Indice

1 1 Il Mar Mediterraneo 3
1 2 La circolazione nel Mediterraneo 4
1.3 Mediterranean Forecasting System 10
1.4 Il modello numerico.
1.5 Il sistema osservativo.
1.5.1 Scomposizione del segnale di SSH
1.6 Assimilazione dei dati
1.6.1 Matrice di covarianza dell'errore di forecast
1.6.2 Assimilazione di SSH
1.6.3 Assimilazione di XBT
1.6.4 Schema di assimilazione in MFSTEP
1.7 Obbiettivi della tesi
Capitolo 2
2.1 Progetto Argo
2.2 Breve storia delle boe derivanti ad assetto neutro
2.3 Aspetti tecnici delle boe derivanti PALACE
2.4 Rete di distribuzione dati e archivio dati
2.5 Argo nel Mediterraneo
2.6 I dati Argo nel Mediterraneo nel periodo gennaio-giugno 2003 36
Capitolo 3 AA
Capitolo 5 44
3.1 Teoria della validazione
3.2 Definizione attributi di qualità del forecast
3.3 Misure scalari di Accuratezza
3.4 Calcolo di Funzioni Empiriche Ortogonali- Empirical Orthogonal
Functions- EOF
3.5 Risultati
3.5.1 Confronto qualitativo tra osservazioni e modello
3.5.2 Indici di accuratezza 55
3.5.3 Confronto osservazioni-modello tramite EOF
Capitolo 4
4.1 Confronto qualitativo tra i dati Argo e le analisi di MFS 72

4.2 Confronto quantitativo	
4.3 Risultati confronto quantitativo	
4.4 Introduzione al problema di assimilazione	91
Capitolo 5	
5.1 La simulazione	
5.2 Confronto Simulazione e Modello Operativo.	
5.3 Risultati Statistici	
5.3.1 Temperatura	
5.3.2 Salinità	
5.4 Considerazioni generali sul confronto quantitativo	
Capitolo 6	123
Conclusioni	
Bibliografia	125

"... poiché tutte le nostre misure ed osservazioni della realtà non sono altro che approssimazioni, lo stesso deve valere per ogni calcolo che da queste dipenda, e il più alto scopo di tutte le elaborazioni numeriche riguardo a fenomeni concreti deve essere di approssimare bene, per quanto possibile, la realtà." Gauss (1806)

Capitolo 1

1.1 Il Mar Mediterraneo.

Il Mediterraneo è un bacino semi chiuso localizzato alle medie latitudini tra Europa, Africa e Asia. Ha una superficie di 2.5 milioni di m² e un volume di 3.7 milioni di m³, con una profondità media di 1500 m ed una massima di 5020 m, raggiunta nella fossa Ellenica. Ha una estensione longitudinale di circa 3900 km ed una latitudinale di circa 600 km, che si riduce in prossimità del Canale di Sicilia permettendo di distinguere due bacini, quello occidentale e quello orientale. Il Canale di Sicilia ha una profondità non superiore ai 500 m e ciò impedisce alle acque profonde di passare lo stretto, influendo notevolmente sulla circolazione generale. Il bacino occidentale comunica con l'Atlantico Settentrionale attraverso lo Stretto di Gibilterra che ha una profondità media di 320 metri e una lunghezza minima di 22 km. L'interazione con l'oceano Atlantico è di primaria importanza per la circolazione e la stessa esistenza del Mediterraneo. Negli ultimi anni si sta radicando la convinzione che anche il Mediterraneo eserciti un ruolo primario sulla circolazione atlantica, contribuendo con le sue masse d'acqua ad alta salinità alla formazione di acque profonde nel Mar di Labrador (Robinson, 2001).



Figura 1.1.1: Morfologia del Mediterraneo.

Il bacino orientale comunica a Nord con il Mar di Marmora, tramite lo stretto dei Dardanelli, e quindi al Mar Nero, tramite lo stretto del Bosforo. Tuttavia la scarsa profondità di tali stretti rende gli scambi molto limitati.

All'interno di questi bacini principali si riconoscono da Ovest verso Est i seguenti mari: il Mare di Alboran, il bacino Algero-Provenzale, il Mar Tirreno, il Mar Ionio, il Mar Adriatico, il bacino Levantino e il Mar Egeo.

1.2 La circolazione nel Mediterraneo

La circolazione nel Mediterraneo è guidata essenzialmente da tre forzanti che sono lo stress del vento, i flussi di acqua e calore all'interfaccia aria-mare, e il sistema di entrata/uscita di acqua dallo Stretto di Gibilterra. Questi fattori interagiscono tra loro per generare un sistema di circolazione complesso che interessa diverse scale spaziali: dalla scala di bacino alla mesoscala.

Il Mediterraneo è caratterizzato da un bilancio negativo di acqua e calore (Pinardi, 1993); per cui soltanto i flussi entranti e uscenti dallo Stretto di Gibilterra consentono al bacino Mediterraneo di riequilibrarsi su una scala temporale di diverse decadi. Ciò produce una circolazione di tipo antiestuarino, con acque superficiali che entrano, e acque profonde che escono, con un'interfaccia posta a 150 metri di profondità.

E' uso comune individuare le diverse masse d'acqua in base al rapporto tra temperatura e salinità. Nel Mediterraneo sono presenti:

- Acqua Atlantica Modificata (MAW): viene caratterizzata solo dalla sua bassa salinità (36.5 e 38.5 rispettivamente ad ovest ed est del Canale di Sicilia). Entra da Gibilterra occupando i primi 100 m della colonna d'acqua nello stretto. Successivamente al suo ingresso si trasforma in acqua mediterranea superficiale: l'acqua Atlantica superficiale subisce evaporazione e viene rimescolata con l'acqua più densa degli strati inferiori, con conseguente progressivo aumento della salinità. Il suo percorso può essere tracciato come un minimo di salinità prima superficiale poi sottosuperficiale. Nel bacino levantino si trova ad una profondità di 50 m.
- Acqua Levantina Intermedia (LIW): è caratterizzata da temperature tra 15-17.5 °C e salinità tra 38.95 e 39.10 psu. Si forma nel bacino orientale mediante processi di convezione nel periodo invernale ad una profondità compresa tra 200-300 m. Da qui si propaga verso il bacino occidentale dove si trova ad una profondità intermedia tra i 300-700 m. La LIW è tracciata come un massimo di salinità sottosuperficiale con media di 38.4 psu. E' essenzialmente la LIW che costituisce la massa d'acqua uscente dallo stretto di Gibilterra.
- Acqua Mediterranea Profonda (MDW): è caratterizzata da temperature tra 12.6-13.3 °C e salinità tra 38.4-38.65 psu. Si forma nel periodo invernale nel Golfo del Leone per il bacino occidentale (WMDW), nell'Adriatico e nell'Egeo per il bacino orientale (EMDW). La EMDW non riesce a passare il Canale di Sicilia per cui non entra mai in contatto con WDMW, benché questa ultima riesca in parte a mescolarsi con LIW ed a uscire da Gibilterra.

La circolazione nel Mediterraneo è il risultato di una complessa interazione tra la variabilità di mesoscala, quella stagionale e le scale stagionali ed interannuali (Pinardi et al., 1997).

Nel Mediterraneo occidentale il segnale di variabilità stagionale è grande e riguarda l'intensità e la direzione delle correnti e la variabilità di mesoscala. Nel Mediterraneo orientale la variabilità interannuale è di grandezza paragonabile al segnale stagionale e coinvolge cambiamenti delle caratteristiche delle masse d'acqua profonde ed intermedie (Korres et al., 2000). In questa zona, anche la variabilità di mesoscala è molto accentuata, ne sono un esempio i vortici, che compaiono e si rinforzano senza un'apparente regolarità.

Per semplicità di esposizione si può suddividere la circolazione del Mediterraneo in tre grandi categorie in funzione delle scale spaziali e temporali:

- Circolazione a scala di bacino.
- Circolazione a scala di sotto-bacino.
- Circolazione di mesoscala.

La circolazione a scala di bacino è dominata dalla circolazione termoalina; si ha quindi che tutti i movimenti verticali e orizzontali delle masse d'acqua a scala di bacino sono principalmente dovuti a differenza di densità (figura 1.2.1). Nel Mediterraneo questo tipo di circolazione è a 'nastro trasportatore' ed è di due tipi:



Figura 1.2.1: Schema della circolazione termoalina (Pinardi e Masetti, 2000).

- il primo, (nastro giallo-rosso) quello zonale, corrisponde alla acque atlantiche che entrano da Gibilterra. Queste attraverso interazioni aria-

mare subite lungo il percorso verso le coste del Medio Oriente (in una zona compresa tra Cipro e Rodi) si trasformano in acque Levantine. Esse si dirigono poi verso occidente fuoriuscendo nuovamente da Gibilterra. La variabilità è a scala stagionale.

 Il secondo, il 'nastro trasportatore' meridionale (nastro blu), è guidato dalle acque profonde. La profondità del riciclo è più alta visto che interessa la formazione di acque profonde e non intermedie, come nel caso precedente. Questo 'nastro trasportatore', più costante nel tempo, presenta una variabilità interannuale.

La circolazione termoalina è forzata dai processi di formazione di acque profonde ed intermedie. Il processo di formazione dipende da tre fattori principale.

- L'arrivo di masse continentali di aria fredda e secca sul mare.
- La presenza di una circolazione ciclonica che rende la stratificazione dell'acqua molto debole al centro del vortice.
- L'aumento di salinità dell'acqua dovuto ad un fenomeno di formazione del ghiaccio o alla risalita d'acqua salata da strati profondi.. Questo processo, praticamente puntiforme nello spazio, regola tutta la circolazione termoalina dell'oceano globale e del Mediterraneo (Pinardi, 1993).

Nel Mediterraneo esistono diversi regioni dove ci sono le condizione sopra descritte per la formazione di acque profonde. Nel bacino occidentale tale regione si colloca nel Golfo del Leone intorno ai 42° N e 5° E. La formazione di acque profonde, in questa area, avviene in risposta al vento di maestrale che soffia durante l'inverno, e che permette la perdita di importanti quantità di calore all'interfaccia aria-mare (fino a $1000W/m^2$) (Pinardi, 1993).

La situazione nel bacino orientale è più complessa. Nel Mar Adriatico esistono due zone di formazione di acque profonde, una nel Nord e una nel Sud del bacino. La acque inabissate escono attraverso lo Stretto di Otranto senza mescolarsi con la WMDW, la profondità del Canale di Sicilia ne impedisce il contatto.

Recenti osservazioni indicano che a nord-est del bacino levantino, zona di formazione di LIW, si possono creare le condizioni per la formazione 'Levantin Deep Water' (LDW); per lo più durante inverni particolarmente freddi (Robinson et al., 2001).

Un ulteriore zona di formazione di acque profonde nel Mediterraneo orientale è nel Mar Egeo (AGDW). Queste acque profonde hanno assunto un ruolo di primaria importanza dopo il 1987 (Eastern Mediterranean Trasient), quando hanno cambiato la circolazione termoalina, apportando modifiche su scala di bacino alle acque mediterranee. Questa fonte genera masse d'acqua più calda, salata e densa rispetto alle masse che esistevano precedentemente, di origine adriatica. La nuove acque profonde si posizionano nelle aree profonde dello Ionio andando a bloccare la LIW che viaggia verso ovest.

Esistono, nel Mediterraneo, zone di formazione di acque intermedie, che avvengono attraverso processi analoghi, ma che sprofondano solo fino allo strato del termoclino (posizionato tra i 300 e 500 metri di profondità). La LIW, che si colloca in questo strato, si forma durante gennaio-febbraio nel giro di Rodi per poi propagarsi verso occidente.

Oltre il passaggio di Creta, la corrente LIW si biforca: un ramo dirige verso nord, vero lo stretto di Otranto, l'altro si snoda al centro dello Ioni arrivando al Canale di Sicilia. Nel bacino occidentale la LIW circuita ciclonicamente nel Tirreno, ne fuoriesce a sud della Sardegna per poi costeggiarla verso Nord arrivando al Golfo del Leone. Una parte di essa partecipa alla formazione di acque profonde, mentre la rimanente fuoriesce dallo stretto di Gibilterra.

Andando verso scale spazio-temporali più brevi si ha la circolazione a scala di sotto-bacino. Si illustrerà solo la circolazione superficiale, che interessa i primi 100 metri, dove le strutture sono più evidenti(figura 1.2.2). Nel Mare di Alboran.



Figura 1.2.2: Schema di circolazione generale nel Mediterraneo.

l'acqua di origine atlantica forma due giri anticiclonici, chiamati appunto giri di Alboran. Proseguendo verso est la corrente di acqua atlantica prende il nome di corrente Algerina. Questa corrente è ben definita e profonda lungo le coste marocchine ma diventa più ampia ed irregolare nel procedere verso est lungo la scarpata algerina e fino al Canale di Sicilia (Robinson et al., 2001).

La corrente algerina è molto instabile e forma meandri dai quali si staccano vortici sia ciclonici che anticiclonici. I vortici si spostano verso est e solo quelli anticiclonici aumentano di dimensioni raggiungendo un diametro di 50-100 km (Millot, 1991), come messo in evidenza da osservazioni satellitari.

A sud della Sardegna la corrente di acqua Atlantica si biforca: un ramo si dirige verso nord, circola ciclonicamente nel Tirreno e arriva al bacino Liguro-Provenzale. La corrente in questa zona è chiamata Liguro-Provenzale-Catalana. L'altro ramo passa il canale di Sicilia e assume il nome di corrente Ionica-Atlantica e attraversa lo Ionio a metà dividendolo in due aree: una a nord composta dal giro ciclonico dello Ionio occidentale e l'altra a sud, composta da uno o più giri anticiclonici.

A sud di Creta la corrente di acqua Atlantica prende il nome di 'Mid Mediterranean Jet' e a sua volta si divide in due rami: uno circola in direzione nord verso il giro di Rodi, l'altro continua il suo percorso zonale fino alle coste del Libano dove si biforca nuovamente. In questa sono presenti molte strutture di sotto-bacino: il giro ciclonico di Rodi, il giro anticiclonico di Mersa-Matruh ed il sistema di vortici anticiclonici collocati a sud-est del bacino levantino: fra essi c'è il vortice di Shikmona a sud di Cipro.

Le componenti principali della circolazione a scala di sottobacino sono quindi i giri semipermanenti il cui forzante principale è il rotore del vento (Pinardi et al., 1993). I flussi di calore ad acqua rafforzano l'azione del vento, in particolare durante l'inverno e per i giri ciclonici del bacino.

Le strutture di circolazione di mesoscala non possono essere raffigurate come fatto per il caso della circolazione termoalina e la circolazione di sotto-bacino poiché le scale temporali diminuiscono ulteriormente, e vortici di segno opposto appaiono e scompaiono nella durate di poche settimane. I vortici di mesoscala sono stati campionati per la prima volta da Robinson et al. (1986) per il bacino levantino, da Paschini et al. (1993) per il Mare Adriatico e recentemente analizzati in grande dettaglio dai dati di altimetria da satellite (Ayoub et al., 1997). In generale si osserva che i vortici di mesoscala hanno un diametro pari a 4-5 volte il raggio locale di deformazione di Rossby (circa 10 km nel Mediterraneo). Si formano per instabilità idrodinamica delle forti correnti a getto (corrente Algerina, corrente Atlantico-Ionica, 'Mid Mediterranean Jet', ecc.) e normalmente sono evanescenti lontano dalla sorgente di energia che li ha creati.

1.3 Mediterranean Forecasting System

La previsione delle correnti a scala di bacino è di primaria importanza per l'implementazione di modelli a scala ragionale che possano essere utilizzati per la gestione sostenibile delle risorse marine e per la protezione dell'ambiente costiero (Pinardi et al., 2003).

Al fine di perseguire questi obbiettivi, nel 1995 è iniziato un programma di ricerca nel Mediterraneo sotto l'egida di EuroGOOS (www.eurogoos.org, Pinardi e Flemming, 1998) chiamato MFS-Mediterranean Forecasting System, il cui scopo principale è di valutare la predicibilità dell'ecosistema marino a livello di produzione primaria dalle scale di bacino alla costa.

Il primo progetto europeo ha avuto inizio nel 1998 sotto il nome di MFSPP-Mediterranean Forecasting System Pilot Project (Pinardi et al., 2002) ed ha visto la nascita delle prime previsioni di correnti, temperatura, salinità e altri parametri fisici nel Mediterraneo. Si sono inoltre iniziati i primi studi di simulazioni ecologiche della produttività primaria. La seconda parte del programma è MFSTEP (Mediterranean Forecasting System Toward Environmental Predictions) ed ha lo scopo di consolidare ed espandere le attività di previsione a scala di bacino fino alle aree regionali e di piattaforma continentale. La terza parte che inizierà nel 2005 si prefigge lo sviluppo di previsioni di produttività primaria, ed altri parametri biogeochimici in alcune zone costiere campione.

Il programma MFS si basa sull'ipotesi che la circolazione a scala di bacino influenzi direttamente, o indirettamente, le variazioni idrodinamiche e le proprietà degli ecosistemi sia nelle aree costiere sia nelle zone di piattaforma.

Affinché le previsioni siano consistenti con la realtà è necessario avere, oltre ad un modello di circolazione generale (OGCM), un sistema d'assimilazione in grado di utilizzare i dati acquisiti per inizializzare la previsione. In seguito si descriveranno gli elementi essenziali dell'OGCM, del sistema osservativo e dello schema di assimilazione dei dati.

1.4 Il modello numerico.

Il modello numerico attualmente utilizzato in MFSTEP è una versione del Modular Ocean Model (MOM) adattato al Mediterraneo da Roussenov et al. (1995) e da Korres et al. (2000). Il modello è basato su equazioni tridimensionali primitive, alle differenze finite. Le equazioni primitive sono equazioni di Navier-Stockes per i fluidi geofisici, ossia considerano un sistema rotante e la gravità, soggette ad alcune approssimazioni fondamentali. L'approssimazione di Boussinesq permette di trattare il fluido come incompressibile ed è giustificata dalle variazioni abbastanza piccole della densità nell'oceano: il che tuttavia esclude la possibilità di trattare le onde acustiche. Un'ulteriore approssimazione è quella idrostatica; ossia si fa dipendere il gradiente di pressione verticale soltanto dalla forza di gravità.

Il modello, inoltre, impone una superficie libera rigida in modo da filtrare le onde di gravità esterne la cui velocità (intorno a 250m/s) imporrebbero un allungamento dei tempi di calcolo senza che la fisica dei processi fisici di interesse (circolazione di bacino e di mesoscala) ne sia condizionata. Tuttavia ad ogni time step viene calcolata diagnosticamene l'altezza del livello del mare, come descritto da Pinardi et al. (1995). Il livello del mare è proporzionale alla pressione sulla superficie rigida, influenzata a sua volta, dalla dinamica interannuale e dal forzante meteorologico. Le onde di gravità esterne, la risposta alla pressione superficiale atmosferica e le maree, sono considerati trascurabili per il calcolo dell'anomalia del livello del mare nelle regioni profonde del mare (Pinardi et al., 1995).

Il modello ha una risoluzione orizzontale di 1/8 di grado; si hanno quindi 363 punti in longitudine (da -5° E a 36° E) e 113 in latitudine (da 30.5° N a 44.5 °N). I livelli verticali sono 31, arrivando ad una profondità massima di 3850 metri. La distribuzione degli intervalli verticali non è omogenea ma tale da realizzare una maggiore risoluzione nello strato superficiale. Ciò si giustifica facilmente in quanto la variabilità temporale del termoclino è maggiore rispetto agli strati

profondi, e quindi è necessaria una griglia più risolta per riuscire a riprodurre i fenomeni a scala più piccola che vi hanno luogo.

Il modello viene quindi integrato con un passo di 30 minuti, partendo da una condizione iniziale, fornita dall'analisi, dei campi di temperatura, salinità e velocità, imponendo la condizione al contorno del forzante atmosferico.

1.5 Il sistema osservativo.

L'attuale sistema osservativo del progetto MFSTEP presenta due componenti principali che sono il sistema volontario VOS-XBT e la telemetria satellitare. In questa tesi non vengono trattate i sistemi osservativi volti al biomonitoraggio, in quanto l'obbiettivo del lavoro concerne gli aspetti prettamente fisici del sistema.

I dati di XBT consistono in profili verticali di temperatura, sono raccolti lungo rotte di navi commerciali ogni 10 miglia nautiche in acque internazionali più profonde di 200 m. Hanno una risoluzione di 0.6 m e raggiungono una profondità massima di 460 o 760 m in base al tipo di sonda. I dati vengono trasmessi via satellite (sistema Argos) o mediante la rete di telefonia mobile dagli operatori a bordo. I centri ENEA-La Spezia e Global Teleconnection System (GTS) raccolgono i dati e ne effettuano un primo controllo di qualità sulle posizioni dei profili; un secondo controllo viene effettuato prima dell'utilizzo dal centro di previsione, attraverso un processo di depeaking e smoothing (Manzella et al., 2003).

L'altezza del livello del mare (SSH) viene misurata da sensori posti sui satelliti ERS-2, TOPEX/POSEIDON e JASON 1. I dati vengono raccolti da Collection e Localization Satellitaire (CLS) di Tolosa, Francia. Questa misura viene corretta rispetto agli errori nell'orbita dei satelliti e se ne sottrae una media climatologia di cinque anni (1993-1997) in modo da attenere una anomalia rispetto al livello medio del mare. I dati vengono pubblicati ogni mercoledì relativamente ai 22 giorni precedenti. In figura 1.5.1 sono riportate le tracce coperte dal satellite JASON 1 durante la settimana 11/11/03-17/11/03.



Figura 1.5.1: Dati di SLA nel periodo 11/11/03-17/11/03.

Il campo di temperatura superficiale viene misurato dai sensori AVHRR-14 AVHRR-15 posti su satelliti geostazionari NOAA. Il Centre Meteorologie Spatiale (CMS) di Meteo France a Tolosa e l'Istituto di Fisica dell'Atmosfera (IFA) de CNR di Roma raccolgono i dati relativi al Mediterraneo occidentale e orientale rispettivamente. Questi dati sono medie settimanali centrate sul lunedì e sono interpolati sulla griglia MOM mediante analisi oggettiva.

A queste misure si aggiungono i dati delle previsioni meteorologiche che sono impiegate nel sistema previsionale marino per valutare il forzante atmosferico. Questi dati meteorologici consistono in pressione media al livello del mare, componenti meridionali e zonali del vento misurato a 10m, temperatura di condensazione a 2 m e copertura nuvolosa totale. Il centro di raccolta è Meteo France, Tolosa; questi campi sono a disposizione ogni sei ore e sono utilizzati per forzare l'OGCM.

1.5.1 Scomposizione del segnale di SSH

Il segnale altimetrico può essere decomposto in questo modo (Le Traon, 2002):

$$SSH = N + \eta + \Sigma + \varphi; \qquad (1.5.1)$$

dove N è il geoide, η la topografia dinamica, Σ l'errore di misura (dovuto all'errore nell' orbita, alle correzioni atmosferiche, etc.) e φ rappresenta le componenti ad alta frequenza del livello del mare dovute alle maree e all'effetto barometrico inverso. La topografia dinamica è il segnale connesso con la circolazione termoalina e con la circolazione forzata dal vento; è perciò necessario isolare η dal segnale complessivo di *SSH*. Nel caso del Mediterraneo, il segnale ad alta frequenza φ viene sottratto dai dati prima che vengano distribuiti: ciò è reso possibile dalla bassa ampiezza del segnale.

La topografia dinamica contiene una componente barotropica ed una baroclina, ossia:

$$\eta = \frac{1}{H} \left\{ f \frac{\Psi}{g} + \frac{1}{\rho} \int_{-H}^{0} \rho \, z dz + \frac{H}{\rho_0} \int_{-H}^{H} \rho \, dz \right\}$$
(1.5.2)

dove ρ è la densità dell'acqua, ρ_0 il suo valore costante, *H* la profondità del mare assunta costante, *f* è il parametro di Coriolis, *g* l'accelerazione di gravità, Ψ la funzione di flusso barotropica espressa in Sverdrup $\left(10^6 \frac{m^3}{s}\right)$ (Pinardi et al., 1993). Il primo termine a destra dell' equazione (1.5.2) è la componente

barotropica mentre l'ultimo è il termine della pressione di fondo. L'integrale nel termine centrale rappresenta l'energia potenziale, ma è normalmente definito come il termine baroclino.

Il segnale di topografia dinamica può essere scomposto nella media, chiamata $\overline{\eta}$, e nella componente variabile η' , anche chiamata SLA. Poiché la forma del geoide è costante nel tempo, è possibile sottrarre N, per quanto non conosciuta, sottraendo la media dalla topografia dinamica.

In questo modo l'unica variabile disponibile in tempo reale (quindi operativamente) è la SLA. Questo processo richiede la conoscenza della media di SSH lungo le traccia del satellite per un periodo consistente di tempo. La media di SSH contiene sia N che η' e non è attualmente possibile separare i due segnali.

L'equazione della topografia dinamica (1.5.2) mostra che il segnale di SLA contiene il modo barotropico e i modi baroclini del moto. La componente geostrofica della variazione del livello del mare è principalmente dovuta ai cambiamenti termoalini stagionali e riguarda le componenti lente del moto alla scala di bacino. I processi all'interfaccia aria-acqua (per esempio l'azione del vento) producono moti baroclini che sono fortemente modulati dai campi di

vorticità alla scala di sotto-bacino e di mesoscala: questi moti compongono le principali strutture di circolazione nel Mediterraneo (Pinardi e Masetti, 2000).

Assimilare dati di SLA in un modello generale di circolazione significa perlopiù correggere la componente baroclina del moto utilizzando la relazione geostrofica alla scala di bacino: valori alti di SLA corrispondono ad alti valori sottosuperficiali di temperatura (Pinardi et al. 2003). Ciò implica che la pendenza delle isoterme e della SLA abbiano segni opposti.

La prima dimostrazione che la SLA è correlata con la profondità del termoclino è stata sviluppata per l'oceano Atlantico da Mellor e Ezer (1991); Masina (2001) ha mostrato che questo concetto è applicabile anche alle zone tropicali degli oceani.

L'equazione (1.5.2) definisce la dipendenza della topografia dinamica dalla stratificazione verticale della densità; da ciò consegue che l'assimilazione di SLA debba ripercuotersi anche sui valori di temperatura e salinità della colonna d'acqua.

1.6 Assimilazione dei dati

L'assimilazione dei dati è il processo di determinazione della migliore stima della realtà, partendo sia dalla osservazioni che dalla soluzioni numeriche dell'equazioni dei fluidi geofisici.

Nel contesto oceanografico e atmosferico, lo stato "vero" del sistema è considerato essere lo stato dell'oceano o dell'atmosfera senza le sue componenti veloci del moto (non sono considerate le onde sonore e le onde di gravità veloci). Inoltre, si assume che l'oceano sia quasi in uno stato geostrofico, e che non siano presenti marcati cambiamenti nelle sue variabili dinamiche nello spazio di pochi punti griglia. Ciò implica che osservazioni puntuali nello spazio e nel tempo possano dare un' informazione sullo stato "vero" del sistema.

Si assuma quindi che il vettore di stato "vero" sia:

$$X = \begin{vmatrix} T \\ S \\ U \\ V \\ W \\ \rho \\ \Psi \\ \eta \end{vmatrix}$$
(1.6.1)

dove, partendo dall'alto verso il basso, i simboli indicano: i campi tridimensionali di temperatura, salinità, le velocità zonali, meridionali e verticali, e la densità dell'acqua; mentre i campi bidimensionali sono la funzione di flusso barotropica e l'elevazione della superficie libera (SSH). Alcune di queste variabili sono prognostiche, ossia la loro evoluzione nel tempo può essere scritta da un'equazione contenente la derivata temporale, altre sono semplicemente diagnostiche, come la densità, la velocità verticale e l'elevazione della superficie libera. La SSH è considerata diagnostica perché l'assimilazione considera solo la componente lenta dell'oscillazione del livello del mare, come spiegato in precedenza.

Si supponga che si abbiano solo due approssimazione di X, che verrà ora considerato un valore scalare x, corrispondente ad una variabile di stato contenuta in (1.6.1). La prima approssimazione sia la soluzione numerica del modello, chiamata x^b , l'errore associato sia $E^b = x^b - x$; la seconda approssimazione sia l'osservazione y^o , presa nello stesso punto di x^b , con un errore $E^o = y^o - x$. Si assuma che la distribuzione della probabilità dell'errore sia Gaussiana e abbia la forma:

$$p(E) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{E^2}{2\sigma^2}\right)}$$
(1.6.2)

dove $\sigma^2 = \langle E^2 \rangle$, e le virgolette indicano l'operatore di valore atteso (Daley, 1996). La distribuzione della probabilità congiunta delle due approssimazioni è:

$$p(E^{b})p(E^{o}) = \frac{1}{\sigma_{b}\sigma_{o}\sqrt{2\pi}}e^{-\left(\frac{E^{b^{2}}}{2\sigma_{b}^{2}} + \frac{E^{o^{2}}}{2\sigma_{o}^{2}}\right)}$$
(1.6.3)

Si vuole quindi trovare il massimo di questa equazione, che significa minimizzarne l'esponente. Si chiami I il termine all'esponente:

$$I = \frac{1}{2} \left[\frac{(x^{b} - x)}{\sigma_{b}^{2}} + \frac{(y^{o} - x)}{\sigma_{o}^{2}} \right]$$
(1.6.4)

Si ha il minimo di *I* in corrispondenza del valore x^a , chiamato analisi:

$$x^{a} = x^{b} + \left(\frac{\sigma_{b}^{2}}{\sigma_{b}^{2} + \sigma_{o}^{2}}\right) \left(y^{o} - x^{b}\right)$$
(1.6.5)

Il secondo termine tra parentesi è chiamato misfit, e corrisponde alla differenza tra la soluzione del modello e l'osservazione. Il valore dell'analisi, o migliore stima, è quindi la media pesata della di x^b e il misfit (innovazione) tra osservazione e modello.

Generalmente le osservazioni sono distribuite diversamente rispetto ai punti griglia del modello; per cui non sono immediatamente rapportabili alle variabili di stato contenute in *X*. Ciò implica che il vettore di innovazione sia scritto nella forma più generale:

$$D = Y^{o} - H(X^{b}) \tag{1.6.6}$$

dove H è l'operatore di osservazione e D, Y^o , X^b sono vettori multivariati nello spazio quadridimensionale. Normalmente H() è una semplice interpolazione, ma nel caso di variabili diagnostiche, come SSH, o di altre quantità indirettamente derivate, può essere un'equazione complessa delle variabili di stato contenute in X^b . La generalizzazione di (1.6.4), considerando (1.6.6), nel caso multivariato è:

$$I = \frac{1}{2} \left[\left(X^{b} - X \right)^{T} B^{-1} \left(X^{b} - X \right) + \left(Y^{o} - H(X) \right)^{T} R^{-1} \left(Y^{0} - H(X) \right) \right] \quad (1.6.7)$$
$$X^{a} = X^{b} + K \left(Y^{o} - H(X^{b}) \right) \quad (1.6.8)$$

dove

$$K = BH^{T} \left(HBH^{T} + R \right)^{-1}$$
(1.6.9)

è la matrice di Kalman.

La matrice di covarianza dell'errore è definita come:

$$B = \left\langle \left(X^{b} - X \right) \! \left(X^{b} - X \right)^{T} \right\rangle$$
(1.6.10)

e rappresenta la varianza dell'errore per tutte le variabili di stato del modello e le loro correlazioni incrociate. La matrice di covarianza dell'errore di osservazione è rappresentata da R.

E' importante sottolineare la presenza di forti similitudini tra l'equazione semplice (1.6.5) e quella più complessa (1.6.8); in tutti e due i casi, il peso del vettore di innovazione è dato dall'errore nel modello (di background) diviso per la somma degli errori delle osservazioni e del modello. Nel caso in cui le variazioni temporali di *B* siano parametrizzate solo come funzione delle variazioni nei campi dell'errore, l'equazione (1.6.8) descrive lo schema di Interpolazione Ottimale (OI). Uno dei problemi centrali dell'assimilazione dei dati consiste nel trovare una rappresentazione degli aspetti multivariati, e della variabilità spaziale e temporale della matrice *B*. Normalmente si fa una scelta a priori della forma di *B*, basata sulla conoscenza dei processi principali descritti dal modello ed inclusi nelle osservazioni.

Il calcolo dell'analisi (1.6.8) può essere fatto in maniera intermittente, cioè, si collezionano osservazioni per un certo intervallo di tempo, per poi introdurle alla fine del ciclo.

L'analisi è quindi usata per inizializzare i campi del modello, e per avviare il processo di previsione. Le previsioni oceanografiche utilizzano le previsioni meteorologiche del forzante atmosferico, e risolvono, per un determinato numero di passi temporali δt , un'equazione non lineare del tipo:

$$X_{t+\delta t}^{b} = M\left(X_{t}^{a}\right) \tag{1.6.11}$$

I campi prodotti dalla previsione, X^{b} , verranno quindi usati come prima approssimazione nel successivo ciclo di assimilazione, dove verranno fusi con le osservazioni per produrre l'analisi successiva.

1.6.1 Matrice di covarianza dell'errore di forecast

La maggiore complicazione nella modellizzazione di *B*, deriva dal fatto che lo stato vero del sistema è sconosciuto; si devono quindi utilizzare delle approssimazioni, che in generale sostituiscono lo stato "vero" del sistema con un valore noto, ritenuto vicino al valore "vero". Le approssimazioni più comuni sono:

$$B = \left\langle \left(X^{f} - X^{a} \right) \left(X^{f} - X^{a} \right)^{T} \right\rangle$$
(1.6.12)

$$B = \left\langle \left(X^{f} - \overline{X}^{f} \right) \left(X^{f} - \overline{X}^{f} \right)^{T} \right\rangle$$
(1.6.13)

$$B = \left\langle \left(X^{o} - \overline{X}^{o} \right) \left(X^{o} - \overline{X}^{o} \right)^{T} \right\rangle$$
(1.6.14)

dove la barra sopra i vettori indica la media.

La prima equazione indica che l'errore è stimato in base alla differenza tra la previsione e l'analisi; questa procedura è probabilmente l'approssimazione più corretta. La seconda e la terza approssimazione implicano che l'errore sia dovuto alla varianza delle variabili di stato, rispetto a una media predefinita. Da un punto di vista matematico, l'approssimazione peggiore è l'ultima, in quanto presuppone che la varianza contenuta nelle osservazioni sia la stessa del modello. Tuttavia, la maggior parte dei sistemi di assimilazione applicano questa approssimazione. Infatti, il modello potrebbe aver deviato della realtà, e l'utilizzo della covarianza dedotta dal modello stesso potrebbe avere severe conseguenze sull'efficienza dell'algoritmo di assimilazione.

In seguito si assumerà *B* nella forma (1.6.13) o (1.6.14), che può essere riscritta come:

$$B = \left\langle \widetilde{X}' \ \widetilde{X}'^{T} \right\rangle \tag{1.6.15}$$

dove i primi indicano le anomalie rispetto ad una media, e \tilde{X} si riferisce ai campi del modello o delle osservazioni. L'informazione importante contenuta in *B* è la covarianza tra tutte le variabili dinamiche del sistema. In altre parole, *B* è multivariata ed è composta da blocchi di matrici:

$$B = \begin{vmatrix} \langle T'T' \rangle & \langle T'S' \rangle & \dots & \dots & \langle T'\eta' \rangle \\ \langle S'T' \rangle & \langle S'S' \rangle & \dots & \dots & \langle S'\eta' \rangle \\ \langle U'T' \rangle & \dots & \langle U'U' \rangle & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \langle \eta'T' \rangle & \dots & \dots & \dots & \langle \eta'\eta' \rangle \end{vmatrix}$$
(1.6.16)

dove le tilde e molti blocchi di matrici sono state trascurati per semplicità. Ogni blocco in (1.6.16) spiega la varianza dell'errore associata ad ogni campo, e rappresenta la correlazione incrociata tra variabili di stato nello spazio.

De Mey e Benkiran (2002) hanno proposto che una possibile scomposizione di (1.6.16) sia :

$$B = SBrS^{T} \tag{1.6.17}$$

dove *S* contiene le Funzioni Ortogonali Empiriche (EOF) verticali multivariate, e Br è una matrice di correlazione bidimensionale, calcolata analiticamente. Ciò significa che le componenti orizzontali e verticali della matrice di covarianza dell'errore sono state separate. Questa scomposizione di *B* può essere applicata all'oceano aperto, e funziona specialmente per l'assimilazione quasigeostrofica di SSH (De Mey et Robinson, 1986).

Nel contesto di MFS, le EOF verticali multivariate sono state calcolate partendo da dati storici (Sparnocchia et al., 2003), ed è stato applicata l'approssimazione (1.6.14).

La matrice *S* rappresenta la variabilità dell'errore di tutte le variabili di stato e le relative correlazioni incrociate rispetto alla verticale. Sostituendo (1.6.17) in (1.6.9), otteniamo:

$$K = SBrS^{T}H^{T} (HSBrS^{T}H^{T} + R)^{-1}$$
(1.6.18)

Si noti che Hr=HS è una matrice bidimensionale, che proietta l'operatore di osservazione nell'ampiezza delle EOF verticali. Il problema multidimensionale posto in (1.6.18) può essere ridotto di ordine, poichè le EOF rilevanti sono certamente un numero minore rispetto ai livelli verticali del modello. Le EOF di una matrice generica individuano uno spazio ortogonale di vettori che descrivono la varianza della matrice, rispetto ad un sistema di basi che raccolgono una quantità decrescente della varianza totale; in altre parole, i primi modi individuati dalle EOF descrivono la maggior parte della varianza, ed è quindi possibile trascurare i modi restanti. La ragione fisica per cui questa procedura matematica è applicabile all'oceano, è che al di sotto del termoclino la varianza delle variabili dinamiche è bassa, ed entro la scala stagionale, la varianza tra le masse d'acqua superficiali, intermedie e profonde, è incorrelata.

L'algoritmo di assimilazione usato nel sistema previsionale di MFS è il SOFA, acronimo di System for Ocean Forecasting and Analysis (De Mey and Benkiran, 2000). Il SOFA consiste essenzialmente in un sistema di Interpolazione Ottimale di Ordine Ridotto (ROOI), dove la riduzione di ordine è ottenuta considerando solo i modi più importanti delle EOF verticali. In questo caso, la matrice del guadagno di Kalman può essere scritta come:

$$K^{ROOI} = \tilde{S}Kr \tag{1.6.19}$$

$$Kr = Br\widetilde{S}^{T}H^{T}\left(H\widetilde{S}Br\widetilde{S}^{T}H^{T} + R^{*}\right)^{-1}$$
(1.6.20)

dove le dimensioni si \tilde{S} sono minori di S.

La più importante implicazione pratica di questo schema di assimilazione è che, sebbene il vettore d'innovazione possa essere disponibile per un numero limitato di variabili, la correzione indotta riguarderà tutte le variabili di stato, la cui correlazione incrociata con le osservazioni disponibili è significativa. Da ciò consegue che l'assimilazione di diverse tipologie di osservazioni utilizza diverse formulazioni della matrice B e diversi gruppi di EOF.

Nei prossimi paragrafi verrà brevemente discussa l'assimilazione dei dati di SLA e XBT.

1.6.2 Assimilazione di SSH

Si supponga che siano disponibili solo dati di SSH, e che la matrice *B* sia scritta come:

$$B = \begin{vmatrix} \langle T'T' \rangle & \langle T'S' \rangle & \langle T'\Psi' \rangle \\ \langle S'T' \rangle & \langle S'S' \rangle & \langle S'\Psi' \rangle \\ \langle \Psi'T' \rangle & \langle \Psi'S' \rangle & \langle \Psi'\Psi' \rangle \end{vmatrix}$$
(1.6.21)

Ciò significa che il misfit tra osservazione ed analisi di SSH sarà proiettato nei profili verticali di temperatura e salinità, e sull'ampiezza della funzione di flusso Ψ : la correzione, quindi, riguarderà tre variabili dinamiche del sistema e non la SSH in sé stessa. Per questo motivo la SSH nel contesto dell'assimilazione è considerata una variabile diagnostica.

Nel contesto di MFS, si utilizzano EOF verticali trivariate ricostruite dal vettore di stato:

$$X^{a} = \begin{vmatrix} T \\ S \\ \Psi \end{vmatrix}$$
(1.6.22)

e si considera la forma (1.6.21) di B. Come è possibile vedere nell'equazione della topografia dinamica (1.3.2), la SLA è funzione di T, S e Ψ . Inoltre, applicando l'equazione geostrofica scritta in (1.3.2), si ha che l'operatore di osservazione H() contiene esplicitamente la stessa equazione (Pinardi et al., 1993). In questo modo si ottiene che la correzione sia proiettata sulle variabili dinamiche T, S e Ψ , pesate per il corrispettivo coefficiente di scala, in modo da riprodurre il corretto valore di η o SSH.



Figura 1.6.1: EOF trivariata adimensionale utilizzata in SOFA per l'assimilazione di SLA.

In figura 1.6.1 è rappresentata la EOF verticale trivariata, utilizzata nell'algoritmo di assimilazione SOFA. La parte di EOF relativa a temperatura e salinità si applica a profondità maggiori di 120 metri; in questo modo si isola la componente barotropica della topografia dinamica.

Un'unica EOF trivariata è applicata a tutto il bacino per tutto l'anno. Si è quindi supposto che la matrice di correlazione dell'errore abbia una struttura analoga in tutto il Mediterraneo, e che non risenta dei cicli stagionali. La validità di questa assunzione verrà testata nel corso di questa tesi.

1.6.3 Assimilazione di XBT

I profili di XBT costituiscono una preziosa informazione per ogni sistema previsionale. Come già è stato descritto, i dati vengono raccolti in tempo reale per essere disponibili all'assimilazione.

In questo caso, la matrice di covarianza dell'errore può essere scritta come:

$$B = \begin{vmatrix} \langle T'T' \rangle & \langle T'S' \rangle \\ \langle S'T' \rangle & \langle S'S' \rangle \end{vmatrix}$$
(1.6.23)

anche chiamata matrice di correlazione incrociata delle masse d'acqua. Quindi per l'assimilazione di XBT sono utilizzate EOF bivariate calcolate considerando la variabilità orizzontale in diverse sotto regioni del bacino (Sparnocchia et al., 2003).

Le EOF bivariate proiettano il misfit di temperatura in una correzione della salinità del modello. E' quindi essenziale che le EOF contengano una informazione statistica appropriata al contesto fisico a cui sono applicate. Se infatti le EOF non fossero rappresentative dei processi fisici in atto, si introdurrebbe un errore in tutte le variabili dinamiche correlate con l'osservazione.

1.6.4 Schema di assimilazione in MFSTEP.

La sequenza è chiamata "analisi-previsione" e prepara settimanalmente le condizioni iniziali per ogni previsione. Per ogni ciclo vengono effettuate due assimilazioni diverse: la prima è di tipo "filtro equidistante" e riguarda un set di

dati, per esempio XBT, la seconda è in modalità "filtro" e riguarda un set di dati diverso, per esempio SLA: al ciclo successivo le procedure si invertono (Demirov et al., 2003).



Figura 1.6.2: Procedura analisi-previsione.

Il "filtro equidistante" utilizza osservazioni della settimana precedente e successiva al giorno predisposto per l'analisi; questo filtro deve necessariamente essere utilizzato solo per analisi nel passato. La modalità "filtro" utilizza osservazioni solo nel passato, e viene quindi applicato per produrre l'analisi di inizializzazione della previsione (figura 1.6.2).

Le due modalità di filtro si basano su lo stesso algoritmo di Interpolazione Ottimale di Ordine Ridotto, ma le EOF multivariate verticali sono diverse nel caso di XBT o SLA. Si fa questo per soddisfare il bisogno formale che la proiezione dell'operatore di osservazione *Hr* nello spazio nullo di Kalman sia marginale, e che le EOF abbiano un impatto relativamente alto sulle osservazioni (Demirov et al., 2003). Lo stesso set di EOF non può soddisfare le condizioni di spazio nullo per SLA e XBT, in quanto le misure di SLA contengono un'informazione che riguarda la variabilità di temperatura e salinità sotto lo strato di rimescolamento, mentre gli XBT danno la temperatura di questo strato.

Infine, le osservazioni di SST sono assimilate continuamente attraverso la correzione del flusso di calore totale alla superficie:

$$Q^{M} = Q^{D} + C(SST^{D} - SST^{M}); \qquad (1.6.24)$$

dove Q^{D} è il flusso totale di calore alla superficie, calcolato usando il forzante meteorologico, Q^{M} è il flusso totale di calore alla superficie usato nelle simulazioni del modello, SST^{D} sono le osservazioni da satellite e SST^{M} sono i valori del modello. La costante empirica *C* è posta uguale a 70 $Wm^{-2}K^{-1}$.

1.7 Obbiettivi della tesi

Come è stato discusso in questo capitolo, il sistema previsionale di MFS è formato da molteplici componenti. E' quindi necessario che ogni aspetto del sistema venga monitorato per individuare, e possibilmente correggere, eventuali malfunzionamenti. Fin dalle prime fasi del progetto pilota di MFS sono state condotte diverse analisi volte a valutare le singole componenti del sistema. Qui di seguito ne viene riportata una breve sintesi.

L'impatto del sistema VOS-XBT sulla funzionalità dell'assimilazione è stato oggetto dello studio di Raichic et al. (2003). Nel corso di questo lavoro sono stati condotti diversi esperimenti di assimilazione basati su sistemi VOS-XBT alternativi allo schema applicato durante il progetto pilota di MFS. Alcuni dei risultati più interessanti riguardano l'individuazione delle rotte XBT, che esercitano un maggiore impatto sull'assimilazione, e la frequenza ottimale di campionamento.

La consistenza tra i dati di XBT e SLA è stata analizzata da Vignudelli et al. (2003). La rotta XBT Genova-Palermo è virtualmente coincidente con la traccia 0044 del satellite TOPEX/Poseidon; questo ne ha facilitato il confronto. Sono stati utilizzati i dati XBT e le misure di altimetria per calcolare l'altezza sterica del mare, e quindi il misfit tra i valori ottenuti. Questa analisi ha dimostrato una sostanziale coerenza tra i dati di altezza sterica calcolati da XBT e SLA, ma ha altresì sottolineato la complementarietà di questi dati per ottenere una descrizione dimanicamente completa della circolazione. In questa analisi sono stati utilizzati dati climatologici di correlazione tra temperatura e salinità per poter ricostruire i profili di densità, e quindi l'altezza sterica a partire di dati di sola temperatura. Gli autori individuano in questa procedura una delle principali fonti di errore dell'analisi condotta.

Un'ulteriore fase critica del sistema osservativo riguarda la trasmissione e la gestione dei dati in tempo reale. Nello specifico si è riscontrato che la procedura di sotto campionamento delle misure XBT, applicata durante il primo periodo della fase operativa di MFSPP, non era pienamente in grado di gestire profili invernali caratterizzati da un prevalente modo barotropico (Manzella et al., 2003). Inoltre, il limite di 15 misure per profilo, reso necessario dalla scarsa velocità di connessione satellitare, si è rivelato non sufficiente per risolvere la struttura verticale di temperatura; si è pertanto deciso di trasmettere i dati di XBT a piena risoluzione utilizzando, dove possibile, la rete di telefonia cellulare.

Il lavoro di Demirov et al. (2003) presenta i risultati di una verifica d'insieme del sistema previsionale. Sono stati calcolati indici di qualità e accuratezza del forecast rispetto all'analisi e ad osservazioni di temperatura e SLA non ancora assimilate. In questo caso la verifica ha riguardato sia la capacità complessiva del sistema previsionale di produrre forecast vicini alle osservazioni che la capacità del modello numerico di produrre forecast migliori della persistenza, ossia lo stato del sistema al primo giorno di previsione.

In questa tesi viene affrontato il problema di validazione del sistema previsionale nel suo complesso. Il punto di riferimento per questa analisi sono i dati di 8 boe Argo dislocate nel Mediterraneo orientale. Ciò che differenzia questa analisi dalle precedenti, è la disponibilità di dati di temperatura e di salinità totalmente indipendenti dal sistema previsionale. L'obbiettivo è quindi di verificare l'accuratezza delle analisi prodotte, l'andamento nel tempo dell'errore e la sua distribuzione nello spazio. Inoltre, si è valutata l'efficienza del sistema di assimilazione SOFA, confrontando i risultati della validazione del modello operativo con i risultati ottenuti da una simulazione in cui l'algoritmo di assimilazione è messo 'a riposo'.

La natura lagrangiana delle boe Argo permette, inoltre, di ricostruire i campi di velocità lagrangiana che hanno determinato gli spostamenti osservati. Si è quindi potuto verificare la consistenza dei campi di velocità del modello con gli spostamenti delle boe Argo.

La tesi è organizzata secondo questo schema:

- Capitolo 2. Descrizione Argo, dati e formati.
- Capitolo 3. Validazione del Sistema Previsionale.
- Capitolo 4. Confronto dei campi di velocità delle boe Argo e del modello.
- Capitolo 5. Esperimento di Simulazione.
- Capitolo 6. Conclusioni.

Capitolo 2

Nel corso di questa tesi si sono utilizzati i dati prodotti dal sistema di assimilazione di MFS e i dati Argo disponibili nel Mediterraneo nei primi 6 mesi del 2003. Nel Capitolo 1 è stata presentata la struttura del sistema previsionale MFS, in questo Capitolo viene descritto il progetto Argo, la tipologia dei dati disponibili e la loro distribuzione.

2.1 Progetto Argo.

Argo è un progetto internazionale volto al monitoraggio in tempo reale di temperatura, salinità e correnti degli oceani a scala globale. Il nome Argo è stato scelto per sottolinearne la complementarietà con il progetto Jason di altimetria satellitare. I dati Argo della struttura verticale di salinità e temperatura forniscono le informazioni necessarie per interpretare correttamente la variabilità associata alle misure satellitari dell'altezza del mare. L'obbiettivo primario è la comprensione della variabilità climatica a scala decennale e la sua prevedibilità, ma il progetto si presta ad un ampio numero di applicazioni per l'analisi ad alta risoluzione dello stato fisico degli oceani.

Il disegno iniziale della rete Argo è basato sull'esperienza acquisita dall'attuale sistema osservativo, in particolare grazie all'altimetro TOPEX/Poseidon, e sulla stima dei requisiti di modelli numerici ad alta risoluzione. Quando il progetto Argo sarà a regime, ogni anno saranno disponibili 100.000 profili di temperatura e salinità e altrettante misure di velocità fornite da circa 3000 boe distribuite in modo quasi uniforme su tutti gli oceani.

Tutti i dati Argo sono pubblicamente accessibili in tempo reale attraverso il sistema di teletrasmissione dati del GTS (Global Telecommunication System), e nel giro di qualche mese i dati vengono messi a disposizione dopo un controllo di qualità. Gli obiettivi del progetto Argo dovranno essere raggiunti durante il Global Ocean Data Assimilation Experiment-GODAE, che assieme a CLIVAR (CLImate VARiability) e GCOS/GOOS (Global Climate Observing System/Global Ocean Observing System) costituiscono il maggiore supporto scientifico e operativo per

il progetto Argo. In questo modo si vuole enfatizzare l'integrazione del progetto Argo con il sistema osservativo globale.

La pianificazione internazionale del progetto Argo è coordinata dall'Argo Science Team. Le principali nazioni che hanno attivato progetti di costruzione o dislocamento di boe Argo sono Australia, Canada, Francia, Giappone, U.K., U.S.A. e l'Unione Europea; nel corso del 2002 sono state rilasciate complessivamente da questi Stati più di 700 boe; ad oggi sono attive 943 boe (figura 2.1.1).



Figura 2.1.1: Argo attualmente attive nel mondo. I cerchi verdi indicano la posizione del rilascio della boa.

2.2 Breve storia delle boe derivanti ad assetto neutro.

Il primo impulso allo sviluppo di boe derivanti ad assetto neutro fu la necessità di misurare le correnti sottosuperficiali. Si riteneva infatti che i flussi profondi degli oceani fossero troppo piccoli per poter essere misurati da correntometri meccanici (Stommel 1955).

Stommel negli Stati Uniti e Swallow in Inghilterra lavorarono contemporaneamente e indipendentemente al progetto, ma fu Swallow nel 1955 a

sviluppare il primo prototipo funzionante di boa derivante ad assetto neutro. Il prototipo consisteva essenzialmente in un tubo di alluminio meno compressibile dell'acqua, in modo da essere in equilibrio di densità con il fluido ad una specifica profondità. La boa era localizzata tramite un sistema acustico a corto raggio (5 km) installato su una nave d'appoggio.

Questi strumenti furono ampiamente utilizzati nello studio della circolazione profonda, e rivelarono per la prima volta la struttura delle correnti di mesoscala che dominano l'oceano aperto.

Per quanto la tecnologia di localizzazione acustica sia stata migliorata nel corso degli anni, soltanto nel 1970 fu realizzata la prima boa localizzabile a grande distanza, circa 1500 km, e quindi applicabile allo studio delle correnti a scala di bacino. Questo prototipo prese il nome di SOFAR e fu ideato da Webb e Rossby. Il bisogno di installare a bordo del SOFAR un trasmettitore acustico efficiente a bassa frequenza, più il necessario gruppo energetico, impose allo strumento dimensioni molto grandi: 400 kg e 5 metri di lunghezza. Per localizzare le boe si utilizzarono stazioni mobili di ascolto fornite dalla Marina Militare degli Stati Uniti; successivamente furono disposti sistemi di ascolto ancorati in modo da aumentare la precisione di localizzazione.

Lo sviluppo di boe derivanti RAFOS (SOFAR letto al contrario) da parte di Rossby (1986) rese la localizzazione acustica a scala di bacino molto più semplice; le enormi boe SOFAR furono rimpiazzate da piccole boe derivanti ad assetto nullo che funzionavano come stazioni di ascolto. Le boe RAFOS, disposte nelle vicinanze di una rete di trasmettitori acustici a bassa frequenza (spesso costituiti da SOFAR) erano in grado di misurare la propria posizione. Alla fine della missione risalivano in superficie per trasmettere a terra i dati di posizione tramite il sistema satellitare Argos.

Le prime boe profilanti autonome ALACE (Autonomous Lagrangian Circulation Explorer) furono realizzate da Davis nel 1992. Questi strumenti eseguivano dei cicli di immersione/emersione, ed una volta in superficie potevano localizzarsi e trasmettere i dati tramite il sistema di teletrasmissione dati Argos. In questo modo non era più necessaria una rete acustica di localizzazione.

Le boe utilizzate nel progetto Argo sono di tipo P-ALACE, (Profiling Alace); la boa PALACE è quindi uno strumento ALACE in grado inoltre di misurare profili di temperatura e salinità.

2.3 Aspetti tecnici delle boe derivanti PALACE.

Il principio di funzionamento delle boe PALACE, illustrato in figura 2.4.1, si basa sul cambio di volume indotto dalla pressione esercitata da un pistone su una membrana elastica esterna. La variazione di volume, così provocata, cambia la densità complessiva dello strumento. Si ha quindi un controllo sulla forza di galleggiamento e di conseguenza sui moti verticali della boa derivante.

I sensori di pressione e i sistemi di calcolo interni allo strumento permettono di programmare il ciclo della boa PALACE in modo che vengano rispettati sequenze di immersione ed emersione a determinate profondità e intervalli di tempo.

La determinazione delle caratteristiche del ciclo di immersione/emersione delle boe derivanti è a discrezione del comitato scientifico che si occupa del progetto locale per il quale si intende usare le Argo. Il ciclo è caratterizzato da il tempo complessivo di durata dell'immersione, dalla profondità alla quale la boe deriva, dalla profondità da cui la boe comincia a misurare temperatura e salinità, e infine dal programma di decimazione applicato ai dati.



Figura 2.4.1: Schema di funzionamento PALACE.

La scelta di questi parametri è vincolata dal contesto fisico in cui si vogliono utilizzare le boe profilanti e da contingenze economiche. Nell'oceano si applicano generalmente cicli della durata di 10-14 giorni, con profondità di deriva a 2000 m. In alcuni casi la profondità di deriva è minore rispetto alla profondità da cui comincia la misura di temperatura e salinità; in figura 2.4.2 è riportato una schematizzazione di questo ciclo. In figura 2.4.3 è mostrato in tre dimensioni un ciclo analogo.



Figura 2.4.2: Schematizzazione del ciclo Argo applicato alle boe Apex rilasciate nell'oceano Atlantico.

Nel contesto del progetto Argo sono utilizzate principalmente due tipi di boe PALACE:

- PROVOR, prodotta dal MARTEC, France.
- APEX, prodotta da WEBB Co (Falmouth, USA).

In tabella 2.4.1 sono riportate alcune caratteristiche tecniche dei due strumenti.

I sensori di temperatura e salinità (conducibilità) possono essere di due tipi (SBE-41 e FSI). E' importante mettere in luce il fatto che i sensori della salinità subiscono un processo di deterioramento nel tempo che porta ad un malfunzionamento del sensore di conducibilità. E' stato misurato che il deterioramento porta ad un errore che non supera i 0.01 PSU nel corso di due anni. L'errore di misura dei sensori di temperatura è di 0.01 °C, l'errore nel posizionamento satellitare è nell'ordine dei 100 m.



Figura 2.4.3: Ciclo boe Argo.

Тіро	APEX (WEBB)	PROVOR(MARTEC)
Dimensioni	16.5x130 26kg	*
Batteria	alcaline	Litio
Numero di cicli performati	Più di 150	Più di 100
Tempo di consegna	4 settimane	4 mesi
Procedura di 'Start'	Semplice	Complessa
Costo	14,500 USD	11,000 USD

Tabella 4.2.1: caratteristiche tecniche delle boe APEX e PROVOR

2.4 Rete di distribuzione dati e archivio dati.

Esistono due centri globali di raccolta dei dati Argo: il server dati US GODAE, localizzato a Monterey in California e il server IFREMER di Brest, Francia. Lo scopo di questi centri (chiamati GDacs) è di raccogliere automaticamente tutti i dati disponibili di profili e traiettorie delle boe, nonché i metadata rilasciati dai centri regionali o nazionali (Dac). Entrambi i GDacs sono aggiornanti contemporaneamente in modo da assicurarne la consistenza reciproca.

Tutti i centri di disseminazione dei dati Argo utilizzano il formato NetCDF. La definizione delle dimensioni e delle variabili contenute nei file NetCDF Argo è accettata a scala globale ed è mantenuta costante in modo da facilitare l'utilizzo operativo dei dati Argo. La scelta del formato NetCDF è dovuta diverse considerazioni

- è ampiamente utilizzato dalla comunità degli utenti.
- è autodescrivente ed esistono molteplici strumenti di utilizzo.
- è efficiente ed affidabile.

I due centri GDacs forniscono un accesso ai dati sia FTP che WWW. L'accesso FTP, più funzionale per la fruizione automatica dei dati, è strutturato:

- Geograficamente (per ogni bacino oceanico).
- Per centro di raccolta dati e per piattaforma (Dac e identificativo boa Argo).
- Per la data di creazione dei file (accesso agli ultimi 12 mesi).

I file corrispondono a singole boe e sono divisi in quattro categorie:

- Dati dei profili; sono archiviati come NetCDF. Contengono tutte le informazioni relative alle misure di salinità, temperatura, pressione e i corrispettivi controlli di qualità. Inoltre sono registrate le coordinate del primo posizionamento effettuato in superficie dopo ogni risalita. Possono essere dati in tempo reale o ritardato.
- Dati delle traiettorie: sono archiviati in NetCDF. Vi sono riportati le coordinate dei successivi posizionamenti effettuati in superficie, oltre che le misure di temperatura e salinità eventualmente raccolte.
- Metadata: archiviati in ASCII. Riguardano tutte le informazioni la boa e la struttura del ciclo adottato (profondità di drifting, tempo di risalita e discesa, ecc.).
- Dati tecnici: archiviati in NetCDF. Contengono tutti i parametri tecnici registrati durante la vita della boa.

In figura 2.4.1 è presentata la struttura gerarchica di archiviazione dei file per centro di raccolta.



Figura 2.4.1: Struttura gerarchica archivio dati Argo.

2.5 Argo nel Mediterraneo.

Nel contesto del progetto MFSTEP saranno dislocate complessivamente 20 boe Argo nel Mar Mediterraneo, delle quali 4 sono già operative al largo del Golfo del Leone. Questi strumenti si andranno ad aggiungere alle 10 boe Argo della Marina Militare degli Stati Uniti che sono tuttora presenti nel Mediterraneo orientale. Sia i dati Argo di MFSTEP che gli altri sono di pubblico accesso e vanno a costituire un unico sistema osservativo del Mar Mediterraneo.

La scelta di perseguire un programma Argo (MedArgo) da parte di MFSTEP è dettata dall'esigenza di avere profili di salinità e temperatura in modo automatico e in tempo reale su tutto il bacino per l'assimilazione nei modelli previsionali. Diversi lavori, tra i quali Molcard et al. (2003), hanno rimarcato l'importanza dello schema di campionamento che deve essere ottimizzato rispetto alle caratteristiche fisiche del bacino e alle finalità per cui le Argo sono applicate.

Un aspetto importante del ciclo di immersione/emersione è il tempo di deriva in profondità, T_{drift} . Molcard (2003) ha evidenziato la necessità di mantenere T_{drift} $< T_L$, dove T_L è il tempo caratteristico lagrangiano; per tempi superiori la componente di turbolenza lagrangiana copre interamente l'informazione sul campo medio di velocità, rendendo i dati di velocità non utilizzabili in un modello di circolazione generale. Il limite inferiore di T_{drift} è, invece, imposto da considerazioni economiche, in quanto una maggiore frequenza di campionamento corrisponde a una minore autonomia dello strumento, e dalla volontà di contenere l'errore di misura delle traiettorie.

L'incertezza legata alla misura delle traiettorie, e quindi delle velocità, è proporzionale a T_z , il tempo impiegato nei moti verticali di immersione ed emersione. Durante questa fase la boa continua a derivare con una velocità che non corrisponde alle correnti sottosuperficiali alla profondità di riferimento, introducendo quindi un errore nella loro stima. Griffa et al. (2003) ha stimato la grandezza di questo errore mediante simulazioni basate su dati climatologici di velocità; si sono riscontrati errori anche del 100% per T_{drift} di tre giorni. Poiché la velocità di immersione ed emersione dipende dalle caratteristiche tecniche dello strumento, le uniche possibilità di ridurre l'errore sulla stima delle velocità sottosuperficiali, alla profondità di riferimento, consistono in aumentare T_{drift} o nel dislocare le boe Argo in aree del bacino caratterizzate da forti correnti profonde: in questo modo il percorso della boa alla profondità di riferimento sarà abbastanza grande da ridurre l'influenza della deriva durante i moti verticali.

A seguito di questi studi si è deciso, nell'ambito di MFSTEP, di dislocare le boe Argo nelle aree del Mediterraneo dove sono presenti le correnti più importanti.. Le boe MedArgo eseguiranno cicli della durata di 5 giorni, con profondità di deriva fino a 350 m.. La profondità di deriva non corrisponde alla profondità da cui la boa comincia a misurare temperatura e salinità, ma la boa derivante esegue un ulteriore inabissamento fino a 700m, subito prima di iniziare la misura. Le differenze rispetto al ciclo applicato nell'oceano Atlantico sono motivate dal diverso tempo lagrangiano e dalla diversa topografia che caratterizza i due bacini.

I dati utilizzati nel corso di questa tesi riguardano i primi 6 mesi del 2003. In questo periodo erano attive nel Mediterraneo 8 boe Argo rilasciate dalla Marina Militare degli Stati Uniti. Non è stato possibile ottenere i 'metadata' e i file delle traiettorie, per cui l'intera analisi si è basata sui file dei profili di temperatura e salinità. Per questo motivo non si conoscono alcune specifiche del ciclo di immersione, in particolare la profondità di deriva. Ciò ha comportato alcune limitazioni nel contesto dell'analisi dei campi di velocità sottosuperfiale, come verrà illustrato nel Capitolo 4. In figura 2.5.1 è riportato il diagramma del ciclo delle 8 boe U.S. Navy che è stato possibile ricostruire a partire dai dati disponibili.



Figura 2.5.1: Specifiche del ciclo delle boe Argo della Marina Militare degli Stati Uniti che è stato possibile determinare.

2.6 I dati Argo nel Mediterraneo nel periodo gennaio-giugno 2003.

Il data set Argo impiegato nel corso di questa tesi consiste in 245 profili raccolti nel Mediterraneo orientale da 8 boe profilanti nel periodo da 01/01/2003 a 16/06/2003 (tabella 2.6.1).

Ad ogni profilo sono associate le seguenti informazioni:

- 30 misure di pressione, temperatura e salinità.
- latitudine e longitudine.
- data e ora.
- numero identificativo dello strumento.

Per quantificare l'errore della misure è necessario fare alcune considerazioni:
Numero di	primo profilo eseguito		ultimo profilo eseguito		Numero
identificazione della boa	data	ora	data	ora	profili eseguiti
1900025	01/01/2003	7.10	15/06/2003	7.11	32
1900026	01/01/2003	8.53	15/06/2003	8.35	34
1900027	01/01/2003	8.53	15/06/2003	8.51	34
6900087	03/01/2003	9.44	12/06/2003	9.41	32
6900088	03/01/2003	14.31	02/06/2003	11.24	29
6900089	03/01/2003	13.46	12/06/2003	13.59	32
6900090	05/01/2003	10.32	15/04/2003	10.19	19
6900091	03/01/2003	12.34	12/06/2003	12.54	33

Tabella 2.6.1: Boe profilanti Argo usate nell'analisi di questa tesi.

- I parametri fisici (pressione, temperatura, salinità) sono misurati durante la fase ascensionale del ciclo della boa, e non corrispondono esattamente ai dati di un ipotetico profilo verticale corrispondente al punto di emersione.
- La localizzazione spaziale (latitudine e longitudine) e temporale (data e ora) è relativa alla prima connessione satellitare effettuata in superficie e non al punto di emersione.

Ne consegue che l'errore di misura deve essere la somma dell'errore strumentale e dell'errore di localizzazione.

Usando le informazioni relative ai posizionamenti si sono ricostruiti i percorsi delle 8 boe Argo durante i sei mesi dell'analisi. I tratti colorati in figura 2.6.1 corrispondono ai percorsi effettuati nel mese di gennaio 2003. Per meglio illustrare la tipologia di dati utilizzati, in figura 2.6.2 sono riportate le misure di 2 boe Argo nel mese di gennaio.

La distribuzione verticale dei dati non è omogenea, ma presenta una maggior densità nei livelli superficiali; in figura 2.6.3 è illustrata la distribuzione verticale di tutti i 245 profili. Il numero limitato di misure verticali disponibili è il prodotto di un sottocampionamento, che normalmente viene applicato ai dati in modo da minimizzarne i tempi di trasmissione satellitare. Il sistema satellitare Argos attualmente impiegato ha una velocità di trasferimento pari a 0.2 bytes/sec. Come si vede da figura 2.6.3, le profondità^{*} a cui le misure corrispondono sono piuttosto costanti; la maggiore variabilità osservata è relativa alla prima ed ultima misura.

^{*} Per convertire i dati di pressione in misure di profondità si è applicata l'algoritmo UNESCO Tech. Pap. in Mar. Sci., No. 44, 55 pp.

Al fine di valutare la distribuzione spaziale dei dati, si è utilizzata la divisione del Mediterraneo applicata nel contesto di MFS. In figura 2.6.4 è mostrata la sovrapposizione tra le localizzazioni dei profili e la divisione in regioni di MFS.

I dati interessano le regioni 8, 11, 12 e 13. L'istogramma del numero totale di profili per regione è riportato in figura 2.6.5, mentre in figura 2.6.6 è riportata la distribuzione temporale dei dati nelle regioni.

Questi dati saranno utilizzati nei prossimi capitoli per un confronto con i risultati del modello, in modo da analizzare la qualità del modello e del sistema di assimilazione dei dati di altimetria.



Figura 2.6.1. Percorsi delle 8 boe derivanti Argo durante i 6 mesi dell'analisi.



Figura 2.6.2: profili T e S boa Argo 6900087 (a) e 6900089 (b).



Figura 2.6.3: Distribuzione verticale di tutte le misure effettuate delle 8 boe Argo nei primi sei mesi del 2003.



Figura 2.6.4: Distribuzione dei dati nelle regioni MFS 8, 11, 12 e 13.





Figura 2.6.5: Distribuzione temporale dei profili nelle regioni 8, 11, 12 e 13.

Capitolo 3.

3.1 Teoria della validazione.

La validazione del forecast è il processo di determinazione della qualità del sistema previsionale. Per quanto esistano diverse procedure, tutte richiedono una misura della relazione tra forecast e osservazione della variabile in esame. Ogni procedura, quindi, si basa necessariamente sul confronto tra singole coppie di valori, costituite dal risultato della previsione e la corrispettiva osservazione (Wilks, 1995).

Da un punto di vista generale, la validazione del forecast implica l'analisi della distribuzione condizionata di previsioni e osservazioni. Quindi, ogni data set di verifica consiste in un insieme di coppie osservazione/previsione; la distribuzione condizionata di queste coppie può essere caratterizzata in base alla frequenza relativa delle possibili combinazioni di forecast e osservazioni.

Si denoti il forecast con y_i , che può assumere ognuno degli I valori $y_1, y_2, ..., y_I$ e la corrispondente osservazione con o_j che a sua volta può assumere ognuno degli J valori $o_1, o_2, ..., o_J$. Allora la distribuzione condizionata di forecast e osservazioni è

$$p(y_{i}, o_{j}) = \Pr\{y_{i}, o_{j}\} = \Pr\{y_{i} \bigcap o_{j}\},$$

$$i = 1, ..., I; \ j = 1, ..., J.$$
(3.1.1)

Questa funzione di distribuzione della probabilità discreta bivariata associa una probabilità ad ognuna delle I x J possibili combinazioni di forecast e osservazioni. Perfino nel caso più semplice, per il quale I = J = 2, questa distribuzione condizionata può essere difficile da usare direttamente.

Usando la definizione di probabilità condizionata:

$$\Pr\{E_1 \mid E_2\} = \frac{\Pr\{E_1 \bigcap E_2\}}{\Pr\{E_2\}}$$
(3.1.2)

la distribuzione condizionata può essere scomposta in due modi che danno informazioni complementari sul problema di validazione.

Dal punto di vista del forecast il più intuitivo dei due è:

$$p(y_i, o_j) = p(o_j | y_i)p(y_i);$$

$$i = 1, ..., I; j = 1, ..., J$$
(3.1.3)

La prima parte di questa scomposizione consiste in un set di I distribuzioni condizionate, $p(o_j | y_i)$, ognuna delle quali rappresenta la probabilità di verificarsi di tutte le J osservazioni o_j dato un forecast y_i . In altre parole, la probabilità condizionata specifica quanto spesso un evento climatico si è verificato, quando la previsione y_i è stata prodotta. La parte restante è la distribuzione incondizionata di $p(y_i)$, che specifica la frequenza relativa di ognuna delle previsioni y_i ; in altre parole, quanto spesso ognuno dei I valori possibili di forecast è realizzato.

L'altra possibile scomposizione della probabilità condizionata di previsione e osservazione è:

$$p(y_i, o_j) = p(y_i | o_j) p(o_j);$$

$$i = 1, ..., I; j = 1, ...J$$
(3.1.4)

In questo caso la distribuzione condizionata $p(y_i | o_j)$ esprime la probabilità che ognuno dei valori possibili di forecast y_i sia previsto rispetto alla osservazione futura o_j . La distribuzione incondizionata di $p(o_j)$ consiste semplicemente nella frequenza relativa dei J eventi climatici o_j nel data set di validazione; questa probabilità è comunemente chiamata climatologia.

Per quanto queste due scomposizioni della probabilità condizionata di osservazioni e previsioni possano aiutare a organizzare concettualmente l'informazione della validazione, esse non riducono la dimensionalità, o gradi di libertà, del problema di validazione. Poiché la somma di tutte le distribuzioni condizionate deve essere 1, il problema è completamente specificato da ognuna delle $(I \times J) - 1$ forme di queste probabilità. Sebbene le scomposizioni (3.1.3) e (3.1.4) rappresentino diversamente l'informazione contenuta in (3.1.1), $(I \times J) - 1$ distinte probabilità sono comunque necessarie per specificare completamente il problema della qualità del forecast.

Perfino nel caso più semplice, I = J = 2, la determinazione completa del problema richiede una misura di validazione tridimensionale. Di conseguenza, è uso comune riassumere le capacità di un sistema previsionale mediante indici scalari. Per quanto questi indici possano rivelarsi utili, è inevitabile che parte dell'informazione contenuta nella distribuzione condizionata di previsioni e osservazioni venga scartata, visto che la dimensionalità del problema di verifica è stata ridotta.

3.2 Definizione attributi di qualità del forecast.

La qualità del forecast può quindi essere espressa mediante scalari che rappresentano attributi differenti del sistema previsionale. Molti di questi aspetti sono tra loro correlati.

Qui di seguito si riporta una lista di questi aspetti scalari.

- 1. Accuratezza: si riferisce alla corrispondenza media tra valori singoli di forecast e l'evento che predicono. Misure scalari dell'accuratezza riassumono in un singolo numero la qualità di una serie di previsioni
- 2. Bias: misura la corrispondenza tra le media del forecast e la media delle osservazioni. Fa riferimento ad errori sistematici che riguardano il sistema previsionale.
- 3. Affidabilità o Calibrazione: riguarda le relazioni tra il forecast e l'osservazione media corrispondente ad un valore specifico di forecast. Le misure di affidabilità raggruppano le coppie di osservazione/previsione in gruppi rispetto al valore della variabile di forecast, e caratterizzano la distribuzione condizionale dell'osservazione dato il forecast; $p(o_i | y_i)$.

- 4. Risoluzione: si riferisce alla misura in cui il forecast raggruppa gli eventi osservati in gruppi differenti gli uni dagli altri. E' correlata con l'affidabilità, in quanto anch'essa riguarda la probabilità condizionata dell'osservazione dato il forecast, $p(o_j | y_i)$. Nello specifico, la risoluzione valuta la differenza tra la media condizionata delle osservazioni rispetto a diversi valori di forecast, mentre l'affidabilità concerne la media condizionata stessa delle osservazioni rispetto a una previsione.
- 5. Discriminazione: è l'opposto della risoluzione. Le misure di discriminazione riassumono la distribuzione condizionata del forecast data l'osservazione, ossia $p(y_i | o_j)$. L'attributo di discriminazione riflette la capacità del sistema previsionale di produrre forecast diversi rispetto a variazioni nei valori osservati.
- 6. Raffinatezza (*Sharpness*): è una proprietà solo del forecast; fa riferimento alla distribuzione incondizionata $p(y_i)$. I sistemi previsionali che variano raramente dalle climatologie mostrano una scarsa raffinatezza. Come esempio estremo, un forecast costituito solamente da climatologie non esibirebbe alcun attributo di raffinatezza.

I test di validazione condotti in questa tesi si sono incentrati sul calcolo degli indici scalari di accuratezza, con la volontà di non scorporare le osservazioni dai loro gruppi fondamentali che sono costituiti dai singoli profili. In questo modo si è cercato di conservare l'informazione legata alla struttura verticale dei campi di temperatura e salinità che sono l'oggetto di questa analisi.

Si precisa inoltre che i dati del modello usati in questa sezione non sono riferiti propriamente alla previsione bensì ai dati di analisi, ovvero la stima migliore dello stato del sistema data dalla fusione delle osservazioni e il modello tramite lo schema di assimilazione. Si parlerà quindi di qualità del modello e non delle previsioni, mentre l'uso delle tecniche sviluppate in questa tesi per valutare la previsione vera e propria sarà fatto in un lavoro che seguirà questa tesi.

3.3 Misure scalari di Accuratezza.

Per stimare l'accuratezza del modello previsionale si è utilizzato l'indice scalare dell'errore quadratico medio o 'Mean Square Error' (MSE) definito come:

$$MSE = \frac{1}{N} \sum_{k=1}^{N} \left(misfit \right)^2$$
(3.3.1)

dove il misfit è la differenza tra l'osservazione e il corrispondente campo del forecast.

Si definisce il misfit come:

$$misfit = \vec{y}^{o}(x_r, y_r, z_r, t_r) - H\left(\vec{x}^{a}(x_m, y_m, z_m, \tilde{t}_r)\right); \qquad (3.3.2)$$

dove \vec{y}^{o} è il vettore delle osservazioni, costituito dal valore di temperatura e di salinità registrato nel punto dello spazio r, \vec{x}^{a} è il corrispondente vettore della analisi e H() è un operatore linearizzabile; nel caso in esame H() è semplicemente un operatore di interpolazione.

Dato che la distribuzione spaziale e temporale delle osservazioni è discordante rispetto a quella del modello, si è usato un operatore di interpolazione per portare il risultato del forecast nel punto spazio-temporale in cui l'osservazione è stata raccolta. I risultati del modello usati sono medie giornaliere mentre i dati sono riferiti all'ora del giorno alla quale il profilo è stato raccolto. La discrepanza temporale tra osservazione e previsione è stata assunta come non significativa in quanto un intervallo temporale dell'ordine di grandezza massima di un giorno è al di sotto della risoluzione del nostro sistema, ovvero dei processi rappresentati dal modello.

Nel nostro caso, H() è rappresentato da una interpolazione bidimensionale cubica per l'orizzontale, basata sui 16 punti griglia più vicini alla posizione della osservazione, e una interpolazione unidimensionale lineare per la verticale. E' importante sottolineare che per l'orizzontale i dati del modello sono stati portati sulla posizione dei profili; mentre l'interpolazione verticale ha riguardato sia le osservazioni che i dati del modello, in quando le boe profilanti hanno intervalli di acquisizione dei dati non regolari. Tuttavia, la scala verticale di interpolazione è stata scelta in modo da privilegiare la distribuzione verticale delle osservazioni, che mostra un leggera oscillazione attorno ad un andamento regolare (figura 3.3.1). In questo modo per ogni profilo Argo è stato stimato il profilo interpolato dai dati del modello, riducendo così il calcolo del misfit a:

$$misfit = \vec{y}^{o}(x_{r}, y_{r}, z_{r}, t_{r}) - \vec{x}^{int\,erp}(x_{r}, y_{r}, z_{r}, t_{r}); \qquad (3.3.3)$$

Seguendo la scomposizione algebrica di Murphy (1995) è possibile riscrivere (3.3.1) nel seguente modo:

$$MSE(\vec{x}, \vec{y}) = (\vec{y} - \vec{x})^2 + s_y^2 + s_x^2 - 2s_y s_x r_{yx}; \qquad (3.3.4)$$

dove $(\overline{y} - \overline{x})$ è il bias del forecast, s_y e s_x la deviazione standard della distribuzione marginale delle osservazioni e delle previsioni e r_{yx} è la correlazione tra osservazioni e forecast del campione in esame. In (3.3.4) la barra indica la somma sul numero di realizzazioni dei dati e del modello a disposizione.

Questa scomposizione è di grande utilità in quanto permette di separare le componenti dell' MSE, individuando da un lato il bias e dall'altro evidenziando il contributo della deviazione standard e del coefficiente di correlazione. E' facile vedere che con $s_y = s_x$, $r_{yx} = 1$ riduce il MSE al semplice bias. In altre parole, un alto coefficiente di correlazione riduce il MSE.

Al di là delle pure implicazioni statistiche, il coefficiente di correlazione r_{yx} assume una grande importanza fisica se il MSE viene calcolato profilo per profilo. Per meglio valutare questo punto è bene considerare il comportamento del coefficiente di correlazione che è definito come segue:

$$r_{yx} = \frac{\text{cov}(y, x)}{s_{y}s_{x}} = \frac{\frac{1}{n}\sum_{k=1}^{n} \left[\left(y_{k} - \overline{y} \right) \left(x_{k} - \overline{x} \right) \right]}{\left\{ \frac{1}{n}\sum_{k=1}^{n} \left(y_{k} - \overline{y} \right)^{2} \right\}^{\frac{1}{2}} \left\{ \frac{1}{n}\sum_{k=1}^{n} \left(x_{k} - \overline{x} \right)^{2} \right\}^{\frac{1}{2}}}$$
(3.3.5)

Questo coefficiente è altresì chiamato Anomaly Correlation coefficient (AC) ed è una stima della correlazione lineare di osservazioni e previsioni. Nel caso in esame, ossia confrontando profili verticali di temperatura e salinità è interessante valutarne il grado di coerenza. In altre parole, questo indice dà la possibilità di stimare quanto il modello riconosca la struttura verticale delle differenti masse d'acqua individuate dalle boe profilanti. L'ultimo indice preso in considerazione è la deviazione standard dell'errore o Root Mean Squared error (RMS) definito come:

$$RMS = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(misfit \right)^2}; \qquad (3.3.6)$$

Il tipo di informazione contenuta in questo indice non differisce da quella nel MSE, ma ha il pregio di essere di facile comprensione: l'unità di misura del RMS è la stessa della parametro in esame.



Figura 3.3.1: Le linee tratteggiate si riferiscono ai livelli di interpolazione, i punti blu indicano tutte le profondità di misura delle 8 boe Argo nei sei mesi studiati.

3.4 Calcolo di Funzioni Empiriche Ortogonali- Empirical Orthogonal Functions- EOF.

La distribuzione della varianza nei dati delle Argo e del modello può essere stimata mediante EOF delle osservazioni e dei dati del modello. Poiché questa procedura non implica alcun confronto diretto tra coppie di osservazione e previsione, non può essere formalmente considerata una verifica del sistema previsionale. Tuttavia questa analisi permette di ridurre la dimensionalità del problema di confronto, fornendo un'immagine sintetica delle strutture verticali di temperatura e salinità e associandone il peso rispetto alla varianza complessiva.

Le EOF sono normalmente applicate allo studio delle serie temporali, il metodo è anche chiamato analisi delle componenti principali (Morisson, 1976); questa tecnica è ampiamente utilizzata nella meteorologia ed oceanografia in quanto permette di ridurre l'ordine del problema, concentrando l'informazione in un piccolo numero di EOF (Lorenz, 1956).

Per semplicità viene prima descritto il caso del calcolo di EOF univariate per la salinità.

Ad ogni profilo è stato associato il vettore di stato \vec{x} di *m* elementi, dove *m* sono i 29 livelli usati per l'interpolazione verticale. Formalmente \vec{x} è dato da:

$$\vec{x}^{T} = [x_1, ..., x_m] = \left[\frac{\delta S_1}{\sigma_1}, ..., \frac{\delta S_m}{\sigma_m}\right];$$
(3.4.1)

dove $\delta S_i = (S_i - \overline{S}_i)$, i=1...m, sono le anomalie della variabile generica S dalla media di ogni livello; le anomalie vengono poi normalizzate per i corrispondenti valori di deviazione standard σ_i , definiti come:

$$\sigma_{i} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (S_{i} - \overline{S_{i}})_{j}^{2}}; \qquad (3.4.2)$$

n è il numero di misure ad ogni livello.

La normalizzazione per i valori di deviazione standard ad ogni livello di profondità è resa necessaria dall'esigenza di contenere la maggiore variabilità nei livelli superficiali dell'oceano, in modo da preservare il segnale derivante dai processi profondi che risulterebbe altrimenti coperto.

Le EOF verticali sono quindi definite dalla matrice di correlazione del vettore di stato \vec{x} , scritta come:

$$C = \frac{1}{n} X^{T} X ; (3.4.3)$$

dove,

$$X = \begin{vmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & \dots & \dots & \vdots \\ \dots & \dots & \dots & \dots & \vdots \\ x_{m,1} & \dots & \dots & x_{m,n} \end{vmatrix};$$
(3.4.4)

è composto dai singoli vettori \vec{x} ai differenti tempi *n* e quindi *C* apparirà in questo modo,

$$C = \begin{vmatrix} 1 & \frac{1}{n} \sum_{j=1}^{n} (x_{2}x_{1})_{j} & \dots & \frac{1}{n} \sum_{j=1}^{n} (x_{m}x_{1})_{j} \\ \frac{1}{n} \sum_{j=1}^{n} (x_{1}x_{2})_{j} & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{n} \sum_{j=1}^{n} (x_{1}x_{m})_{j} & \dots & \dots & 1 \end{vmatrix};$$
(3.4.5)

La matrice di correlazione è definita simmetrica e non negativa, quindi possiede una serie di autovalori positivi i cui autovettori sono ortogonali; questi autovettori sono appunto le EOF della matrice costituita dai vettori di stato. In altre parole:

$$C = U\Lambda U^{T}$$

dove U contiene gli autovettori e Λ gli autovalori corrispondenti a ciascun autovettore.

Una indicazione della varianza spiegata da una singola EOF è data dall'autovalore associato. Pertanto si può calcolare il parametro di Percentuale di Varianza Spiegata (PVE) in modo da associare ad ogni EOF il suo peso. Questo indice è espresso come,

$$PVE_{i} = \frac{\lambda_{i}}{\sum_{k=1}^{m} \lambda_{k}} \cdot 100; \qquad (3.4.6)$$

Poiché si prendono gli autovalori più grandi, i modi corrispondenti raccolgono la maggior parte di varianza nella colonna d'acqua (Morrison, 1976)

Le EOF bivariate sono calcolate in maniera analoga, di seguito è riportata una breve descrizione del procedimento applicato.

Il vettore di stato \vec{x} per il calcolo delle EOF bivariate viene definito come:

$$\vec{x}^{T} = \begin{bmatrix} t_1, \dots, t_m, s_1, \dots, s_m \end{bmatrix} = \begin{bmatrix} \frac{\delta T_1}{\sigma_1}, \dots, \frac{\delta T_m}{\sigma_m}, \frac{\delta S_1}{\sigma_1}, \dots, \frac{\delta S_m}{\sigma_m} \end{bmatrix}; \quad (3.4.7)$$

dove m=1, ..., M sono ancora i livelli verticali e questa volta T indica la temperatura e S la salinità.

In forma sintetica, il vettore di stato X per tutti i tempi n=1,...,N è definito:

$$X = \begin{vmatrix} T \\ S \end{vmatrix} \tag{3.4.8}$$

dove:

$$T = \begin{vmatrix} t_{1,1} & t_{1,2} & \dots & t_{1,n} \\ t_{2,1} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ t_{m,1} & \dots & \dots & t_{m,n} \end{vmatrix}$$
(3.4.9)

.

$$S = \begin{vmatrix} s_{1,1} & s_{1,2} & \dots & s_{1,n} \\ s_{2,1} & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{m,1} & \dots & \dots & s_{m,n} \end{vmatrix}$$
(3.4.10)

Si ha quindi che la matrice di correlazione è uguale a:

$$C = \frac{1}{n} X^{T} X = \frac{1}{n} |T \quad S| \cdot \begin{vmatrix} T \\ S \end{vmatrix} = \frac{1}{n} \begin{vmatrix} T \cdot T & S \cdot T \\ T \cdot S & S \cdot S \end{vmatrix}$$
(3.4.11)

La matrice *C* può essere idealmente divisa in quattro quadranti: il primo rappresenta la variabilità temporale della sola temperatura, il secondo e il terzo la variabilità combinata di T e S e il quarto la variabilità temperale della salinità. Poiché *C* ha dimensioni $2m \times 2m$, ogni singola EOF sarà costituita da 2m elementi. Si è considerata la prima metà del vettore come EOF della temperatura e la seconda metà come EOF della salinità.

3.5 Risultati

I risultati sono prima mostrati in maniera qualitativa e in seguito sono mostrati gli indici di accuratezza e le EOF.

3.5.1 Confronto qualitativo tra osservazioni e modello

Un quadro d'insieme delle osservazioni, e separatamente del modello, è presentato in figura 3.5.1. La figura mostra il generale andamento dei profili di salinità e temperatura ottenuti dalle boe derivanti Argo e dal modello. Da questa

prima figura è già possibile inquadrare alcuni aspetti fondamentali del confronto e della qualità del modello.

I profili di temperatura del modello sembrano riprodurre il generale andamento dei profili Argo; in particolare le medie totali, calcolate per ogni livello di profondità su i tutti i mesi dell'analisi, delineano una comune struttura verticale di temperatura. Il caso della salinità appare invece più complesso; l'andamento generale dei profili Argo è abbastanza omogeneo, mentre non lo è l'andamento dei profili del modello. Alcuni di questi profili presentano un minimo di salinità a 200 metri di profondità che non è registrato dalle Argo e che apparentemente interessa solo una parte dei dati interpolati dal modello.



Figura 3.5.1: Raffigurazione d'insieme dei 245 profili Argo e dei rispettivi profili del modello. La linea rossa indica la media.



Figura 3.5.2: Diagramma Temperatura–Salinità per i dati delle Argo e del modello.

Un'informazione analoga può essere ricavata dalla figura 3.5.2: la nube di dati del modello è spostata verso valori a più bassa salinità. In particolare si può notare che la maggior parte di questi dati è confinata nell'intervallo di temperatura 19 – 14 °C. Questo intervallo generalmente corrisponde alle temperature registrate attorno ai 200 metri di profondità.

3.5.2 Indici di accuratezza

L'analisi del misfit tra i dati del modello e delle Argo si è basata sul calcolo degli indici, già descritti, cercando di separare il segnale legato al tempo e allo spazio. Si sono quindi calcolati gli indici per ogni profilo, per le medie settimanali, per le medie settimanali nelle regioni considerate, e infine si è cercato di individuare la componente legata alla profondità nella colonna d'acqua. Gli indici MSE, AC, Bias e RMS sono sempre presentati in un'unica figura basata su

una struttura comune a tutti i casi elencati, in modo da facilitarne la comprensione.



Figura 3.5.3: Statistiche di misfit relative a tutti i profili esaminati.

Temperatura

In figura 3.5.3 sono riportate le statistiche calcolate su tutti i profili di temperatura disponibili. Il segnale settimanale è calcolato in due modi: in base alla media dei dati (media settimanale) e considerando come unico profilo tutte le misure raccolte in una settimana (cumulativo settimanale). Questa seconda procedura permette di aumentare i gradi di libertà dell'analisi e fornisce valori sistematicamente "peggiori" rispetto alle statistiche calcolate sulle medie settimanali.

L'andamento complesso del segnale giornaliero è il prodotto dalla dislocazione spaziale disomogenea delle boe, per cui valori vicini nel tempo si riferiscono a profili registrati in regioni differenti del Mediterraneo; mentre la presenza di discontinuità è dovuta all'assenza di alcuni dati nel profilo, si è deciso di non interpolare i valori mancanti per evidenziare la disomogeneità del campione.

In generale si osserva che il MSE (e quindi RMS) registra i valori maggiori durante il mese di gennaio, il RMS assume valori attorno a 0.8 – 1 °C, per poi diminuire nei mesi centrali dell'analisi: tra febbraio e maggio il RMS è di 0.5–0.6 °C. Un ulteriore aumento è registrato tra maggio e giugno. Facendo riferimento a (3.3.4) si nota che la componente maggiore dell'errore è introdotta dal bias. Il bias non è costante, ma è ben individuabile una tendenza crescente lungo l'analisi; il modello sovrastima la temperatura nei mesi invernali e la sottostima nei mesi primaverili.

L'Anomalia di Correlazione è generalmente superiore a 0.8, ossia il modello riproduce bene l'andamento della struttura verticale di temperatura. E' significativo che i minimi vengano registrati tra marzo e aprile; in questo periodo la colonna d'acqua subisce veloci mutamenti della stratificazione interna che sono riprodotti dal modello con maggiori difficoltà.

In figura 3.5.4 è mostrata la scomposizione del segnale settimanale, non mediato, nelle componenti regionali. Gli indici statistici calcolati nelle regioni 11, 12 e 13 mostrano un andamento abbastanza simile, mentre la regione 8 è caratterizzata da RMS più bassi, dell'ordine di 0.5 °C fino ad aprile.

Un'ulteriore scomposizione del segnale settimanale è presentata nelle figure 3.5.5 e 3.5.6. La variabilità dell'errore nello strato superficiale (intendendo i primi 200 metri) è maggiore che nei livelli intermedi, ma l'ampiezza del RMS è generalmente paragonabile (minore tra febbraio e aprile).



Figura 3.5.4: Statistiche di misfit per le regioni 8, 11, 12, 13.



Figura 3.5.5; Statistiche di misfit per le regioni 8, 11, 12 e 13 relative alla strato superficiale della colonna d'acqua.



Figura 3.5.6 Statistiche di misfit per le regioni 8, 11, 12 e 13 relative allo strato intermedio della colonna d'acqua.

In figura 3.5.6 si nota che il RMS nello strato intermedio è generalmente elevato, intorno a 1° C; anche in questo caso si di osserva una minore entità del misfit nella regione 8.

I dati presentati mostrano una generale congruenza con i risultati di analisi precedenti, in particolare con il lavoro di Demirov et al. (2003). In questo articolo sono stati confrontati i dati di temperatura da XBT, non ancora assimilati, con le previsioni. L'analisi di Demirov si riferisce a un periodo in cui era attiva l'assimilazione sia di SLA che di XBT, mentre nel periodo considerato in questa tesi, lo schema di assimilazione è basato solamente sui dati di SLA.

L'errore commesso nello strato intermedio (200–600 m) senza assimilazione di XBT risulta essere nell'ordine di 1° C, mentre nel periodo in cui gli XBT erano assimilati l'errore nello strato intermedio (400 m) era generalmente contenuto in 0.5° C. Tuttavia la distribuzione spaziale e temporale dei dati utilizzati nelle due analisi è diversa e pertanto parte delle differenze osservate sono da ricondurre al diverso campionamento applicato.

<u>Salinità</u>

Le statistiche di misfit della salinità appaiono più complesse rispetto al caso della temperatura, come era già stato possibile intuire dalla figura 3.5.1. I due aspetti fondamentali che si evincono dalla figura 3.5.7 sono: valori di RMS generalmente elevati e AC generalmente negativa.

Nel periodo tra gennaio e febbraio si osservano valori di RMS attorno 0.4 psu. Nello stesso periodo si rilevano valori negativi di AC, prossimi a -1, che testimoniano la sostanziale differenza tra la struttura verticale di salinità del modello e quella rilevata dalle Argo.

Dal mese di marzo si osserva una diminuzione dell'entità del misfit che si attesta su valori intorno a 0.2 psu di RMS; in corrispondenza a questo andamento si osserva un generale aumento dell'AC e una diminuzione del bias.

L'andamento irregolare dell'Anomalia di Correlazione (AC) è dovuto alla disomogeneità spaziale dell'indice. Come si vede in figura 3.5.8, nella regione 8 il modello riesce a riprodurre strutture di salinità abbastanza affini ai dati Argo, contrariamente a quanto non faccia nelle regioni 11, 12 e 13. Tuttavia il RMS calcolato per i dati della regione 8 non indica una migliore accuratezza del modello in questa regione.

In figura 3.5.9 e 3.5.10 sono mostrati gli indici relativi allo strato superficiale e profondo della colonna d'acqua.

Nei primi 200 m della colonna d'acqua il modello commette l'errore maggiore, nei mesi invernali il RMS è di 0.6 psu; l'Anomalia di Correlazione mostra un andamento completamente irregolare, indicando la maggiore variabilità associata a questo strato, che a differenza della temperatura prevarica i limiti del cambio di stagione.

La salinità del modello nello strato intermedio mostra invece una sistematica anti-correlazione con la i dati Argo nei mesi di gennaio e febbraio che riguarda tutte le 4 regioni interessate dalla analisi: il RMS assume valori, intorno a 0.2-0.4 psu, che sono particolarmente elevati per la profondità considerata.

Le discrepanza osservate tra i dati di salinità delle Argo e del modello indicano la presenza di errori sistematici e sostanziali che riguardano il sistema previsionale.

Per quanto gli indici statistici siano affetti dall'errore di misura delle osservazioni (vedi Capitolo 2), è evidente che l'errore del modello non possa essere spiegato da questa incertezza, se non in minima parte.



Figura 3.5.7; Statistiche di misfit relative a tutti i profili esaminati.



Figura 3.5.8: Statistiche di misfit per le regioni 8, 11, 12 e 13.



Figure 3.5.9; Statistiche per le regioni 8, 11, 12 e 13 per lo strato superficiale della colonna d'acqua.



Figura 3.5.10: Statistiche per le regioni 8, 11, 12 e 13 per lo strato intermedio della colonna d'acqua.

3.5.3 Confronto osservazioni-modello tramite EOF

Data la natura lagrangiana delle misure Argo non si è ritenuto opportuno utilizzare il data set iniziale per il calcolo delle EOF, in quanto le misure delle boe non sono linearmente indipendenti entro il limite del tempo caratteristico lagrangiano; due misure consecutive di una stessa boa sono separate da 5 giorni, assumendo $T_L=7$ giorni (nello strato sottosuperficiale) si ha che il data set è costituito da un insieme di coppie di profili le cui misure non sono indipendenti. I calcoli delle EOF si sono quindi basati su delle super-osservazioni costituite dalle medie settimanali dei dati.

Per aiutare la comprensione e l'interpretazione dei risultati si è riportata la forma dimensionale dei dati, che è stata ottenuta rimoltiplicando le EOF per la deviazione standard che è il fattore di normalizzazione, si veda (3.4.1). Per ogni figura sono riportati i primi 3 modi e le corrispettive percentuali di varianza spiegata.

In figura 3.6.1-4 sono riportate le EOF univariate e bivariate di temperatura e salinità per i dati Argo e per i corrispondenti dati del modello. Le EOF permettono di valutare in modo sintetico come sia distribuita la varianza nel tempo nei due campioni di dati. Tuttavia poiché si sono utilizzate le medie interne dei dati per il calcolo delle anomalie (3.4.1), i risultati ottenuti non contengono alcuna informazione riguardo il bias dei due data set rispetto ad un valore comune.

Le EOF di temperatura calcolate dai dati Argo mostrano un massimo di variabilità confinato negli strati superficiali della colonna d'acqua. La superficie possiede valori di deviazione standard elevati poiché risente maggiormente, rispetto ai livelli profondi, della variabilità del ciclo stagionale. E' dalla superficie che l'energia viene trasferita all'oceano tramite i flussi di calore, acqua e momento. Nel caso della salinità a questi fenomeni se ne aggiungono altri legati ai processi avvettivi lungo la colonna d'acqua; le EOF di salinità (Argo) mostrano un leggero massimo sottosuperficiale che è il prodotto dell'avvezione di Acqua Levantina Intermedia (LIW). Il minimo superficiale che si osserva nella prima EOF di salinità è legato alla presenza di Acqua Atlantica Modificata (MAW) che sembra interessare i primi 100 m della colonna d'acqua. Tuttavia è importante attenersi a considerazioni il più possibile generali, in quanto il

campione di dati utilizzato è estremamente eterogeneo ed è strutturato in modo da non poter separare il segnale temporale da quello spaziale. La variabilità osservata è senz'altro prodotta, almeno in parte, da cicli stagionali, ma una componente altrettanto importante riguarda la dislocazione spaziale del data set. La varianza osservata è anche il prodotto della disomogeneità a scala di sotto bacino dei campi di temperatura e salinità. Lo scopo di questa analisi è quindi di permettere un confronto qualitativo con i dati del modello che dovrebbero risentire di simili variazioni spaziali e temporali.

Il confronto tra e EOF di salinità del modello e dei dati Argo riassume bene i risultati già osservati. La presenza di un minimo sottosuperficiale di salinità è chiara. Il modello riproduce una struttura verticale di salinità che è inconsistente con i dati Argo e che riguarda tutto il bacino orientale. La prima EOF di salinità del modello contiene più del 90% della varianza complessiva lasciando presagire la presenza di un errore sistematico.

Nel periodo considerato l'assimilazione dei dati ha riguardato soltanto la SLA; per cui le correzioni di temperatura e salinità sono state introdotte mediante una EOF trivariata comune per tutto il Mediterraneo e per tutto l'anno (Capitolo 1). L'efficacia di questo schema di assimilazione non è mai stata testata per la salinità nel Mediterraneo orientale e, data la forte semplificazione introdotta dall'utilizzo di EOF comuni per tutto il bacino, è plausibile che in questo modo si introduca una parte importante dell'errore osservato.

Le EOF hanno una forte influenza sul sistema di assimilazione SOFA in quanto costituiscono l'unica possibilità di evolvere nel tempo e nello spazio la matrice di covarianza dell'errore del forecast e, nel caso specifico di assimilazione di dati superficiali, svolgono la funzione di proiettare la correzione sulla verticale (De Mey et al., 2002).

La temperatura risente dello stesso problema, tuttavia questo è meno evidente in quanto le differenze nella stratificazione verticale di temperatura tra le diverse regioni del Mediterraneo sono contenute; per cui l'introduzione di una EOF comune influisce sulla semplificazione di strutture secondarie del profilo di temperatura.

Un ulteriore aspetto del problema è costituito dalle intercorrelazioni tra i diversi parametri. L'assimilazione di SLA utilizza EOF trivariate che esprimono le correlazioni di T, S e Ψ con sé stesse e le intercorrelazioni reciproche. Applicare un'unica EOF a tutto il bacino significa assumere che anche le intercorrelazioni siano costanti nel tempo e nello spazio; questa è un'assunzione forte e può contribuire all'errore osservato.







Figura 3.5.2: EOF univariate di temperatura e salinità dati Modello.



Figura 3.5.3: EOF bivariate di Temperatura e Salinità dati Argo.



Figura 3.5.4: EOF bivariate di Temperatura e Salinità dati Modello.

Capitolo 4

4.1 Confronto qualitativo tra i dati Argo e le analisi di MFS.

Si sono utilizzate le coordinate e le date associate ad ogni profilo per calcolare i campi di velocità lagrangiana tra due punti d'emersione. Queste velocità rendono conto dello spostamento medio orizzontale della boa. L'assenza di dati di traiettorie superficiali (vedi Capitolo 2) non ha infatti permesso di separare gli spostamenti effettuati in profondità da quelli in emersione, per cui ogni velocità calcolata è la media delle velocità con cui la boa ha derivato in superficie, in profondità, e durante gli spostamenti verticali di immersione ed emersione (vedi figura 2.5.1).

Questo problema ha fortemente condizionato l'analisi svolta, e in congiunzione con la mancanza delle informazioni 'metadata', ha reso necessaria la formulazione di alcune schematizzazioni. Innanzitutto si è assunto che le boe abbiano derivato alla profondità di 600 m, poiché tutti i profili di temperatura e salinità partono approssimativamente da questa profondità. E' possibile che le boe abbiano derivato ad una profondità minore per poi inabissarsi ulteriormente prima di cominciare ogni profilo di misura (come mostrato in Figura 2.4.2), ma l'assenza di informazioni rende opportuno attenersi alla possibilità più semplice.

Un'altra incertezza è legata al tempo di emersione: ragionevolmente la boa impiega 12 ore per effettuare la risalita, la deriva superficiale e la successiva discesa in profondità (Carval, 2003). Quindi il 10% del tempo di un ciclo completo (5 giorni) è impiegato per la misura dei profili di temperatura e salinità e
per la connessione satellitare. Nel contesto di questa analisi preliminare si è trascurato questo fattore e si è proceduto a confrontare le velocità e i percorsi delle boe con i campi delle velocità euleriane del modello alla profondità di 600 m.

Nel tentativo di effettuare questa prima analisi qualitativa, per ogni mese considerato si sono calcolate le medie mensili dei campi di velocità del modello. Anche questa è una forte schematizzazione del processo di confronto tra boe e modello che contribuirà a produrre differenze sostanziali tra il modello e i dati delle boe. Il confronto avviene quindi tra le strutture costanti di media mensile di circolazione del modello e i campi di velocità euleriane dedotte dalle boe lagrangiane.

Il calcolo della velocità euleriana delle boe è fatto in base alla semplice formula:

$$v_{m}^{o}(n) = \frac{\Delta r_{m}^{o}}{\Delta t} = \frac{r_{m}^{o}(n) - r_{m}^{o}(n-1)}{\Delta t}$$
(4.1.1)

dove m=1,...,M rappresenta il numero delle boe derivanti (8 nel caso in esame), n=1,...,N é il numero dei cicli effettuati della durata Δt (5 giorni) e $r_m(n)$ è il vettore posizione della boa derivante *m* al ciclo numero *n*. I vettori $v_m(n)$ sono riportati nelle figure alle posizioni $r_m(n-1)$.

In figura 4.1.1-6 sono mostrati i risultati ottenuti per i mesi da gennaio a giugno 2003. Dalle figure si evince che:

- La boa 6900087 rimane intrappolata nell'anticiclone del Peloponneso per tutti i sei mesi analizzati; il modello riesce grosso modo a individuare questa componente di sotto bacino della circolazione, tuttavia si osserva una maggiore variabilità nei dati del modello che non nel percorso osservato della boa.
- Il percorso della boa 6900089 nei mesi invernali risulta particolarmente consistente con i campi di velocità del modello che ben riproduce la corrente dell'Asia Minore; il percorso della boa è ben simulato anche lungo il Giro di Rodi che la boa inizia a seguire in marzo.
- La boa 1900025 è collocata approssimativamente nella posizione climatologica del 'Mid Mediterranean Jet' (vedi Figura 1.2.2) a sud del Giro Rodi, qui si osserva una maggiore imprecisione del modello che nei mesi invernali colloca questa corrente più a sud, verso le coste meridionali del Levantino, rispetto al percorso della boa.
- La boa 6900091 ha derivato seguendo il 'Mid Mediterranean Jet' fino in prossimità delle coste di Cipro, dove ha cominciato a muoversi in

concomitanza con una corrente occidentale di sottocosta. Le medie mensili dei campi di velocità del modello sono abbastanza consistenti con lo spostamento osservato, in particolare sono ben riprodotti i movimenti in prossimità delle coste di Cipro.

- La boa 1900026 è stata interessata prevalentemente da circolazioni instabili di mesoscala collocate nel centro del bacino Levantino. Le medie mensili dei campi di velocità del modello non riproducono queste strutture, se non nei mesi maggio e giugno quanto la boa comincia a derivare in prossimità delle coste Medio Orientali.
- La boa 1900027 deriva in prossimità della boa 1900026. Come nel caso precedente, si osserva una maggiore consistenza dei campi di velocità del modello in prossimità delle coste Medio Orientali.
- Le boe 6900088 e 6900090 derivano in una zona che è ai bordi della griglia del modello. In questa zona le analisi del modello dei campi di velocità non sono attendibili.

In generale si osserva un sostanziale consistenza tra le strutture generali di circolazione individuate dalle boe e dal modello. In particolar modo, trovano conferma i caratteristici giri e jet a scala di sotto bacino che formano la circolazione generale a scala di bacino (Robinson et al., 2001). Queste componenti sono caratterizzate da una elevata energia cinetica, si osservano correnti di 10 cm/s alla profondità di 600 m, e da una importante variabilità. L'individuazione di queste strutture da parte del modello numerico è evidentemente complessa e i risultati ottenuti sono incoraggianti.

4.2 Confronto quantitativo.

Lo scopo di questa seconda analisi è di valutare l'andamento quantitativo dei campi di velocità del modello rispetto alle di traiettorie lagrangiane o di deriva delle boe Argo. Allo scopo di quantificare l'errore commesso si è reso necessario introdurre una simulazione lagrangiana delle traiettorie consistenti con i campi di velocità del modello. Chiameremo queste le 'boe modello'.

Per calcolare le traiettorie lagrangiane delle 'boe modello' consistenti con i campi di velocità euleriani si è usato un semplice modello teorico. Innanzitutto i campi di velocità del modello sono stati interpolati sulla posizione di rilevamento della boa alla superficie, usando i quattro punti griglia più vicini alla posizione della boa Argo. Si è proceduto quindi al successivo calcolo dello spostamento ipotetico, o traiettoria, prodotto dalla velocità interpolata inizialmente sulla posizione della boa reale. Se $r_m(t)$ è il vettore delle coordinate della boa modello m al tempo t, la posizione $r_m(t + \Delta t)$ è calcolata come:

$$r_m(t + \Delta t) = r_m(t) + H_t(u_b)\Delta t; \qquad (4.2.1)$$

dove H_t () è l'operatore di interpolazione dei campi di velocità del modello, \vec{u}_b , al tempo t, e $\Delta t = 12$ ore. Il metodo applicato è pseudo-lagrangiano in quanto non si è tenuto conto della diversa natura delle velocità del modello e delle Argo. La profondità di riferimento utilizzata è di 600 m.

La simulazione delle traiettorie delle boe è stata condotta per cicli singoli della durata di 5 giorni, ossia ogni 5 giorni la posizione iniziale della 'boa modello' veniva riportata esattamente nella posizione data dal successivo posizionamento delle Argo. In simboli, al tempo iniziale t_0 , $\vec{r}_m^{\ b}(t=t_0) \equiv \vec{r}_m^{\ o}$ dove il termine a destra dell'uguale è la posizione osservata. Si è quindi impedito che la simulazione della traiettoria della 'boa modello' si allontanasse troppo dalle traiettorie osservate. Va osservato che non si tratta di una serie continua di simulazioni, bensì di singole simulazioni scollegate una dall'altra della durata uguale a un ciclo di immersione-emersione della boa Argo.

Una formulazione più consistente del modello di simulazione delle traiettorie dovrebbe includere una parametrizzazione della turbolenza lagrangiana. In generale si ha infatti che la velocità lagrangiana \overline{u}_{lagr} può essere scritta come:

$$u_{lagr} = u_{eu} + u' \tag{4.2.2}$$

dove \overline{u}_{eu} è la velocità media euleriana del modello. La differenza tra i due campi, u', rappresenta la fluttuazione attorno alla media indotta dalla turbolenza lagrangiana che è conosciuta solo in alcune regione dell'oceano e difficilmente rappresentabile in maniera analitica. Questo problema può essere aggirato minimizzando l'importanza della turbolenza lagrangiana tramite l'utilizzo di Δt piccoli rispetto al tempo integrale di scala lagrangiano T_L (tipicamente dell'ordine di 1-3 giorni per l'oceano superficiale e 7-10 giorni per l'oceano profondo) (Griffa, 2003). Tuttavia Δt non può essere preso piccolo a piacere in quanto dipende dalla risoluzione spaziale del modello; deve essere un tempo abbastanza lungo da permettere alla boa derivante di spostarsi di una distanza "sensibile" rispetto alla griglia spaziale del modello utilizzato. Assumendo una velocità media di 10 cm/s si ha uno spostamento di mezzo punto griglia nel corso di 12 ore, nel caso del modello usato in questa tesi che ha una griglia a 1/8 di grado di risoluzione in latitudine e longitudine. Si è scelto quindi di prendere \vec{u} 'uguale a zero e Δt uguale a 12 ore.

Infine, le traiettorie delle 'boe modello' sono state usate per ricalcolare le velocità medie pseudo-lagrangiane e quindi confrontarle con le velocità delle boe calcolate in precedenza. Le velocità sono così definite:

$$\vec{v}_{m}^{b}(n) = \frac{\Delta \vec{r}_{m}}{\Delta t} = \frac{\vec{r}_{m}^{b}(n) - \vec{r}_{m}^{b}(n-1)}{\Delta t}$$
(4.2.3)

$$\vec{v}_{m}^{o}(n) = \frac{\vec{\Delta r}_{m}}{\Delta t} = \frac{\vec{r}_{m}^{o}(n) - \vec{r}_{m}^{o}(n-1)}{\Delta t}$$
(4.2.4)

dove, in questo caso, Δt è l'intervallo di tempo tra due localizzazioni successive di una boa Argo ed è il tempo complessivo di una simulazione delle traiettorie, ovvero 5 giorni.

In figura 4.2.1-3 sono riportati i risultati dell'analisi. I risultati sono divisi in base alle singole boe, in questo modo si è voluta conservare l'informazione dei tracciati percorsi.

Sono stati inoltre calcolati degli indici statistici della differenza tra i dati della simulazione e le osservazioni. Poiché i dati hanno una natura vettoriale, la differenza è stata quantificata in base alla distanza normalizzata tra il vettore osservazione e il vettore simulazione; dove il vettore di normalizzazione è dato dalle osservazioni.

Nello specifico si sono confrontati i vettori velocità delle formule 4.2.3 e 4.2.4; quindi l'indice di qualità è dato da:

$$Eru = \frac{\sqrt{\left(u^{\circ} - u^{b}\right)^{2} + \left(v^{\circ} - v^{b}\right)^{2}}}{\sqrt{\left(u^{\circ^{2}} + v^{\circ^{2}}\right)}}$$
(4.2.5)

E' possibile verificare l'efficienza della simulazione pseudo lagrangiana, confrontando l'indice *Eru* con:

$$Eru^{*} = \frac{\sqrt{\left(u^{o} - \tilde{u}^{b}\right)^{2} + \left(v^{o} - \tilde{v}^{b}\right)^{2}}}{\sqrt{\left(u^{o^{2}} + v^{o^{2}}\right)}}$$
(4.2.6)

dove $\vec{\tilde{v}}^{b} = (\tilde{u}^{b}, \tilde{v}^{b})$ è ottenuta interpolando il campo di velocità del modello sulla posizione iniziale della boa Argo; questa è di fatto una approssimazione puramente euleriana che non tiene conto dell'evoluzione spaziale e temporale dei campi di velocità lungo la deriva della boa.



Figura 4.1.1:Gennaio, confronto tra gli spostamenti delle boe Argo e la media mensile dei campi di velocità del modello.



Figura 4.1.2: Febbraio, confronto tra gli spostamenti delle boe Argo e le medie mensili dei campi di velocità del modello.



Figura 4.1.3: Marzo, confronto tra gli spostamenti delle boe Argo e la media mensile dei campi di velocità del modello.



4.1.4: Aprile, confronto tra gli spostamenti delle boe Argo e la media mensile dei campi di velocità del modello.



4.1.5: Maggio, confronto tra gli spostamenti delle boe Argo e la media mensile dei campi di velocità del modello.



4.1.6: Giugno, confronto tra gli spostamenti delle boe Argo e la media mensile dei campi di velocità del modello.

4.3 Risultati confronto quantitativo

Sono state eseguite 32 simulazioni della durata di 5 giorni per le boe profilanti 1900025, 1900026, 1900027, 6900087, 6900089 e 6900091. I risultati sono riportati nelle figure 4.3.1-4.3.6.

Le figure sono così strutturate:

- in alto; con il rosso sono contrassegnati i tracciati delle 'boe modello' per ogni simulazione, in blu è riportato il percorso reale delle boe Argo.
- al centro; i vettori rossi rappresentano le velocità di deriva delle 'boe modello', i vettori blu le velocità lagrangiane delle boe Argo.
- in basso; in rosso è mostrato l'indice *Eru*, in blu *Eru**.

Le simulazioni delle traiettorie delle 'boe modello' indicano chiaramente la difficoltà connessa alla predizione degli spostamenti delle boe lagrangiane. Per quanto il modello riproduca le principali strutture di circolazione, si osservano ricorrenti errori nel loro posizionamento; il metodo psedo lagrangiano applicato risente fortemente di questo errore e produce risultati non sempre attendibili. Tuttavia è interessante notare che l'indice Eru è generalmente minore di Eru^* , si ha quindi conferma che la simulazione pseudo-lagrangiana produce risultati migliori rispetto alla semplice interpolazione dei campi euleriani.

La mancanza di importanti informazioni riguardo alcune specifiche del ciclo Argo ha inevitabilmente compromesso l'attendibilità dei risultati ottenuti; è quindi necessario ripetere le simulazioni eseguite una volta che i dati mancati siano resi disponibili.







Figura 4.2.3: Boa Argo1900027 (vedi paragrafo 4.3)..







4.4 Introduzione al problema di assimilazione

Negli ultimi decenni si sta assistendo ad un forte sviluppo della tecnologia e della diffusione degli strumenti lagrangiani. Se nel passato questi dati venivano utilizzati prevalentemente per valutare da un punto di vista statistico le caratteristiche della circolazione generale oggi sta aumentando l'interesse legato ad uso operativo di questi dati, nel contesto dei sistemi previsionali marini. In particolare nell'ambito del progetto MFS si è deciso di implementare un sistema di assimilazione dei dati Argo.

E' possibile dividere, da un punto di vista operativo, il problema dell'assimilazione Argo in due sistemi differenti: l'assimilazione dei profili di temperatura e salinità e l'assimilazione dei campi di velocità.

I profili delle boe Argo possono ragionevolmente essere considerati come delle misure puntuali di tipo euleriano, in quanto il tempo impiegato per effettuare la risalita è abbastanza contenuto da poter approssimare il profilo effettivamente percorso con un profilo verticale. Applicando questa procedura l'assimilazione dei profili Argo si riduce a un problema paragonabile al caso dei CTD.

L'assimilazione dei campi di velocità richiede invece una sostanziale riformulazione del problema dell'analisi. Nel contesto dell'Interpolazione Ottimale, Lorenc (2000) ha così formulato il problema:

$$u^{a} = u^{b} + BG^{T} (GBG^{T} + R)^{-1} (y - H(u^{b}))$$
(4.4.1)

dove u^a è il vettore delle velocità del modello dopo l'assimilazione, u^b è il vettore delle velocità del modello prima dall'assimilazione (first guess) e \vec{y} è il vettore delle osservazioni. H() è l'operatore di osservazione, R è la matrice di covarianza dell'errore di osservazione, B è la matrice di covarianza dell'errore del modello e infine:

$$G = \frac{\delta H(u^b)}{\delta u^b} \tag{4.4.2}$$

è la derivata dell'operatore di osservazione (matrice di sensitività).

Come è stato evidenziato da Molcard et al. (2002) un problema centrale dell'assimilazione dei dati lagrangiani è definire la forma dell'operatore di osservazione H() che deve rendere conto della relazione non lineare tra le velocità euleriane del modello e le velocità stimate dalla deriva delle boe. Si noti che l'approssimazione pseudo-lagrangiana, equivale a prendere un operatore di

osservazione H() lineare; ne consegue quindi che (4.4.2) si riduce ad una costante.

Il lavoro di Molcard et al. (2002) ha, tra le altre cose, presentato un confronto tra due simulazioni di assimilazione: una pseudo-lagrangiana e una completamente lagrangiana. I risultati mostrano un sostanziale miglioramento dell'efficienza dell'assimilazione quando la natura lagrangiana degli strumenti è pienamente considerata. Un ulteriore risultato incoraggiante di questo lavoro riguarda il confronto tra l'impatto di misure fisse e l'impatto di misure lagrangiane sull'efficienza dell'assimilazione. Le misure lagrangiane sembrano avere, stando ai dati della simulazione, una maggiore influenza sull'assimilazione rispetto ai correntometri, specialmente quando la quantità di dati disponibili è limitata. Considerate le difficoltà pratiche connesse alle misure euleriane dei campi di velocità nell'oceano aperto questo risultato è di grande importanza e mette in luce l'importanza che l'assimilazione delle boe lagrangiane potrebbe avere nel correggere il campo di velocità del modello.

Capitolo 5

5.1 La simulazione

Le componenti principali di ogni sistema previsionale sono due: l'algoritmo di assimilazione e il modello numerico. Per poter separare l'errore osservato nei singoli contributi delle due componenti, è necessario isolarne una. Si è quindi condotto un esperimento nel quale l'assimilazione è stata messa 'a riposo'. Nel corso di questo esperimento, d'ora in poi chiamato simulazione, sono state rieseguite le previsioni dal 10 ottobre 2002 al 16 giugno 2003 utilizzando lo stesso forzante atmosferico del forecast operativo, ma inizializzando di volta in volta il vettore di stato del modello con il forecast precedente.

In equazione (1.4.8) è stata definita la forma generale dell'assimilazione, ovvero la combinazione ottimale di osservazioni e modello:

$$X_t^a = X_t^b + K(Y_t^o - H(X_t^b))$$

Nel corso dell'esperimento di simulazione si è invece considerato;

$$X_t^a = X_t^b \tag{5.1.1}$$

Quindi il modello numerico, M(), rimasto invariato, integra le seguenti equazioni:

$$X_{t+1}^{b} = M\left(X_{t}^{a}\right) \equiv M\left(X_{t}^{b}\right)$$
(5.1.2)

Prima di presentare i risultati è opportuno fare alcune considerazioni sull'esperimento di simulazione.

La data di inizio della simulazione, il 10 ottobre 2002, è anteriore di tre mesi rispetto al primo dato Argo. In questo modo si è cercato di far perdere al sistema previsionale l'informazione delle precedenti assimilazioni, in quanto la condizione iniziale della simulazione è l'assimilazione del 10 ottobre 2002. Tuttavia, 3 mesi non sono sufficienti per ottenere risultati completamente indipendenti rispetto alle precedenti assimilazioni; per cui lo scopo della simulazione non è di valutare l'efficienza in sé del modello numerico, ma piuttosto di analizzare come l'efficienza cambia in assenza di assimilazione.

Un'ulteriore osservazione riguarda lo stato del sistema osservativo nei mesi dell'analisi. Il programma VOS-XBT è stato sospeso nel settembre 2000, ad eccezione di un'unica rotta nell'Adriatico Meridionale che è stata mantenuta. Di fatto si può dire che non vi sia alcuna influenza dei dati XBT sul sistema di assimilazione nell'area e nel periodo interessate da questo lavoro. Quindi i dati del forecast operativo usati per la validazione risentono soltanto dell'assimilazione di SLA. Per quanto questa situazione non sia ottimale per il sistema previsionale, ha permesso di focalizzare l'attenzione sullo schema di assimilazione di SLA, che come è stato introdotto nel Capitolo 1, è separato rispetto allo schema XBT.

5.2 Confronto Simulazione e Modello Operativo.

Al fine di valutare la consistenza delle previsioni della simulazione e del modello operativo, sono state calcolati le differenze dei campi medi mensili di temperatura e salinità. In figura 5.2.1-5.2.12 sono riportati i risultati. Per ogni mese da gennaio a giugno 2003 si sono confrontate le medie mensili dei campi temperatura e salinità alle profondità di 15 m , 200 m e 600 m in modo da poter valutare come l'assimilazione incide sui livelli verticali del modello. La differenza calcolata ha questa forma:

$$T' = \overline{T} \mod operativo - \overline{T}$$
 simulazione
 $S' = \overline{S} \mod operativo - \overline{S}$ simulazione

dove T' è la differenza di temperatura e S' è la differenza di salinità. Si ha quindi che una differenza positiva implica un valore maggiore nei campi del modello operativo rispetto alla simulazione (e viceversa).



Figura 5.2.1: Differenze di Salinità tra simulazione ed assimilazione, Gennaio: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.2: Differenze di Salinità tra simulazione ed assimilazione, Febbraio: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.3: Differenze di Salinità (psu) tra simulazione ed assimilazione, Marzo: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.4: Differenze di Salinità (psu) tra simulazione ed assimilazione, Aprile: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.5: Differenze di Salinità (psu) tra simulazione ed assimilazione, Maggio: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.6: Differenze di Salinità (psu) tra simulazione ed assimilazione, Giugno: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.7: Differenze di Temperatura (°C) tra simulazione ed assimilazione, Gennaio: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.8: Differenze di Temperatura (°C) tra simulazione ed assimilazione: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.9: Differenze di Temperatura (°C) tra simulazione ed assimilazione: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.10: Differenze di Temperatura (°C) tra simulazione ed assimilazione, Aprile: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2. Differenze di Temperatura (°C) tra simulazione ed assimilazione, Maggio: (a) 15 m, (b) 200, (c) 600 m.



Figura 5.2.12: Differenze di Temperatura (°C) tra simulazione ed assimilazione, Giugno: (a) 15 m, (b) 200, (c) 600 m.

La prima considerazione che si può fare sui risultati è che l'assimilazione produce un effetto complesso sui campi di temperatura e salinità, e non è possibile identificare un bias costante rispetto ai dati della simulazione; si osservano anomalie sia positive che negative in tutti i casi considerati.

L'ampiezza massima delle differenze si registra alla profondità intermedia considerata (200 m), sia per la temperatura che per la salinità. Ciò è consistente con lo schema di assimilazione dei dati SLA, che riguarda lo strato della colonna d'acqua a profondità superiore di 100 m. Si ha quindi che nei livelli superficiali l'assimilazione non agisce direttamente, ma la correzione viene veicolata dalla dinamica interna del modello. Tuttavia, va considerato che la temperatura superficiale nel modello operativo viene 'rilassata' con i dati SST da satellite, mentre nel caso della simulazione si è applicata una correzione basata sulla climatologia. Ciò contribuisce alle differenze calcolate.

In profondità le variazioni indotte dall'assimilazione sono piuttosto limitate in quanto l'entità della correzione apportata è proporzionale alla varianza del campo alla profondità relativa.

L'informazione riguardo la varianza dei parametri temperatura e salinità alle profondità della griglia del modello, è contenuta nella matrice di covarianza dell'errore del forecast *B*, che nell'algoritmo di assimilazione SOFA è decomposta nella sua componente orizzontale e verticale (vedi Capitolo 1); la componente verticale è costituita da EOF. Quindi la struttura della correzione dell'assimilazione sui dati del modello è fatta in base a una statistica della varianza dei parametri calcolata a priori. Nel caso dell'assimilazione di SLA questa statistica è trivariata e rende conto delle variazioni e intercorrelazioni nel tempo di Ψ , *T* ed *S*. In figura 1.6.1 è riportata la EOF usata nell'assimilazione di SLA, la cui struttura è coerente con l'entità massima della correzione apportata dall'assimilazione alla profondità di circa 200 metri.

L'unica regione del Mediterraneo a cui queste considerazioni non si possono applicare è l'Adriatico meridionale e la parte a ridosso delle coste Italiane del Mar Ionio; ossia in corrispondenza dell'unica traccia di XBT ancora attiva. Da marzo in poi è particolarmente chiaro il contributo di queste misure sulla correzione introdotta dall'assimilazione. Si osservano forti differenze sia di temperatura che di salinità che sono localizzate lungo la traccia percorsa dalla nave di opportunità. Il contributo di queste misure risulta facilmente individuabile alla profondità di 600 m, dove i dati di SLA producono effetti contenuti sulla correzione dell'assimilazione. In particolare si nota come le misure di XBT correggano il modello introducendo valori sistematicamente più altri di temperatura e salinità rispetto alla simulazione.

5.3 Risultati Statistici

I dati della simulazione sono stati utilizzati per eseguire di nuovo l'analisi di accuratezza già condotta sui dati del modello operativo. In questo modo è possibile valutare quantitativamente l'efficacia dell'assimilazione; ossia quanto l'assimilazione riesce ad avvicinare i dati del modello alle misure di temperatura e salinità delle boe Argo. Per la una descrizione degli indici e delle procedure applicate si veda il Capitolo 3.

In figura 5.3.1 sono riportati tutti i profili esaminati delle boe Argo, del modello operativo, e della simulazione. L'informazione contenuta in questa figura è di fondamentale importanza rispetto al lavoro di questa tesi. Risulta infatti evidente che i profili di temperatura e in particolar modo di salinità, interpolati dai dati della simulazione, approssimano meglio le osservazioni Argo di quanto non faccia il modello operativo. Ciò equivale a dire che l'assimilazione introduce un errore nel modello piuttosto che correggerlo.

Il calcolo delle EOF univariate e bivariate per i dati della simulazione (figure 5.2.2 e 5.2.3) ha prodotto risultati che concordano con questa visione d'insieme. Si nota che la simulazione riproduce una struttura della varianza di temperatura e salinità in cui le caratteristiche inconsistenti con i dati Argo, delle EOF del modello operativo, sono mitigate. E' tuttavia evidente che i dati della simulazione risentano ancora dell' assimilazione precedente al 10 ottobre 2002.

Di seguito sono riportate le statistiche calcolate per la temperatura e la salinità.

5.3.1 Temperatura

Le figure sono presentate in analogia al capitolo 3.

In figura 5.5.2 sono riportate le statistiche di misfit calcolate su tutto il bacino. Nelle figure 5.3.3-5.3.6 sono riportate le statistiche di misfit relative alle singole
regioni e ai due strati considerati della colonna d'acqua. I dati della simulazione mostrano statistiche di misfit abbastanza simili a quelle relative ai dati del modello operativo. Il confronto tra le statistiche calcolate sui dati della simulazione e del modello operativo può essere così riassunto:

- il RMS è generalmente minore per i dati della simulazione, in particolar modo nei periodi gennaio-febbraio e maggio-giugno. Le differenze sono dell'ordine di 0.1 ° C.
- il bias non mostra un andamento crescente nei dati della simulazione, ma rimane costantemente negativo intorno a 0.5 °C. La differenza massima nei valori di bias tra la simulazione e il modello operativo è di 0.3 ° C e si riscontra nello strato superficiale della colonna d'acqua nel periodo gennaio-febbraio.
- L'Anomalia di Correlazione ha un andamento simile per i dati della simulazione e del modello operativo. Solo nello strato intermedio si osservano delle differenze; i dati della simulazione mostrano una AC leggermente maggiore con i dati Argo rispetto ai dati del modello operativo nel periodo maggio-giugno.

5.3.2 Salinità

In figura 5.5.6 sono riportate le statistiche di misfit calcolate su tutto il bacino. Nelle figure 5.3.7-5.3.10 sono riportate le statistiche di misfit relative alle singole regioni e ai due strati considerati della colonna d'acqua.

Le principali differenze rispetto alle statistiche di misfit calcolati sui dati del modello operativo sono:

- nel periodo gennaio-febbraio si osservano le maggiori differenze nel RMS.
 Per i dati della simulazione il RMS risulta minore di 0.2 psu rispetto al modello operativo; se si considerano i primi 200 m della colonna d'acqua questa differenza è di 0.3 psu. Nel periodo aprile-giugno i dati del modello operativo hanno un RMS leggermente minore rispetto ai dati della simulazione, specialmente nello strato intermedio della colonna d'acqua;
- nel caso dei dati della simulazione il bias assume valori abbastanza costanti attorno a 0.2 psu, mentre i dati del modello operativo mostrano un andamento decrescente da gennaio a giugno. Si nota che nei mesi maggio e giugno il bias ha valori leggermente minori per il modello operativo rispetto alla simulazione;

l'Anomalia di Correlazione (AC) ha un andamento simile per i dati della simulazione e del modello operativo. L'AC assume valori negativi prossimi a -1 per i primi mesi dell'analisi per poi assumere valori intorno a 0 nel periodo maggio-giugno. Sia il modello operativo che la simulazione non riescono a cogliere la struttura di salinità osservata dalle boe Argo, in particolar modo nelle regioni 11, 12 e 13.



Figura 5.3.1: Figura complessiva dei profili Argo, del Modello operativo e della Simulazione.



Figura 5.2.2: EOF univariate calcolate sui dati della simulazione.



Figura 5.2.3: EOF bivariate calcolate sui dati della simulazione.



Figura 5.3.2: Statistiche di misfit calcolate sui tutti i profili disponibili.



Figura 5.3.3: Statistiche del misfit relative alle regioni 8, 11, 12 e 13



Figura 5.3.4: Statistiche di misfit per le regioni 8, 11, 12 e 13 relative allo strato superficiale della colonna d'acqua.



Figura 5.3.5: Statistiche di misfit per le regioni 8, 11, 12 e 13 relative allo strato intermedio della colonna d'acqua.



Figura 5.3.6: Statistiche di misfit calcolate sui tutti i profili.



Figura 5.3.7: Statistiche di misfit calcolate per le regioni 8, 11, 12 e 13.



Figura 5.3.8: Statistiche di misfit per le regioni 8, 11, 12 e 13 relative allo strato superficiale della colonna d'acqua.



Figura 5.3.9 Statistiche di misfit per le regioni 8, 11, 12 e 13 relative allo strato intermedio della colonna d'acqua.

:

5.4 Considerazioni generali sul confronto quantitativo.

I dati delle boe derivanti Argo hanno costituito il termine di paragone per verificare quanto l'assimilazione contribuisca ad avvicinare il modello alle osservazioni, che a meno dell'errore di misura, rappresentano lo stato 'vero' dei campi di temperatura e salinità nel punto ed all'istante della misura..

Il problema di verifica è stato diviso in due fasi:

- 1. calcolo delle statistiche di misfit tra i dati del modello operativo e le osservazioni Argo (Capitolo 3).
- 2. calcolo delle statistiche di misfit tra i dati della simulazione e le osservazioni Argo.

Questa analisi si riduce ad una valutazione dell'efficienza dell'assimilazione, poiché l'assimilazione costituisce l'unica differenza tra il modello operativo e la simulazione.

I risultati ottenuti mostrano che l'assimilazione non è efficace nel caso della temperatura e che introduce un errore importante nei campi di salinità.

Lo schema di assimilazione SOFA per i dati di SLA presenta due aspetti 'critici' che sono la definizione dell'operatore di osservazione H() e la decomposizione della matrice di covarianza dell'errore del forecast *B*:

- l'operatore di osservazione H() applica l'equazione (1.3.2), mette quindi in relazione la topografia dinamica, stimata dalla SLA, con la funzione di flusso e la struttura verticale di densità, da cui si ricava la struttura di temperatura e salinità;
- nel caso dell'assimilazione di SLA la matrice di covarianza dell'errore del forecast B è espressa come (1.4.21). Questa matrice viene scomposta nelle sue componenti verticali e orizzontali. La struttura verticale della varianza è introdotta mediante una EOF trivariata che è stata calcolata a priori dai risultati di una simulazione. Questa statistica è unica per tutto il bacino per tutto l'anno.

Le inefficienze dello schema di assimilazione riscontrate nel corso di questo studio sono probabilmente legate a questi due aspetti di SOFA.

In particolar modo l'utilizzo di una EOF trivariata, comune per tutto il Mediterraneo e per tutto l'anno, sembra essere una semplificazione molto forte che non tiene conto delle differenze nella struttura termoalina che caratterizzano i diversi bacini del Mediterraneo.

Capitolo 6

Conclusioni

I dati Argo costituiscono una nuova rete di osservazione dello stato fisico degli oceani a supporto del consolidamento dei sistemi previsionali marini. Questa tesi ha riguardato l'analisi dei primi dati Argo disponibili nel Mar Mediterraneo e il loro confronto con il sistema previsionale Mediterranean Forecasting System (MFS). Per la prima volta è stato possibile disporre di un insieme di dati consistente di misure completamente indipendenti rispetto a MFS; ciò ha permesso di eseguire una rigorosa validazione del sistema previsionale nel suo insieme.

I risultati ottenuti indicano chiaramente la necessità di migliorare alcuni aspetti di MFS, in particolar modo dell'algoritmo di assimilazione SOFA. Pertanto la conclusione di questa tesi non vuole essere altro che l'individuazione di alcune direttive secondo cui procedere nello sviluppo del progetto:

- L'assimilazione dei dati di SLA è fortemente vincolata dall'utilizzo di una singola EOF trivariata costante nel tempo per tutto il bacino Mediterraneo. Risulta evidente la necessità di calcolare nuove statistiche relative alle singole regioni del Mediterraneo che non siano mediate rispetto alla variabilità annuale. Ossia l'assimilazione di SLA dovrebbe basarsi su diversi set di EOF in grado di evolvere, in modo il più possibile consistente con i processi fisici, la matrice di covarianza dell'errore del campo di base o background.
- I dati Argo sono un insieme prezioso di informazioni sullo stato fisico del Mediterraneo, per cui è necessario ampliare SOFA in modo da assimilare le misure di temperatura e salinità delle boe profilanti. Altrettanto

importante e necessaria è l'assimilazione delle misure lagrangiane di velocità.

Una questione non risolta da questa tesi è la determinazione dell'errore di misura delle boe Argo, con particolare riferimento all'errore di localizzazione dei profili. Nell'ambito di questa tesi si è potuto ragionevolmente assumere che le discrepanze osservate tra le analisi di MFS e i dati Argo fossero solo in minima parte dovute all'errore di misura. Tuttavia un successivo lavoro, volto all'assimilazione dei dati Argo, non può prescindere dall'analisi dell'errore di osservazione, in quanto la costruzione della matrice R di covarianza dell'errore di osservazione è un prerequisito essenziale dell'assimilazione stessa. La matrice R è generalmente diagonale in quanto gli errori delle osservazioni sono normalmente non correlati tra loro; la possibilità di applicare questa semplificazione ai dati di temperatura e salinità delle boe Argo è da verificarsi.

In questa tesi non è stato trattato il problema della valutazione dell'importanza delle inefficienze osservate, in quanto il giudizio deve propriamente riguardare l'influenza che il malfunzionamento ha sull'utente del sistema previsionale; soltanto l'utilizzo determina il valore delle analisi e delle previsioni. E' quindi importante che, durante una fase di piena operatività del sistema previsionale, sia attivato un sistema di validazione permanente che tenga conto delle applicazioni finali del sistema previsionale. In questo modo si potrà propriamente valutare il ruolo del servizio fornito sulla comunità degli utenti e i limiti di confidenza dello stesso.

Il sistema previsionale di MFS è ancora in una fase attiva di sviluppo. La più consistente innovazione, che sarà resa operativa nel 2004, è stata l'implementazione di un modello numerico OPA a 1/16 di grado di risoluzione (Tonani 2003). L'ottimizzazione del sistema di assimilazione SOFA è da considerarsi una priorità, la cui realizzazione, in concomitanza all'introduzione del modello OPA, costituisce un sensibile avanzamento verso gli obiettivi del Mediterranean Forecasting System.

Bibliografia

Artegiani A., Bregant D., Paschini E., Pinardi, N, Raicich, F and Russo A. (1997), The Adriatic Sea General Circulation. Part II: Baroclinic Circulation Structure, *Journal of Physical Oceanography*, 27, 1515-1532.

Ayoub N., Le Traon P. Y. and De Mey P. (1997), Combining ERS-1 and Topex/Poseidon data to observe the variable oceanic circulation in the Mediterranean Sea, *Journal of Marine Systems*.

Carval T. (2003), Argo User's Manual, Version 2.0.

Daley R. (1996), Atmospheric data analysis. Cambridge University Press, 457 pp.

Demirov E., Pinardi N., Fratianni C., Tonani M., Giacomelli L. and De Mey P. (2003), Assimilation Scheme of Mediterranean Forecasting System: Operational implementation. *Annales Geophysicae* 21, 1, 189-204.

De Mey, P. and Benkiran, M., 2002. A multivariate reduced order Optimal Interpolation Method and its application to the Mediterranean basin-scale circulation, in: *Ocean Forecasting: conceptual basis and applications*. Springer Verlag, pp. 472.

De Mey P. and Robinson A. (1987), Assimilation of altimeter eddy fields in a limited area quasi-geostrophic model, *Journal of Physical Oceanography*, 17, 2279-2293.

Griffa, A. and Molcard, A. (2003), Intermediate Currents; Recommendations for the profiler cycle characteristics, Task 4120, MFSTEP EU Project EVK3-2002-00075

Korres G., Pinardi N. and Lascaratos A. (2000), The ocean response to lowfrequency interannual variabilità in the Meditarranean sea. Part I: Sensitivity Experiments and Energy Analysis, *Journal of Climate*, 13, 705-731.

Korres G., Pinardi N. and Lascaratos A. (2000), The ocean response to low-frequency interannual variability in the Meditarranean sea. Part II: Empirical Orthogonal Function Analysis, *Journal of Climate*, 13, 732-745.

Le Traon P.-Y. (2002), Satellite oceanography for ocean forecasting, in: *Ocean Forecasting: conceptual basis and applications*. Springer Verlag, pp. 472.

Lorenc A. C. (2000), A Bayesian approach to observation quality control in variational and statistical assimilation, In: Molcard A., Piterbarg L., Griffa A., Özgökmen T.M. and Mariano A.J. (2002), Assimilation of drifter position for the

reconstruction of the Eulerian circulation field, *Journal of Geophisical Reserch*, 108(C3), 3056-3077.

Lorenz E. N. (1956), Empirical Orthogonal function and statistical weather prediction, Scientific Report No1, Statistical Forecasting Project.

Manzella G. M. R., Scoccimarro E., Pinardi N.and Tonani M. (2003), Improved near real-time data management procedures for the Mediterranean ocean Forecasting System-Voluntary Observing Ship program, *Annales Geophysicae*, 21, 1, 49-62.

Masina S., Pinardi N. and Navarra A. (2001), A global ocean temperature and altimeter data assimilation system for studies of climate variability. *Climate Dynamics*, 17, 687-700.

Mellor G. L.and Ezer T. (1991), A Gulf Stream model and an altimetry assimilation scheme, *Journal of Geophysical Research*, 96(C5), 8779-8795.

Molcard A., Piterbarg L., Griffa A., Özgökmen T.M. and Mariano A.J. (2002), Assimilation of drifter position for the reconstruction of the Eulerian circulation field, *Journal of Geophisical Reserch*, 108(C3), 3056-3077.

Morrison. D.F. (1976), Metodi di analisi statistica multivariata, Casa Editrice Ambrosiana .

Murphy A. H. (1995), The Coefficients of Correlation and Determination as Measures of Performance in Forecast Verification, *Weather and Forecasting*, 7, 681-698.

Paschini E., Artegiani A. and Pinardi N. (1993), The mesoscale eddy field of the middle Adriatic Sea, *Deep Sea Reserch*.

Pinardi, N. and Flemming, N. C. (1998), The Mediterranean Forecasting System Science Plan, *EuroGOOS Publication* No.11, Southampton Oceanography Centre, Southampton.

Pinardi, N. and Masetti E. (2000), Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: a review, *Paleooceanografia Paleoecologia Paleoclimatologia*, 158, 153-174.

Pinardi, N. and Navarra, A. (1993), Baroclinic wind adjustment processes in the Mediterranean sea, *Deep-sea research II*, 40, 6, 1299-1326.

Pinardi N., Rosati A. and Pacanowski R.C. (1995), The sea surface pressure formulation of rigid lid models. Implications for altimetric data assimilation studies. *Journal of Marine Systems*, 19, 109-119.

Raicich F. and Rampazzo A. (2003), Observing System Simulation Experiment for the assessment of temperature sampling strategies in the Mediterranean Sea, *Annales Geophysicae*, 21, 1, 151-166.

Robinson A. R.(1986), Data Assimilation, Mesoscale Dynamics and Dynamical Forecasting. Advanced Physical Oceanographic Numerical Modelling, J.J. O'Brien, editor, R. Reidel Publishing Co.

Robinson A.R., Lesile W.G., Theocharis A., Lascaratos A. (2001), Mediterranean Sea Circulation, *Indira*.

Roussenov V., Stanev E., Artale V. and Pinardi N. (1995), A seasonal model of the Mediterranean Sea, *Journal of Geophysical Research*, 100(C7), 13, 515-538.

Sparnocchia S., Pinardi N. and Demirov E. (2003), Multivariate Empirical Orthogonal Function analysis of the upper thermocline structure of the Mediterranean sea from observation and model simulation. *Annales Geophysicae* 21, 1, 167-188.

Tonani M. (2003), Studio di predicibilità della circolazione del Mar Mediterraneo, Tesi di Dottorato di Ricerca in Geofisica, XV Ciclo, Università degli studi di Bologna

Vignudelli S., Cipollini P., Reseghetti F., Fusco C., Gasparini G. P. and Manzella G. M. R. (2003), Comparison between XBT data and TOPEX/Poseidon satellite altimetry in the Ligurian-Tyrrhenian area, *Annales Geophysicae*, 21, 1, 123-135.

Wilks D. S. (1995), Statistical Methods in the Atmospheric Science, *Academic Press*, 233-283.