day of the assimilation period (September 28), respectively experiments 'NS-1' and 'NS-2'.

The impact of the assimilation radius is clear. The assimilation process provides a surface temperature warmer in most cases, except at two locations: (54°N, 7°E) and (54.75°N, 6°E), where it is slightly colder.

The bottom temperature on the left side of Figure 5.8 is less homogeneous than the surface temperature, especially in the stratified part of the North Sea.

A comparison of the left and right panels of Figure 5.8 for the bottom temperature indicates that the data assimilation has roughly the same effect as at the surface i.e. to locally diminish the bottom temperature. These effects will be quantified further in this section.

At stations 6 and 7 the bottom bias is negative, the right panel of Figure 5.8 indicates that the bottom temperature modelled with assimilation is

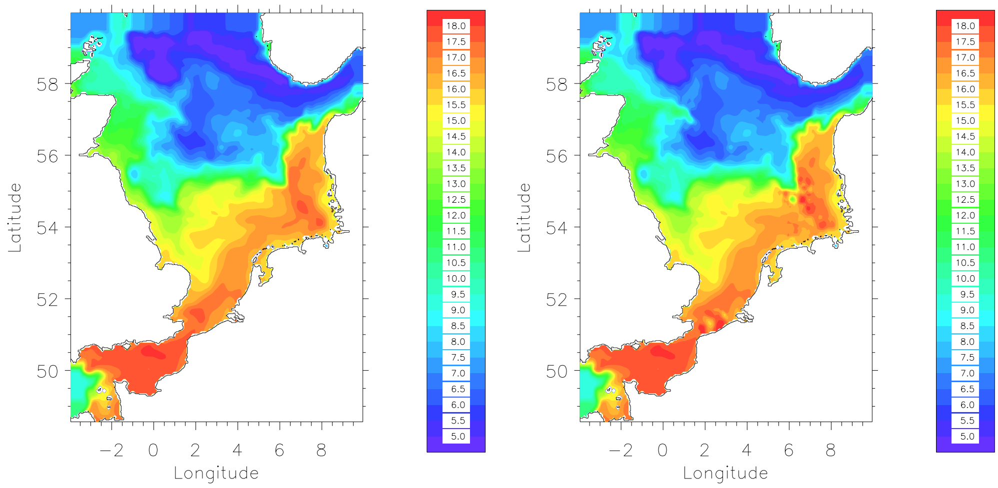

Figure 5.8: Bottom temperature modelled without (left) and with (right) data assimilation on the last day of the assimilation period (September 28), respectively experiments 'NS-1' and 'NS-2'.

warmer. In the German Bight the bias at the bottom of stations 9, 10, 11 and 15 is positive and the data assimilation induces a warming of the temperature. However, at station 13 it is negative. Station 11 is the only station at which the bias is positive at the surface and negative at the bottom.

Hovmöller diagrams of the temperature

Time-depth series of the temperature modelled without and with data assimilation are examined for a well-mixed and a summer stratified station (respectively stations 10 and 19). The time-series of the data that will be assimilated at stations 10 and 19 are shown on Figure 5.9. They start on the 2nd of September 2001 (day 245).

In order to identify the missing features that could be incorporated by the assimilation process we first compare the data that will be assimilated to the non assimilative runs from day 255 (12th of September) on, which is the starting day of the assimilation process and then we examine if they are incorporated in the assimilative runs.

Non assimilative and assimilative model results for the well-mixed station 10 are presented on Figure 5.10. The temperature profile simulated without data assimilation is very close to the assimilated data (left panel of Figure 5.9), the only difference between the assimilated data and the free run consists of the assimilated data being warmer around day 255 (12th of September) with a cooling starting on day 258 (15th of September) while in the free run the cooling starts on day 254 (11th of September). When data are

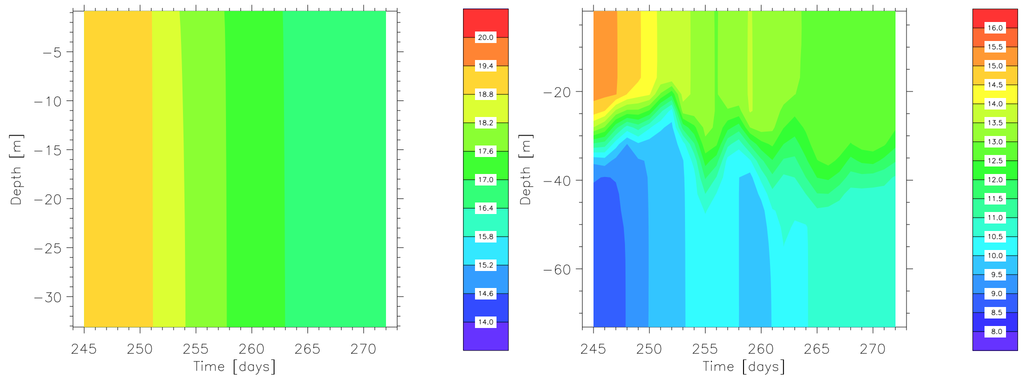

Figure 5.9: Time-depth series of the temperature data that will be assimilated at station 10 (left) and at station 19 (right).

assimilated, the modelled temperature shows that the cooling starts later on, on day 256 (13th of September).

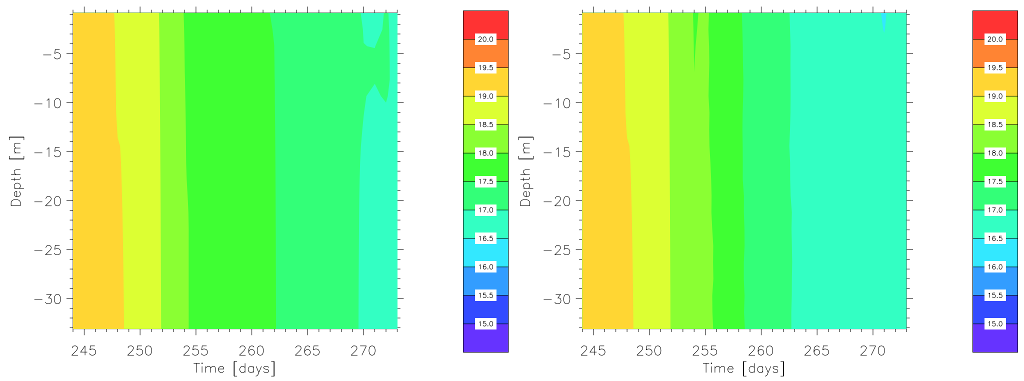

Figure 5.10: Time-depth series of the temperature modelled without (left) and with (right) data assimilation at station 10, respectively experiments 'NS-1' and 'NS-2'.

At station 10 the differences between the data and the free run are very small so that the assimilation process has not a strong impact on the modelled temperature. However, even if the differences are small the assimilation process is able to correct the model for minor missing features.

Figure 5.11 presents time-depth series of non assimilative and assimilative temperatures at stratified station 19. The assimilated data at station 19 (right panel of Figure 5.9) show a warm patch between days 254 (11th of September) and 264 (21st of September). Its extent is not well represented in the non assimilative run. After assimilation, around days 260 (17th of September) and 265 (22nd of September), the upper mixed layer is warmer than in the non assimilative run. However the patch is not homogeneous, contrary to the data. After assimilation, the temperature below the thermocline is closer to the observations, with a well captured warmer pattern after day 264 (21st of September) which was not present in the non assimilative run.

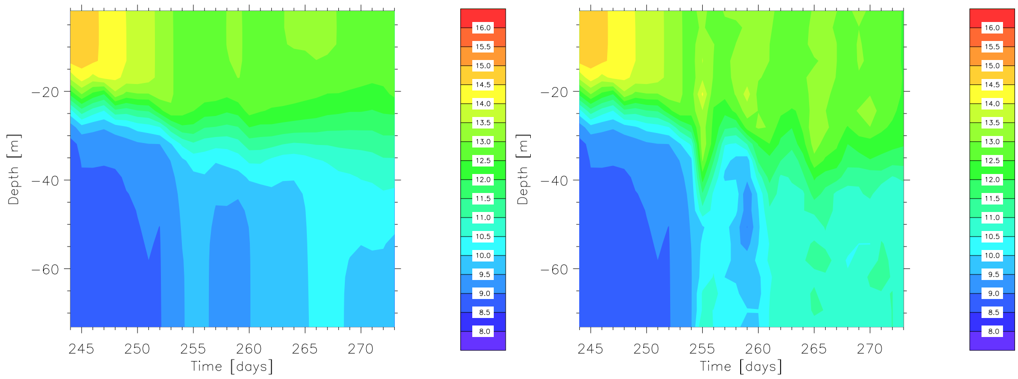

Figure 5.11: Time-depth series of the temperature modelled without (left) and with (right) data assimilation at station 19, respectively experiments 'NS-1' and 'NS-2'.

There are more missing features in the modelled temperature without assimilation at station 19 than at station 10. At station 19, the assimilation process captures globally well the data features that were not present in the non assimilative run.

5.3.2 Quantitative assessment of the impact of the assimilation process

The root mean square error and the bias between the assimilative runs and the assimilated data have been computed at each station, at the bottom and at the surface. The mean is computed in time. Results are presented in Table 5.2.

At the surface, the largest root mean square error appears at stations 13 and 15 with a value of 0.22°C while at the bottom, the maximum error is at station 19, with a value of 0.35°C. The smallest root mean square error appears at station 7 with surface and bottom errors of 0.03°C. In most cases the model bias has the same sign at the surface and at the bottom, except at stations 6, 11 and 19.

	RMS [$^\circ C$]		BIAS [$^\circ C$]	
Station	Sfce	Bot.	Sfce	Bot.
Stat. 6	0.06	0.07	0.04	-0.04
Stat. 7	0.03	0.03	0.00	-0.01
Stat. 9	0.06	0.05	0.03	0.01
Stat. 10	0.03	0.03	-0.01	-0.02
Stat. 11	0.05	0.03	0.01	-0.01
Stat. 13	0.22	0.24	0.07	0.13
Stat. 15	0.22	0.19	0.00	-0.04
Stat. 19	0.19	0.35	-0.04	0.30

Table 5.2: Root mean square error and model bias between the model with data assimilation and the assimilated data, at the surface and at the bottom.

A comparison with the equivalent results for the non assimilative runs in Table 5.1 indicates that the minimum root mean square difference in both cases appears at the same station, station 7. Without assimilation, the maximum surface root mean square error appeared at station 11. After assimilation, it is reduced by an order of magnitude; the bottom value at station 11 is also reduced by an order of magnitude. At station 19, the assimilation process reduces the root mean square error by a factor of two. Finally, station 13 is the only station at which the root mean square error of the assimilative runs is slightly larger than without assimilation.

At well mixed stations, the assimilation process reduces the root mean square error between the model and the assimilated data by up to one order of magnitude. A strong decrease in terms of values of the root mean square errors can even be observed at stations were the gap between model and data was already small before assimilation. In the stratified area, data assimilation reduces the error between the model and the data by a factor of two.

5.3.3 Comparison of the standard deviation of the temperature ensemble without and with data assimilation

In order to highlight the reduction of standard deviation produced by the analysis step of the Kalman filter, we compare the standard deviation of the temperature ensemble modelled without and with data assimilation. The comparison is shown at the surface for the whole North Sea domain and on time-depth series at stations 10 and 19.

Standard deviation of the surface temperature

The standard deviation of the surface temperature ensemble on the last day of the assimilation period is presented on Figure 5.12, without data assimilation on the left panel and with data assimilation on the right panel.

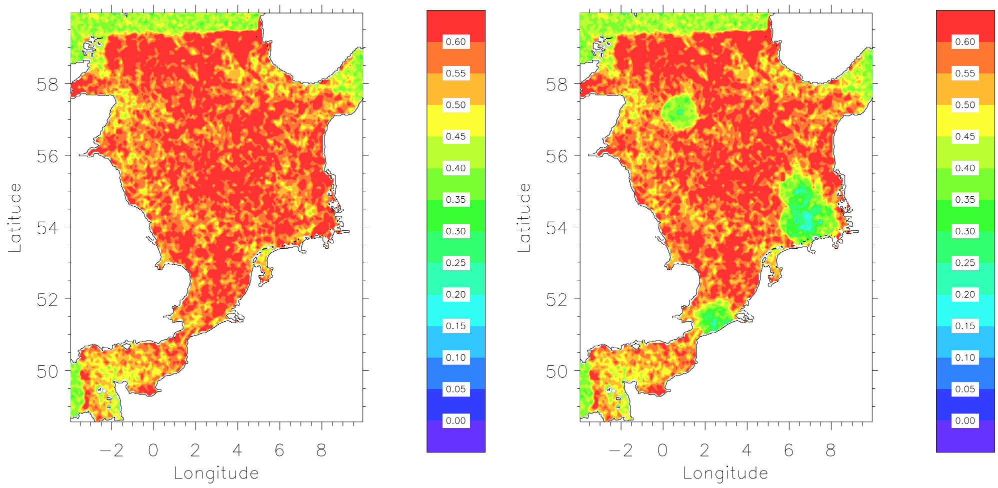

Figure 5.12: Standard deviation of the temperature ensemble modelled at the surface without (left) and with (right) data assimilation on the last day of the period, September 28, respectively experiments 'NS-1' and 'NS-2'.

At the surface, for the non assimilative runs, the standard deviation of the ensemble is of about 0.6°C. At the locations where an assimilation takes place, it is reduced to about 0.3°C.

There is a clear impact of the relaxation conditions applied at the open boundaries. The effect of a horizontal decorrelation length of 10 kilometres is also captured. The assimilation process reduces the standard deviation of the ensemble at a distance equal to the assumulation radius of 50 kilometres around the station at which the assimilation occurs:

- in the Southern Bight of the North Sea (stations 6 and 7),
- in the German Bight (stations 9, 10, 11, 13, 15),
- in the central part of the North Sea (station 19).

Hovmöller diagrams of the temperature standard deviation

Time-depth series of the ensemble standard deviation without and with data assimilation are presented for station 10 on Figure 5.13 and on Figure 5.14 for station 19.

A comparison of the left panel of Figure 5.13 to its right panel indicates a reduction of the ensemble spread from the first day of assimilation, day 255 (12th of September) until the end of the assimilation period.

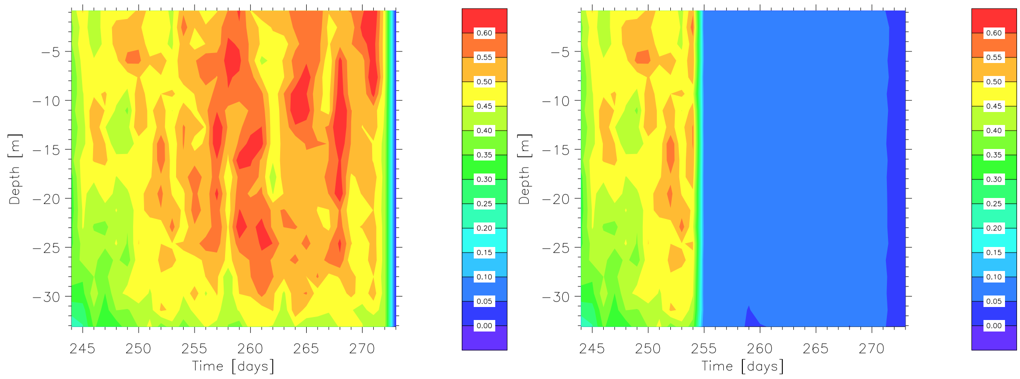

Figure 5.13: Time-depth series of the standard deviation of the temperature ensemble modelled without (left) and with (right) data assimilation at station 10, respectively experiments 'NS-1' and 'NS-2'.

At first sight, at station 19 (Figure 5.14), the time-depth series of the ensemble standard deviation show the same patterns as at the well-mixed station 10. However, the standard deviation at station 10 has a "vertically well mixed structure" while at station 19, a "stratified structure" is present. These features stem from the fact that the ensemble Kalman filter integrates in time each ensemble member with the fully non linear equations of the dynamical model.

At both stations, the ensemble standard deviation is slightly larger at the surface than at the bottom. When vertically correlating the error terms defining the ensemble members, the standard deviation of the ensemble is initially defined at the bottom and propagated along the vertical following equation (2.28). As a result the standard deviation is slightly larger at the surface than at the bottom.

Finally, at both stations, the standard deviation of the ensemble without assimilation increases as time passes. This feature is related to the way the ensemble is generated. Perturbations are added to the temperature deviation around the mean at each sampling time step.

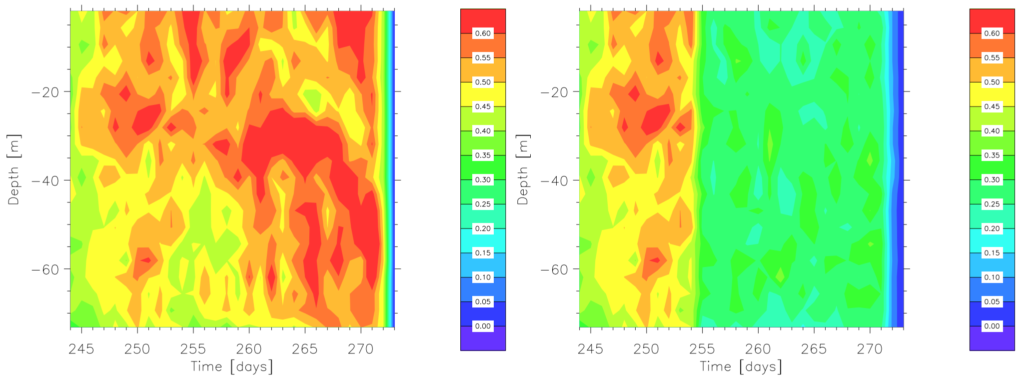

Figure 5.14: Time-depth series of the standard deviation of the temperature ensemble modelled without (left) and with (right) data assimilation at station 19, respectively experiments 'NS-1' and 'NS-2'.

5.3.4 Comparison of the innovations to the ensemble standard deviation

In order to check that the sampled model error is sufficiently large to incorporate the gap between the model and the data, the root mean square of the innovations can be compared to the mean of the standard deviation of the ensemble.

These two quantities are compared for the reference simulation 'NS-2' when data are assimilated. The means are computed in the vertical from the first day of assimilation, day 255 (12th of September), until the end of the assimilation period, day 272 (28th of September).

Results are presented on Figure 5.15 at the well mixed station 10 (left panel) and at the stratified station 19 (right panel). The standard deviation of the ensemble is quite different from station 10 to station 19. This is related to the reduction of the ensemble standard deviation by the assimilation process and corresponds to the differences between the right panels of Figures 5.13 and 5.14. The innovations are much larger at the well mixed station than at the stratified station. This stems from the discrepancies between the model and the data at the level of the thermocline of station 19. It also illustrates the challenges of data assimilation in coastal areas where a wide range of model-data discrepancies can be present from one area to another.

At station 10, the mean standard deviation is always larger than the root mean square of the innovations. On the other hand, at station 19, there are three days (day 255, day 257, and day 259) at the beginning of the assimilation period where the rms of the innovations is significantly larger than

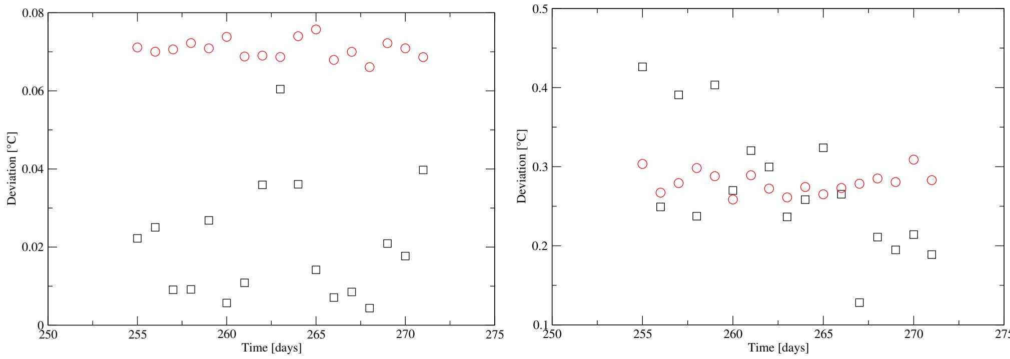

Figure 5.15: Time series of the root mean square of the innovations (black squares) and mean of the modelled standard deviation (red circles) at station 10 (left) and at station19 (right).

the mean of the standard deviation with values around 0.3°C for the mean of the standard deviation and reaching 0.4°C for the rms of the innovations. After a few assimilation steps, the temperature profiles are improved and these discrepancies are reduced. At the end of the assimilation period, after day 265 (22nd of September), the root mean square of the innovations is of the same order of magnitude as the mean of the standard deviation of the temperature ensemble.

At station 10, the standard deviation of the ensemble is sufficiently large to represent the gap between the model and the observations. At station 19, a few assimilation time steps are necessary to reduce the gap between the model and the data. When this is achieved, the standard deviation of the ensemble becomes of the same order of magnitude as the innovations. As a result, the multiplicative factors η_T and γ_T can be considered as adequate to represent the model error in both well mixed and stratified regimes.

5.3.5 Choice of the assimilation radius and number of ensemble members

As we work with only eight temperature profiles for assimilation in the North Sea, we have to maximize their effect by using the largest assimilation radius consistent with the physical features of the neighboring temperature field. In the framework of this study the assimilation of temperature profiles has an impact on the neighboring temperature in the limit of the assimilation cut-off radius which has been chosen on the basis of the horizontal features of the temperature field; the location of the temperature fronts being the main limiting factor to the size of the assimilation radius (Figure 5.8). Moreover, stations are overlapping in the German Bight.

Furthermore, as already mentioned, a key parameter for model error sampling is the number of ensemble members. In the case of the North Sea 3-D simulations, because of computing time considerations we can not afford more than 50 ensemble members (the model runs get rejected from the machine after 24 hours). So, the problem of finding the largest assimilation radius for a given number of ensemble members arises.

Numerical experiments have been performed with 30, 40 and 50 ensemble members in combination with assimilation radii of 30, 50 and 70 km. The choice is based on a qualitative and a quantitative assessment. In the qualitative assessment, we examine the features of the surface temperature in the German Bight area located between 55°N and 56°N in latitude and 5°E and 8°E in longitude and where five assimilation stations (9, 10, 11, 13, 15) are located and could overlap depending on the value of the assimilation radius. In the quantitative assessment we compute the root mean square errors between the assimilated data and the temperature modelled for different combinations of the assimilation radius and the number of ensemble members.

Qualitative assessment of the assimilation radius and number of ensemble members

We show the surface temperature in the German Bight for four limit cases involving the combination of the smallest number of ensemble members (30) with respectively the smallest assimilation radius (30 km) and the largest assimilation radius (70 km) and then, the combination of the largest number of ensemble members (50) with respectively the smallest (30 km) and the largest (70 km) assimilation radius.

The results obtained with 30 ensemble members are shown on Figure 5.16. They correspond respectively to experiments 'NS-3' and 'NS-5'.

A comparison of the two panels of Figure 5.16 clearly captures the difference of size of the assimilation radius. The combination of a small number of ensemble members with a large assimilation radius (right panel of Figure 5.16) generates patches of temperature much colder than their neighborhood.

Figure 5.17 represents the temperature modelled with 50 ensemble members and assimilation radii of 30 km (left) and 70 km (right)

The right panel of Figure 5.17 shows results obtained with an assimilation radius of 70 km and 50 ensemble members. Compared to the right panel of Figure 5.16, increasing the number of ensemble members reduces the patchiness and provides a more realistic surface temperature. On the left

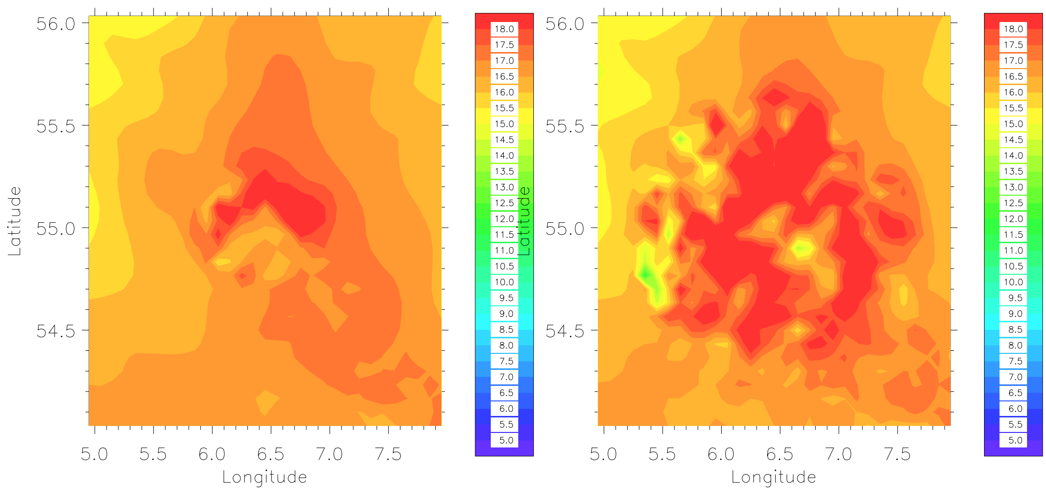

Figure 5.16: Zoom (German Bight) on the surface temperature modelled with 30 ensemble members and an assimilation radius of 30 km (left) and 70 km (right) on the last day of the assimilation period (September 28, 2001), respectively experiments 'NS-3' and 'NS-5'.

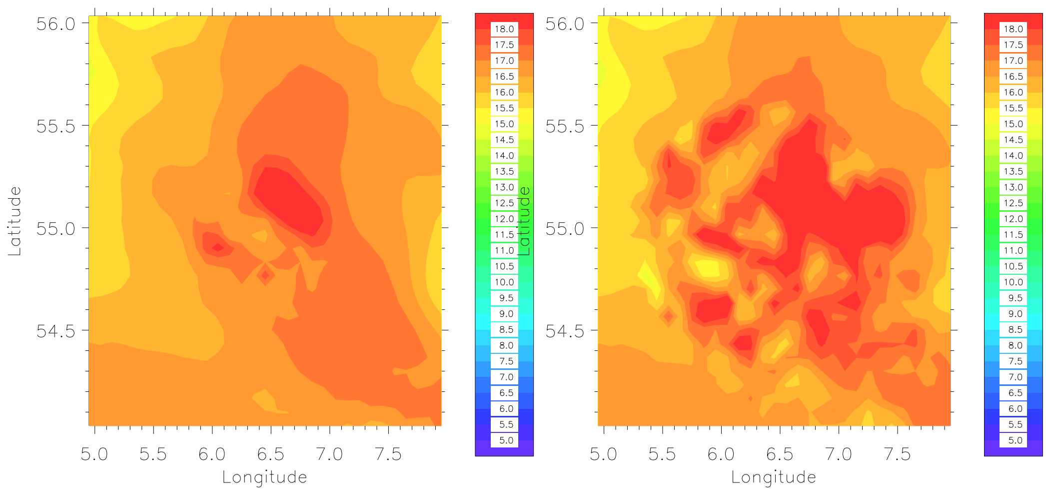

Figure 5.17: Zoom (German Bight) on the surface temperature modelled with 50 ensemble members and an assimilation radius of 30 km (left) and 70 km (right) on the last day of the assimilation period (September 28, 2001), respectively experiments 'NS-9' and 'NS-10'.

panel of the figure an assimilation radius of 30 km is used. In that case the information propagated from the assimilation process remains quite limited.

This first assessment of the surface temperature in the German Bight indicates that to avoid unrealistic patches in the temperature field 50 ensemble members are necessary. However, such undesirable features re-appear with the use of an assimilation of 70 km. An assimilation radius of 30 km limits the propagation of information from the assimilation process. Therefore, at first sight the combination of 50 ensemble members with an assimilation radius of 50 km is a good choice as reference for the simulations.

Quantitative assessment of the assimilation radius and number of ensemble members

To further examine an adequate combination of the number of ensemble members and of the assimilation radius, the root mean square error between the temperature modelled with assimilation and the assimilated data has been computed for two pairs of stations located close together. This is the case of stations 6 and 7 in the Southern Bight, they are respectively distant by 7.11 km in latitude; as well as of stations 9 and 10 in the German Bight, they are distant by 11.11 km in latitude and 25.45 km in longitude. This root mean square error has been computed for the nine possibilities of combinations of (30, 40, 50) ensemble members and of (30km, 50km, 70km) for the assimilation radius.

Figure 5.18 presents results for stations 6 and 7 that are located in the Southern Bight of the North Sea.

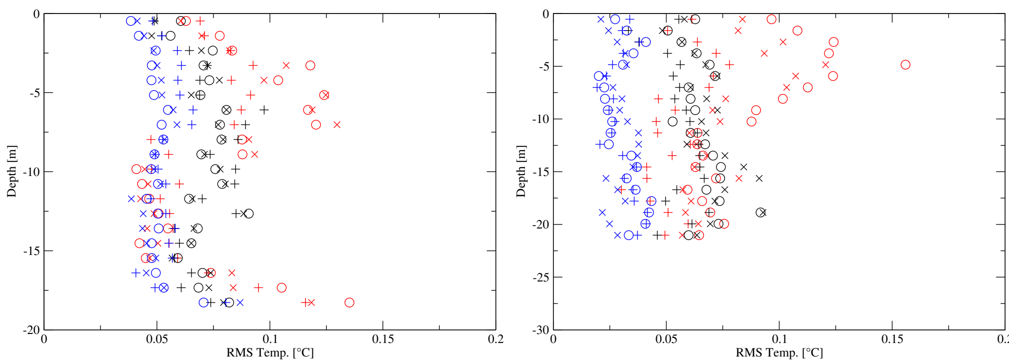

Figure 5.18: Vertical profile of the root mean square error between the temperature modelled with assimilation and the data at station 6 (left) and 7 (right). Results corresponding to 30 ensemble members are in red, to **40** ensemble members are in **black**, to 50 ensemble members are in blue. An assimilation radius of 30 km is represented by a (+), of 50 km by a (x) and of 70 km by a (◯).

A general feature of the results at stations 6 and 7 is that the spread of the root mean square differences along the water column is larger in the case of 30 ensemble members (red symbols), varying from 0.05°C to 0.15°C. At both stations, but more clearly at station 7, the root mean square errors are smaller with 50 ensemble members (blue symbols). However, surprisingly, at station 6, simulations with 30 members (red symbols) give results as good as those obtained with 50 members (blue symbols) between 10 and 15 metres depth.

At station 6, in the case of simulations with 50 ensemble members (blue symbols), results with an assimilation radius of 30 km (blue plusses) give at almost all depths the largest error with respect to assimilation radii of 50 km (blue crosses) or 70 km (blue circles).

At station 7, in the case of simulations with 50 ensemble members, results with an assimilation radius of 50 km (blue crosses) give the smallest errors between 0 and 5 metres depth and between 15 and 20 metres depth while an assimilation radius of 70 km (blue circles) performs better between 5 and 15 metres depth.

Concerning the German Bight, results for station 9 and 10 are presented on Figure 5.19.

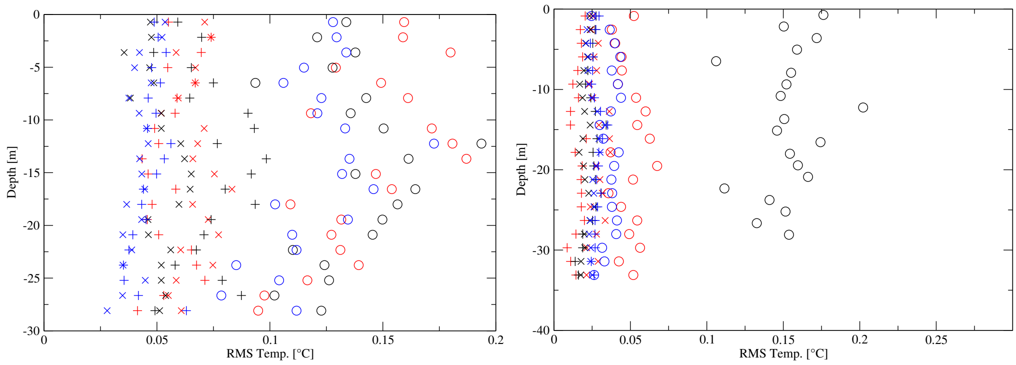

Figure 5.19: Vertical profile of the root mean square error between the temperature modelled with assimilation and the data at station 9 (left) and 10 (right). Results corresponding to 30 ensemble members are in red, to **40** ensemble members are in **black**, to 50 ensemble members are in blue. An assimilation radius of 30 km is represented by a (+), of 50 km by a (x) and of 70 km by a (◯).

At station 9 (left panel of Figure 5.19), the largest root mean square errors appear for an assimilation radius of 70 km (circles) in combination with 30 ensemble members above 15 metres depth (red circles) and with 40 ensemble members below 15 metres depth; while the smallest errors are obtained with an assimilation radius of 50 km combined with 50 ensemble members (blue crosses).

Surprisingly at station 10 (right panel of Figure 5.19), the largest spread and values in the root mean errors appear for simulations performed with 40 ensemble members and an assimilation radius of 70 km (black circles). It was expected that the root mean square error with 40 ensemble members be larger than with 30 members and smaller than 50 members. The smallest root mean square errors at all depths are provided by the simulations performed with 30 ensemble members combined with an assimilation radius of 30 km (red plusses). In particular, for simulations performed with 30 ensemble members (red symbols), there is a strong decrease of the root mean square error for the simulations performed with an assimilation radius of 30 kilometres (red plusses) in place of 70 kilometres (red circles).

To summarize the quantitative assessment reported in the previous paragraphs, there is not a combination of the assimilation radius with the number of ensemble members performing clearly better than the others at all stations. Features at stations located in the Southern Bight indicate that 30 ensemble members are too few. Moreover, simulations with 50 ensemble members combined with an assimilation radius of 30 km give sometimes large discrepancies. At these locations better performances are obtained from simulations with 50 ensemble members combined with assimilation radii of 50 km or 70 km.

The characteristics of the root mean square errors in the German Bight indicate that 40 ensemble members are not always adequate and that an assimilation radius of 70 km may be too large. Furthermore, in most cases, an assimilation radius of 50 km combined with 50 ensemble members performs well.

Generally speaking, it is clear that augmenting the number of ensemble members provides a more precise representation of the model error and that the performance of the filter is therefore improved. For practical reasons of limits in computing time allowed for one simulation, the maximum number of ensemble members we can afford is 50.

The assimilation radius plays an important role as it limits the influence of the assimilated data. The choice of the assimilation radius is essentially governed by the features of the temperature in the North Sea region under study and by the geographical locations of the stations. For these reasons, we focus on the Southern and German Bights where observations may overlap. Several factors can justify why an assimilation radius of 50 km is adequate. Firstly the chosen horizontal decorrelation length between ensemble members is of 10 km. Then, in the Southern Bight, stations 6 and 7 are separated by 7 km in latitude. In the German Bight, stations 9 and 10 are distant by 11 km in latitude and 25 km in longitude.

5.4 Considering ensemble errors of the model and of the observations: the low rank square root algorithm

In this section we present results obtained with the low rank square root algorithm allowing an ensemble representation of the errors of the model as well as of the data. Though this algorithm provides analysed model states slightly more distant from the assimilated data than the square root algorithm, it has the advantage of diminishing the constraints on the model and of being more realistic.

The standard deviation of the ensemble of observations is half that of the model with $\gamma_{T_{obs}} = \eta_{T_{obs}} = 0.7°C$. There is neither vertical correlation nor horizontal correlation applied to the observations' ensemble members.

At first, the results obtained from the two algorithms are discussed and compared to an independent set of observations. Then the vertical correlation of the ensemble members is assessed by computing correlation coefficients.

5.4.1 Comparison of the results obtained without and with considering ensemble errors for the observations to a partial set of independent observations

The results discussed in this section are obtained with the low rank square root algorithm and correspond to experiment 'NS-11'. They are compared to the free run ('NS-1) and to the reference simulation ('NS-2') performed with the square root algorithm and discussed in detail in the previous sections. An independent set of observations is available only at stations 10, 11 and 15. The assessment of the low rank square root algorithm presented here is based on comparisons of the temperature profiles on the last day of assimilation, the 28th of September. It is expected that the temperature profiles modelled using the square root algorithm will be closer to the assimilated data because they are considered as perfect.

Stations 10 and 11

Figure 5.20 presents the temperature profiles at stations 10 (left) and 11 (right).

At station 10, as highlighted from previous discussions, the temperature modelled with data assimilation is very close to the assimilated data (thick black line). The assimilated data are colder than the model without assimilation (red crosses) and the observations are colder than the assimilated

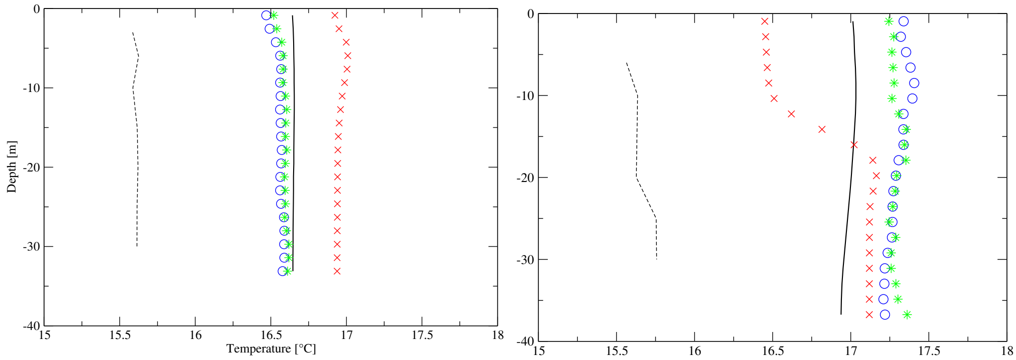

Figure 5.20: Temperature profiles on the last day of the assimilation period (September 28, 2001) at station 10 (left) and station 11 (right), without data assimilation 'NS-1' (red crosses - ×), with SR 'NS-2' (blue circles - ○) and LRSR 'NS-11' (green stars - $*$) data assimilation, assimilated profile (thick straight black line), independent observations (dashed black line).

data. Results from the low rank square root algorithm (green stars) are very slightly closer to the assimilated data than the square root algorithm (blue circles). This feature was not expected but the difference is quite small and the 28th of September is the only day at which this happens. At all other dates (independent observations available earlier in September not shown here), the square root results are closer to the assimilated data than those of low rank square root algorithm. A possible explanation to this is the influence of assimilated data at neighboring stations in the German Bight.

At station 11, the temperature modelled without assimilation (red crosses) shows a strong unstable thermocline. This instability is not present in the assimilated data (thick black line) but slightly captured by the independent observations (dashed black line). Once again, the features of the assimilated data are well reproduced by the assimilation process with square root (blue circles) and low rank square root (green stars) profiles typical of a well mixed station.

At stations 10 and 11, features of the assimilated data are well captured by the assimilation process with square root and low rank square root profiles very close to the assimilated data. However, at both stations, the independent observations are warmer than the assimilated data by about 1°C.

Stations 15 and 19

The temperature profiles at stations 15 and 19 are shown on Figure 5.21.

At station 15 (left panel of Figure 5.21), the assimilated data (thick black

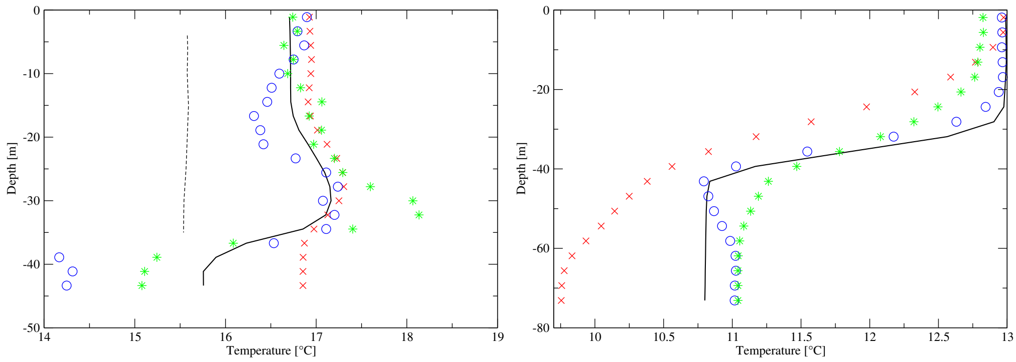

Figure 5.21: Temperature profiles on the last day of the assimilation period (September 28, 2001) at station 15 (left) and station 19 (right), without data assimilation 'NS-1' (red crosses - x), with SR 'NS-2' (blue circles - ○) and LRSR 'NS-11' (green stars - $*$) data assimilation, assimilated profile (thick straight black line), independent observations (dashed black line).

line) present a feature of unstable temperature stratification at 30 metres depth while it is absent from the independent observations (dashed black line). The temperature profile modelled without assimilation (red crosses) already shows a slight similar feature but not as markedly as the assimilated data. The assimilation reinforces this feature when the square root algorithm is applied as it is expected to provide a profile (blue circles) closer to the assimilated data. In the profile computed with the low rank square root algorithm (green stars), this feature is strengthened, with a temperature of 18°C at 35 metres depth and of 15°C at 40 metres. At 40 metres depth, the temperature profile of the square root algorithm (blue circles) is of 14°C while that of the low rank square root algorithm (green stars) is closer to the independent observations.

Results at station 19 (right panel of Figure 5.21) show that the largest difference between the non assimilative model (red crosses) and the assimilated data (thick black line) takes place at the bottom. Results obtained with the square root algorithm (blue circles) reproduce sharply the position of the thermocline in the assimilated data. They show a very good agreement with the assimilated data up to 50 metres depth. Below that depth, they are diverging from the assimilated data. This could be due to the fact that the corrections applied to the temperature are too strong and perturb the model dynamics which, in turn, influences the temperature at that station. This could be further investigated by examining the impact on the horizontal components of the velocity.

The square root algorithm provides usually results closer to the assimilated data as they are considered as perfect. However, the corrections applied to the temperature may perturb the model dynamics in the neighborhood and induce spurious features in the assimilative temperature profiles.

5.4.2 Influence of the vertical correlation of the model error

The vertical correlation of the ensemble members has already been discussed in the case of the square root algorithm. Nevertheless, we come back to that issue in the case of the low rank square root algorithm by computing the correlation coefficient between the assimilated data and the temperature modelled without and with vertical correlation. In the case of a non vertically correlated model error $\alpha_v = 0$ while when it is vertically correlated its value at station 10 is equal to 0.9 and at station 19 it is equal to 0.6.

Figure 5.22 presents the correlation coefficient between the assimilated data and the temperature modelled with a vertically non correlated model error and with a vertically correlated model error at station 10 (left) and at station 19 (right).

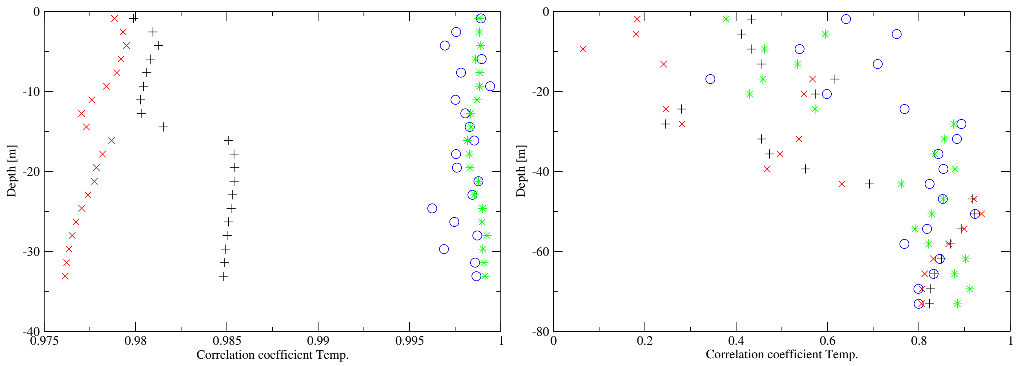

Figure 5.22: Correlation coefficients between the assimilated data and the temperature modelled with a *non vertically correlated* model error for simulations without data assimilation 'NS-12' (black plusses) and with data assimilation 'NS-13' (blue circles). Correlation coefficients between the assimilated data and the temperature modelled with a **vertically correlated** model error for simulations without data assimilation 'NS-1' (red crosses), with data assimilation 'NS-11' (green stars). At station 10 (left) and at station 19 (right).

At station 10, in all cases, the correlation coefficient between the assimilated data and the model is high, with values ranging from 0.97 to 0.99. The lowest values correspond to the temperature modelled without data assimilation (red crosses and black plusses). Without assimilation, the model results obtained with a non vertically correlated model error (black plusses) are slightly better, the largest difference taking place at the bottom (the difference between the two correlation coefficients is of the order of 0.01). When the temperature profiles are modelled with data assimilation, the correlation coefficient between the modelled temperature and the assimilated data indicates that the results are very slightly better at all water depths in the case of a vertically correlated model error (green stars). The maximum difference between the correlation coefficients represented by blue circles and green stars is of 0.005!

Results for station 19 are presented on the right panel of Figure 5.22. Without data assimilation, in the upper mixed layer, the correlation coefficients are again larger in the case of a non vertically correlated model error (black plusses). When data are assimilated, for the upper mixed layer the values of the correlation coefficient are larger with a non vertically correlated model error (blue circles) than with a vertically correlated model error (green stars). Below the thermocline the correlation coefficients are larger with a vertically correlated model error (green stars) than with a non vertically correlated model error (blue circles).

At well mixed stations, the assimilative results are slightly better (in terms of correlation coefficient between the model and the assimilated data) in the case of a vertically correlated model error. This is probably related to the dynamics of well mixed stations where the temperature is homogeneous along the whole water column.

5.5 Conclusions of model error sampling and temperature estimation of the 3-D case

This chapter was dedicated to a detailed study of the assimilation of temperature profiles in 3-D North Sea numerical simulations with a horizontal resolution of four nautical miles. Eight temperature profiles generated from model runs with a one nautical mile horizontal resolution have been assimilated.

Two algorithms of the ensemble Kalman filter, the square root algorithm and the low rank square root algorithm, have been applied. In the square root algorithm the observations are considered as perfect so that it allows to account for the maximum effect of the assimilation process. Hence it is used to assess the effect of the assimilation on the modelled temperature and to determine an adequate set of parameters for ensemble generation.

The choice of parameters for model error sampling has been studied in detail. Vertical and horizontal standard deviations of 1°C have been chosen for the temperature ensemble. They were derived from the temperature features of free model runs and of the assimilated data. The root mean square errors between the free model runs and the assimilated data vary from 0.2°C to 0.9°C. Such model-data discrepancies illustrate the challenges of data assimilation in coastal areas.

Vertical and horizontal decorrelation scales have been determined on the basis of temperature profiles representative of the two North Sea regimes in summer and from time series of the surface temperature at stations located close together. A vertical decorrelation length of 10 metres is chosen. A horizontal decorrelation length of 10 km between the error terms used to generate the ensemble members.

The use of a relatively small number of ensemble members can produce unwanted correlations between dynamically unrelated fields. Because of the computing cost associated to simulations with a high horizontal resolution (our run gets rejected from the machine after 24 hours), the number of ensemble members that can be used to represent the model error is limited to 50. Moreover, as the North Sea temperature has a high variability, the impact of observations located at a long distance from the point where the Kalman filter analysis is applied should be small. This can be achieved by localizing the analysis. A cut-off radius is used to limit the extent to which the assimilated data influence the neighboring temperature field.

Qualitative and quantitative analyses of the simulated North Sea temperatures using different combinations of the number of ensemble members and assimilation radius have been performed. The analysis of the root mean square error between the assimilated data and the temperature modelled from these numerical experiments has shown that the most adequate configuration consists of using 50 ensemble members with a cut-off radius of 50 kilometres.

The temperature profiles derived from the assimilative runs and the assimilated data have been quantitatively compared (root mean square errors and bias). Results indicate that the ensemble Kalman filter is adequate to reproduce to a good level of precision data features absent from the model in a 3-D coastal application where well mixed and summer stratified dynamical regimes coexist. In particular, the assimilation process reduces the root mean square error at a typical well mixed station by one order of magnitude, from 0.30°C to 0.03°C at the surface and from 0.29°C to 0.03°C at the bottom. At the stratified station, the root mean square error in the surface temperature remains of the order of 0.2°C at the surface and is reduced by a factor of three at the thermocline (from 1.5°C to 0.5°C), and by a factor of two at the bottom (from 0.7°C to 0.35°C). Furthermore, a comparison of the standard deviation of the temperature ensemble modelled without and with assimilation shows an immediate reduction induced by the Kalman filtering step.

In the low rank square root algorithm an ensemble representation is considered for the observations' errors, this allows to examine the impact of data assimilation in more realistic configurations. Results from both algorithms have been compared to independent observations. The square root algorithm provides model results closer to observations as they are assumed perfect while the low rank square root algorithm is more realistic.

Chapter 6

Combined model state and parameter estimation with the ensemble Kalman filter in the 1-D case: the North Sea CS station

In model state estimation examined in the previous chapters, the error stemming from all possible sources is represented by perturbation of the temperature that is further corrected at the analysis step. That approach to error representation can be complemented by taking further account of the errors generated by approximate values for model parameters. This chapter deals with the issue of combined model state and parameter estimation in the 1-D vertical test case at the CS station.

Prior to the simultaneous estimation of the model temperature and of a set of selected model parameters, it is necessary to examine the sensitivity of the model to the values of these parameters. This gives a primary idea of the expected impact of the parameters' adjustment on the modelled temperature and provides hints on how to define the range of perturbation of these parameters for ensemble generation. Results are reported in the first section of the chapter. From considerations on model error sources developed in chapter 5, the model parameters chosen for adjustment are the optical attenuation and the latent and sensible heat exchange coefficients.

For combined model state and parameter estimation three numerical experiments are performed. Their aim is to determine what is the best configuration of the model state vector in terms of the number of parameters that can be simultaneously adjusted. In the first experiment the temperature, horizontal currents and optical attenuation are simultaneously estimated, in the second one we estimate the model variables together with latent and sensible heat exchange coefficients while in the last one the temperature, horizontal currents and all coefficients are simultaneously adjusted. In the second section we compare the estimation of the parameters obtained from the different configurations of the model state vector. The third section is dedicated to an analysis of the temperature estimation. The numerical experiments are listed in Table 8.7 in Appendix F.

6.1 Study of the temperature sensitivity to the values of the parameters

6.1.1 Sensitivity of the temperature profile to the optical attenuation coefficient

The model parameter associated with the absorption of the solar radiation in the upper mixed layer is the optical attenuation coefficient. For reasons of ease of interpretation, we will also discuss in terms of the optical attenuation length which is its inverse. Three numerical simulations have been performed to assess the effect of the optical attenuation on the temperature. In the first two experiments, the attenuation coefficient is respectively 20% smaller and 20% larger than the default value, and in the third experiment it is 40% smaller.

The temperature profiles corresponding to different values of the optical attenuation length on August 15th are presented on Figure 6.1.

The default value of the optical attenuation coefficient corresponds to an optical attenuation length of 14 metres. An optical attenuation length of 12 metres corresponds to an optical attenuation coefficient 20% larger than the default value, and a length of 18 metres to an optical attenuation coefficient 20% smaller than the default value. The optical attenuation length of 23 metres corresponds to an optical attenuation coefficient 40% smaller. As the surface temperature of the non assimilative run is over-estimated, it is expected that the assimilation process increases the optical attenuation length, it is the reason why the results for an attenuation coefficient 40% larger are not considered here.

As expected, the temperature modelled with a shorter optical attenuation length (red curve) is slightly warmer at the surface and slightly colder at the bottom while a longer optical attenuation length (blue curve) allows the fixed-energy incoming solar radiation to warm deeper water layers so that

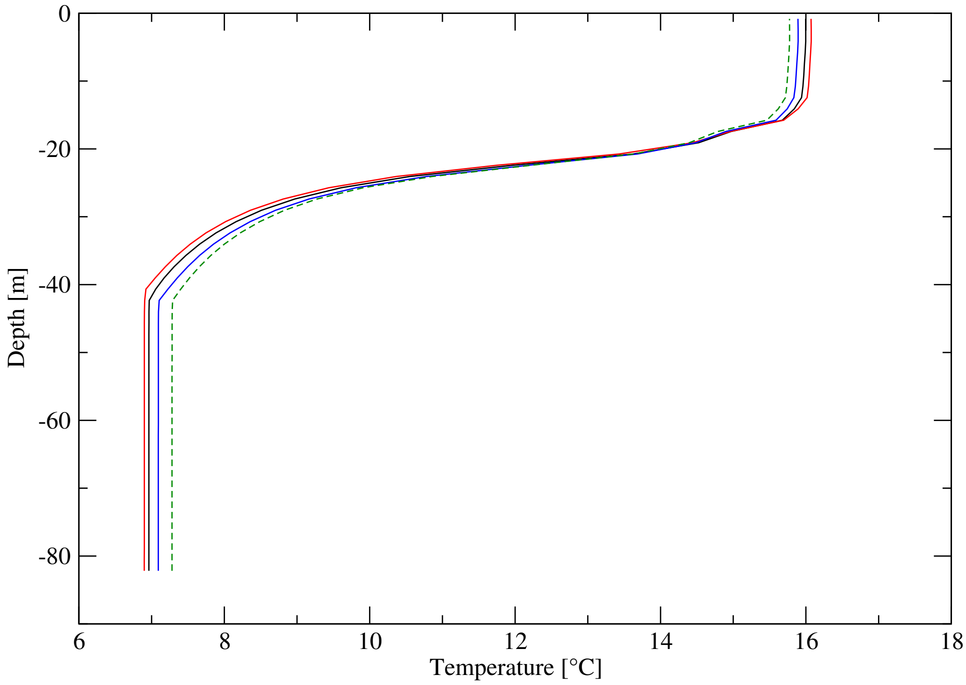

Figure 6.1: Temperature profiles on August 15th. The black curve corresponds to the default value of the optical attenuation length of 14 metres, the red curve corresponds to a decreased optical attenuation length of 12 metres, the blue curve correspond to an increased optical attenuation length of 18 metres and finally the green dashed curve corresponds to an optical attenuation length of 23 metres.

the surface water is colder and the bottom layer is warmer.

Between the temperature profile modelled with an optical attenuation coefficient larger by 20% with respect to the default value (red curve) and a temperature profile modelled with an optical attenuation coefficient smaller by 40% with respect to the default value (green dashed curve), there is a gap of 0.3°C at the surface and of 0.4°C at the bottom.

The bias between the model and the observations at the surface presented on Figure 4.18 in the case of the default run performed without data assimilation is of about 0.25°C. This indicates that perturbations of 40% around the default value can be considered as adequate.

6.1.2 Sensitivity of the temperature profile to the surface heat exchange coefficients

One way to determine the range of perturbation for the neutral surface heat exchange coefficients as a function of the surface temperature is to rely on the differences between the values of the neutral and stratified Kondo formulations. This difference is related to the ratio of the sea-air temperature difference to the wind speed at 10 metres above the sea surface and is given by equation (3.17). The values of the latent and sensible heat exchange coefficients have been computed using the neutral and the stratified formulations.

Figure 6.2 allows to compare the differences between the neutral and the stratified formulation of the latent heat exchange coefficient (left panel) and of the sensible heat exchange coefficient (right panel).

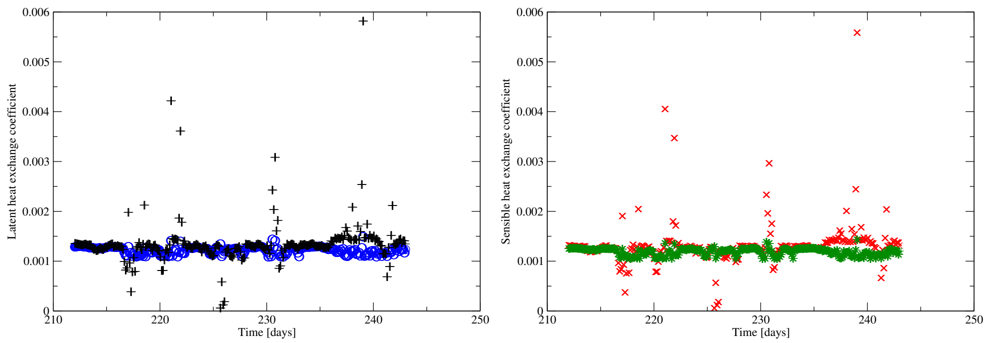

Figure 6.2: Time series of the latent heat exchange coefficient (left panel), the blue circles correspond to the neutral formulation while the black plusses correspond to the stratified formulation. Sensible heat exchange coefficient (right panel), the green stars represent the neutral formulation and the red crosses the stratified formulation.

For both coefficients, the mean values in the stratified formulation (black plusses for the latent heat coefficient and red crosses for the sensible heat coefficient) are about the same as in the neutral formulation (blue circles for the latent heat coefficient and green stars for the sensible heat coefficient).

The magnitude of the heat exchange coefficients in the Kondo stratified formulation is correlated to small wind speeds. Their values are much larger when the sea-air temperature difference is positive than when it is negative. Hence they are located at the same time steps for both coefficients. The maximum values of these departures reach a value of 0.006 which is four times the mean value of 0.0015. In order these coefficients not to take negative values, such a range for the ensemble perturbations is too large. A standard deviation of 15% is prescribed for the generation of their ensemble.

6.2 Study of the parameter estimation in the 1-D case

The assessment of combined estimation of model state and parameters is split into two sections. This section focuses on the adjustment of the parameters. The effects on the estimation of the temperature will be dealt with in the next section.

Results from 1-parameter, 2-parameters and 3-parameters adjustment cases are compared. They correspond respectively to the simultaneous adjustment of the temperature, horizontal currents and optical attenuation, of the temperature, horizontal currents and heat exchange coefficients and of the temperature, horizontal currents, optical attenuation and heat exchange coefficients.

As it allows an ensemble representation of the observations' errors the low rank square root algorithm provides a more realistic representation of the system, therefore it is used for combined model state and parameter estimation. The model error is vertically correlated and the parameters governing the ensemble generation correspond to those of the reference simulation.

6.2.1 Effects on the optical attenuation coefficient

An ensemble for the optical attenuation coefficient is generated by adding error terms to a first guess corresponding to its default value of $0.07m^{-1}$ previously considered in the simulations of model state estimation. These error terms are generated with a horizontal correlation.

Results for the optical attenuation from the simultaneous estimation of the temperature, horizontal currents and optical attenuation coefficient (experiment 'CPA-1') are compared to those of the simultaneous estimation of the temperature, horizontal currents, optical attenuation and heat exchange coefficients (experiment 'CPA-3'). The first experiment ('CPA-1') is performed with the $AD1$ state vector which contains the model state variables and the optical attenuation coefficient that are simultaneously adjusted at the analysis step. The $AD3$ state vector contains the model state variables, the optical attenuation coefficient and the surface heat exchange coefficients i.e. the set of examined model parameters. It is used for the third experiment ('CPA-3').

The time series of the optical attenuation coefficient modelled with the $AD1$ state vector are represented by the magenta line in Figure 6.3 while those of the $AD3$ state vector are in green. During the four first days of August when no assimilation is applied, the optical attenuation coefficient is represented by a horizontal green line corresponding to its default value.

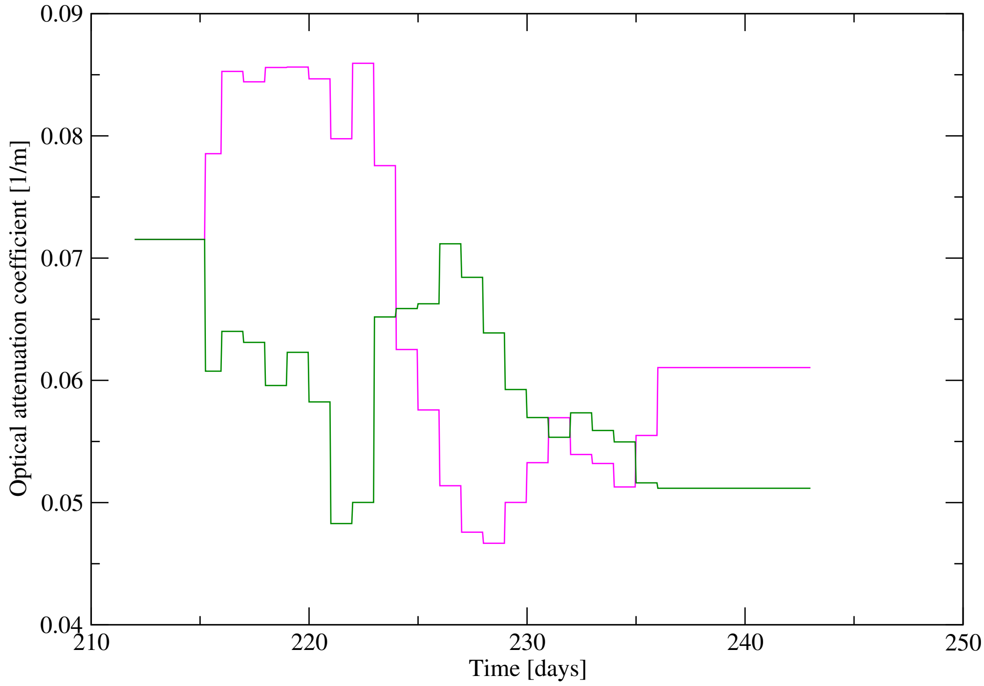

Figure 6.3: Time series of the optical attenuation coefficient computed with the $AD1$ state vector (magenta line) 'CPA-1' and the $AD3$ state vector (green line) 'CPA-3'.

The magenta line ($AD1$ state vector) indicates that after the beginning of the assimilation process the optical attenuation coefficient increases from $0.07m^{-1}$ to $0.085m^{-1}$, corresponding to a 21% increase, until day 225 (13th of August). From that day on it decreases steeply to reach a value of $0.061m^{-1}$ corresponding to an optical attenuation length of 16 metres.

The green line corresponds to the results obtained with the simultaneous adjustment of the set of model parameters ($AD3$ state vector). From the first assimilation day on the optical attenuation coefficient decreases. At the end of the assimilation period it reaches a value of $0.051m^{-1}$ which corresponds to an attenuation length of 20 metres in place of the default value of 14 metres.

Roughly speaking the $AD1$ and $AD3$ adjusted optical attenuation coefficients show opposite behaviour. In order to better understand the link between the assimilated temperature data at 10 metres depth and the adjusted optical attenuation coefficients, the optical attenuation coefficients are scaled to the same range as the temperature data. This is shown in Figure 6.4.

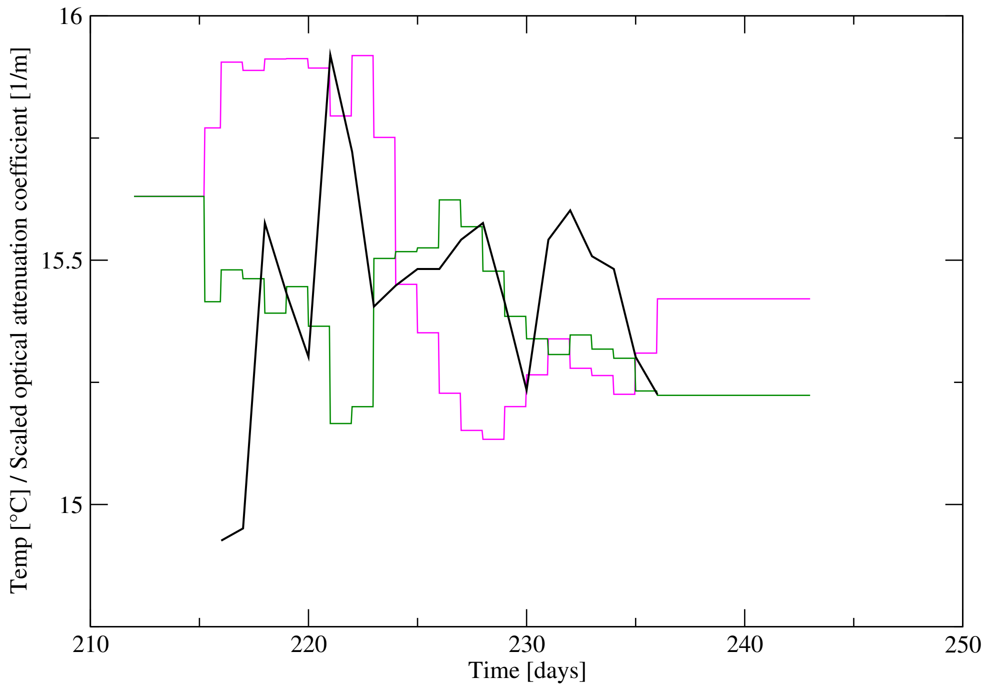

Figure 6.4: Time series of the scaled optical attenuation coefficient computed with the $AD1$ state vector (magenta line) and the $AD3$ state vector (green line) and of the temperature data at 10 metres depth (black line).

At the beginning of the assimilation period, on day 221 (9th of August), the temperature data rise to their maximum value around 16°C. This feature is reproduced by the optical attenuation coefficient adjusted with the $AD1$ state vector (magenta line). Then the temperature data decrease until day 230 (18th of August) to a value of 15.25°C; that trend is followed by the $AD3$ adjusted coefficient (green line) which is correlated to the temperature data between days 222 (10th of August) and 230 (18th of August).

Finally, from day 230 (18th of August) until the end of the assimilation period, the analysed values of the optical attenuation coefficient have a smaller variation range. One possible explanation for this "stabilization" could be that the error terms added to generate the ensemble are not time varying and are added to a good initial guess [Song and Grizzle, 1995] [91].

Because of the too high location of the thermocline the surface temperature is over-estimated by the non assimilative model run. To contribute to a surface temperature decrease, the optical attenuation coefficient should decrease so that the associated optical attenuation length increases. The results obtained with the $AD3$ state vector are slightly better than those of the $AD1$ state vector as the adjustment process reduces the optical attenuation coefficient since the beginning of the assimilation and the optical attenuation length is therefore increased.

6.2.2 Effects on the surface heat exchange coefficients

The effects on the heat exchange coefficients of the simultaneous estimation of the model state and heat exchange coefficients (experiment 'CPA-2') are compared to those of the simultaneous estimation of the model state, optical attenuation and heat exchange coefficients (experiment 'CPA-3'). The simulation 'CPA-2' is performed with the $AD2$ state vector and involves the simultaneous adjustment of the temperature, horizontal currents and surface heat exchange coefficients in their neutral formulation. The optical attenuation coefficient is kept constant to its default value. For the simulation 'CPA-3' performed with the $AD3$ state vector, the temperature, horizontal currents, optical attenuation and heat exchange coefficients in their neutral formulation are simultaneously adjusted.

An ensemble of heat exchange coefficient is generated by adding error terms to their neutral formulation. These error terms have a horizontal correlation. The stratified formulation is considered as more realistic than the neutral formulation as it is dependent on the ratio of the sea air temperature difference to the wind speed which play a large role in the heat exchange at the sea surface. The adjusted values of the neutral heat exchange coefficients are compared to their stratified values which were extracted from a simulation with temperature estimation ($AD0$ state vector).

Latent heat exchange coefficient

Results for the latent heat coefficient are presented in Figure 6.5. The left panel corresponds to the $AD2$ state vector and the right panel corresponds to the $AD3$ state vector. As their Kondo parameterisations are alike, results for the sensible heat coefficient are very similar and are not shown.

The latent heat coefficients computed with a stratified formulation (black plusses) have larger values than those derived from the adjustment of the neutral formulation (blue circles for $AD2$ and green crosses for $AD3$). This can be attributed to the magnitude of the prescribed standard deviation that had to be limited to avoid negative values at the ensemble generation.

Compared to the adjustment with the $AD2$ state vector (blue circles), the points departing from the mean to larger values are slightly more numerous

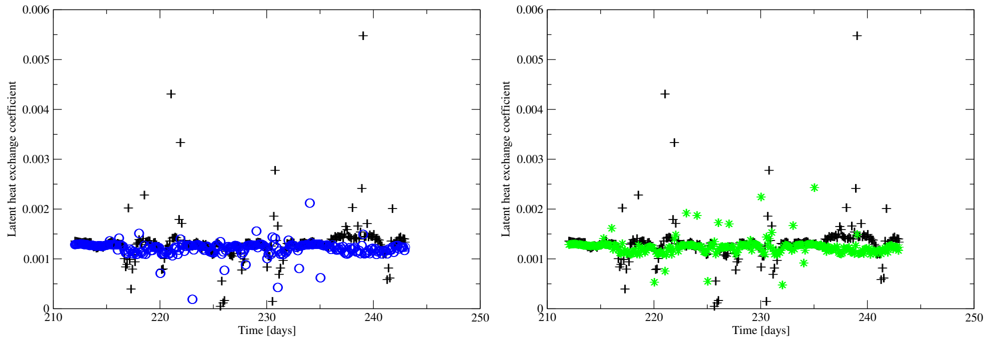

Figure 6.5: Time series of the latent heat exchange coefficient estimated with the $AD2$ (left) 'CPA-2' and with the $AD3$ (right) 'CPA-3' state vector. The stratified formulation is represented by black plusses, the blue circles correspond to the $AD2$ adjustment and the green stars to the $AD3$ adjustment.

in the case of the adjustment performed with the $AD3$ state vector (green stars), and the maximum value is slightly larger (0.025 with $AD3$ against 0.0205 with $AD2$).

However, the adjustment of the neutral heat exchange coefficients neither with the $AD2$ nor with the $AD3$ state vector is able to produce coefficients clearly correlated in time and in magnitude with the stratified coefficients.

In the case of the $AD2$ estimation state vector, the heat exchange coefficients are assumed to be the sole parameters responsible for the missing parameterization of the surface processes. This could induce misleadings in their adjustment.

Sensible heat fluxes

The results of the estimation of heat exchange coefficients in themselves do not allow to clearly conclude which estimation state vector is the most adequate. In order to get some more hints, we have computed the sensible heat fluxes provided by three formulations of the sensible heat exchange coefficient: the stratified formulation without adjustment of the coefficient but update of the sea surface temperature by the assimilation process ($AD0$ state vector), the adjusted neutral formulation from the $AD2$ and $AD3$ state vectors. We focus on the sensible heat fluxes as they are directly proportional to the sensible heat exchange coefficient and to the sea air temperature difference. They are computed following equation (3.17). Results are presented on Figure 6.6.

The gaps between the stratified and neutral formulations of the heat

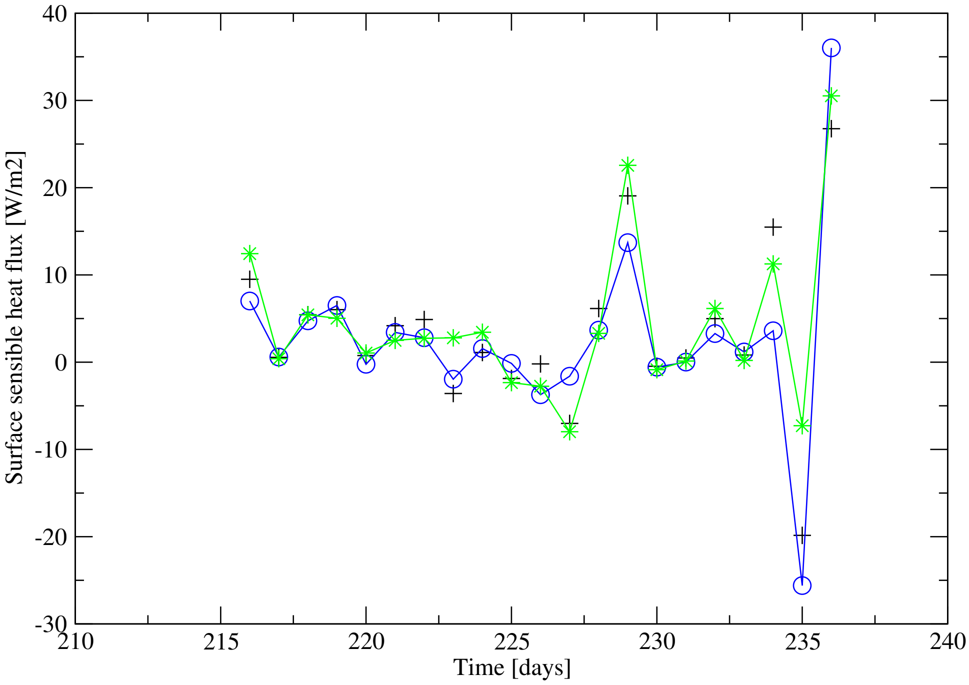

Figure 6.6: Time series of the sensible heat flux at the surface computed with a non adjusted stratified formulation (black plusses), the $AD2$ (blue circles) 'CPA-2' and $AD3$ (green stars) 'CPA-3' adjusted neutral formulations.

exchange coefficients are dominated by the sea-air temperature differences in small wind conditions and are even smoother in the sensible heat flux than in the sensible heat coefficient (not shown here but with similar trends to the latent heat coefficient). This is interpreted in terms of the sensible heat exchange process. When the sea-air temperature difference is large and in small wind conditions, the sensible heat flux is the main process for balancing the temperature difference between the two fluids.

Figure 6.6 tends to indicate that values exceeding the stratified formulation occur mainly with the $AD3$ state vector (green stars). However it is not always the case, especially at the end of the assimilation period.

No independent data of the values of the parameters are available for these simulations. Temperature results from simulations performed without parameter estimation but with data assimilation and a more realistic stratified formulation of the latent heat exchange coefficients constitute the reference for the interpretation of the parameters estimation. Therefore, the root mean square errors between the sensible heat fluxes derived respectively with the $AD2$ and $AD3$ state vectors with respect to the $AD0$ configuration have been computed. The values are respectively 4.13 $W.m^{-2}$ and 3.67 $W.m^{-2}$ so that slightly better results are obtained when the $AD3$ state vector is used.

As for the adjustment of the optical attenuation coefficient, slightly better results in the estimation of the surface heat exchange coefficients are obtained when the set of selected parameters are simultaneously adjusted ($AD3$ state vector).

6.3 Effect of the simultaneous model state and parameters estimation on the temperature

In this section we assess the effects on the temperature of the three configurations of the model state vector for the simultaneous estimation of the model state and parameters. The three numerical experiments 'CPA-1', 'CPA-2' and 'CPA-3' correspond respectively to the three state vectors 'AD1', 'AD2' and 'AD3'. We examine the temperature profile at the end of the assimilation period as well as time series of the temperature at 10 metres depth.

The temperature profiles on the last day of the assimilation period (25th August) are presented on the left panel of Figure 6.7. They allow to compare the temperature profile modelled with data assimilation but no parameter adjustment i.e. with the $AD0$ state vector (brown diamonds), the observations that are assimilated (black plusses) and the temperature profiles modelled with the $AD1$ state vector (magenta crosses), $AD2$ state vector (blue circles) and $AD3$ state vector (green stars).

From the temperature profiles presented on the left panel of Figure 6.7, the following remarks can be made. The differences in the surface temperature estimation between the different simulations are of about 0.1°C. Close to the surface the temperature from the simulation with the $AD2$ state vector (blue circles) is the closest to the observations. However below 60 metres depth where no more observations are available, the temperature difference between the simulation performed with the $AD2$ state vector and the other simulations increases to 0.31°C.

The right panel of Figure 6.7 focuses on time series of the temperature at 10 metres depth. There is no clear hint of one estimation state vector providing an estimation closer to the observations. However, the green stars corresponding to the $AD3$ state vector follow more closely the observations, especially in the middle of the assimilation period, from day 222 (10th of August) until day 234 (22nd of August). The results with the $AD1$ state

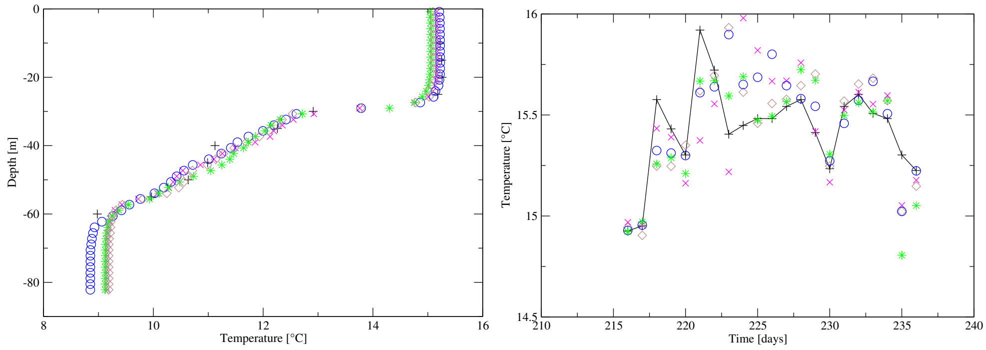

Figure 6.7: Temperature profile (left) at the end of the assimilation period, August, 25th. Time series of the temperature at 10 metres depth (right). The black plusses represent the assimilated observations, the brown diamonds the temperature modelled with the $AD0$ state vector 'CS-13'. The temperature modelled with the $AD1$ state vector is represented by magenta crosses 'CPA-1', with the $AD2$ state vector by blue circles 'CPA-2'and with the $AD3$ state vector by green stars 'CPA-3'.

vector (magenta crosses) present the largest variations.

As the differences between the different model configurations in temperature profiles as well as in time series do not allow to conclude, the root mean square error between the temperature modelled respectively with the $AD0$, $AD1$, $AD2$ and $AD3$ state vectors and the assimilated observations has been computed. Results are presented on Figure 6.8.

Around the thermocline, at 25 and 30 metres depth, the largest root mean square errors appear with the $AD2$ state vector. At 10 metres depth, the root mean square error between the observations and the temperature modelled with the different estimation state vectors are the following:

- Without assimilation: 0.49°C,
- $AD0$: 0.19°C,
- $AD1$: 0.21°C,
- $AD2$: 0.19°C,
- $AD3$: 0.18°C.

Although the differences are of the order of 0.01°C which is one order of magnitude smaller than the estimated root mean square error of the observations (which is of the order of 0.1°C), they indicate a slight benefit from

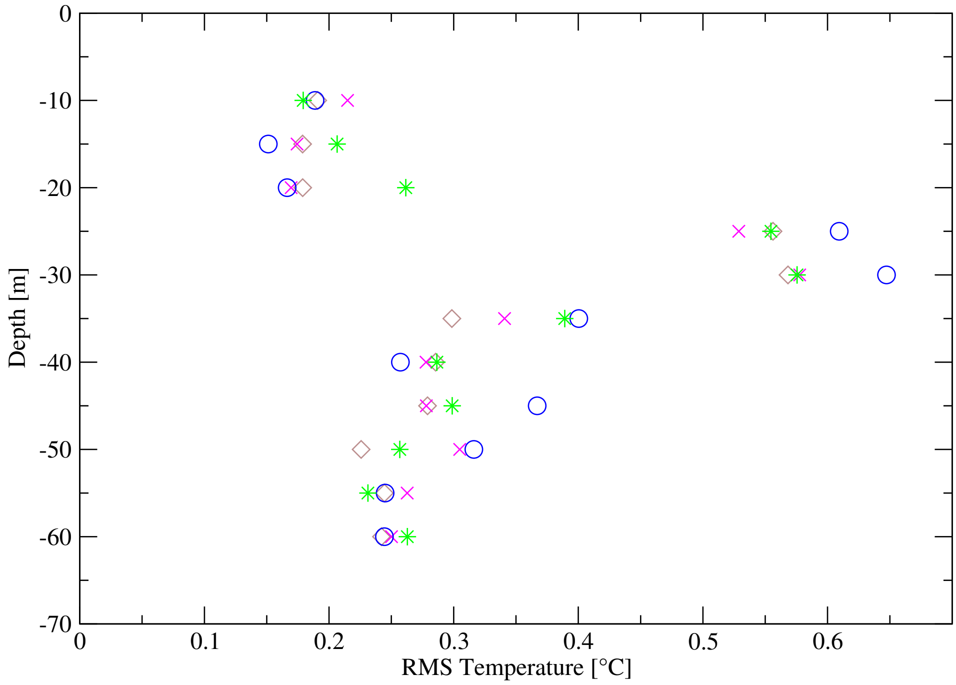

Figure 6.8: Root mean square error between the assimilated data and the temperature modelled respectively with the $AD0$ state vector (brown diamonds) 'CS-13', the $AD1$ state vector (magenta crosses) 'CPA-1', the $AD2$ state vector (blue circles) 'CPA-2' and the $AD3$ state vector (green stars) 'CPA-3'.

the $AD3$ state vector. Finally, as the the combined estimation provides a more realistic model error representation, it does not degrade the estimation of the model state.

6.4 Conclusions of combined model state and parameter estimation the 1-D case

Combined model state and parameter estimation with the ensemble Kalman filter has been studied in a 1-D vertical implementation of the COHERENS model at the North Sea CS station which is stratified in summer. The parameter adjustment is based on the state vector augmentation method. The simulations are based on a 1-D simplified physical model inducing a wrong representation of the temperature profile. The parameters primarily linked to the temperature of the upper mixed layer are considered for adjustment, the chosen parameters are the optical attenuation coefficient and the surface heat exchange coefficients.

In order to assess the influence of the simultaneous adjustment of several parameters, simulations were performed with different augmented state vectors comprising the model state plus only one of the parameters, two of them or the set of selected parameters.

As no independent data for the parameters are available, the interpretation of the estimation of the optical attenuation coefficient is based on the reduction of the discrepancy between the modelled temperature of the upper mixed layer and the assimilated observations. The estimated optical attenuation coefficient is more realistic when the set of selected parameters is considered ($AD3$ state vector). Moreover, after a few assimilation steps, the estimated values have a smaller variation range.

Concerning the heat exchange coefficients, we compare the adjusted values of the coefficients in the neutral formulation to the more realistic stratified formulation. These results are further assessed by comparison of the sensible heat fluxes derived from the temperature field derived from an adjusted neutral formulation of the coefficients to those computed from a temperature modelled with a stratified formulation of these coefficients. As in the case of the optical attenuation coefficient, the heat exchange coefficients are slightly more realistic when the set of selected parameters are adjusted simultaneously with the model state ($AD3$ state vector).

The impact of combined model state and parameter estimation on the temperature representation is relatively small with respect to the "sole" model state estimation. This is attributed to the fact that after model state estimation, there was not much room left for further improvement of the temperature of the upper mixed layer. However, results indicate a slight benefit in the temperature representation when the set of selected parameters are simultaneously estimated ($AD3$ state vector). The aim of these simulations was to assess if parameter estimation in a coastal model could be performed with the ensemble Kalman filter. The obtained results constitute a preliminary step towards applications to the estimation of other physical parameters and ecological parameters. They also indicate that the model state estimation is not degraded by the combined adjustment of the model state and parameters but slightly improved as this configuration corresponds to a more realistic model error representation.

Chapter 7

Combined model state and parameter estimation with the ensemble Kalman filter in the 3-D case : the North Sea

In chapter 5, we have focused on model state estimation on the whole North Sea domain. Such an approach treats globally the different sources of model error in the representation of the North Sea temperature. However, within specified error bounds, a model formulation (for a given set of parameters) may not be able to reproduce the data. In that case, the model parameters may be the cause of the model error so that model parameter estimation may be a tool to determine model inaccuracies.

The results presented in that chapter aimed at testing the details of the filter implementation, at defining an adequate set of parameters for model error sampling, and at showing that the ensemble Kalman filter could be successfully used for data assimilation in coastal models.

This chapter is dedicated to the simultaneous estimation of model state and parameters on the whole North Sea domain with a special focus on the spatial variability of the parameter estimation. As in the 1-D case, the estimation of parameters is based on the augmentation of the model state vector. As they are primarily involved in processes governing the temperature of the upper mixed layer, we have chosen to estimate the optical attenuation and the latent and sensible heat exchange coefficients.

The simulations of parameter estimation reported in this chapter constitute a first step towards more realistic estimations of model parameters. Compared to the 1-D vertical test case, proxy data are assimilated instead of "real" observations. As a result these 3-D simulations present several drawbacks. First of all, they were generated with default constant values of the optical attenuation coefficient. Then, they do not provide information from processes not incorporated in the model such as the optical properties of the water masses. As a result, it is not expected to derive "true" values for the model parameters but rather to examine how the ensemble Kalman filter is performing. Then, the estimated values of the parameters are not compared to observations of the parameters but assessed in terms of the assimilated temperature data.

Several numerical experiments with different combinations of the model state vector and parameters are performed. Their aim is to highlight the benefit of combined model state and parameter estimation with respect to the "sole" model state estimation.

As it allows an ensemble representation of the observations' errors, the low rank square root (LRSR) algorithm is applied for combined model state and parameter estimation. The parameters for ensemble generation are the same as for the LRSR reference simulation 'NS-11'.

In the first section of this chapter we determine the range of perturbation of the parameters' values for generating the model ensemble. Then, the objectives of the numerical experiments are described. In the third section, results of the estimation of the optical attenuation coefficient and surface heat exchange coefficients are discussed. In section 4, the benefit of combined model state and parameter estimation in the representation of the temperature profiles is highlighted. The conclusions are given in the fifth section. The simulations and their features are summarized in Table 8.8 in Appendix G.

7.1 Determination of the range of perturbation of the selected parameters in the 3-D case

In this section we examine the impact on the temperature profiles of different values of the selected model parameters. The sensitivity of the temperature to different values of these parameters will give us some hints to define the standard deviation of the ensemble for each of these parameters.

7.1.1 Sensitivity of the temperature to the optical attenuation coefficient

At first we examine the sensitivity of the temperature to the optical attenuation coefficient. Four sensitivity experiments are performed. In the first two experiments the optical attenuation coefficient is respectively decreased by 40% and 20% while in the third and fourth experiments it is increased by respectively 20% and 40%.

Temperature profiles on the 25th of September are presented on Figure 7.1 at station 10 (left) and at station 19 (right). This date has been chosen as it is the day at which the differences between the cases are the largest.

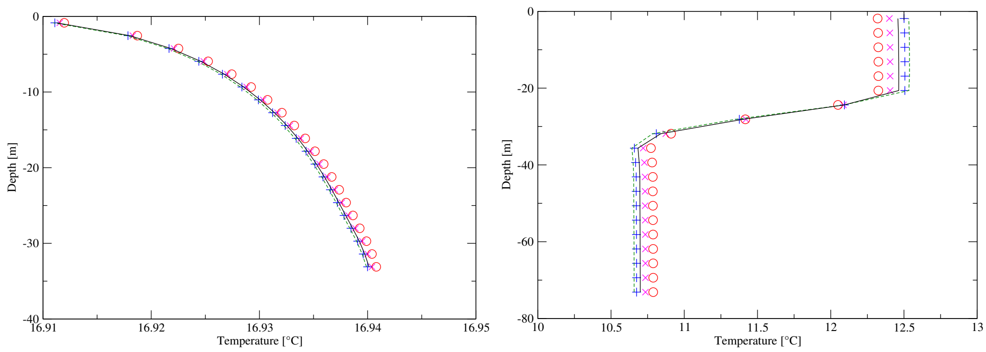

Figure 7.1: Temperature profiles on the 25th of September at station 10 (left) and at station 19 (right). The black curve is for the default optical attenuation coefficient corresponding to an attenuation length of 17 metres. Magenta crosses correspond to a coefficient decrease of 20%, the red circles to a decrease of 40%, the blue plusses to an increase by 20% and the dashed green line to an increase by 40%.

At station 10, any change of the optical attenuation coefficient induces almost unnoticeable changes in the temperature profile with respect to the reference value (black curve). At well mixed stations the diffusion in the water column is high so that the optical attenuation coefficient has a negligible influence.

At the stratified station 19, the temperature sensitivity to the values of the optical attenuation coefficient is slightly larger than at station 10, it varies by up to 0.25°C. When the optical attenuation coefficient increases (blue plusses and dashed green line), the optical attenuation length is decreased and the temperature of the upper mixed layer increases while the corresponding temperature below the thermocline decreases. The effects are proportional to the magnitude of variation of the optical attenuation coefficient. A decrease by 40% with respect to the default value corresponds to an optical attenuation length of 28 metres while an increase by 40% corresponds to an optical attenuation length of 12 metres.

A variation of the optical attenuation coefficient of ±40% with respect to the default value corresponds to optical attenuation lengths varying between 28 metres and 12 metres. In the well-mixed zone the temperature profile is decreasing slightly more strongly in the first 10 metres; below that depth it is almost constant along the vertical. As already mentioned, this is related to the dissipation of the incoming radiation along the whole water column in well mixed zones and explains why the optical attenuation length does not play a critical role in the shape of the temperature profile at station 10. At the stratified station, things are a little bit different. Results show slight changes in the temperature of the upper mixed layer and of the temperature below the thermocline. The changes vary linearly with the values of the optical attenuation coefficient.

7.1.2 Sensitivity of the temperature field to the surface heat exchange coefficients

The COHERENS model allows to use a neutral or a stratified Kondo formulation of surface heat exchange coefficients. The objective of the parameter estimation that will be performed in this study is to compare the adjusted neutral formulation to the stratified formulation. In order to define the range of the ensemble for the heat exchange coefficients, we compare their values in the two formulations without data assimilation.

Results for the latent heat exchange coefficients are presented on Figure 7.2 at stations 10 (left) and 19 (right) (there is no adjustment of these parameters by data assimilation).

At station 10, the mean value of the latent heat exchange coefficient in the neutral formulation is around 0.00125 while in the stratified formulation it is around 0.0014. Latent heat coefficients in the neutral formulation range between 0.0011 and 0.0013 while in the stratified formulation, they vary from 0.0013 to 0.0017.

At the stratified station 19, the latent heat exchange coefficients in the neutral formulation are in the variation range of the stratified case. This variation range is comprised between 0.001 and 0.0014 (neglecting the extreme value of 0.0018).

The sensible heat exchange coefficients in the neutral and stratified formulation at stations 10 and 19 are presented on Figure 7.3.

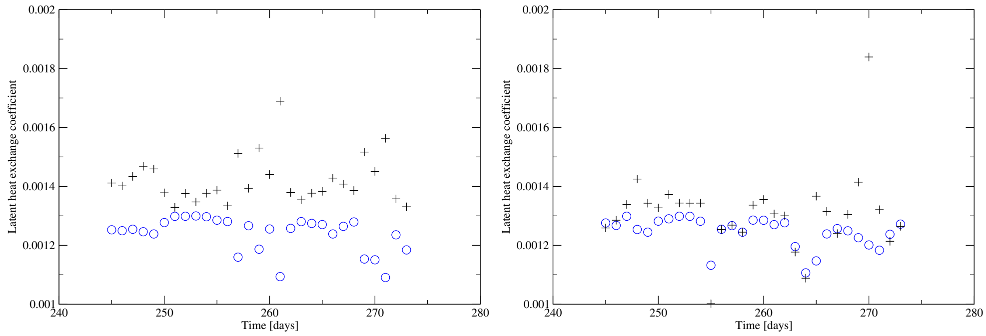

Figure 7.2: Time series of the latent heat exchange coefficient at station 10 (left) and at station 19 (right). The blue circles correspond to the neutral formulation and the black plusses to the stratified formulation.

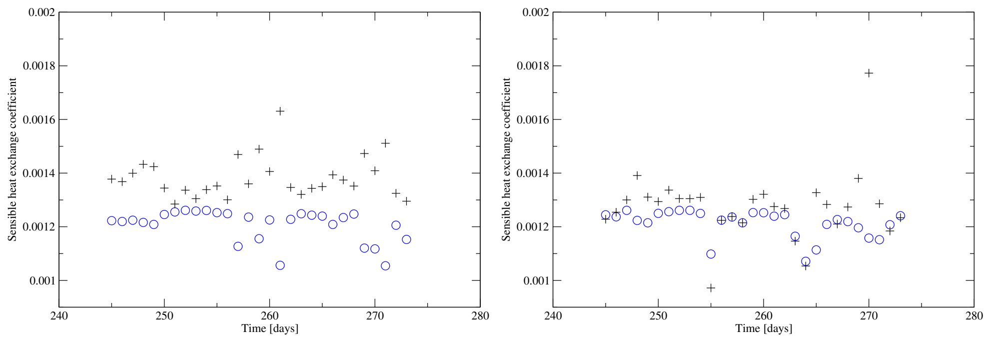

Figure 7.3: Time series of the sensible heat exchange coefficient at station 10 (left) and at station 19 (right). The blue circles correspond to the neutral formulation and the black plusses to the stratified formulation.

At station 10, the features of the sensible heat exchange coefficients are the same as those of the latent heat coefficients with values slightly smaller in the case of the sensible heat coefficient.

At the stratified station 19, the sensible heat coefficients show the same features as the latent heat coefficients. Like at station 10 their values are slightly smaller than those of the latent heat coefficient, with a minimum value below 0.001.

For a given station, the sensible and latent heat coefficients show the same features. The stratified formulation is obtained by multiplying the neutral formulation by the same factor for the latent and sensible coefficients. This factor depends on the ratio of the sea-air temperature difference to the wind speed at 10 metres height (equation 3.23). The differences of magnitude of the coefficients in the neutral formulation stem from the coefficients weighting the wind speed (equations (3.21) and (3.22)).

If we neglect the point on day 270 (27th of September), the spread of the stratified values of the heat exchange coefficients is smaller at station 19 than at station 10. At station 10, the differences between the mean values of the two formulations are the largest because the stratified formulation provides coefficients larger than the neutral formulation.

7.2 Objectives of the numerical experiments

One objective of the simulations is to evaluate the ensemble Kalman filter for combined model state and parameter estimation in cases where one parameter or a set of parameters are simultaneously adjusted with model state variables. Then, as the North Sea contains zones in a well mixed regime and other zones stratified in summer, the spatial variability of the parameter estimation will be examined. In the framework of this study, we also examine if a further benefit of the estimation of a set of selected model parameters with respect to the "sole" model state estimation can be obtained for the representation of the temperature. Therefore the parameters for sampling the model error of the temperature (vertical and horizontal standard deviations of the ensemble, decorrelation lengths, ...) have the same value as in the case of 'sole' model state estimation and the model state vector containing the temperature, salinity and horizontal currents remains simultaneously analysed.

The generation of ensembles for the parameters is based on the sensitivity of the model to values of the parameters provided in the first section of this chapter. The optical attenuation coefficient is perturbed with a standard deviation of 20% while the neutral heat exchange coefficients are perturbed with a standard deviation of 15% around their mean values.

7.3 Study of the variability of the adjusted model parameters in the 3-D case

In this section, we assess the estimation of the model parameters as well as the spatial variability of the adjusted parameters. We compare the results of the combined estimation of model state variables with one parameter and with a selected set of them. The spatial variability of the estimation is examined by comparison of the results at stations representative of the two North Sea regimes, station 10 and station 19.

7.3.1 The optical attenuation coefficient

Estimated optical attenuation coefficient

At first we focus on the estimation and spatial variability of the optical attenuation coefficient. Results from numerical experiments 'NPA-1' and 'NPA-3' are compared. Experiment 'NPA-1' is performed with the $AD1$ state vector while experiment 'NPA-3' is performed with the $AD3$ state vector. In the case of the $AD1$ state vector, the optical attenuation coefficient is the sole estimated parameter whereas in the $AD3$ state vector it is simultaneously estimated with the heat exchange coefficients. For both $AD1$ and $AD3$ state vectors the selected model parameters are simultaneously estimated with the model state variables.

An ensemble for the optical attenuation coefficient is generated by adding error terms to a first guess corresponding to its default value of $0.06m^{-1}$ previously considered in the simulations of model state estimation. These error terms are generated with a horizontal correlation.

Results from these two experiments are presented on Figure 7.4, on the left panel for station 10 and on the right panel for station 19. Note that the vertical scale of the two graphs is different.

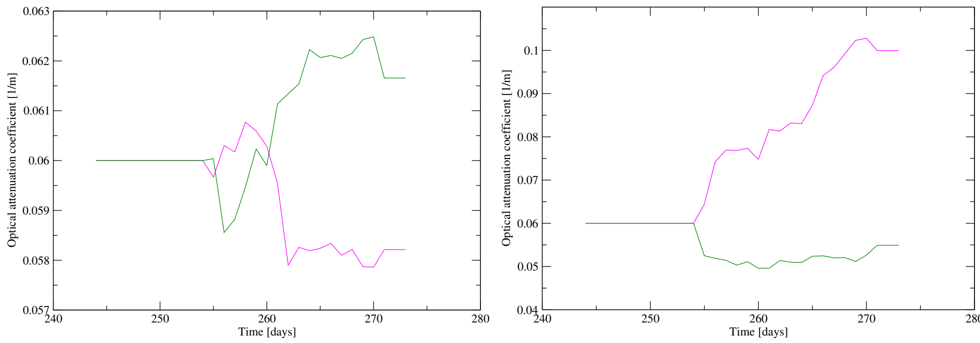

Figure 7.4: Time series of the optical attenuation coefficient at stations 10 (left) and 19 (right), computed with the $AD1$ state vector (magenta line) and the $AD3$ state vector (green line).

Both estimation experiments 'NPA-1' and 'NPA-3' are performed with the same assimilated temperature data at a given station. However, they result in estimations of the optical attenuation coefficient of opposite tendencies and differing in magnitude. At station 10 (left panel of Figure 7.4), the estimation with the $AD1$ state vector (magenta line) provides an increasing optical attenuation coefficient during the five first days of the assimilation and then the optical attenuation coefficient decreases to stabilize around 0.058 m^{-1}, i.e. to an optical attenuation length of 17 metres.

Results from the simulation performed with the $AD3$ state vector (green line) at station 10 show a first decrease of the optical attenuation that is followed by an increase stabilizing around 0.062 m^{-1} corresponding to an optical attenuation length of 16 metres. Though the tendencies shown by the two experiments at station 10 are in opposite direction, they both indicate a variation of the same small magnitude, about 3%, of the default value.

In the stratified area of the North Sea at station 19 (right panel of Figure 7.4), the estimation with the $AD1$ state vector (magenta line) shows an increasing optical attenuation coefficient which stabilizes at 0.1 m^{-1} corresponding to an optical attenuation length of 10 metres. Experiment performed with the $AD3$ state vector (green line) provides a decreasing optical attenuation coefficient of 0.05 m^{-1}, corresponding to an optical attenuation length of 20 metres. The magnitude of variation between the two experiments provides optical attenuation lengths differing by a factor of two at station 19.

At station 10, though the estimated values of the optical attenuation coefficient differ in tendency depending on the estimation state vector, they are not significantly different in magnitude. Moreover the stabilize for both state vectors. At station 19, the estimation of the optical attenuation coefficient derived with the $AD3$ state vector has a small range of variation. These features are probably related to the time constant values of the error terms used to generate the ensemble and to the fact that the ensemble is generated from a best guess estimate [Song and Grizzle, 1995] [91].

The estimated values of the optical attenuation coefficient are interpreted in terms of the discrepancy between the temperature of the upper mixed layer modelled without and with data assimilation. At station 10, table 5.1 indicates a mean model error of 0.24°C between the model without assimilation and the data at the surface. This corresponds to an over-estimation of the surface temperature by the model. In terms of optical attenuation length, a decrease of the temperature by the assimilation can be obtained by an increase corresponding to a decrease of the optical attenuation coefficient. This is derived with the $AD1$ state vector (magenta line). At station 19, the mean model error is of -0.11°C at the surface (see table 5.1). Hence the model under-estimates the surface temperature. So, one expects an increase of the surface temperature by the assimilation process at that station. This can be achieved by increasing the optical attenuation coefficient. Such a tendency is again resulting from the estimation with the $AD1$ state vector (magenta line).

Finally, contrary to the 1-D vertical test case, at the two stations under consideration for the 3-D simulations, slightly more realistic results are obtained when the optical attenuation coefficient is the sole parameter simultaneously adjusted with the model state. Whether the station is well mixed or stratified, the corresponding estimations of the optical attenuation length may vary from 17 metres ($AD1$ state vector at station 10) to 10 metres ($AD1$ state vector at station 19).

Interpretation of the estimation of the optical attenuation coefficient in terms of the assimilated surface temperature data

In order to better understand the effect of the temperature assimilation on the optical attenuation coefficient, we examine time series of the surface temperature of the assimilated data, and of the $AD0$, $AD1$ and $AD3$ surface temperature estimations. Results for stations 10 and 19 are presented on Figure 7.5.

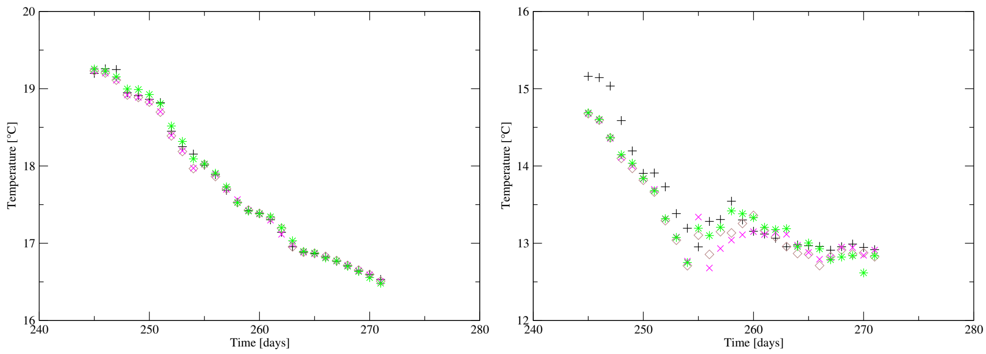

Figure 7.5: Time series of surface temperature at stations 10 (left) and 19 (right). The black plusses represent the assimilated data, the brown diamonds the $AD0$ state vector ('NS-11'), the magenta crosses the $AD1$ state vector, and the green stars represent the $AD3$ state vector.

At station 10, from day 255 on (12th of September) corresponding to the beginning of the assimilation process, the model results for the three state vectors virtually stick to the assimilated data (black plusses).

At station 19, during the first twelve days of the assimilation period, from day 255 until day 266 (23rd of September), results from the $AD3$ state vector (green stars) nearly coincide with the assimilated data (black plusses). For the remaining of the assimilation period, from day 267 on, results obtained with the $AD1$ state vector (magenta crosses) are slightly closer to the assimilated data (black plusses) than those derived from the $AD3$ state vector.

7.3.2 The surface heat exchange coefficients

As the assimilated temperature data were generated from simulations using the more realistic stratified formulation for the heat exchange coefficients, we examine if that "stratified" information is incorporated in the adjusted neutral formulation of the coefficients.

Estimated surface heat exchange coefficients

Model parameter estimation using the ensemble Kalman filter is further conducted by analyzing the estimation derived for the neutral formulation of the latent and sensible heat exchange coefficients.

Results correspond to experiments 'NPA-2' and 'NPA3' respectively performed with the $AD2$ and $AD3$ state vectors. With the $AD2$ state vector, the heat exchange coefficients are the sole parameters that are estimated while the $AD3$ state vector combines the simultaneous estimation of the set of chosen parameters.

Results for the latent heat exchange coefficient are presented on Figure 7.6. The left panel corresponds to station 10 while the right panel corresponds to station 19.

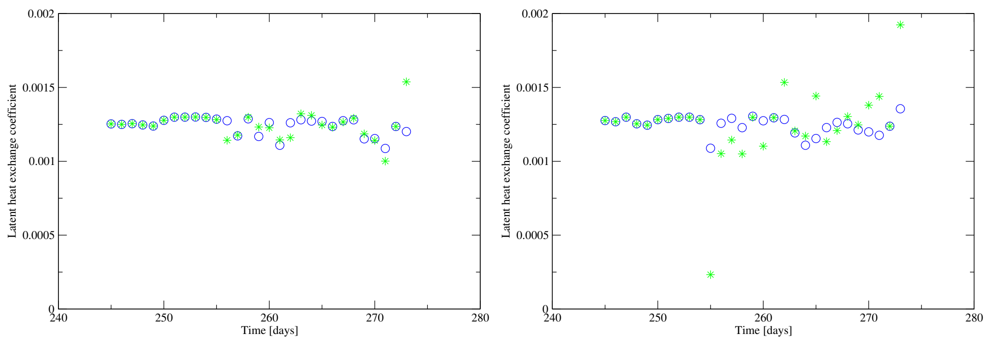

Figure 7.6: Time series of the latent heat exchange coefficient at stations 10 (left) and 19 (right), computed with the $AD2$ state vector (blue circles) and the $AD3$ state vector (green stars).

A comparison of the two panels of Figure 7.6 indicates differences in the extreme values of the coefficients from one station to the other while their mean values are about the same. The largest excursions around the mean occur at station 19 while the mean value of the analyzed coefficients is of about 0.0012 at both stations.

At both stations, the estimation of the latent heat exchange coefficient with the $AD2$ state vector (blue circles) provides values in a smaller range than those obtained with the $AD3$ state vector (green stars).

The estimation of the sensible heat exchange coefficients is presented on Figure 7.7 (left panel for station 10 and right panel for station 19).

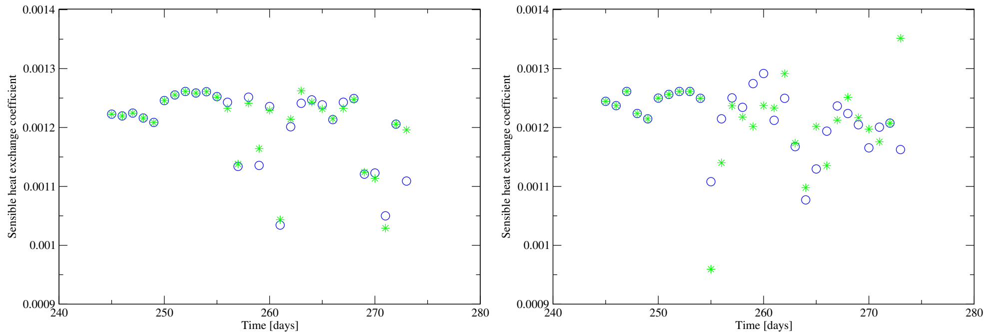

Figure 7.7: Time series of the sensible heat exchange coefficient computed with the $AD2$ state vector (blue circles) and the $AD3$ state vector (green stars) at station 10 (left) and at station 19 (right).

In the case of the sensible heat exchange coefficients, at both stations, the values resulting from the $AD2$ estimation (blue circles) are similar to those derived from the $AD3$ estimation (green stars).

The following remarks concerning the estimation of the heat exchange coefficients are worth mentioning. First of all, the estimations derived from the $AD2$ and $AD3$ state vectors are close together, especially for the sensible heat coefficient.

At a given station, the mean values of the latent and sensible coefficients are slightly different. This is attributed to different weighting coefficients for the neutral formulation of the latent and sensible heat exchange coefficients following equations (3.21) and (3.22) and given in Table 8.4. These weighting coefficients are slightly smaller for the sensible heat coefficient than for the latent heat coefficient.

Comparison of the estimations of the heat exchange coefficients to values derived from the stratified formulation

An ensemble of heat exchange coefficient is generated by adding error terms to their neutral formulation. These error terms have a horizontal correlation. As the assimilated temperature data were generated with the more realistic stratified formulation depending on the sea-air temperature difference, further interpretation of the adjusted values of the heat exchange coefficients is sought by comparison to their stratified formulation.

Results for the latent heat coefficient are presented on Figure 7.8 where station 10 is presented on the left panel and station 19 on the right panel. For the sensible heat coefficients, the same features as for the latent heat coefficients appear when the estimated values are compared to the stratified formulation, so that results for the sensible heat exchange coefficients are not shown.

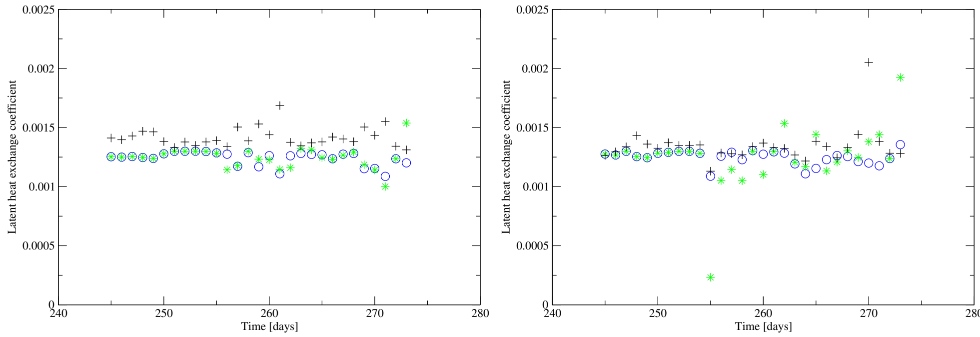

Figure 7.8: Time series of the latent heat exchange coefficient at stations 10 (left) and 19 (right), computed with the $AD2$ state vector (blue circles), the $AD3$ state vector (green stars) and the stratified formulation (black plusses).

One impact of the adjustment of the latent heat coefficients at station 10 is to increase the dispersion of their values. The mean value of the analysed neutral coefficients at station 10 remain smaller than in the stratified formulation. This can not be attributed to a too small range of perturbation allowed by the ensemble generation as the range of extreme values of the $AD3$ estimation is at least as large as that of the stratified formulation.

At station 19, the gap between the analysed values and the stratified formulation is smaller than at station 10. Except on day 269 (26th of September), there is a closer agreement between the estimated latent heat exchange coefficients ($AD2$ and $AD3$ state vectors) and the stratified formulation than at station 10.

At the well mixed station, both for the latent and sensible heat exchange coefficients, the estimation derived from the ensemble Kalman filter provides mean as well as extreme values different from those obtained from the stratified formulation. However at the stratified station, these values are sometimes close together indicating that the assimilation process corrects the neutral formulation of the heat exchange coefficients in the expected direction. This is interpreted as the fact that the assimilation process choses values for the coefficients amongst an ensemble of possible values in such a way that the model presents a higher flexibility allowing it to come closer to the true ocean state.

Interpretation of the estimation of the heat exchange coefficients in terms of the assimilated surface temperature data

In order to understand the impact of the temperature assimilation on the neutral formulation of the heat exchange coefficients, we compare time series of the surface temperature of the assimilated data and of assimilative model results derived with the $AD0$, $AD2$ and $AD3$ state vectors. The simulation performed with the $AD0$ state vector uses a stratified formulation of the heat exchange coefficients whereas in the case of the $AD2$ and $AD3$ state vectors the neutral formulation is adjusted by data assimilation.

Results are presented on Figure 7.9 at stations 10 (left) and 19 (right).

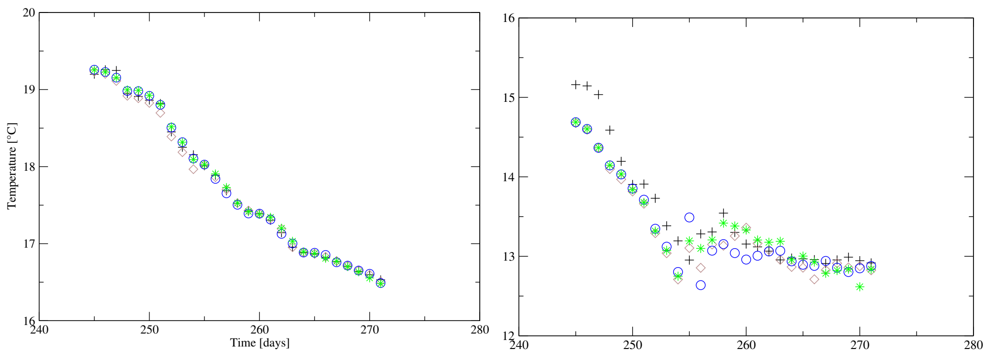

Figure 7.9: Time series of the surface temperature at stations 10 (left)and 19 (right). The black plusses correspond to the assimilated data, the brown diamonds to the temperature modelled without parameter estimation ($AD0$ state vector, 'NS-11'), the blue circles to the $AD2$ state vector and the green stars to the $AD3$ state vector.

At station 10, even without assimilation the modelled surface temperature is very close to the data, the effect of the assimilation is to have it virtually sticked to the assimilated data. At station 19, before the beginning of the assimilation process the gap between the model and data is larger than at station 10. Once the assimilation starts the model comes very close to the data.

7.4 Effect on the temperature of the combined model state and parameter estimation in the 3-D case

7.4.1 Temperature profiles

In chapter 5 dedicated to 3-D model state estimation, we have examined in detail how the ensemble Kalman filter could provide a better representation of the North Sea temperature. In this chapter we examine the combined estimation of model state and parameters. When performing this combined estimation our aim in terms of the modelled temperature is to assess if there is a benefit of the combined estimation with respect to the "sole" model state estimation and what it is. So we keep the same representation of the model error as in the case of "sole" model state estimation.

In order to highlight the benefit of the combined estimation of model state and parameters with respect to the "sole" estimation of model state, we compare results from experiment 'NS-11' performed with the $AD0$ state vector to experiments 'NPA-1', 'NPA-2' and 'NPA-3'. Experiment 'NS-11' is performed with the same ensemble parameters that are used for the model error representation in the combined model state and parameter estimation. We focus on the temperature profiles of the 28th of September 2001, the last day of assimilation, so that all the information available from the data has been incorporated to improve the model.

Figure 7.10 presents the temperature profiles on the 28th of September 2001 computed with the four estimation state vectors used in the framework of this study, $AD0$, $AD1$, $AD2$ and $AD3$. Results at well mixed station 10 are on the left panel while those of the summer stratified station 19 are on the right panel.

At station 10, the temperature profile modelled with the $AD0$ state vector, i.e. without parameter adjustment (brown diamonds) is the closest to the assimilated data (black plusses). At station 19, the temperature profiles derived with any state vector are almost identical except that there is a departure to a slightly warmer temperature at two points below the surface with the $AD1$ state vector (magenta crosses) which come closer to the assimilated data. Then, in the thermocline region around 40 metres depth, the temperature modelled with the $AD3$ state vector is very close to the assimilated data. Moreover, from 55 metres depth to the bottom, the temperature profile modelled with the $AD3$ state vector is very close to the assimilated data.

At station 19, there is a slight benefit in the temperature representation

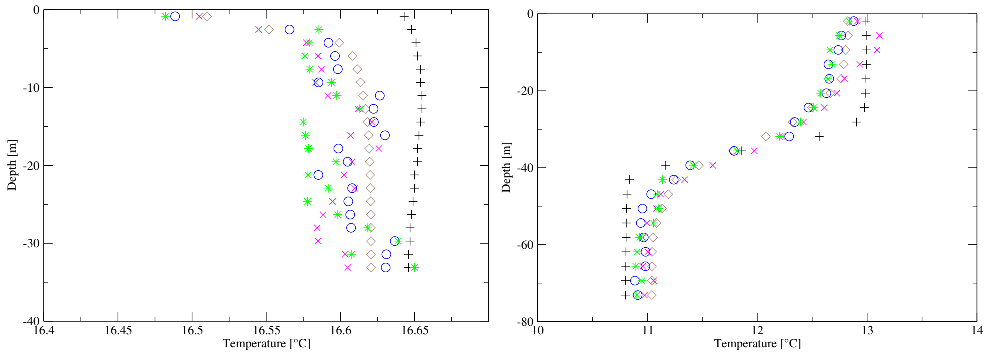

Figure 7.10: Temperature profiles on the 28th of September at stations 10 (left) and 19 (right). The brown diamonds correspond to the temperature computed with the $AD0$ state vector ('NS-11'), the magenta crosses to the $AD1$ vector, the blue circles to $AD2$, the green stars to $AD3$ and the black plusses to the assimilated data.

when the $AD1$ and $AD3$ state vectors are used. In the case of the $AD1$ state vector, the improvement takes place at three points just below the surface (magenta crosses) while with the $AD3$ state vector the improvement appears in the position of the thermocline and at the bottom of the water column. These small differences are confirmed by the values of surface root mean square errors between the model and the assimilated data given in Table 7.1. This could be related to the fact that station 19 is stratified so that the optical attenuation coefficient plays a more important role in the representation of the temperature of the upper mixed layer than at a well-mixed station.

7.4.2 Hovmöller diagrams of the near surface and bottom temperature

The temperature profiles on the 28th of September shown on Figure 7.10 indicate that at station 10 the modelled temperature profile is warmer just below the surface and then the incoming radiation is quickly diffused along the whole water column and presents a temperature profile typical of a well-mixed station (from the 3rd point below the surface on). At station 19, the temperature profiles modelled with the $AD1$ state vector (magenta crosses) are slightly different from the others at the three points just below the surface, whereas slightly better results are obtained at the bottom with the $AD3$ state vector. So we investigate time-series of the near surface temperature at station 10 and at three vertical levels below the surface at station 19. Time series of the bottom temperature are also examined at both stations.

Near surface temperature

Results are shown on Figure 7.11, the left panel corresponds to station 10 and the right panel to station 19. The water depths under consideration are 4 metres at station 10 and of 9 metres at station 19.

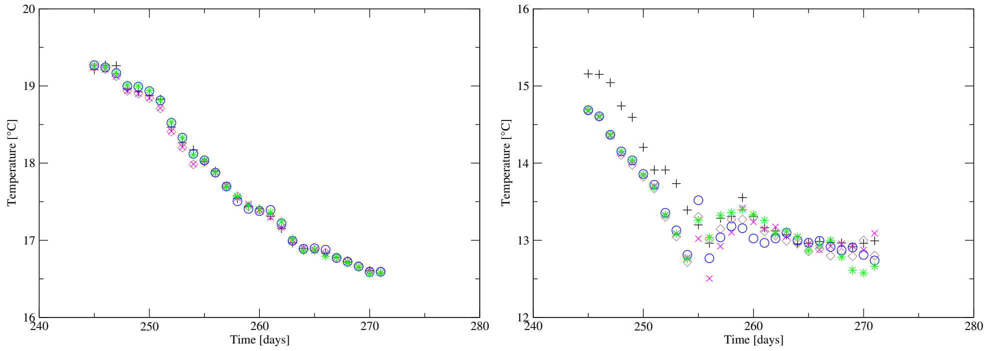

Figure 7.11: Time series of the near surface temperature at station 10, depth of 4 metres (left) and at station 19, depth of 9 metres (right). The brown diamonds correspond to the temperature computed with the $AD0$ state vector ('NS-11'), the magenta crosses to the $AD1$ vector, the blue circles to $AD2$, the green stars to $AD3$, and the black plusses to the assimilated data.

At station 10, the temperature modelled with the $AD0$ state vector (brown diamonds) is already in very good agreement with the assimilated data so that the adjustment of the model parameters has virtually no effect on the temperature representation.

At station 19, before the beginning of the assimilation, the near surface modelled temperature is clearly under-estimated. At the beginning of the assimilation period on day 255 (12th of September), the temperature modelled with the $AD2$ state vector (blue circles) remains slightly too cold until day 261 (18th of September) and on the two last days of the assimilation period (days 270 and 271, 27th and 28th of September). The results closest to the assimilated data are obtained with state vectors including the optical attenuation coefficient ($AD1$ and $AD3$, respectively magenta crosses and green stars). As already mentioned, this is attributed to the link between the optical attenuation coefficient and the representation of the temperature of the upper mixed layer at a stratified station.

Bottom temperature

Time series of the bottom temperature modelled with the default and the three augmented state vectors are shown on Figure 7.12.

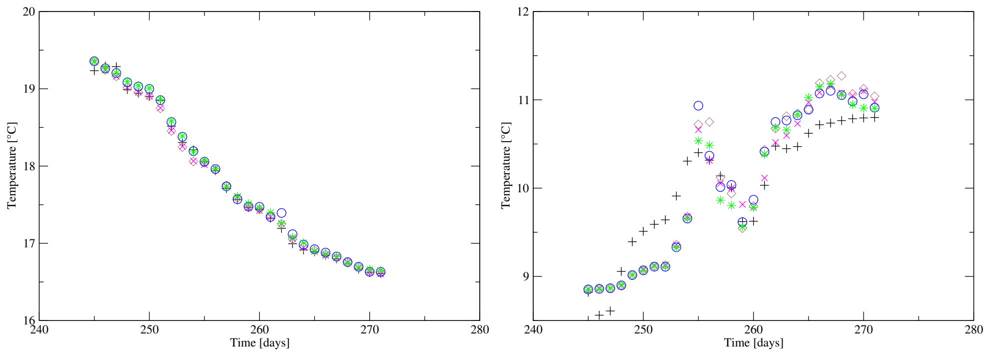

Figure 7.12: Time series of the bottom temperature at station 10 (left) and at station 19 (right). The brown diamonds correspond to the temperature computed with the $AD0$ state vector ('NS-11'), the magenta crosses to the $AD1$ vector, the blue circles to $AD2$, the green stars to $AD3$, and the black plusses to the assimilated data.

At the well-mixed station 10 there is virtually no effect of the adjustment of parameters related to surface processes on the bottom temperature except on day 262 (19th of September) where the bottom temperature modelled with the $AD2$ state vector is slightly warmer than the temperature modelled with any other state vector and warmer than the assimilated data.

At the summer stratified station 19, before the beginning of the assimilation on day 255 (12th of September), the modelled temperature under-estimates the assimilated data from day 248 until day 254. After the beginning of the assimilation the modelled temperature over-estimates the assimilated data by about the same order of magnitude as it under-estimated it before assimilating the data. As the parameters selected for adjustment are involved in the representation of the temperature of the upper mixed layer, their effect on the bottom temperature is negligible.

7.4.3 Root-mean square error between the modelled temperature and the assimilated data

In order to assess more precisely the performance of the different state vectors, we have computed the root mean square errors between the assimilated data and the temperature modelled using the three state vectors at the eight stations. The mean is computed in time from the beginning until the end of the assimilation period (12th of September 2001 until 28th of September 2001).

Root-mean square errors of the temperature profiles

The root mean square error as a function of depth for stations 10 and 19 is presented on Figure 7.13.

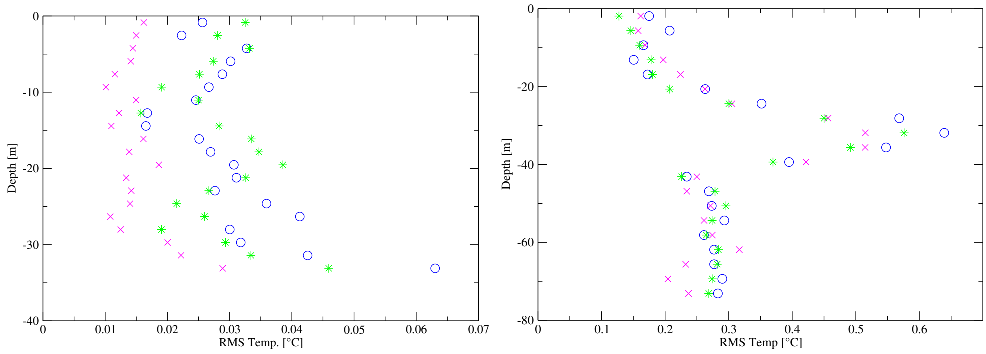

Figure 7.13: Root mean square error between the assimilated data and the modelled temperature at stations 10 (left) and 19 (right). The magenta crosses correspond to the $AD1$ state vector, the blue plusses to the $AD2$ state vector and the green stars to the $AD3$ state vector.

At station 10, the root mean square errors between the model and the assimilated data are very small, of the order of 0.02°C. The smallest errors clearly appear in the case of the $AD1$ state vector (magenta crosses). Then there are slight differences between the $AD2$ and $AD3$ state vectors: at the two first points below the surface and between 15 and 25 metres depth, the $AD2$ state vector gives smaller errors than the $AD3$ state vector while at all other depths it is the contrary.

At the well-mixed station, it was not expected that the optical attenuation coefficient would play a role ($AD1$ state vector). However the root mean square errors between the different state vectors are very small, around 0.02°C.

The right panel deals with station 19 where the differences between the model and the assimilated data are one order of magnitude larger than at station 10. At first sight the $AD3$ state vector seems to provide slightly better results than $AD1$ or $AD2$ at most water depths. The largest errors take place at the level of the thermocline with a maximum root mean square value of 0.6°C for the $AD2$ state vector.

At the stratified station, it is expected that the optical attenuation coefficient plays a larger role than at the well-mixed station. The $AD3$ state vector allows to adjust the main parameters involved in the temperature of the upper mixed layer.

Root mean square errors of the surface temperatures

In order to get a better idea of the performance of the three state vectors, surface results at each station are presented in Table 7.1. The results from the $AD0$ state vector are based on a stratified formulation of the heat exchange coefficients whereas with the $AD2$ and $AD3$ state vectors their neutral formulation is adjusted by the assimilation process. The smallest values are highlighted in bold.

Stat.	AD0 [$°C$]	AD1 [$°C$]	AD2 [$°C$]	AD3 [$°C$]
St. 6	0.07	0.04	**0.03**	0.04
St. 7	0.04	**0.03**	0.04	0.05
St. 9	0.02	**0.02**	0.05	0.05
St. 10	0.17	**0.02**	0.03	0.03
St. 11	0.16	0.13	0.14	**0.12**
St. 13	0.10	0.11	0.10	**0.09**
St. 15	0.14	**0.10**	0.17	0.12
St. 19	0.17	0.16	0.18	**0.13**

Table 7.1: Surface root mean square errors between the assimilated data and the temperature modelled with the default ($AD0$) and the three augmented state vectors ($AD1$, $AD2$, $AD3$).

At the surface there is not a clear benefit from one state vector with respect to another. However, at almost all stations, a slightly better agreement between the modelled temperature and the assimilated data is obtained using augmented state vectors comprising the optical attenuation coefficient. This suggests that the default value of this coefficient is not totally adequate and that it played a more important role than the heat exchange coefficients in the misrepresentation of the temperature of the upper mixed layer.

The results from table 7.1 are plotted on a chart (Figure 7.14). Values for the model runs without data assimilation are also presented.

The assimilation process clearly improves the representation of the surface temperature, with root mean square errors of 0.3°C reduced to less than 0.1°C (e.g. stations 6 and 10). At stations 6, 7, 9 and 10, the root mean square error between the model results with "sole" model state estimation ($AD0$ in brown)

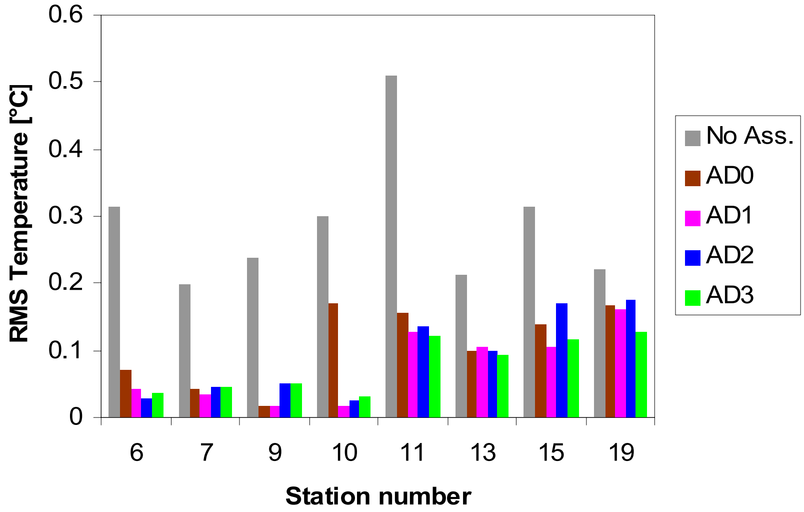

Figure 7.14: Root mean square error between the modelled temperature and the assimilated data at the surface for each station. Model results without assimilation are represented in grey, with the $AD0$ state vector in brown, with the $AD1$ state vector in magenta, with the $AD2$ state vector in blue and with the $AD3$ state vector in green.

and the assimilated data is reduced to less than 0.1°C. At these stations, there is hardly room left for further improvement by parameter adjustment. In that case, estimating the parameters can even slightly degrade the surface temperature estimation (at stations 9 and 10).

On the other hand, at stations 11, 13, 15 and 19, the root mean square errors after "sole" model state estimation are comprised between 0.1°C and 0.2°C . At these stations, the combined model state and parameter estimation is able to provide a slight further improvement of the surface temperature. Though the differences are quite small, the smallest root mean square errors are obtained in most cases with the $AD3$ model state vector involving the set of selected parameters.

Root mean square errors of the bottom temperatures

The root mean square errors between the modelled temperature and the assimilated data are presented in Table 7.2.

Stat.	AD0 [$°C$]	AD1 [$°C$]	AD2 [$°C$]	AD3 [$°C$]
St. 6	0.10	0.10	0.08	**0.08**
St. 7	0.05	0.05	**0.04**	0.05
St. 9	0.02	**0.02**	0.04	0.03
St. 10	0.03	**0.03**	0.06	0.05
St. 11	0.09	**0.09**	0.10	0.10
St. 13	0.17	**0.09**	0.11	0.17
St. 15	**0.17**	0.47	0.51	0.24
St. 19	0.33	**0.24**	0.28	0.27

Table 7.2: Bottom root mean square errors between the assimilated data and the temperature modelled with the default ($AD0$) and the three augmented state vectors ($AD1$, $AD2$, $AD3$).

At the bottom the root mean square errors between the modelled temperature and the assimilated data does not show a benefit of the combined model state and parameter estimation ($AD1$, $AD2$, $AD3$) with respect to the "sole" model state estimation ($AD0$).

Results from 3-D simulations with "sole" state estimation ($AD0$ state vectors) show that the improvement of the modelled temperature obtained from the temperature assimilation without estimation of the model parameters is significant. As one of the objective of combined model state and parameter estimation was to assess the benefit that could be gained with respect to the "sole" model stat estimation we have kept the same representation of the model error so that there was not much room left for an improvement of the model state by the adjustment of the parameters. Moreover, the selected parameters are involved in the representation of the upper mixed layer so that their effect on the bottom temperature is negligible.

The results from table 7.2 are plotted on a chart (Figure 7.15), complemented by the results without data assimilation.

As at the surface, the assimilation process reduces strongly the root mean square error between the assimilated data and the bottom temperature, especially at station 15 where the discrepancy is reduced from around 0.86°C to 0.17°C. At well mixed stations 6, 7, 9 and 10, there is no room left for further improvement by parameter estimation as the root mean square differences are less than 0.1°C. At station 15, parameter estimation degrades the

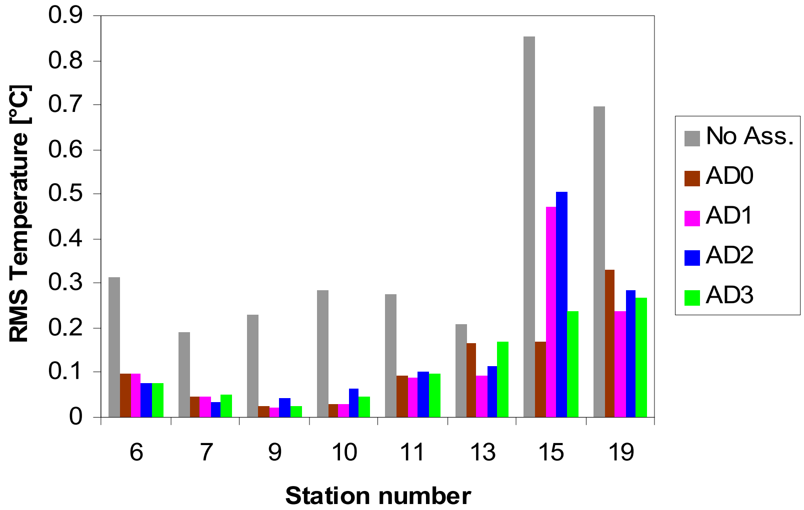

Figure 7.15: Root mean square error between the modelled temperature and the assimilated data at the bottom for each station. Model results without assimilation are represented in grey, with the $AD0$ state vector in brown, with the $AD1$ state vector in magenta, with the $AD2$ state vector in blue and with the $AD3$ state vector in green.

temperature representation whereas at stations 11, 13 and 19 the smallest errors appear for the $AD1$ state vector which indicates that the surface heat exchange coefficients do not help to improve the bottom temperature.

Final remarks on the root mean square errors between the modelled temperatures and the data

The root mean square errors between the modelled temperature and the assimilated data indicate that the "sole" temperature estimation (without simultaneous adjustment of the parameters) strongly reduces these discrepancies. There are stations (6-7-9-10) where there is almost no room left from further improvement, with remaining discrepancies of about 0.01°C. At stations where a further improvement is possible (11-13-15-19), results obtained with a simultaneous adjustment of all the parameters with the model state are slightly better ($AD3$ state vector).

Furthermore, in these 3-D simulations, the assimilated data are not stemming from observations but from model runs with a higher horizontal resolution. As a result, no information from processes not represented by the model is incorporated.

7.5 Conclusions of combined model state and parameter estimation in the 3-D case

This chapter was dedicated to combined model state and parameter estimation in 3-D simulations of the North Sea domain. The method is based on the state vector augmentation. The estimated parameters are those primarily involved in the representation of the temperature of the upper mixed layer: the optical attenuation and surface heat exchange coefficients. Three augmented state vectors are considered. The first one is augmented with the optical attenuation coefficient, the second one is augmented with the surface heat exchange coefficients while the third one contains the set of selected parameters. The adjusted heat exchange coefficient follow the neutral Kondo formulation and are compared to the stratified Kondo formulation considered as more realistic and used to generate the assimilated data.

As the area covered by the model presents a well-mixed regime and a summer stratified regime, the main objective of the chapter was to investigate the spatial variability of the estimation of the parameters. Temperature profiles were assimilated in the well mixed zone (Southern and German Bights) and at a summer stratified station in the central part of the domain. Results were examined at a station representative of the well mixed regime (station 10) and at the summer stratified station 19.

Results of the adjustment of the optical attenuation coefficient indicate that larger variations of the coefficient are observed in the stratified zone than in the well-mixed area. As the ensemble for the optical attenuation coefficient is generated from a good initial guess with constant error terms the estimated values have a smaller range of variation at the end of the assimilation period. The tendencies in the estimated optical attenuation coefficient are in opposite direction for the $AD1$ and $AD3$ state vectors. Further interpretation of the adjusted coefficients in terms of the assimilated surface data indicates that the $AD1$ state vector provides results slightly more realistic than the $AD3$ state vector. In that case the optical attenuation length stabilizes to around 17 metres depth in the well mixed area while 10 metres are adopted in the summer stratified zone.

The formulations of the latent and sensible heat coefficients present the same temperature dependency so that the temperature assimilation has nearly the same effect on both of them and our study focused on the latent heat coefficient. The impact of parameter adjustment on the latent heat coefficients is to increase the dispersion of the values. At the well mixed station there is no clear correlation between the adjusted neutral formulation of the coefficients and their stratified formulation. In the stratified zone there is a better agreement between the adjusted values and the ones derived from the stratified formulation.

The improvement in temperature representation from combined estimation of model state and parameters with respect to "sole" model state estimation has also been investigated. As the aim was to highlight the benefit of the combined model state and parameter estimation we have kept the same error representation of the temperature as in the "sole" model state estimation. Moreover, the "sole" model state estimation is performed with the stratified formulation of the heat exchange coefficients. At some stations, the impact of the combined estimation of parameters and model state with respect to the "sole" estimation of model state is relatively small. This is related to the fact that, at these stations, after model state estimation, there was not much room left for further improvement (of the order of 0.01°C). Moreover, the assimilated data are generated from model runs. As a result, the assimilation process does not incorporate information from dynamical processes that are absent from model runs and that could be present in "real" observations.

At the well mixed station the assimilation improves the temperature in such a way that there is hardly room for further improvement by parameter estimation. At the summer stratified station, there is a slight benefit on the temperature representation from the model state vectors including the optical attenuation coefficient ($AD1$ and $AD3$ state vectors). From root mean square errors between the assimilated data and the temperature modelled at station 19, the $AD3$ state vector might be considered as slightly better than the $AD1$ state vector, with root mean square errors of 0.13°C with the $AD3$ state vector in place of 0.16°C with the $AD1$ state vector.

These preliminary results have shown that the ensemble Kalman filter allows to estimate model parameters in realistic North Sea simulations. They also indicate that the methodology can be further applied to the estimation of other physical and ecological parameters.

Chapter 8

Conclusions and perspectives

8.1 Conclusions

Data assimilation allows combining information from available observations and a dynamical model to refine the model representation of the processes. There are several motivations for assimilating data in shallow water models: the improvement of operational forecasts and the provision of associated measures of uncertainty as well as the estimation of model parameters and the design of observations networks. Because of the wide range of spatial and temporal scales of coastal processes, assimilating data in coastal models remains challenging. Moreover one region can be dominated by one type of forcing and a neighboring one by a completely different type of dynamical regime.

Starting from the COHERENS model and two square root algorithms that were already implemented, the work presented here examines how a sequential data assimilation method such as the ensemble Kalman filter can be applied to North Sea simulations. It focuses on the sampling of the model error, on model state estimation and on combined model state and parameter estimation. The first square root algorithm considers data as perfect whereas the second one (the low rank square root algorithm) allows an ensemble representation of the observations' errors.

Results are obtained for a 1-D case study and 3-D simulations, each with 50 ensemble members. The 3-D simulations are performed with a horizontal resolution of 4 nautical miles. As model error sampling is critical in ensemble Kalman filter applications, several numerical experiments were conducted in order to derive adequate values for the sampling parameters such as the number of ensemble members, the horizontal and vertical decorrelation scales as well as the standard deviation of the members with respect to the ensemble mean. In 3-D simulations, a cut-off radius is needed to limit the influence zone of the assimilated data.

Because temperature is one of the most important ocean parameters, both for physical and biogeochemical processes, temperature profiles are assimilated. In the 1-D test case, they are extracted from the 'North Sea Project' data base while for the North Sea simulations they are provided from a model run with the same setup but performed with a higher horizontal resolution. The numerical experiments reported here focus on temperature assimilation during a summer month so that the southern part of the North Sea is well-mixed and the central part is thermally stratified.

Data assimilation allows to improve the model state representation and to estimate model parameters. Both approaches are investigated. In a first step the estimation focuses merely on the model state estimation. Then the state vector is augmented with parameters primarily related to the representation of the temperature in the upper mixed layer, the optical attenuation and latent and sensible heat exchange coefficients. A neutral formulation of the heat exchange coefficients is corrected by the assimilation process and is compared to a formulation in terms of the sea-air temperature difference that is considered as reference. The benefits of combined model state and parameter estimation for the model state representation are assessed. Furthermore, the existence of the two North Sea regimes offers the opportunity to investigate the spatial variability of the adjusted model parameters from a well mixed region to a stratified zone.

The 1-D test case is performed at one station of the 'North Sea Project' surveys referred to as the CS station. Results from model state estimation in the 1-D test case indicate that the assimilation process allows to represent observations' features absent in the model. They mainly consist of a better position of the thermocline and a partial representation of internal oscillations. Moreover, some of the corrections brought by assimilation stay incorporated in the model when the assimilation is stopped.

In the case of combined model state and parameter estimation at the CS station, results are in better agreement with the assimilated observations when the set of selected model parameters are simultaneously adjusted than when the model parameters are adjusted separately.

The simulations on the whole North Sea domain indicate that a good configuration for model runs consists of 50 ensemble members combined with an assimilation cut-off radius of 50 km. The impact of the vertical correlation of the model error terms has been quantitatively assessed. Though they do not strongly affect the results, vertically correlated model error terms provide slightly better temperature profiles than non correlated model errors, in both well mixed and stratified areas. Some of the temperature profiles are compared to independent observations. Model results obtained with the square root algorithm are closer to the assimilated data while the low rank square root algorithm provides in some cases model results closer to the independent observations.

Results from combined model state and parameter estimation for the 3-D simulations indicate that the estimated values of the optical attenuation coefficient have a smaller range of variation at the end of the assimilation period. The estimation of the optical attenuation coefficient suggests that results are slightly better when the optical attenuation coefficient is the sole paramater simultaneously adjusted with the model state. In that case the optical attenuation length stabilizes to around 17 metres depth in the well mixed area (nearly the default value), while 10 metres are adopted in the summer stratified zone. Concerning the heat exchange coefficients, as the assimilated temperature data are derived from a model run with a stratified formulation of the coefficients, we compare their adjusted neutral formulation to their stratified formulation. At the well mixed station, the assimilation process increases the dispersion of the values. In the stratified zone, there is a good correlation between the adjusted coefficients and their stratified formulation. At the well mixed stations the "sole" model state estimation does not leave room for further improvement of the temperature by parameter estimation whereas at the summer stratified station the near surface temperature and position of the thermocline are slightly improved when the set of selected parameters are simultaneously estimated.

More generally, the improvements in the results obtained from the 1-D and 3-D simulations of the North Sea indicate that the ensemble Kalman filter is an adequate method for data assimilation in a North Sea model. They also provide detailed information on the specific features of model error sampling required for a successful North Sea application. The temperature modelled with data assimilation is closer to the observations, in particular at the thermocline with root mean square errors reduced from about 1.5°C to 0.5°C. Its surface value can be slightly further improved when a set of model parameters related to upper mixed layer processes are estimated in combination with the model state (with root mean square errors reduced from 0.17°C to 0.13°C at the summer stratified station) .

8.2 Perspectives

The work presented here has examined how the ensemble Kalman filter could be applied for the assimilation of temperature profiles in a North Sea model. It also highlighted the benefit of such an assimilation on the temperature representation and investigated the spatial variability of the estimation of a set of model parameters primarily related to the temperature of the upper mixed layer. However many questions and perspectives of work remain open.

First of all, the assimilation of satellite data of sea surface temperature could be integrated in the operational version of the COHERENS model to improve the temperature forecasts for the North Sea provided by MUMM (Management Unit of the North Sea Mathematical Models).

Another application of data assimilation is the assessment of networks of observations and the performance of OSSEs (Observing Systems Simulation Experiments). The benefit on the temperature representation from different networks could be examined and a criterium to assess their adequacy (location, density of measurement points) could be determined.

In the 3-D simulations, a cut-off radius is used to limit the impact of an assimilated data to points dynamically related. However, this method could be refined by using a smoother assimilation radius.

The model error representation could be adapted depending on the model features needing improvement. For example, if a more sharp representation of the thermocline is sought, the error term used for the ensemble generation could be based on the vertical gradient of temperature rather than on its profile.

The Kalman filter provides an optimal state if the model noise is white. However, it is robust with respect to the definition of errors and, even in the case of biased model error, it is able to correct the model efficiently. General speaking, the use of coloured (time correlated) model noise allows a better performance of the filter provided the state vector augmentation technique including the model noise is applied. The benefit of considering time correlated model error could be investigated and its impact on the filter performance could be assessed.

Then, model parameter estimation opens many perspectives. Here, we focused on the spatial variability of the estimation of model parameters. Another interesting aspect would be to investigate the temporal variability of their estimation by comparing a winter to a summer situation (e.g. January and August), in particular for heat exchange coefficients that are functions of the wind speed. Moreover, we focused on the estimation of a selected set of parameters related to the temperature of the upper mixed layer. The surface drag coefficient involved in the exchange of momentum at the sea surface could be added to this set of parameters and its role could also be examined. Furthermore, since the heat flux is calculated using the modelled sea surface temperature, there is an indirect influence of the turbulence scheme on the surface forcing and on the depth mean temperature [Luyten et al., 2003] [63]. Therefore, a further application of parameter estimation by assimilation of temperature data is to adjust the coefficients parameterizing the turbulence. Finally, the adjustment of ecological parameters offers a wide range of application for data assimilation in coastal models. To this aim, the assimilation of sea surface temperature data could be investigated. As ecological models deal with concentrations and model parameters take positive values, the use of a log-normal distribution for parameter error sampling would be more relevant [Annan, 2005] [5]. This kind of distribution should be implemented in the model in the next steps of combined model state and parameter estimation.

Another important source of model error that is not investigated in this study stems from the forcings, taking them into account could further improve the model forecasts. A multivariate analysis is applied to variables dynamically correlated i.e. temperature, salinity and horizontal currents. However, the modifications in their structure has not been investigated here. Their study would help to further understand the impact of data assimilation on the model dynamics.

Appendices

Appendix A - Symbols for the description of the ensemble Kalman filter

Symbol	Description	Dimension
n	Dimension of the model state vector	-
m	Dimension of the observations vector	-
N	Number of ensemble members	-
d	Observations vector	$m \times 1$
ψ^{tr}	True state of the system	$n \times 1$
ψ^{f}	Forecast state of the system	$n \times 1$
ψ^{a}	Analysed state of the system	$n \times 1$
M	Model integration operator	$n \times n$
H	Observations operator	$m \times n$
R	Observations error covariance matrix	$m \times m$
P	Model error covariance matrix	$n \times n$
P^f	Forecast state error covariance matrix	$n \times n$
P^a	Analysed state error covariance matrix	$n \times n$
PH^T	-	$n \times m$
HPH^T	Mapping operator of the model error covariance matrix to the observations space	$m \times m$
K	Kalman gain matrix	$n \times m$
A	Matrix of model ensemble members	$n \times N$
$\bar{A}$	Mean of the model ensemble members	$n \times N$
A'	Ensemble of model perturbations matrix	$n \times N$
D	Observations ensemble members matrix	$m \times N$
E	Ensemble of observations perturbations matrix	$m \times N$

Table 8.1: Symbols for the description of the ensemble Kalman filter.

Appendix B - Equation of state for the density

McDougall et al., [2003] [69] propose an equation of state using the potential temperature. With a precision of 0.003 $kg.m^{-3}$ the density is given by:

$$\rho(S,T,p) = P_1(S,T,p)/P_2(S,T,p)$$

$$\begin{aligned} P_1 = a_0 \quad & +a_1T + a_2T^2 + a_3T^3 + a_4S + a_5ST + a_6S^2 \\ & +a_7p + a_8pT^2 + a_9pS + a_{10}p^2 + a_{11}p^2T^2 \end{aligned}$$

$$\begin{aligned} P_1 = 1 \quad & +b_1T + b_2T^2 + b_3T^3 + b_4T^4 + b_5S + b_6ST + b_7ST^3 \\ & +b_8S^{3/2} + b_9S^{3/2}T^2 + b_{10}p + b_{11}p^2T^3 + b_{12}p^3T \end{aligned}$$

where a_i and b_i are empirical parameters given in Table 8.2 below.

a_0	999.843699	b_1	$7.28606739 \times 10^{-3}$
a_1	7.3521284	b_2	$-4.60835542 \times 10^{-5}$
a_2	$-5.45928211 \times 10^{-2}$	b_3	$3.68390573 \times 10^{-7}$
a_3	$3.98476704 \times 10^{-4}$	b_4	$1.80809186 \times 10^{-10}$
a_4	2.96938239	b_5	$2.14691708 \times 10^{-3}$
a_5	$-7.23268813 \times 10^{-3}$	b_6	$-9.27062484 \times 10^{-6}$
a_6	$2.12382341 \times 10^{-3}$	b_7	$-1.78343643 \times 10^{-10}$
a_7	$1.04004591 \times 10^{-2}$	b_8	$4.76534122 \times 10^{-6}$
a_8	$1.03970529 \times 10^{-7}$	b_9	$1.63410736 \times 10^{-9}$
a_9	5.1876188×10^{-6}	b_{10}	$5.30848875 \times 10^{-6}$
a_{10}	$-3.24041825 \times 10^{-8}$	b_{11}	$-3.03175128 \times 10^{-16}$
a_{11}	$-1.2386936 \times 10^{-11}$	b_{12}	$-1.27934137 \times 10^{-17}$

Table 8.2: Values of the empirical parameters in the McDougall et al., [2003] [69] equation of state.

Appendix C - Values of parameters for sea surface drag and heat exchange coefficients

	a_d	b_d	p_d
$W_{10} < 2.2$	0.0	1.08	-0.15
$2.2 \leq W_{10} < 5$	0.774	0.0858	1.0
$5 \leq W_{10} < 8$	0.867	0.0667	1.0
$8 \leq W_{10} < 25$	1.23	0.025	1.0
$25 \leq W_{10}$	0.0	0.073	1.0

Table 8.3: Empirical values of the parameters used in the Kondo, [1975] [53] formulation of the surface drag coefficient.

	a_e	b_e	c_e	p_e
$W_{10} < 2.2$	0.0	1.23	0.0	-0.16
$2.2 \leq W_{10} < 5$	0.969	0.0521	0.0	1.0
$5 \leq W_{10} < 8$	1.18	0.01	0.0	1.0
$8 \leq W_{10} < 25$	1.196	0.008	-0.0004	1.0
$25 \leq W_{10}$	1.68	-0.016	0.0	1.0
	a_h	b_h	c_h	p_h
$W_{10} < 2.2$	0.0	1.185	0.0	-0.157
$2.2 \leq W_{10} < 5$	0.927	0.0546	0.0	1.0
$5 \leq W_{10} < 8$	1.15	0.01	0.0	1.0
$8 \leq W_{10} < 25$	1.17	0.0075	-0.00045	1.0
$25 \leq W_{10}$	1.652	-0.017	0.0	1.0

Table 8.4: Empirical values of the parameters used in the Kondo, [1975] [53] formulation of surface heat exchange coefficients.

Appendix D - Numerical experiments for model error sampling and model state estimation at the North Sea CS station

N.	D.A.	Algo.	Noens	γ_T	ξ_T	η_T	Δt S.	Δt Ass.
CS-1	N	SR	50	1°C	10m	1°C	1 day	-
CS-2	Y	SR	50	1°C	10m	1°C	1 day	1 day
CS-3	Y	SR	50	1°C	10m	0.5°C	1 day	1 day
CS-4	Y	SR	50	1°C	20m	1°C	1 day	1 day
CS-5	Y	SR	50	1°C	2m	1°C	1 day	1 day
CS-6	Y	SR	50	1°C	80m	1°C	1 day	1 day
CS-7	Y	SR	10	1°C	10m	1°C	1 day	1 day
CS-8	Y	SR	200	1°C	10m	1°C	1 day	1 day
CS-9	Y	SR	50	1°C	10m	1°C	1 day	3 hours
CS-10	Y	SR	50	1°C	10m	1°C	1 day	6 hours
CS-11	Y	LRSR	50	1°C	10m	1°C	1 day	1 day
CS-12	N	LRSR	50	1°C	-	1°C	1 day	1 day
CS-13	Y	LRSR	50	1°C	-	1°C	1 day	1 day

Table 8.5: Description of the 1-D numerical experiments for model state estimation at the North Sea CS station: experiment number, assimilation of data, algorithm, number of ensemble members, γ_T: magnitude of ensemble standard deviation, ξ_T: vertical decorrelation length, η_T: magnitude of vertical standard deviation, Δt S.: sampling time step, Δt Ass.: assimilation time step.

Appendix E - Numerical experiments for model error sampling and model state estimation on the North Sea domain

N.	D.A.	Algo.	Noens	μ_T	γ_T	ξ_T	η_T	A. Rad.
NS-1	N	SR	50	10km	1°C	10m	1°C	-
NS-2	Y	SR	50	10km	1°C	10m	1°C	50km
NS-3	Y	SR	30	10km	1°C	10m	1°C	30km
NS-4	Y	SR	30	10km	1°C	10m	1°C	50km
NS-5	Y	SR	30	10km	1°C	10m	1°C	70km
NS-6	Y	SR	40	10km	1°C	10m	1°C	30km
NS-7	Y	SR	40	10km	1°C	10m	1°C	50km
NS-8	Y	SR	40	10km	1°C	10m	1°C	70km
NS-9	Y	SR	50	10km	1°C	10m	1°C	30km
NS-10	Y	SR	50	10km	1°C	10m	1°C	70km
NS-11	Y	LRSR	50	10km	1°C	10m	1°C	50km
NS-12	N	LRSR	50	10km	1°C	-	1°C	-
NS-13	Y	LRSR	50	10km	1°C	-	1°C	50km

Table 8.6: Description of the 3-D numerical experiments for model state estimation on the North Sea domain: experiment number, assimilation of data, algorithm, number of ensemble members, μ_T: horizontal decorrelation length, γ_T: magnitude of horizontal standard deviation, ξ_T: vertical decorrelation length, η_T: magnitude of vertical standard deviation, A. Rad.: assimilation radius.

Appendix F - Numerical experiments for combined model state and parameter estimation at the North Sea CS station

N.	S.V.	Algo.	γ_T	ξ_T	η_T	γ_{exch}	γ_{opt}
CPA-1	AD1	LRSR	1°C	10m	1°C	-	$0.015m^{-1}$
CPA-2	AD2	LRSR	1°C	10m	1°C	0.00015	-
CPA-3	AD3	LRSR	1°C	10m	1°C	0.00015	$0.015m^{-1}$

Table 8.7: Description of the 1-D numerical experiments for combined model state and parameter estimation at the North Sea CS station: experiment number, estimation state vector, algorithm, γ_T: magnitude of standard deviation of the temperature ensemble, ξ_T: vertical decorrelation length of temperature, η_T: vertical standard deviation of temperature, γ_{exch}: magnitude of standard deviation of the heat exchange coefficients, γ_{opt}: magnitude of standard deviation of the optical attenuation coefficient.

Appendix G - Numerical experiments for combined model state and parameter estimation on the North Sea domain

N.	S.V.	Algo.	μ_T	γ_T	ξ_T	η_T
NPA-1	AD1	LRSR	5km	1°C	10m	1°C
NPA-2	AD2	LRSR	5km	1°C	10m	1°C
NPA-3	AD3	LRSR	5km	1°C	10m	1°C

μ_{exch}	γ_{exch}	μ_{opt}	γ_{opt}
-	-	5km	$0.012m^{-1}$
5km	0.00015	-	-
5km	0.00015	5km	$0.012m^{-1}$

Table 8.8: Description of the 3-D numerical experiments for combined model state and parameter estimation on the North Sea domain: experiment number, estimation state vector, algorithm, μ_T: horizontal decorrelation length of temperature, γ_T: magnitude of horizontal standard deviation of temperature, ξ_T: vertical decorrelation length of temperature, η_T: magnitude of vertical standard deviation of temperature, μ_{exch}: horizontal decorrelation length of heat exchange coefficients, γ_{exch}: magnitude of horizontal standard deviation of heat exchange coefficient, μ_{opt}: horizontal decorrelation length of the optical attenuation coefficient, γ_{opt}: magnitude of horizontal standard deviation of the optical attenuation coefficient.

Appendix H - Methods for the analysis of results

When a set of data has a sufficiently strong tendency to cluster around a central value, it may be characterized by its first and second moments. The first moment is related to the mean of the data set $\psi_1, \cdots \psi_N$:

$$\bar{\psi} = \frac{1}{N}\sum_{i=1}^{N}\psi_i \tag{8.1}$$

which estimates the central value around which clustering occurs. Having a distribution's central value, the next most useful characteristic is its width, measuring the dispersion or variability around the mean:

$$\sigma^2 = var(\psi) = E[\psi^2] - E^2[\psi] = \frac{1}{N-1}\sum_{i=1}^{N}(\psi_i - \bar{\psi})^2, \tag{8.2}$$

or its square root, the standard deviation defined as:

$$\sigma(\psi) = \sqrt{var(\psi)}. \tag{8.3}$$

The variance of the ensemble is the averaged square spread of each ensemble member with respect to the ensemble mean:

$$\sigma_{ens}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\psi_i - \bar{\psi})^2, \tag{8.4}$$

where N represents the number of ensemble members. The model standard deviation σ_{ens} is defined as the square root of the ensemble variance. As the Kalman filter is variance-minimizing, the modelled standard deviation decreases when data are assimilated.

The averaged error between a model field T_{mod} and an observations' field T_{obs} can be represented by the temporally averaged error of model results with respect to the observations:

$$\frac{1}{n}\sum_{i=1}^{n}(T_{mod}(i) - T_{obs}(i)) = \bar{T}_{mod} - \bar{T}_{obs}, \tag{8.5}$$

where n is the number of times the model and observations fields are compared (each time observations are available).

When considering a model field T_{mod} and and an observations' field T_{obs} defined on n grid points, their spatial variability can be described by their respective standard deviations:

$$\sigma_{mod} = [\frac{1}{n}\sum_{k=1}^{n}(T_{mod}(k) - \bar{T}_{mod})^2]^{\frac{1}{2}} \tag{8.6}$$

$$\sigma_{obs} = [\frac{1}{n}\sum_{k=1}^{n}(T_{obs}(k) - \bar{T}_{obs})^2]^{\frac{1}{2}} \quad (8.7)$$

The similarity between the model and observations is given by the correlation coefficient C_{cor}. It has a maximum value of 1 when the model and observations fields vary simultaneously, indicating a good correlation:

$$C_{cor} = \frac{1}{n}\frac{\sum_{k=1}^{n}(T_{mod}(k) - \bar{T}_{mod})(T_{obs}(k) - \bar{T}_{obs})}{\sigma_{mod}\sigma_{obs}} \quad (8.8)$$

The correlation coefficient gives a measure of the ability of the model to simulate the variability of the observed data but does not indicate how far away the model forecast is from the observations [Haugen and Evensen, 2002] [40].

In order to verify if the impact of the assimilated data is correct, i.e. to measure if the model prediction and the assimilated data are correctly weighted in a statistical sense when computing the analysis, the model prediction can be compared to the assimilated data and the results will appear the better the closer they are to the assimilated data. However, an assimilation scheme should not just interpolate the observations but rather smooth them consistently with prescribed error statistics of the model and of the observations [De Mey et al.] [22].

Furthermore the assimilated data can be used consistently for verification by comparing the innovation statistics (the model-minus-observation residuals) to the predicted error statistics from the assimilation scheme. This provides insight whether the error statistics is prescribed correctly in order to properly represent the actual errors in the model [De Mey et al.] [22]. The innovation is defined as the residual between the observations and the model:

$$I_k = d_k - H\bar{\psi}_k^f, \quad (8.9)$$

the root mean square of the innovations provides a measure of the deviation between the model and the observations:

$$rms_k = \sqrt{\frac{1}{m}(d_k - H\bar{\psi}_k^f)^T(d_k - H\bar{\psi}_k^f)}, \quad (8.10)$$

where m is the number of measurements. The modelled standard deviation provides the estimate of the model error statistics:

$$\sigma_{ens} = \sqrt{\frac{1}{m}trace(HP^fH^T)}. \quad (8.11)$$

Bibliography

[1] Aksoy, A., F. Zhang, and J.W. Nielsen-Gammon: Ensemble-based simultaneous state and parameter estimation with MM5, *Geophysical Research Letters*, **33**, L12801, 2006 - doi: 10.1029/2006GL026186.

[2] Anderson, J.L.: An Ensemble Adjustment Kalman Filter for Data Assimilation, *Monthly Weather Review*, **129**, 2884-2903, 2001.

[3] Anderson, J.L.: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D*, **230**, 99-111, 2007.

[4] Annan, J.D. and J.C. Hargreaves: Efficient parameter estimation for a highly chaotic system, *Tellus*, **56A**, 520-526, 2004.

[5] Annnan, J.D., J.C. Hargreaves, N.R. Edwards, R. Marsh: Parameter estimation in an intermediate complexity Earth system model using an ensemble Kalman filter, *Ocean Modelling*, **8**, 135-154, 2005.

[6] Barth, A., A. Alvera-Azcárate, K.-W. Gurgel, J. Staneva, A. Port, J.-M. Beckers, and E.V. Stanev: Ensemble perturbation smoother for optimizing tidal conditions by assimilation of High-Frequency radar surface currents - application to the in German Bight, *Ocean Dynamics*, **6**, 161-178, 2010.

[7] Blackadar, A.K.: The vertical distribution of wind and turbulent exchange in a neutral atmosphere, *Journal of Geophysical Research*, 1962.

[8] Blumberg, A.F. and G.L. Mellor: A description of a three-dimensional coastal ocean circulation model, *In: N.S. Heaps (editor), Three-dimensional Coastal Ocean Models, Coastal and Estuarine Sciences, Vol. 4, American Geophysical Union, Washington D.C.*, 1-16, 1987.

[9] Bouttier, F. and P. Courtier: Data assimilation concepts and methods, *ECMWF Meteorological Training Course Lecture Series*, 2002.

[10] Brasseur, P., P. Bahurel, L. Bertino, F. Birol, J.-M. Brankart, N. Ferry, S. Losa, E. Remy, J. Schröter, S. Skachko, C.-E. Testut, B. Tranchant, P.J. van Leeuwen, J. Verron: Data assimilation in operational ocean forecasting systems: the MERCATOR and MERSEA developments, 2005.

[11] Brasseur, P., J. Ballabrera-Poy, J. Verron: 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, 1999.

[12] Breivik, O. and O. Saetra: Real time assimilation of HF radar currents into a coastal ocean model, *Journal of Marine Systems*, **28**, 161-182, 2001.

[13] Brent, R.P.: Algorithms for Minimization Without Derivatives, Chapter 4, *Prentice-Hall, Englewood Cliffs, N.J.*, 1973.

[14] Brusdal, K., J.M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P.J. van Leeuwen, E. Dombrovsky, J. Verron: A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, *Journal of Marine Systems*, **40-41**, 253-289, 2003.

[15] Burgers, G., P.J. van Leeuwen, G. Evensen: Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, **126**, 1719-1724, 1998.

[16] Cañizares, R., A.W. Heemink and H.J. Vested: Application of advanced data assimilation methods for the initialisation of storm surge models, *Journal of Hydraulics Research*, **36**, 655-674, 1998.

[17] Charnock, H., K.R. Dyer, J.M. Huthnance, P.S. Liss, J.H. Simpson, P.B. Tett: Understanding the North Sea System, *Cambridge University Press*, 1994.

[18] Chen, C., P. Malanotte-Rizzoli, J. Wei, R.C. Beardsley, Z. Lai, P. Xue, S. Lyu, Q. Xu, J. Qi and G.W. Cowles: Application and comparison of Kalman filters for coastal ocean problems: An experiment with FVCOM, *Journal of Geophysical Research*, **114**, C05011, 2009 - doi: 10.1029/2007JC004548.

[19] Deleersnijder, E.L.: Upwelling and upsloping in three-dimensional marine models, *Applied Mathematical Modelling*, **13**, 462-467, 1989.

[20] Deleersnijder, E. and K.G. Ruddick: A generalized vertical coordinate for 3-D marine models, *Bulletin de la société royale des sciences de Liège*, **61**, 489-502, 1992.

[21] Deleersnijder E. and P.J. Luyten: On the practical advantages of the quasi-equilibrium version of the Mellor and Yamada level 2.5 turbulence closure applied to marine modelling, *Applied Mathematical Modelling*, **18**, 1994.

[22] De Mey, P., T. Awaji, M.J. Bell, A.F. Bennett, P. Brasseur, G. Evensen, K. Haines, I. Fukumori, O. M. Smedstad, D. Stammer and A. Weaver: Approaches to Data Assimilation within GODAE.

[23] Derber, J.: A variational continuous assimilation scheme, *Monthly Weather Review*, **117**, 2437-2446, 1989.

[24] Eben, K., P. Juru, J. Resler, M. Belda, E. Pelikán, B.C. Krüger, and J. Keder: An ensemble Kalman filter for short-term forecasting of tropospheric ozone concentrations, *Quarterly Journal of the Royal Meteorological Society*, **131**, 3313-3322, 2005.

[25] Echevin, V. and P. De Mey: Horizontal and Vertical Structure of the Representer Functions for Sea Surface Measurements in a Coastal Circulation Model, *Journal of Physical Oceanography*, **30**, 2627-2635, 2000.

[26] Ehrendorfer, M. and J.J. Tribbia: Optimal Prediction of Forecast Error Covariances through Singular Vectors, *Journal of the Atmospheric Sciences*, **54**, 286-313, 1997.

[27] Eknes M., Evensen, G.: An Ensemble Kalman Filter with a 1-D marine ecosystem model, *Journal of Marine Systems*, **36**, 75-100, 2002.

[28] Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**, 10,143-10,162, 1994.

[29] Evensen, G. and P.J. van Leeuwen: Assimilation of Geosat Altimeter data for the Agulhas current using the Ensemble Kalman Filter with a quasi-geostrophic model, *Monthly Weather Review*, **124**, 85-96, 1996.

[30] Evensen, G. and P.J. van Leeuwen: An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Review*, **128**, 1852-1867, 2000.

[31] Evensen, G.: The ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343-367, 2003.

[32] Evensen, G.: Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, **54**, 539-367, 2004.

[33] Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, *Control Systems Magazine, IEEE*, **29**, 83-104, 2009.

[34] Freeman, N.G., A.M. Hale and M.B. Danard: A modified sigma equations' approach to the numerical modelling of Great Lakes hydrodynamics, *Journal of Geophysical Research*, **77**, 1050-1060, 1972.

[35] Gelb, A.: Applied optimal estimation, *MIT University press*, 1974.

[36] Ghil, M. and P. Malanotte-Rizzoli: Data assimilation in meteorology and oceanography, *Advances in Geophysics*, **33**, 141-266, 1991.

[37] Gill, A.E.: Atmosphere-Ocean dynamics, *International Geophysics Series, Academic Press, Orlando*, **30**, 1982.

[38] Hamill, T.M., J.S. Withaker and C. Snyder: Distance-dependent filtering of background covariances estimates in an ensemble Kalman filter, *Monthly Weather Review*, **129**, 2884-2903, 2001.

[39] Hanea, R.G., G.J.M. Velders, A.J. Segers, M. Verlaan and A. Heemink: A Hybrid Kalman Filter Algorithm for Large-Scale Atmospheric Chemistry Data assimilation, *Monthly Weather Review*, **135**, 140-151, 2007.

[40] Haugen, V.E. and G. Evensen: Assimilation of SLA and SST data into an OGCM for the Indian Ocean, *Ocean Dynamics*, **52**, 133-151, 2002.

[41] Heemink, A.W., K. Bolding and M. Verlaan: Storm surge forecasting using Kalman filtering, *Journal of the Meteorological Society of Japan*, **75**, 305-318, 1997.

[42] Holt, J.T., I.D. James: A simulation of the Southern North Sea in comparison with measurements from the North Sea Project. Part 1: Temperature, *Continental Shelf Research*, **19**, 1087-1112, 1999.

[43] Hossain, M.S. and W. Rodi: A turbulence model for buoyant flows and its application to vertical buoyant jets, *In W. Rodi (editor), Turbulent buoyant jets and plumes, HMT-Series, Vol. 6, Pergamon Press, Oxford*, 121-178, 1982.

[44] Hoteit, I., D.-T. Pham, G. Triantafyllou and G. Korres: A New Approximate Solution of the Optimal Nonlinear Filter for Data Assimilation in Meteorology and Oceanography, *Monthly Weather Review*, **138**, 317-334, 2008.

[45] Houtekamer, P., H.L. Mitchell: A sequential ensemble Kalman filter technique, *Monthly Weather Review*, **129**, 2776-2790, 1998.

[46] Houtekamer, P., H.L. Mitchell: Data assimilation using an ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, **129**, 123-137, 2001.

[47] Houtekamer, P.L., H.L. Mitchell: Ensemble Kalman filtering, *Quarterly Journal of the Royal Meteorological Society*, **131**, 3269-3289, 2005.

[48] Jazwinski, A., H.: Stochastic Processes and Filtering Theory, *Academic Press, New York*, 1970.

[49] Kalman, R.E.: A new approach to linear filter and prediction theory, *Journal of Basic Engineering*, **82 D**, 35-45, 1960.

[50] Kepert, J.D.: On ensemble representation of the observation-error covariance in the Ensemble Kalman Filter, *Ocean Dynamics*, **54**, 561-569, 2004 - doi: 10.1007/s10236-004-0104-9.

[51] Keppenne C.L. and M.M. Rienecker: Initial Testing of a Massively Parallel Ensemble Filter with the Poseidon isopycnal ocean general circulation model, *Monthly Weather Review*, **130**, 2951-2965, 2002.

[52] Keppenne C.L. and M.M. Rienecker: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter, *Journal of Marine Systems*, **40-41**, 363-380, 2003.

[53] Kondo, J.: Air-sea bulk transfer coefficients in diabatic conditions, *Boundary Layer Meteorology*, **9**, 91-112, 1975.

[54] Kourafalou, V.H., L.-Y. Oey, J.D. Wang and T.N. Nee: The fate of river discharge on the continental shelf. 1. Modeling the river plume and the inner shelf coastal current, *Journal of Geophysical Research*, **101**, 3415-3434, 1996.

[55] L'Ecuyer P., and Côté S.: Implementing a Random Number Package with Splitting Facilities, *ACM Transactions on Mathematical Software*, **17**, 98-111, 1991.

[56] Lenartz, F., C. Raick, K. Soetaert and M. Grégoire: Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea, *Journal of Marine Systems*, **68**, 327-348, 2007.

[57] Leeuwenburgh, O.:Assimilation of along-track altimeter data in the tropical Pacific region of a global OGCM ensemble, *Quarterly Journal of the Royal Meteorological Society*, **131**, 2455-2472, 2005.

[58] Leeuwenburgh, O.: Validation of an EnKF System for OGCM Initialization Assimilating Temperature, Salinity, and Surface Height Measurements, *Monthy Weather Review*, **125-139**, 125-, 2007.

[59] Lisaeter, K.A., G. Evensen, and S. Laxon: Assimilating synthetic CryoSat sea ice thickness in a coupled ice-ocean model, *Journal of Geophysical Research*, **12**, C07023, 2007 - doi: 10.1029/2006JC003786.

[60] Ljung, L.: Asymptotic-behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Transactions on Automatic Control*, **29**, 36-50, 1979.

[61] Lorenc, A.C.: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-VAR, *Quarterly Journal of the Royal Meteorological Society*, **129**, 3183-3203, 2003 - doi: 10.1256/qj.02.132.

[62] Luyten, P.J., E. Deleersnijder, J. Ozer and K.G. Ruddick: Presentation of a family of turbulence closure models for stratified shallow water flows and preliminary application to the Rhine outflow region, *Continental Shelf Research*, **16**, 101-130, 1996.

[63] Luyten, P.J., J.E. Jones and R. Proctor: A Numerical Study of the Long- and Short-Term Temperature Variability and Thermal Circulation in the North Sea, *Journal of Physical Oceanography*, **33**, 37-56, 2003.

[64] Luyten, P., I. Andreu-Burillo, A. Norro, S. Ponsar, and R. Proctor: A new version of the European public domain code COHERENS, *Proceedings of the Fourth International Conference on EuroGOOS*, 474-481, 2005

[65] Luyten, P.: COHERENS - A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation. Version 2.1.2 RBINS-MUMM Report, Royal Belgian Institute of Natural Sciences, 2011.

[66] Madsen, H. and R. Cañizares: Comparison of extended and ensemble Kalman filters for data assimilation in coastal area modelling, *International Journal for Numerical Methods in Fluids*, **31**, 961-981, 1999.

[67] Magri, S.: Assimilation de données dans un modèle d'écosystème marin couplé à un modèle de couche de mélange océanique de la mer Ligure, *Thèse de doctorat, Université Joseph-Fourier - Grenoble 1*, 2002.

[68] Martinsen E.A. and H. Engedahl: Implementation and testing of a lateral boundary scheme as an open boundary condition in a barotropic ocean model, *Coastal Engineering*, **11**, 603-627, 1987.

[69] McDougall, T.J., D.R. Jackett, D.G. Wright, and R. Feistel: Accurate and computationally efficient algorithms for potential temperature and density of seawater, *Journal of Atmospheric and Oceanic Technology*, **20**, 730-741, 2003.

[70] Mellor G.L. and T. Yamada: A hierarchy of turbulence closure models for planetary boundary layers, *Journal of the Atmospheric Sciences*, **31**, 1791-1806, 1974.

[71] Meng, Z. and F. Zhang: Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part II: Imperfect Model Experiments, *Monthly Weather Review*, **135**, 1403-1423, 2007.

[72] Mesinger, F. and A. Arakawa: Numerical methods used in atmospheric models, *Global Atmospheric Research Program (GARP) publication series No. 17*, **Vol. 1**, 1976.

[73] Mourre, B., P. De Mey, F. Lyard, C. Le Provost: Assimilation of sea level data over continental shelves: an ensemble method for the exploration of model errors due to uncertainties in bathymetry, *Dynamics of Atmospheres and Oceans*, **38**, 93-121, 2004.

[74] Nativk, L.J. and G. Evensen: Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1. Data assimilation experiments, *Journal of Marine Systems*, **40-41**, 127-153, 2003.

[75] Nerger, L., S. Danilov, G. Kivman, W. Hiller, and J. Schröter: Data Assimilation with the Ensemble Kalman Filter and the SEIK Filter applied to a Finite Element Model of the North Atlantic, *Journal of Marine Systems*, **65**, 288-298, 2007.

[76] Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D., Kosro, P.M.: Assimilation of surface velocity data into a primitive equation coastal model, *Journal of Geophysical Research*, **107**, 2002 - doi: 10,1029/2000JC000511.

[77] Ospar Commission: Quality Status Report 2000 - Region II Greater North Sea, 2000.

[78] Otto, L., J.T.F. Zimmerman, G.K. Furnes, M. Mork, R. Saetre and G. Becker: Physical Oceanography of the North Sea, *Netherlands Journal of Sea Research*, **26**, 161-238, 1990.

[79] Owen, A.: A three-dimensional model of the Bristol Channel, *Journal of Physical Oceanography*, **10**, 1290-1302, 1980.

[80] Parent, L., C.E. Testut, J.M. Brankart, J. Verron, P. Brasseur and L. Gourdeau: Comparative assimilation of Topex/Poseidon and ERS altimeter data and of TAO temperature data in the Tropical Pacific Ocean during 1994-1998, and the mean sea surface height issue, *Journal of Marine Systems*, **40-41**, 381-401, 2003 - doi: 10.1016/S0924-7963(03)00022-5.

[81] Pham, D.T., J. Verron and M.C. Roubaud: A singular evolutive extended Kalman filter for data assimilation in oceanography, *Journal of Marine Systems*, **16**, 323-340, 1998.

[82] Phillips, N.A. : A coordinate system having some special advantages for numerical forecasting, *Journal of Meteorology*, **14**, 184-185, 1957.

[83] Raick-Blum, C: Mathematical modelling of the Ligurian Sea ecosystem using models of different complexities. Application of a Kalman filter to improve models reliability, *Ph. D, Université de Liège*, 2006.

[84] Robinson, A.R., P.F.J. Lermusiaux and N. Quincy Sloan III: Chapter 20. Data assimilations, *The Sea*, 1998.

[85] Robinson, A.R. and P.F.J. Lermusiaux: Overview of data assimilation, *Harvard Reports in Physical/Interdisciplinary Ocean Science*, **62**, 2000.

[86] She, J., B. Amstrup, K. Borenas, E. Buch, L. Funkquist, P. Luyten and R. Proctor: ODON: Optimal Design of Observational Networks, *Scientific Report*, 2006.

[87] Simon, E., and L. Bertino: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment, *Ocean Science*, **5**, 495-510, 2009.

[88] Simpson, J.H. and J.R. Hunter: Fronts in the Irish Sea, *Nature*, **250**, 404-406, 1974.

[89] Smagorinsky, J.: General experiments with the primitive equations - I. The basic experiment, *Monthly Weather Review*, **91**, 99-164, 1963.

[90] Snyder, C. and F. Zhang: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter, *Monthly Weather Review*, **131**, 1663-1677, 2003.

[91] Song, Y. and J.W. Grizzle: The Extended Kalman Filter as a Local Asymptotic Observer for Discrete-Time Nonlinear Systems, *Journal of Mathematical Systems, Estimation and Control*, **5**, 59-78, 1995.

[92] Szunyogh, E., J. Kostelich, G. Gyarmati, D.J. Patil, B.R. Hunt, E. Kalnay, E. Ott, and J.A. Yorke: Assessing a local ensemble Kalman filter: Perfect model experiments with the national center for environmental prediction (NCEP) global model, *Tellus*, **57**, 528-545, 2005.

[93] Talagrand, O.: Assimilation of Observations, an Introduction, *Journal of the Meteorological Society of Japan*, **75**, 191-209, 1997.

[94] Testut, C.E., P. Brasseur, J.M. Brankart and J. Verron: 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, 2003 - doi: 10.1016/S0924-7963(03)00022-8.

[95] Tippett, M.K., J.L. Anderson, C.H. Bishop: Ensemble Square Root Filters, *Monthly Weather Review*, **131**, 1485-1490, 2003.

[96] van Haren, H.: Properties of vertical current shear across stratification in the North Sea, *Journal of Marine Research*, **58**, 465-491, 2000.

[97] van Leeuwen, P.J.: Comment on "Data Assimilation Using an Ensemble Kalman Filter Technique", *Monthly Weather Review*, **127**, 1374-1377, 1999.

[98] van Leeuwen, P.J.: An Ensemble Smoother with Error Estimates, *Monthly Weather Review*, **129**, 709-728, 2001.

[99] Verlaan, M. and A.W. Heemink: Tidal flow forecasting using reduced rank square root filters, *Stochastic Hydrology and Hydraulics*, **11**, 349-368, 1997.

[100] Verlaan, M.: Efficient Kalman Filtering Algorithms for Hydrodynamic models, *PhD thesis*, 1998.

[101] Verlaan, M., A. Zijderveld, H. de Vries and J. Kroos: Operational storm surge forecasting in the Netherlands: Developments in the last decade, *Philosophical Transactions of the Royal Society A*, **363**, 1441-1453, doi: 10.1098/rsta.2005.1578, 2005.

[102] Vermeulen, P.T.M. and A.W. Heemink: Model-Reduced Variational Data Assimilation, *Monthly Weather Review*, **134**, 2888-2899, 2006.

[103] Warrach, K.: Modeling the thermal stratification in the North Sea, *Journal of Marine Systems*, **14**, 151-165, 1998.

[104] Welch, G. and G. Bishop: An Introduction to the Kalman Filter, 2006.

[105] Whitaker, J.S., T.M. Hamill: Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, **130**, 19 13-1924, 2002.

[106] Whitaker, J.S., T.M. Hamill, X. Wei, Y. Song and Z. Toth: Ensemble data assimilation with the NCEP Global Forecast System, *Monthly Weather Review*, **136**, 463-482, 2008 - doi: 10.1175/2007MWR2018.1.

[107] Zhang, F., C. Snyder and J. Sun: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter, *Monthly Weather Review*, **132**, 1238-1253, 2004.

[108] Zheng, F., and J. Zhu: Balanced multivariate model errors of an intermediate coupled model for ensemble Kalman filter data assimilation, *Journal of Geophysical Research*, **113**, C07002, 2008 - doi: 10.1029/2007JC004621.