Available online at www.sciencedirect.com
SCIENCE DIRECT
Ocean Modelling 13 (2006) 255–270
Ocean Modelling
www.elsevier.com/locate/ocemod

Coupling a two-way nested primitive equation model and a statistical SST predictor of the Ligurian Sea via data assimilation

A. Barth [a,*], A. Alvera-Azcárate [a], J.-M. Beckers [b], M. Rixen [c]

[a] University of South Florida, College of Marine Science, St. Petersburg, FL 33701, USA
[b] University of Liege, GHER, MARE, Institut de Physique B5, Sart Tilman, 4000 Liège, Belgium
[c] NURC, Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

Received 31 August 2005; received in revised form 9 February 2006; accepted 9 February 2006
Available online 7 March 2006

Abstract

A primitive equation model and a statistical predictor are coupled by data assimilation in order to combine the strength of both approaches. In this work, the system of two-way nested models centred in the Ligurian Sea and the satellite-based ocean forecasting (SOFT) system predicting the sea surface temperature (SST) are used. The data assimilation scheme is a simplified reduced order Kalman filter based on a constant error space. The assimilation of predicted SST improves the forecast of the hydrodynamic model compared to the forecast obtained by assimilating past SST observations used by the statistical predictor. This study shows that the SST of the SOFT predictor can be used to correct atmospheric heat fluxes. Traditionally this is done by relaxing the model SST towards the climatological SST. Therefore, the assimilation of SOFT SST and climatological SST are also compared.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Data assimilation; Two-way nested model; Reduced-rank Kalman filter; Ligurian Sea

1. Introduction

Data assimilation is traditionally used to combine model dynamics and observations in a statistically optimal way (e.g. Ghil and Malanotte-Rizzoli, 1991). Therefore assimilation of observations improves hindcasts and nowcasts of the ocean state than otherwise obtained by the model alone. Observational constraints are necessary to reduce uncertainties and imperfections of the ocean model. Due to the obvious lack of future observations, the model forecast cannot be controlled by observations and the predictive skill degrades as the forecast lead time increases. The error growth is not only caused by the chaotic nature of the system but also by the biases and drifts of the model. The latter contribution to the error can be reduced by considering different

* Corresponding author. Tel.: +1 727 553 3508; fax: +1 727 553 1189.
E-mail address: abarth@marine.usf.edu (A. Barth).

1463-5003/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2006.02.003

models with different imperfections (and different strengths). Data assimilation provides the statistical frame for merging different model results.

Statistical predictors have been applied in ocean forecasting as an alternative or complementary approach to physical models. Various applications of statistical predictors have been tested and numerous implementations have been applied to the El Niño Southern Oscillation (ENSO). To predict the state of the system the analogue prediction methods (e.g. Sugihara and May, 1990) look for the similar past states, the so-called analogues. The evolution is inferred from the evolution of the analogues. Based on analogues, Drosdowsky (1994) predicted the Southern Oscillation Index time series. Van den Dool (1994) used constructed analogues to forecast surface weather over the US from monthly 700 mb height. The approach of Keppenne and Ghil (1992) relies on a maximum entropy method to predict the Southern Oscillation Index. A linear inverse model was also used to predict sea surface temperature (SST) anomalies in the tropical Indo-Pacific region (Penland and Magorian, 1993; Penland and Matrosova, 2001). Zhang et al. (1993) proposed an empirical orthogonal function (EOF) iteration scheme to predict El Niño 3 SST anomalies and the Southern Oscillation Index. Another class of statistical predictor is based on a canonical correlation analysis. This method was also used to predict the 3-month mean SST in several regions of the tropical Pacific and Indian oceans (Barnston and Ropelewski, 1992), total precipitation for island stations in the tropical Pacific (He and Barnston, 1996) and temperature and precipitation in Hawaii and Alaska (Barnston and He, 1996). Knaff and Landsea (1997) developed an ENSO statistical prediction method, which is based on the optimal combination of persistence, month-to-month trend of initial conditions, and climatology.

Statistical models of the atmosphere have also been coupled to ocean general circulation models. These systems are often referred to hybrid coupled models (Neelin, 1989; Neelin, 1990). Several hybrid coupled models are used to predict the ENSO (e.g. Latif and Flügel, 1991; Graham et al., 1992; Barnett et al., 1993; Balmaseda et al., 1994). More recent ENSO prediction systems are based on physical models for both, the ocean and the atmosphere (e.g. Ji et al., 1994; Kirtman et al., 1997; Yu et al., 2002).

This paper describes the implementation of a hybrid coupled model with a primitive equation model of the ocean. The major difference with previous approaches is that the statistical predictor forecasts SST instead of the state of the atmosphere. Therefore, the coupling between the statistical predictor and the ocean model can not be realised through air–sea fluxes. Instead, the statistical SST model is connected via data assimilation to the ocean model.

SST is an important model parameter impacting and causing several processes such as ocean–atmosphere heat transfer, surface mixing controlled by stratification, and deep water formation through buoyancy changes. Accurate SST is crucial to successfully model these processes which depend on SST. Data assimilation that improves SST will have a direct effect on these processes. SST changes are also the consequence of several ocean processes such as the dynamics of the diurnal and seasonal thermocline, coastal up- and downwelling and, like any tracer, the flow of the ocean surface layer. In principle, SST information can also improve these processes, but they require the knowledge of the covariance of SST and the state of ocean and atmosphere at previous times. Only the Kalman Smoother and 4D-Var data assimilation take these covariances into account.

In this paper, we will first introduce the statistical satellite-based ocean forecasting (SOFT) predictor developed by Álvarez et al. (2004). The applied methodology as well as the accuracy of its forecast will be briefly discussed. Then the SST predicted by the SOFT system will be assimilated into the GHER (GeoHydrodynamics and Environment Research) ocean model. The skill of this hybrid modelling system composed of the GHER primitive equation (PE) model and the statistical predictor will be assessed. Finally, we will discuss the results and the conclusions in the perspective of operational forecasts.

2. Models and data assimilation

2.1. GHER model

The GHER model (Beckers, 1991) is used for the present doubly two-way nesting implementation. The free surface, hydrostatic, primitive equations under the classical Boussinesq and β-plane approximations are

solved. The surface momentum and heat fluxes are computed by the formulae using ECMWF wind, air temperature, dew point temperature, cloud coverage and sea level pressure.

The domain is discretised horizontally along parallels and meridional lines. In the vertical, the model uses a double-sigma coordinate (Beckers, 1991; Beckers et al., 2002). The two-way nesting system is composed of three models as described in Barth et al. (2005). A coarse grid primitive equation model of 1/4° resolution is implemented covering the whole Mediterranean Sea. Within this grid a 1/20° resolution model of the Liguro–Provençal basin and the northern part of the Tyrrhenian Sea is embedded. A third, fine resolution model of 1/60° is nested in the latter one and simulates the dynamics of the Ligurian Sea (Fig. 1). The boundary values of the parent model are applied as Dirichlet (clamped) boundary conditions. A sponge layer within the nested model reduces numerical noise at the boundary. At the Strait of Gibraltar, the climatological transport, temperature and salinity are imposed.

2.2. Statistical model

The statistical model of Álvarez et al. (2004) is applicable to ocean properties that are measured at a high temporal and spatial coverage. Satellite sensors meet these coverage requirements. The statistical predictors have been applied to remotely sensed surface properties such as SST and sea level anomaly. The basic characteristics of this predictor pertinent to the present work, are briefly reviewed. This statistical model is based on an EOF decomposition and on a genetic algorithm (Barth, 1992; Hernandez et al., 1995). The EOF decomposition significantly reduces the degrees of freedom in the system and therefore the number of variables to predict.

The $m \times n$ matrix $\mathbf{D}$ contains weekly SST of the Ligurian Sea from 1 March 1993 to 4 October 1999 where n is the number of time instance and m is the number of grid points representing the sea. The temporal mean SST $\bar{\mathbf{D}}$,

$$\bar{D}_i = \frac{1}{n}\sum_{j=1}^{n} D_{ij} \tag{1}$$

has been removed from the SST data to give the anomaly:

$$X_{ij} = D_{ij} - \bar{D}_i \tag{2}$$

where the indexes i and j correspond to space and time respectively. These anomalies $\mathbf{X}$ have been decomposed into a set of $r = 14$ EOFs describing the evolution of SST as a sum of time-modulated patterns:

$$X_{ij} = \sum_{k=1}^{r} A_{kj} U_{ik} + \varepsilon_{ij} \tag{3}$$

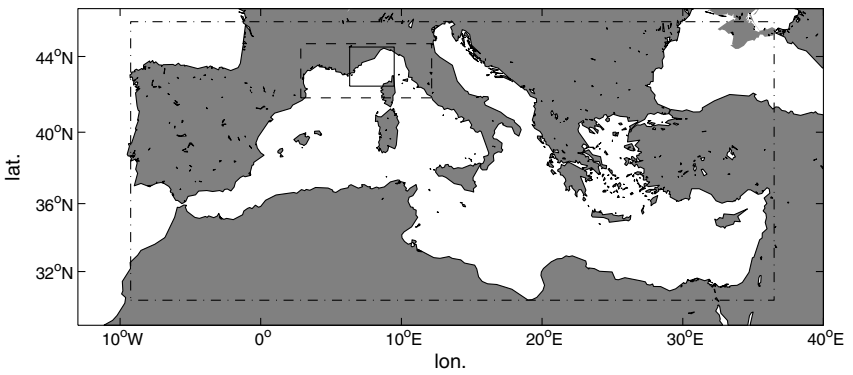

Fig. 1. Domain of the three models. The Mediterranean Sea model is delimited by the dashed-dotted line. The dashed line represents the boundary of the Liguro–Provençal model. The Ligurian Sea model lies inside the solid line rectangle.

where U_{ik} is the kth EOF, A_{kj} is the amplitude of the kth EOF and ε_{ij} is the truncation error. This error is different from zero since the r is smaller than the rank of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. It can be shown (Davis, 1976) that no other set of basis functions can produce a total error, $\|\boldsymbol{\varepsilon}\|^2 = \mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon})$, smaller than the error obtained by the EOFs.

The amplitudes of the EOFs were filtered by a singular spectrum analysis in order to extract the deterministic part of the signal. For a deterministic time series a_j, $j = 1, \ldots, n$, Takens' theorem (Takens, 1981) proves the existence of a smooth map $\gamma : \mathbb{R}^l \to \mathbb{R}$ relating the elements of the time series by

$$a_{j+1} = \gamma(a_j, a_{j-1}, a_{j-2}, \ldots, a_{j-l}) \tag{4}$$

where l is the embedding dimension. This theorem is applied to each of the 14 amplitudes time series:

$$A_{j+1k} = \gamma_k(A_{jk}, A_{j-1k}, A_{j-2k}, \ldots, A_{j-lk}) + \xi_{jk} \tag{5}$$

with $l = 8$. The embedding dimension l has been chosen to maximise the forecast skill. The error term ξ_{jk} is due to the fact that the time series may still contain noise. The DARWIN algorithm (Álvarez et al., 2000, 2001) tries to minimise the misfit ξ_{jk} over the optimisation period ranging from 1 March 1993 to 4 October 1999. The function is represented analytically implying a succession of the four binary operators $(+, -, *, /)$ and a set of constants. Starting from an initial set (a "population") of randomly chosen functions, the DARWIN algorithm tries to determine which function minimises the residual ξ by interchanging parts of the analytical expression of two members of the populating ("crossing") and randomly changing parts of a single function ("mutation").

Once the function γ fits sufficiently well the known evolution of this training data set, γ can be used to predict the EOF amplitudes. From the predicted amplitudes, the complete SST scene is reconstructed by using the spatial EOFs.

2.3. Data assimilation

The data assimilation procedure is explained in Barth et al. (in press). The state vector consists of sea surface elevation, temperature and salinity. Novel in this approach is that these variables from all embedded models are assembled into one multigrid state vector. The velocity, however, does not belong to the state vector but is also corrected by the assimilation. The velocity covariance is parameterised with the geostrophic balance linking the velocity to sea surface elevation, temperature and salinity.

The error space is estimated by an ensemble run of 200 members by perturbing the initial conditions and atmospheric forcings as described in Barth et al. (in press). The perturbations of the initial conditions are created on a grid covering the whole Mediterranean Sea. These perturbations are interpolated on the large-scale and intermediate model grids. Each member is integrated for 14 days using the Mediterranean and Provençal model. The final ensemble is interpolated on the Ligurian Sea grid.

The 50 dominant EOFs are taken as the error covariance of the model forecast. The data are assimilated using the Singular Evolutive Extended Kalman filter analysis (SEEK; Pham et al., 1998; Brasseur et al., 1999). In most assimilation experiments, long-range correlations are suppressed by a covariance localisation. This method consists of performing the assimilation in each water column independently using only nearby observations (Brankart et al., 2003).

In regions where the model domains overlap, coarse and fine resolution variables are both included in the state vector. Since the error covariance is based on the model results, this redundancy is correctly represented in the error covariance and the consistency between the grids is maintained after assimilation. More details of the implemented data assimilation procedure can be found in Barth et al. (in press).

3. Application of the SOFT predictor

Generally, operational models run in two modes: the hindcast mode and the forecast mode. They are illustrated in Fig. 2a. In the former mode, the model assimilates the available data to obtain a model state as

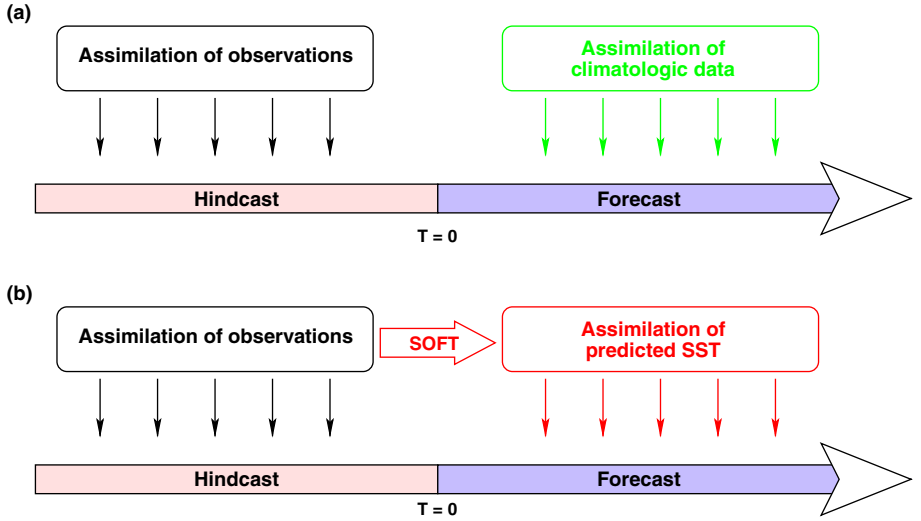

Fig. 2. Application framework for the SOFT predictor. Instead of performing an unconstrained model forecast or a model forecast with assimilation of climatological data (a), the forecast of the SOFT system can be assimilated into the operational model during the forecast (b).

accurate as possible at the end of the hindcast. When the model has reached the present time, the run turns into the forecast mode. In this phase, no observations constrain the model. Climatological data may also be assimilated in order to constrain the model integration. For example, the model SST may be nudged towards the climatological SST to correct for errors in the atmospheric heat fluxes. In the MERCATOR system, climatological data are additionally assimilated at depth when the model deviation from climatology is too large (Bahurel et al., 2004).

The climatology has some obvious drawbacks; it does not contain extreme or rare events. If the model predicts such an event, the assimilation of the climatology will attenuate it. Real and unreal events will be affected. The SOFT predictor can be integrated with such operational models as shown in Fig. 2b. The model forecast can be constrained by predicted ocean properties. Here we focus on the prediction of SST. One question that we have addressed in this work is does the assimilation of predicted SST give a more skillful forecast than the assimilation of climatological SST.

4. Assimilation experiments

In order to assess the impact of the assimilation of predicted SST, five different experiments were carried out. All the experiments start on 5 July 2000 and simulate the dynamics for 60 days. They use the same initial conditions and atmospheric forcing. The five experiments are:

`Free`: The system of nested models is run without data assimilation but with realistic initial conditions and ECMWF atmospheric forcing.

`AssimObs`: The hydrodynamic model is constrained by weekly remotely sensed SST from DLR (Deutsches Zentrum für Luft- und Raumfahrt). This model run is intended to validate and assess the capabilities of the system of nested models.

`AssimPred`: In this experiment, the hydrodynamic model assimilates the SST from the SOFT predictor. The objective is to study the model behaviour when it is constrained by the results of another model and to quantify the expected benefit of assimilating predicted SST compared with no assimilation and with the assimilation of observed SST. The model results are not completely independent of the observations since the SOFT prediction is based on observations from

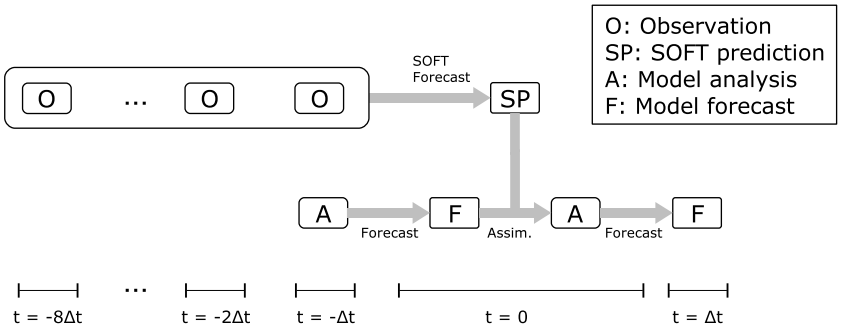

Fig. 3. Dependence of AssimPred, the SOFT predictor and observations. In particular the model analysis at $t = 0$ and the model forecast at $t = \Delta t$ depends on observations from $t = -8\Delta t$ to $t = -\Delta t$.

previous weeks. Therefore the analysis and subsequent model forecast is only influenced by observations from previous weeks and not by current observations (Fig. 3).

AssimAmpl: The predicted SOFT SST is a linear combination of the dominant spatial EOFs. This implies that the rejected EOF components always have zero amplitude. The SST of the model, however, can contain components of these rejected EOFs. When assimilating the reconstructed EOF field (AssimPred) the variance of these components is reduced by the assimilation of the predicted SST that has zero variance in this space. The approach in this experiment is to consider these components as "unobserved" by directly assimilating the predicted EOF amplitudes and not the actual SST. The observation model performs the following operation on the state vector $\mathbf{x}^f$:

$$\mathbf{a}^f = \mathbf{U}^{\mathrm{T}}\left(\mathbf{H}\mathbf{x}^f - \mathbf{y}_m^{\mathrm{SST}}\right) \tag{6}$$

where the vector $\mathbf{a}^f$ contains the EOF amplitudes predicted by the hydrodynamic model and $\mathbf{y}_m^{\mathrm{SST}}$ is the annual mean SST in the Ligurian Sea used to compute the spatial EOFs $\mathbf{U}$. The matrix $\mathbf{H}$ is the observation operator extracting the SST from the model state vector. This operator is the same as used in the two previous experiments.

AssimClim: In this experiment, the time interpolated SST from the MEDAR/Medatlas climatology (Rixen et al., 2001; MEDAR-Group, 2002) is assimilated into the model. For consistency with the other experiment only the SST in the Ligurian Sea is used.

All four assimilation experiments use the same error covariance matrix described in Barth et al. (in press). The covariance localisation was used for the experiments AssimObs, AssimPred and AssimClim. The covariance localisation technique is based on the distance between a model grid point and each single scalar observation. However, EOF amplitudes are not local observations since they involve the temperature of all surface grid points in the domain. For the assimilation of the amplitudes, the original global assimilation scheme was used.

5. The predicted SST compared to climatology and persistence

The SST predictor produced a SST forecast of the Ligurian Sea from 11 October 1999, to 28 August 2000. The overall skill assessment of the method compared to the persistence of the observations during this period was made by Álvarez et al. (2004), who concluded that the predictor is more accurate than persistence. However, it is useful to take a closer look at the performance of the predictor in the time interval of the data assimilation experiments since Álvarez et al. (2004) showed that the forecast skill of the predictor depends on the seasons. Fig. 4 shows the RMS error of the predicted SST, of the SST observed one week ago (persistence) and

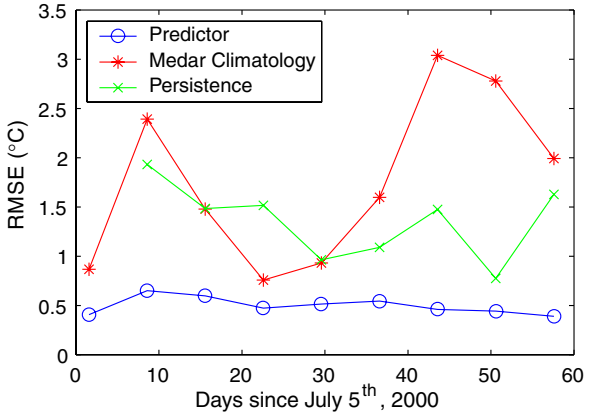

Fig. 4. The RMS error based on the observed SST of the SOFT system (circle), persistence (cross) and climatology (stat).

of the climatological SST compared to the DLR SST observations. The monthly MEDAR/Medatlas climatology is interpolated linearly to obtain weekly SST in the Ligurian Sea.

During this time period the predicted SST is substantially better than the persistence or the climatology. The RMS error of the predictor (Fig. 4) is also better than the RMS error of the hydrodynamic model (Fig. 5), which justifies the assimilation of the predicted SST.

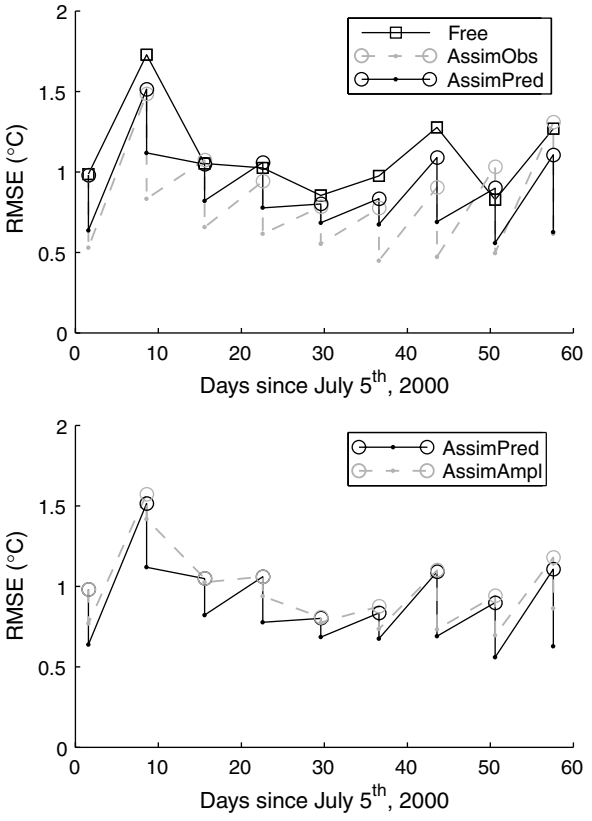

Fig. 5. RMS error between the observations and the SST of the free model run (square), the model run with assimilation of observed SST (`AssimObs`) and of the predicted SST (`AssimPred` and `AssimAmpl`). The dots correspond to the RMS error of the analysis and the empty circles to the forecast.

The climatology is too cold, especially at the end of the study period. Interannual variability has an important impact on the SST. This variability is smoothed out by a monthly climatology and causes the large discrepancy between the climatology and the observations.

6. Results

The RMS error as a function of time between the observed DLR SST and the five experiments are shown in Figs. 5 and 6. At each observation time, two estimates of the ocean state exist for the model run with assimilation: the model forecast ($\mathbf{x}^f$ in the standard notation) and the analysis ($\mathbf{x}^a$). The time averaged RMS error between the observed SST and both estimates for the four assimilation experiments and the free model integration are summarised in Table 1.

The rapid error growth after an assimilation (Fig. 4) attracted our attention. The error growth over 7 days separating two consecutive assimilation cycles is related to the heat flux formulation used in the model. The bulk flux formulation ''pulls'' the model SST towards thermal equilibrium between ocean and atmosphere since the sensible, latent and long-wave radiative heat fluxes depend on SST. Under typical July conditions for the Ligurian Sea (e.g. 75% relative humidity, 1013 HPa air pressure, 15% cloud fraction, 3 m s^{-1} wind speed, 20.5 °C SST and 21 °C air temperature), the sum of these three heat fluxes is 82 W m^{-2} (cooling). If the SST is increased by 1 °C through the assimilation (leaving the state of the atmosphere unchanged), the sum of these three heat fluxes is 136 W m^{-2}. The difference (53 W m^{-2}) corresponds to a cooling of the ocean surface layer. Since the model thermocline depth is approximately 10 m in summer, the surface temperature drops by 1 °C in only approximately 9 days (assuming the 53 W m^{-2} is constantly applied during this period). The ocean SST thus reaches its thermal equilibrium with the atmosphere in about 9 days and undoes the impact of the assimilation after this period. In this reasoning the heat transfer at the bottom of the mixed layer is neglected.

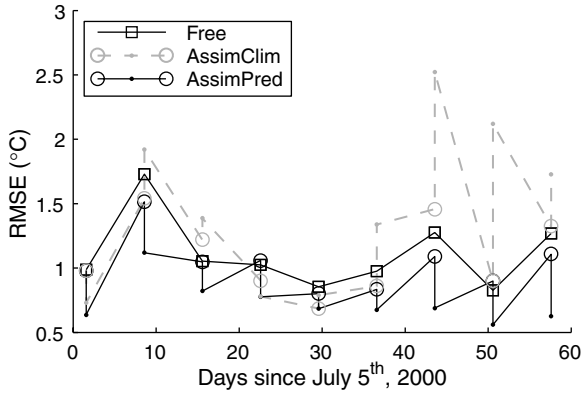

Fig. 6. RMS error between the observations and the SST of model run with assimilation of the MEDAR SST. The dots correspond to the RMS error of the analysis and the empty circles correspond to the forecast. `Free` and `AssimPred` are shown again for comparison.

Table 1
The time averaged SST RMS error of the different experiments

	Forecast	Analysis
`Free`	1.111	–
`AssimObs`	1.031	0.580
`AssimPred`	1.037	0.732
`AssimAmpl`	1.063	0.883
`AssimClim`	1.096	1.479

The RMS error between the observed SST and the analysis of the model run with observed SST assimilation, AssimObs, is obviously not a comparison with independent data. But all other RMS errors in Table 1 are skill assessments based on independent data since the model forecast at a given time has not yet been influenced by the observations of this time and the analysis of the experiments AssimPred and AssimAmpl assimilate the predicted SST of the analysis time. This predicted SST is also independent of the observed SST at the same time.

The model without assimilation predicts the SST with an average RMS error of 1.11 °C. The analysis of AssimPred is better, on average, than the free model; the RMS of the analysis is reduced by 0.4 °C. This improvement is significant at a 95% confidence level using a paired t-test.

The best results are of course obtained when assimilating the observations, but these are not available in a real operational forecast. However, the error reduction of the forecast from AssimPred is comparable to the improvement of the forecast from AssimObs. The RMS error of the analysis in AssimObs is the smallest because the observations have been assimilated. The improvement of the analysis in this experiment can be achieved in a real operational forecast because AssimPred does not use the observed SST. The difference between the forecast from AssimObs and the analysis of AssimPred highlights the benefit of statistical predictors in constraining ocean models.

The AssimAmpl experiment produces similar results as the AssimPred experiment. However, the improvement of the AssimAmpl experiment is not as large as the assimilation of the reconstructed SST. This difference suggests that the errors of the rejected EOF amplitudes are not negligible. In particular, the variance of the model SST within this space is too high and the concentration of the SST variance in the space formed by the dominant EOFs improves the result. The assimilation of EOF amplitudes is also technically more difficult since it implies non-local observations.

Large differences between the climatological SST and the observed SST exist. Thus, the failure of the experiment AssimClim is not a surprise. Most of the time, the model forecast is closer to the observations than the climatology. The assimilation of the climatological SST pulls the model results away from the observations. This is particularly evident during a strong warming event at the end of the assimilation period. This warming event is absent from the climatology but is present in the atmospheric heat flux forcing. Therefore the free model is able to predict this event but the assimilation of the climatological SST substantially degrades the model results during this period.

An error in the heat flux will mainly appear as a SST bias in the Ligurian Sea given the small extent of the model domain. But the free model, forecast and analysis of AssimSST all have a space and time average bias lower than 0.14 °C, indicating that the total heat input of the Ligurian Sea over the simulation period is sufficiently accurate. However, the time series of the bias reveals that a large positive SST bias from day 15 to 50 is compensated by a negative bias at the beginning and end of the simulated time period. The average magnitude of the bias, defined as follows, is 0.57 °C.

$$\sqrt{\langle\langle \mathbf{y}^0 - \mathbf{H}\mathbf{x}\rangle_s^2\rangle_t} = 0.57\ ^\circ\mathrm{C} \tag{7}$$

where $\langle\ \rangle_s$ and $\langle\ \rangle_t$ denotes an average over space and time respectively, and $\mathbf{x}$ are the results of the free model run. This average bias contributes to 50% of the total SST RMS error:

$$\sqrt{\langle\langle (\mathbf{y}^0 - \mathbf{H}\mathbf{x})^2\rangle_s\rangle_t} = 1.14\ ^\circ\mathrm{C} \tag{8}$$

This analysis shows that the heat flux error, although not persistent in time, creates momentary important SST biases in the model. As pointed out by Alvera-Azcárate et al. (in press) such error behaviour can be explained by a timing error (also called phase error) in the atmospheric fields.

Since SST and atmospheric heat fluxes are so tightly connected, an improved understanding of SST will also benefit estimates of the heat flux. From the heat content correction of the upper ocean introduced by the assimilation we can compute the correction in the total heat content introduced by the assimilation. In Fig. 7 the heat content correction per m^2 of the Ligurian Sea is shown. Since the horizontal advection of heat is small compared to the surface heat flux, this correction can be interpreted as an error in the total atmospheric heat forcing introduced in the model during the one-week forecast. This quantity is shown in Fig. 7

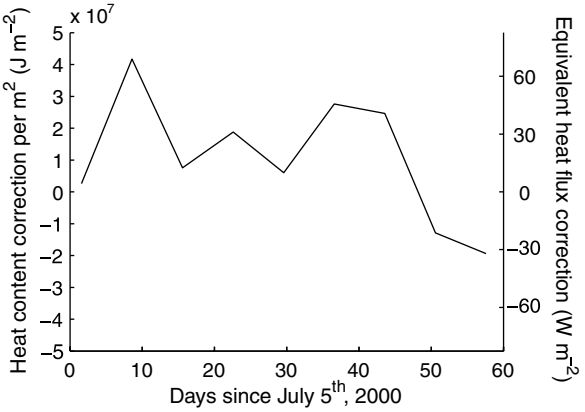

Fig. 7. Vertically integrated and spatial averaged heat content correction and equivalent heat flux correction.

on the left axis. The standard deviation of the equivalent heat flux correction is 35 W m^{-2}. Assimilation of SST can thus be used to estimate the spatial average of the heat flux error. The corrected heat flux could be used for subsequent model runs.

7. Discussion

The SOFT predictor relies on a given number of past observations, the embedding dimension. For SST prediction in the Ligurian Sea, Álvarez et al. (2004) used an embedding dimension of eight. After assimilating eight weekly SST images of the Ligurian Sea should not the model forecast be better than the forecast of the statistical predictor? If the hypotheses of the Kalman filter are verified, this should be the case for a reasonably good ocean model, since the Kalman filter should provide the best estimate of the ocean state taking into account the dynamics and all previous observations. But in ocean modelling these hypotheses are not exactly satisfied. For instance, the dynamics are nonlinear, the model and observations are sometimes biased, the model error is correlated in time, and the error covariances specified in the assimilation scheme are only approximations.

In particular, the reduced rank approximation of the error covariance matrix can only take into account model error within a certain error subspace. Any error outside this space is ignored. Therefore, a part of the signal in the observations is systematically rejected.

The manifold of the model trajectory is generally different from the manifold of the true dynamics. This can be highlighted by comparing the model climatology and the climatology based on observations. Often the model error covariances are not well specified and the model error bias and its time correlation are neglected. This leads to the situation where the model falls back on its manifold after assimilation (Judd and Smith, 2001). In this context, it makes sense to use different models with different dynamics, biases and drifts. If the models are independent, one can suppose that by merging different model results the problems associated with the biases and drifts are reduced (e.g. by superensemble approach).

The SOFT system forecasts the SST using an algebraic function of the l (the embedding dimension) past SST scenes of the Ligurian Sea. The operations (the four elementary operations $+$, $-$, $*$, / are implemented) of this function and their order are determined by a genetic algorithm. This approach is very different from ocean modelling based on differential equations such as the primitive equations. Therefore, the models can be considered as independent with respect to their dynamics, but, they are influenced by the same observations in this case. After taking into account l SST images of the Ligurian Sea, the forecast of the primitive equations model with SST assimilation, and the forecast of the statistical predictor are affected by different model errors which can be reduced by merging them via data assimilation.

Of course, this procedure is not limited to two models. The Kalman filter analysis can be extended to any number of independent models. An option that has not been considered here is to use the analysis as the initial state of the statistical predictor. In this case, the model coupling would be bi-directional. But it is not clear how

the statistical model behaves when it is fed with the SST of the analysis since the statistical model was "trained" with observed SST with different scales and noise characteristics than the SST of the analysis.

8. Application to dynamically evolving error spaces

So far the correlations between the model forecast and the pseudo-observations have been neglected. But in the forecasting scenario suggested by Fig. 2b, the SOFT forecast and the PE model forecast both use the same SST observations. An error in these SST data implies also an error in the SOFT and PE forecasts. The errors of both forecasts are thus correlated. Their correlation and covariance involve the model dynamics since cause and effect of the error are at different time. The approximation to neglect this covariance is at least consistent with our assumption of a fixed error space since the error modes are not propagated through the model dynamics between the assimilation cycles. However, in an assimilation system with an evolving error space, the estimation of this covariance is desirable. First we show how this covariance can be determined in the frame of the extended Kalman filter. It turns out that the estimation of this covariance, which depends on the model dynamics, does not require additional model runs. Next we show how it can be taken into account during the analysis.

8.1. Estimation of the error covariance

The $n \times m$ matrix $\mathbf{C}_n^f$ is the covariance between the model forecast $\mathbf{x}_n^f$ and the data to be assimilated $\mathbf{y}_n^0$:

$$E[(\mathbf{x}_n^f - \mathbf{x}_n^t)(\mathbf{y}_n^0 - \mathbf{H}\mathbf{x}_n^t)^{\mathrm{T}}] = \mathbf{C}_n^f \tag{9}$$

The vector $\mathbf{x}_n^t$ denotes the true state and $\mathbf{H}$ is the observation operator. Since the SOFT system involves the Ligurian Sea SST at l different times, it is convenient to define the predictor's state vector $\mathbf{z}_n$ by

$$\mathbf{z}_n = \begin{pmatrix} \mathbf{y}_n^0 \\ \vdots \\ \mathbf{y}_{n-l}^0 \end{pmatrix} \tag{10}$$

This state vector $\mathbf{z}_n$ is composed by l SST scenes of the Ligurian Sea. With this definition, the prediction step can be expressed by

$$\mathbf{z}_n = M^p(\mathbf{z}_{n-1}) \tag{11}$$

The evolution of the true predictor state vector $\mathbf{z}_n^t$ is governed by the following stochastic equation which takes the model error $\boldsymbol{\eta}_n^p$ into account:

$$\mathbf{z}_n^t = M^p(\mathbf{z}_{n-1}^t) + \boldsymbol{\eta}_{n-1}^p \tag{12}$$

The assimilated pseudo-observations are extracted through the operator $\mathbf{H}^p$.

$$\mathbf{y}_n^0 = \mathbf{H}^p \mathbf{z}_n \tag{13}$$

The covariance between the state vectors of the model $\mathbf{x}_n^{f,a}$ and the predictor $\mathbf{z}_n$, $\boldsymbol{\Gamma}_n^{f,a}$ and the covariance of the model state vector and $\mathbf{y}_n^0$, $\mathbf{C}_n^{f,a}$ is given by

$$\boldsymbol{\Gamma}_n^{f,a} = E\left[(\mathbf{x}_n^{f,a} - \mathbf{x}_n^t)(\mathbf{z}_n - \mathbf{z}_n^t)^{\mathrm{T}}\right] \tag{14}$$

$$\mathbf{C}_n^{f,a} = \boldsymbol{\Gamma}_n^{f,a} \mathbf{H}^{p\mathrm{T}} \tag{15}$$

As in the Extended Kalman filter, we assume that the error dynamics of the PE model state and the predictors state can be approximated by the linearised equations:

$$\mathbf{x}_n^f - \mathbf{x}_n^t = \mathbf{M}_{n-1}(\mathbf{x}_{n-1}^a - \mathbf{x}_{n-1}^t) + \boldsymbol{\eta}_{n-1} \tag{16}$$

$$\mathbf{z}_n - \mathbf{z}_n^t = \mathbf{M}_{n-1}^p(\mathbf{z}_{n-1} - \mathbf{z}_{n-1}^t) + \boldsymbol{\eta}_{n-1}^p \tag{17}$$

where $\mathbf{M}_n$ and $\mathbf{M}_n^p$ are the tangent linear model of the PE and the statistical model respectively. From these two equations and from the definition (14), we can derive an evolution equation for $\mathbf{\Gamma}_n^f$:

$$\mathbf{\Gamma}_n^f = \mathbf{M}_{n-1}\mathbf{\Gamma}_{n-1}^a\mathbf{M}_{n-1}^p \tag{18}$$

where we have used the fact that the PE model error $\boldsymbol{\eta}_n$ and the predictors model error $\boldsymbol{\eta}_n^p$ are uncorrelated.

8.2. Analysis scheme

The analysis scheme is derived in Appendix A. The standard Kalman filter analysis can be applied but with modified observations $\mathbf{y}'$, observation error covariance $\mathbf{R}'$ and observation operator $\mathbf{H}'$. These modified quantities are related to the original quantities by

$$\mathbf{y}' = \mathbf{y}^0 - \mathbf{C}^{f\mathrm{T}}\mathbf{P}^{f-1}\mathbf{x}^f \tag{19}$$

$$\mathbf{R}' = \mathbf{R} - \mathbf{C}^{f\mathrm{T}}\mathbf{P}^{f-1}\mathbf{C}^f \tag{20}$$

$$\mathbf{H}' = \mathbf{H} - \mathbf{C}^{f\mathrm{T}}\mathbf{P}^{f-1} \tag{21}$$

The analysis can be carried out with the classical Kalman filter analysis:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}'^{\mathrm{T}}\left(\mathbf{H}'\mathbf{P}^f\mathbf{H}' + \mathbf{R}'\right)^{-1} \tag{22}$$

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}' - \mathbf{H}'\mathbf{x}^f) \tag{23}$$

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}'\mathbf{P}^f \tag{24}$$

The covariance $\mathbf{\Gamma}_n^a$ is updated according to the following equations:

$$\mathbf{\Gamma}_n^a = \mathbf{\Gamma}_n^f + \mathbf{K}\left(\mathbf{H}^p\mathbf{P}_n^p - \mathbf{H}\mathbf{\Gamma}_n^f\right) \tag{25}$$

where $\mathbf{P}^p$ is the error covariance of the predictors state $\mathbf{z}$:

$$\mathbf{P}_n^p = E\left[\left(\mathbf{z}_n - \mathbf{z}_n^t\right)\left(\mathbf{z}_n - \mathbf{z}_n^t\right)^{\mathrm{T}}\right] \tag{26}$$

8.3. Reduced rank covariance

If the model error covariance $\mathbf{P}^f$ is expressed as in the SEEK filter (Pham, 2001) in a low dimensional space $\mathbf{L}_n$:

$$\mathbf{P}_n^{f,a} = \mathbf{L}_n\tilde{\mathbf{P}}_n^{f,a}\mathbf{L}_n^{\mathrm{T}} \tag{27}$$

where $\tilde{\mathbf{P}}_n^{f,a}$ is the error covariance in this low dimensional space, then from the forecast equation (18) and from the analysis equation (25), the covariance $\mathbf{\Gamma}_n^{f,a}$ can always be expressed as

$$\mathbf{\Gamma}_n^{f,a} = \mathbf{L}_n\tilde{\mathbf{\Gamma}}_n^{f,a} \tag{28}$$

if initially $\mathbf{\Gamma}_0^{f,a} = 0$. The matrix $\tilde{\mathbf{\Gamma}}_n^{f,a}$ has only a size of $r \times m$. Furthermore, the forecast step to compute $\tilde{\mathbf{\Gamma}}_n^f$ no longer involves any additional PE model integrations:

$$\tilde{\mathbf{\Gamma}}_n^f = \tilde{\mathbf{\Gamma}}_{n-1}^a\mathbf{M}_{n-1}^p \tag{29}$$

since the error space is itself evolved in time by $\mathbf{L}_n = \mathbf{M}_{n-1}\mathbf{L}_{n-1}$. In summary, the assimilation of pseudo-observations is not limited to an assimilation method based on constant error space as used in the present work. The derivations above show that the assimilation of pseudo-observations (correlated with the state vector) can also be integrated at little additional costs in assimilation methods derived from the Extended Kalman Filter using an evolving error space.

9. Conclusions

The main conclusion from this work is that the forecast of a statistical predictor, like the SOFT system developed by Álvarez et al. (2004), can be used to improve the forecast skill of models based on primitive equations. The SST forecast obtained by the model assimilating the predicted SST is closer to the observations than the SST forecast by the model without assimilation. The accuracy of the SST forecast was improved by 0.3 °C when the forecast of the experiment `AssimObs` is compared to the analysis of the experiment `AssimPred`. This corresponds to a 30% reduction of the average RMS error.

SST is a key parameter for heat exchange between the ocean and atmosphere. In our experiments, the assimilation of predicted SST appears to reduce the effect of errors and biases in the atmospheric heat fluxes. The assimilation of the results obtained from statistical predictors might have an application for operational forecast. Currently, operational forecasts are unconstrained model runs. But systematic errors due to biases can be detected and therefore corrected by comparing the forecasts with other models forecasts. By assimilating the forecasts of specialised statistical predictors, a model can still be constrained even in forecast-mode. This reduces the error growth due to model imperfections.

Another way to constrain a forecast model run is to assimilate climatological SST. It is a common practise in ocean modelling to weakly nudge the model towards climatological SST to reduce the long-term impact of biased heat fluxes. Since the SST is a highly variable parameter, the assimilation of climatological SST does not improve the capacity of the model to forecast SST. In the present case, the model without assimilation gives a better SST forecast than the interpolated climatological SST. For short-range forecasts, the SST predicted by the SOFT system is therefore a better choice.

To our knowledge, this is also the first time that a data assimilation method is used to combine the results of a statistical prediction model and a primitive equation model. Traditionally, data assimilation merges observations with model dynamics. Here we show that with the framework of data assimilation it is also possible to combine the results of different models of very different natures. The improvements on the prediction skill of this hybrid system are encouraging but further research is necessary, for example, to assess and to control the impact of the assimilation on the unobserved model state vector and to explore the possibilities of this method with error space evolving data assimilation schemes.

Acknowledgements

Alberto Alvarez is acknowledged for providing the data of the SOFT predictor. The present work was carried out with the support of the Fund for Research Training in Industry and Agriculture (FRIA, Belgium). The National Fund of Scientific Research (FNRS, Belgium) is acknowledged for the financing of a supercomputer. The work was realised within the framework of the Satellite-based Ocean ForecasTing project (SOFT, EVK3-CT-2000-00028) of the European Union. We would like to thank the anonymous reviewers for their constructive criticism and valuable suggestions. This is MARE publication 79.

Appendix A. Derivation of the analysis equations

In the following development the time index n is dropped because all quantities are taken at the same time. Since in our case, the model state $\mathbf{x}^f$ and the pseudo-observations to be assimilated $\mathbf{y}^0$ are not independent, it is convenient to introduce the vector $\mathbf{z}$ defined by

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}^f \\ \mathbf{y}^0 \end{pmatrix} \tag{A.1}$$

The true state of the system $\mathbf{x}^t$ is related to this vector $\mathbf{z}$ by

$$\mathbf{E}\mathbf{x}^t = \mathbf{z} + \boldsymbol{\zeta} \tag{A.2}$$

where the matrix $\mathbf{E}$ is defined as

$$\mathbf{E} = \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \end{pmatrix} \tag{A.3}$$

where $\mathbf{I}$ is the identity matrix and $\boldsymbol{\zeta}$ is an unknown error term with the following error covariance $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} \mathbf{P}^f & \mathbf{C}^f \\ \mathbf{C}^{f\mathrm{T}} & \mathbf{R} \end{pmatrix} \tag{A.4}$$

In this general case, the analysis $\mathbf{x}^a$ and its error covariance $\mathbf{P}^a$ are given by

$$\mathbf{x}^a = \mathbf{P}^a \mathbf{E}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{z} \tag{A.5}$$

$$\mathbf{P}^a = (\mathbf{E}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{E})^{-1} \tag{A.6}$$

The inverse of the block matrix $\mathbf{S}$ is

$$\mathbf{S}^{-1} = \begin{pmatrix} \mathbf{P}^{f\,-1} + \mathbf{P}^{f\,-1}\mathbf{C}^f \mathbf{R}'^{-1} \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} & -\mathbf{P}^{f\,-1}\mathbf{C}^f \mathbf{R}'^{-1} \\ -\mathbf{R}'^{-1}\mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} & \mathbf{R}'^{-1} \end{pmatrix} \tag{A.7}$$

where

$$\mathbf{R}' = \mathbf{R} - \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} \mathbf{C}^f \tag{A.8}$$

Thus, the error covariance of the analysis is obtained by

$$\mathbf{P}^{a\,-1} = \mathbf{P}^{f\,-1} + \mathbf{P}^{f\,-1}\mathbf{C}^f \mathbf{R}'^{-1} \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} - \mathbf{P}^{f\,-1}\mathbf{C}^f \mathbf{R}'^{-1}\mathbf{H} - \mathbf{H}^{\mathrm{T}} \mathbf{R}'^{-1} \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}'^{-1} \mathbf{H} \tag{A.9}$$

$$= \mathbf{P}^{f\,-1} + \mathbf{H}' \mathbf{R}'^{-1} \mathbf{H}'^{\mathrm{T}} \tag{A.10}$$

where we have defined $\mathbf{H}'$ as

$$\mathbf{H}' = \mathbf{H} - \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} \tag{A.11}$$

Eq. (A.10) finally leads to

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{K}\mathbf{H}'\mathbf{P}^f \tag{A.12}$$

where

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}'^{\mathrm{T}} (\mathbf{H}' \mathbf{P}^f \mathbf{H}' + \mathbf{R}')^{-1} \tag{A.13}$$

In order to compute $\mathbf{x}^a$, we need to evaluate the following product:

$$\mathbf{E}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{z} = \mathbf{P}^{f\,-1} \mathbf{x}^f - \mathbf{H}'^{\mathrm{T}} \mathbf{R}'^{-1} \mathbf{C}^f \mathbf{P}^{f\,-1} \mathbf{x}^f + \mathbf{H}'^{\mathrm{T}} \mathbf{R}'^{-1} \mathbf{y}^0 \tag{A.14}$$

$$= \mathbf{P}^{f\,-1} \mathbf{x}^f + \mathbf{H}'^{\mathrm{T}} \mathbf{R}'^{-1} \mathbf{y}' \tag{A.15}$$

where

$$\mathbf{y}' = \mathbf{y}^0 - \mathbf{C}^{f\mathrm{T}} \mathbf{P}^{f\,-1} \mathbf{x}^f \tag{A.16}$$

Finally one obtains for $\mathbf{x}^a$,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}' - \mathbf{H}' \mathbf{x}^f) \tag{A.17}$$

In summary, the assimilation of correlated observations is formally equivalent to the traditional Kalman analysis equations where the observations, the observation operator and the observation error covariance are substituted by (A.16), (A.11) and (A.8) respectively. It is also interesting to note that the transformed observations $\mathbf{y}'$ are no longer correlated with the state vector $\mathbf{x}^f$.

References

Álvarez, A., Lopez, C., Riera, M., Hernandez-Garcia, E., Tintoré, J., 2000. Forecasting the SST space-time variability of the Alboran Sea with genetic algorithms. Geophysical Research Letters 27 (17), 2709–2712.

Álvarez, A., Orfila, A., Tintoré, J., 2001. DARWIN: An evolutionary program for nonlinear modelling of chaotic time series. Computer Physics Communications 136, 334–349.

Álvarez, A., Orfila, A., Tintoré, J., 2004. Real-time forecasting at weekly timescales of the SST and SLA of the Ligurian Sea with a satellite-based ocean forecasting (SOFT) system. Journal of Geophysical Research 109, C03023.

Alvera-Azcárate, A., Barth, A., Bouallègue, Z.B., Vandenbulcke, L., Rixen, M., Beckers, J.-M., in press. Forecast verification of a 3D model of the Ligurian Sea. The use of discrete wavelet transforms and EOFs in the skill assessment of spatial forecasts. Journal of Marine Systems.

Bahurel, P., Dombrowsky, E., Lellouche, J.-M., the Mercator Project Team, 2004. Mercator Ocean Monitoring and Forecasting System, near-real-time assimilation of satellite and in situ data in different operational ocean models. In: Marine Environmental Monitoring and Predictions. 37th Liège Colloquium, oral presentation.

Balmaseda, M.A., Anderson, D.L.T., Davey, M.K., 1994. ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. Tellus 46A, 497–511.

Barnett, T.P., Latif, M., Graham, N., Flugel, M., Pazan, S., White, W., 1993. ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. Journal of Climate 6, 1545–1566.

Barnston, A.G., He, Y., 1996. Skill of canonical correlation analysis forecasts of 3-month mean surface climate in Hawaii and Alaska. Journal of Climate 9, 2579–2605.

Barnston, A.G., Ropelewski, C.F., 1992. Prediction of ENSO episodes using canonical correlation analysis. Journal of Climate 7, 1316–1345.

Barth, A., Alvera-Azcárate, A., Beckers, J.-M., Rixen, M., Vandenbulcke, L., in press. Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea. Journal of Marine Systems.

Barth, A., Alvera-Azcárate, A., Rixen, M., Beckers, J.-M., 2005. Two-way nested model of mesoscale circulation features in the Ligurian Sea. Progress in Oceanography 66, 171–189.

Barth, N.H., 1992. Oceanographic experiment design II: genetic algorithms. Journal of Atmospheric and Oceanic Technology 9, 434–443.

Beckers, J.-M., 1991. Application of a 3D model to the Western Mediterranean. Journal of Marine Systems 1, 315–332.

Beckers, J.-M., Rixen, M., Brasseur, P., Brankart, J.-M., Elmoussaoui, A., Crépon, M., Herbaut, C., Martel, F., Van den Berghe, F., Mortier, L., Lascaratos, A., Drakopoulos, P., Korres, P., Pinardi, N., Masetti, E., Castellari, S., Carini, P., Tintore, J., Alvarez, A., Monserrat, S., Parrilla, D., Vautard, R., Speich, S., 2002. Model intercomparison in the Mediterranean. The MedMEx simulations of the seasonal cycle. Journal of Marine Systems 33–34, 215–251.

Brankart, J.M., Testut, C.E., Brasseur, P., Verron, J., 2003. Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean modes: application to a 1993–96 hindcast of the North Atlantic Ocean circulation. Journal of Geophysical Research 108 (C3).

Brasseur, P., Ballabrera, J., Verron, J., 1999. Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model. Journal of Marine Systems 22 (4), 269–294.

Davis, R.E., 1976. Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. Journal of Physical Oceanography 6 (3), 249–266.

Drosdowsky, W., 1994. Analog (nonlinear) forecasts of the oscillation index time series. Weather and Forecasting 9, 78–84.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Advances in Geophysics 33, 141–265.

Graham, N.E., Barnett, T.P., Latif, M., 1992. Considerations of the predicatability of ENSO with a low-order coupled model. In: Proceedings of the 16th Annual Climate Diagnostic Workshop, NOAA, Los Angeles. pp. 323–327.

He, Y., Barnston, A.G., 1996. Long-lead forecasts of seasonal precipitation in the Tropical Pacific Islands using CCA. Journal of Climate 9, 2020–2035.

Hernandez, F., Le Traon, P.-Y., Barth, N., 1995. Optimizing a drifter cast strategy with a genetic algorithm. Journal of Atmospheric and Oceanic Technology 12, 330–345.

Ji, M., Kumar, A., Leetmaa, A., 1994. An experimental coupled forecast system at the National Meteorological Center: some early results. Tellus 46A, 398–418.

Judd, K., Smith, L.A., 2001. Indistinguishable states II: imperfect model scenarios. Physica D 2–4 (151), 125–141.

Keppenne, C.L., Ghil, M., 1992. Adaptive spectral analysis and prediction of the Southern Oscillation index. Journal of Geophysical Research 97 (D18), 20449–20554.

Kirtman, B.P., Shukla, J., Huang, B., Zhu, Z., Schneider, E.K., 1997. Multiseasonal predictions with a coupled tropical ocean–global atmosphere system. Monthly Weather Review 125, 789–808.

Knaff, J.A., Landsea, C.W., 1997. An El Niño southern oscillation climatology and persistence (CLIPER) forecasting scheme. Weather and Forecasting 12, 633–652.

Latif, M., Flügel, M., 1991. An investigation of short-range climate predictability in the tropical Pacific. Journal of Geophysical Research 96 (C2), 2661–2673.

MEDAR-Group, 2002. MEDATLAS/2002 database. Mediterranean and Black Sea database of temperature salinity and bio-chemical parameters. Climatological Atlas. Ifremer edition 4 CD-ROM.

Neelin, J.D., 1989. Interannual oscillations in an ocean general circulation model coupled to a simple atmosphere model. Philosophical Transactions of the Royal Society of London 329A, 189–205.

Neelin, J.D., 1990. A hybrid coupled general circulation model for El Niño studies. Journal of the Atmospheric Sciences 47, 674–696.

Penland, C., Magorian, T., 1993. Prediction of Niño 3 sea-surface temperatures using linear inverse modeling. Journal of Climate 6, 1067–1076.

Penland, C., Matrosova, L., 2001. Expected and actual errors of linear inverse model forecasts. Monthly Weather Review 129, 1740–1745.

Pham, D.T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. Monthly Weather Review 129, 1194–1207.

Pham, D.T., Verron, J., Roubaud, M.C., 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. Journal of Marine Systems 16 (3–4), 323–340.

Rixen, M., Beckers, J.-M., Brankart, J.-M., Brasseur, P., 2001. A numerically efficient data analysis method with error map generation. Ocean Modelling 2 (1–2), 45–60.

Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344, 734–741.

Takens, F., 1981. Ch. Detecting strange attractors in fluid turbulence. In: Dynamical Systems and TurbulenceLecture Notes in Mathematics, vol. 898. Springer-Verlag, New York.

Van den Dool, H.M., 1994. Searching for analogues, how long must we wait? Tellus A 46, 314–324.

Yu, J.-Y., Mechoso, C.R., McWilliams, J.C., Arakawa, A., 2002. Impacts of the Indian Ocean on the ENSO cycle. Geophysical Research Letters 29 (8), 46-1–46-4.

Zhang, B., Lie, J., Sun, Z., 1993. A new multidimensional time series forecasting method based on the EOF iteration scheme. Advances in Atmospheric Sciences 10, 2430–2437.