Fifth Framework Programme of the European Commission 1998-2002
Energy, Environment and Sustainable Development (EESD) Programme

FerryBox

From On-line Oceanographic Observations to Environmental Information

Quantification of the Value of Assimilated Ferrybox Data into Pre-operational Models

Contract number : **EVK2-2002-00144**

Deliverable number : **D-5-3**

Revision : **2.0**

Co-ordinator:

Professor Dr. Franciscus Colijn

GKSS Research Centre
Institute for Coastal Research
Max-Planck-Strasse
D-21502 Geesthacht

http://www.ferrybox.org

Document Reference Sheet

This document has been elaborated and issued by the European FerryBox Consortium.

P 1	**GKSS**	**GKSS Research Centre** **Institute for Coastal Research**	**Coordinator**
P 2	**NERC.NOC**	**NERC.NOC – National Oceanography Centre** **Southampton University and National Environment Res. Council** formerly NERC.SOC – Southampton Oceanography Centre	
P 3	**NIOZ**	**Royal Netherlands Institute of Sea Research**	
P 4	**FIMR**	**Finnish Institute of Marine Research**	
P 5	**HCMR** (formerly NCMR)	**Hellenic Centre for Marine Research** (formerly National Centre for Marine Research)	
P 6	**NERC.POL**	**Proudman Oceanographic Laboratory**	
P 7	**NIVA**	**Norwegian Institute for Water Research**	
P 8	**HYDROMOD**	**HYDROMOD Scientific Consulting**	
P 9	**CTG** **(formerly CIL)**	**Chelsea Technology Group** **(formerly Chelsea Instruments Ltd.)**	
P 10	**IEO**	**Spanish Institute of Oceanography**	
P 11	**EMI**	**Estonian Marine Institute** (in cooperation with the Estonian Maritime Academy)	

The European FerryBox Project was co-funded by the European Commission under the Fifth Framework Programme of the European Commission 1998-2002 - Energy, Environment and Sustainable Development (EESD) Programme under contract no. EVK2-2002-00144

Document Control Table

Project acronym:	FerryBox		Contract no.:	EVK2-2002-00144	
Deliverable No.:	D-5-3			Revision:	2.0
WP number and title:	FerryBox WP-5	Applications of Ferrybox data			
Work Package Manager:	Roger Proctor – Proudman Oceanographic Laboratory				
Work Package Team:					
Document title:	Quantification of the Value of Assimilated Ferrybox Data into Pre-operational Models				
Document owner:	European FerryBox Project Consortium				
Document category:	Deliverable				
Document classification:	PU – Public				
Status:	Final				
Purpose of release:	Deliverable for the European Commission				
Contents of deliverable:	Quantification of the Value of Assimilated Ferrybox Data into Pre-operational Models				
Pages (total):	21	Figures:	15	Tables:	2
Remarks:	Updated for publication on the FerryBox report CD and website				
Main author / editor:	Roger Proctor		FerryBox WP-5 Leader		NERC.POL
Contributors:	FerryBox WP-5 Team				
Main contacts:	FerryBox project coordinator:		Contact for this report:		
	Professor Dr. Franciscus Colijn GKSS Research Centre Institute for Coastal Research Max-Planck-Strasse D-21502 Geesthacht, Germany Tel.: +49 4152 87 – 1533 Fax.: +49 4152 87 – 2020 E-mail: franciscus.colijn@gkss.de		Dr. Roger Proctor POL, Proudman Oceanographic Laboratory 6 Brownlow Street UK- Liverpool L35DA, United Kingdom Tel.: +44 151 795 4856 Fax.: +44 151 795 4801 E-mail: rp@pol.ac.uk		
Project website:	http://www.ferrybox.org				

Table of Contents

List of Figures

List of Tables

Quantification of the Value of Assimilated Ferrybox Data into Pre-operational Models

G. Korres[1], K.Nittis[1], I.Hoteit[2] and G.Triantafyllou[1]

I. Andreu-Burillo[3], R. Proctor[3]

[1] *Hellenic Centre for Marine Research, Greece*

[2] *Scripps Institution of Oceanography, USA*

[3] *Proudman Oceanographic Laboratory, UK.*

Ferrybox data has the potential to be assimilated into numerical models to improve the performance of these models. In this report we consider 2 case studies as a first assessment of this capability. Each study implements different data assimilation schemes (reported fully in D-5-2) in different geographical areas. In our first case study, Sea Surface Salinity data from the Athens to Iraklion ferry are assimilated into a 3-dimensional model of the Aegean Sea and its influence compared to the influence of assimilating satellite-derived Sea Surface Temperature and Sea Surface Anomaly. In our second case study we assimilate Sea Surface Temperature measured from the Irish Sea ferry into a 3-dimensional model of the Irish Sea and compare its influence to the assimilation of satellite-derived Sea Surface Temperature. The first study was conducted by the Hellenic Centre for Marine Research (HCMR) and the second by the Proudman Oceanographic Laboratory (POL).

1 HCMR

The Hellenic Centre for Marine Research implemented an assimilation scheme based on the Singular Evolutive Extended Kalman (SEEK) filter into an Aegean Sea configuration of the POM (Princeton Ocean Model) model. This system was validated for a 6-month period (January – June 2004) during which weekly satellite SLA, SST data and daily Ferrybox SSS data were jointly assimilated into the model.

1.1 The Aegean Sea Numerical Model

The Aegean Sea model is based on the Princeton Ocean model and is part of the Poseidon system (Soukisian et al., 2002). The computational grid covers the geographical area 20oE-29 °E and 30 °N – 41 °N with a resolution of 1/20°. The model is forced with hourly surface fluxes of momentum, heat and water provided by the ETA high resolution (1/10°) regional atmospheric model (Papadopoulos et al., 2002) issuing forecasts for 72 hours ahead.

Boundary conditions at the western and eastern open boundaries of this model are provided by a large scale model covering the whole Eastern Mediterranean Sea with a resolution of 1/10°. A detailed description of the modelling system and a preliminary evaluation of its forecasting skill are provided by Nittis et al. (2001).

1.1.1 The SEEK Filter

Assume that an ensemble of system states describing the probability distribution of a given system is available. An EOF (Empirical Orthogonal Functions) decomposition of the ensemble usually provides those few dominant directions that describe the spreading of the ensemble. The SEEK filter uses these dominant directions to construct its correction subspace at the initial time and then evolves them in time according to the EK filter equations.

Let the filter's error covariance matrix P_i^a be factorised as $L_i U_i L_i^T$. The SEEK filter operates in two stages:

Forecast stage: The model is integrated forward to obtain the forecast state X^f at time t_i starting from the analysis state X^a available at time t_{i-1}:

$$X^f(t_i) = M\left(t_{i-1}, t_i\right) X^a(t_{i-1}) \qquad (1)$$

where $M(t_{i-1}, t_i)$ is the model transition operator.

Correction stage: The analysis state at time t_i is estimated as:

$$X^a(t_i) = X^f(t_i) + K_i\left[Y_i^o - H_i X^f(t_i)\right] \qquad (2)$$

where the Kalman gain matrix K_i is given by

$$K_i = L_i U_i L_i^T \mathrm{H}_i^T R_i^{-1} \qquad (3)$$

and

$$U_i^{-1} = \rho U_{i-1}^{-1} + L_i^T \mathrm{H}_i^T R_i^{-1} \mathrm{H}_i L_i \qquad 0 < \rho < 1 \qquad (4)$$

In Eq.2, Y_i^o is the observations vector and H_i is the observational operator which projects the model state onto the locations of the observations. In eq. 3, H_i is the gradient of H_i and R_i is the observations error covariance matrix. The latter is usually taken as a diagonal matrix by assuming that the observations are spatially uncorrelated. Note that in cases of a linear observational operator, $\mathrm{H}_i = H_i$. The filter's correction basis L_i evolves in time according to:

$$L_i = \mathrm{M}(t_{i-1}, t_i)\, L_{i-1} \qquad (5)$$

where $\mathrm{M}(t_{i-1}, t_i)$ is the gradient of $M(t_{i-1}, t_i)$ evaluated at $X^a(t_{i-1})$. The factor ρ appearing in (4) is a forgetting factor, which stands for the increment of uncertainty during the model integration in order to attenuate the impact of model errors. Practically, with $\rho < 1$ recent data are exponentially more weighted than old data.

1.1.2 Semi-evolutive SEEK Filter

In order to further reduce the computational burden of the SEEK filter, the evolution of the correction basis was only performed for the first dominant EOFs while the rest of the modes were kept invariant. This idea was first proposed by Hoteit et al. (2002) who found that this approximation has a rather limited impact on the filter's performance, especially when the dynamical system evolves through a stable regime.

1.1.3 Localisation of the Filter Gain

The SEEK filter is usually initialized by an error covariance matrix which is approximated using a finite set of state vectors extracted from a long enough model simulation. Correlations resumed by the last EOFs, which remain invariant in the present study, might be therefore very often noisy, especially in strongly variable high resolution systems. In order to deal with this problem, a strategy is to retain the short-range correlation structure, which is the most reliable, and to filter out long-range correlations. This can be done through a Schur product (an element by element multiplication) of the error covariance matrix with a correlation function γ as suggested by Houtekamer and Mitchell (2000). In this approach, the gain K_i of the original Kalman filter is reformulated as:

$$K_i = \left(\gamma \circ P_i^f\right) H^T \left[H\left(\gamma \circ P_i^f\right) H^T + R\right]^{-1} \qquad (6)$$

where the notation $\gamma \circ P_i^f$ denotes the Schur product of the covariance matrix P_i^f with the localization function γ.

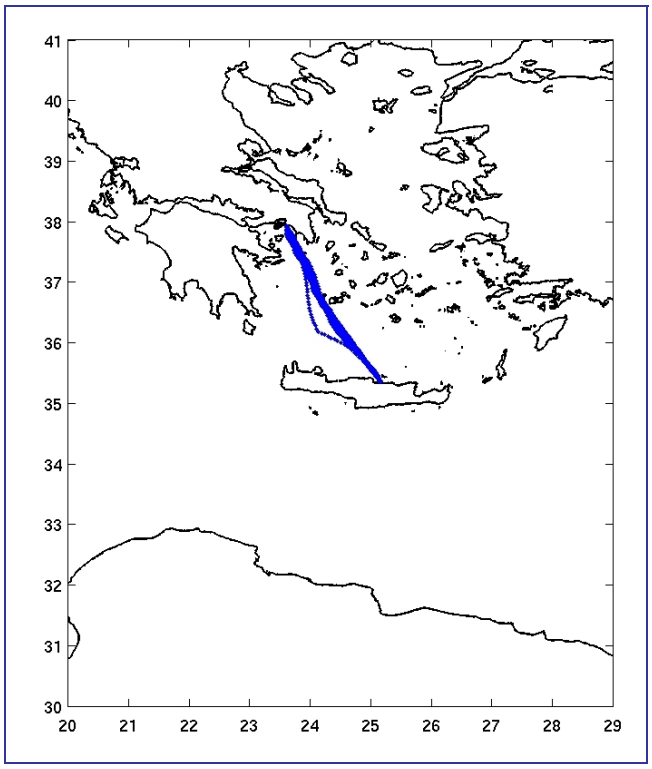

Figure 1-1: Athens – Iraklion ferry track.

1.1.4 Model Error Statistics along the Ferry Track

In order to assess the auto/cross correlation length scales along the ferry track, we consider surface slices of the model error correlation matrix showing the spatial distribution of auto and cross correlations between the SSS at the grid point located at 36.5^o N 24.3 ^{o}E and the rest of the grid points model variables (SST, U, V, SSH). This particular grid point was selected because it lies approximately in the centre of the composite ferry tracks shown in Figure 1-1.

In order to test the statistical significance of the correlation maps, we assume that the distribution of $\frac{\rho}{\sqrt{1-\rho^2}}\sqrt{n^*-2}$ is t with n^*-2 degrees of freedom where ρ is the auto/cross correlation. The effective sample size n^* due to serial correlations existing in the time series is determined according to Bretherton et al. (1999) as $n^* = \frac{1-\rho^2}{1+\rho^2}$ for autocorrelation and $n^* = \frac{1-\rho_1\rho_2}{1+\rho_1\rho_2}$ for cross-correlation between two time series (ρ is spatial mean lag-one autocorrelation).

The model error correlation matrix is constructed by taking daily model state samples for a 2-year period (2002 – 2003). Typical spatial average lag-one autocorrelation is 0.96 for SSS, 0.99 for SST, 0.69 and 0.67 for surface U (longitudinal) and V (meridional) velocities respectively and finally 0.88 for SSH. Thus the effective sample size is 23 for the SSS autocorrelation field and 13, 141, 152, 57 for SSS-SST, SSS-U, SSS-V and SSS-SSH cross-correlation fields respectively. According to the previous effective sample sizes, the threshold values (at the 99% significance level) are estimated to be ±0.46 (SSS autocorrelation), ±0.68 (SSS-SST cross-correlation), ±0.21 (SSS-U cross-correlation), ±0.20 (SSS-V cross-correlation) and ±0.34 (SSS-SSH cross-correlation).

The salinity autocorrelation map shown in Figure 1-2 indicates strong anisotropy (the correlation ellipse is oriented along the NW – SE direction) but presents the expected behaviour of decorrelation with radial distance.

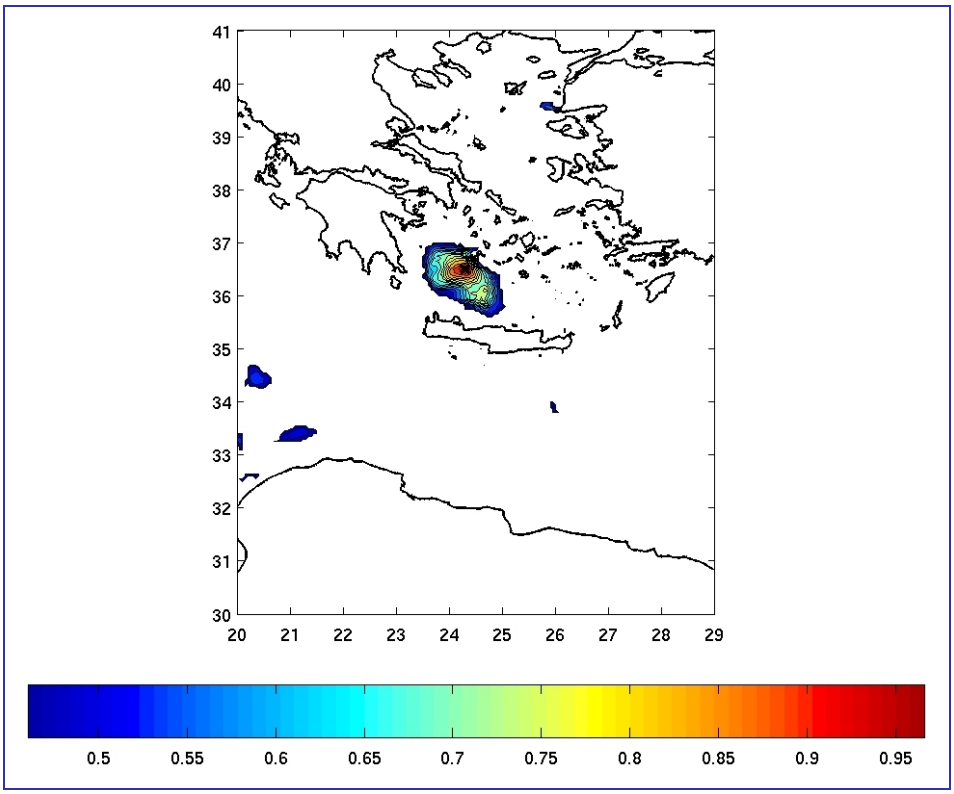

Figure 1-2: Horizontal correlation between SSS and SSS at point 24.3 °E – 36.5 °N (indicated by x). The correlation map has been given a threshold value of ±0.46 and has been estimated by the model run 2002 – 2003.

The model estimated decorrelation length scale along the average ferry track is 90-100 km as shown in Figure 1-3. This is 2-3 times larger compared to the decorrelation length scale estimated directly from the Ferrybox SSS data (Kontoyiannis and Ballas 2005). The rest of the cross-correlation maps show statistically insignificant correlations with the exception of SSS-SSH cross correlation field presented in Figure 1-4. Although this field does not show the typical decorrelation with radial distance behaviour (the maximum cross-correlation does not occur on the spot), there are significant cross-correlations all-over the central Cretan Sea.

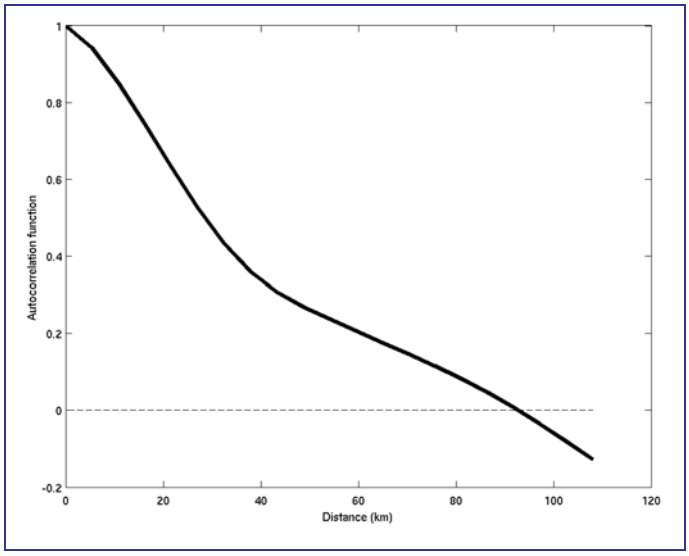

Figure 1-3: SSS spatial autocorrelation function along the average ferry track as estimated by the model run 2002 – 2003.

Overall the study of the model error statistics indicates that the assimilation of Ferrybox SSS observations into the Aegean Sea model will have a rather localised effect on the model analysis considering the present limited geographical coverage of the ferry tracks.

1.2 Results and Discussion

The model has been first spun-up from climatology for a 3-year period (1999-2001) forced with the ETA atmospheric model 24-hour forecasts. A 2-year run of the model (2002-2003) was then performed in order to estimate the multivariate EOFs needed for the SEEK filter algorithm. Finally, an additional integration of 6 months (January –June 2004) initialised from the previous experiment constitutes the “free-run” of the model.

The time window chosen for the assimilation experiment extends from January-June 2004 during which SST, SLA and Ferrybox SSS data products are available. The SLA data are merged T/P & ERS altimetric data on a weekly basis provided by the AVISO project and mapped on a $1/3^{o}$ Mercator grid. The mean SSH added to these SLA maps is derived from a long run (2000-2003) of the Aegean Sea model. The SST data consist of weekly composite AVHRR observations on a $1/8^{o}$ grid provided by the MFSTEP EU project. Finally the SSS data are available on a daily basis along the ferry route from Piraeus to Heraklion. However due to Ferrybox malfunctioning, these data are often sparse during the study period. Overall the availability of all observational data during the time window selected is shown in Figure 1-5.

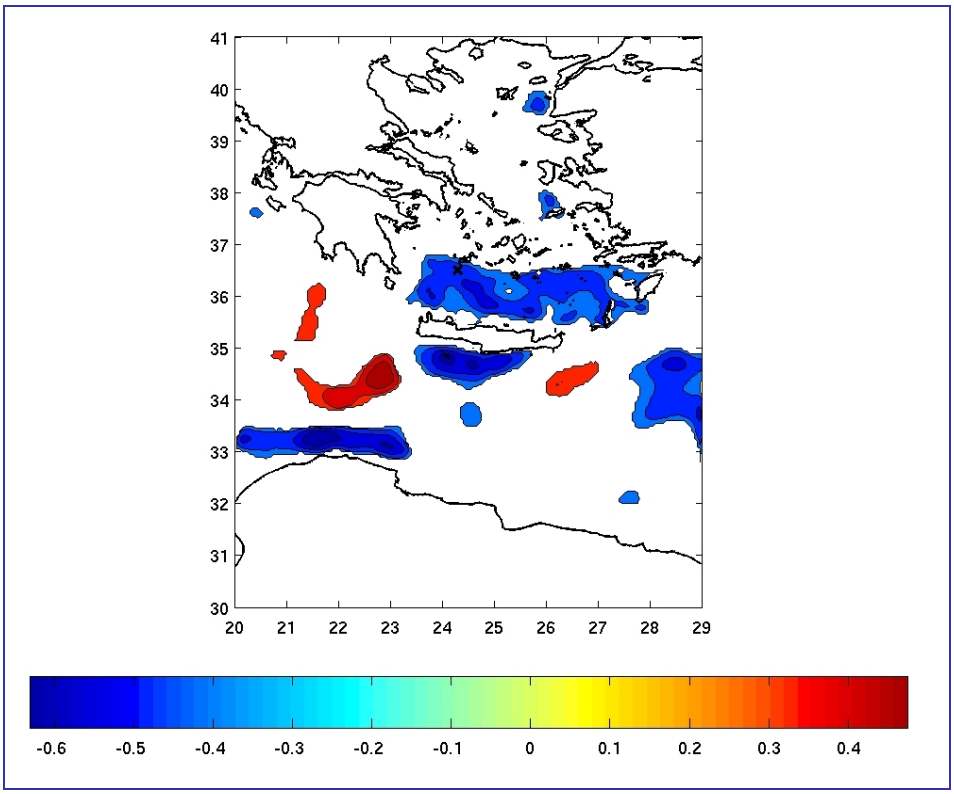

Figure 1-4: Horizontal correlation between SSH and SSS at point 24.3° E – 36.5° N (indicated by x). The correlation map has been given a threshold value of ±0.34 and has been estimated by the model run 2002 – 2003.

In the assimilation runs, the rank of the filter's covariance matrices was set to 60. The first 10 modes are evolved according to the system's dynamics while the rest of the modes are kept invariant. For the localization of the filter gain, we have defined a radius of influence of 150 km outside of which the correlations are set to zero. The assimilation system assumes an accuracy of 2.5 cm on the SLA data, 0.5 °C on SST data and 0.08 on SSS data.

An evaluation of the behaviour of the assimilation system is shown in Table 1-1 in terms of the RMS misfits between observational data and model estimates for the free-run and the forecast and analyzed fields as estimated by the filter.

Table 1-1: RMS error (free run, forecast and analysis fields of the SEEK filter) on SSH, SST and SSS averaged over Jan-Jun 2004.

	Free run	Forecast	Analysis
SSH [cm]	5.17	4.20	3.32
SST [°C]	0.932	0.766	0.486
SSS	0.330	0.094	0.068

Regarding SSH, the filter analysis is almost 2 cm more accurate on average than the free-run. The filter improves also the estimation of the model SST by 50%. Most significant improvement was obtained on the estimation of the SSS probably because of the daily availability of these data. Overall, the assimilation system was able to fit the data within the specified observational errors.

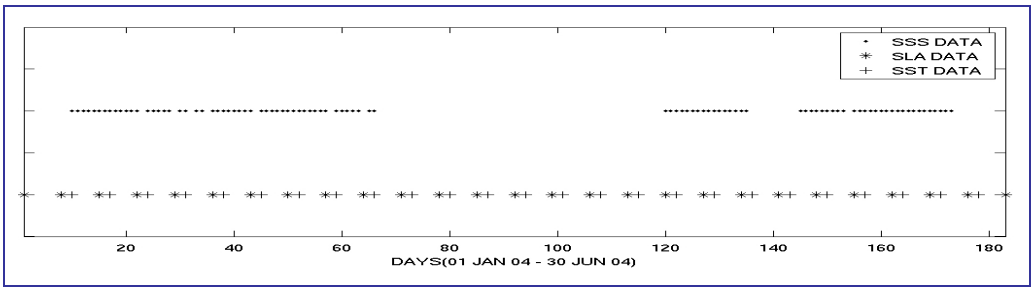

Figure 1-5: Time distribution of assimilated observational data.

Time series of the RMS misfits for SSH, SST and SSS are presented in Figure 1-6. The forecast and analysis SSH misfits with respect to the observations generally follow the free-run errors but reduced by more than 20% and 35% respectively. Regarding SST (Figure 1-6(b)), the model free-run was significantly deviated from the observations toward the end of the assimilation window due to the model inability to correctly describe a strong warming event occurring in the Aegean Sea. The filter greatly enhances the behaviour of the model and efficiently stabilizes the RMS error of both SST forecast and analysis during the warming period.

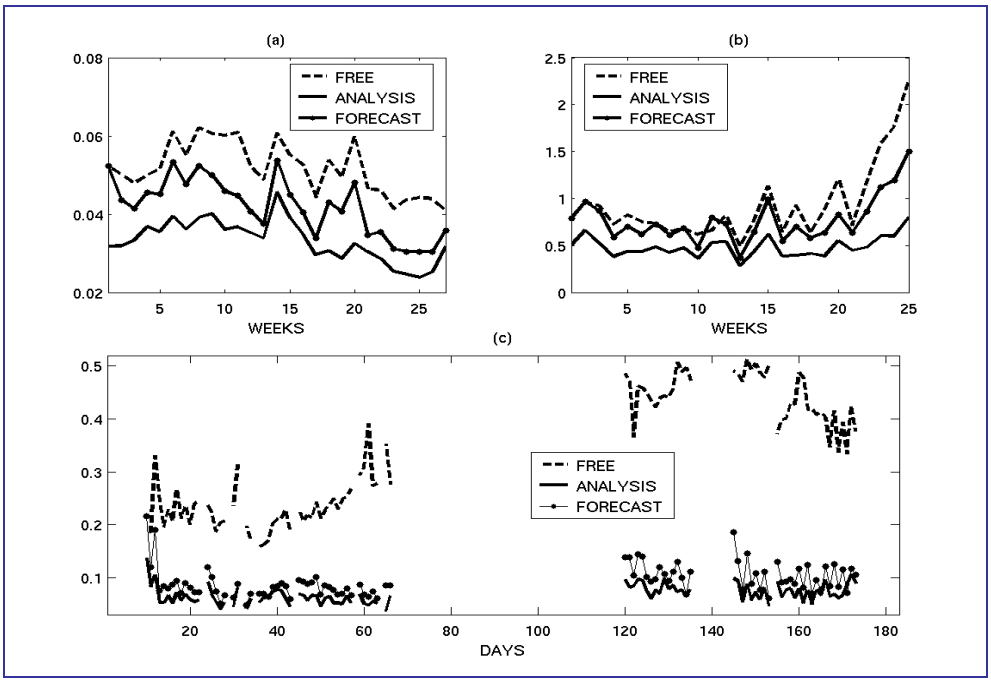

Figure 1-6: (a) RMS misfits (in meters) time series between the model SSH (free run, analysis and forecast of the assimilation system) and the altimetric data. (b) RMS misfits (in °C) time series between the model SST (free run, analysis and forecast of the assimilation system), and the AVHRR SST data. (c) RMS misfits time series between the model SSS (free run, analysis and forecast of the assimilation system) and the Ferrybox SSS data.

The spatial distribution of the average RMS errors for SSH and SST as obtained from for the model free-run and the assimilation system are shown in Figure 1-7 and Figure 1-8, respectively. Regarding the SSH error, one can see that the model was unable to reproduce several mesoscale features (Rhodes cyclonic gyre, Mersha Matruh cyclone) mainly occurring within the Levantine basin south of Crete.

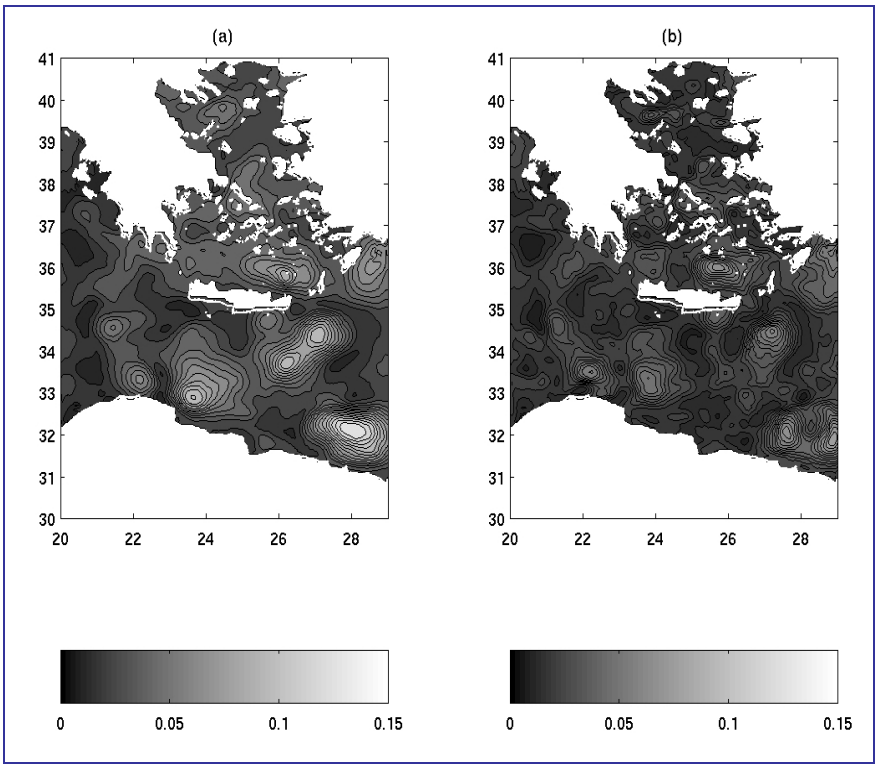

Figure 1-7: Distribution of the average RMS misfits (in meters) between altimetric data and (a) the model free run SSH, and (b) the assimilation system analysis SSH.

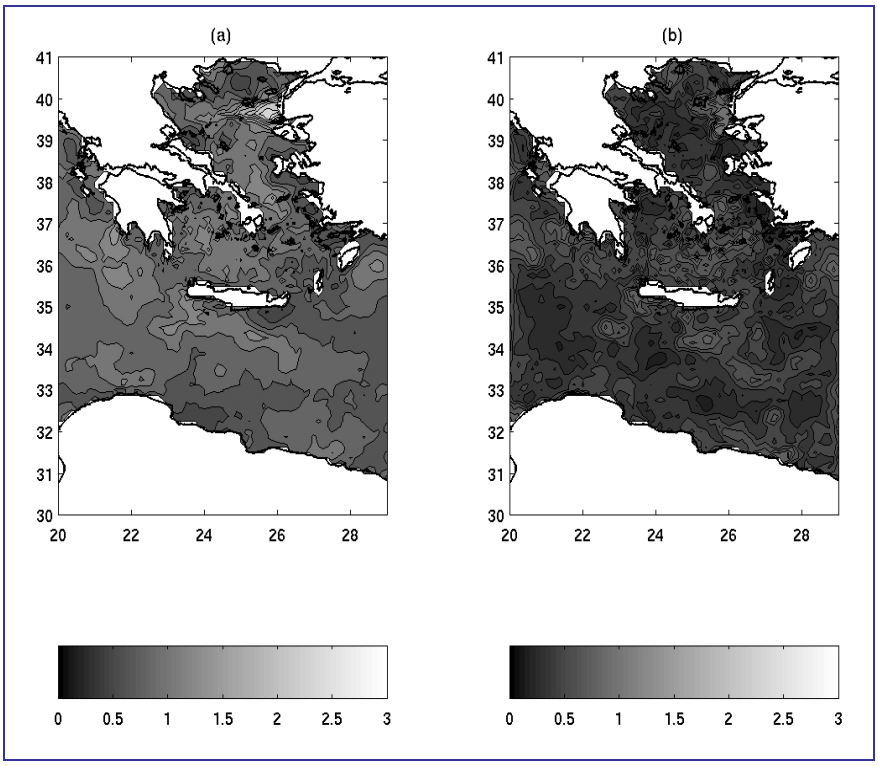

Figure 1-8: Distribution of the average RMS misfits (in °C) between AVHRR SST data and (a) the model free run SST, and (b) the assimilation system analysis SST.

The assimilation system significantly improved this picture while correctly introducing the Mersha-Matruh anticyclone and relocating the Rhodes gyre. The system, however, still needs improvements especially in areas like the Cretan Sea where the error reduction was practically insignificant. The model free-run error for SST shows maxima in the area close to Dardanelles (almost 3 °C) and along the Black Sea waters route within the North Aegean and secondary maxima along the Asia Minor route within the Aegean Sea. Both areas of high errors are probably related to the highly specified nonlinear open boundaries. The assimilation system is very skilful and drastically reduces the SST RMS error all over the modeling area by approximately 50%.

Overall the results obtained so far with the assimilation system are very encouraging, but clearly call for further improvements in the model parameterizations (Dardanelles outflow) and the assimilation scheme. As for the latter we believe that the assimilation system should be extended to the coarse resolution Eastern Mediterranean model which controls a large part of the errors present in the southern sector of the Aegean Sea model. Finally the definition of a realistic SSH climatology needed to convert SLA data to absolute signals remains open for near future studies.

1.3 References

Bretherton, C.S., M.Widmann, V.P.Dymnikov, J.M.Wallace and I.Blade, (1999). The Effective number of degrees of freedom of a time-varying field. J.Climate, 12, 1990-2009.

Hoteit I., D.T. Pham and J. Blum, (2002). A simplified reduced Kalman filtering and application to altimetric data assimilation in the Tropical Pacific. J. Mar. Sys. 36, 101-127.

Houtekamer P. and H.Mitchell, (2001). A sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. Mon. Weather Rev., 123-137.

Jazwinski A.H., 1970: Stochastic processes and filtering theory. Academic Press, New York, 376 pp.

Kontoyiannis H. and D.Ballas, (2005). The scales of the surface temperature and salinity fields in the southern Aegean Sea as derived from Ferry-box space-series measurements. 4th EuroGOOS Confererence, 6-9 June 2005, Brest – France.

Nittis N., Zervakis V., Perivoliotis L., Papadopoulos A., and G.Chronis, (2001). Operational monitoring and forecasting in the Aegean Sea: system limitations and forecasting skill evaluation. Mar. Pol. Bul., 43, 154-163.

Papadopoulos A., Kallos G., Katsafados P., Nickovic S., (2002). The Poseidon weather forecasting system: an overview. Glob. Atmos. Ocean Sys. 8, 218-237.

Soukissian T., Chronis G., Nittis K., and C.Diamanti, (2002). Advancement of operational oceanography in Greece: the case of the POSEIDON system. Global Atmos. Ocean Sys. 8, 119-133.

Wunsch, C., The Ocean Circulation Inverse Problem, Cambridge University Press, 437 pp., 1996.

2 NERC.POL

The Proudman Oceanographic Laboratory implemented an Ensemble Kalman Filter (EnKF) to assimilate satellite derived SST as well as Ferrybox SST data into POLCOMS. An Ensemble Optimal Interpolation (EnOI) scheme was implemented afterwards to allow the use of a system that could be easily set up and had less memory and computing time requirements than the EnKF.

2.1 The EnKF

The Ensemble Kalman Filter (Evensen, 1994) is an adaptation of the classical Kalman Filter that represents the evolution of the *probability density function* (*pdf*) of the system using ensemble simulations. These are started from a set of initial conditions (*a priori ensemble*) of mean that of a first guess and variance given by the estimated uncertainty of that first guess. The simulations are then integrated forward in time, giving a succession of ensemble forecasts (*a posteriori ensembles*). Model error can be parameterized by adding a stochastic term to the variables that are supposed to account for most of that error. The best estimate of the system state at any time is provided by the ensemble mean at that time:

$$\overline{x}^{e} = \sum_{i=1}^{N} x_i$$

where x_i represents the state vector corresponding to the i^{th} member of the ensemble.

Following Evensen 2003, the analysis step can then be formulated as:

$$x_i^a = x_i^b + K_e\left(y_i^o - H\ x_i^b\right)$$

expressing that for each ensemble member i of an ensemble of n members ($i=1, \ldots, n$) the analysis, $\boldsymbol{x}^a_i$, results from a correction of the forecast, x^b_i. The correcting term is the product of the gain matrix $\boldsymbol{K}_e$ by the misfit between the observations y^o_i and their model equivalent $\boldsymbol{Hx}^b_i$. The Kalman or gain matrix takes the expression

$$K_e = P_e^f H^T \left(H\ P_e^f H^T + R\right)^{-1}$$

where the matrix $\boldsymbol{P}^f_e$ is obtained through the formula:

$$P_e^f = \overline{(x^f - \overline{x^f}^e)\ (x^f - \overline{x^f}^e)^T}^e$$

and $\overline{\ }^e$ indicates the average over the ensemble.

Similarly, the analyzed covariance matrix is obtained from:

$$P_e^a = \overline{(x^a - \overline{x^a}^e)\ (x^a - \overline{x^a}^e)^T}^e .$$

A difficulty that was encountered in the application of the EnKF forecast-analysis scheme implemented with POLCOMS was the increased requirements of memory allocation both in forecast and analysis mode. This was due to the introduction of new variables that represented the effects of model error that we wanted to parameterize, on the one hand. This was a limitation in the computing platforms available at NERC.POL, which are Massivelly Parallel computers, as the computations involving the new variables were not completely parallelized and therefore the variables couldn't be memory distributed. Although less acute, the problem was the same for the analysis step, which again had to be performed in one processor (the master processor). For a given size of the state vector, this would impose limitations on the number of members in the ensemble and the number of observations that could actually be used for assimilation (in case where a great number of observations were available).

Ensemble Optimal Interpolation

$$x^a = x^b + K_s\left(y^o - H\ x^b\right)$$

With less memory and CPU requirements, the EnOI scheme carries only one forecast and uses a prescribed static ensemble to compute the background error covariances needed to estimate the gain matrix, that in this case can be noted $\boldsymbol{K}_S$. The analysis corrects the forecast following:

The static forecast can be chosen following the season or specific situation (dynamic regime). For the present study, the ensemble consisted of a group of model states obtained during the winter period of a reference simulation. Due to the great size of the state vector, the number of members composing the static ensemble had to be restricted to 15. This would affect the quality of the analyses, which improved with the size of the static ensemble.

For the same reason, the number of observations used for assimilation was limited to a maximum of 6000 points in the observation space.

2.2 Observational Data Sets Used

Two sets of observations were used:

(a) The SAF SST satellite products, generally available for the area and period concerned at 02h00 10h00 12h00 and 20h00. These correspond to the central times of each mosaic obtained from up to 3 consecutive satellite orbits. As orbits are 102 minutes apart, the time difference between data in a same mosaic can be slightly larger than 200 minutes.

(b) The Ferrybox SST data used were extracted from time series along ferry tracks between Birkenhead and Belfast, as well as between Birkenhead and Dublin throughout the period from 18 October until 17 November 2004. These data were provided by ferry lines that run 6 days a week, between Birkenhead and Belfast/Dublin and back. The ferry departs between 10-11 a.m.(p.m.) and crosses the Irish Sea in about 6 hours. It samples temperature, salinity, turbidity and chlorophyll at 2 m depth every minute, which gives a spatial resolution of ~100 m on the boat's track, for which time and position information are available. The report for WP-2 gives further details about this dataset. For the application presented here, SST data from were sampled along the track at a spatial resolution close to that of the model.

An example of the SAF and Ferrybox SST data used in an analysis is given in. This shows the low coverage of the SAF data, due to cloud cover, for the corresponding time.

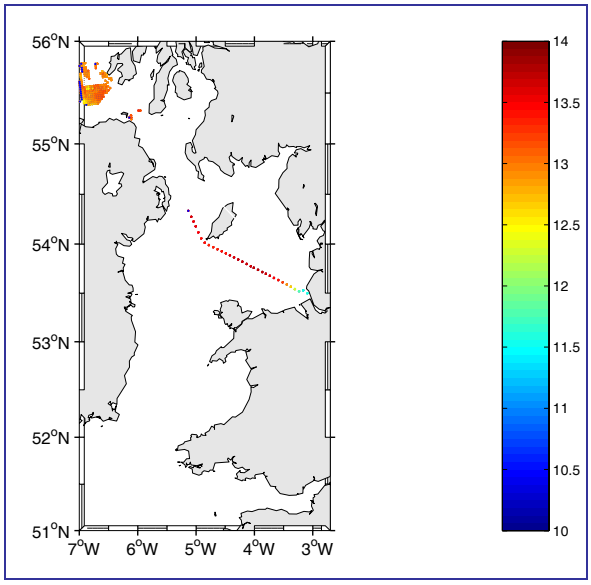

Figure 2-1: SAF and Ferrybox SST data available in a time-window of 5 hours around 2 a.m. on 20 October 2004. The Ferry track corresponds to a Belfast – Birkenhead track.

2.3 Free Simulations

In this study, POLCOMS was set up at a resolution of 1.8 km in the Irish Sea region, covering the area 7° W- 2.75° W, 51° N-56° N with 301 x 173 horizontal grid points and 32 sigma levels. Initial and boundary conditions were extracted from the output of a wider-area, 12 km resolution model, on 1 January 2004 at 0 hrs. A 2.5 turbulent closure scheme (Mellor-Yamada-Galperin) computed eddy viscosities and diffusivities, while sea surface heat and momentum fluxes were calculated from ERA40 atmospheric fields using bulk formulae. Two *free* simulations (*i.e.* without assimilation) were performed for the year, one using climatological freshwater input and another with real river input. In this report only the results obtained from the simulation that used real river input will be presented, as they are considered to better represent reality. This simulation will be referred to as $Free_{realriv}$ hereafter.

2.4 Constrained Simulations

The constrained simulations were started from the output of the free simulation corresponding to 18 October 2004 at 0 hrs. This is the date at which the time series of high quality Ferrybox data starts. From this initial date and until 17 November 2004, the model was integrated in periods of 24 hours, giving an output of the state vector at 0 hrs each day. This forecast was then used as the input for the assimilation step, that produced the analyzed state vector as the corrected state with which to start the model integration of the following day. The concatenation of forecast and analysis steps was managed by a UNIX script that launched them in a sequential way, also providing the necessary observation files at analysis times and using the forecast as an analysis in case of failure of this step (due to lack of observations or other unforeseen incident that would disable pre-operational functioning).

The state vector consisted of the three-dimensional (3D) temperature, salinity, instantaneous zonal and meridian current components, and the two-dimensional (2D) sea surface elevation.

Two constrained simulations were run. A first simulation assimilating SAF satellite SST products (SAF_real) and a second one (experiment SAF_FB_real) including the same satellite data combined with Ferrybox data obtained within a similar time window (Figure 2-2). This discarded more than half of the Ferrybox data available each day for the track concerned, but it enabled a fair estimation of the information content of this data set with respect to the satellite product used. Table 2-1 presents the simulations performed.

Table 2-1: Set of simulations runs.

Free simulation	Free_real
Assimilation of SAF SST	SAF_real
Assimilation of SAF and Ferrybox SST	SAF_FB_real

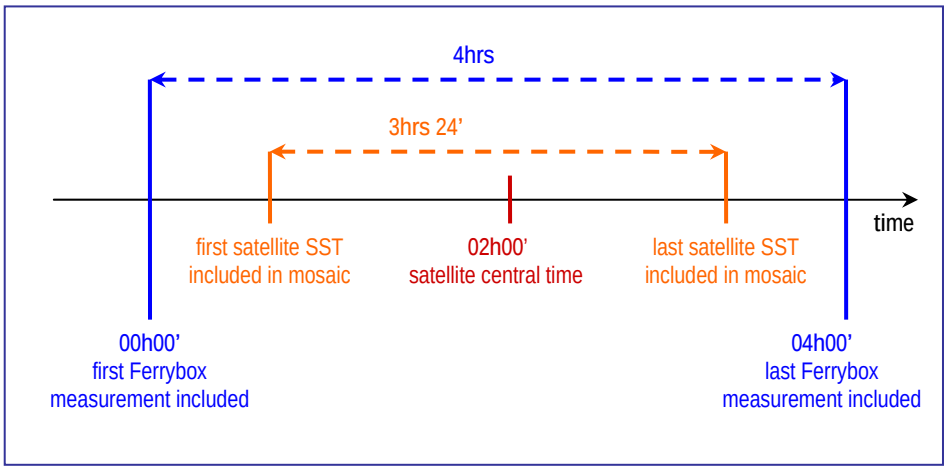

Figure 2-2: Time windows of the SAF satellite SST and the Ferrybox data used for assimilation.

2.5 Results

Model output from the free and constrained simulations was systematically compared to the SAF and Ferrybox data available at assimilation times. This not only includes the data actually assimilated, but some additional data that were discarded for assimilation, in order to avoid too high computing load. Results show that assimilation of SAF data produces a substantial decrease of the forecast RMS error (comparison of Figure 2-3 and Figure 2-4). Figure 2-5 gives us a first impression of the improvement of adding Ferrybox data to SAF satellite data, with both forecast and analysis RMS substantially reduced. The number of observations used in each element of the assimilation is shown in Figure 2-6.

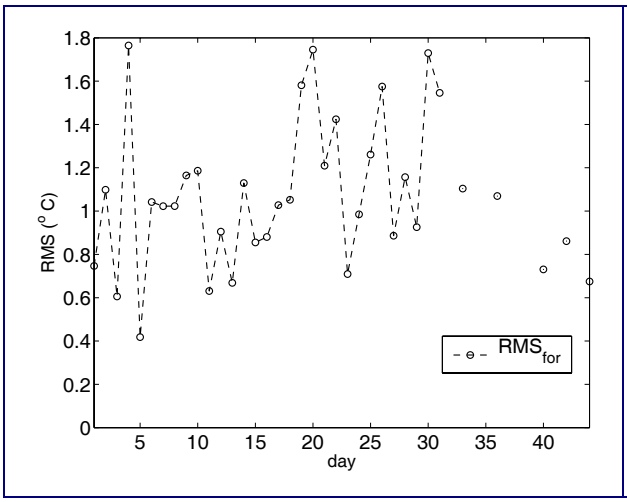

Figure 2-3: Forecast RMS error for simulation Free_real with respect to the SAF and Ferrybox available data within the times windows chosen around each forecast step.

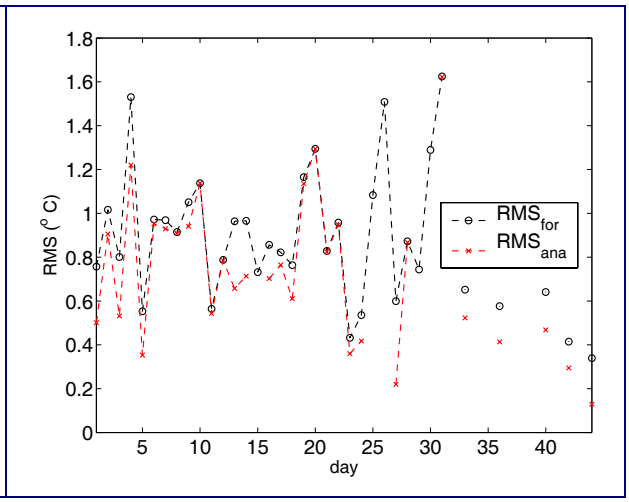

Figure 2-4: Forecast RMS error for simulation SAF_real with respect to the SAF and Ferrybox data set available data within the times windows chosen around each forecast (analysis) step.

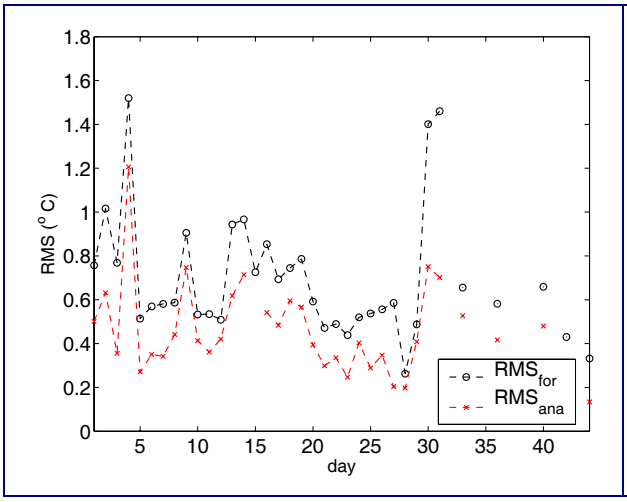

Figure 2-5: Forecast RMS error for simulation SAF_FB_real with respect to the SAF and Ferrybox data set available data within the times windows chosen around each forecast (analysis) step.

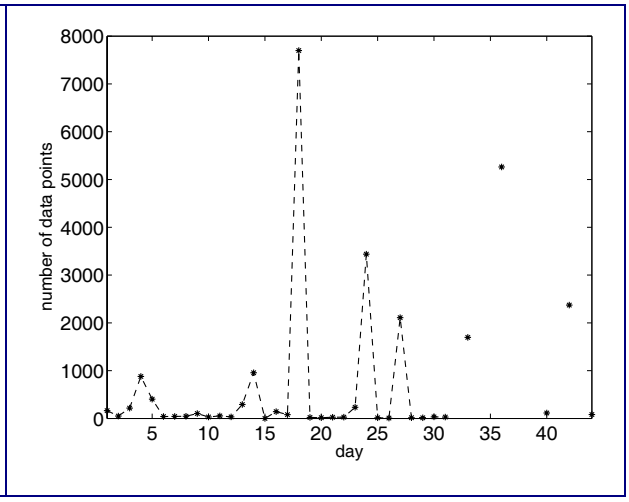

Figure 2-6: Number of observations available at each analysis time. This is the total number of observations used to compute the RMS values presented in figures 2-3 to 2-5.

The impact of assimilating Ferrybox SST data upon the horizontal representation of the surface temperature is illustrated in Figure 2-7, that shows the the forecast on 21/10/04 for simulations SAF_real and SAF_FB_real. The data available for assimilation, within the fixed time window, in the preceding analysis step are those presented in Figure 2-1. As we can see from Figure 2-7 inclusion of the Ferrybox data introduces structures that were not present in the simulation assimilating SAF SST data only. The temperature is raised up to 1 °C in some areas off the Liverpool Bay with fine scale, frontal structures revealing.

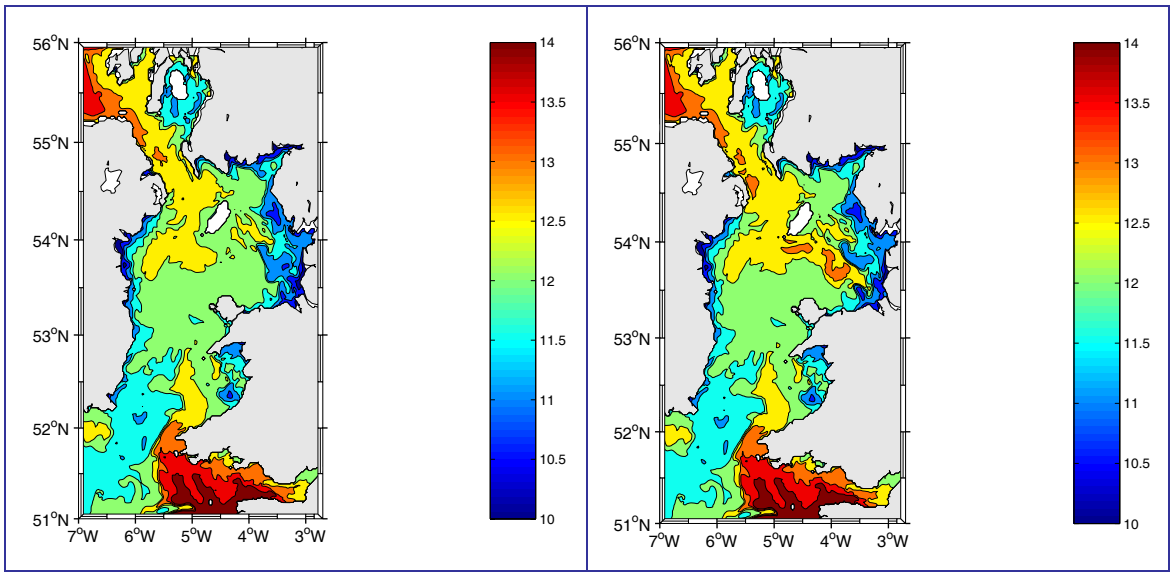

Figure 2-7: Forecast for 21 October 2004 for the simulations (SAF_real (left) and SAF_FB_real (right). Assimilation of Ferrybox data corresponding to the line Birkenhead – Belfast on that date (Figure 2-1) increases the temperatures off the Liverpool Bay up to 1 °C and introduces a frontal structure around 54° N, between 3° W and 4° W.

2.6 Conclusions

In this work, we have investigated the impact of assimilating Ferrybox SST data, upon the representation of the surface temperature in a 1.8 km setup of the Irish Sea. One of the questions we aimed at answering was if the Ferrybox dataset could contribute with some information that was not present in the widely used satellite products. In order to tackle this question, an experiment was designed in which satellite products were assimilated in a first simulation. These were combined with Ferrybox data available in a similar time window for assimilation in a second simulation.

Although the time window used to select the Ferrybox data actually used for assimilation left out most of the data available from this set, the results obtained when these observations are used show a great decrease of the RMS error and the appearance of fine scale structures not present in the results obtained assimilating SAF data only. This can be explained by the actual scarcity of SAF data in this region in periods of heavy cloud coverage. In those times, the in situ, high resolution and regular Ferrybox data can provide precious information for the areas surrounding the track with high (anti-)correlation values.

Even when a lot of other observational data sets are available, the Ferrybox SST is an important piece of information, due to its high accuracy and high-resolution sampling. Finally, this set provides unique information at sampling frequencies required by smaller scale process studies.

2.7 References

Evensen, G., (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143-10162.

3 Summary

Both case studies indicate that Ferrybox data can have an impact on the performance of 3-dimensional models by reducing the RMS errors. However both studies indicate that the influence of the Ferrybox data on the solution is geographically limited to an area adjacent to the ferry route.

The most positive value of the Ferrybox data is that it is not constrained by atmospheric conditions, its availability being limited by ferry downtime, unlike the satellite-derived surface properties. This means that, in more northerly latitudes at least with considerable cloud cover, the Ferrybox provides considerably more useful data than satellites.