REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-10-2009	Journal Article	

4. TITLE AND SUBTITLE
Super-Ensemble Techniques: Application to Surface Drift Prediction During the DART06 and MREA07 Campaigns

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
060153N

6. AUTHOR(S)
L. Vandenbulcke, J. Beckers, F. Lenartz, A. Barth, P. Poulain, M. Aidonidis, J. Meyrat, F. Ardhuin, M. Tonani, C. Fratianni, L. Torrisi, D. Pallela, J. Chiggiato, M. Tudor, J. Book, P. Martin, G. Peggion, M. Rixen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
73-9858-A8-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/JA/7330-09-9070

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The prediction of surface drift of floating objects is an important task, with applications such as marine transport, pollutant dispersion, and search-and-rescue activities. But forecasting even the drift of surface waters is very challenging, because it depends on complex interactions of currents driven by the wind, the wave field and the general prevailing circulation. Furthermore, although each of those can be forecasted by deterministic models, the latter all suffer from limitations, resulting in imperfect predictions. In the present study, we try and predict the drift of two buoys launched during the DART06 (Dynamics of the Adriatic sea in Real-Time 2006) and MREA07 (Maritime Rapid Environmental Assessment 2007) sea trials, using the so-called hyper-ensemble technique: different models are combined in order to minimize departure from independent observations during a training period; the obtained combination is then used in forecasting mode. We review and try out different hyper-ensemble techniques, such as the simple ensemble mean, least-squares weighted linear combinations, and techniques based on data assimilation, which dynamically update the model's weights in the combination when new observations become available. We show that the latter methods alleviate the need of fixing the training length a priori, as older information is automatically discarded.

15. SUBJECT TERMS
super-ensemble, multi-model, surface drift

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jeffrey Book
Unclassified	Unclassified	Unclassified	UL	19	19b. TELEPHONE NUMBER *(Include area code)* 228-688-5251

Progress in Oceanography 82 (2009) 149–167

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Super-ensemble techniques: Application to surface drift prediction

L. Vandenbulcke [a,*], J.-M. Beckers [a], F. Lenartz [a], A. Barth [a], P.-M. Poulain [b], M. Aidonidis [c], J. Meyrat [c], F. Ardhuin [c], M. Tonani [d], C. Fratianni [d], L. Torrisi [e], D. Pallela [e], J. Chiggiato [f,j], M. Tudor [g], J.W. Book [h], P. Martin [h], G. Peggion [i], M. Rixen [j]

[a] GeoHydrodynamics and Environment Research, Université de Liège, Bat. B5, 17, Allée du 6 Août, 4000 Liège, Belgium
[b] Istituta Nazionale di Oceanografia e di Geofisica sperimentale (OGS), Trieste, Italy
[c] Service Hydrographique et Océanographique de la Marine, 13 rue du Chatellier, 29200 Brest, France
[d] Istituto Nazionale di Geofisico e Vulcanologia, Bologna, Italy
[e] Palazza A.M., Viale dell'Universita' 4, 000185 Rama, Italy
[f] ARPA Emilia Ramagna, Servizia Idro Meteorologica, Viale Silvani 6, 40122 Balagna, Italy
[g] DHMZ Meteorologicol and Hydrological Service, Zagreb, Crootio
[h] US Naval Research Labaratary, Stennis Space Center, MS 39529, United States
[i] University of New Orleans, 2000 Lakeshore Dr., New Orleans, LA 70148, United Stotes
[j] NATO/SACLANT Undersea Research Centre, La Spezia, Italy

ARTICLE INFO

Article history:
Received 22 January 2009
Received in revised farm 3 June 2009
Accepted 8 June 2009
Available online 13 June 2009

Keywords:
Super-ensemble
Multi-model
Surface drift

ABSTRACT

The prediction of surface drift of floating objects is an important task, with applications such as marine transport, pollutant dispersion, and search-and-rescue activities. But forecasting even the drift of surface waters is very challenging, because it depends on complex interactions of currents driven by the wind, the wave field and the general prevailing circulation. Furthermore, although each of those can be forecasted by deterministic models, the latter all suffer from limitations, resulting in imperfect predictions. In the present study, we try and predict the drift of two buoys launched during the DART06 (Dynamics of the Adriatic sea in Real-Time 2006) and MREA07 (Maritime Rapid Environmental Assessment 2007) sea trials, using the so-called hyper-ensemble technique: different models are combined in order to minimize departure from independent observations during a training period; the obtained combination is then used in forecasting mode. We review and try out different hyper-ensemble techniques, such as the simple ensemble mean, least-squares weighted linear combinations, and techniques based on data assimilation, which dynamically update the model's weights in the combination when new observations become available. We show that the latter methods alleviate the need of fixing the training length *a priori*, as older information is automatically discarded.

When the forecast period is relatively short (12 h), the discussed methods lead to much smaller forecasting errors compared with individual models (at least three times smaller), with the dynamic methods leading to the best results. When many models are available, errors can be further reduced by removing colinearities between them by performing a principal component analysis. At the same time, this reduces the amount of weights to be determined.

In complex environments when meso- and smaller scale eddy activity is strong, such as the Ligurian Sea, the skill of individual models may vary over time periods smaller than the forecasting period (e.g. when the latter is 36 h). In these cases, a simpler method such as a fixed linear combination or a simple ensemble mean may lead to the smallest forecast errors. In environments where surface currents have strong mean-kinetic energies (e.g. the Western Adriatic Current), dynamic methods can be particularly successful in predicting the drift of surface waters. In any case, the dynamic hyper-ensemble methods allow to estimate a characteristic time during which the model weights are more or less stable, which allows predicting how long the obtained combination will be valid in forecasting mode, and hence to choose which hyper-ensemble method one should use.

1. Introduction

The prediction of the drift of objects floating at the surface of the ocean has various applications, for example tracking of floating mines or pollutants such as tar balls, dispersion of algae blooms,

* Corresponding author.
E-mail address: luc.vandenbulcke@ulg.ac.be (L. Vandenbulcke).

0079-6611/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pocean.2009.06.002

marine transport, search-and-rescue activities, etc. However, due to multiple reasons whose effects add up, drift prediction remains a very challenging task. Even small errors in estimation can drastically change the subsequent particle trajectories (Griffa et al., 2004). Even when one predicts the drift of buoys configured to closely track the drift of surface waters, and hence only the ocean current should be taken into account, it is still useful to also take surface wind, waves, tides, etc. into consideration. Indeed, most ocean current models do not include wave-driven currents at all, and wind-driven currents are not fully accurate. However, all these currents contribute (in a complex way) to the real surface water drift, and furthermore interact with one another. Thus missing dynamics in ocean current models can be partially accounted for using empirical methods with direct model predictions of the forcing fields (winds and waves) for these dynamics. Finally, when one tries to predict the drift of various floating objects, other parameters should be considered, such as the specific hydrodynamic drifter response.

Even though most of these contributions can be forecast by deterministic models (albeit with some limitations inherent to the models), there is not yet a deterministic method to combine them in order to reproduce the floating object drift; completely-coupled deterministic models that take all these processes into consideration are just now under development. In the present study, we instead use multi-model methods to try and empirically combine individual models of different processes that are all directly or indirectly related to surface drift. Super-ensembles (SE), which combine different models of the same physical processes, were applied within the atmospheric community by Krishnamurti et al. (1999) some years before the oceanic community took on. Other atmospheric studies followed, see e.g. Shin and Krishnamurti (2003a,b); Williford et al. (2003); Yun et al. (2003, 2005); Mutemi et al. (2007). Nowadays, other communities also apply the technique (e.g. oceanography, hydrology, paleoclimatology, etc.), as they all realize its low cost, but large benefit. Generally speaking, the technique could be applied to every field where different concurrent models aim at predicting the same variable, or even where different models predict different variables which are all somehow related to the desired output variable. In the latter case, the technique is rather called hyper-ensemble (HE); it was first introduced in the oceanic community (Rixen and Ferreira-Coelho, 2007).

In the present study, we forecast surface drift using linear HE methods both with static and dynamic weights, the latter allowing the weights to evolve smoothly in time. Section 2 is devoted to the description of the models and observational data used in two experiments: the DART06 sea trial in the Adriatic Sea, and the MREA07 campaign in the Ligurian Sea. The HE methods are described in Section 3. We will then focus on two case studies, one where the drifter is predominately influenced by a mean-kinetic-energy environment (the Western Adriatic Current) and one where the drifter is predominately influenced by an eddy-kinetic-energy environment (Ligurian Sea). The results are then shown in Section 4 and a summary and the conclusions are given in Section 5.

2. Models and data

Surface drift of floating objects depends on various factors. It is strongly determined by the ocean surface currents. However, the hydrodynamic models used to forecast the currents have chaotic components, have incomplete representations of the underlying physics, and have uncertainties on forcing fields and model parameters. For a complete discussion of error causes in hydrodynamic models, see e.g. Lermusiaux et al. (2006). The hydrodynamic models used in both experiments have high resolutions (between 1/16° and 1/100°), and therefore represent many smaller scale processes that are difficult to correctly phase and forecast. The fact that the models have energies at such scales is ultimately important for successful HE modeling, but phase problems can easily lead to higher forecast errors than for lower resolution models (no energy at these scales) if the higher resolution models are not corrected in some way. On top of this, even with this high resolution, many phenomena at yet smaller scales are not represented, whereas the real surface drift depends on every scale present.

Paldor et al. (2004) shows that instantaneous winds have more influence on surface drift than climatic surface currents; Rixen and Ferreira-Coelho (2007) confirm this by showing that in an atmospheric–oceanic hyper-ensemble, the (weighted) wind model has more importance; ocean advection has less impact. However, the wind-driven surface current is still poorly understood. Observations show, in addition to inertial oscillations, a drift of the order of 2–4% of the wind speed with directions that vary from 0° to 30° to the right of the wind in the Northern Hemisphere, and to the left in the Southern Hemisphere (Tsahalis, 1979). These variations may be understood as the combination of a wave-induced Stokes drift, roughly aligned with the wind, and a drift due to the wind-driven current. The magnitude and deflection angle of this current depend strongly on the vertical structure of turbulence. For example, the classical Ekman (1905) theory with a constant eddy viscosity give a 45° deflection angle, while linear eddy viscosity profiles give deflections of the order of 10° (Madsen, 1977). Recent evidence for strong mixing in the upper ocean [e.g. (Agarwal et al., 1992)] suggest that the eddy viscosity profile may be piecewise-linear with a strong surface value. This should produce a surface current limited to about 0.5% of the wind speed in open ocean conditions without stratification, and about 1% with a strong stratification. Given that the surface Stokes drift (see below) is of the order of 1.2% of the wind speed, the total surface drift explained by models with realistic mixing is of the order of 2% of the wind speed (Rascle et al., 2006; Rascle and Ardhuin, 2009). This is generally on the low side of the reported values for surface drift. This difference may be due to fetch variations (e.g. laboratory compared to field conditions), convergence-related biases (such as caused by Langmuir circulations) or yet unknown processes. As a "rule-of-thumb", we will consider that the wind sets up a surface current of roughly 3% of the wind speed, 15° to the right of the downwind direction. But similarly to the ocean models mentioned before, the atmospheric models used to forecast the wind field suffer of their own limitations: they are also chaotic, also have only an incomplete representation of the real atmospheric physics, etc.

The wave theory leads to the so-called Stokes drift, which induces a movement of water particles in the direction of the waves. The displacement velocity depends on the ratio of wave height and wavelength; it also strongly decreases with depth and becomes negligible at a depth equal to a fourth of the wavelength. The Coriolis force induces yet another net transport, the so-called Hasselmann drift, which depends on the turbulence, and has a direction opposed to the Stokes drift. The sum of vertically-integrated net transports of the Stokes and Hasselmann drifts is zero, leading to a zero net water transport. However, the different vertical profiles for Stokes and Hasselmann drifts indicate that the former is more important than the latter at the surface, leading to a net surface transport in the direction of the waves (below the surface, there is a transport in the opposed direction).

Finally, surface drift still depends on other phenomena such as tides.

Most of the drifters used in the DART06 and MREA07 experiments were CODE drifters manufactured by Technocean (model Argodrifter). CODE designs were developed by Davis (1985) to measure the currents in the first meter under the sea surface. More details about these drifters can be found in Poulain (1999) and Ursella et al. (2006). Measurements with dye (D. Olsen, Personal Communication) and through direct measurements of relative flow

(P.-M. Poulain, Personal Communication) revealed that the CODE drifters follow the surface currents to within 2–3 cm/s. The wind-driven components of the CODE drifter velocities, including Ekman currents and slip, were recently assessed by Poulain et al. (2009) and related statistically to ECMWF winds. Using complex linear regression models, they found that the wind-driven currents amount to 1% of wind speed and are rotated by 28° to the right of the wind.

The majority of the drifters were localized by Glob al Positioning System (GPS) at hourly intervals. Their data were telemetered via the Argos system orbiting on the NOAA satellites. The drifter positions were edited for outliers using automatic statistical and manual procedures (Barbanti et al., 2007; Ursella et al., 2006).

Finally, lets note that the HE methods, inclusive the "tricks" (discussed in Section 3), might actually also account for the slip and leeway response of the particular drifters considered.

2.1. DART06 experiment

We first try and predict the displacement of drifters launched in the Adriatic Sea during the DART06 sea trials; drift data for the region were compiled by Veneziani et al. (2007). During this campaign, extensive data sets were collected by multiple means, and made available in near real-time. Drifters were launched and data was made available in near real-time by Istituto Nazionale di Oceanografia e di Geofisica sperimentale (OGS) and the NATO/SACLANT Undersea Research Centre (NURC). Model predictions of the Gargano region (41°45 N, 16°E) were used to direct the launching of pairs of drifters with the goal of maximizing the coverage of the sampling area. Some drifters were found to separate at locations and in the directions given by the model finite-size Lyapunov exponents (FSLE) (Haza et al., 2007). The trajectories are shown in Fig. 1; we will focus only on drifter a06956 (Barbanti et al., 2007) flowing around the Gargano peninsula as it exhibits a typical behavior. We consider only the first week of the drifter trajectory, as afterward at least one model does not cover the area anymore.

At the same time, a wide range of atmospheric, ocean and wave models were provided operationally. However, increasing the complexity of the problem could lead to less accurate results if over-fitting occurs (Everitt, 2002), and hence only two wind models and two hydrodynamic models are used in the HE combinations (i.e. no wave models are used). The following models were used in the present study:

1. Meteo France Aladin, output fields provided by the *Service Hydrographique et Océanographique de la Marine* (SHOM), http://www.cnrm.meteo.fr/aladin. The horizontal resolution is 0.1°; hourly model outputs are available. This model is further referred to as *Aladin-FR*. The predicted drift is obtained from the following rule-of-thumb: the time interval multiplied by 3% of the wind speed, with a direction 15° to the right.
2. Aladin/Croatia, run by the Meteorological and Hydrological Service of Croatia (see Ivatek-Sahdan and Tudor (2004) and http://meteo.hr/index_en.php). The horizontal resolution is 0.03° and

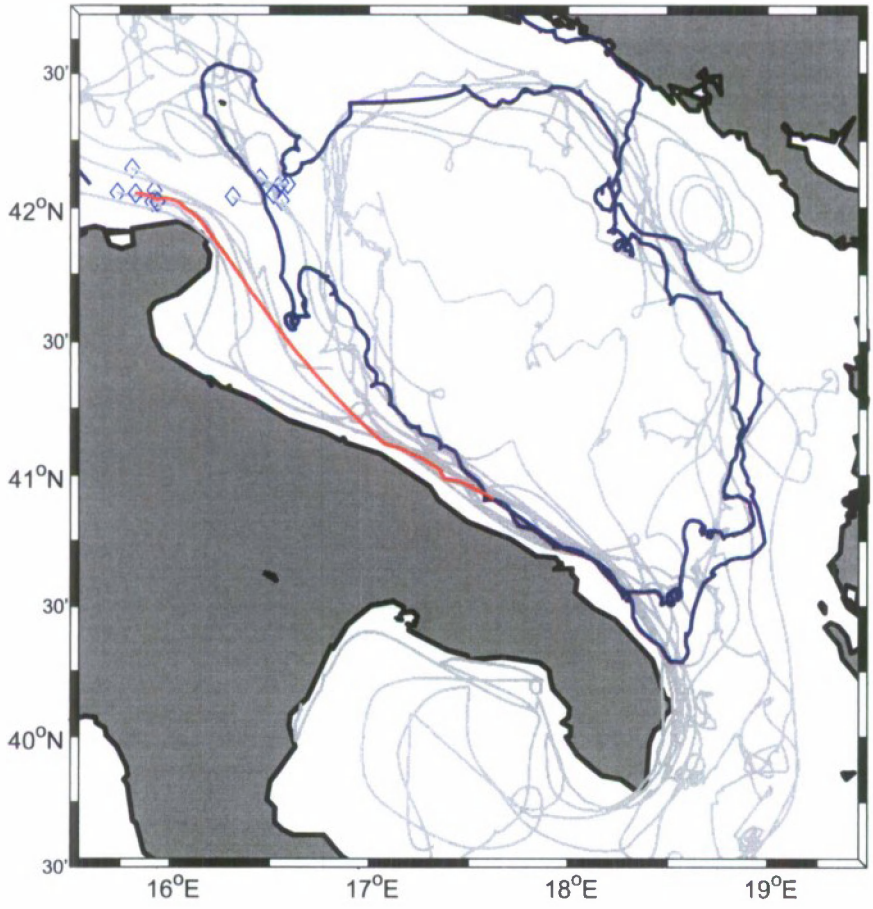

Fig. 1. Trajectories of the drifters launched during DART06. The dark track corresponds to drifter a06956 studied later in this paper, and called "track 1" further on; the first week of data, which is effectively used in this study, is in red. All other tracks are gray. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the "time resolution" (i.e. when the model outputs are saved to disk) is 3 h. This model is further referred to as *Aladin-HR*. The predicted drift is again obtained by the same rule-of-thumb.

3. AdriaROMS, an operational ocean forecasting system for the Adriatic Sea run by the HydroMeteorological Service of ARPA Emilia Romagna, Bologna, Italy (see e.g. Chiggiato and Oddo (2008) and references herein, and http://www.arpa.emr.it/sim/?mare), further referred to as *ROMS*. The resolution is 0.025° and 3 h.
4. NRL (Navy Research Laboratory) regional Navy Coastal Ocean Model. NCOM was implemented over the Adriatic sea (Martin et al., 2009), and subsamples were made available in near real-time; here we use the area2 subset covering the central Adriatic Sea only, with a horizontal resolution of 0.08° and time resolution of 3 h. It is further referred to as NCOM_D06.

The reader is referred to the official documentation of the relevant operational centers or above cited journal papers for descriptions of the models. All in all, with the constant (bias) model added, there are 5 weights to determine in order to obtain a linear HE (which may be real or complex numbers depending on the method used), or less if principal component analysis (PCA, see Section 3.3) is applied beforehand.

2.2. MREA07 experiment

We also try out the hyper-ensemble techniques with data from the MREA07 experiment in the Ligurian Sea. This campaign also aimed at collecting a vast amount of observations, and drifters data were again provided by NURC and OGS. The trajectories are shown in Fig. 2 [see (Zanasca et al., 2007)]. We focus only on the entire track a74875 later in this study.

At the same time, multiple models were applied to the domain. We again use two atmospheric models and two hydrodynamic models in our ensemble. In order to add some complexity, we will also include a Stokes drift model, even though remembering that it might be correlated to the wind contribution. Furthermore, observed drifter trajectories (see Fig. 2) indicate that the inertial oscillations are quite important. Hence, we also add a synthetic model corresponding to a circular trajectory. This was not necessary in the case of the DART06 experiment, where the considered drifter is mainly constrained by the relatively strong Western Adriatic Current (WAC), leaving little contribution to inertial oscillations. In the Ligurian Sea, the inertial period is about 17.9 h. Of course, this synthetic model by itself will not be able to represent real drifter trajectories, because it lacks the correct amplitude and phase. However, when this is corrected for during the training period, and a bias model is also considered, the obtained synthetic forecast may correspond surprisingly well to reality, particularly if other currents, winds, etc. are weak. In an ensemble of models, the synthetic model may compensate incorrect (e.g. dephased) inertial oscillations of some models.

All in all, the following models were used:

1. Meteo France Aladin (provided by SHOM). This model is further referred to as *Aladin-FR*; predicted drift is obtained from the same rule-of-thumb as before. Horizontal and time resolution are 0.1° and 1 h, respectively

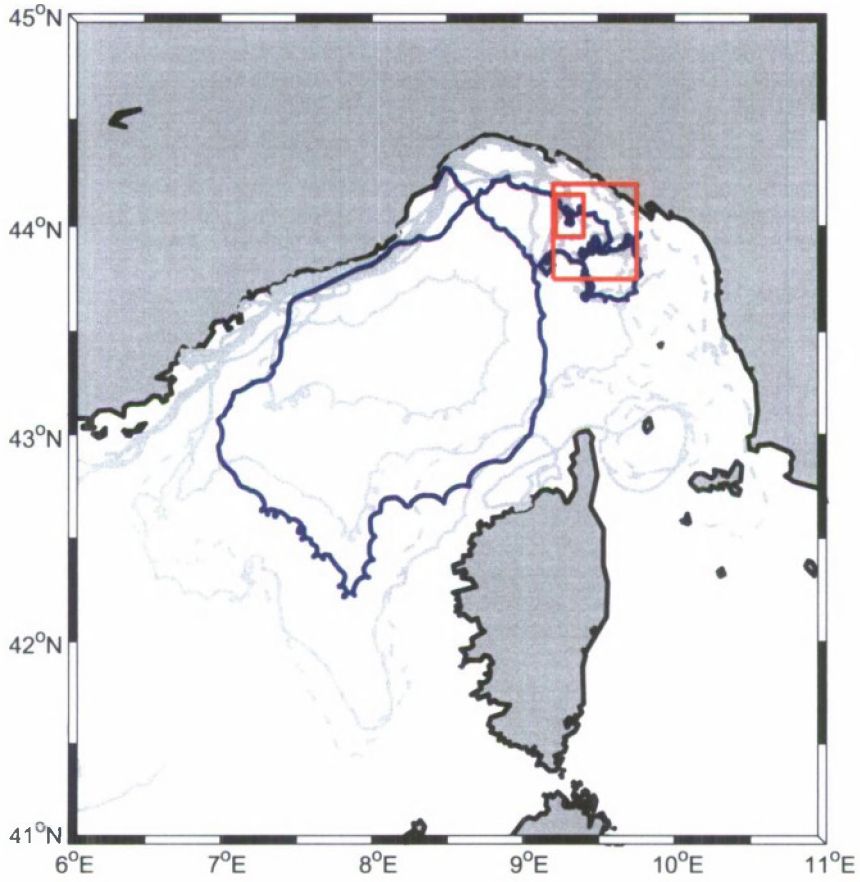

Fig. 2. Trajectories of the drifters launched during MREA07. The dark track corresponds to drifter a74875 (Zanasca et al., 2007) also studied later in the paper, and called "track 5". The two red boxes correspond to later Fig. 14 (largest box) and Fig. 16 (smallest box). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. COSMO-ME (www.cosmo-model.org/content/tasks/operational/default.htm) run operationally by CNMCA – Italian Meteorological Service (http://www.meteoam.it), further referred to as *COSMO-ME*; and again drift is obtained from the rule-of-thumb. The resolutions are 0.03° and 1 h.
3. Mediterranean Forecasting System run by INGV, Bologna, Italy, see Pinardi et al. (2003) and (http://www.bo.ingv.it/mfs/) for the whole forecasting system, and Tonani et al. (2008) for the model itsef, further referred to as *MFS*. Resolutions are 0.0625° and 1 day.
4. NRL NCOM (see Coelho et al. (in press)), further referred to as *NCOM_M07*, with resolutions 0.005° and 1 h.
5. WaveWatch III (SHOM), further referred to as *CMO WW3*. The resolution is 0.1° and 3 h. The predicted drift is obtained as the time interval multiplied by the velocity: the latter is obtained from the wave model as $3.2\frac{H_s^2}{T_m^2}$, where H_s is the significant wave height and T_m the mean period of a broad spectrum of waves (Carniel et al., 2002).
6. a synthetic model of inertial oscillations with a period 17.9 h.

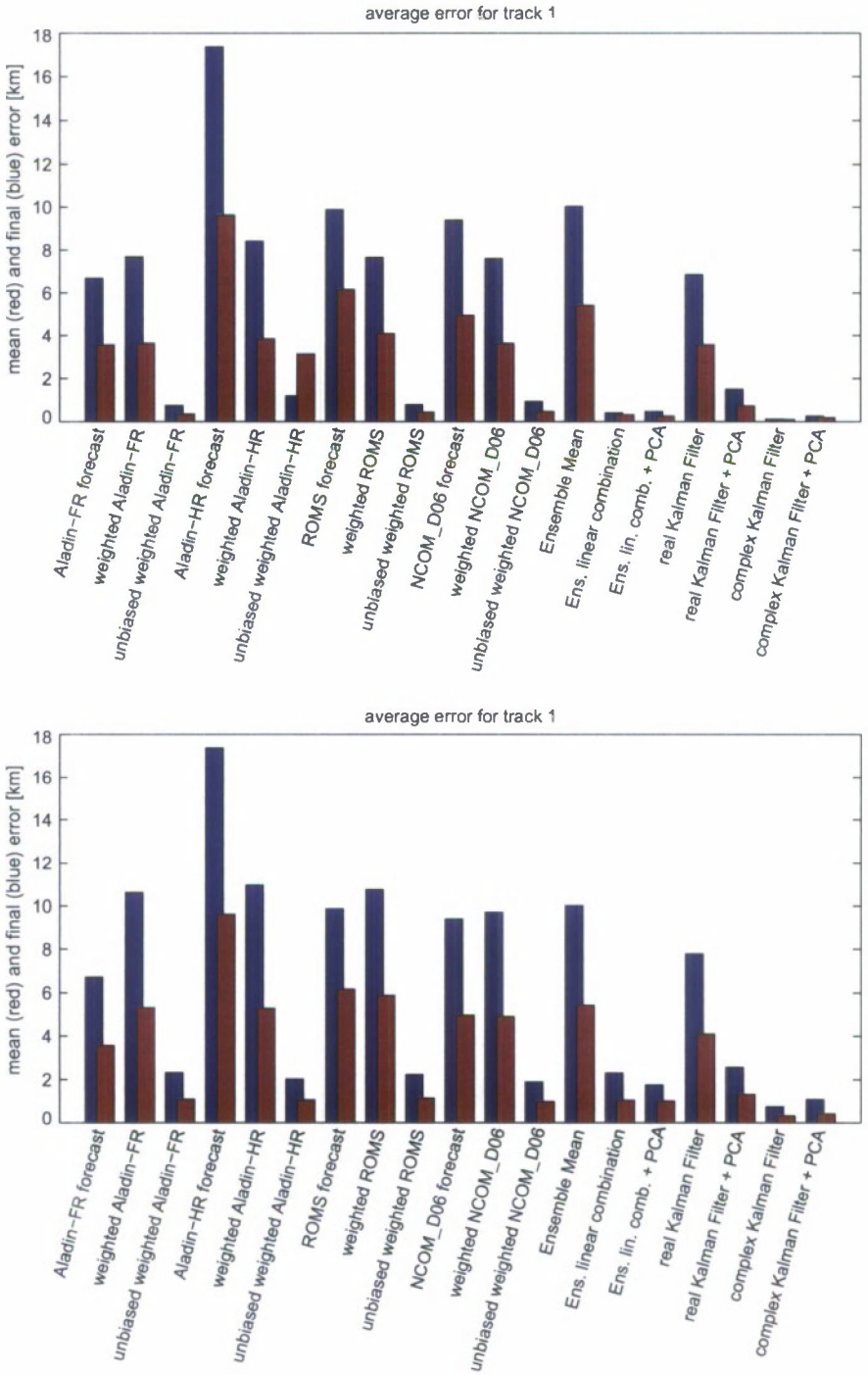

Fig. 3. DART06 experiment: average (over all segments) final (blue) and hourly average (red) error [km] for the drifter position after 12 h, using various HE methods, during the last 12 h of the training period (upper panel) and during the forecast (lower panel). The results are averaged over all 12-h segments of track a06956. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Thus, considering a bias model, at most seven (real or complex) weights are to be determined with the HE methods.

3. Hyper-ensemble methods

Super-ensembles and hyper-ensembles are techniques which aim at combining multiple models (of respectively the same and different physical processes) in order to provide a forecast with a higher skill. The optimal combination is obtained during a training period, and minimizes the distance to independent observations. Thus, SE techniques can be considered as data assimilation methods, as they aim to optimally combine different sources of information (in this case, multiple models, and observations). The main question for these techniques is whether the obtained combination will still be optimal in the forecasting mode, i.e. one needs to know a characteristic time during which the combination is stable, which means, a characteristic time during which none of the model's skill significantly changes. Krishnamurti et al. (1999) proposed to use an unbiased linear combination of the available models, optimal (in the least-squares sense) with respect to observations during a training period of *a priori* chosen length; all observations have equal importance. Rixen and Ferreira-Coelho (2007) applied the technique in the ocean, also adding non-linear combinations of the models (i.e. using neural networks and genetic algorithms), but found little improvement over the linear combination. This can be understood as the combination is determined over the same training period, either by linear or non-linear methods. Thus, not much is changed with respect to the combination being (staying) appropriate (or not) in forecasting mode. However, Shin and Krishnamurti (2003a); Rixen et al. (in press) introduced dynamically evolving weights in a linear combination of models, using data assimilation techniques (Kalman filter and particle filter) adapted to the super-ensemble paradigm. The latter techniques are able to train the weights on a time-scale corresponding to their natural characteristic time, discarding older information automatically. The weight's rate of change is determined by the respective (and evolving) uncertainties of the weights themselves, of individual models and of observations. Hence, these techniques were shown to yield significantly better results than more simple techniques. Of course, if one desires to obtain a forecast further away in the future than this characteristic time, no optimal combination can possibly be obtained, and without other *a priori* knowledge, one should probably just use a simple ensemble mean of the model forecasts.

In the current study, we try to forecast the motion of surface drifters. Their position can be elegantly represented using complex numbers, the longitude being the real part, and the latitude the imaginary part. The used HE methods are described hereunder in the context of our application.

3.1. Individual models

The simplest SE technique is called "best model"; it simply selects the model which performs best during the whole training period, and uses that one to obtain the forecast, discarding all other models. Although potentially useful information is neglected, this method is often used in practice.

A variant on this method is to multiply each model by a complex number determined during the training period. This corresponds to stretching and rotating the drift vector predicted by the model. When considering wind models, the multiplication thus allows to "optimize" the rule-of-thumb mentioned above (surface drift velocity of 3% of the wind velocity, 15° to the right).

A third variant also removes the bias by searching for an optimal combination of the considered model and a synthetic, constant model (i.e. bias); both models are also multiplied by complex factors.

3.2. Ensemble mean

The next method is the simple "ensemble mean". It does not use a training period or observations and thus, cannot really be considered as a SE technique; however, it is also a widely used method, since long known to provide better forecasts than individual models (Kalnay and Ham, 1989).

3.3. Least-squares linear combinations

Another technique consists of finding a linear combination of the models, minimizing (in the least-squares sense) its departure

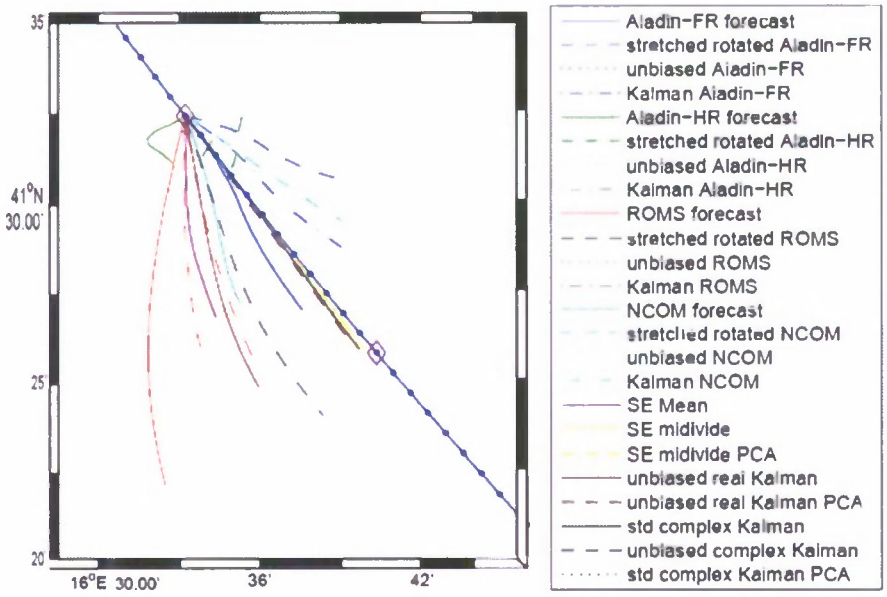

Fig. 4. Results of the forecast by all HE methods for a particular 12-h segment in the track a06956 showed in Fig. 1, with the training period starting 24 h after the drifter's deployment. The forecast starts at the brown diamond; the pink diamond represents the real drifter position at the end of the forecast. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from observations during the training period. Again, the weights are complex numbers, which corresponds to stretching and rotating each model in order for the final combination to be optimal. Two variants of this method are also used in our study. First, we add again a constant model, thus adding an unbiasing capability to our ensemble. Second, we remove some of the colinearities between the models. To this purpose, we perform principal component analysis (PCA) on the models, and decide to remove a certain percentage of variability, e.g. 10%. For example, when considering seven models, they would be transformed into seven principal components, of which the last 2 ones might be discarded. This has the further advantage of reducing the amount of weights that need to be determined (see below).

3.4. Non-linear combinations

Another class of SE methods use non-linear combinations of models, e.g. by feeding individual models as input to a neural

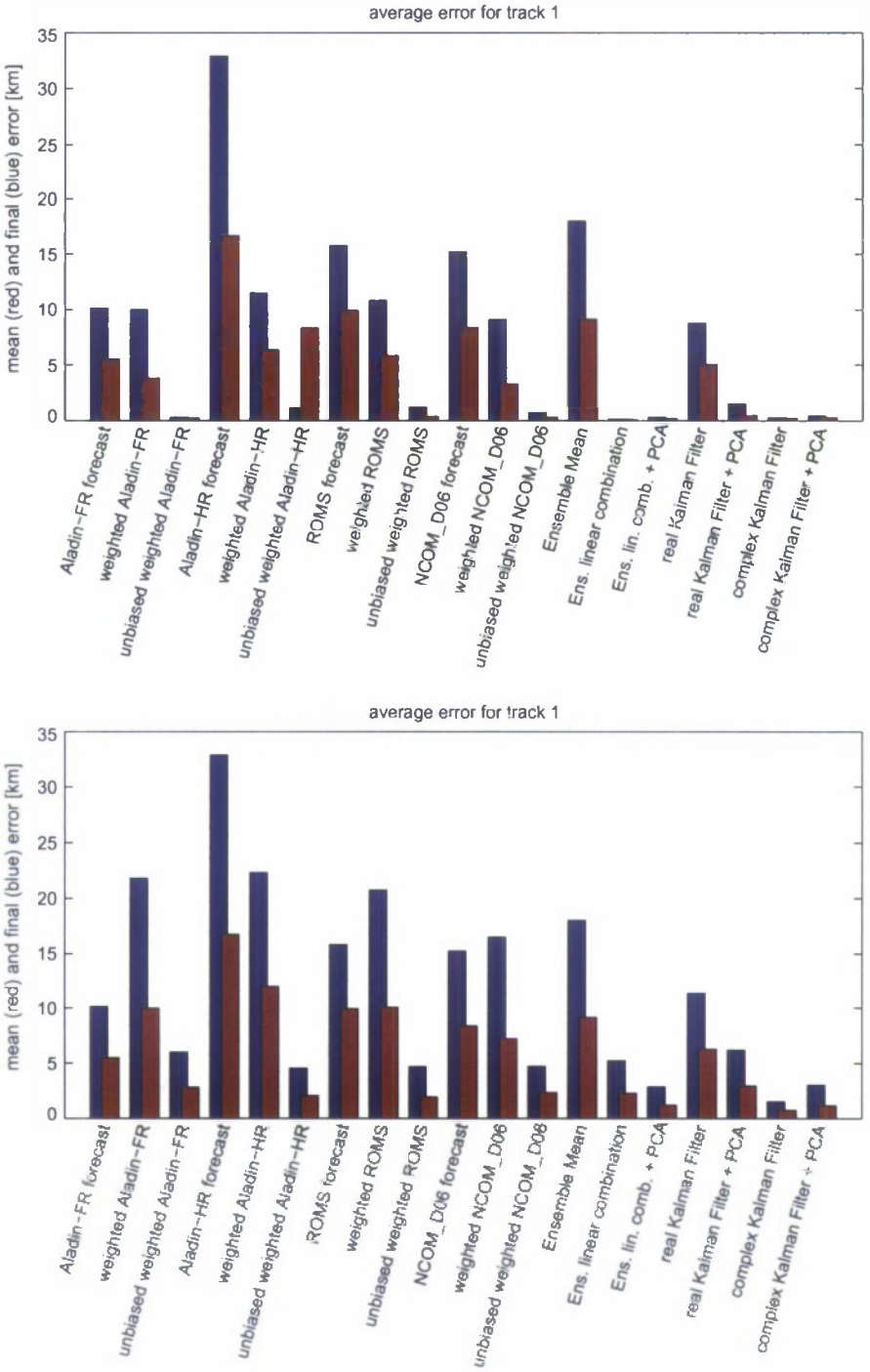

Fig. 5. DART06 experiment: average final (blue) and hourly average (red) error [km] for the drifter position after 24 h, using various HE methods, for the hindcast (upper panel) and forecast (lower panel) modes. The results are averaged over all 24-h segments of the track described before. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

network or genetic algorithm. However, as mentioned before, this does not change the fundamental fact that the combination is determined to be optimal during a defined training period, and one just hopes that it will still be adapted to the forecast period. Even though the non-linear combination *might* be better than the linear one, in practice, improved results in forecasting mode were not observed (Rixen and Ferreira-Coelho, 2007). This might be due to the fact that, compared to the linear combination (where one weight per model has to be determined), more parameters must be determined for those non-linear methods, even if one uses e.g. a neural network with a relatively simple architecture. Even with linear methods, the more models are included in the SE, the more weights need to be determined, and hence, smaller ensembles may lead to better results (for an illustration, see e.g. Maeng-Ki et al. (2004)). Thus, some improvements might appear with non-linear methods if one has a large amount of observations during the training period (and if no over-fitting problems appear). However, this is not the case in our study, and hence, we will not consider non-linear methods any further.

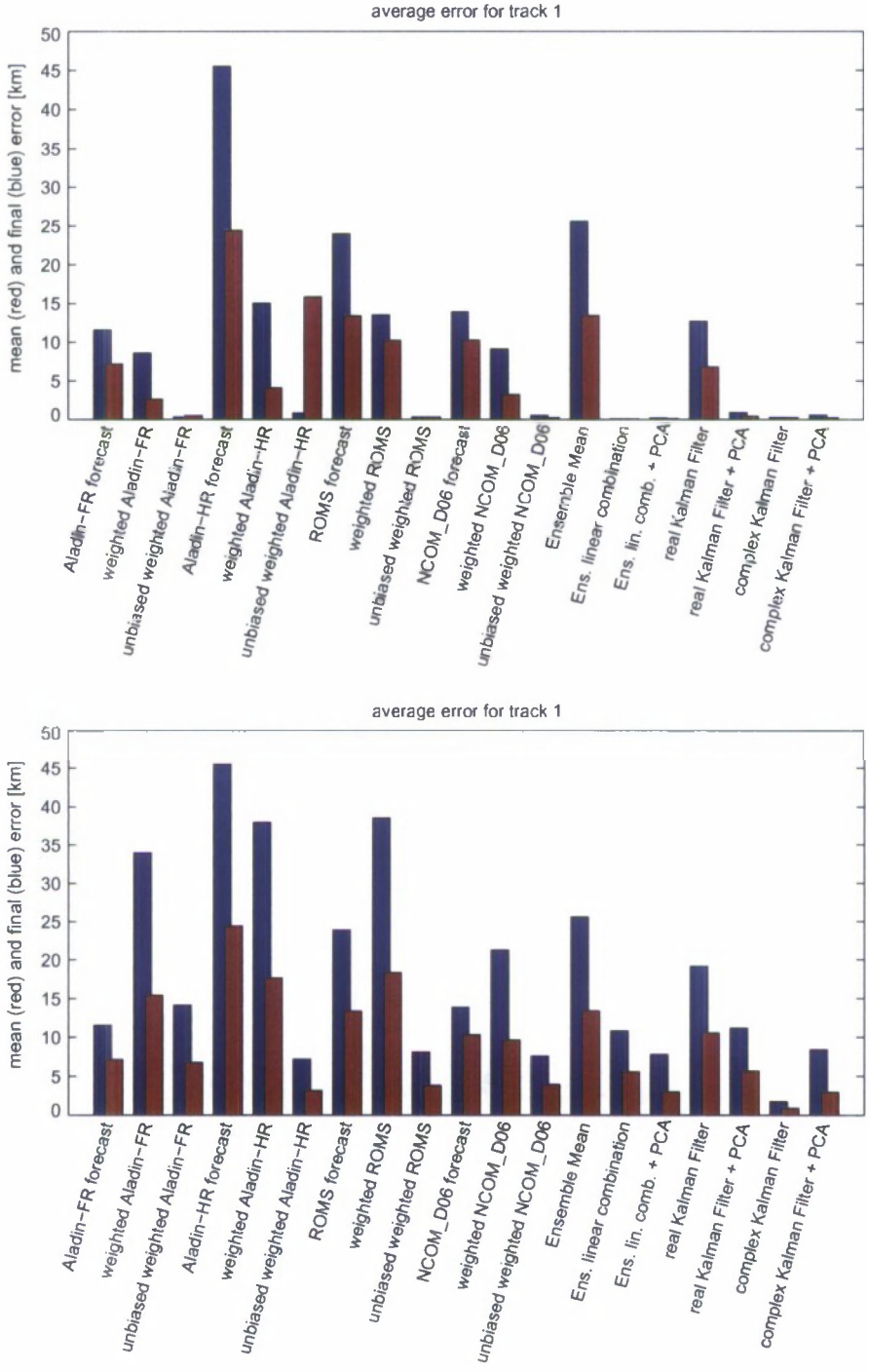

Fig. 6. DART06 experiment: average final (blue) and hourly average (red) error [km] for the drifter position after 36 h, using various HE methods, for the hindcast (upper panel) and forecast (lower panel) modes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

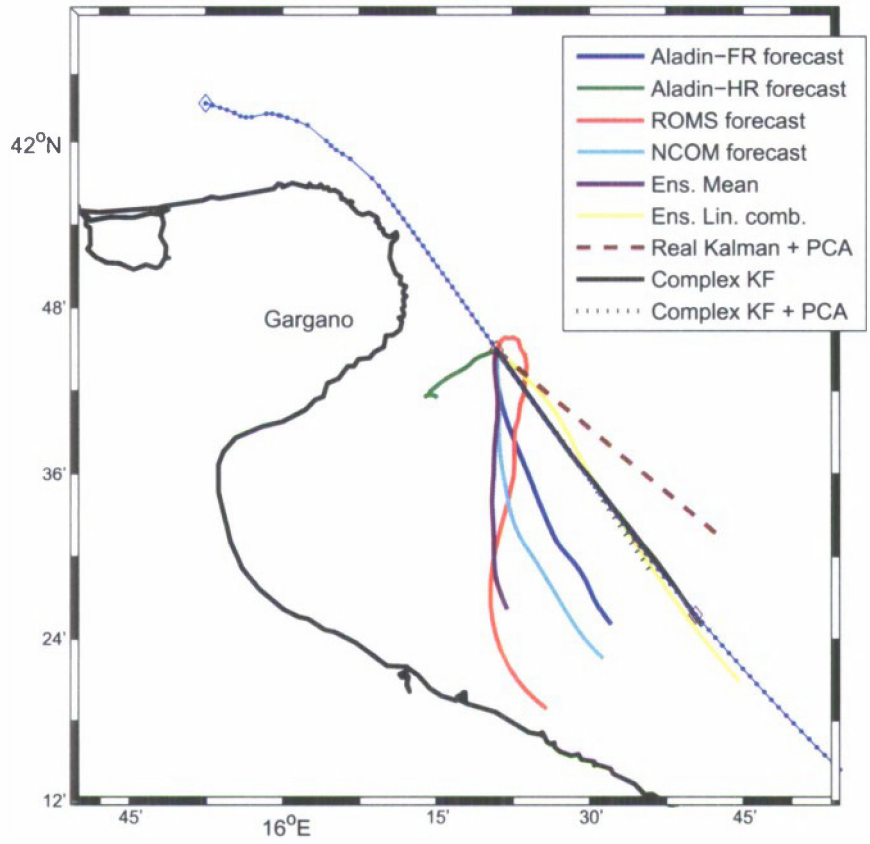

Fig. 7. Results of the forecast by selected HE methods for a particular 36-h segment in track a06956. Same color codes as Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

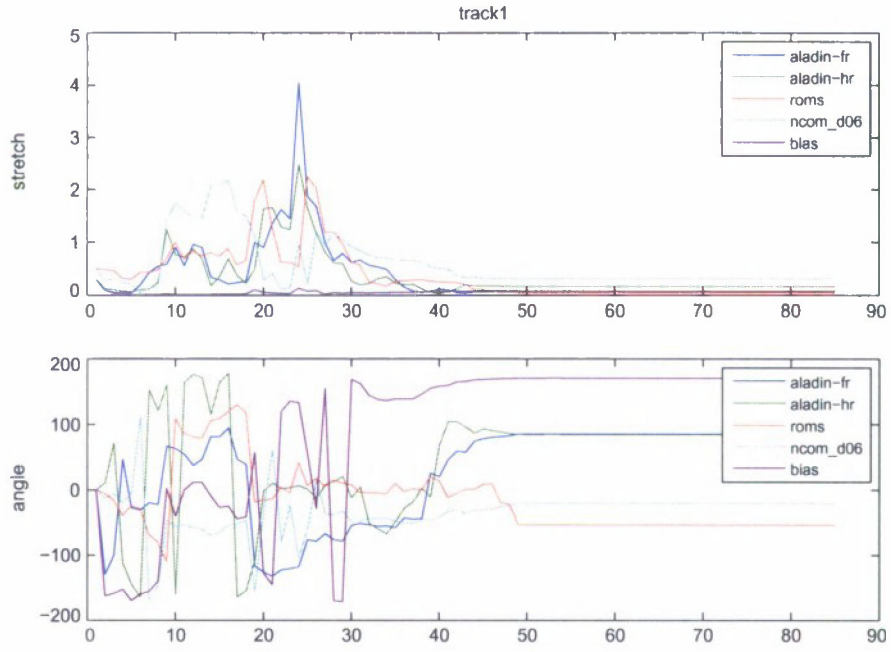

Fig. 8. Evolution of model weights with the ACEKF filter as a function of time (hours) from the start of the training period, corresponding to the first 3 days of track 1. The complex weights are represented by their magnitude and the angle they form with the eastward axis (positive clockwise).

3.5. Dynamical methods

In all previous methods (except the simple ensemble mean, obviously), the length of the training period had to be chosen *a priori* and all observations during the training period have an equal importance. More complicated methods can be thought of, e.g. where the observation's importance decreases exponentially with time. However, it would be more useful to have a method automatically adapting the weights to skill changes in models. This can be approximated with common data assimilation (DA) techniques: starting from our best guess, the weights are adapted during the training period, when observations are available, up to present time. Afterward, the weights are frozen and used during the forecasting period. All DA algorithms could be implemented; we will restrict ourselves to sequential DA and the Kalman filter (Kalman, 1960). As one might easily get confused by the unusual content of the different matrices in the Kalman filter equations, we briefly write them down and explain them below:

Forecast

$$\mathbf{x}^f(t_i) = \mathbf{M}_{t_i}\mathbf{x}^a(t_{i-1}) \quad (1)$$

$$\mathbf{P}^f(t_i) = \mathbf{M}_{t_i}\mathbf{P}^a(t_{i-1})\mathbf{M}_{t_i}^{\mathbf{T}} + \mathbf{Q} \quad (2)$$

Analysis

$$\mathbf{K} = \mathbf{P}^f(t_i)\mathbf{H}^{\mathbf{T}}[\mathbf{R} + \mathbf{H}\mathbf{P}^f(t_i)\mathbf{H}^{\mathbf{T}}]^{-1} \quad (3)$$

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}[\mathbf{y}^o - \mathbf{H}\mathbf{x}^f(t_i)] \quad (4)$$

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{K}\mathbf{H}\mathbf{P}^f(t_i) \quad (5)$$

$\mathbf{x}$ is the state vector, which contains the weights attributed to the models in the SE combination; its error covariance matrix is $\mathbf{P}$. Superscript f denotes its forecasted state after prediction steps; superscript a stands for analyzed state after the correction steps using observations. We have no *a priori* knowledge about the weight's evolution in time, and hence, the "model" matrix $\mathbf{M}$ is chosen as the identity matrix at all times; the state vector prediction step is trivial. Another choice would have been to include an exponential decrease of the model weights toward $\frac{1}{N}$, (N being the amount of models), or even more complicated relaxation schemes. In any case, as weights obviously do evolve in time, the chosen constant model $\mathbf{M}$ contains errors; they are represented by the random vector η, and have a covariance matrix $\mathbf{Q}$. Although not mathematically constrained, intuitively, one expects model's weights to sum approximately to 1, and to lie somewhere in or close to the [0–1] range. Hence, we estimated a reasonable standard deviation of the (model) error for individual weights to be 0.1; the non-diagonal elements of $\mathbf{Q}$ are put to zero. Furthermore, the errors affecting the state vector of weights have a covariance matrix denoted by $\mathbf{P}$; the initial standard deviation is chosen as 0.7 (as we expect a relatively bad initial guess of weights), and again, non-diagonal elements in $\mathbf{P_0}$ are put to zero (though they will become non-zero in time). The choices for the values of $\mathbf{Q}$ and $\mathbf{P}$ were validated by cross-correlation. Let's also note that the prediction step for $\mathbf{P}$ allows it to increase by $\mathbf{Q}$ at each timestep, in accordance with our intuition that the errors on weights increase with time.

Observations are represented by the vector $\mathbf{y}$; in our case they are observed surface drifts. The observation operator $\mathbf{H}$ linking the state vector space with the observation space, contains the individual model forecasts of surface drift (whereas usually, when one assimilates e.g. temperature in a primitive equation model, $\mathbf{H}$ is just an interpolation operator).

The observations' error covariance matrix is denoted $\mathbf{R}$, and contains three contributions: instrumental errors on the observations themselves (supposed small in our experiments), representativity errors due to the fact that the model does not represent all the physical processes included in the observations, and errors in the observation operator $\mathbf{H}$. Thus, $\mathbf{R}$ essentially contains the (unknown) errors affecting all the individual, physical models; these errors should be carefully estimated as $\mathbf{R}$ is a critical parameter in the filter's functioning. However, this is a very difficult task, requiring also more information than simply each model's forecast: the errors and shortcomings of individual models are precisely the reason why we use an HE method for! Hence, in the present study, $\mathbf{R}$ was again chosen by cross-correlation.

In oceanography, usually, the state vector contains hundreds of thousands of points, so that low-rank approximations of the Kal-

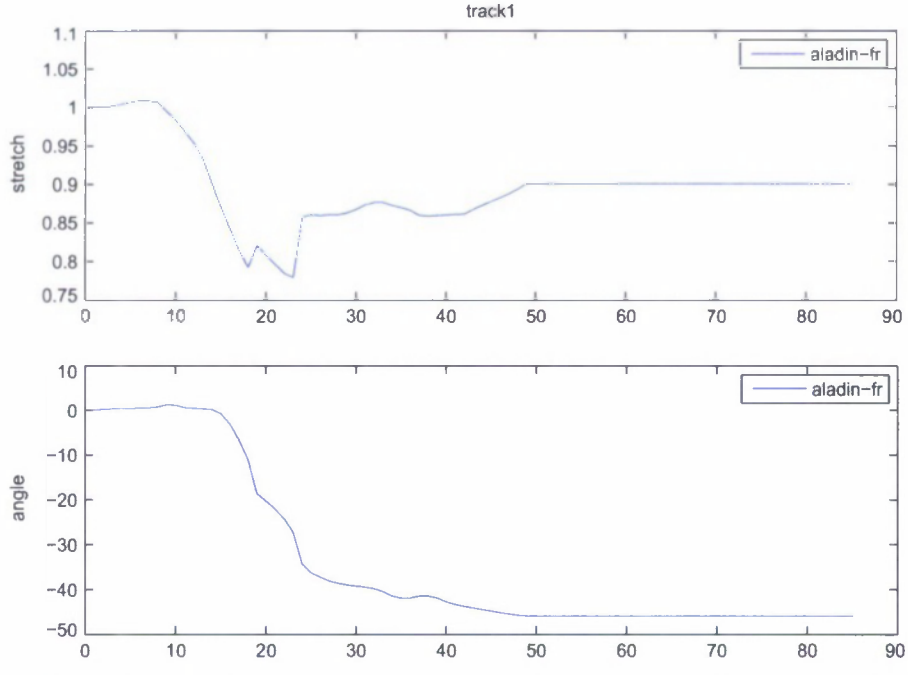

Fig. 9. Same as Fig. 8 but for the weights with the ACEKF filter in a singleton ensemble comprising only the Aladin (SHOM) wind model.

man filter must be implemented, such as the SEEK filter (Pham et al., 1998), the Ensemble Kalman filter (Evensen, 1994), etc. However, here, the state vector is very small, and hence the original, complete Kalman filter can be implemented. Thus, apart from the hypothesis of a linear model and a Gaussian weight distribution, no further assumptions have to be made. Finally, it should also be noted that at the end of the training period, the resulting weight vector, obtained with the Kalman filter, is strictly identical to the one that would have been obtained with the Kalman smoother (the same observations having been taken into account) or with the 4D-Var filter [see e.g. (Bennett, 1992)].

The equations written above are valid for real numbers, and hence we use them with real weights (i.e. the individual models are multiplied with a real number before being summed together). However, to use complex numbers as with the previous SE methods, the equations must be adapted into the so-called Augmented

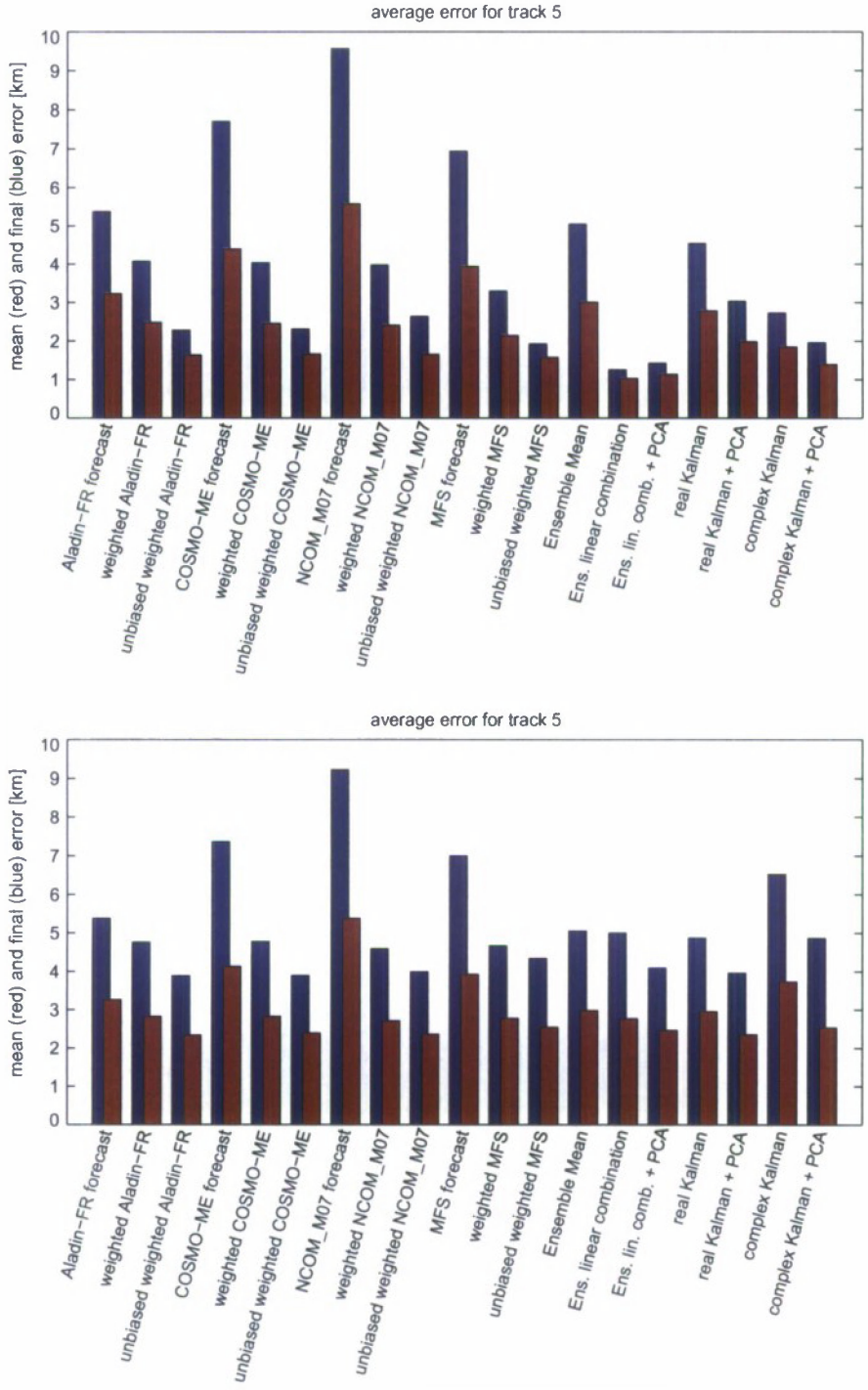

Fig. 10. MREA07 experiment: average (over all segments) final (blue) and hourly average (red) error [km] for the drifter position after 12 h, using various HE methods, during the hindcast (upper panel) and the forecast (lower panel). The results are averaged over all 12-h segments of the considered track. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Complex Extended Kalman filter (ACEKF) (Goh and Mandic, 2007), where all the initial vectors and matrices, as well as the model matrix, are "augmented" in the following way:

$$\mathbf{M}^{aug} = \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{M}^* \end{bmatrix} \tag{6}$$

with the asterisk denoting the complex conjugate. Vectors thus become matrices of double length, and width equal to 2; matrices have double length and width. For our study, all initial covariance matrices are chosen identically as above, but are then augmented. During the hindcast period, the state vector covariance matrix $\mathbf{P}^{aug}$ progressively becomes fully filled, with non-zero covariances between the real and imaginary parts.

Thus, using the ACEKF, we have a tool allowing to dynamically evolve complex weights during the hindcast period, and automatically take covariances between longitude and latitude increments into account. Finally, let's note that the previously mentioned "tricks" (unbiasing, reduction *via* PCA) can also be applied for the dynamical methods; our initial guess for the state vector is simply taken as the result of the corresponding least-squares linear combination method.

Other dynamical methods can be thought of. For example, if one supposes that the weights of the models in the combination do not have normal probability density functions, the Kalman filter should not be used. Particle filters (see e.g. Doucet et al. (2000), or van Leeuwen (2003) for an implementation in oceanography) alleviate this hypothesis of gaussianity. In our SE paradigm, one particle is one specific linear combination of models. The cost is that one has to use a relatively large ensemble of particles in order to ensure convergence. As in our experiment, the model **M** is the identity model, this is not necessarily a limitation; however, in the present study, the most time-consuming step is the spatial and temporal interpolation in relatively massive (physical) models output files. The results of a standard Sequential Importance Resampling (SIR) filter were similar to those of the Kalman filter (see Section 4), and about 1000 particles were required for convergence, leading to much longer computing times.

4. Results

For the two experiments, drifter observations and model forecast fields are interpolated in order to have one position every hour. Each hour, model velocity fields are also interpolated spatially to the exact drifter location. During the training as well as the forecast period, we use model casts with at least 24 h forecast lead time. In other words, we do not use a model hindcast for the training period, but use a forecast at least 24 h old. This ensures that models' skills are not artificially higher during the training

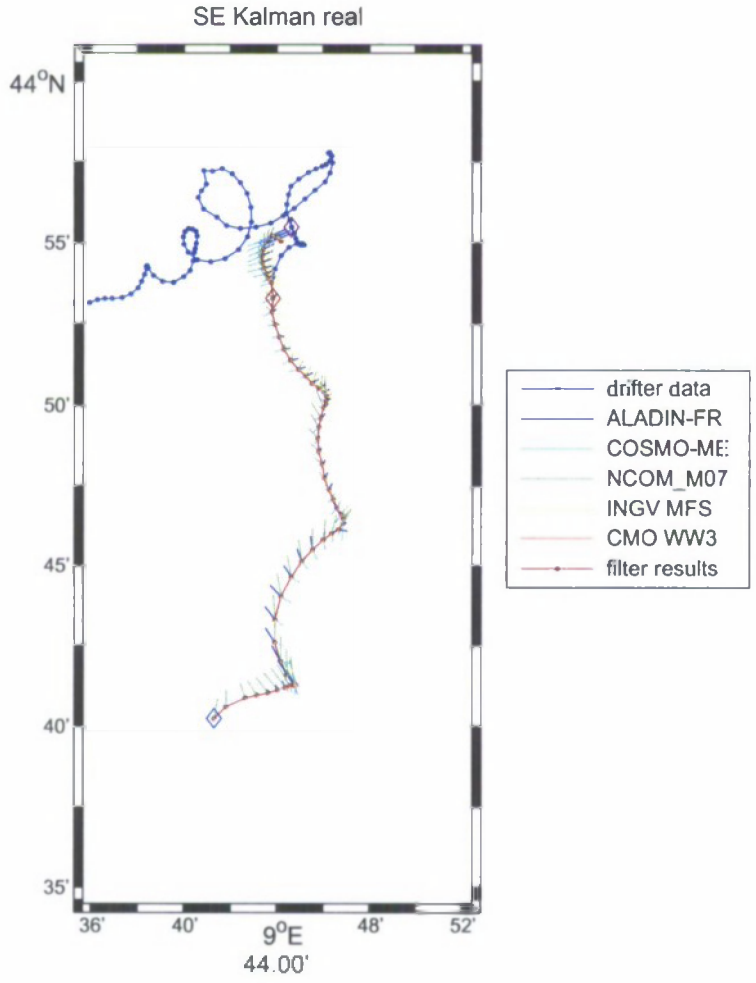

Fig. 11. Training and forecast using the Kalman filter. Training starts at the blue diamond; hourly displacements predicted by each individual model are represented by a colored segment. The actual forecast starts at the brown diamond, the pink diamond represents the real drifter position at the end of the forecast. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

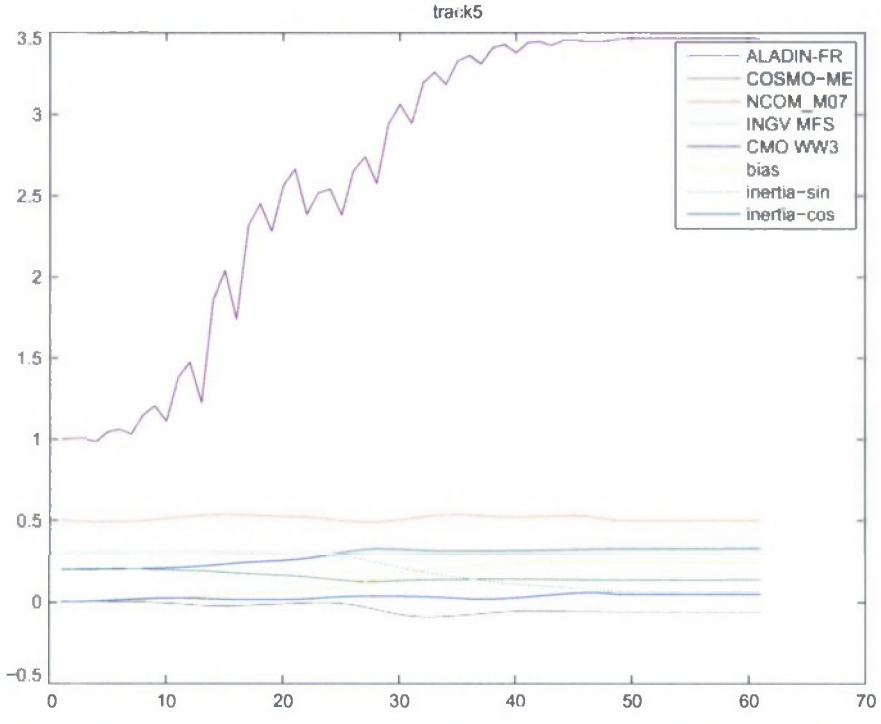

Fig. 12. Time-evolution of the absolute value and angle of the weights obtained with the Kalman filter method shown in Fig. 11.

due to the fact that data is assimilated during hindcasts. Our training period is chosen as 48 h (keeping in mind that dynamical methods can discard older information). Forecasts are obtained for three horizons: 12 h, 24 h and 36 h.

4.1. DART06 experiment

Fig. 3 shows the position error after 12 h of forecast (blue bars)[1] and the hourly mean error during these 12 h (red bars), for each of the HE methods, averaged over the first week (i.e. five daily forecasts) of the drifter track starting on 11 March 2006 (the first track in Fig. 1), when it flows along the Gargano peninsula. This track is the most rectilinear one of the experiment, but this does not necessarily make model predictions correspond more accurately with observations. Indeed, at the end of the first week, at least one model predicted that the drifter would hit the shore, which was not the case. The upper panel shows the results in a hindcast period (i.e. a non-independent pseudo-forecast obtained during the last 12 h of the training period, which means the weights should be particularly well adapted); the lower panel shows the results in the independent forecast. These results are typical for all the tracks in the WAC. After 12 h, all individual (wind or current) models have errors of 6.5–17 km, and of course perform equally well during hindcast and forecast (on average). In general, multiplying an individual model by a weight (obtained during the training) improves the hindcast slightly. The absolute value of the weights in question is generally comprised between 0.8 and 1.2; the angle is small for the ocean models and sometimes larger for the wind models.

Adding a bias model improves the results very significantly, with errors dropping to less than 1 km and 2 km in hindcast and forecast mode respectively. This can be understood as the trajectory is very linear, and hence the bias model takes a lot of the weight (i.e. we are using persistence); the considered model functions as a correction to the bias or persistence model. In summary, correcting any of the models for bias and multiplying it with a weight, yields much better forecasts than the common "best model", or, for that matter, "ensemble mean" strategies.

Combining all the models improves results only slightly compared to unbiased, weighted individual models; and adding the PCA "trick" does not improve the forecast skill a lot either in this case, albeit that the latter method yields the smallest forecast error of all static methods.

When real weights are evolved during the training period with a real-number Kalman filter, results are relatively bad (final error about 8 km). Indeed, real weights only allow stretching the drifter displacement vectors predicted by the model, but not rotating them. When one adds PCA, the first principal components are oriented toward the direction with largest variations, and hence the rotation induced by complex weights is less critical; results are better, comparable to those of the linear combination with complex weights. Finally, when one updates complex weights with the ACEKF, the best results are obtained, and the predicted drifter position is very close on the real position (error smaller than 1 km). In this case, adding PCA does not bring any improvement; the only benefit would be to remove redundant information, which appears unnecessary here.

As an example, Fig. 4 shows the results of the forecast by all HE methods for the third 12-h segment in the track discussed above. The real drifter trajectory is represented in blue, with hourly data represented by a dot. The training stops at the brown diamond; 12 h later, the drifter is at the pink diamond. All four individual models bring the drifter too much southward; but the unbiased, weighted, and particularly the dynamical methods can cope with this and correct the forecast.

Figs. 5 and 6 show the results for 24 h and 36 h forecasts respectively, for the same drifter. Results and comparisons between the different HE methods are qualitatively the same, although of

[1] For interpretation of color in Figs. 1–7, 10, 11, 13, 14 and 17, the reader is referred to the web version of this article.

course the forecast error gets larger as the forecast length increases. Even more than for a 12 h hindcast, the 36 h hindcast now almost coincides with the 48 h training period, and thus the linear combination is yielding very good results during this hindcast. An example of a 36 h forecast is shown in Fig. 7. It can be seen that none of the individual models are very successful, hence the ensemble mean is not accurate either. Corrected individual models, not shown in the figure for clarity, are closer to the real drifter than the respective uncorrected models. However, the ensemble linear combination is even closer, particularly when adding PCA. The real Kalman Filter is unable to rotate models, hence the results are not perfect, as explained higher. Finally, one can see that among all HE methods, the ACEKF filters (with or without PCA) forecast the drifter position most accurately.

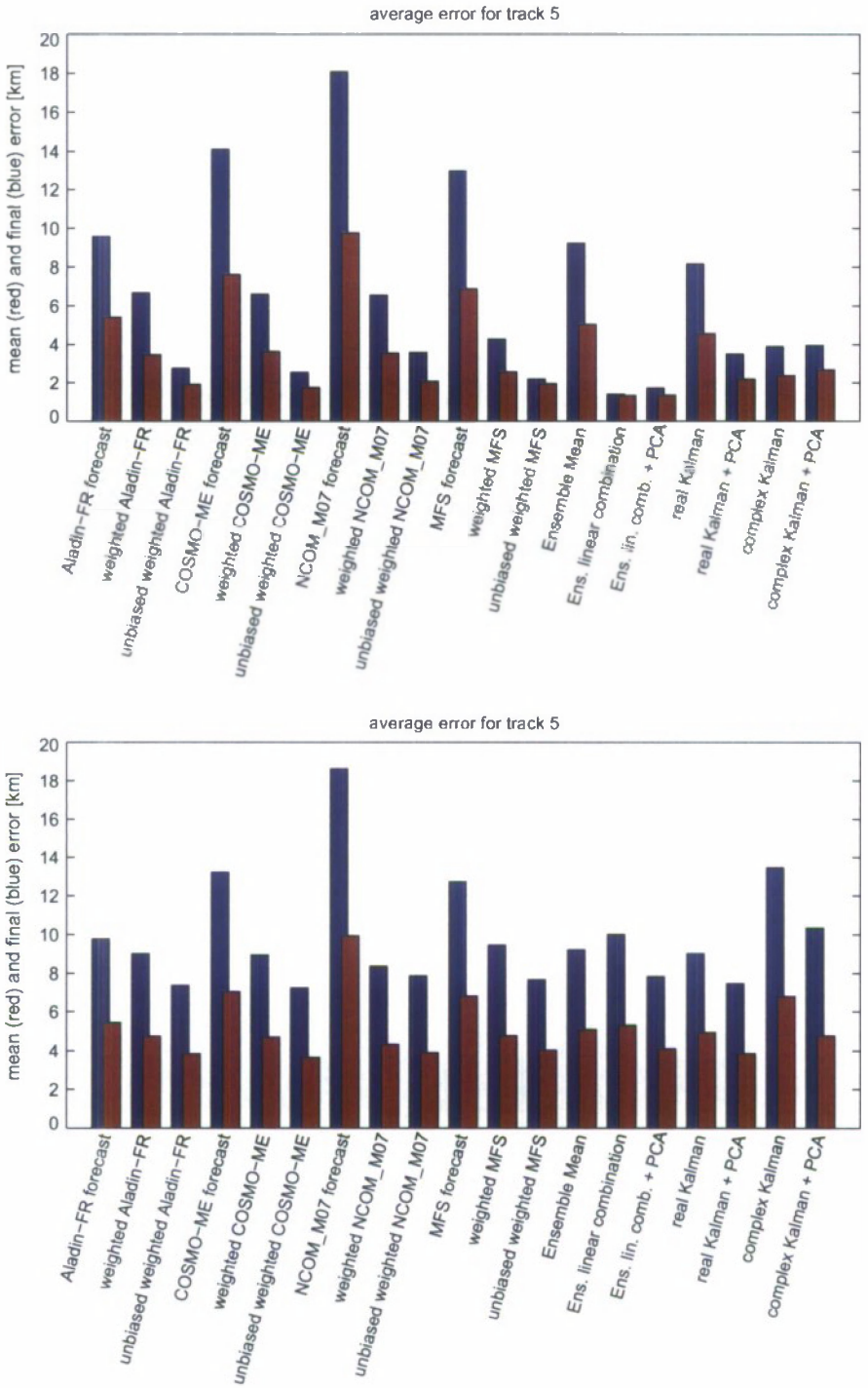

Fig. 13. MREA07 experiment: average final (blue) and hourly-average (red) error [km] for the drifter position after 24 h, for both the hindcast (upper panel) and the forecast (lower panel). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, to illustrate the concept of the characteristic time during which a HE combination remains valid, Fig. 8 shows the evolution of the complex weights during the first 3 days of the considered track. It can be seen that the weights undergo rapid changes starting at hour 8; at least one model probably undergoes a strong change in skill at that time. This is verified using the com-

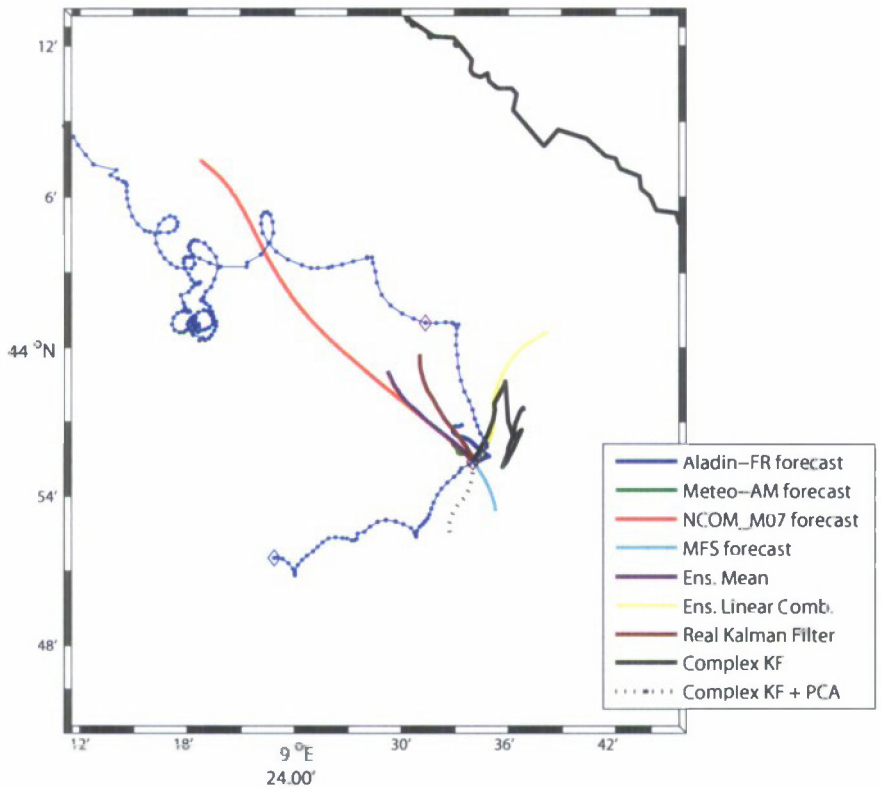

Fig. 14. Results of the forecast by selected HE methods for a particular 24-h segment in the track a74875 (see the largest red box in Fig. 2). The training starts at the blue diamond, the forecast at the brown diamond; the pink diamond represents the real drifter position at the end of the forecast. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

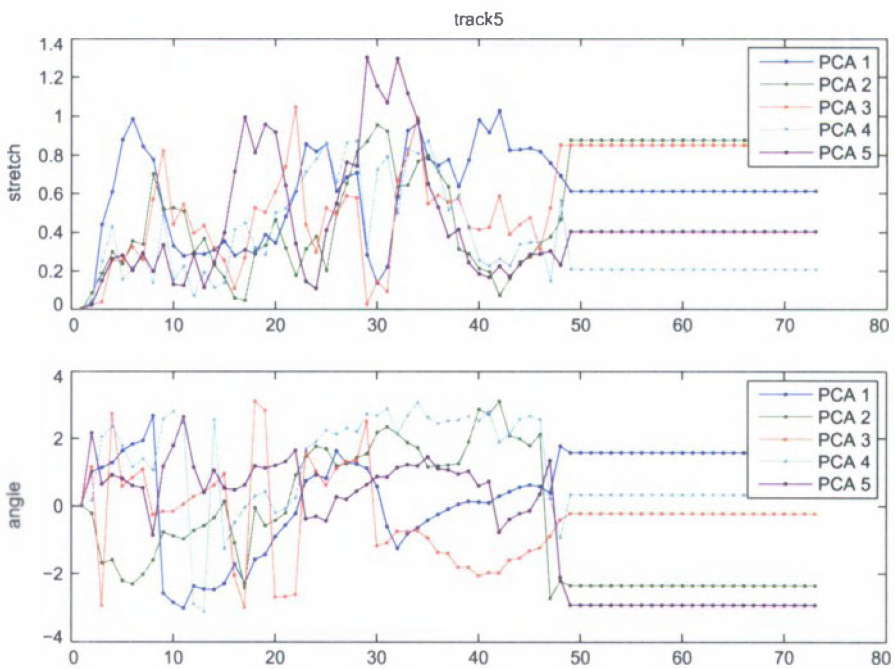

Fig. 15. Time-evolution of the absolute value and angle of the weights obtained with the ACEKF method with PCA (result showed in Fig. 14 in dashed black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plex Kalman filter but just on single models. For example, the obtained weight evolution of the Aladin (SHOM) wind model is shown in Fig. 9; it can indeed be seen that from hour 8, the drift predicted by that model has to be strongly attenuated (by about 20%), the adjustment taking about 10 h.

From Fig. 8, it can be seen that similar rapid changes occur around hours 20 and 25; but elsewhere, and particularly after hour 25, the weights are modified only slowly. Thus, as only small changes happen after hours 25 (except the continuing adjustment), and onward to hour 48 these changes become even smaller, one can suppose that the models' skills are relatively constant during these 23 h. This gives us some confidence to use HE methods for the forecast, rather than the ensemble mean. In particular, for the track considered in Figs. 8 and 9, the characteristic time of HE validity is at least 24 h. This should be related to the Lagrangian autocorrelation time, which is about half a day to 1 day (Poulain and Zambianchi, 2007; Rubio et al., in press).

The absolute value of the final weights obtained at hour 48 (the end of the training period) are about 0.4 for NCOM_D06, and less for the three other models, although no model gets a negligible weight. Furthermore, the bias model obtains about 0.1, i.e. the same weight as the ROMS and ALADIN (SHOM) models. The ocean models undergo relatively small rotations, whereas the atmospheric wind models are turned by about 90°.

4.2. MREA07 experiment

The results in the Ligurian basin are less straightforward, as could already be expected from Fig. 2, particularly because most of the trajectories closely follow the coastline; hence, an error in one of the individual models could lead the simulated trajectory into land.

Fig. 10 shows the error bars for "track 5" (shown in Fig. 2), concerning the 12 h forecast. Conclusions are, again, similar to those obtained in the DART06 experiment. In particular, the best results are now obtained with the real-number Kalman filter with the PCA trick. All HE methods yield better results than the simple ensemble mean, except the ACEKF (without PCA). In general, it can be seen that PCA reduces the forecast errors. As shown later, this is also the case of the 24-h and 36-h forecasts. Hence, one might suspect that some models present colinearities (which need to be removed) or that there are simply too many weights (seven complex numbers) to be determined. For the 12-h forecast, when comparing the real and complex Kalman filters respectively, the advantage of having less degrees of freedom to determine outbalances the fact that drift vectors can only be stretched, and not rotated.

An example of result obtained with the Kalman Filter method is detailed in Fig. 11; the time-evolution of the weights is shown in Fig. 12. One can see from Fig. 11 that none of the individual models

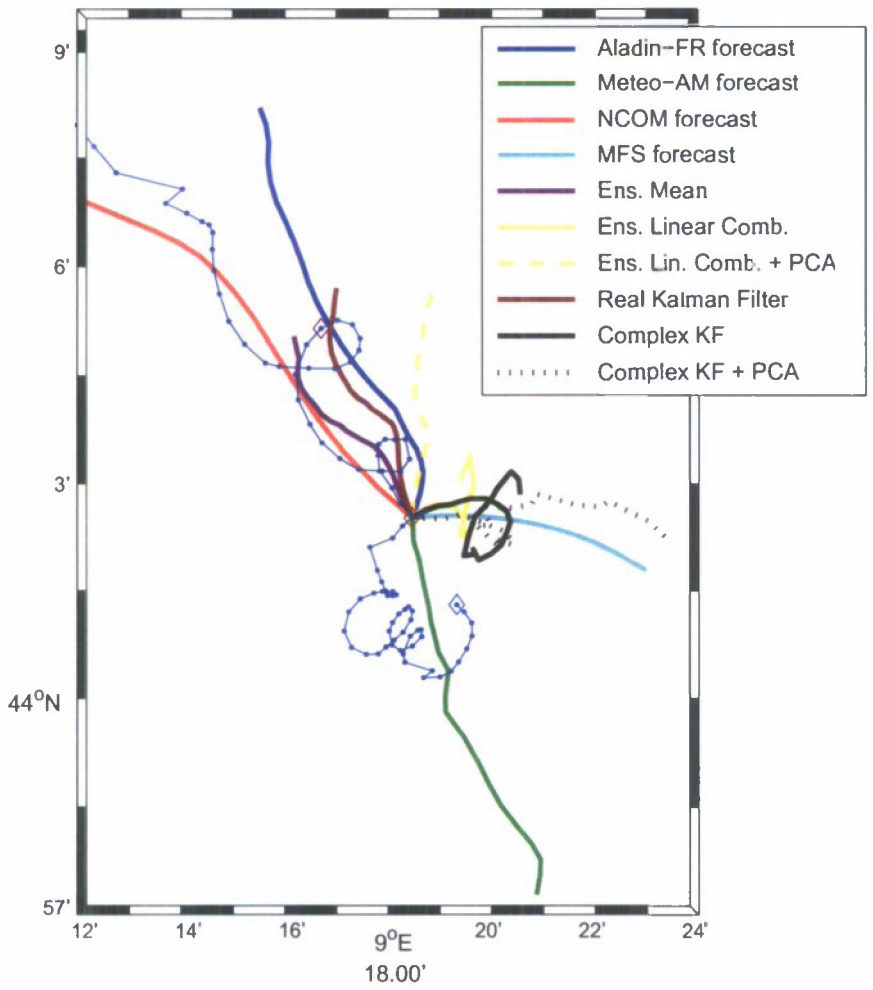

Fig. 16. Results of the forecast by selected HE methods for a particular 24-h segment in the track a74875 (see the smaller red box in Fig. 2). Same color codes as Fig. 14. The NCOM_M07 forecast extends to 44°15'N, 9°03'W but is cut off for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is quite accurate; most predicted displacements are too small (except for NCOM_M07, which has correct amplitudes but is badly orientated most of the time, moreover with changing error direction). However, the weights adapt permanently to the latest information; one can see that for this particular segment, the SHOM (Aladin-France) model obtains a larger weight; furthermore, the bias also becomes more and more important. The circular models keep low weights at all times, but as the weight of the COSMO-ME and even more of the INGV MFS model are decreasing over time, the latter ultimately obtains a weight similar to the synthetic inertial oscillations model. Finally, we notice the very large factor affecting the wave model; one should remember that the displacement itself forecasted by this model is much smaller.

The results for a 24 h forecast are shown in Fig. 13. During the hindcast, the unbiased, weighted individual models, the unbiased linear combination and the ACEKF combination all perform relatively well (and better than the ensemble mean). However, in forecast mode, the ensemble mean method leads to a smaller error than

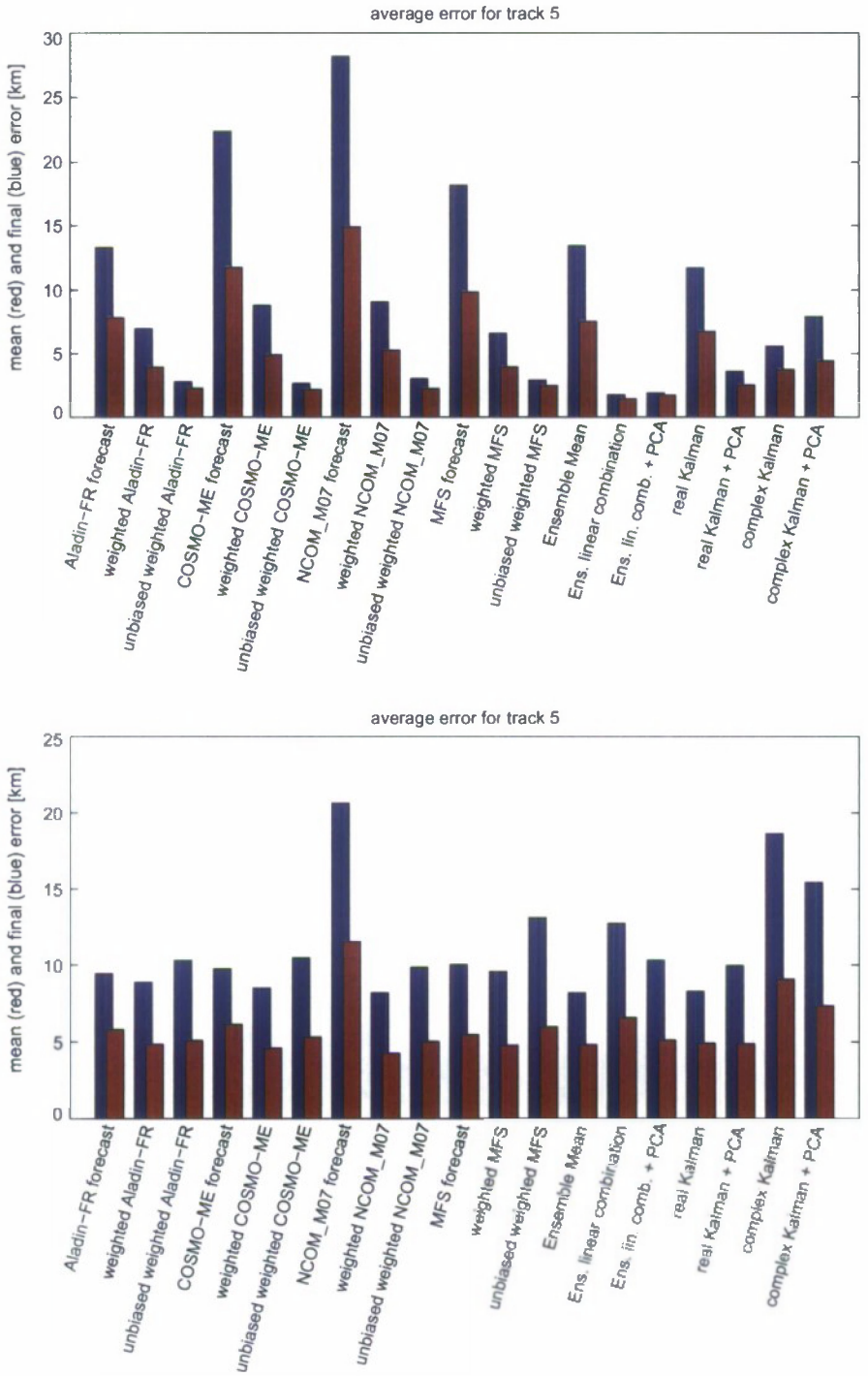

Fig. 17. MREA07 experiment: average final (blue) and hourly-average (red) error [km] for the drifter position after 36 h, for both the hindcast (upper panel) and the forecast (lower panel). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the linear combination! With PCA, the linear combination is somewhat better; the Kalman filter with real weights also performs reasonably well. All this indicates that the characteristic time during which the obtained combinations are valid, has approximately been reached. The ACEFK combination, where more degrees of freedom are present, yields a much larger error than the real-number Kalman; again, PCA allows to somewhat improve its performance. Fig. 14 illustrates this for a particular segment starting 20 days after the drifter launch (some HE methods are not shown for the clarity of the figure). In this example, the ensemble mean and ensemble linear combination are outperformed by the real Kalman filter; however, the weights in the complex Kalman filter do lead to an erratic forecast. With PCA, the obtained trajectory is less erratic, but still completely incorrect. The instability of complex weights, even with PCA, is further illustrated in Fig. 15 showing their time-evolution; all components weights undergo large variations, with each component sometimes being important, sometimes negligible. One more example is given in Fig. 16, starting 26 days after the drifter launch; similar conclusions again apply.

The situations gets even worse when trying to predict the drift at 36 h. Results are shown in Fig. 17. In forecasting mode, the ensemble mean methods now yields the smallest errors; all other methods have errors of the same order, or larger, as individual models. This clearly indicates that the obtained combinations are not valid anymore after (less than) 36 h; results may be somewhat better or somewhat worse, depending purely on luck. For some other tracks (not shown), the results are somewhat better, and some HE methods still perform relatively well, leading to results similar or slightly better than the ensemble mean. However, one might conclude that, using the mentioned models, the surface drift predictability limit in the Ligurian Sea during the MREA07 experiment was somewhere between 24 and 36 h.

5. Conclusion

In the present study, we examined how hyper-ensemble (HE) methods can improve the forecast of surface drift over forecasts obtained with a single model, or with the mean of different models. We used linear combinations of atmospheric and ocean models, as well as a wave model and synthetic models (circular or constant, corresponding to inertial oscillations or to bias). We first examined the most common "combinations", such as the ensemble mean or the "best past model". Another method is to determine the value of the weights during a training period, by least-squares minimization of the distance to observed surface drift. We also implemented the Kalman filter, a data assimilation method allowing to dynamically change the value of weights when new observed drifts become available. The latter method also allows to estimate a characteristic time during which the model's skills are approximately constant, and hence help us to decide whether or not a HE method should be used or not.

Surface drift can be represented by complex numbers; furthermore, if one also uses complex weights in the linear combination, this allows to stretch and to rotate the predicted drift vectors. The Kalman filter has to be adapted for using complex numbers, leading to the so-called ACEKF filter; covariances between real and imaginary parts are automatically generated.

Whenever the forecast period was short enough, the HE lead to strongly improved results, with the final position error reduced by at least a factor 3 compared to individual models. It was also shown that dynamical methods, such as the ACEKF, yield the smallest forecast error; as mentioned before, the time-evolution of the weights also provides insight into the HE and models performance. When many models are available (seven in our MREA07 experiment), it is useful to reduce the amount of weights to determine, e.g. by applying a principal component analysis and removing colinearities between models.

We showed the benefit of adding one or more synthetic models (a constant model adds unbiasing to the ensemble; a circular model can add or correct inertial oscillations). However, more models imply more degrees of freedom to determine during the training period, and this may render the ensemble unstable.

In general, forecasting the drift up to 12 h is always possible (in both domains), and HE methods significantly improve results over individual models. In particular, adding a synthetic bias model to a single weighted model strongly decreases errors, indicating that at this forecasting timescale, persistence is very useful. Adding more models and combining them with dynamical methods such as the Kalman filter allows to further improve results. However, after 24, and especially 36 h, forecasting might become more problematic, at least in a complex environments with strong meso- and smaller scale eddy activity, such as the Ligurian Sea. Indeed, we showed that model skills may change significantly over such a time period, and hence the weighted combination of models obtained during the training period is not optimal during the whole forecasting period. In particular, adding a bias model to a single model does not increase skill, indicating that persistence is not useful anymore, and that the role of primitive equation models become truly crucial for scales longer than 12 h: the issue of good model performance cannot be avoided by super-ensembles! In the Ligurian Sea, HE methods performed poorly for 24 or 26 h forecasts, and a simple ensemble mean or an unbiased linear combination lead to better results than a Kalman filter method. Hence, it might be better to use "average" weights obtained during a longer training period rather than adapting to the most recent data. However, the Kalman filter methods at least allow to know how fast weights change in time, so that one can decide which HE technique to use. Thus, when observations and different models are available, we recommend to implement a dynamical HE method such as the Kalman filter, and to examine the time-evolution of the weights. If they are stable during a lapse of time corresponding to the desired forecast horizon, then the results of the dynamical method should be used. If they are stable during a much longer period, methods with *a priori* fixed training lengths will yield approximately the same results. To the contrary, if the weights are varying very quickly (compared to the desired lead time of the forecast), one should use average weights.

In the case of a 2-dimensional variable such as surface velocity (or drift), the question whether one should use real or complex weights depends on the size of the ensemble (i.e. the degrees of freedom to be fixed). Generally, complex weights provide better results as model-predicted drifts can be both multiplied and rotated. However, twice as many parameters are to be fixed during the training period. If many different models are present, or insufficient training data is available, one could then obtain better results with real weights. PCA generally helps to decrease the amount of degrees of freedom, and might thus allow using complex weights where otherwise, real weights would have led to the best results.

In order to use hyper-ensembles operationally, one needs to centralize all the forecasts, as well as the observations. The HE computations themselves are performed very fast; a large part of the effort goes to correctly reading and interpolating the forecasts from the individual models into the HE algorithm. Provided that these issues are resolved, an operational HE forecast can be provided, as has already been demonstrated (Rixen et al., 2008). As more models are implemented in various regions, we hope the HE techniques will improve forecasts at reduced cost.

Acknowledgement

The authors wish to acknowledge the participants of the DART06 and MREA07 experiments involved with the drifter

deployment/recovery logistics and the drifter data processing. Thanks to all the people who kindly provided numerical model products. Pierre-Marie Poulain was partially supported by the Office of Naval Research under grant N000140310291. The work of J.W. Book and P.J. Martin was supported by the Office of Naval Research as part of the research program "Dynamics of the Adriatic in Real-Time" under Program Element Number 0602435N. Meteo France is acknowledged for providing results to us.

References

Agarwal, Y., Terray, E., Donelan, M., Hwang, P., Williams, A.I., Drennan, W., Kahma, K., Kitagorodskii, S.A., 1992. Enhanced dissipation of kinetic energy beneath surface waves. Nature 359, 219–220.

Barbanti, R., Gerin, R., Poulain, P.-M., <http://doga.ogs.trieste.it/doga/sire/dart/database_dart/>, 2007. Dart Drifter Database: 11 March 2006 to 1 January 2007.

Bennett, A.F., 1992. Inverse Methods in Physical Oceanography. Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge University Press.

Carniel, S., Umgiesser, G., Sclavo, M., Kantha, L., Monti, S., 2002. Tracking the drift of a human body in the coastal ocean using numerical prediction models of the oceanic atmospheric and wave conditions. Science and Justice 42, 143–151.

Chiggiato, J., Oddo, P., 2008. Operational ocean models in the Adriatic Sea: a skill assessment. Ocean Science 4 (1), 61–71. <http://www.ocean-sci.net/4/61/2008/>.

Coelho, E., Peggion, G., Rowley, C., Jacobs, G., Allard, R., Rodriguez, E., in press. A note on NCOM temperature forecast error calibration using the ensemble transform. Journal of Marine Systems. Available online 28 February 2009.

Davis, R.E., 1985. Drifter observation of coastal currents during CODE. The method and descriptive view. Journal of Geophysical Research 90, 4655–4741.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering statistics and computing. Statistics and Computing, 197–208.

Ekman, V., 1905. On the influence of the earth's rotation on ocean currents. Arkiv for Mathematik, Astronomiocho Fysik. 2 (11).

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99 (C5), 10143–10162.

Everitt, B., 2002. Cambridge Dictionary of Statistics. Cambridge University Press.

Goh, S., Mandic, D., 2007. An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks. Neural Computation 19 (4), 1039–1055.

Griffa, A., Piterbarg, L.I., Ozgokmen, T., 2004. Predictability of lagrangian particle trajectories: effects of smoothing of the underlying eulerian flow. Journal of Marine Research 62, 1–35.

Haza, A.C., Griffa, A., Martin, P., Molcard, A., Ozgokmen, T.M., Poje, A.C., Barbanti, R., Book, J.W., Poulain, P.-M., Rixen, M., Zanasca, P., 2007. Model-based directed drifter launches in the Adriatic sea: results from the DART experiment. Geophysical Research Letters 34, L10605. doi:10.1029/2007GL029634.

Ivatek-Sahdan, S., Tudor, M., 2004. Use of high-resolution dynamical adaptation in operational suite and research impact studies. Meteorologische Zeitschrift 13, 1–10.

Kalman, R., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82 (D), 35–45.

Kalnay, E., Ham, M., 1989. Forecasting forecast skill in the Southern Hemisphere. In: Preprints of the 3rd International Conference on Southern Hemisphere Meteorology and Oceanography, Buenos Aires, 13–17 November.

Krishnamurti, T., Kishtawal, C., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C., Gadgil, S., Surendran, S., 1999. Improved weather and seasonal climate forecasts from Multimodel Superensemble. Science 285, 1548–1550.

Lermusiaux, P., Chiu, C.-S., Gawarkiewicz, G., Abbot, P., Robinson, A., Miller, R., Haley, P., Leslie, W., Majumdar, S., Pang, A., Lekien, F., 2006. Quantifying uncertainties in Ocean Predictions. Oceanography 19 (1), 92–105 (Special issue on "Advances in Computational Oceanography").

Madsen, O., 1977. A realistic model of the wind-induced Ekman boundary layer. Journal of Physical Oceanography 7, 248–255.

Maeng-Ki, K., In-Sik, K., Chung-Kyu, P., Kyu-Myong, K., 2004. Superensemble prediction of regional precipitation over Korea. International Journal of Climatology 24, 777–790.

Martin, P., Book, J., Burrage, D., Rowley, C., Tudor, M., 2009. Comparison of model-simulated and observed currents in the Central Adriatic during DART. Journal of Geophysical Research 114, C01S05. doi:10.1029/2008JC004842.

Mutemi, J., Ogallo, L., Krishnamurti, T., Mitra, A., Vijaya Kumar, T., 2007. Multimodel based superensemble forecasts for short and medium range NWP over various regions of Africa. Meteorology and Atmospheric Physics 95, 87–113.

Paldor, N., Dvorkin, Y., Mariano, A., Özgökmen, T., Ryan, E., 2004. A practical, hybrid model for predicting the trajectories of near-surface ocean drifters. Journal of Atmospheric and Oceanic Technology 21, 1246–1258.

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. October.

Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Le Traon, P.-Y., Maillard, C., Manzella, G., Tziavos, C., 2003. The Mediterranean ocean forecasting system: first phase of implementation (1998–2001). Annales Geophysicae 21, 1–18.

Poulain, P.-M., 1999. Drifter observations of surface circulation in the Adriatic Sea between December 1994 and March 1996. Journal of Marine Systems 20, 231–253.

Poulain, P.-M., Gerin, R., Mauri, E., Pennel, R., 2009. Wind effects on drogued and undrogued drifters in the Eastern Mediterranean. Journal of Atmospheric and Oceanic Technology, doi:10.1175/2008JTECHO618.1.

Poulain, P.-M., Zambianchi, E., 2007. Surface circulation in the central mediterranean sea as deduced from lagrangian drifters in the 1990s. Continental Shelf Research 27, 981–1001.

Rascle, N., Ardhuin, F., 2009. Drift and mixing under the ocean surface revisited: stratified conditions and model-data comparisons. Journal of Geophysical Research 114, C01S05, doi:10.1029/2007JC004466.

Rascle, N., Ardhuin, F., Terray, E., 2006. Drift and mixing under the ocean surface: a coherent one-dimensional description with application to unstratified conditions. Journal of Geophysical Research, 111.

Rixen, M., Book, J., Carta, A., Grandi, V., Gualdesi, L., Stoner, R., Ranelli, P., Cavanna, A., Zanasca, P., Baldasserini, G., Trangeled, A., Lewis, C., Trees, C., Grasso, R., Giannechini, S., Fabiani, A., Merani, D., Berni, A., Leonard, M., Martin, P., Rowley, C., Hulbert, M., Quaid, A., Goode, W., Preller, R., Pinardi, N., Oddo, P., Guarnieri, A., Chiggiato, J., Carniel, S., Russo, A., Tudor, M., Lenartz, F., Vandenbulcke, L., in press. Improved ocean prediction skill and reduced uncertainty in the coastal region from multi-model super-ensembles. Journal of Marine Systems. Available online 28 February 2009.

Rixen, M., Ferreira-Coelho, E., 2007. Operational surface drift prediction using linear and non-linear hyper-ensemble statistics on atmospheric and ocean models. Journal of Marine Systems 65, 105–121.

Rixen, M., Ferreira-Coelho, E., Signell, R., 2008. Surface drift prediction in the Adriatic Sea using hyper-ensemble statistics on atmospheric, ocean and wave models: uncertainties and probability distribution areas. Journal of Marine Systems 69, 86–98.

Rubio, A., Taillandier, V., Garreau, P., in press. Reconstruction of the mediterranean northern current variability and associated cross-shelf transport in the gulf of lions from satellite-tracked drifters and model outputs. Journal of Marine Systems. Available online 28 February 2009.

Shin, D., Krishnamurti, T., 2003a. Short- to medium-range superensemble precipitation forecasts using satellite products: 1. Deterministic forecasting. Journal of Geophysical Research, 108.

Shin, D., Krishnamurti, T., 2003b. Short- to medium-range superensemble precipitation forecasts using satellite products: 2. Probabilistic forecasting. Journal of Geophysical Research, 108.

Tonani, M., Pinardi, N., Dobricic, S., Pujol, I., Fratianni, C., 2008. A high-resolution free-surface model of the Mediterranean Sea. Ocean Science 4, 1–14.

Tsahalis, D., 1979. Theoretical and experimental study of wind- and wave-induced drift. Journal of Physical Oceanography 9, 1243–1257.

Ursella, L., Poulain, P.-M., Signell, R.P., 2006. Surface drifter derived circulation in the northern and middle Adriatic Sea: response to wind regime and season. Journal of Geophysical Research 112, C03S04. doi:10.1029/2005JC003177.

van Leeuwen, P., 2003. A variance-minimizing filter for large-scale applications. Monthly Weather Review 131, 2071–2084.

Veneziani, M., Griffa, A., Poulain, P., 2007. Historical drifter data and statistical prediction of particle motion: a case study in the Central Adriatic Sea. Journal of Atmospheric and Oceanic Technology 24, 235–254.

Williford, C.E., Krishnamurti, T., Torres, R., Cocke, S., 2003. Real-time multi-model superensemble forecasts of Atlantic tropical systems of 1999. Monthly Weather Review 131, 1878–1894.

Yun, W., Stefanova, L., Krishnamurti, T., 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. Journal of Climate 16, 3834–3840.

Yun, W., Stefanova, L., Mitra, A., Vijaya Kumar, T., Dewar, W., Krishnamurti, T., 2005. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus 57, 280–289.

Zanasca, P., Gerin, R., Poulain, P.-M., 2007. MREA07-LASIE07 Drifter Database: 14 May 2007 to 19 October 2007. <http://doga.ogs.trieste.it/sire/drifter/mrea07-lasie07_database/>.