



ISPRA

Istituto Superiore per la Protezione
e la Ricerca Ambientale

Tesi di tirocinio di formazione e orientamento

(Convenzione stipulata tra ISPRA e l'Università degli Studi Roma Tre)

ANALISI E RICOSTRUZIONE DELLE SERIE STORICHE DEI DATI MAREOGRAFICI ED ONDAMETRICI DELL'ISPRA

Ing. Marco Picone

Tutor ISPRA: Ing. Gabriele Nardone

ANNO 2008-2009

Data	Firma Tirocinante	Firma del Tutor ISPRA	Firma Responsabile Servizio ISPRA

Abstract

L'informazione ondametrica e mareografica è fondamentale in tutti gli studi, i progetti, i lavori e l'esercizio di opere di ingegneria costiera e marittima. I database RON e RMN presentano una serie di gap di breve e lunga durata che rendono l'analisi statistica dei dati inefficiente. In questo lavoro viene effettuata un'analisi dei dati mancanti nei dataset delle reti nazionali gestite dal Servizio Mareografico dell'ISPRA e viene proposte delle metodologie di ricostruzione dei missing values della Rete Ondametrica Nazionale. Per i dati di temperatura dell'acqua si adotta il metodo statistico dell'imputazione multipla dei dati basato su un modello multivariato Bayesiano che utilizza come covariate i dati analoghi misurati dai mareografi. La ricostruzione dei dati di altezza d'onda significativa prende spunto, invece, dalla rielaborazione delle informazioni grezze (HNE) inviate dalle boe.

Sommario

Indice delle Figure.....	iv
Indice delle Tabelle.....	vii
INTRODUZIONE.....	1
Progetto formativo.....	1
Obiettivi.....	2
CAPITOLO 1: Fonti di Dati.....	4
1.1 - La rete di monitoraggio dell'ISPRA.....	4
1.2 - La rete mareografica nazionale.....	6
1.2.1 - Cenni storici.....	6
1.2.2 - Descrizione della rete.....	9
1.3 - La rete ondametria nazionale.....	15
1.3.1 - La rete ondametria nazionale dal 1989 al 1999.....	15
1.3.2 - La rete ondametria nazionale dal 1999 al 2002.....	16
1.3.3 - La rete ondametria nazionale dopo il 2002.....	18
CAPITOLO 2: Missing Values nel database RON.....	21
2.1 - I settori di costa.....	21
2.2- Tipologie di dati acquisiti dalla rete ondametria.....	23
2.3.1 - La temperatura dell'acqua.....	26
2.3.2 - L'altezza d'onda.....	29
2.3 - Acquisizione dei dati.....	31
2.3.1 - File HNE per acquisizioni nel dominio del tempo.....	32
2.3.2 - File MAIN, relative tabelle SQL e dati spettrali.....	32
2.3.3 - File di sistema, SIS e relativa tabella.....	33
2.3.4 - File di sistema, Status e relativa tabella.....	34
2.4 - Elaborazione dei dati.....	34

2.4.1 - Modulo di Post-Processing nella stazione a terra.....	34
2.4.2 - Modulo di Post-Processing nel Centro di controllo e gestione.....	37
2.5 - Assenza dei dati ondametrici e qualità del dataset.....	39
2.5.1 - Dati anomali.....	39
2.5.2 - Dati mancanti.....	41
2.6 - Metodi statistici di ricostruzione dei dati mancanti.....	53
2.6.1 - Imputazione Multipla: Teoria.....	55
2.6.2 - Imputazione Multipla: Metodo.....	59
CAPITOLO 3: Ricostruzione dei missing values della Temperatura dell'acqua.....	62
3.1 - Dati RON (anni 2002-2006).....	62
3.2 - Boe e Mareografi (anni 2002-2006).....	69
3.3 - Imputazione multipla dei dati mancanti.....	74
3.3.1. - Mar Tirreno.....	79
3.3.2. - Mar Adriatico.....	82
3.3.3. - Mar Ionio.....	85
CAPITOLO 4: Ricostruzione dei missing values.....	87
dell'altezza d'onda significativa.....	87
4.1 - Dati HNE (anni 2002-2006).....	88
4.2 - Ricostruzione da file HNE.....	97
CONCLUSIONI.....	100
Sviluppi futuri.....	102
Bibliografia.....	104

Indice delle Figure

Fig. 1. 1 - La Rete Ondametrica Nazionale e la Rete Mareografica Nazionale	5
Fig. 1. 2 - RMN: Carta delle stazioni all'origine della rete.....	6
Fig. 1. 3 – RMN: Sistema locale e sistema centrale	9
Fig. 1. 4 - RMN: gli strumenti in uso in un mareografo.....	13
Fig. 1. 5 - RON: Periodo 1 e Periodo 2	15
Fig. 1. 6 - RON: Le Boe utilizzate nella rete nel corso degli anni	18
Fig. 1. 7 - Numero di eventi significativi registrati dalle boe fino al 2003	18
Fig. 1. 8 - RON: Caratteristiche in acqua della boa TRIAXYS	19
Fig. 2. 1 - Tratti di costa individuati in base alla posizione geografica di ciascun ondametro e di criteri di omogeneità e orientamento del litorale.....	22
Fig. 2. 2 - Settori di costa estesi.....	23
Fig. 2. 3 - Rappresentazione schematica dei profili di temperatura del mare in funzione della profondità degli strati superficiali inferiori al metro di profondità.....	27
Fig. 2. 4 – Profilo verticale di temperature negli strati superficiali (inferiori ai 10 m di profondità) durante la notte (sinistra) e il giorno (destra, in presenza di forte radiazione solare e vento leggero).....	28
Fig. 2. 5 - Meccanismo di formazione delle onde da vento.....	30
Fig. 2. 6 - Incidenza dei dati mancanti per ogni anno sul totale per ogni boa.....	43
Fig. 2. 7 – Percentuale di dati mancanti riferiti a gap lunghi sul totale.....	46
Fig. 2. 8 – Inoperatività annua.....	49
Fig. 2. 9 - Inoperatività complessiva	52
Fig. 2. 10 - Ritardo globale medio.....	52
Fig. 2. 11 - Rappresentazione schematica del metodo di imputazione multipla, in cui <i>m</i> è il numero di imputazioni.....	55
Fig. 2. 12 - Schematizzazione dell'algorithmo MI	59
Fig. 3. 1 - Andamento nel tempo della temperatura dell'acqua per ogni boa.....	63
Fig. 3. 2 - Dati aggregati delle boe suddivise per bacini	66

Fig. 3. 3 - Correlogrammi relativi ai dati di temperatura dell'acqua (Boa di Cetraro) con Lag piccolo (a) e grande (b).....	68
Fig. 3. 4 - La temperatura dell'acqua misurata dalla boa e dal mareografo. La differenza tra le misure è legata alla differenza di profondità del sensore	70
Fig. 3. 5 - Andamento nel tempo delle temperature dell'acqua misurate dalla Boa di Cetraro e dal Mareografo di Palinuro	72
Fig. 3. 6 – Boa di Cetraro e Mareografo di Palinuro: (a) Correlazione tra i dati di temperatura dell'acqua misurati da Boe e Mareografi; (b) Andamento dei residui.....	73
Fig. 3. 7 - Correlogramma dei residui dei dati di temperatura dell'acqua misurati alla boa di Cetraro e al mareografo di Palinuro.....	73
Fig. 3. 8 - Variazione dello standard error dei primi 5000 dati mancanti ricostruiti nel dataset della boa di Capo Comino in funzione del numero di imputazioni (da 10 a 100).....	75
Fig. 3. 9 - Grafici tipo - (a) Andamento della serie osservata e della serie ricostruita; (b) andamento degli standard error; (c) boxplot degli standard error.....	77
Fig. 3. 10 - Ingrandimento delle serie storiche reali e ricostruite in una settimana (Boa di Cetraro): a- Sovrapposizione dei dati reali e ricostruiti; b- Andamento delle medie mobili.....	78
Fig. 3. 11 - Capo Comino - Ricostruzione dei Dati e Standard error	79
Fig. 3. 12 - Capo Gallo - Ricostruzione dei Dati e Standard error	80
Fig. 3. 13 - Capo Linaro - Ricostruzione dei Dati e Standard error	80
Fig. 3. 14 - Cetraro - Ricostruzione dei Dati e Standard error	81
Fig. 3. 15 - La Spezia - Ricostruzione dei Dati e Standard error.....	81
Fig. 3. 16 - Ponza - Ricostruzione dei Dati e Standard error.....	82
Fig. 3. 17 - Ancona - Ricostruzione dei Dati e Standard error	83
Fig. 3. 18 - Monopoli - Ricostruzione dei Dati e Standard error.....	83
Fig. 3. 19 - Ortona - Ricostruzione dei Dati e Standard error	84
Fig. 3. 20 – Punta della Maestra - Ricostruzione dei Dati e Standard error	84
Fig. 3. 21 - Catania - Ricostruzione dei Dati e Standard error	85
Fig. 3. 22 - Crotone - Ricostruzione dei Dati e Standard error.....	86
Fig. 4. 1 - Esempio di dati contenuti in un file HNE – il piano orizzontale individua la posizione della boa in coordinate Nord-Est; l'asse verticale è relativo alle elevazioni (altezze d'onda).	88

Fig. 4. 2 - Diagrammi HNE - Alghero.....	90
Fig. 4. 3 - Diagrammi HNE - Ancona	90
Fig. 4. 4 - Diagrammi HNE - Catania	91
Fig. 4. 5 - Diagrammi HNE – Capo Comino.....	91
Fig. 4. 6 - Diagrammi HNE - Cetraro.....	92
Fig. 4. 7 - Diagrammi HNE - Civitavecchia.....	92
Fig. 4. 8 - Diagrammi HNE - Crotone.....	93
Fig. 4. 9 - Diagrammi HNE – Punta della Maestra	93
Fig. 4. 10 - Diagrammi HNE - Mazara.....	94
Fig. 4. 11 - Diagrammi HNE - Monopoli.....	94
Fig. 4. 12 - Diagrammi HNE - Ortona.....	95
Fig. 4. 13 - Diagrammi HNE - Palermo	95
Fig. 4. 14 - Diagrammi HNE - Ponza.....	96
Fig. 4. 15 - Diagrammi HNE – La Spezia	96
Fig. 4. 16 - Distribuzione temporale dei valori recuperabili	98
Fig. 4. 17 - Qualità dei dati ricostruibili ad Ortona e Cetraro.....	99

Indice delle Tabele

Tab. 1. 1 - Le Boe e la loro posizione	20
Tab. 1. 2 – RON: Specifiche tecniche della boa TRIAXYS	20
Tab. 2. 1 - Limiti geografici dei settori costieri.....	23
Tab. 2. 2 - Definizione dei periodi e delle soglie di operatività per tutte le stazioni RON.....	25
Tab. 2. 3 - La scala Douglas	30
Tab. 2. 4 - Informazioni contenute nel file MAIN	33
Tab. 2. 5 - Valori eliminati di temperatura dell'acqua e altezza d'onda significativa.....	40
Tab. 2. 6 - Dati mancanti di temperatura dell'acqua (sopra) e di altezza d'onda significativa (sotto) nel dataset RON tra il 2002 e il 2006	42
Tab. 2. 7 - Numero di gap brevi nel dataset RON tra il 2002 e il 2006.....	45
Tab. 2. 8 – Suddivisione dei gap lunghi in 8 classi	48
Tab. 2. 9 - Efficienza percentuale dell'algorithmo MI in funzione del numero di imputazioni e della frazione di dati mancanti.....	60
Tab. 3. 1 - Suddivisione delle boe in base ai bacini	66
Tab. 3. 2 - Specifiche tecniche dei sensori di temperatura montati su mareografi (a) e su Boe (b)	69
Tab. 3. 3 - Correlazioni Boa-Mareografo	71
Tab. 3. 4 - Stazioni di misura nel Mar Tirreno	79
Tab. 3. 5 - Stazioni di misura nel Mar Adriatico	82
Tab. 3. 6 - Stazioni di misura nel Mar Ionio	85
Tab. 4. 1 - Numero di file HNE presenti nel database.....	89
Tab. 4. 2 - Numero dei dati recuperabili con grado di affidabilità medio per ogni boa e ogni anno di osservazione.....	97

Introduzione

Il presente lavoro è stato condotto nell'ambito dell'attività di stage prevista a conclusione del Master di secondo livello in Ingegneria ed Economia dell'Ambiente ed il Territorio dell'Università degli Studi Roma Tre grazie alla convenzione stipulata tra l'Ateneo e l'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA).

Lo studio, svolto presso il Servizio Mareografico (Dipartimento Tutela delle Acque Interne e Marine) dell'ISPRA rappresenta solo il primo passo di un'analisi che partendo dagli attuali dati misurati dalle stazioni delle reti di monitoraggio marine nazionale, potrà portare alla valutazione di importanti fenomeni strettamente legati con le dinamiche oceanografiche.

Progetto formativo

Il cambiamento del clima a livello mondiale ha suscitato un notevole interesse in alcuni Paesi, in primo luogo quelli dell'Unione Europea, che si stanno attrezzando per affrontare questa sfida, mettendo in campo politiche di riduzione delle emissioni di inquinanti e l'adeguamento delle strutture e degli stili di vita ai cambiamenti dell'ambiente e del territorio che inevitabilmente ci saranno.

E' stato scientificamente dimostrato come variazioni globali del clima anche molto contenute possano generare conseguenze estreme come generazione di uragani, desertificazione, innalzamenti dei livelli marini, ecc. Assume sempre maggiore importanza lo studio e l'analisi di dati sintetici che possano consentire la formulazioni di previsioni affidabili ad ampio raggio sia nello spazio che nel tempo. L'analisi delle caratteristiche fisiche dell'ambiente marino (in particolare temperature, livelli, dati ondometrici) costituisce un valido supporto allo studio degli effetti del global warming e consente di individuare scenari di eventi possibili nel caso di persistenza dei fenomeni anche sul lungo periodo. Le attuali Direttive Europee sulle Acque costituiscono un importante riferimento alla definizione di una valida strategia riguardante lo Sviluppo Sostenibile e la Protezione dell'Ambiente Marino, che si basa essenzialmente sul monitoraggio delle caratteristiche fisiche del sistema Mare.

Lo studio del bacino del Mediterraneo centrale risulta estremamente interessante a causa delle peculiarità che esso presenta (quali temperatura, salinità, carenza di scambi con l'esterno, ecc.): la reazione quindi ai cambiamenti climatici potrebbe risultare ben diversa dai bacini oceanici e generare impatti di natura ambientale e sociale difficilmente prevedibili.

L'individuazione di scenari futuri relativamente affidabili generati a partire dall'analisi delle serie storiche presenti negli archivi nazionali, può consentire la definizione di interventi di mitigazione da attuare in aree a rischio e la regolazione delle attività commerciali e turistiche che si sviluppano nelle zone costiere nazionali.

Obiettivi

Il presente lavoro ha lo scopo di analizzare le serie storiche presenti nei database del Settore Mareografico dell'ISPRA relative alle Reti di interesse strategico nazionale (RON "*Rete Ondametrica Nazionale*" e RMN "*Rete Mareografica Nazionale*") e di individuare un modello affidabile di ricostruzione dei numerosi dati mancanti. Esso potrà risultare un valido supporto alla valutazione degli effetti dei cambiamenti climatici sulle caratteristiche fisiche del bacino del Mediterraneo, ed all'individuazione di strategie di adattamento in specifiche aree a rischio.

Lo studio, partendo dall'analisi di un dataset estratto ad hoc di durata limitata a soli cinque anni (dal 1 gennaio 2002 al 31 dicembre 2006) per ridurre i tempi di calcolo, valuta le principali grandezze fisiche significative e la numerosità dei *missing values* e dei valori *anomali* dei dati mareografici e ondametrici nel periodo di riferimento. I dati della rete RON risultano estremamente frammentati a causa di interruzioni di servizio e di errori di trasmissioni. I dati relativi al clima di moto ondosso direzionale e alle temperature dell'acqua risultano essere di ridotta utilità per eventuali analisi ed elaborazioni. E' necessario quindi individuare un modello di ricostruzione dei dati mancanti preciso, flessibile ed efficace.

Grazie alla preziosa collaborazione del Prof. Francesco Lagona (Docente di Statistica presso l'Università degli Studi Roma Tre) è stato possibile implementare un modello che consente di applicare il metodo di *imputazione multipla* alle serie storiche di temperatura acquisite dalle reti RON e RMN.

I dati di altezza d'onda significativa presentano una notevole complessità dovuti alla presenza della componente direzionale difficilmente schematizzabile in modelli se non attraverso l'applicazione di metodi fisicamente basati. L'analisi riguarda quindi solo i dati "grezzi" (file HNE) ricevuti dalle stazioni della rete ondometrica presenti nei server ISPRA dai quali è possibile ricavare informazioni circa alcune tipologie di dati mancanti.

Il progetto si sviluppa quindi nelle seguenti fasi:

- Valutazione e descrizione delle caratteristiche delle reti RON e RMN;
- Valutazione e quantificazione dei dati mancanti nel database RON;
- Implementazione di un modello di ricostruzione dei dati mancanti delle serie storiche di temperatura dell'acqua nel periodo 2002-2006 e valutazione dei risultati;
- Analisi della quantità e qualità dei dati grezzi di altezza d'onda nel periodo 2002-2006.

Fonti di Dati

1.1 - La rete di monitoraggio dell'ISPRA

L'informazione ondametria e mareografica è fondamentale in tutti gli studi, i progetti, i lavori e l'esercizio di opere di ingegneria costiera e marittima, nonché nell'analisi dei relativi impatti ambientali.

Nei mari italiani, dove le maree raggiungono solo raramente ampiezze notevoli, lo studio degli stati di mare interessa l'idraulica e l'idrografia, ed ha ampi risvolti pratici nella conservazione della qualità delle acque lagunari e degli estuari nonché nella valutazione della sicurezza di zone litorali e soprattutto portuali, la progettazione e l'attuazione di interventi per il turismo, la nautica, la realizzazione di opere lungo la riva o in mare aperto, interventi di protezione civile, di messa in sicurezza, allerta della popolazione, etc [5].

L'acquisizione dei dati costituisce quindi un'esigenza irrinunciabile per una conoscenza accurata e completa dei parametri meteo-marini che definiscono istante per istante e punto per punto lo stato di mare.

Il tipo di dati disponibili condiziona sia la metodologia di esecuzione degli studi degli stati di mare che la qualità dei risultati. I requisiti ideali dei dati sono [12]:

- Alto contenuto di informazioni e possibilità di estrarre tutti i parametri caratteristici;
- Affidabilità e precisione;
- Buona copertura del periodo di rilievo;
- Durata estesa nel tempo (rappresentatività per il lungo termine);
- Rappresentatività geografica per il sito in esame.

Nessuna delle fonti usualmente disponibili possiede contemporaneamente tutti questi requisiti; è quindi necessario reperire, analizzare e confrontare i dati e, dove possibile, applicare algoritmi specifici per la ricostruzione dei missing values.

Le principali fonti di dati sono le più disparate: si passa dalla rete nazionale (ondametrica e mareografica) presenti ormai da diversi decenni sul territorio nazionale a fonti private o istituzionali installate per scopi ben precisi che vanno dalla ricerca alla progettazione ed esecuzione di strutture a riva o off-shore.

I dati risultano a volte di ridotta utilità per alcune elaborazioni, a causa dei periodi di funzionamento troppo brevi (spesso inferiori ai tre anni) e discontinui (per interruzioni di servizio, spesso in corrispondenza dei picchi di mareggiata), per errori di installazione ed affidabilità degli strumenti e delle misurazioni. In alcuni casi anche le modalità impiegate per l'acquisizione e l'analisi dei dati hanno compromesso l'affidabilità dei risultati ottenuti. In questo capitolo si descrivono le caratteristiche delle reti di monitoraggio nazionali gestite dall'ISPRA.



Fig. 1. 1 - La Rete Ondametrica Nazionale e la Rete Mareografica Nazionale

1.2 - La rete mareografica nazionale

1.2.1 - Cenni storici

La mareografia, scienza che studia i fenomeni fisici del mare, è da sempre considerata uno dei rami più importanti della oceanografia.

Nei nostri mari, dove solo raramente si verificano fenomeni di notevole ampiezza, lo studio delle maree è finalizzato essenzialmente alla risoluzione di problemi legati alla buona conservazione delle lagune, degli estuari e dei porti, il turismo, la nautica, etc.

Per questo motivo negli ultimi 100 anni in Italia si è data una grande rilevanza alle osservazioni mareografiche, a partire dal 1896, anno in cui iniziano le osservazioni sistematiche e strutturate in reti, fino ad oggi in cui la rete è gestita dal servizio Mareografico dell'ISPRA.

L'importanza che veniva riconosciuta all'osservazione mareografica era, ed è tuttora, indiscutibile. Essa infatti consente, da un lato, di stabilire la relazione che lega l'ampiezza della variazione del livello del mare alla grandezza della forzante, e dall'altro, di conoscere il ritardo del massimo dell'effetto sul massimo della causa [16].

La registrazione delle oscillazioni della superficie del mare, tracciata da un mareografo, permette la misura della distanza verticale compresa tra la superficie del mare e un piano orizzontale assegnato nei singoli istanti dell'intervallo di registrazione.



Fig. 1. 2 - RMN: Carta delle stazioni all'origine della rete

Questo piano è individuato da un caposaldo, esterno al mareografo, solidale con il terreno circostante allo strumento e individuato con precise tecniche topografiche.

Le serie temporali dei valori istantanei registrati dai mareografi permettono di studiare le leggi che regolano le oscillazioni del mare, dovute alle maree o ad altre cause, intorno al livello medio mare. Fin dal XIX secolo l'Associazione Geodetica Internazionale decise di adottare il livello medio del mare come superficie di riferimento per l'altimetria. Infatti dovendo procedere a un lavoro di livellazione geometrica di precisione, esteso a vaste regioni, era indispensabile appoggiarsi ad una superficie di riferimento geodeticamente sicura. L'altezza di un punto venne così riferita al geoide, superficie di equilibrio delle acque del mare sotto l'influenza della forza centrifuga e dell'attrazione di tutte le masse all'esterno ed all'interno della terra. Questo piano di riferimento può essere determinato in qualunque località per la quale si posseggano dati mareografici raccolti durante un periodo adeguato di anni. Il livello medio veniva comunemente segnato su un'apposita targa solidamente fissata su una superficie verticale, oppure individuato determinandone la differenza di livello rispetto ad un caposaldo comunque collocato. Da uno di tali capisaldi, che si trova nell'immediata vicinanza di un mareografo, venivano iniziate le operazioni di livellazione che dovevano svolgersi per determinare l'altezza di punti comunque lontani. Una rete di livellazione che si appoggi a diverse origini può essere inoltre compensata in modo da basare il calcolo delle altitudini su tutte le origini collegate.

Nella storia della mareografia in Italia, funzionarono per primi gli strumenti dell'Istituto Talassografico di Trieste (1859), di Venezia Punta della Salute (1871), di Rimini (1867), e dell'Istituto Idrografico della Marina Militare a Genova (1883). Poi fra il 1896 ed il 1897 iniziarono l'attività anche i mareografi di Imperia, Livorno, Civitavecchia, Napoli (Arsenale e Mandracchio), Messina, Palermo, Catania, Ancona, Ravenna Porto Corsini, Cagliari, e La Maddalena. Successivamente, nel 1920 fu dato inizio all'attività delle stazioni di Reggio Calabria e Vieste [16].

Sotto la gestione del Servizio Mareografico il complesso delle stazioni di misura mareografiche assunse la connotazione di una vera e propria rete razionalmente distribuita sul territorio. La localizzazione delle stazioni nacque da considerazioni sia di carattere cognitivo, al fine di ottenere elementi che consentissero di studiare e prevedere il fenomeno della marea lungo l'intera estensione delle coste italiane, sia di carattere prettamente operativo, ovvero conoscere l'andamento dei livelli del mare nei principali porti (o comunque in porti in cui ciò fosse di interesse) al fine di consentire il mantenimento dei fondali, per condurre in sicurezza le operazioni portuali e di determinare

le quote di progetto delle opere marittime, banchine o frangiflutti, o di verificare loro possibili assestamenti o cedimenti nel tempo. Materia ovviamente di grande interesse per chi, come il Ministero dei Lavori Pubblici sovrintende alla costruzione, manutenzione e sicurezza delle opere. Infatti le stazioni erano affidate alle locali sezioni operative degli Uffici del Genio Civile per le Opere Marittime, competenti per territorio, che provvedevano, una volta installata la stazione di concerto con il Servizio Mareografico, alla loro gestione e manutenzione. In questa fase temporale, le stazioni, per la maggior parte del tipo a cilindro verticale, costituite da un elemento prefabbricato appoggiato su di un imbasamento di pietrame opportunamente livellato, erano dotate di un mareografo a galleggiante. I primi strumenti mareografici, avevano una pennina o matita registratrice connessa ad un galleggiante, in modo da spostarsi lungo una guida rettilinea proporzionalmente alle escursioni verticali del galleggiante stesso che tracciava la curva mareografica su un diagramma o striscia di carta, fatta avanzare da un movimento ad orologeria in direzione perpendicolare alla guida. Un'altra pennina o matita fissa, o un dispositivo di altro genere, tracciava sul diagramma una retta detta linea di base che ne individuava lo "zero" [16].

In prossimità della cabina mareografica e all'interno di essa venivano materializzati un numero di capisaldi (in genere tre) piano altimetrici vincolati tra loro con operazioni di livellazione di alta precisione.

La grande attenzione posta alla regolazione strumentale e alla conservazione del livello di riferimento, ha condotto, nel tempo al rilevamento di serie temporali di marea di straordinaria precisione ed affidabilità. Purtroppo una grande parte del materiale cartaceo storico, soprattutto quello che si trovava presso la sede centrale del Servizio Mareografico del Ministero dei Lavori Pubblici è andata perduta e quindi di questa parte della storia della Rete Mareografica Nazionale non restano che le informazioni sintetiche elaborate e pubblicate sugli Annali Idrologici.

Nel 1986 la Rete Mareografica, ancora sotto la gestione del Ministero dei Lavori Pubblici, fu completamente ristrutturata. Il numero delle stazioni venne razionalizzato e ridotto a 25: Imperia, Genova, Livorno, Civitavecchia, Napoli, Salerno, Palinuro, Messina, Palermo, Porto Empedocle, Catania, Reggio Calabria, Crotone, Taranto, Otranto, Bari, Vieste, Ortona, Ancona, Ravenna, Venezia Diga Sud Lido, Trieste, Cagliari, Carloforte, Porto Torres, a cui sono da aggiungere le stazioni di Genova e Brindisi dell'Istituto Idrografico della Marina Militare, la stazione di Trieste dell'Istituto Talassografico e la stazione di Rimini gestita dall'Amministrazione Comunale. Nel contempo fu effettuata la sostituzione

di tutte le apparecchiature mareografiche meccaniche con strumenti elettronici a galleggiante a registrazione locale su supporto di memoria rimovibile alimentate a pannelli solari [16].

Da quando, nel 1989, con l'entrata in vigore della legge n.183 sulla difesa del suolo, le competenze in materia di mareografia sono state trasferite alla Presidenza del Consiglio dei Ministri, e nello specifico al Servizio Idrografico e Mareografico Nazionale, quest'ultimo ha gestito la Rete Mareografica con la collaborazione dei propri Uffici Compartimentali che provvedevano all'attivazione del servizio di manutenzione in caso di malfunzionamento e al ritiro periodico dei supporti rimovibili di memoria allo stato solido e al loro invio presso la Direzione Generale di Roma, dove i dati venivano decodificati, validati ed archiviati [16].

1.2.2 - Descrizione della rete

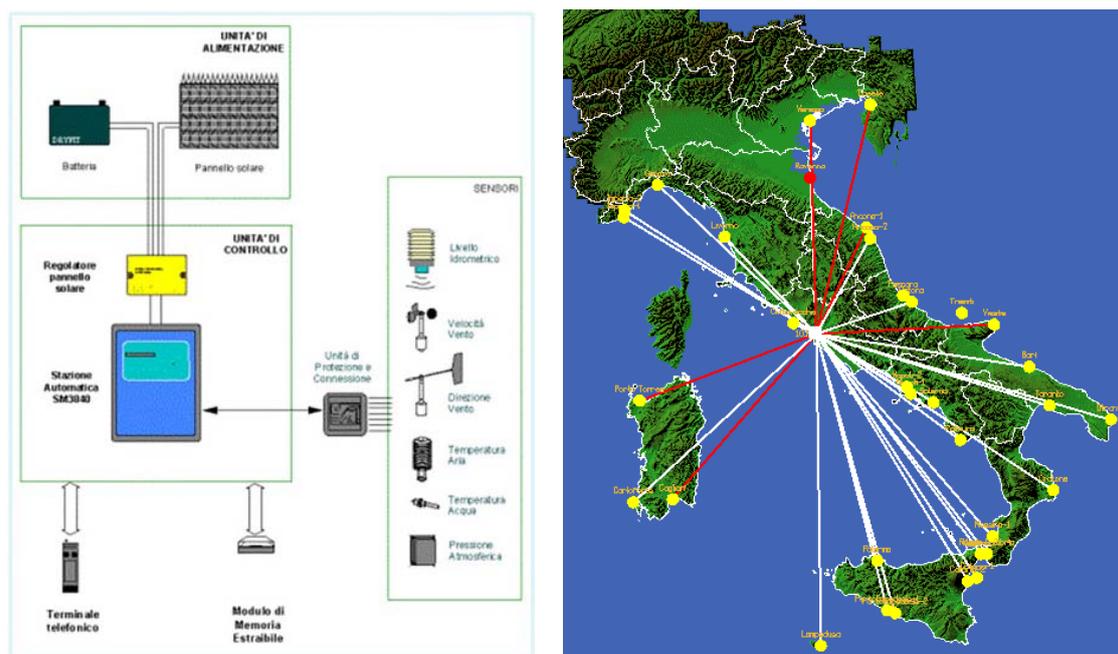


Fig. 1.3 – RMN: Sistema locale e sistema centrale

La Rete Mareografica Nazionale è composta di 26 nuove stazioni di misura uniformemente distribuite sul territorio nazionale ed ubicate prevalentemente all'interno delle strutture portuali: Trieste, Venezia Lido, Ancona, Ravenna, Pescara, Ortona, Isole Tremiti, Vieste, Bari, Otranto, Taranto, Crotona, Reggio Calabria, Messina, Catania, Porto Empedocle, Lampedusa, Palermo, Palinuro, Salerno, Napoli, Cagliari, Carloforte, Porto Torres,

Civitavecchia, Livorno, Genova ed Imperia. Le nuove stazioni che compongono la RMN sono composte da due strumenti di misura della marea: uno (principale) ad ultrasuoni con guida tubolare del fascio e compensazione di temperatura, ed uno (secondario) a galleggiante e registrazione cartacea per la verifica puntuale delle misure, l'analisi di eventi o fenomeni particolari ed il recupero dei dati nei casi di mancato funzionamento dello strumento principale. I mareografi sono riferiti ad un caposaldo quotato con livellazione di alta precisione, opportunamente verificato, sulla base del più vicino caposaldo IGM. Le stazioni sono dotate anche di un sensore anemometrico (velocità e direzione del vento a 10 metri dal suolo), di un sensore barometrico, di un sensore di temperatura dell'aria e di un sensore di temperatura dell'acqua. Tutte le stazioni sono dotate di un sistema locale di gestione e memorizzazione dei dati e di un apparato di trasmissione in tempo reale alla sede centrale dell'ISPRA a Roma. Da quando la nuova Rete Mareografica Nazionale è pienamente operativa, l'ISPRA mette a disposizione degli utenti informazioni aggiornate relative a serie storiche, osservazioni in tempo reale, previsione dei dati di marea astronomica, analisi dei dati a fini progettuali e scientifici. I dati mareografici e le costanti di marea locali aggiornate vengono pubblicati su un bollettino annuale [16].

La rete è composta da una serie di stazioni periferiche e da una centrale operativa con sede presso il Servizio Mareografico dell'ISPRA.

Le nuove stazioni periferiche acquisiscono le misure dai sensori installati, elaborano i dati rilevati e li memorizzano su una memoria locale e un supporto magnetico estraibile. Per svolgere tale attività sono dotate di una stazione di rilevazione automatica e da un set di sensori di misura.

La Stazione Meteorologica Automatica acquisisce i dati dai sensori a intervalli predefinitibili nella fase di configurazione (tramite la tastiera della stazione o direttamente dal "Front-End" di acquisizione). Oltre alle grandezze meteorologiche viene acquisito anche il valore della tensione della batteria.

I dati provenienti dai sensori analogici vengono convertiti in grandezze digitali da un convertitore A/D da 14 bit. L'uso di questo tipo di convertitore in luogo dei comuni convertitori da 12 bit consente una maggiore precisione della misura, oltre che una maggiore risoluzione [16].

La validazione dei dati acquisiti viene effettuata dalla stazione stessa, utilizzando algoritmi dipendenti dalla misura effettuata e dal sensore utilizzato. I dati considerati non validi vengono marcati e viene generato un allarme.

I dati validi vengono elaborati dalla stazione utilizzando i più comuni algoritmi statistici (valori medi, massimi e minimi giornalieri, etc.). L'intervallo di elaborazione dei parametri statistici può essere predefinito nella configurazione della stazione, e può essere diverso da sensore a sensore e, per lo stesso sensore, da funzione a funzione.

I dati acquisiti ed elaborati vengono memorizzati sia nella memoria RAM che equipaggia la stazione che nella Memory Card estraibile. Entrambe le memorizzazioni vengono gestite come archivi ciclici: in caso di esaurimento della memoria il dato più recente sostituisce quello più vecchio. La Memory Card garantisce la conservazione dei dati acquisiti dalla stazione in caso di problemi di comunicazione con il Front-End, per periodi molto lunghi, anche in caso di mancanza dell'alimentazione.

La comunicazione con il Front-End è bidirezionale e può essere gestita dal centro con procedura di chiamata ciclica o estemporanea delle stazioni, oppure direttamente dalle stazioni nei casi in cui sia necessario segnalare condizioni di allarme tecnico (malfunzionamenti) o operativo (superamento di soglie etc.). In caso di problemi nella comunicazione con il Centro i dati vengono memorizzati dalla stazione. Quando viene ristabilita la comunicazione possono essere automaticamente trasmessi tutti i dati memorizzati dall'ultima chiamata andata a buon fine [16].

La stazione è dotata di una porta seriale dedicata all'interfacciamento con dispositivi esterni (terminali di manutenzione o PC portatili) ed è provvista di un orologio-datario con calendario perpetuo e correzione automatica dell'anno bisestile. Ogni volta che il dispositivo entra in funzione la data e l'ora dell'intervento vengono registrate nella memoria della stazione e ritrasmesse al Front-End.

L'unità di alimentazione della stazione comprende un pannello solare da 50 W, un regolatore/caricabatterie, che regola la tensione generata dal pannello solare e gestisce la carica della batteria da 63 Ah.

Ciascuna stazione mareografica è dotata di un riferimento plano-altimetrico in prossimità della porta d'ingresso, facilmente accessibile, ma difficilmente asportabile, tale da poter determinare geometricamente la posizione dello "Ø" idrometrico. Ciascun riferimento è materializzato con un caposaldo le cui caratteristiche sono conformi a quelle indicate dall'Istituto Geografico Militare Italiano. Per ciascun caposaldo è determinata la posizione planimetrica in coordinate Gauss-Boaga e geografiche con allacciamento alla rete trigonometrica istituzionale IGMI e la posizione altimetrica in quota assoluta con allacciamento alla nuova rete altimetrica istituzionale IGMI, posteriore al 1942.

La strumentazione adottata per il rilevamento dei dati è costituita da [16]:

- sensori di livello ad ultrasuoni. Al fine di eliminare eventuali perturbazioni dovute alla presenza di altri apparati all'interno del pozzetto di calma il sensore è provvisto di una "guida d'onda" in materiale plastico che convoglia le onde ultrasonore in un fascio cilindrico. Il sensore è completamente controllato da un microprocessore che gestisce le principali funzioni dello strumento.
- sensori di velocità e direzione vento. La caratteristica più importante di tali sensori è il trasduttore elettro-ottico, che elimina i contatti striscianti di dinamo e potenziometri comunemente utilizzati. Le parti in movimento si riducono quindi al solo elemento sensibile (coppe e banderuola). Il risultato è un notevole aumento dell'affidabilità dei sensori. I sensori del vento sono installati su un palo anemometrico ribaltabile da 10m. Tale scelta ha i seguenti vantaggi: è un palo autoportante che non richiede controventatura, e quindi è installabile senza problemi nello spazio limitato normalmente disponibile in prossimità delle cabine mareografiche; grazie alla cerniera posta in prossimità della metà del palo la traiettoria di ribaltamento è un cerchio di 5 metri il cui centro posto a 5m di altezza: pertanto lo spazio libero richiesto a livello del terreno può essere molto limitato; la semplicità meccanica rende il palo compatibile con l'aggressività dell'ambiente marino: l'abbattimento richiede il movimento di una sola cerniera che può essere facilmente sbloccata anche in presenza di ossidazione, semplicemente grazie alla massa e al notevole braccio i leva della parte in movimento;
- sensore di temperatura dell'aria. Il sensore utilizza una termoresistenza Pt100 come elemento sensibile. Per evitare problemi legati alla caduta di tensione per effetto dei conduttori il collegamento è effettuato tramite 4 fili. Uno schermo in materiale plastico protegge l'elemento sensibile dalla radiazione solare. Il sensore di temperatura aria è installato su un apposito sostegno ricavato sul palo anemometrico, a un'altezza di 3,5 m circa.
- sensore di temperatura dell'acqua. Il sensore è di tipo Pt100, il materiale è acciaio inox. Il sensore è installato all'interno della cabina, calato tramite un cavo d'acciaio con un peso a un'estremità all'interno del pozzetto di calma, ad un'altezza di 1m inferiore al livello minimo di marea.
- sensore di pressione atmosferica. E' un sensore di classe molto elevata, utilizzato nelle applicazioni aeroportuali, che richiedono un'elevatissima precisione nella misura della pressione, necessaria per la taratura degli altimetri. Il sensore è dotato

di un circuito di compensazione che consente di limitare al massimo l'errore in temperatura (minore di 0,45 hPa tra 0 e 40 °C). Viene installato all'interno della cabina mareografica, nel contenitore in vetroresina che protegge le apparecchiature di comunicazione.

Tutte le parti che compongono la stazione sono in grado di funzionare nei campi di temperatura da -30 °C a +50 °C e di umidità dallo 0 al 100%. L'elettronica della stazione è installata, insieme al sensore di pressione atmosferica e al terminale telefonico, in una cabina in vetroresina munita di serratura a chiave.

La Rete Mareografica Nazionale è dotata di una unità centrale per:

- la configurazione ed il monitoraggio delle stazioni mareografiche interconnesse;
- l'acquisizione ed la validazione dei dati rilevati;
- l'archiviazione, l'elaborazione e la rappresentazione dei dati meteomarini registrati.



Fig. 1. 4 - RMN: gli strumenti in uso in un mareografo

La configurazione della rete consente la modifica di tutti i parametri caratteristici delle stazioni senza intervento diretto in loco. Ogni modifica nella configurazione della rete o delle stazioni viene memorizzata in un apposito archivio datato.

L'acquisizione dei dati dalle stazioni automatiche avviene tramite linea telefonica commutata e/o GSM o tramite la lettura della Memory Card estraibile che equipaggia le stazioni. La trasmissione dei dati può essere avviata secondo diverse procedure [16]:

- tramite un'interrogazione delle stazioni;
- in modo estemporaneo su richiesta dell'operatore che gestisce il Front-End;
- in modo estemporaneo su iniziativa della stazione periferica, in caso di segnalazione di allarmi tecnici o operativi.

Vengono inoltre trasmessi dalla stazione eventuali messaggi di errore, l'indicazione della tensione delle batterie, l'orario. In ogni momento è possibile visualizzare la situazione del processo di acquisizione; in particolare è possibile per ogni stazione visualizzare la data e l'ora dell'ultima chiamata e analizzare la relativa trasmissione, compresi i segnali di stato e

gli eventuali allarmi. Nei casi in cui problemi di comunicazione dovessero rendere impossibile il colloquio con la stazione anche per lunghi periodi, una volta ripristinato il collegamento il software consente l'aggiornamento degli archivi inserendo nel database tutti i dati mancanti.

L'archiviazione dei dati dipende dalla tipologia di dato trasmesso; essi infatti possono provenire [16]:

- in tempo reale dalle stazioni automatiche: si tratta di dati correnti raccolti automaticamente dal Front-End agli intervalli prefissati;
- dagli archivi RAM e/o Memory Card delle stazioni automatiche: sono raccolti automaticamente dal Front_End agli intervalli prefissati;
- dalle memory card: dati registrati sui moduli di memoria allo stato solido lette nell'apposita interfaccia del Front-End;
- da ingresso manuale: dati introdotti direttamente dall'operatore tramite interfaccia grafica.

I dati sono archiviati nel database in un particolare formato interno compatto contenente per ogni misura:

- riferimenti temporali: data e ora
- riferimento all'intervallo di elaborazione
- riferimento alla stazione
- riferimento al sensore
- riferimento al tipo: corrente ciclico, estemporaneo, di archivio, manuale, da memoria asportabile od elaborato internamente al centro validità.

I dati presentati in forma tabellare o grafica sono aggiornati automaticamente man mano che arrivano dalla periferia o vengono elaborati internamente.

1.3 - La rete ondamettrica nazionale

La Rete Ondamettrica Nazionale è stata installata sotto la tutela della Direzione Generale delle Opere Marittime del Ministero dei Lavori Pubblici nel giugno 1989. Del primo set di strumenti di misura (Fig. 1.5) facevano parte otto boe direzionali tipo pitch-roll Datawell MKI con le relative stazioni locali di acquisizione ed un centro di acquisizione a Milano dove i dati venivano scaricati trimestralmente [16].

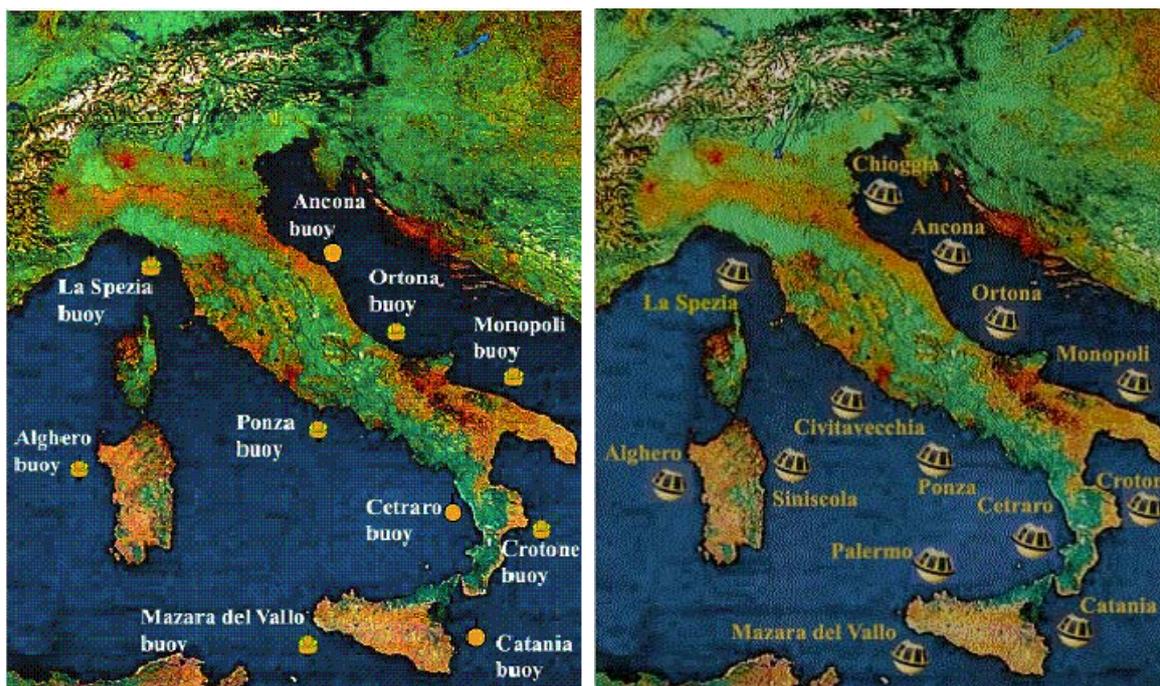


Fig. 1. 5 - RON: Periodo 1 e Periodo 2

La Rete Ondamettrica, così come quella Mareografica è passata nel 1996 sotto l'amministrazione del Dipartimento per i Servizi Tecnici Nazionali – Servizio Idrografico e Mareografico Nazionale (DSTN-SIMN) nell'ambito della Presidenza del Consiglio dei Ministri. La centrale di acquisizione dei dati è stata spostata a Roma e sono iniziati i progetti per il consolidamento ed il potenziamento delle reti.

Nel 2003 tutta la struttura tecnico-amministrativa del DSTN associata alle reti marine è transitata nell'Agenzia per la Protezione dell'Ambiente e per i servizi Tecnici (APAT) ora Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA).

1.3.1 - La rete ondamettrica nazionale dal 1989 al 1999

La Rete Ondamettrica Nazionale è attiva dal luglio 1989. La RON originaria era composta da otto boe direzionali di tipo pitch-roll, dislocate al largo di La Spezia, Alghero, Ortona,

Ponza, Monopoli, Crotona, Catania e Mazara. Ogni boa, ancorata su fondali dell'ordine di 100 metri, segue il movimento della superficie dell'acqua e permette di determinare l'altezza e la direzione delle onde. Gli strumenti sono dotati di un sistema di localizzazione che utilizza il satellite ARGOS per il controllo continuo della posizione [16].

1.3.2 - La rete ondometrica nazionale dal 1999 al 2002.

Nel 1999 sono state aggiunte alla rete preesistente le boe a traslazione di Cetraro ed Ancona ed è stata sostituita con una ulteriore boa a traslazione la boa pitch-roll di Catania. La Rete Ondometrica Nazionale è quindi configurata su dieci stazioni di misura, di cui sette costituite da una boa direzionale Datawell-Wavec di tipo pitch-roll e tre costituite da una boa direzionale Datawell-Waverider di tipo a traslazione. Le stazioni sono completate da un centro di ricezione ed elaborazione a terra dei dati inviati via radio dalla boa.

La boa Datawell-Wavec segue il movimento della superficie del mare e mediante il sensore Hippy 120 ne misura l'elevazione e le inclinazioni riferite ad una piattaforma inerziale, contenuta all'interno del sensore, che individua il piano di riferimento orizzontale. Per le boe Wavec l'elevazione è ottenuta dalla doppia integrazione dell'accelerazione verticale misurata da un accelerometro montato sulla piattaforma. Le inclinazioni sono ottenute dalla misura del seno degli angoli formati tra gli assi x e y del sistema di riferimento solidale con la boa ed il piano di riferimento orizzontale. Un sistema di bussole misura le componenti del campo magnetico lungo gli assi del sistema di riferimento solidale con la boa: H_x , H_y , H_z . Da tali componenti del campo magnetico e dalle suddette inclinazioni si ricavano: il valore assoluto e l'inclinazione del campo magnetico terrestre, l'orientamento della boa e le due pendenze locali della superficie dell'acqua nelle due direzioni nord-sud ed est-ovest, che consentono di definire la direzione di propagazione delle onde [16].

La boa Datawell-Waverider misura il moto ondoso mediante il sensore Hippy 40. Una bussola inerziale misura le componenti del campo magnetico H_x e H_y lungo gli assi x e y del sistema di riferimento solidale con la boa, e la direzione dell'asse z, perpendicolare a x e y. In tal modo la posizione della boa rispetto al sistema fisso di coordinate nord, ovest e verticale è completamente determinato. L'elevazione è ottenuta dalla doppia integrazione dell'accelerazione verticale misurata da un accelerometro montato su una piattaforma stabilizzata. Dalle accelerazioni misurate per mezzo di altri due accelerometri lungo le direzioni x e y solidali con il riferimento mobile della boa vengono calcolate le accelerazioni lungo gli assi fissi orizzontali nord ed ovest. Le tre accelerazioni sono poi

sottoposte a doppia integrazione fino ad ottenere i corrispondenti spostamenti e opportunamente filtrate. I valori misurati: elevazione, inclinazioni, H_x , H_y , H_z per le boe tipo pitch-roll ed elevazione e spostamenti in direzione nord e ovest per le boe a traslazione, sono trasmessi, via radio, in continuo dalla boa e vengono ricevuti nella stazione a terra. Il ricevitore esegue il controllo sulla qualità dei dati e li pre-elabora. I dati vengono acquisiti normalmente per periodi di 30 minuti ogni tre ore e con intervalli inferiori nel caso di mareggiate particolarmente significative [16].

Dalle tre serie temporali di elevazione, pendenza nord-sud e pendenza est-ovest per le boe Datawell-Wavec e di elevazione, spostamento nord e spostamento ovest per le boe Datawell-Waverider, vengono ottenuti [16]:

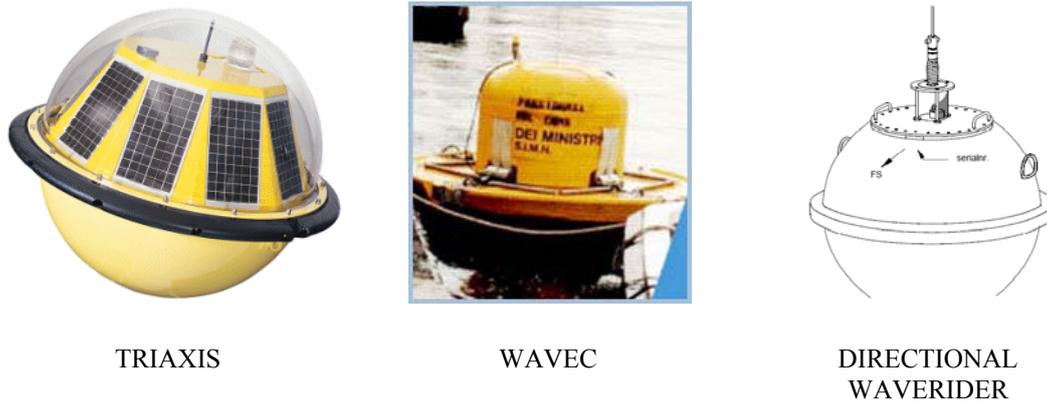
1) I parametri sintetici:

- H_s (metri) altezza d'onda significativa spettrale;
- T_p (secondi) periodo di picco;
- T_m (secondi) periodo medio;
- D_m (gradi N) direzione media di propagazione.

2) I parametri spettrali per banda di frequenza:

- densità di energia;
- direzione media di propagazione;
- dispersione direzionale (spread);
- asimmetria (skewness);
- curtosi.

Nei casi di mareggiate particolarmente significative, con valori di H_s superiori a 3,0 m per Ancona, Pescara, Monopoli, Crotona e Catania, 4,5 m per La Spezia, Ponza, Cetraro e Mazara del Vallo e 5,0 m per Alghero, l'acquisizione dei dati avviene automaticamente in continuo e i dati sono prodotti ogni mezz'ora. Nella stazione a terra, al ricevitore sono collegati due calcolatori mediante i quali si acquisiscono i dati grezzi trasmessi dalla boa e i dati elaborati, e si trasmettono sintesi dei dati al Centro di controllo e gestione della rete presso la Direzione del Servizio Idrografico e Mareografico Nazionale.



TRIAxis

WAVEC

DIRECTIONAL WAVERIDER

Fig. 1. 6 - RON: Le Boe utilizzate nella rete nel corso degli anni



Fig. 1. 7 - Numero di eventi significativi registrati dalle boe fino al 2003

1.3.3 - La rete ondometrica nazionale dopo il 2002

All'inizio del 2002 sono iniziate le attività di potenziamento della RON volte alla realizzazione di un sistema affidabile di monitoraggio e diffusione dei dati in tempo reale. Dal punto di vista dei parametri osservati non ci saranno variazioni rispetto alle grandezze misurate dal 1989, tranne per l'aggiunta delle elaborazioni nel dominio del tempo (analisi zero-crossing). Minori differenze sono in gran parte dovute alla migliore affidabilità dei dati conseguente all'incremento delle capacità di elaborazione del processore sulla boa, all'uso di sensori allo stato solido e all'aumento della frequenza di campionamento dei

sensori. Gran parte delle modifiche della rete è stata orientata a migliorare l'efficienza generale del sistema, dotando le boe di dispositivi di data-logging per recuperare i dati non trasmessi a terra, potenziando sensibilmente i centri locali di ricezione con elaboratori di tipo industriale e dotando il sistema di warning per i disormeggi di un sistema automatico di posizionamento GPS con trasmissione satellitare della posizione via Inmarsat D+. Il numero delle stazioni è stato infine portato a 14, aggiungendo 4 nuove boe posizionate a Capo Linaro (Civitavecchia, Tirreno Centrale), Capo Gallo (Palermo, Sicilia), Punta della Maestra (alto Adriatico) e Capo Comino (Sardegna Orientale). Gli ondometri direzionali impiegati nella RON dal terzo trimestre 2002 sono tutti del tipo TRIAXYS (Tab. 1.2 e Fig 1.7) [7], fabbricati dalla ditta canadese Axys Technologies Inc. modificati per rispondere alle particolari esigenze della RON. La rete è comunque in grado di poter gestire tutte le differenti tipologie di ondometro impiegate correntemente in Italia. Attualmente è in fase di realizzazione la nuova rete caratterizzate dalla presenza di nuove boe [16].

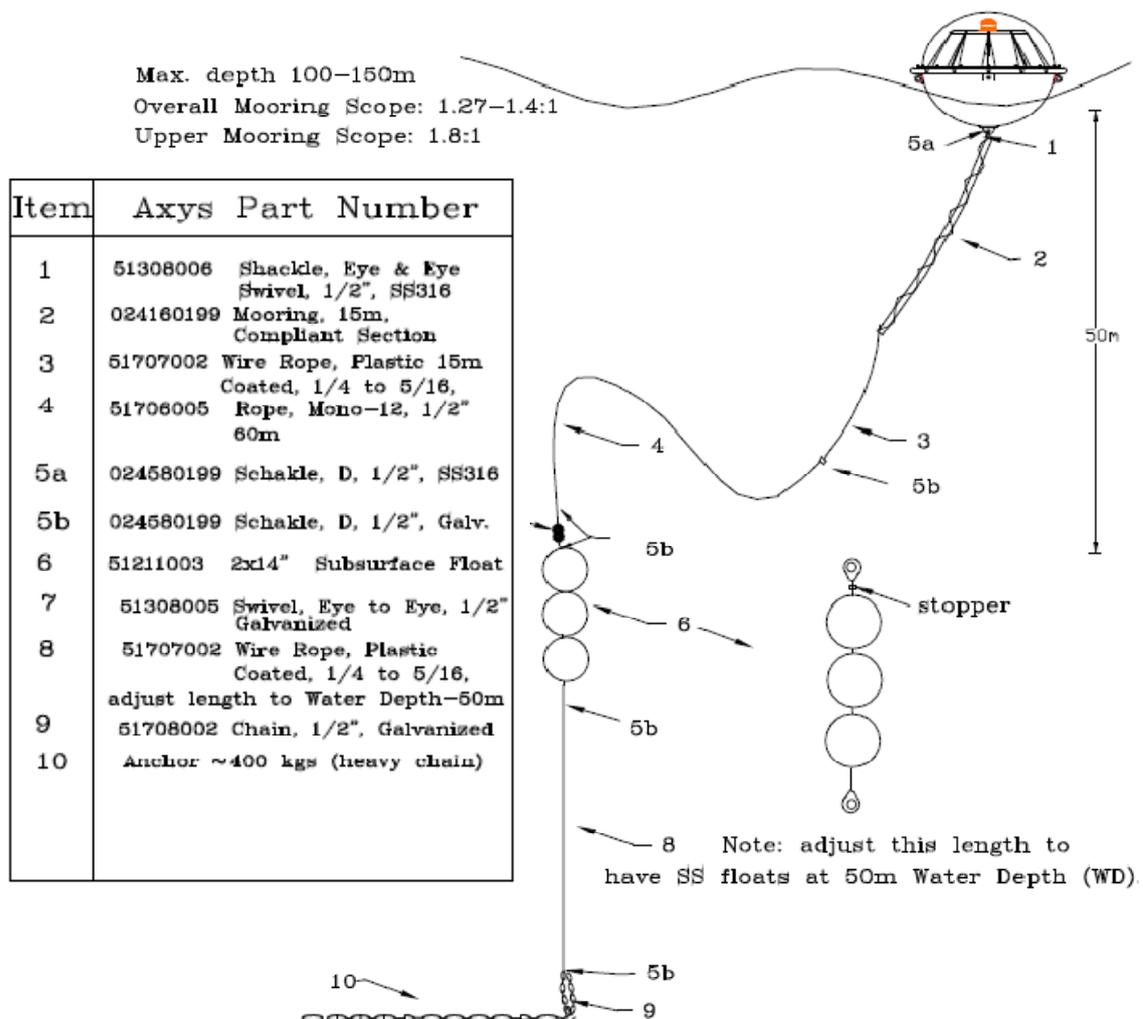


Fig. 1. 8 - RON: Caratteristiche in acqua della boa TRIAXYS

	CODICE WMO	LATITUDINE	LONGITUDINE	PROFONDITA'
CATANIA	61 207	37°26'.4N	15°08'.8E	70 m
MAZARA	61 208	37°31'.0N	12°32'.0E	87 m
PALERMO	61 209	38°15'.5N	13°20'.0E	125 m
CROTONE	61 210	39°01'.4N	17°13'.2E	95 m
CETRARO	61 211	39°27'.2N	15°55'.1E	110 m
C.po COMINO	61 212	40°37'.0N	09°53'.5E	115 m
ALGHERO	61 213	40°32'.1N	08°06'.0E	95 m
PONZA	61 214	40°52'.0N	12°57'.0E	100 m
MONOPOLI	61 215	40°58'.5N	17°22'.6E	90 m
CIVITAVECCHIA	61 216	42°00'.0N	11°46'.6E	92 m
ORTONA	61 217	42°24'.4N	14°32'.2E	70 m
ANCONA	61 218	43°49'.9N	13°42'.6E	70 m
LA SPEZIA	61 219	43°55'.7N	09°49'.6E	90 m
P.ta MAESTRA	61 220	44°58'.0N	12°38'.0E	30 m

Tab. 1. 1 - Le Boe e la loro posizione

Diameter		0.9m (36")	
Weight (Including 4 Batteries)		197 Kg (435 lbs)	
Weight (without Batteries)		90 Kg (199 lbs)	
Sensors		3 accelerometers; 3 rate gyros; 1 flux-gate compass	
HEAVE	Range	± 20 m	
	Resolution	better than 2%	
	Accuracy	1 cm	
	Period	1.6 - 30 seconds	
DIRECTION	Range	0-360°	
	Accuracy	± 1°	
	Period	1.6 - 30 seconds	
SEA SURFACE TEMPERATURE	Range	-5 to +50°C	
	Accuracy	0.1°C	
	Operating Temperature Range	-30 to + 60°C	
	Storage/Transit Temperature	-40 to +70°C	
REAL TIME DATA PROGRAMMABLE CONFIGURATION	Obstruction Light	Amber, 3 nautical mile visibility, ODAS sequence	
	Sampling Frequency	4Hz	
	Standard Data Transmission	Multiple transmissions of each data set; Directional wave characteristics and spectra; Extensive wave statistics; Sea surface temperature; Battery voltage; Solar current; Error checking	
	Transmission Options	VHF or ARGOS	
	VHF Transmitter	30-39 MHz or 39-50MHz; synthesized frequency selection	
	Transmitter Power	3 watts @ 12 VDC; VHF Range up to 10 nautical miles	
	Batteries	4 gel cells	
	Solar Panels:	10 @ 6 watts	
	Position:	GPS/ARGOS (optional)	

Tab. 1. 2 – RON: Specifiche tecniche della boa TRIAXYS

Missing Values nel database RON

La Rete Ondametrica Nazionale ha generato tra il 1999 ed il 2008 una enorme quantità di dati per lo più organizzati in forma di file e tabelle di database. Il database conta alcune decine di milioni di record al momento (fino al 31 marzo 2008).

In questo capitolo vengono descritte le tipologie dei dati disponibili e le modalità di acquisizione ed elaborazione. Vengono inoltre classificati i dati mancanti rilevati tra il 2002 e il 2006, periodo in cui la rete, completamente rinnovata in tutte le sue componenti, ha funzionato con continuità pur presentando un notevole numero di missing values.

2.1 - I settori di costa.

La Rete Ondametrica Nazionale ha il compito di indagare lo stato di mare dei bacini italiani. La disposizione delle boe è quindi funzionale a tale scopo: i dati relativi ad ogni boa sono rappresentativi della zona di mare limitrofa e dei litorali sottesi [4].

Il criterio per la definizione dei “tratti costieri” corrispondenti alle boe è che il clima ondoso misurato al largo possa essere utilmente trasportato a costa con metodi numerici che correggano la geometria della propagazione considerando fenomeni come lo shoaling, la rifrazione, la riflessione, etc.

Dato l’orientamento geografico delle coste in relazione alle posizioni delle boe appare evidente che esistono tratti per i quali il criterio descritto non è applicabile; ad esempio tutto il lato nord della Sardegna è completamente schermato rispetto alle direzioni di provenienza delle onde che vengono osservate nelle due boe di Alghero e Capo Comino. In tali tratti ci si può aspettare di dover ricorrere a metodi puramente numerici (es. modelli numerici tipo Wave Model o Wavewatch) per stimare il clima sia al largo che a costa [4].

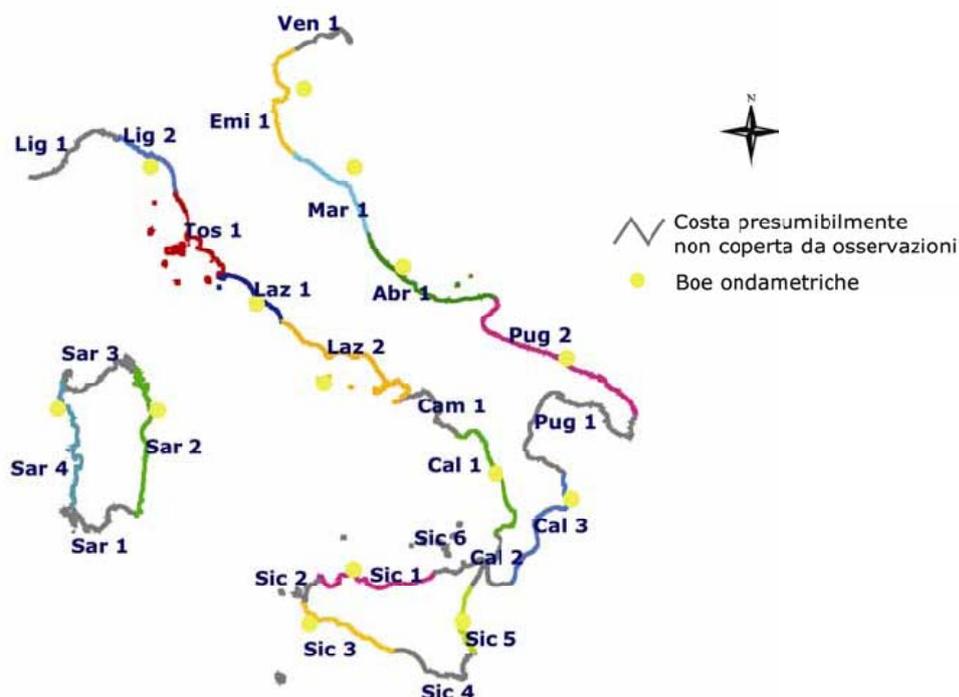
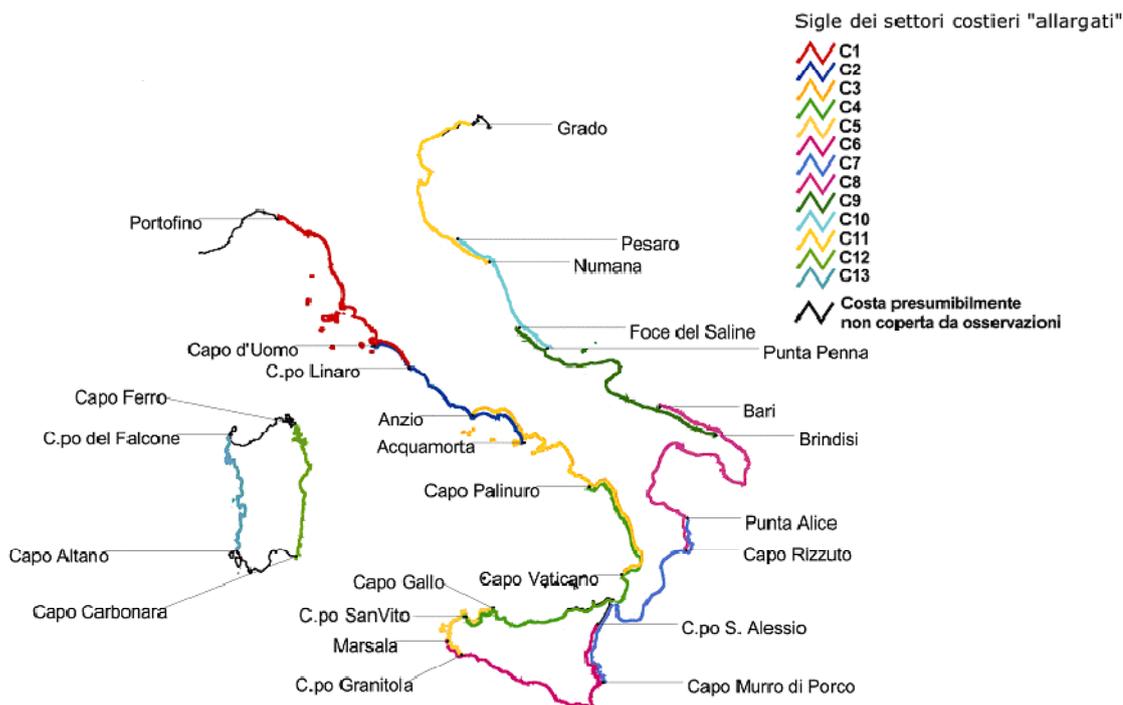


Fig. 2. 1 - Tratti di costa individuati in base alla posizione geografica di ciascun ondometro e di criteri di omogeneità e orientamento del litorale

Tutta la costa italiana è stata suddivisa in tratti costieri (Fig. 2. 1). I tratti monitorati sono: Lig-2, Laz-1, Laz-2, sar-2, Sar-4, Cal-1, Sic-1, Sic-3, Sic-5, Cal-3, Pug-2, Abr-1, Mar-1, Emi-1. I restanti tratti, non coperti da boe, sono: Lig-1, Tos-1, Cam-1, Cal-2, Sic-6, Sic-2, Pug-1, Ven-1.

Mentre quasi tutto l'Adriatico è coperto in modo uniforme, molte zone scoperte sono presenti nel Mare Tirreno, e nello Ionio [4].

Dall'analisi dei climi ondosi emerge tuttavia la presenza di notevoli affinità nelle caratteristiche direzionali di diversi tratti costieri non necessariamente vicini. E' stata dunque introdotta una categoria di elementi costieri più grandi, chiamati "settori costieri", all'interno dei quali è possibile confrontare i climi ondosi di tratti differenti. L'idea è che dalle caratteristiche climatiche comuni a due tratti di costa coperti da una boa, separati da un tratto non rappresentabile da nessuno strumento vicino, si possano in qualche modo trarre delle informazioni utili per fare ipotesi sul possibile clima ondoso nel tratto scoperto [4]. Sono stati definiti 13 settori costieri codificati da C1 a C13, rappresentati in Fig. 2. 2. La descrizione dei 13 elementi in termini di delimitazione geografica, codice identificativo, presenza di boe ondametrische e tratti costieri corrispondenti è presentata nella Tab. 2. 1. I limiti geografici di ciascun settore sono stati scelti a partire dalla sovrapposizione di tratti di costa monitorati da ondometro RON e da criteri di omogeneità e orientamento del litorale [4].



Settore Costiero	Boe rappresentative	Sigla	Tratti costieri
Portofino - C.po Linaro	La Spezia - Civitavecchia	C1	Lig2-Tos1-Laz1
Capo d'Uomo - Acquamorta	Civitavecchia - Ponza	C2	Laz1-Laz2
Anzio - C.po Vaticano	Ponza - Cetraro	C3	Laz2-Cam1-Cal1
C.po Palinuro - C.po S.Vito	Cetraro - Palermo	C4	Cal1-Cal2-Sic1-Sic6
C.po Gallo - C.po Granitola	Palermo - Mazara del Vallo	C5	Sic1-Sic2-Sic3
Marsala - C.po S.Alessio	Mazara del Vallo - Catania	C6	Sic3-Sic4-Sic5
C.po Murro di Porco - P.ta Alice	Catania - Crotona	C7	Sic5-Sic6-Cal3
C.po Rizzuto - Bari	Crotona - Monopoli	C8	Cal3-Pug1-Pug2
Brindisi - Foce del Saline	Monopoli - Ortona	C9	Pug2-Abr1
P.ta Penna - Pesaro	Ortona - Ancona	C10	Abr1-Mar1
Numana - Grado	Ancona - Chioggia	C11	Mar1-Emi1-Ven1
C.po Carbonara - C.po Ferro	Siniscola	C12	Sar1-Sar2
C.po Ferro - C.po Spartivento Sardo	Alghero	C13	Sar3-Sar4-Sar5

Tab. 2. 1 - Limiti geografici dei settori costieri

2.2 - Tipologie di dati acquisiti dalla rete ondametria

Il sistema di archiviazione della RON era stato originariamente progettato per fornire un insieme omogeneo (dataset statistico triorario) di dati di altezza significativa (H_{m0}), direzione media totale di provenienza (α), periodo di picco (T_p), periodo medio (T_m), relativi al periodo di acquisizione di 30 minuti con periodo di campionamento triorario, da

allegare alla pubblicazione del Bollettino Ondametrico Trimestrale. Quando lo stato del mare era caratterizzato da H_{m0} eccedenti un livello di soglia $S1$, nel bollettino venivano anche pubblicati i valori dei parametri spettrali dell'analisi di Fourier su 30 minuti delle serie originali di *Heave*, *Pitch*, *Roll* con campionamento di 0,78 secondi.

In aggiunta a questo archivio, venivano regolarmente elaborati i files contenenti tutte le informazioni relative alla elaborazione dei dati, tra cui i parametri spettrali e 20 minuti di tutte le serie temporali misurate dalla boa a 0,78 secondi. In condizioni di mare calmo questi files completi erano registrati ogni tre ore [4].

Quando le condizioni del mare peggioravano, a partire da un livello di soglia $S2$, il sistema passava al regime di acquisizione ogni trenta minuti. In questa modalità i dati sono registrati in continuo (la durata della misura è di circa trenta minuti quindi il sistema interrompe le misurazioni solo per trasmettere a terra i dati). Le misure in continuo venivano archiviate separatamente.

I grandi progressi nelle capacità di elaborazione dei sistemi di calcolo e nella trasmissione dei dati hanno prodotto, in quindici anni, una significativa evoluzione dei sistemi di rilevamento in mare aperto a cui la RON si è adeguata, attraversando successive fasi di aggiornamento tecnologico: nel 1996 è stato introdotto il monitoraggio della temperatura superficiale dell'acqua (*Tmp*), nel 1999, alle otto stazioni originali sono state aggiunte due nuove stazioni con boe accelerometriche direzionali Datawell tipo Waverider in grado di operare in telemisura [4].

Nel 2002 l'intera rete è stata coinvolta in un impegnativo progetto di potenziamento finalizzato al monitoraggio in tempo reale dello stato del mare lungo le coste italiane. Sono state introdotte 4 ulteriori stazioni di misura, aggiornati i sistemi di trasmissione ed acquisizione dei dati, e tutta la precedente strumentazione è stata sostituita con nuove boe accelerometriche Axys di tipo Triaxys.

Questi cambiamenti hanno avuto notevoli conseguenze nelle qualità e nella quantità delle misure registrate successivamente. La qualità dei dati è certamente migliorata grazie alle aumentate capacità di elaborazione e di misura delle boe, mentre l'introduzione del sistema di telemisura e di sei nuove stazioni ha permesso un miglior controllo del sistema, sebbene lo abbia reso più complesso ed in qualche misura più difficile da gestire.

Durante il periodo 1989-2002 (periodo I) la RON ha funzionato con due diversi regimi operativi: nei periodi in cui lo strumento misurava altezze significative inferiori ad un livello di soglia fissato, la stazione registrava solo i dati con cadenza trioraria, mentre quando le misure superavano il livello di soglia, l'acquisizione diventava continua e i dati

venivano registrati con cadenza semioraria. In passato gran parte delle elaborazioni realizzate a partire dai dati della RON sono state basate sui dati triorari, anche a causa della non facile procedura necessaria per estrarre le informazioni semiorarie dall'archivio [4].

Grazie all'evoluzione dei sistemi di elaborazione, trasmissione ed archiviazione dei dati, il nuovo sistema di misura della RON opera dal 2002 (periodo II) esclusivamente in modalità continua. Questo permette di acquisire informazioni continue su base semioraria con due vantaggi principali: il primo riguarda una maggiore semplicità nell'analisi statistica dell'insieme dei dati rispetto al caso di dati cronologicamente disomogenei, il secondo è dato da migliore attenzione dedicata allo studio delle fasi crescenti e decrescenti delle mareggiate e dei fenomeni ondosi associati a periodi più lunghi in assenza di onde di vento (infragravity waves) [4].

Stazione	Periodo I	Periodo II	S I (m)	S II (m)
Alghero	01/06/1989 – 15/06/2002	15/06/2002 – marzo 2008	3,0	5,0
Ancona	10/03/1999 – 05/08/2002	05/08/2002 – marzo 2008	2,0	3,0
Siniscola		01/07/2002 – marzo 2008	3,0	4,5
Catania	01/06/1989 – 25/01/2002	25/01/2002 – marzo 2008	2,0	3,0
Cetraro	28/02/1999 – 08/02/2002	08/02/2002 – marzo 2008	3,0	4,5
Civitavecchia		01/07/2002 – marzo 2008	3,0	4,5
Crotone	01/06/1989 – 08/02/2002	08/02/2002 – marzo 2008	2,5	3,0
La Spezia	01/06/1989 – 12/07/2002	12/07/2002 – marzo 2008	2,5	4,5
Mazara	01/06/1989 – 27/03/2002	27/03/2002 – marzo 2008	3,0	4,5
Monopoli	01/06/1989 – 27/06/2002	27/06/2002 – marzo 2008	2,0	3,0
Palermo		01/07/2002 – marzo 2008	3,0	4,5
Ortona	01/06/1989 – 10/04/2002	10/04/2002 – marzo 2008	2,0	3,0
Ponza	01/06/1989 – 15/07/2002	15/07/2002 – marzo 2008	3,0	4,5
Chioggia		01/07/2002 – marzo 2008	2,0	3,0

Tab. 2. 2 - Definizione dei periodi e delle soglie di operatività per tutte le stazioni RON

2.3.1 - *La temperatura dell'acqua*

La Temperatura Superficiale del mare è una variabile di rilevante interesse scientifico e di grosso impatto socio-economico, poiché è un parametro chiave nella determinazione dei flussi di calore fra mare ed atmosfera responsabili dei principali fenomeni meteorologici e climatici a varie scale. El Niño è forse l'esempio più conosciuto di come il cambiamento della temperatura superficiale del mare incida direttamente sul clima e sulle attività economiche legate all'ambiente (ad esempio nei cicli biologici).

L'oceano è un'immensa riserva di calore. L'osservazione precisa della SST come variabile su lunghi periodi di tempo è, quindi, critica per lo sviluppo di modelli climatici; su base operativa, lo è anche per migliorare le previsioni meteorologiche a varie scale. Le osservazioni *in situ* (tipicamente da boe) sono ad oggi poche e disomogenee, ma necessarie, non solo per avere una stima precisa del dato puntuale di temperatura dell'acqua, ma anche per calibrare le osservazioni remote, come le misure satellitari, che vengono largamente utilizzate per indagare le dinamiche climatiche e marine.

Le boe ondometriche sono gli strumenti di riferimento per le misure dirette (*in situ*) di temperatura marina ed effettuano “misure di superficie” tramite sensori termometrici posti ad una profondità pari a 45 centimetri (pari alla profondità del fondo della boa su cui è posto il sensore), dando la cosiddetta *bulk temperature* (T_{bulk}). La differenza tra la *skin temperature* (temperatura riferita ai primi micron di profondità, usualmente misurata tramite osservazioni remote) e la T_{bulk} è pari a circa 0,2 K durante le ore notturne (T_{skin} minore di T_{bulk}), ma può arrivare ad alcuni gradi centigradi durante il giorno in condizioni di intenso irraggiamento solare (T_{skin} maggiore di T_{bulk}) [21].

Tale differenza è dovuta alla concomitanza dei due fenomeni rappresentati schematicamente in Fig. 2. 3: il raffreddamento superficiale (profili A e B), ed il riscaldamento diurno (profilo B).

Il primo fenomeno, dovuto agli scambi di calore attraverso l'interfaccia aria-acqua, causa usualmente differenze modeste (comprese tra 0,1 e 0,5°C): la temperatura dei primi millimetri del mare è minore di quella di qualche millimetro più profonda. Questo fenomeno è presente durante tutto l'arco della giornata. Il secondo fenomeno, oltre ad originare variazioni maggiori (può arrivare fino a 3°C), è molto più complesso da modellizzare. In questo caso, gli strati più superficiali del mare si riscaldano maggiormente a causa dell'irraggiamento solare (ovviamente solo nelle ore diurne) [21].

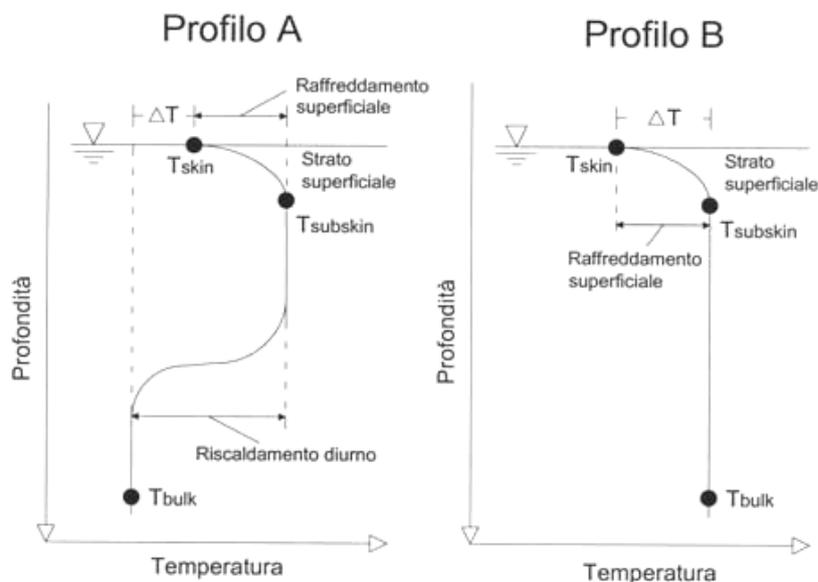


Fig. 2.3 - Rappresentazione schematica dei profili di temperatura del mare in funzione della profondità degli strati superficiali inferiori al metro di profondità.

L'entità del riscaldamento indotto da tale fenomeno e la profondità dello strato da esso interessato, dipende non solo dall'irraggiamento solare, ma anche dai meccanismi di rimescolamento indotti dalla turbolenza che caratterizza lo strato limite marino. Man mano che si scende in profondità, l'influenza dell'irraggiamento solare diminuisce e la temperatura che tende a non subire grandi escursioni nell'arco della giornata, varia quasi esclusivamente per il solo effetto di trasmissione di calore tra gli strati superficiali e quelli sottostanti.

Risulta quindi più facile confrontare le osservazioni ondametriche con i dati mareografici di temperatura dell'acqua. Infatti, questo tipo di misurazioni in situ avviene tramite sensori termometrici posti ad una profondità di circa 2 metri. Nonostante l'ubicazione non ideale dei mareografi (spesso posizionati in aree portuali chiuse) i dati di temperatura dell'acqua risultano essere numerosi e continui nel tempo e quindi estremamente affidabili per la ricostruzione dei dati mancanti della rete ondometrica. A questi dati è possibile accostare anche le misure di temperatura dell'aria (ricavabili sempre dalle stazioni mareografiche) che mostrano una grande variabilità giornaliera e stagionale, ma che rappresentano un ottimo indicatore dell'irraggiamento solare.

Nel Bacino del Mediterraneo, inoltre, i valori di temperatura misurati a profondità paragonabili a quelle dei mareografi risultano essere dei predittori particolarmente affidabili poiché, contrariamente a quanto succede nei bacini oceanici, il gradiente verticale di temperatura è particolarmente basso.

Infatti, mentre gli oceani, anche quelli con temperature superficiali elevate (26-30°), presentano temperature molto basse già a 200/300m di profondità (con valori che si attestano intorno 0° o di poco superiori anche a latitudini sub-equatoriali), il Mediterraneo invece, pur presentando fosse come quella delle Baleari o quella del Tirreno profonde fino a 4000m, non ha temperature che scendono sotto i +13/14°. Questo dipende dal fatto che con l'Atlantico non c'è una comunicazione profonda per la presenza della soglia di Gibilterra che arriva fino a 320m dalla superficie. Poiché l'acqua fredda tende a rimanere negli strati inferiori dell'Atlantico, essa non potrà mai raggiungere il Mediterraneo. Inoltre anche se 320m potrebbero essere sufficienti a garantire un certo mescolamento di acque fredde profonde, questo non avviene perché la circolazione lungo la soglia è dovuta soprattutto alla differenza di salinità dei due mari. Significa che quando le due masse d'acqua si incontrano, quella mediterranea più salata e quindi più densa, tenderà ad infilarsi sotto quella atlantica mentre quella atlantica ci scorrerà sopra. Motivo per il quale da ovest arriveranno solo acque mediamente temperate e di superficie che andranno rapidamente soggette a riscaldamento.

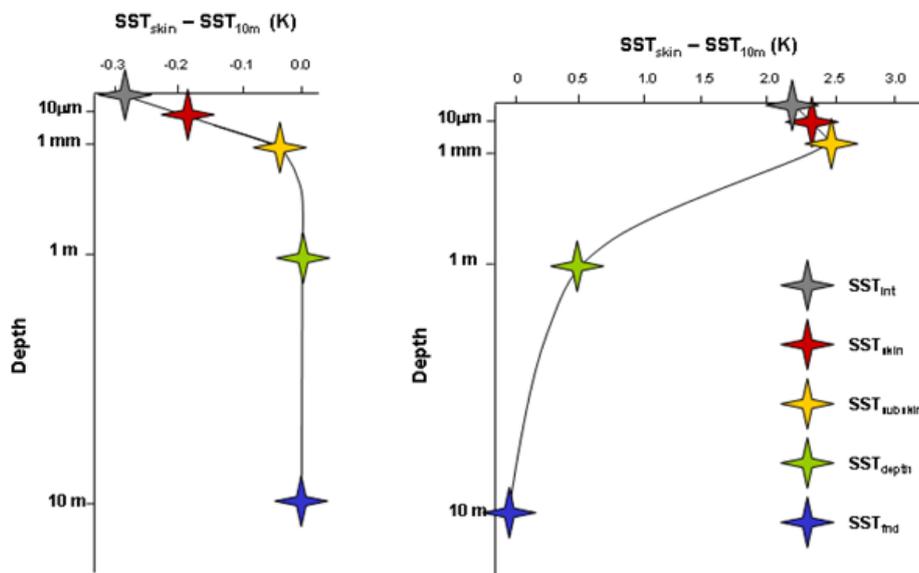


Fig. 2. 4 – Profilo verticale di temperature negli strati superficiali (inferiori ai 10 m di profondità) durante la notte (sinistra) e il giorno (destra, in presenza di forte radiazione solare e vento leggero)

2.3.2 - *L'altezza d'onda*

L'altezza d'onda rappresenta il parametro principale per la definizione dello "stato di mare". Consente infatti, di valutare il contenuto energetico associato ad una determinata mareggiata. In base alle caratteristiche dei sensori montati sulle boe Triaxys, la RON è in grado di valutare tutte le onde di periodo compreso tra 1,6 e 30 secondi: è quindi possibile studiare lo stato di mare derivante dalle sole onde superficiali di vento trascurando gli effetti derivanti dalle onde di mare morto e dalle onde di marea.

Le onde superficiali sono generate dal vento tramite una sequenza di processi di proporzioni variabili nel tempo: si tratta quindi di un trasferimento di quantità di moto ed energia dall'atmosfera al mare [13]. Quando il vento comincia a soffiare a bassa velocità, la turbolenza produce fluttuazioni random di pressione che sviluppano sulla superficie marina piccole onde dette capillari (*ripples*), di lunghezza d'onda dell'ordine del cm. Queste piccole onde sono instabili e possono aumentare nel tempo sia l'ampiezza che la lunghezza d'onda, a causa dello stress superficiale esercitato tangenzialmente dal flusso d'aria e per la presenza di vortici turbolenti su piccole scale.

Ma la causa più importante è la forzante diretta del vento: soffiando sulle onde, il vento genera differenze di pressione lungo il profilo ondosio, in particolare sulle parti più ripide dell'onda. Esprimendo le fluttuazioni come serie di Fourier, il lavoro di pressione deriva dalle componenti che hanno lunghezza d'onda paragonabile a quella delle onde ed in fase con la pendenza: il risultato è l'aumento dell'ampiezza e della lunghezza dell'onda. Ne consegue che aumenta la pressione esercitata e l'instabilità assume una crescita di tipo esponenziale.

Si raggiunge un punto di equilibrio tra vento e onde quando non è possibile ulteriore trasferimento di energia, o meglio c'è un bilancio tra l'input di energia dal vento, il trasferimento dell'energia tra le varie componenti ondose e la dissipazione per rottura delle onde. La distribuzione di energia raggiunge una forma di saturazione più o meno globale sia come livello (presenza di una frequenza di picco) che come forma analitica. L'onda raggiunge la velocità del vento e non può viaggiare più veloce perché comincerebbe la sua erosione per attrito ed effetto frenante. Si parla di condizioni di *fully developed sea* (f.d.s.): non può essere prodotta per le onde una frequenza minore di quella minima del vento, data la sua velocità.

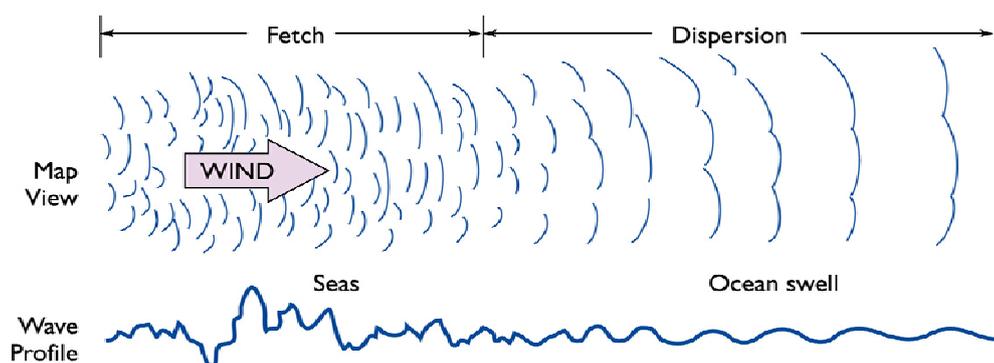


Fig. 2. 5 - Meccanismo di formazione delle onde da vento

Le onde superficiali dipendono, quindi, per la loro generazione da quattro parametri

- Velocità del vento
- Direzione del vento
- Durata dell'azione del vento
- Area di azione del vento (*fetch*)

Venti più veloci, unidirezionali, che soffiano da più tempo e su aree maggiori generano onde con lunghezze d'onda più grandi.

La classificazione dello stato ondoso si basa quindi su quella del vento.

Un scala indicativa dello stato del mare generalmente utilizzata nelle previsioni meteomarine è la *scala Douglas*. Lo stato del mare è lo stato del moto ondoso locale dovuto agli effetti combinati del mare vivo (o mare di vento) e del mare lungo (o mare morto). La tabella seguente ne riporta le caratteristiche.

FORZA	DESCRIZIONE	Hs
0	Calmo	-
1	Quasi calmo	0 - 0,04 m
		0,04 - 0,10 m
2	Poco mosso	0,10 - 0,50 m
3	Mosso	0,50 - 0,80 m
		0,80 - 1,25 m
4	Molto mosso	1,25 - 1,80 m
		1,80 - 2,50 m
5	Agitato	2,50 - 3,20 m
		3,20 - 4 m
6	Molto agitato	4 - 5 m
		5 - 6 m
7	Grosso	6 - 7 m
		7 - 9 m
8	Molto grosso	9 - 14 m

Tab. 2. 3 - La scala Douglas

E' possibile definire e calcolare a partire dai dati osservati quattro parametri sintetici rappresentativi dello stato di mare:

- l'altezza d'onda significativa, che rappresenta una misura del livello energetico del moto ondoso;
- il periodo dominante, corrispondente al picco dello spettro di frequenza;
- il periodo medio delle onde;
- la direzione di propagazione dominante.

Proprio a causa dei rapporti di causa-effetto, è facile mettere in relazione i dati direzionali di altezza d'onda significativa con i dati direzionali anemometrici derivanti dalle misurazioni mareografiche caratterizzate da una maggiore continuità delle osservazioni. Occorre tuttavia valutare le condizioni di esposizione al vento dei singoli mareografi. Condizioni di copertura dei mareografi quali presenza di edifici alti, condizioni orografiche del territorio sfavorevoli, situazioni locali varie possono ridurre la qualità delle informazioni. Inoltre è opportuno valutare lo stato di vento nelle immediate vicinanze della boa e su tutta la zona denominata *fetch efficace* di ogni stazione ondometrica. Ciò è possibile solo tramite modelli fisicamente basati di trasferimento dei dati di vento dal punto a riva in cui sono posizionati i mareografi più vicini alla boa e la zona di mare afferente alla boa stessa.

2.3 - Acquisizione dei dati

I valori misurati dalle boe della rete ondometrica (elevazione, inclinazioni, temperature) sono trasmessi in continuo dalla boa alla stazione a terra la quale contiene un sistema radio (tipicamente un ricevitore DIREC), un microprocessore ed un timer. Il ricevitore acquisisce i dati in formato codificato, il microprocessore li decodifica, esegue il controllo di qualità dei dati e li elabora. Al ricevitore sono inoltre collegati due calcolatori mediante i quali i dati grezzi, i dati elaborati e i dati sintetici vengono trasmessi al Centro di Controllo e Gestione della rete presso il servizio Mareografico dell'ISPRA dove vengono organizzati in una serie di file. Di seguito vengono descritte le tipologie dei dati disponibili, la loro nomenclatura il tracciato record e le eventuali tabelle in cui sono conservati.

2.3.1 - *File HNE per acquisizioni nel dominio del tempo*

I file con suffisso .HNE contengono le acquisizioni nel dominio del tempo della durata di circa 20 minuti con cadenza semioraria. Sono alla base del processo di elaborazione e sono disponibili dal 1999. I file rappresentano un patrimonio importantissimo contenendo la registrazione del moto ondoso direzionale così come rilevato presso l'ondametro e si presta quindi a numerose rielaborazioni ad analisi storiche.

I dati acquisiti a bordo delle boe sono assemblati in n-ple "HNE" e si riferiscono all'elevazione (Heave) ed agli spostamenti per Nord-sud ed Est-ovest. I file sono nominati secondo lo schema [11]:

SSS2_YYYYMMGG_HHMM.hne (es: cat2_20020331_2330.hne)

in cui:

- SSS è il codice della boa;
- YYYYMMGG e HHMM sono codici identificativi della data e dell'orario di misurazione.

I dati acquisiti all'interno del file seguono le seguenti convenzioni:

- Codice iniziale relativo alle informazioni contenute nel file, eventuali errori e numerazione dei pacchetti;
- Heave [cm]: Positivo se lo scostamento è verso l'alto, negativo verso il basso
- NS [cm]: Positivo se lo scostamento è verso Nord, negativo verso Sud
- EO [cm]: Positivo verso Est e negativo verso Ovest.

Ciascun record del file corrisponde ad un campione di 1,28 Hz

In condizioni ottimali in ciascun file devono essere presenti 1536 campioni pari ad un intervallo di acquisizione di $(1/1,28) \cdot 1536 = 1200$ secondi = 20 minuti.

2.3.2 - *File MAIN, relative tabelle SQL e dati spettrali*

I file con suffisso .YYYS (YY=Anno, S=Codice stazione) contengono le elaborazioni nel dominio delle frequenze ottenuto sottoponendo i file HNE ad una Fast Fourier Transform.

Sono disponibili file Main a partire dal 1995.

Il file contiene quattro tabelle contenenti ciascuna le informazioni riportate in Tab. 2. 4.

La seconda tabella contiene 128 record per ogni chiave composta da u codice rappresentativo della stazione, la data e l'ora di inizio registrazione e tutte le informazioni utili all'analisi spettrale (autospettri degli spostamenti, cospettri e spettri in quadratura).

La terza tabella contiene 58 record per ogni chiave composta da un codice rappresentativo della stazione, la data e l'ora di inizio registrazione e i parametri statistici relativi all'analisi spettrale (spread direzionale, asimmetria, curtosi).

La quarta tabella contiene 12 record per ogni chiave composta dalla sigla della stazione e dalla data e ora di inizio della registrazione da cui sono tratti. Per ognuno dei record sono descritti i valori relativi ad una banda di frequenza di ampiezza tale che la sua energia sia pari a circa 1/12 (un dodicesimo) dell'energia di tutto lo spettro [11].

2.3.3 - File di sistema, SIS e relativa tabella

Si tratta della tabella contenente i parametri estratti dal file "SYSTEM" relativo ai dati delle Boe Datawell Waverider. E' l'unica sorgente che rende disponibile la tensione delle batterie a bordo della boa e la temperatura dell'acqua marina. Tali parametri vengono poi riportati anche nella tabella principale del file Main [11].

Dati generali	Stazione	Testo
	Data	Testo formattato
Parametri sintetici	Hm0	Altezza d'onda significativa in metri
	Tm	Periodo medio in secondi
	Tp	Periodo di picco in secondi
Analisi Spettrale	Hs	Altezza d'onda significativa spettrale in metri
	Tp1	Periodo di picco in secondi
	Tm1	Periodo medio in secondi ottenuto dalla radice quadrata del rapporto tra il momento di ordine 0 e il momento di ordine 1
	Tm01	Periodo medio in secondi ottenuto dal rapporto tra il momento di ordine 0 e il momento di ordine 1
	Tm_10	Periodo medio in secondi ottenuto dal rapporto tra il momento di ordine 1 e il momento di ordine 0
	Tm_20	Periodo medio in secondi ottenuto dalla radice quadrata del rapporto tra il momento di ordine 2 e il momento di ordine 0
	Tm24	Periodo medio in secondi ottenuto dalla radice quadrata del rapporto tra il momento di ordine 2 e il momento di ordine 4
	Tm04	Periodo medio in secondi ottenuto dal rapporto tra il momento di ordine 0 e il momento di ordine 4
	Dmt	Direzione media totale di provenienza in gradi Nord
	Dmp	Direzione media totale di picco in gradi Nord
Parametri nel dominio del tempo	DT Hm	Altezza media in metri
	DT HVar	Varianza dell'altezza in metri
	DT Hmax	Altezza massima in metri
	DT Hs	Altezza significativa in metri
	DT Hdec	Altezza media dei decimi più alti in metri
	DT Hcent	Altezza media dei centesimi più alti in metri
	DT Tm	Periodo medio in secondi
	DT TVar	Varianza del periodo in secondi
	DT Tmax	Periodo massimo in secondi
	DT Ts	Periodo significativo in secondi
	DT Tdec	Periodo medio dei decimi più alti in secondi
	DT Tcent	Periodo medio dei centesimi più alti in secondi

Tab. 2. 4 - Informazioni contenute nel file MAIN

2.3.4 - *File di sistema, Status e relativa tabella.*

Si tratta della tabella contenente i parametri estratti dal file “STATUS” relativo ai dati delle Boe TRIAXYS. E' l'unica sorgente che rende disponibile la tensione delle batterie a bordo della boa e la temperatura dell'acqua marina. Tali parametri vengono poi riportati anche nella tabella principale (tab 2.4) del file Main [11].

2.4 - **Elaborazione dei dati**

I dati grezzi sono soggetti ad una prima elaborazione nel sistema boa-stazione e ad una post-elaborazione presso il Centro di controllo e Gestione.

2.4.1 - *Modulo di Post-Processing nella stazione a terra*

Nella stazione a terra sono previste le seguenti operazioni:

- **Verifica della qualità dei dati**

Il campione di dati misurati è filtrato sulla boa con un filtro passa-basso del 7° ordine di Butterworth con frequenza di taglio pari a 0,6 Hz per prevenire errori dovuti all'aliasing, ed è trasmesso ogni 0,78125 secondi come sequenza di 128 bit. Il blocco di 128 bit viene decodificato dal ricevitore DIREC e vengono individuati e corretti gli errori di trasmissione. Vengono quindi calcolati il valore assoluto e l'inclinazione del campo magnetico. Poiché, entrambi questi valori, in condizioni di funzionamento normale, devono restare pressoché stabili, per la verifica di qualità si procede al loro confronto con i valori medi ottenuti sui precedenti 256 campioni validi. I valori dei dati misurati sono accettati se la deviazione dalla media del valore assoluto del campo magnetico terrestre è inferiore al 10% (l'inclinazione del campo magnetico terrestre ha una deviazione dalla media inferiore a 5.5°). Il valore dell'elevazione viene confrontato con il valore atteso derivato dalla interpolazione dei valori precedenti e successivi. Il criterio generale di accettabilità è che la differenza tra il valore misurato ed il valore atteso sia inferiore a quattro volte la deviazione standard delle differenze nella serie e sia comunque inferiore a 0,7 m, che è la massima differenza assoluta ammessa [3].

- **Assemblaggio dei dati per il calcolo spettrale**

Le terne di valori di elevazione, pendenza N/S e pendenza E/W sono raccolti in blocchi di 256 campioni, ognuno corrispondente ad un intervallo di 200 secondi. Il campionamento inizia di nuovo se sono stati rilevati più di 25 campioni non ritenuti validi (10%) o se risultano mancanti 2 o più campioni adiacenti. In tal caso i dati precedenti all'interno del blocco di 256 sono eliminati. Il numero totale di blocchi di 256 dati impiegati nella successiva elaborazione è pari al numero di blocchi accettati [3].

- **Smussamento degli estremi ("Tapering")**

I primi e gli ultimi 32 campioni del blocco di 256 terne di dati vengono filtrati moltiplicando per la funzione [3]:

$$f_n = 1 - \cos(n\pi/32)$$

Tale operazione evita la fluttuazione della Fast Fourier Transform, utilizzata nella elaborazione, come di seguito illustrato, agli estremi dell'intervallo di osservazione.

- **Analisi spettrale**

Il metodo utilizzato per l'analisi dei dati trasmessi dalla boa consente di ottenere 4 parametri caratteristici della distribuzione direzionale per frequenza $D_f(\theta)$, senza alcuna ipotesi preliminare sulla forma della distribuzione stessa [3]:

$$D_f(\theta) = E(f, \theta) / E(f)$$

dove $E(f, \theta)$ è lo spettro direzionale di energia ed $E(f) = \int_0^{2\pi} E(f, \theta) d\theta$ è la densità di energia. La distribuzione è approssimata con uno sviluppo in serie di Fourier troncato ai primi 4 coefficienti è data dalle seguente relazione:

$$D_f(\theta) = 1/\pi [1/2 + (a_1 \cos \theta + b_1 \sin \theta) + (a_2 \cos 2\theta + b_2 \sin 2\theta)]$$

I coefficienti a_1, b_1, a_2 e b_2 vengono calcolati a partire dai dati registrati nel seguente modo. Mediante una routine di Fast Fourier Transform dai blocchi di dati filtrati di 256 terne di elevazione (z), pendenza nord/sud (x) e pendenza est/ovest (y), si ottengono 9 funzioni di densità spettrale nell'intervallo tra 0,005Hz e 0,635 Hz per 127 bande di frequenza di ampiezza 0,005 Hz [3]:

3 autospettri	$[C_{zz}(f), C_{xx}(f), C_{yy}(f)]$
3 cospettri	$[C_{zx}(f), C_{zy}(f), C_{xy}(f)]$
3 spettri di quadratura	$[Q_{zx}(f), Q_{zy}(f), Q_{xy}(f)]$

La matrice di 127 righe per 9 colonne ottenuta dalla elaborazione del primo blocco viene mediata con quella ottenuta dal secondo e così via fino ad ottenere la matrice rappresentativa dei 30 minuti di osservazione.

Dall'autospettro dell'elevazione $C_{zz}(f)$ è possibile ricavare direttamente le grandezze

- H_S , altezza d'onda significativa;
- T_m , periodo medio;
- T_p , periodo di picco.

Essendo infatti

$$C_{zz}(f) = E(f) = \int_0^{2\pi} E(f, \theta) d\theta$$

indicando con $q_i = \int_0^\infty f^i E(f) df$ il momento spettrale di ordine i-esimo, si ottengono i seguenti valori [3]:

$$H_S = 4 [q_0]^{1/2} ; \quad T_m = q_1/q_0$$

Il valore di T_p è quello corrispondente al massimo valore di $C_{zz}(f)$.

Con le funzioni di densità spettrale ottenute, si calcolano, quindi, i coefficienti a_1 , b_1 , a_2 e b_2 , dello sviluppo in serie di Fourier della funzione $D_f(\theta)$ con le seguenti relazioni [3]:

$$a_1(f) = \int_0^{2\pi} D_f(\theta) \cos \theta d\theta = Q_{zx}(f)/k(f)C_{zz}(f);$$

$$a_2(f) = \int_0^{2\pi} D_f(\theta) \cos 2\theta d\theta = C_{xx}(f) - C_{yy}(f)/k^2(f)C_{zz}(f);$$

$$b_1(f) = \int_0^{2\pi} D_f(\theta) \sin \theta d\theta = Q_{zy}(f)/k(f)C_{zz}(f);$$

$$b_2(f) = \int_0^{2\pi} D_f(\theta) \sin 2\theta d\theta = 2C_{xy}(f)/k^2(f)C_{zz}(f);$$

dove $k(f) = \{C_{xx} + C_{yy}/C_{zz}\}^{1/2}$ è il numero d'onda.

La direzione media di propagazione per banda di frequenza è ottenuta come:

$$\theta_0(f) = \arctan(b_1(f)/a_1(f))$$

Dai coefficienti dello sviluppo in serie di Fourier, mediante una rotazione del sistema di coordinate in modo da avere direzione media nulla, si ottengono le funzioni:

$$m_1(f) = a_1(f) \cos[\theta_0(f)] + b_1(f) \sin[\theta_0(f)]$$

$$m_2(f) = a_2(f) \cos[2\theta_0(f)] + b_2(f) \sin[2\theta_0(f)]$$

$$n_2(f) = b_2(f) \cos[2\theta_0(f)] - a_2(f) \sin[2\theta_0(f)]$$

mediante le quali si calcolano i seguenti valori per banda di frequenza [3]:

- la dispersione direzionale (spread) $\sigma(f) = \{2[1 - m_1(f)]\}^{1/2}$
- l'asimmetria (skewness) $S(f) = -n_2(f)/\{0,5[1 - m_2(f)]\}^{3/2}$
- la curtosi $K(f) = [6 - 8m_1(f) + 2m_2(f)]/[\sigma(f)]^4$.

2.4.2 - Modulo di Post-Processing nel Centro di controllo e gestione

Presso il Centro di Controllo e gestione dell'ISPRA, i dati provenienti da boe ondametriche Datawell tipo Directional Waverider e TRIAXYS vengono processati attraverso un opportuno software che richiede in input un file con estensione *.hne*. Il file di uscita ha la stessa struttura e lo stesso contenuto dei file MAIN, contenente i risultati dell'analisi dei dati nel dominio delle frequenze e nel dominio tempo.

Nella fase iniziale il software prevede un modulo per il controllo dell'attendibilità dei dati delle serie temporali che vengono inseriti come dati di input. In particolare vengono accettati solo i dati il cui codice di errore è uguale a zero (Receiver is locked, no errors) o uguale ad uno (One or two errors of the receiver detected and corrected) e ne viene controllata l'effettiva successione temporale [11]. Nel caso di mancanza di un solo dato questo viene aggiunto sulla base di una semplice interpolazione lineare. Non si è ritenuto opportuno introdurre tecniche di interpolazione più evolute, intendendo con ciò l'utilizzo di algoritmi di ordine superiore nel caso di mancanza di più dati, in quanto tali algoritmi sono ritenuti non calibrabili in maniera affidabile per tutto l'insieme degli stati di mare nei quali la boa si potrebbe trovare ad operare.

Sulla base delle informazioni fornite da questo primo modulo del software si individua la successione di dati più lunga con M valori che al più sarà costituita da 2048 valori nel caso che tutti i dati abbiano passato la prova di affidabilità, i criteri della quale sono stati qui sopra esposti, e vengono presi in considerazione un numero di dati N , uguale al prodotto $256 \cdot K$, avendo definito come K la parte intera del rapporto $M/256$. Al proposito va notato che il software andrà in esecuzione solamente se K risulta maggiore od uguale a uno, altrimenti esso darà un messaggio di errore [11].

Per quanto riguarda l'analisi nel dominio della frequenza il software segue nuovamente l'analisi esposta nel paragrafo (in piena sintonia con quanto esposto al proposito nel capitolo 4 de "Operation and Service Manual for Directional Waverider"), introducendo tre diversi fattori correttivi moltiplicativi delle tre serie temporali che tengono conto

dell'attenuazione dell'energia del segnale dovuta allo smussamento degli estremi ("Tapering"). In particolare si ha:

$$Z_c(t) = Z(t) \frac{H_{snt}}{H_{swt}}$$

avendo indicato con:

H_{snt} l'altezza significativa del segnale senza smussamento degli estremi

H_{swt} l'altezza significativa del segnale con smussamento degli estremi

$Z(t)$ serie temporale originale

$Z_c(t)$ serie temporale corretta

Infine va notato che il calcolo delle medie degli autospettri, dei coospettri e degli spettri in quadratura, che sono calcolati su una serie temporale di 256 campioni in modo da avere una risoluzione in frequenza di 0,01 Hz, viene fatta su K campioni, ove K, alla luce di quanto detto precedentemente, è un numero intero compreso fra 1 e 8 [11].

L'analisi nel dominio del tempo invece è limitata alla serie temporale della elevazione ondosa, sullo spezzone più lungo che, come detto in precedenza, ha M valori. Una volta che tale segnale è stato depurato della componente continua vengono individuate le sue altezze ed i suoi periodi; le prime intese come differenze fra il valore massimo (cresta) e il valore minimo (cavo) fra due attraversamenti dello zero del segnale ed i secondi definiti come intervalli di tempo intercorrenti fra due attraversamenti dello zero. Per attraversamento dello zero si intende quello ove il segnale ha la sua derivata temporale negativa (zero down crossing), come viene suggerito dalla letteratura specializzata (Shore Protection Manual). Sulla base dei valori delle altezze e dei periodi ricavati con il metodo precedentemente descritto è possibile ricavare il valore massimo, quello medio, la varianza, quello significativo, inteso come la media effettuata su un terzo dei valori più alti riscontrati e in analogia quelli del decimo e del centesimo più alti, che nel caso non si avesse un numero sufficiente di eventi vengono posti uguali a zero.

Il software inoltre da i seguenti messaggi di errore [11]:

- errore numero 1: il file di ingresso specificato in input non esiste,
- errore numero 2: il file di uscita specificato in input già esiste,
- errore numero 3: nel file di ingresso specificato in input si hanno meno di 256 dati attendibili consecutivi.

2.5 - Assenza dei dati ondametrici e qualità del dataset

Un aspetto importante dell'analisi delle serie temporali riguarda valutazione della qualità complessiva del dataset ed in particolare, la ricerca e la ricostruzione di dati anomali e di dati mancanti.

2.5.1 - Dati anomali

Nel dataset RON è possibile individuare diverse tipologie di dati anomali [4]:

- errori nella formattazione;
- presenza di valori registrati con boa a terra;
- presenza di spike residui;
- presenza di misure ripetute.

Per ognuno di questi occorre applicare metodi per evidenziare i dati non validi, eliminarli o, dove possibile, correggerli.

Analisi della formattazione

I dati elaborati possono presentare errori relativi ad una formattazione anomala delle stringhe di testo, come nel caso di dati registrati ad orari non standard. Vanno quindi eliminate le registrazioni ad orari ripetuti e corrette, ove possibile, le stringhe relative a serie temporali non standard.

Valori registrati con boa a terra

Durante gli interventi di manutenzione, le boe vengono portate a terra, e, se gli strumenti sono lasciati in funzione, vengono registrati e trasmessi dati non significativi. Le misurazioni relative a tali periodi di dati non validi sono individuabili dalla presenza di valori di H_{m0} nulli o costantemente molto bassi ($H_{m0} < 0.2$ m) per lunghi periodi di tempo e devono essere eliminati [4].

Spike residui

La ricerca di spike (valori anomali nelle serie temporali) all'interno dei file di dati viene normalmente realizzata tramite un test che identifica le stringhe con valori di H_{m0} , T_p o T_m soggette a variazioni eccedenti un limite prefissato rispetto ai valori dei dati immediatamente precedenti e successivi. Valori tipici utilizzabili come soglia per tutte le

stazioni possono essere 1 metro per la H_{m0} , 15 secondi per il T_p e 10 secondi per il T_m [4]. I dati selezionati devono poi essere giudicati validi o meno attraverso un'analisi soggettiva delle serie temporali. Questi spike non sono stati eliminati nell'analisi al fine di valutare la bontà del modello statistico che, in prossimità di questi dati, dovrebbe segnare un errore maggiore. Sono stati, invece, eliminati dati che presentano errori grossolani, come ad esempio valori di temperatura dell'acqua minore di 0°C e maggiore di 50°C (la temperatura dell'acqua generalmente oscilla tra gli 8°C nei mesi invernali e 30°C in estate) e valori di altezza d'onda significativa pari a 0, poiché spesso il valore nullo è stato utilizzato per indicare l'assenza del dato.

Temperatura		Dati eliminati						
Anno	Osservazioni	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro	Catania	Cetraro
2002-2006	87648	16829	17713	0	0	0	5663	4729
2002	17520	8867	10473	0	0	0	3936	3065
2003	17520	3384	2152	0	0	0	1727	757
2004	17568	4572	0	0	0	0	0	0
2005	17520	6	2831	0	0	0	0	0
2006	17520	0	2257	0	0	0	0	907

Temperatura		Dati eliminati						
Anno	Osservazioni	Crotone	La Spezia	Mazara	Monopoli	Ortona	Ponza	P. d. Maestra
2002-2006	87648	23268	13450	9929	17926	23945	11673	488
2002	17520	3718	11777	6592	9234	9981	9869	0
2003	17520	1456	1673	3337	2880	4219	1804	0
2004	17568	0	0	0	0	0	0	488
2005	17520	9015	0	0	0	1038	0	0
2006	17520	9079	0	0	5812	8707	0	0

Hs		Dati eliminati						
Anno	Osservazioni	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro	Catania	Cetraro
2002-2006	87648	12532	13166	27	0	927	6615	4337
2002	17520	8873	10363	0	0	0	3611	3006
2003	17520	3423	2179	0	0	0	1732	1150
2004	17568	0	142	25	0	192	412	0
2005	17520	127	422	2	0	623	787	0
2006	17520	109	60	0	0	112	73	181

Hs		Dati eliminati						
Anno	Osservazioni	Crotone	La Spezia	Mazara	Monopoli	Ortona	Ponza	P. d. Maestra
2002-2006	87648	6277	13605	9747	12681	13829	11942	39
2002	17520	3718	11927	6593	9184	8871	9870	0
2003	17520	1466	1672	3004	2923	4258	1805	0
2004	17568	269	0	80	164	141	102	39
2005	17520	654	0	44	206	283	113	0
2006	17520	170	6	26	204	276	52	0

Tab. 2. 5 - Valori eliminati di temperatura dell'acqua e altezza d'onda significativa

Misure ripetute

La ricerca delle misure ripetute consiste nell'evidenziare tutte le registrazioni che presentano lo stesso valore di ora, H_{m0} , T_p , T_m e direzione (α) anche se in date o in stazioni diverse. Il test è condotto su ogni serie temporale e viene successivamente ripetuto per

verificare la presenza di stringhe uguali nello stesso periodo temporale su stazioni differenti.

2.5.2 - *Dati mancanti*

Per quanto riguarda i dati ondametrici, le cause principali delle indisponibilità dei dati sono

- a) problemi nella trasmissione radio del segnale tra la boa e la stazione a terra,
- b) malfunzionamenti del sistema, tra cui disormeggio della boa, rottura degli apparati di trasmissione/ricezione radio, rottura dei supporti di memorizzazione, operazioni di manutenzione programmata, etc.
- c) dati eliminati in fase di post-processing (despiking, serie di dati ripetuti).

I periodi di malfunzionamento strumentale hanno un ampio spettro di possibili durate, legato alla molteplicità delle possibili cause e al conseguente ritardo nel ripristino dell'operatività e possono provocare l'inattività di una stazione da un minimo di 6 ore fino a qualche mese.

Problemi di trasmissione hanno invece la caratteristica di durare per periodi di tempo limitati ad una o poche rilevazioni consecutive. Lo studio prende in esame la disponibilità complessiva di osservazioni e, nel caso di serie temporali più lunghe, anche la distribuzione statistica delle lunghezze dei periodi di mancanza dato (gap). E' opportuno dividere i gap brevi (inferiori alle 6 ore) dai gap più lunghi (uguali o maggiori di 6 ore). La scelta è stata motivata anche dalla diversa natura delle cause che provocano i malfunzionamenti della rete: i gap brevi sono essenzialmente dovuti a problemi di trasmissione radio del dato dalla boa alla stazione ricevente a terra o a occasionali "spikes" e sono presenti ovviamente solo nelle parti di archivio corrispondenti alle modalità di acquisizione semioraria, mentre i gap lunghi sono legati a guasti degli apparati, operazioni di manutenzione programmata o disormeggi della boa [4]. La tabella di seguito riportata evidenzia la numerosità dei dati mancanti.

Anno	Osservazioni	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro
2002-2006	87648	38328 (43,73%)	49466 (56,44%)	64838 (73,98%)	46541 (53,10%)	56234 (64,16%)
2002	17520	8869 (50,62%)	10475 (59,79%)	17520 (100,00%)	17520 (100,00%)	17520 (100,00%)
2003	17520	3387 (19,33%)	2155 (12,30%)	17520 (100,00%)	17520 (100,00%)	17520 (100,00%)
2004	17568	9813 (55,86%)	3830 (21,80%)	1109 (6,31%)	1745 (9,93%)	2010 (11,44%)
2005	17520	9438 (53,87%)	15486 (88,39%)	11169 (63,75%)	5481 (31,28%)	11410 (65,13%)
2006	17520	6821 (38,93%)	17520 (100,00%)	17520 (100,00%)	4275 (24,40%)	7774 (44,37%)

Anno	Osservazioni	Catania	Cetraro	Crotone	La Spezia	Mazara
2002-2006	87648	29149 (33,26%)	13964 (15,93%)	37883 (43,22%)	41017 (46,80%)	23993 (27,37%)
2002	17520	3938 (22,48%)	3067 (17,51%)	3720 (21,23%)	11779 (67,23%)	6595 (37,64%)
2003	17520	1730 (9,87%)	761 (4,34%)	1459 (8,33%)	1676 (9,57%)	3340 (19,06%)
2004	17568	2516 (14,32%)	2662 (15,15%)	1648 (9,38%)	7960 (45,31%)	2200 (12,52%)
2005	17520	4848 (27,67%)	3463 (19,77%)	15652 (89,34%)	8751 (49,95%)	3620 (20,66%)
2006	17520	16117 (91,99%)	4011 (22,89%)	15404 (87,92%)	10851 (61,93%)	8238 (47,02%)

Anno	Osservazioni	Monopoli	Ortona	Ponza	P. d. Maestra
2002-2006	87648	28707 (32,75%)	48688 (55,55%)	30156 (34,41%)	78259 (89,29%)
2002	17520	9237 (52,72%)	9987 (57,00%)	9871 (56,34%)	17520 (100,00%)
2003	17520	2883 (16,46%)	4223 (24,10%)	1807 (10,31%)	17520 (100,00%)
2004	17568	1423 (8,10%)	8621 (49,07%)	2747 (15,64%)	8179 (46,56%)
2005	17520	5425 (30,96%)	11627 (66,36%)	6713 (38,32%)	17520 (100,00%)
2006	17520	9739 (55,59%)	14230 (81,22%)	9018 (51,47%)	17520 (100,00%)

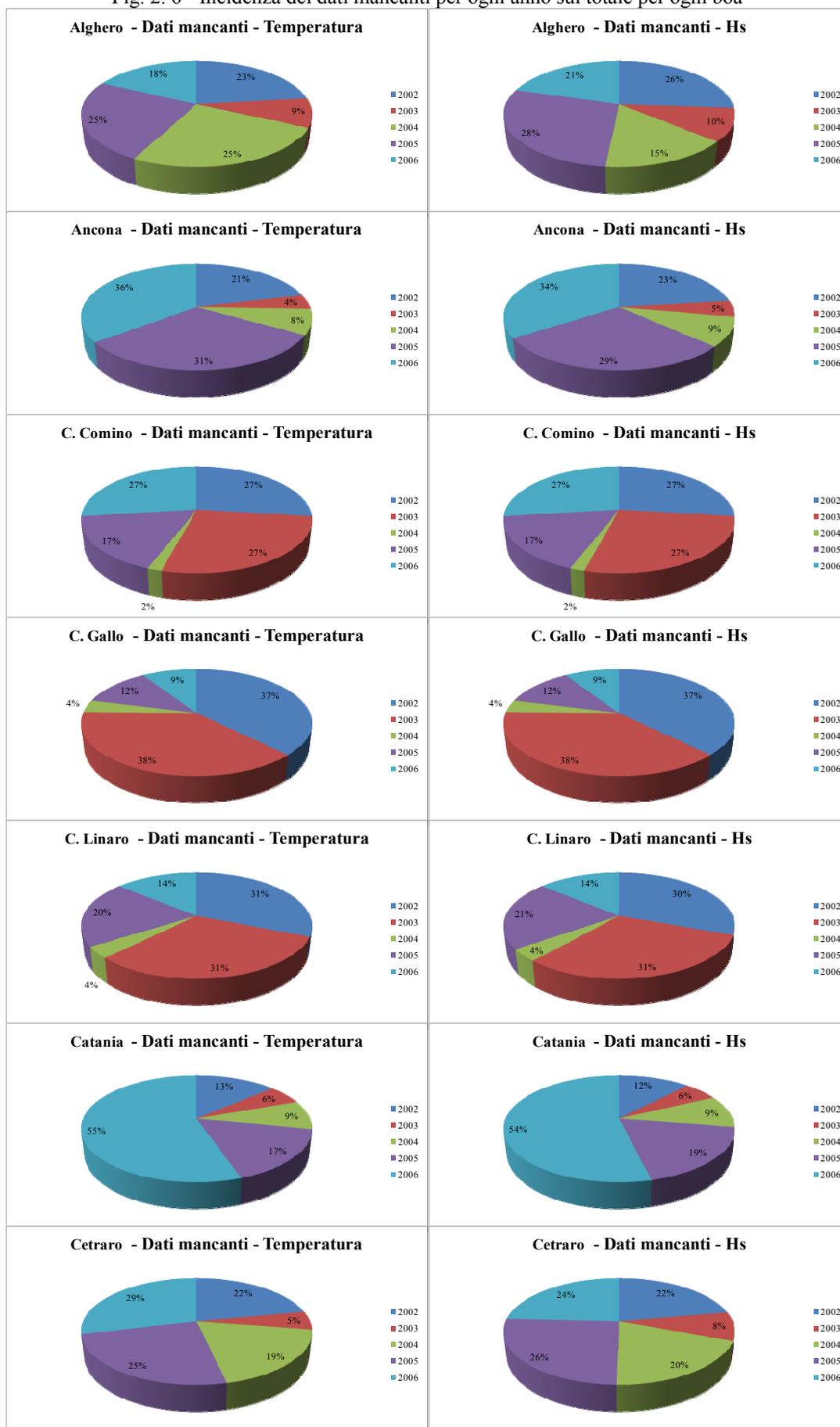
Anno	Osservazioni	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro
2002-2006	87648	34031 (38,83%)	44919 (51,25%)	64865 (74,01%)	46541 (53,10%)	57161 (65,22%)
2002	17520	8875 (50,66%)	10365 (59,16%)	17520 (100,00%)	17520 (100,00%)	17520 (100,00%)
2003	17520	3426 (19,55%)	2182 (12,45%)	17520 (100,00%)	17520 (100,00%)	17520 (100,00%)
2004	17568	5241 (29,83%)	3972 (22,61%)	1134 (6,45%)	1745 (9,93%)	2202 (12,53%)
2005	17520	9559 (54,56%)	13077 (74,64%)	11171 (63,76%)	5481 (31,28%)	12033 (68,68%)
2006	17520	6930 (39,55%)	15323 (87,46%)	17520 (100,00%)	4275 (24,40%)	7886 (45,01%)

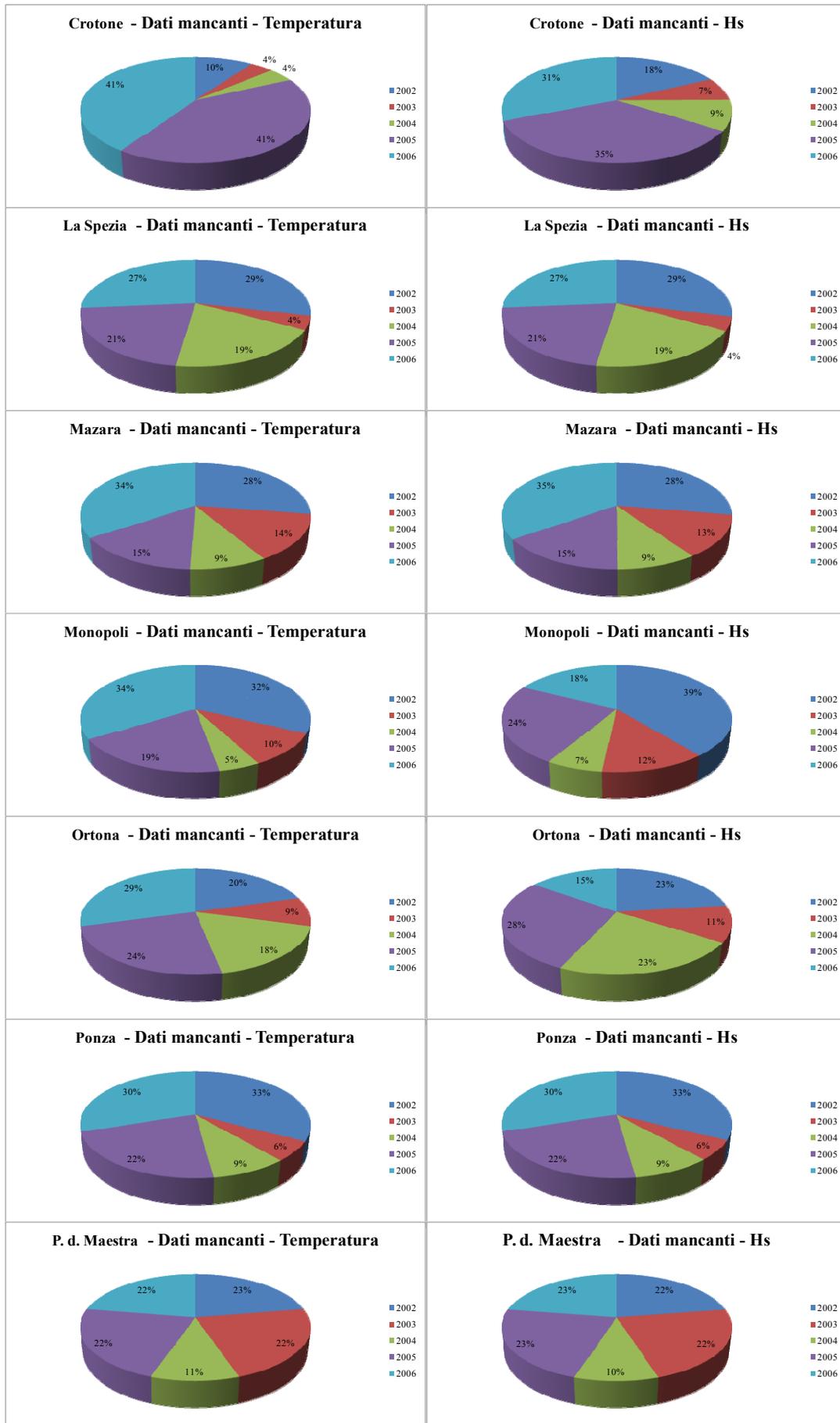
Anno	Osservazioni	Catania	Cetraro	Crotone	La Spezia	Mazara
2002-2006	87648	30101 (34,34%)	13572 (15,48%)	20892 (23,84%)	41172 (46,97%)	23811 (27,17%)
2002	17520	3613 (20,62%)	3008 (17,17%)	3720 (21,23%)	11929 (68,09%)	6596 (37,65%)
2003	17520	1735 (9,90%)	1154 (6,59%)	1469 (8,38%)	1675 (9,56%)	3007 (17,16%)
2004	17568	2928 (16,67%)	2662 (15,15%)	1917 (10,91%)	7960 (45,31%)	2280 (12,98%)
2005	17520	5635 (32,16%)	3463 (19,77%)	7291 (41,62%)	8751 (49,95%)	3664 (20,91%)
2006	17520	16190 (92,41%)	3285 (18,75%)	6495 (37,07%)	10857 (61,97%)	8264 (47,17%)

Anno	Osservazioni	Monopoli	Ortona	Ponza	P. d. Maestra
2002-2006	87648	23462 (26,77%)	38572 (44,01%)	30425 (34,71%)	77810 (88,78%)
2002	17520	9187 (52,44%)	8877 (50,67%)	9872 (56,35%)	17520 (100,00%)
2003	17520	2926 (16,70%)	4262 (24,33%)	1808 (10,32%)	17520 (100,00%)
2004	17568	1587 (9,03%)	8762 (49,87%)	2849 (16,22%)	7730 (44,00%)
2005	17520	5631 (32,14%)	10872 (62,05%)	6826 (38,96%)	17520 (100,00%)
2006	17520	4131 (23,58%)	5799 (33,10%)	9070 (51,77%)	17520 (100,00%)

Tab. 2. 6 - Dati mancanti di temperatura dell'acqua (sopra) e di altezza d'onda significativa (sotto) nel dataset RON tra il 2002 e il 2006

Fig. 2. 6 - Incidenza dei dati mancanti per ogni anno sul totale per ogni boa





Gap brevi (30' - 5.30' ore)

I gap brevi sono prevalentemente legati a problemi di trasmissione radio dei dati dalla boa alla stazione a terra, dunque i fenomeni che possono essere presi in esame nel valutare le distribuzioni osservate sono di tipo geometrico (la distanza tra antenna ricevente e trasmittente, l'altezza s.l.m. della stazione ricevente) la ripidità delle onde più alte, la possibilità di frangimenti al largo, interferenze radio locali.

Temperatura	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro	Catania	Cetraro
2002	1894	1595	-	-	-	113	519
2003	751	241	-	-	-	90	152
2004	682	182	131	146	67	192	137
2005	750	166	538	538	100	564	652
2006	300	-	-	307	1470	10	108
Totale	4377	2184	669	991	1637	969	1568

Temperatura	Crotone	La Spezia	Mazara	Monopoli	Ortona	Ponza	P.d. Maestra
2002	304	1494	1863	1350	1200	1466	-
2003	69	264	227	268	689	570	-
2004	84	195	94	408	640	144	155
2005	107	521	526	855	463	491	-
2006	19	30	614	106	267	119	-
Totale	583	2504	3324	2987	3259	2790	155

Hs	Alghero	Ancona	C. Comino	C. Gallo	C. Linaro	Catania	Cetraro
2002	1891	1712	-	-	-	130	587
2003	748	264	-	-	-	91	178
2004	997	174	130	146	73	192	137
2005	745	391	539	538	36	549	652
2006	286	118	-	307	1454	9	565
Totale	4667	2659	669	991	1563	971	2119

Hs	Crotone	La Spezia	Mazara	Monopoli	Ortona	Ponza	P.d. Maestra
2002	304	1522	1861	1348	1317	1465	-
2003	76	263	171	265	689	568	-
2004	72	195	89	388	620	143	216
2005	1012	521	522	812	560	490	0
2006	41	28	613	172	458	73	0
Totale	1505	2529	3256	2985	3644	2739	216

Tab. 2. 7 - Numero di gap brevi nel dataset RON tra il 2002 e il 2006

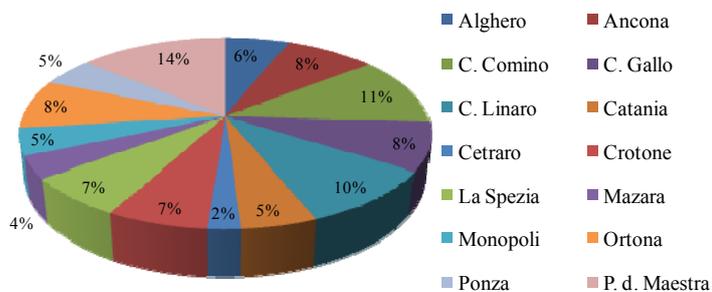
Nei primi mesi del 2002 non è facile stabilire il numero reale di dati mancanti dovuti ad una inefficace trasmissione del dato, perché la frequenza di acquisizione dei dati è trioraria con dati semiorari in caso di eventi di notevole intensità. Nell'intorno dell'altezza di soglia si addensano una quantità di rilevazioni mancanti nel set semiorario, ma la grande parte di tali lacune è semplicemente dovuta alla transizione tra le due modalità operative sopra e sotto soglia: non si può a priori stabilire se un dato non triorario manchi per problemi di trasmissione o perché la misura passa sotto il livello di soglia dell'acquisizione trioraria [4]. Certamente, se i valori prima e dopo la lacuna eccedono largamente entrambi il valore

di soglia, le probabilità che si tratti effettivamente di un dato “missing” diventano consistenti. In ogni caso, nella maggior parte dei casi nel dataset manca un solo valore e, nei primi mesi del 2002, sono frequenti sequenze di 5 dati semiorari mancanti, che corrispondono ad un gap di 2,5 ore. Questo gap non costituisce un dato mancante quando la frequenza di acquisizione è trioraria, ma, ai fini della definizione del modello statistico dei ricostruzione del dataset, vengono considerati come missing values.

Gap lunghi (≥ 6 ore)

I gap di dimensioni maggiori di sei ore sono solitamente associabili a malfunzionamenti del sistema di misura/trasmisione che hanno richiesto un intervento manutentivo di ripristino.

Gap Lunghi 2002-2006 - Temperatura



Gap Lunghi 2002-2006 - Hs

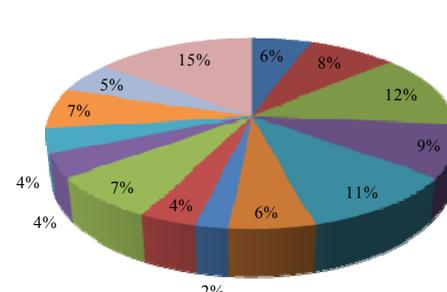


Fig. 2. 7 – Percentuale di dati mancanti riferiti a gap lunghi sul totale

Confrontando la quantità complessiva di dati mancanti sia della temperatura dell’acqua che dell’altezza d’onda significativa (Fig. 2. 7) si osserva che le stazioni con maggior numero di dati mancanti sono Punta della Maestra, Capo Comino e Capo Linaro che da sole contribuiscono con il 38% delle lacune riscontrate in tutta la rete.

Le restanti stazioni contribuiscono con valori inferiori al 10%. Cetraro è la stazione con il minor numero complessivo di osservazioni mancanti, pari al 2% del totale. La distribuzione della lunghezza dei gap osservati, rappresentata in Tab. 2. 8, si ottiene suddividendo l’insieme in 8 classi in modo da riconoscere gap di breve, media, lunga e lunghissima durata [4].

In diverse boe si sono riscontrati periodi estremamente lunghi (superiori ai 6 mesi) di mancanza dati (Ancona 2005, 2006; Capo Comino 2002, 2003, 2005, 2006; Capo Gallo 2002, 2003; Capo Linaro 2002, 2003, 2006; Catania 2006; La Spezia 2002, 2006; Punta della Maestra 2002, 2003, 2005, 2006) dovuti a guasti delle apparecchiature o impedimenti

burocratici. A volte si riscontrano molti valori mancanti nelle serie storiche delle temperature corrispondenti a valori presenti di altezza d'onda significativa (Alghero 2004; Ancona 2005, 2006; Crotone 2005, 2006; Monopoli 2006; Ortona 2006): in questo caso si tratta di guasti al sensore di rivelazione delle temperature non riparati per convenienza economica o per evitare di dover portare a terra la boa e perdere così ulteriori dati di altezza d'onda ritenuti maggiormente importanti.

Per valutare le variazioni annuali nella disponibilità dei dati delle singole stazioni è stato introdotto il numero di giorni di inoperatività annuale, definito come somma del numero di osservazioni semiorarie mancanti della stessa stazione espresse in giorni per ogni ondometro e per ogni tipo di misurazione (Fig. 2. 8) [4].

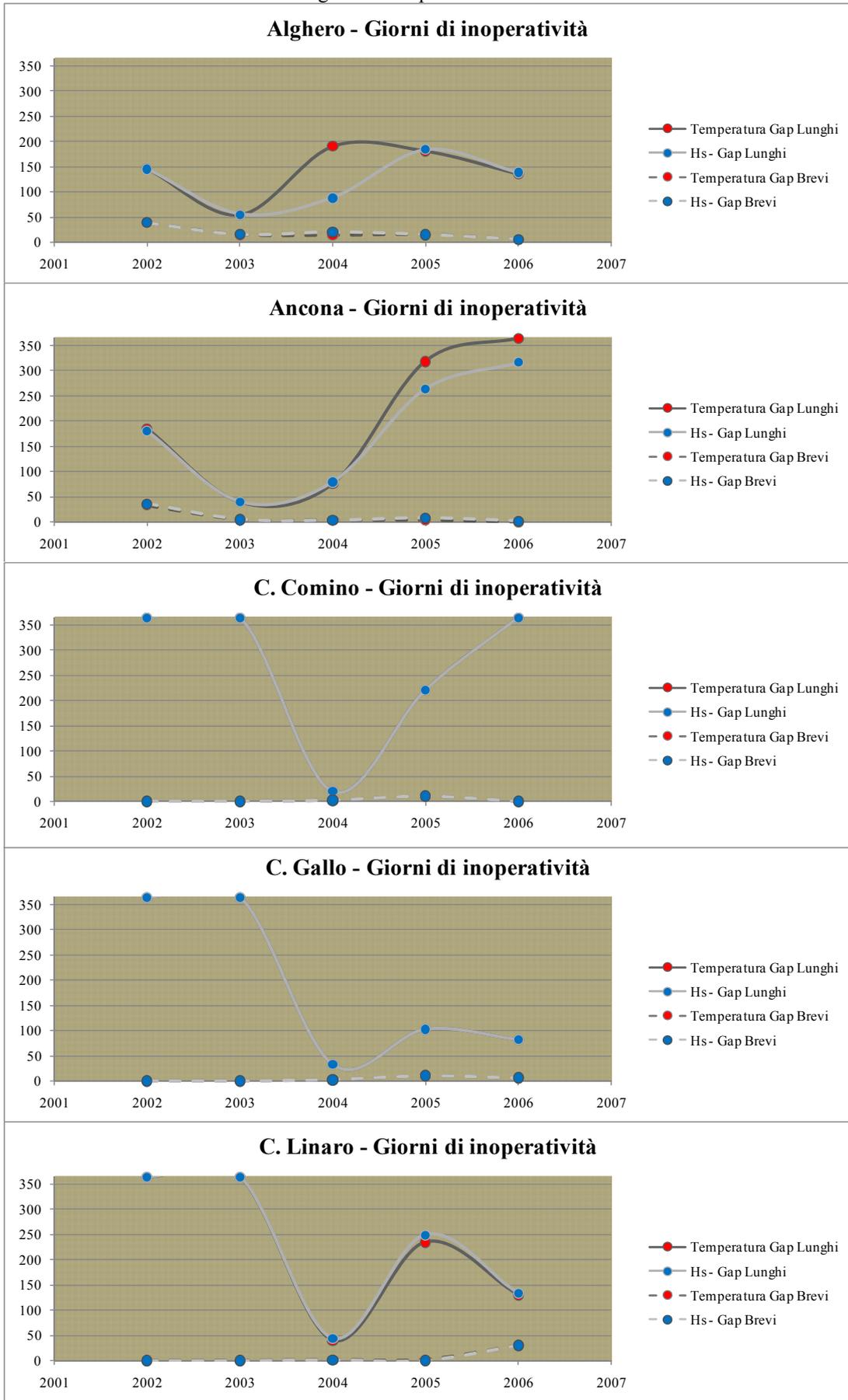
Si può notare come alcune boe abbiano avuto un buon rendimento nel corso degli anni (Cetraro, Crotone, Mazara, Monopoli), mentre altre hanno vissuto fasi altalenanti (Alghero, Ancona, Capo Gallo, Capo Linaro, Catania, La Spezia, Ortona, Ponza) e altre ancora hanno funzionato decisamente male (Capo Comino, Punta della Maestra).

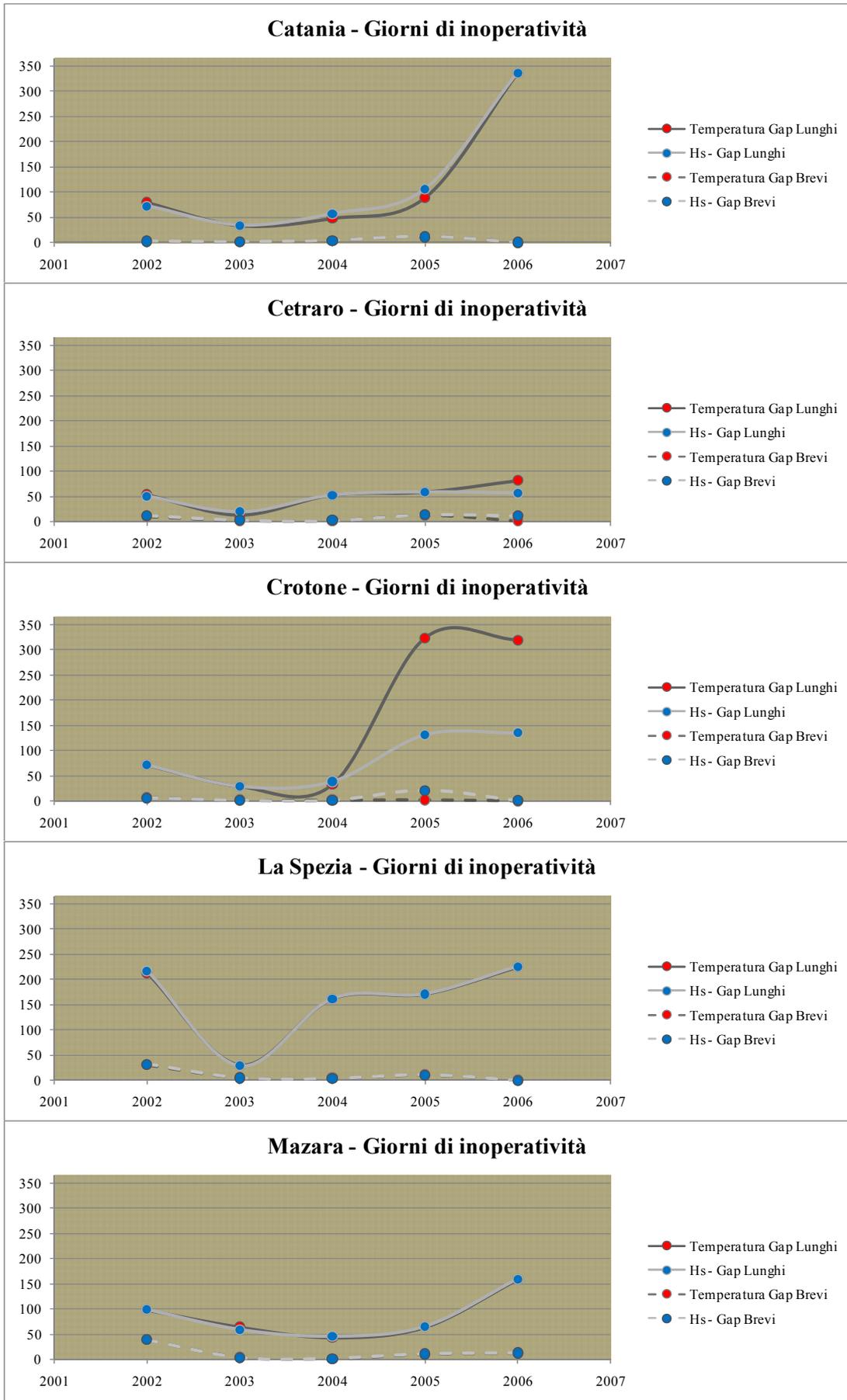
		DURATA							
		BREVE		MEDIA			LUNGA		LUNGHISSIMA
Temperatura	totale	< 1 giorno	< 3 giorni	< 6 giorni	< 12 giorni	< 24 giorni	< 60 giorni	< 120 giorni	> 120 giorni
Alghero	319	260	32	11	8	4	3	1	-
Ancona	90	57	19	6	4	2	1	-	1
C. Comino	13	5	3	1	1	-	-	1	2
C. Gallo	52	31	9	2	1	6	2	-	1
C. Linaro	19	8	3	-	2	2	1	1	2
Catania	48	20	8	8	4	6	-	1	1
Cetraro	54	24	17	3	6	2	2	-	-
Crotone	54	29	14	-	3	3	2	1	2
La Spezia	123	70	31	7	1	7	4	2	1
Mazara	218	187	16	5	4	4	2	-	-
Monopoli	87	54	15	6	4	4	3	-	1
Ortona	193	120	48	9	7	2	4	2	1
Ponza	51	24	5	10	3	6	2	-	1
P. d. Maestra	31	7	10	3	7	2	-	-	2

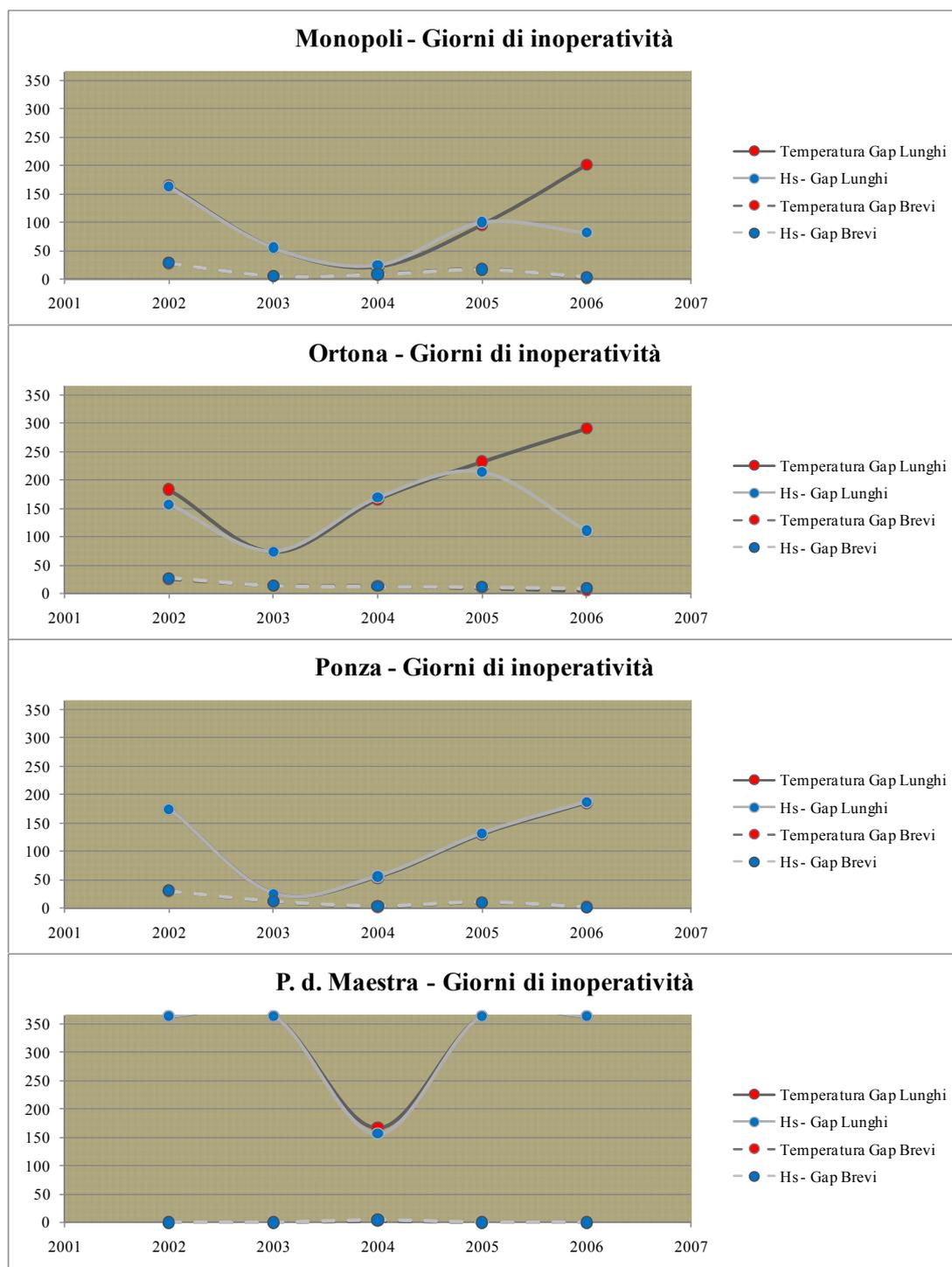
		DURATA							
		BREVE		MEDIA			LUNGA		LUNGHISSIMA
Hs	totale	< 1 giorno	< 3 giorni	< 6 giorni	< 12 giorni	< 24 giorni	< 60 giorni	< 120 giorni	> 120 giorni
Alghero	195	153	24	4	7	3	2	2	-
Ancona	122	74	20	12	4	5	5	1	1
C. Comino	14	6	3	1	1	-	-	1	2
C. Gallo	52	31	9	2	1	6	2	-	1
C. Linaro	29	12	7	2	2	2	1	1	2
Catania	54	26	5	7	8	6	-	1	1
Cetraro	48	21	15	1	8	2	1	-	-
Crotone	67	35	17	2	3	4	6	-	-
La Spezia	121	68	31	7	1	7	4	2	1
Mazara	186	154	17	5	4	4	2	-	-
Monopoli	84	52	14	4	6	5	3	-	-
Ortona	209	128	51	12	7	5	4	2	-
Ponza	50	25	5	9	3	4	3	-	1
P. d. Maestra	39	17	9	3	7	1	-	-	2

Tab. 2. 8 – Suddivisione dei gap lunghi in 8 classi

Fig. 2. 8 – Inoperatività annua







L'andamento annuale dell'inoperatività complessiva della rete espressa in giorni (totale delle osservazioni semiorarie mancanti in tutte le boe) (Fig. 2. 9) differisce sensibilmente dal numero di episodi complessivo (numero totale di gap annuale), evidenziando che nel tempo sono variate le cause che hanno determinato l'assenza del dato. Nel 2004, ad esempio, si registra un minimo nell'inoperatività in corrispondenza di un massimo nel

numero degli episodi registrati: ciò implica che sono stati molti i gap mediamente brevi che hanno portato all'assenza nell'intera rete di circa "1000 giorni" di dati.

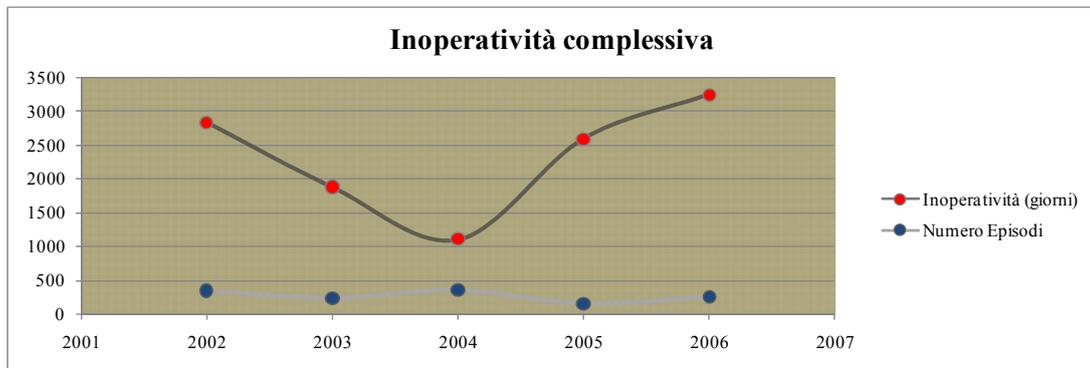


Fig. 2. 9 - Inoperatività complessiva

E' stato dunque introdotto il ritardo medio globale in giorni come il rapporto tra l'indisponibilità complessiva annuale ed il numero di gap lunghi nell'anno considerato in tutte le stazioni [4]. Un valore del ritardo molto minore di 1 indica che ci sono molti piccoli gap, un valore maggiore di 1 indica che i gap sono relativamente pochi e di lunga durata. Dalla Fig. 2. 10 si può osservare come il ritardo medio globale si mantenga pressoché costante tra il 2002 e il 2003, tende a diminuire drasticamente nel 2004 per raggiungere un massimo nel 2005 (il peggior anno per quanto riguarda la qualità dei dati) e decrescere nuovamente nel 2006. In ogni caso, in questo tipo di diagramma pesano eccessivamente i periodi lunghissimi (anche di 2 anni) di interruzione del servizio relativo ad alcune boe. Il ritardo globale medio è un importante indicatore del livello di affidabilità ed obsolescenza della rete; si nota, infatti come gli interventi di ripristino delle stazioni sia elevato e tendenzialmente crescente nel tempo.

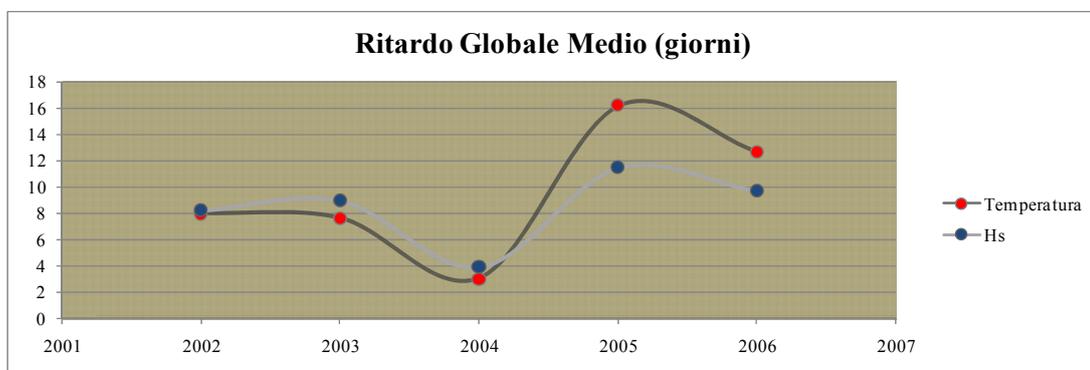


Fig. 2. 10 - Ritardo globale medio

2.6 - Metodi statistici di ricostruzione dei dati mancanti

L'analisi e il trattamento dei dati incompleti costituisce un problema di grande interesse nell'ambito della ricerca sulla qualità dei dati in un'indagine statistica. L'incompletezza dei dati è riconducibile essenzialmente a due cause principali [14]:

- a) la *mancata risposta totale (unit non-response)* che si riferisce alla totale assenza di informazione su talune unità del campione;
- b) la *mancata risposta parziale (MRP nel seguito)* che ha luogo quando, per una data unità, non sono disponibili i valori di alcune variabili, come accade ad esempio per dati ondametrici.

Le strategie più comuni per affrontare il problema della MRP sono essenzialmente quattro [14]:

- 1) Utilizzare le osservazioni parziali;
- 2) Scartare tutte le unità con valori mancanti (*Complete case estimation*);
- 3) Usare tecniche di riponderazione (*Maximum likelihood, Bayesian estimation, etc.*);
- 4) Produrre valori “artificiali” con i quali rimpiazzare i dati mancanti (*Hot deck imputation*).

La prima opzione, che potrebbe essere considerata la più “onesta”, delega essenzialmente all'utente la decisione sull'approccio da seguire nel trattamento delle *MRP*, consentendogli di adattare la strategia alle sue specifiche esigenze, ma esclude la possibilità di utilizzare metodi e software standard (sviluppati cioè per data-set completi) Tuttavia, questa metodologia è da considerarsi la più “disonesta” dal punto di vista statistico, poiché risulta impossibile ottenere inferenze corrette applicando gli stessi pesi ai valori presenti e ai dati mancanti.

La seconda strategia, adottata frequentemente, consiste nel basare le analisi di interesse sulle sole osservazioni complete. Sfortunatamente questo approccio produce serie distorsioni nelle stime a meno che i valori mancanti non costituiscano un campione casuale dell'intero insieme di dati (meccanismo di mancata risposta *Missing Completely at Random* o *MCAR*) o, in altre parole, la non-risposta sia indipendente da tutte le variabili di interesse. Inoltre, anche qualora il meccanismo di mancata risposta sia *MCAR*, il non utilizzo di tutta la dimensione campionaria comporta una diminuzione della precisione delle stime (aumento degli errori standard).

La terza strategia, usualmente adottata per la mancata risposta totale, si basa sull'utilizzo di variabili ausiliarie mediante le quali vengono definite celle di riponderazione per compensare la mancata risposta. Sebbene semplice in linea di principio, questo approccio presenta il non trascurabile svantaggio di richiedere la definizione di un diverso insieme di pesi per ogni variabile soggetta a *MRP*, rendendo peraltro problematica la stima di parametri legati alla relazioni di interdipendenza tra le variabili.

Il quarto approccio al trattamento della *MRP*, cui ci si riferisce comunemente con il nome di *imputazione*, è probabilmente il più popolare. La sostituzione dei valori mancanti con valori prodotti “artificialmente” consente di riprodurre un data-set completo sul quale, utilizzando strumenti standard, diverse analisi possono essere effettuate in modo consistente. Uno degli aspetti più critici di questo approccio consiste tuttavia nel fatto che, una volta prodotto un data-set “rettangolare” completo, i valori imputati tendono ad essere considerati come valori effettivamente osservati. La componente della variabilità delle stime associata alla non-risposta e al suo trattamento viene in tal modo trascurata, con un conseguente effetto di distorsione degli errori standard.

Diversi approcci sono stati proposti in letteratura per affrontare il problema dell'impatto della mancata risposta sulla precisione delle stime nelle indagini statistiche. Il più popolare è probabilmente quello dell'*Imputazione Multipla*, introdotta da Rubin negli anni '70, che consiste essenzialmente nella ripetizione del processo di imputazione m volte (con $m > 1$) e, conseguentemente, nella generazione di un insieme di m data-set completi (Fig. 2. 11) [30]. Rubin ha dimostrato che, purché il processo di imputazione soddisfi alcuni specifici requisiti, inferenze valide possono essere ottenute svolgendo analisi indipendenti sui singoli data-set completati e combinando opportunamente i risultati anche se la necessità di gestire un certo numero di data-set per ogni insieme di dati incompleto da trattare, può costituire un limite da un punto di vista operativo specialmente in presenza di grosse moli di dati.

Un modo tradizionale di classificare i metodi di imputazione è quello di distinguerli in *parametrici* e *non parametrici* a seconda che essi facciano uso o meno di un modello esplicito. Alla prima categoria appartiene, ad esempio, l'imputazione mediante regressione, mentre alla seconda l'imputazione con donatore casuale o con donatore di minima distanza. I metodi di imputazione più diffusi, in ogni caso, si basano sull'assunzione che il meccanismo di mancata risposta sia almeno *Missing at Random (MAR)*, cioè la mancata risposta, condizionatamente ai dati osservati, non dipenda dai valori mancanti; sotto questa ipotesi, infatti, è possibile effettuare inferenze valide sui parametri di interesse senza

specificare un modello per la non risposta. E' evidente che la validità di tale assunzione non è verificabile a meno che non si disponga di opportuni *training set* costruiti mediante ritorno alle unità non rispondenti [14].

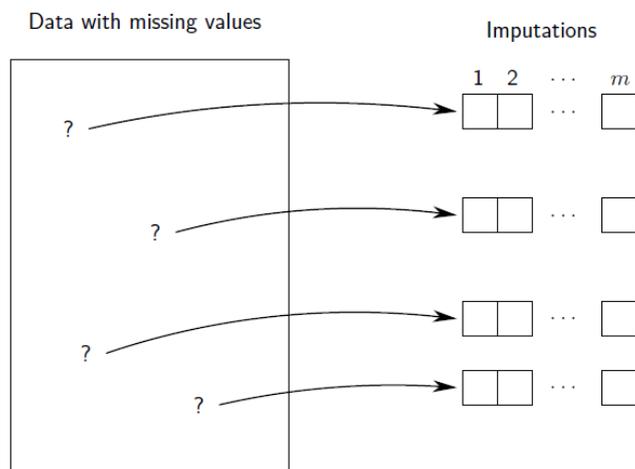


Fig. 2. 11 - Rappresentazione schematica del metodo di imputazione multipla, in cui m è il numero di imputazioni

2.6.1 - Imputazione Multipla: Teoria

L'*imputazione multipla* è essenzialmente un metodo Monte Carlo che consente di effettuare un'ampia classe di analisi inferenziali in presenza di non-risposta mediante analisi standard su diversi data-set completi [14]. Con l'imputazione multipla, ai valori mancanti Y_{mis} sono associati un certo numero m di insiemi di valori artificiali $Y_{mis}^{(1)}$, $Y_{mis}^{(2)}$, ..., $Y_{mis}^{(m)}$ in modo da ottenere m dataset completi che possono essere analizzati con metodi standard per dati completi. La variabilità dei risultati che si ottengono con le m analisi indipendenti effettuate, riflette l'incertezza associata alla mancata risposta e, combinata con la (eventuale) componente di variabilità di origine campionaria, può fornire una misura complessiva di incertezza nelle inferenze sui parametri di interesse.

Da un punto di vista Bayesiano l'imputazione multipla può essere vista come una procedura di generazione di m realizzazioni indipendenti dalla distribuzione predittiva a posteriori dei dati mancanti condizionatamente ai dati osservati $P(Y_{mis}|Y_{obs})$ [14]. Le procedure più comuni sono basate sull'assunzione di un modello parametrico esplicito per i dati completi e di un'opportuna distribuzione a priori (generalmente scelta non

informativa). In questo contesto, la distribuzione predittiva a posteriori dei dati mancanti può essere espressa come:

$$P(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}) = \int P(\mathbf{Y}_{mis} | \mathbf{Y}_{obs}, \theta) P(\theta | \mathbf{Y}_{obs}) d\theta$$

cioè come la media della distribuzione predittiva condizionata $P(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}, \theta)$ di \mathbf{Y}_{mis} dati i parametri θ , rispetto alla distribuzione a posteriori dei parametri “a dati osservati” $P(\theta | \mathbf{Y}_{obs})$. Mentre la generazione dei dati mancanti \mathbf{Y}_{mis} dalla $P(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}, \theta)$ non presenta particolari difficoltà, la distribuzione a posteriori a dati osservati $P(\theta | \mathbf{Y}_{obs})$ è in generale intrattabile. Si è soliti pertanto ricorrere a tecniche di tipo MCMC (Markov Chain Monte Carlo) che consentono di ricondurre il problema della generazione dalla distribuzione $P(\theta|\mathbf{Y}_{obs})$ a quello più semplice della generazione dalla distribuzione “a dati completi” $P(\theta|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$. Uno dei procedimenti più comuni consiste nel seguente schema iterativo, suggerito da Shafer nel 1997 ([14],[15],[28],[39]):

- **I-Step:** dato un insieme di valori correnti $\theta^{(t)}$ per i parametri, generare dalla distribuzione predittiva condizionata $P(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}, \theta^{(t)})$ un insieme di valori dei dati mancanti $\mathbf{Y}_{mis}^{(t+1)}$;
- **P-step:** condizionatamente a $\mathbf{Y}_{mis}^{(t+1)}$, generare dalla distribuzione a posteriori a dati completi $P(\theta | \mathbf{Y}_{obs}, \mathbf{Y}_{mis}^{(t+1)})$ un nuovo insieme di parametri $\theta^{(t+1)}$.

Lo schema descritto è noto in letteratura come *Data Augmentation*. Esso fornisce, a partire da un insieme di parametri iniziali $\theta^{(0)}$, una successione aleatoria $\{\theta^{(t)}, \mathbf{Y}_{mis}^{(t)}\}_{t=1,2,\dots}$ la cui distribuzione stazionaria è $P(\mathbf{Y}_{mis}, \theta | \mathbf{Y}_{obs})$. In particolare le sottosuccessioni $\{\theta^{(t)}\}_{t=1,2,\dots}$ e $\{\mathbf{Y}_{mis}^{(t)}\}_{t=1,2,\dots}$ hanno come distribuzioni stazionarie $P(\theta | \mathbf{Y}_{obs})$ e $P(\mathbf{Y}_{mis} | \mathbf{Y}_{obs})$ rispettivamente.

Nel caso in cui il modello assunto per i dati completi sia normale multivariato si ha

$$\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma}), (\mathbf{Y}_{mis}, \mathbf{Y}_{obs}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

dove $\boldsymbol{\mu}$ è il vettore p -dimensionale delle medie e $\boldsymbol{\Sigma}$ la matrice $p \times p$ di varianze e covarianze. Il passo di *I-step* non presenta particolari difficoltà: in pratica, avendo supposto indipendenti le n osservazioni del dataset da analizzare, esso si riduce, per ogni ciclo t e per ogni record incompleto u_i , alla generazione dalla distribuzione di probabilità condizionata

$$P(\mathbf{Y}_{M(\mathcal{P})} | \mathbf{y}_{iO(\mathcal{P})}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})$$

dove \mathcal{P} è il *pattern* di mancata risposta del record u_i e $O(\mathcal{P}), M(\mathcal{P})$ sono i relativi insiemi di indici che si riferiscono alle variabili osservate e non osservate rispettivamente.

Per quanto riguarda il passo di *P-step*, un'appropriata distribuzione a priori $\pi(\theta)$ per i parametri $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ può essere costruita a partire dalla distribuzione di Wishart inversa che è la distribuzione coniugata naturale per la funzione di verosimiglianza relativa a dati normali multivariati [14]. Più precisamente $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ può essere specificata assumendo che la distribuzione condizionata di $\boldsymbol{\mu}$ data $\boldsymbol{\Sigma}$ sia normale multivariata e che $\boldsymbol{\Sigma}$ abbia distribuzione di Wishart inversa:

$$\begin{aligned}\boldsymbol{\mu} | \boldsymbol{\Sigma} &\sim N(\boldsymbol{\mu}_0, \tau^{-1} \boldsymbol{\Sigma}) \\ \boldsymbol{\Sigma} &\sim W^{-1}(q, \Lambda)\end{aligned}$$

dove $\boldsymbol{\mu}_0, \tau, q, \Lambda$ sono parametri noti detti *iperparametri*. Se non sono disponibili informazioni a priori sui parametri, una appropriata distribuzione a priori non informativa si ottiene ponendo:

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}$$

dove $|\boldsymbol{\Sigma}|$ indica il determinante di $\boldsymbol{\Sigma}$.

Tramite la funzione di verosimiglianza relativa alla distribuzione multinormale basata sui dati $\mathbf{Y}^{(t)} \equiv (\mathbf{Y}_{obs}, \mathbf{Y}_{mis}^{(t)})$ si ottiene finalmente la distribuzione a posteriori a dati completi:

$$\begin{aligned}P(\boldsymbol{\mu}^{(t)} | \boldsymbol{\Sigma}^{(t)}, \mathbf{Y}^{(t)}) &= N(\bar{\mathbf{y}}^{(t)}, n^{-1} \boldsymbol{\Sigma}^{(t)}) \\ P(\boldsymbol{\Sigma}^{(t)} | \mathbf{Y}^{(t)}) &= W^{-1}(n-1, (n\mathbf{S}^{(t)})^{-1})\end{aligned}$$

dove n rappresenta la numerosità del dataset, e $\bar{\mathbf{y}}^{(t)}, \mathbf{S}^{(t)}$ sono rispettivamente la media e la matrice di varianze e covarianze campionarie calcolate sulla base dei dati $\mathbf{Y}^{(t)}$.

Come ultima osservazione sul metodo della *data augmentation* è utile sottolineare che, affinché si possa considerare ogni insieme di valori $\mathbf{Y}_{mis}^{(t)}$ come generato dalla distribuzione predittiva $P(\mathbf{Y}_{mis} | \mathbf{Y}_{obs})$ è necessario che t sia sufficientemente grande da garantire la stazionarietà; per la generazione di dataset multipli, si è soliti pertanto considerare solo valori di t maggiori di un valore prefissato t_0 (*burn-in period*) che si ritiene abbastanza elevato [14]. Inoltre la richiesta di indipendenza per le imputazioni multiple effettuate suggerisce di non utilizzare data-set ottenuti da iterazioni consecutive dello schema descritto. Al contrario, è conveniente sottocampionare dalla successione $\{\boldsymbol{\theta}^{(t)}, \mathbf{Y}_{mis}^{(t)}\}_{t=t_0, t_0+1, \dots}$ estraendo da essa a passo costante K , dove K è abbastanza grande da poter considerare trascurabile la dipendenza tra gli insiemi estratti.

Una volta ottenuto mediante imputazione multipla un insieme di m dataset completi, è semplice effettuare inferenze su un parametro di interesse Q della popolazione di riferimento combinando opportunamente i risultati delle m analisi effettuate sui singoli

data-set completati. Supponiamo che $\hat{Q} = \hat{Q}(Y_{obs}, Y_{mis})$ sia una stima puntuale a dati completi per Q , cioè la stima che si otterrebbe se tutti i dati fossero osservati. Sia inoltre $U = U(Y_{obs}, Y_{mis})$ la varianza associata alla stima \hat{Q} e supponiamo che possa considerarsi valida l'approssimazione [14]:

$$\hat{Q}(Y_{obs}, Y_{mis}) \approx E(Q | Y_{obs}, Y_{mis})$$

$$U(Y_{obs}, Y_{mis}) \approx V(Q | Y_{obs}, Y_{mis})$$

con $E(Q | \cdot)$, $V(Q | \cdot)$ rispettivamente media e varianza a posteriori di Q . Per il t -esimo dataset completato, potranno essere calcolate le quantità $\hat{Q}^{(t)} = \hat{Q}(Y_{obs}, Y_{mis}^{(t)})$ e $U^{(t)} = U(Y_{obs}, Y_{mis}^{(t)})$. La stima puntuale di Q basata sull'imputazione multipla sarà allora ([14], [19], [25], [36], [39]):

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^m \hat{Q}^{(t)}$$

e la varianza associata può essere calcolata come:

$$T = \bar{U} + (1+m^{-1})B$$

dove:

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m U^{(t)} \quad (\text{varianza within})$$

$$B = \frac{1}{m-1} \sum_{t=1}^m (\hat{Q}^{(t)} - \bar{Q})^2 \quad (\text{varianza between}).$$

2.6.2 - *Imputazione Multipla: Metodo*

E' possibile applicare il modello di imputazione multipla (MI) seguendo idealmente tre step successivi (Fig. 2. 12) ([27] [28] [29] [38] [39]):

- Imputazione (*Imputation*): vengono generati (riempiti) $m > 1$ set di valori ricostruiti plausibili;
- Analisi (*Analysis*): sono valutati i singoli dataset imputati
- Combinazione (*Pooling*): si aggregano i risultati dalle analisi degli m dataset.

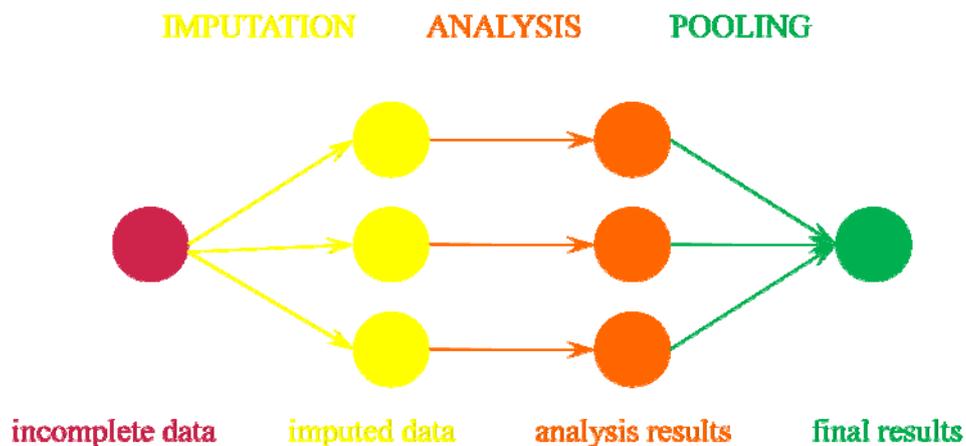


Fig. 2. 12 - Schematizzazione dell'algoritmo MI

Il primo step (imputazione) costituisce probabilmente la fase maggiormente critica poiché richiede la definizione di ipotesi semplificative circa la tipologia di dati mancanti (in genere MAR) e la relazione tra le variabili osservate (lineare, non lineare, etc.).

La generazione dei dati imputati viene di solito effettuata tramite un approccio di tipo *Markov Chain Monte Carlo* (MCMC). La maggior parte dei problemi che si riscontrano in questa fase sono dovuti:

- Al fatto che la mancanza del valore è legata al valore stesso condizionatamente ai valori delle altre covariate (per esempio, valori elevati di altezza d'onda sono mancanti a causa di problemi di trasmissione durante le mareggiate);
- In un determinato istante temporale possono apparire nel dataset solo valori mancanti;
- Le regressioni da effettuare devono essere fatte sul dataset completo; assume quindi una notevole importanza la ricostruzione corretta dei dati.

Lo step di analisi consente di valutare e registrare i valori sintetici (coefficienti della regressione, standard error, etc.) relativi alle singole imputazioni.

L'ultima fase, la combinazione, consente di mediare le singole imputazioni al fine di ottenere i dati ricostruiti e i relativi standard error. E' necessario che nei due precedenti step l'analisi sia stata svolta considerando un valore di $m > 1$: una singola imputazione, infatti, tende a sovrastimare la precisione del modello poiché non si tiene conto della *varianza between* tra le singole imputazioni [26].

L'algorithmo di imputazione multipla risulta particolarmente vantaggioso per una serie di ragioni:

- è complementare e compatibile con i metodi e i software standard di analisi dei dati completi ([15], [26]);
- gli m dataset imputati possono essere usati nelle varie analisi e valutati sia singolarmente che in forma aggregata ([19], [15], [26]);
- le inferenze statistiche ottenute (standard error, p-value, etc.) sono ritenute estremamente valide dalla comunità scientifica poiché includono l'incertezza dovuta alla mancanza dei dati [26];
- risulta efficiente anche per bassi valori di m . In molte applicazioni sono sufficienti da 3 a 5 imputazioni per ottenere risultati eccellenti. Rubin ha infatti dimostrato che l'efficienza di una stima basata su m imputazioni, intesa come capacità predittiva del modello, è approssimativamente ([29], [31]):

$$\left(1 + \frac{\gamma}{m}\right)^{-1}$$

in cui γ è la frazione di dati mancanti per il dataset da ricostruire. La tabella seguente mostra quanto rapidamente si incrementa l'efficienza all'aumentare di m [31].

m	γ				
	.1	.3	.5	.7	.9
3	97	91	86	81	77
5	98	94	91	88	85
10	99	97	95	93	92
20	100	99	98	97	96

Tab. 2. 9 - Efficienza percentuale dell'algorithmo MI in funzione del numero di imputazioni e della frazione di dati mancanti

A fronte di questi vantaggi, l'algoritmo MI presenta anche qualche difetto legato essenzialmente alla complessità del modello. I problemi che si riscontrano riguardano [26]:

- la lentezza di calcolo: il metodo è difficilmente applicabile con le attuali tecnologie per la previsione in tempo reale dei dati;
- la necessità di una grande capacità di immagazzinamento dei dati: l'algoritmo necessita di una memoria pari ad almeno $[1+m \cdot (\% \text{ missing values})]$ volte la dimensione originale del data-set;
- la necessità di definire delle ipotesi semplificative del modello per avere una migliore governabilità del fenomeno e della procedura di imputazione.

E' importante sottolineare che il metodo di imputazione multipla non pretende di *creare* informazioni attraverso valori simulati, ma semplicemente di aggregare le misurazioni osservate per poter effettuare le analisi tramite inferenze su dati completi.

Ricostruzione dei missing values della Temperatura dell'acqua

I dati di temperatura dell'acqua misurati dalle boe risentono fortemente dell'irraggiamento solare. Essi, quindi, presentano una forte componente stagionale ma, a differenza della temperatura dell'aria, non subiscono grandi variazioni nell'arco della giornata. La profondità a cui è posto il sensore (circa 45 cm), infatti, è caratteristica di una fascia di rimescolamento in cui l'acqua superficiale (primi micron di profondità) calda di giorno e fredda di notte, si unisce alle correnti più profonde che presentano andamenti della temperatura generalmente opposti. Si verificano quindi delle fluttuazioni giornaliere di modesta entità.

Il periodo analizzato (2002-2006) è caratterizzato da periodi di funzionamento della rete intermittenti a causa delle operazioni di ammodernamento, dei numerosi interventi di manutenzione e del grande numero di disormeggi. Sono rari i casi in cui le boe hanno funzionato contemporaneamente. La ricostruzione dei missing values appare quindi particolarmente complessa tramite approcci classici di regressione.

Fortunatamente, nello stesso periodo, la rete mareografica ha misurato con continuità la temperatura dell'acqua ed è quindi possibile utilizzare questi dati come predittori.

La ricostruzione dei dati mancanti viene effettuata tramite tecniche di imputazione multipla applicate sui dati relativi all'intero bacino afferente alle coste italiane e a singoli sottobacini quali il Tirreno, lo Ionio e l'Adriatico.

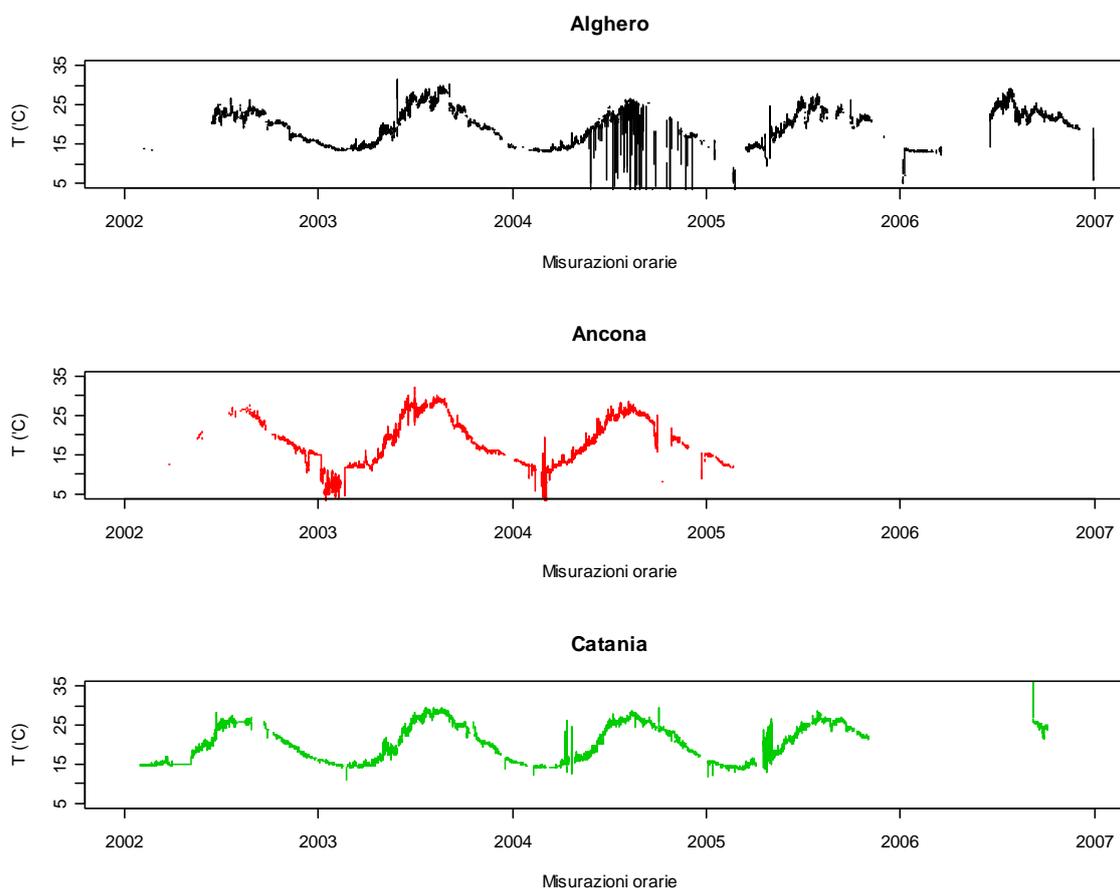
3.1 - Dati RON (anni 2002-2006)

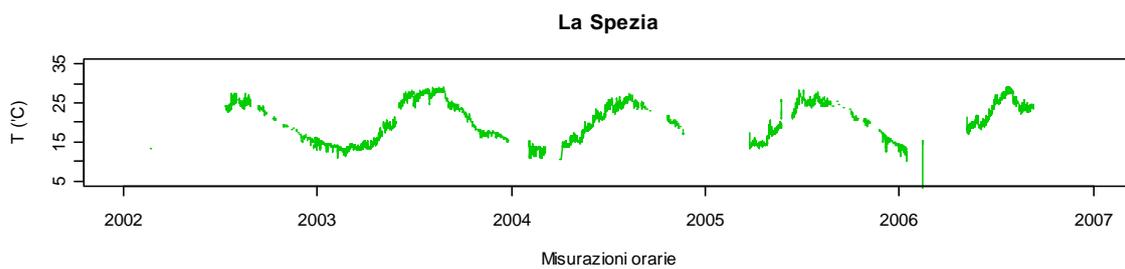
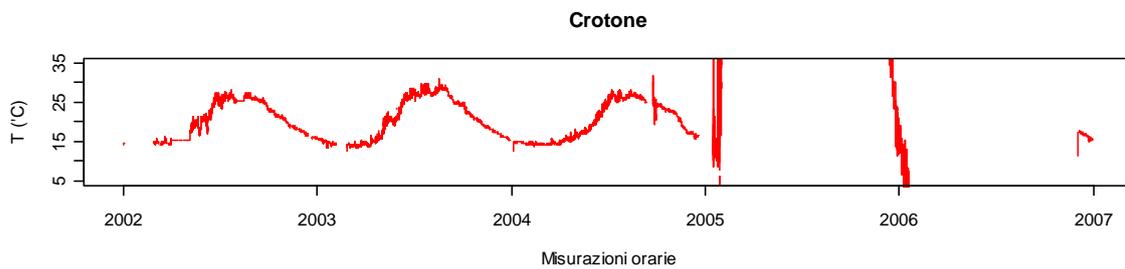
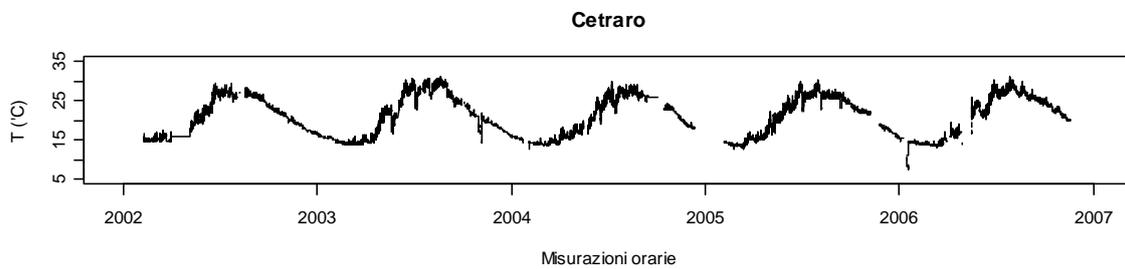
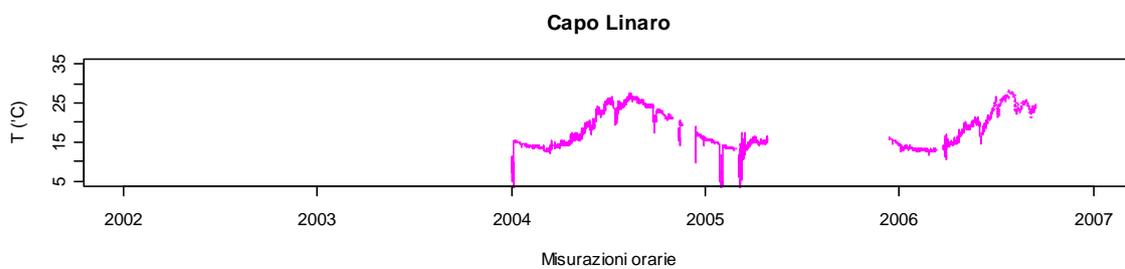
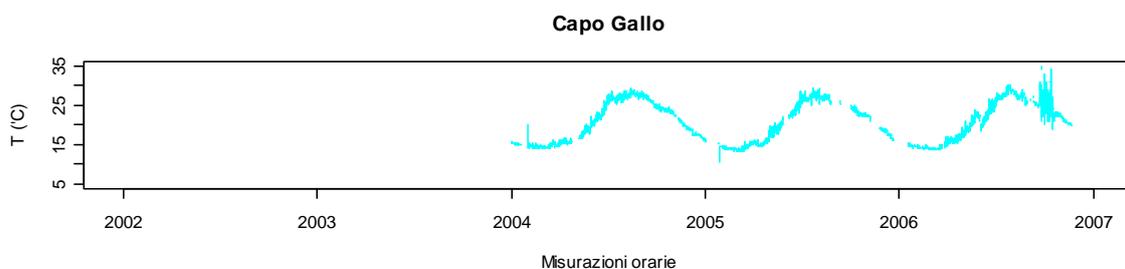
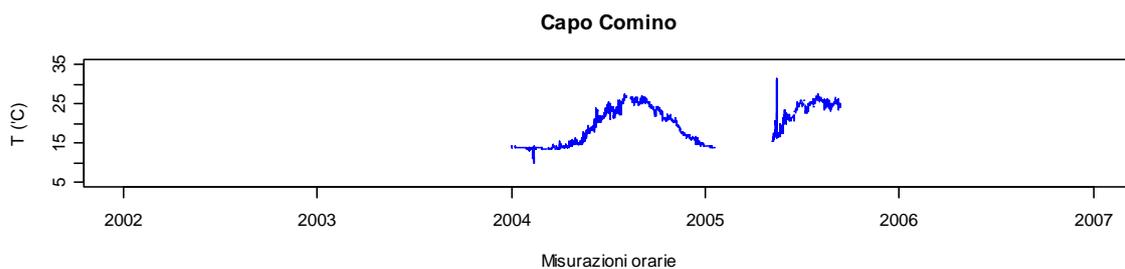
Il dataset delle Temperature dell'acqua della rete RON tra il 2002 e il 2006 si presenta estremamente discontinuo. Al consistente numero di gap dovuti all'assenza totale dei dati, si accosta la presenza di valori anomali (come ad esempio spike residui) e valori stagionali altamente improbabili (temperature inferiori a 10 °C in estate): questi dati non sono stati eliminati, se non tramite un filtro macroscopico che consente di cancellare i dati di

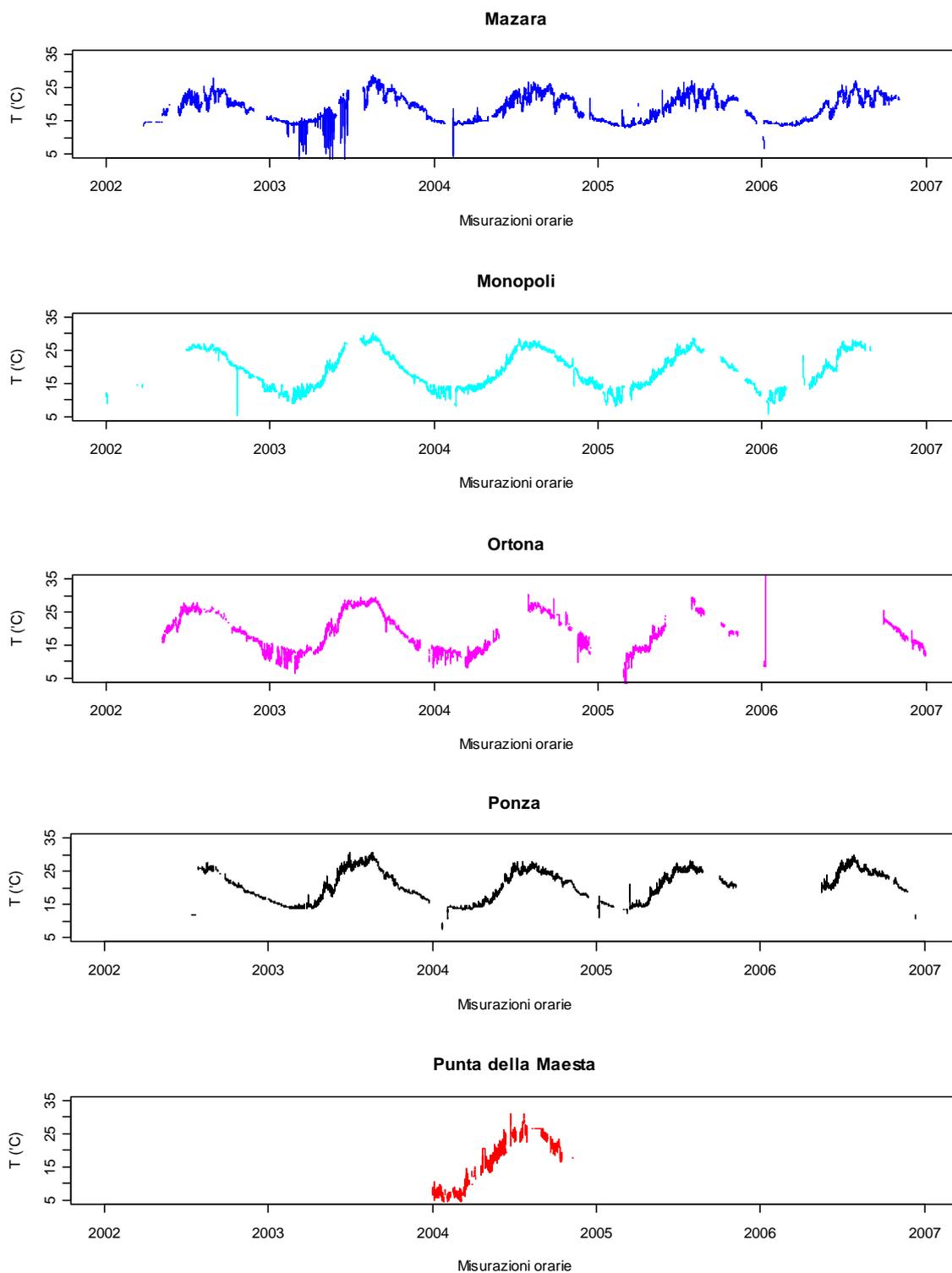
temperatura minori di 0°C e maggiori di 50°C, poiché la robustezza del modello di ricostruzione dei dati mancanti da imputare deve essere tale da identificare e ove possibile correggere anche tutte le anomalie.

Inoltre, come evidenziato dalla Fig. 3. 1, solo alcune boe presentano dati per tutto il periodo di osservazione. Cetraro e Monopoli mostrano una elevata qualità del dato mentre Punta della Maestra e Capo Comino sono caratterizzate da una percentuale di dati mancanti superiori al 70%.

Fig. 3. 1 - Andamento nel tempo della temperatura dell'acqua per ogni boa







I dati sono stati trattati in forma aggregata, considerando in maniera distinta le boe afferenti ai bacini del Tirreno, dell'Adriatico e dello Ionio (Tab. 3. 1).

I diagrammi seguenti (Fig. 3. 2) mostrano un'ottima sovrapposizione dei dati di temperatura: nel Tirreno, ad esempio, tra La Spezia (Liguria) e Capo Gallo (Sicilia), così come tra Ponza (Lazio) e Capo Comino (Sardegna), si registrano le stesse medie stagionali

e gli stessi trend annuali. Infatti, ogni sottobacino del Mediterraneo è relativamente chiuso, la circolazione non è particolarmente attiva, le altezze d'onda sono limitate (possono arrivare fino a 4 metri) segno di una scarsa attività di rimescolamento e l'irraggiamento si può considerare uniforme su tutto il bacino. Si nota anche una stretta correlazione tra la temperatura dell'acqua ed eventi (atmosferici e meteorici) di media e lunga durata che portano a brusche variazioni di temperatura persistenti nel tempo.

Mar Tirreno	Capo Comino Capo Gallo Capo Linaro Cetraro La Spezia Ponza
Mar Adriatico	Ancona Monopoli Ortona Punta della Maestra
Mar Ionio	Catania Crotone

Tab. 3. 1 - Suddivisione delle boe in base ai bacini

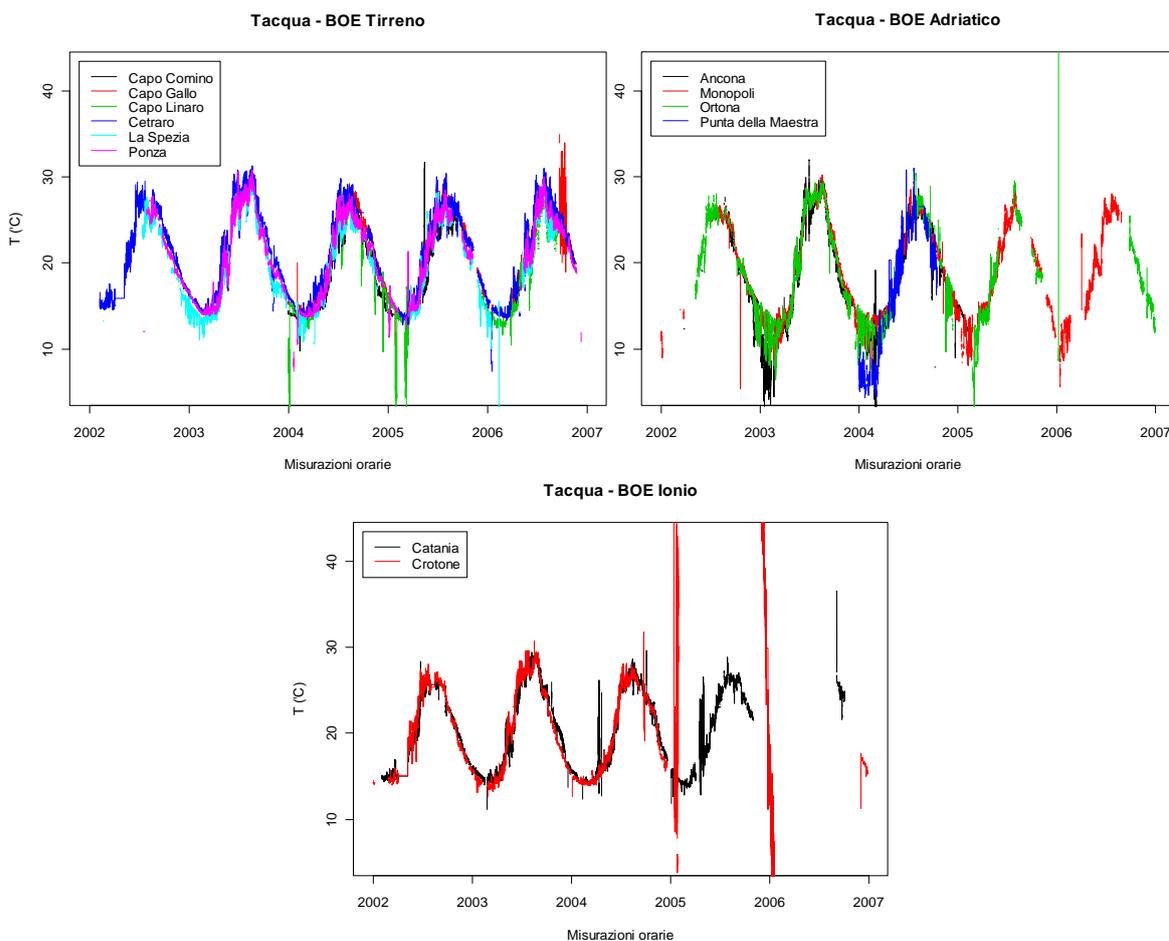


Fig. 3. 2 - Dati aggregati delle boe suddivise per bacini

Per individuare un modello capace di ricostruire in maniera accurata i dati di temperatura dell'acqua, occorre identificare e stimare i parametri caratteristici e risalire al processo generatore dei dati. E' necessario quindi effettuare una prima stima delle componenti stagionali del modello.

La stagionalità è tecnicamente definita come una dipendenza da correlazione di ordine k tra ciascun i -esimo elemento della serie, e l'elemento $(i-k)$ -esimo, ed è misurata tramite l'autocorrelazione. Il termine k è detto ritardo (lag). In pratica se la componente di errore non è troppo elevata, la stagionalità può essere visualizzata come uno schema che si ripete ogni k elementi.

L'analisi della correlazione parte dall'identificazione dei momenti teorici dei processi stocastici, in particolare dell'autocovarianza, intesa come la covarianza tra valori della serie Z in istanti temporali diversi [34]. Normalmente la covarianza misura la tendenza di due grandezze a variare nello stesso senso, in questo caso si utilizza un'unica variabile misurata in due istanti temporali diversi. In formule si ottiene:

$$\gamma(t, t + k) = E[(Z_t - \mu(t))(Z_{t+k} - \mu(t))] = cov(Z_t, Z_{t+k})$$

L'autocovarianza è funzione di due istanti temporali, t e $(t+k)$. In questo senso la varianza risulta essere un caso particolare dell'autocovarianza, ponendo $k=0$. L'autocorrelazione ha il vantaggio, rispetto all'autocovarianza, di essere compresa fra limiti fissi (i valori estremi sono -1 e $+1$). L'autocorrelazione si ottiene semplicemente dividendo l'autocovarianza per il prodotto degli scarti quadratici medi di Z_t e Z_{t-k} .

La funzione di autocorrelazione (ACF) è il coefficiente di correlazione lineare $\rho(k)$ tra le variabili casuali Z_t e Z_{t-k} calcolato al variare di k .

$$\rho(k) = Corr(Z_t, Z_{t-k}) = \frac{cov(Z_t, Z_{t-k})}{\sqrt{Var(Z_t)Var(Z_{t-k})}}$$

I modelli stagionali di serie storiche possono essere esaminati tramite correlogrammi, i quali illustrano graficamente e numericamente la funzione di autocorrelazione (ACF).

Lo studio dell'autocorrelazione è di fondamentale importanza per valutare la cosiddetta "memoria" della serie storica, ovvero quanto a lungo permane un'influenza di una particolare osservazione sulle osservazioni successive.

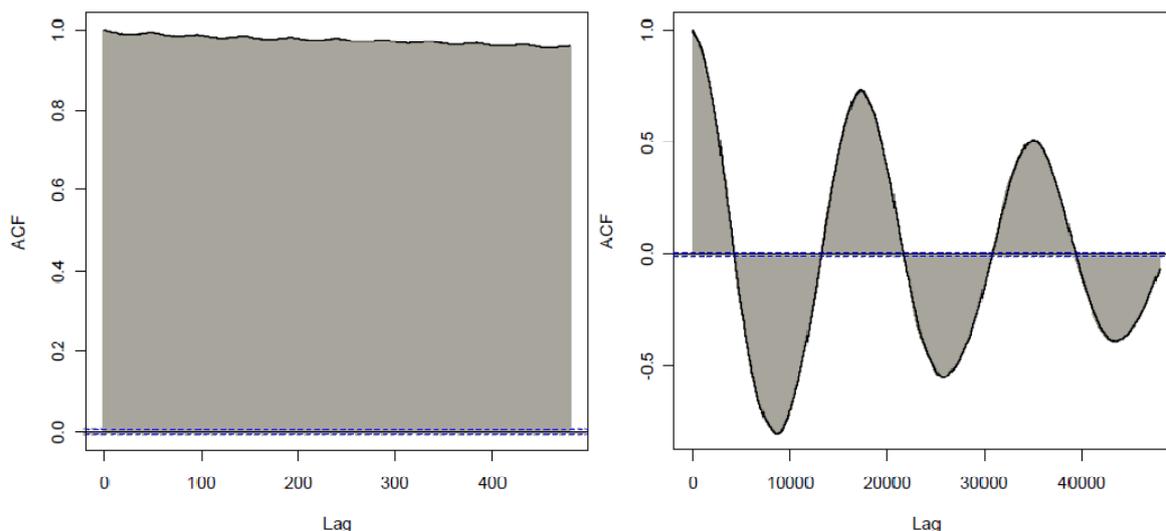


Fig. 3.3 - Correlogrammi relativi ai dati di temperatura dell'acqua (Boa di Cetraro) con Lag piccolo (a) e grande (b)

I correlogrammi consentono di dimostrare la dipendenza temporale delle osservazioni di temperatura dell'acqua della rete RON. Per ogni boa, essi presentano tutti lo stesso andamento rappresentato dalla Fig. 3.3. E' possibile notare una forte dipendenza tra i dati (i valori presenti all'interno dell'intervallo di confidenza sono in numero trascurabile) e la presenza di una marcata componente stagionale sia giornaliera (Lag 48) che annuale (Lag 17520). Come può sembrare ovvio, il primo dato (1 gennaio 2002 ore 0:00 - inverno) è strettamente correlato con i dati relativi alla stessa ora del giorno successivo, ma anche con tutti i dati misurati nei mesi invernali, mentre è inversamente correlato con i dati misurati nei mesi estivi.

3.2 - Boe e Mareografi (anni 2002-2006)

I mareografi misurano la temperatura dell'acqua tramite sensori immersi in pozzetti di calma collegati direttamente al mare tramite canalette profonde. Ci sono notevoli differenze tra i dati misurati dalle stazioni mareografiche e dalle boe ondometriche legate essenzialmente a:

- Posizione delle stazioni: i mareografi solitamente si trovano in aree portuali o comunque lungo i litorali e risentono quindi delle dinamiche costiere (interferenze con opere marittime, correnti costiere, presenza di bassi fondali, etc.) non confrontabili con i fenomeni in mare aperto.
- Sensori di misurazione: nelle stazioni mareografiche è montato un termometro a resistenza di platino tipo 'tip-sensitive' (SIAP TM9557) con puntale in acciaio inossidabile [17] con caratteristiche mostrate in Tab. 3. 2a. La boa, invece, adotta un sensore YSI 44203 [10] con termoresistori che producono un output proporzionale alla temperatura (Tab. 3. 2b)

MAREOGRAFI		BOE	
Termometro SIAP TM9557		Sensore YSI 44203	
Tipo	Stagno ad immersione	Tipo	Termoresistori di precisione
Materiale punta	Acciaio AISI 316	Materiale punta	Acciaio AISI 316
Dimensioni	10 x 300 mm	Dimensioni	8 x 55 mm
Campo di misura	Da -40 a +60 °C	Campo di misura	Da -30 a +50 °C
Accuratezza	0,1 °C	Accuratezza	0,15 °C ±0.12 °C (-30 °C) ±0.07 °C (11 °C) ± 0.02 °C (50 °C)
Alimentazione	20 V	Alimentazione	3 V
Consumo	1 mA	Consumo	< 1 mA

Tab. 3. 2 - Specifiche tecniche dei sensori di temperatura montati su mareografi (a) e su Boe (b)

- Frequenza di misurazione: mentre le boe hanno una frequenza semioraria di misurazione della temperatura dell'acqua, i mareografi effettuano il campionamento ogni ora; questo risulta un problema marginale poiché la variabilità di questa grandezza è modesta. Non si commette quindi un grande errore nel ricavare dati semiorari a partire da quelli orari tramite interpolazione semplice.

- Profondità del campione misurato: i mareografi misurano la temperatura dell'acqua in un pozzetto di calma che pesca acqua generalmente a due metri di profondità, mentre la boa misura la temperatura riferita a strati profondi circa 45 cm. Appare evidente, quindi, che il dato misurato dai mareografi risentirà minimamente dell'irraggiamento solare diretto e tenderà a non presentare le tipiche escursioni termiche tra il giorno e la notte evidenziate dai dati misurati in mare aperto (Fig. 3. 4).
- Numerosità dei dati: i mareografi non presentano, fortunatamente, problemi che invece sono tipicamente riscontrabili nella rete ondometrica, quali ad esempio disormeggi, problemi di trasmissione via radio, dismissione a terra causa manutenzione, urti con natanti e pesci di grosse dimensioni. Per questo motivo, i dati presenti nel dataset RMN sono estremamente affidabili e numerosi (i dati mancanti non sono mai superiori al 10%)

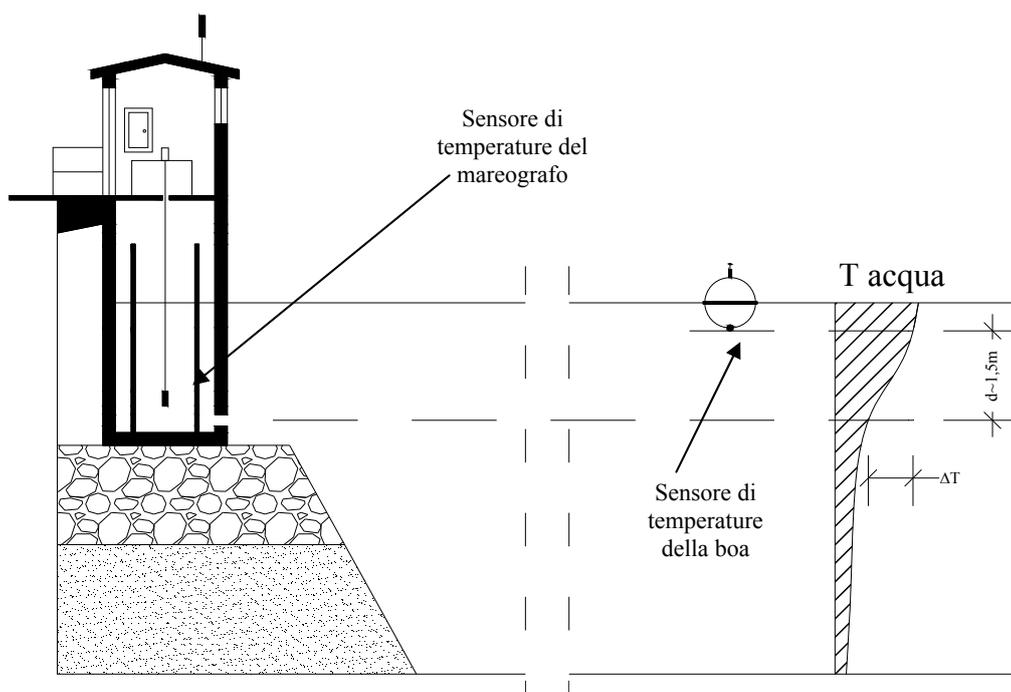


Fig. 3. 4 - La temperatura dell'acqua misurata dalla boa e dal mareografo. La differenza tra le misure è legata alla differenza di profondità del sensore

Ciò nonostante, i mareografi si dimostrano degli ottimi predittori anche in considerazione del fatto che alle profondità tipiche delle misurazioni (ondametriche e mareografiche) il processo generatore della temperatura dell'acqua è lo stesso e non viene influenzato dalle diverse caratteristiche delle stazioni di misura.

Risulta quindi di fondamentale importanza lo studio delle relazioni che intercorrono tra i dati derivanti da entrambe le reti. Particolarmente interessante è la valutazione della

correlazione di misurazioni relative a stazioni relativamente vicine tra di loro e afferenti allo stesso sottobacino marino (Tirreno, Adriatico, Ionio), come riportato nella seguente tabella:

BOA	MAREOGRAFO
Alghero	Porto Torres
Ancona	Ancona
Catania	Catania
Cetraro	Palinuro
Chioggia	Venezia
Capo Linaro	Civitavecchia
Capo Linaro	Napoli
Crotone	Crotone
La Spezia	Livorno
Mazara	Porto Empedocle
Monopoli	Bari
Ortona	Ortona
Capo Gallo	Palermo
Ponza	Napoli
Capo Comino	Napoli

Tab. 3. 3 - Correlazioni Boa-Mareografo

Per ogni coppia Boa-Mareografo sono stati individuati:

- 1) gli andamenti nel tempo delle temperature;
- 2) la correlazione tra i dati di temperatura dell'acqua misurati dalla Boa e dal Mareografo;
- 3) l'andamento dei residui;
- 4) i correlogrammi dei residui con lag piccoli e grandi.

Le coppie di stazioni che sono caratterizzate da un minor numero di dati mancanti e che mostrano una migliore sovrapposizione tra misurazioni ondametriche e mareografiche riguardano

- Cetraro – Palinuro;
- Crotone – Crotone;
- La Spezia – Livorno;
- Monopoli – Bari.

In queste coppie di dati è possibile valutare la quantità del dato presente e l'ottima sovrapposizione delle osservazioni derivanti da boe e mareografi (Fig. 3. 5 esempio riferito alle stazioni di Cetraro e Palinuro), nonché la fortissima correlazione tra le due misurazioni dimostrata dalle rette di regressione in Fig. 3. 6a che presenta un'inclinazione di circa 45°.

I residui tra i due set di misure solo raramente raggiungono valori elevati (intorno ai 4°C) e tipicamente nelle stagioni estive in cui le misure dalle boe mostrano sbalzi termici non riscontrabili nelle osservazioni dai mareografi (Fig. 3. 6b)

L'andamento dei correlogrammi dei residui (Fig. 3. 7) risulta essere molto vicino ad un'armonica smorzata con flessioni periodiche di lag 24 dovute esclusivamente a variazioni giornaliere. I correlogrammi decrescono tendenzialmente a zero secondo una curva esponenziale. Valori prossimi allo zero si raggiungono molto prima rispetto ai correlogrammi illustrati nel paragrafo precedente (Fig. 3. 3). Introducendo quindi nell'analisi i dati dei mareografi si può eliminare un trend di stagionalità relativo alle variazioni annuali (inverno-estate) della temperatura. Inoltre questi diagrammi dimostrano la dipendenza delle osservazioni non solo nel tempo, ma anche nello spazio, tra stazioni lontane o addirittura posizionate agli estremi dello stesso bacino

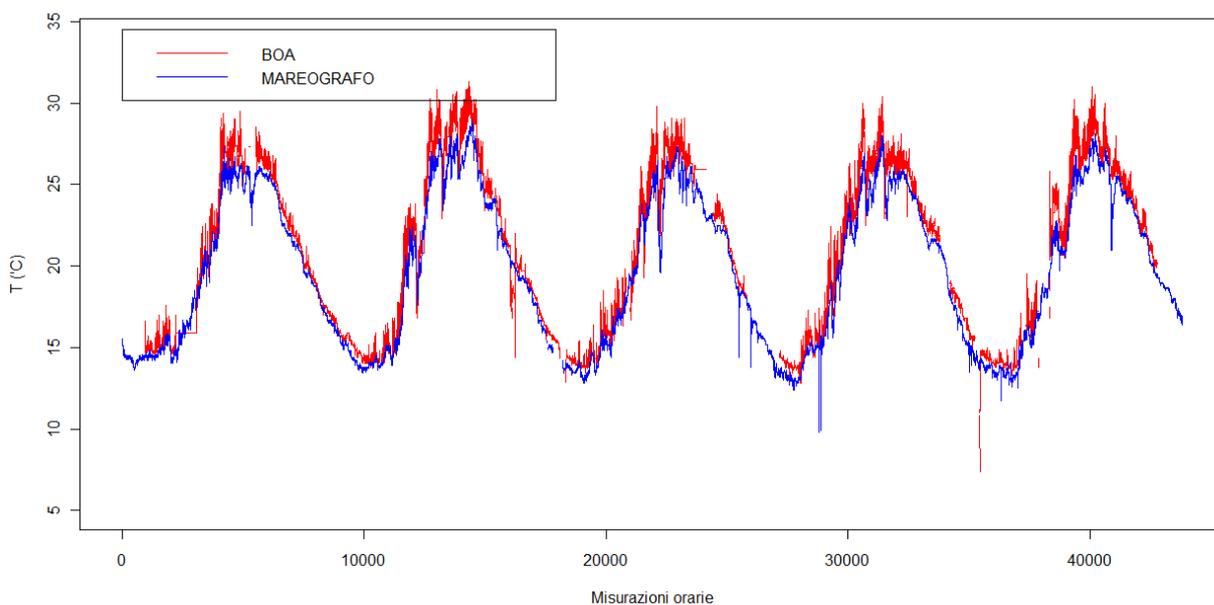


Fig. 3. 5 - Andamento nel tempo delle temperature dell'acqua misurate dalla Boa di Cetraro e dal Mareografo di Palinuro

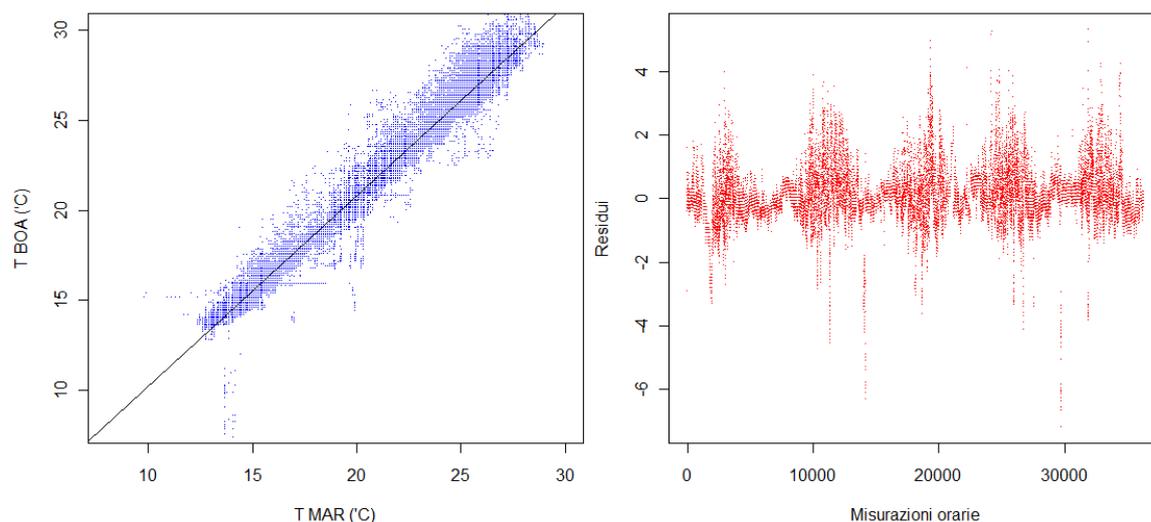


Fig. 3. 6 – Boa di Cetraro e Mareografo di Palinuro: (a) Correlazione tra i dati di temperatura dell'acqua misurati da Boe e Mareografi; (b) Andamento dei residui

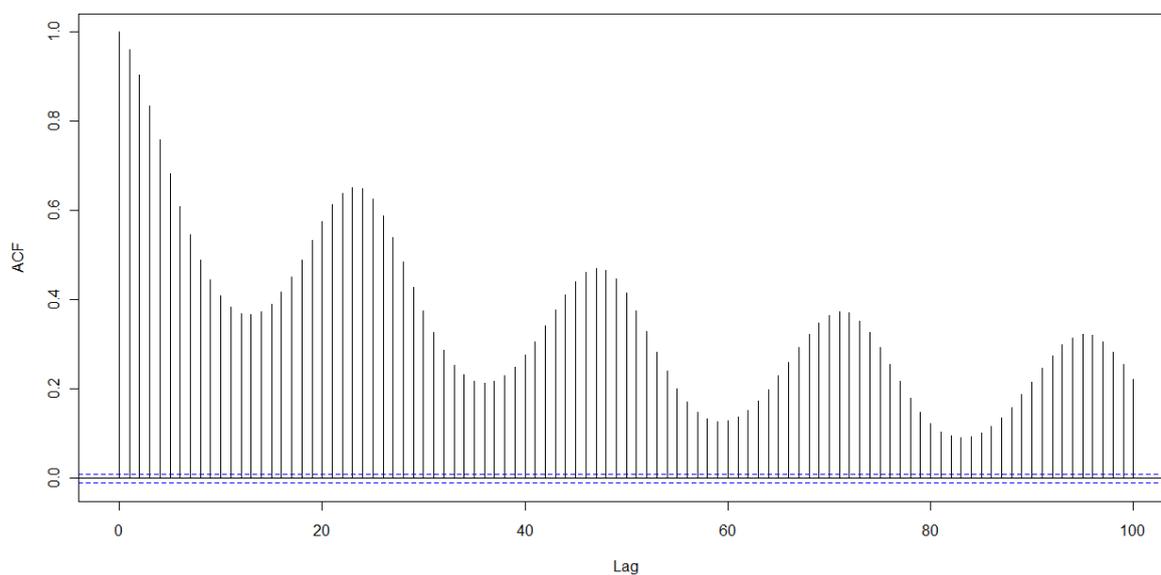


Fig. 3. 7 - Correlogramma dei residui dei dati di temperatura dell'acqua misurati alla boa di Cetraro e al mareografo di Palinuro

3.3 - Imputazione multipla dei dati mancanti

Il modello di ricostruzione dei dati mancanti si basa su metodi di imputazione multipla. Come dimostrato nel paragrafo precedente, date le serie storiche prese in esame, è possibile ipotizzare dei forti legami dimensionali, spaziali e temporali tra i dati. L'applicazione di modelli semplici di interpolazione dei dati mancanti potrebbe risultare quindi estremamente inadeguato. E' possibile, invece, accostare i metodi di imputazione multipla a modelli autoregressivi che tengano conto delle relazioni tra il singolo dato valutato nella boa i al tempo t e altri dati relativi a strumenti di misura j (boe o mareografi) al tempo τ .

In questo paragrafo verrà individuato solo il modello di imputazione multipla che consente la ricostruzione dell'intero dataset (relativo sia a boe che mareografi) e l'individuazione degli standard error associati ad ogni dato imputato. Vengono quindi introdotte delle ipotesi forti sulla distribuzione dei dati e dei missing values:

- Dato mancante casualizzato (*Non-Response Mechanism* di tipo *Missing at Random MAR*): si assume che la probabilità che il dato sia mancante, condizionatamente ai valori delle altre covariate, dipenda dai dati osservati e non dai dati mancanti. In questo caso i valori mancanti possono essere ricostruiti sfruttando proprio tale collegamento con i dati osservati [30];
- Dati statisticamente indipendenti nel tempo: l'osservazione al tempo t è statisticamente indipendente dall'osservazione al tempo $t-\tau$. Questa ipotesi forte, introdotta a scopo semplificativo solo per la definizione del modello di imputazione multipla, dovrà essere eliminata in futuro con l'introduzione di modelli autoregressivi;
- Ipotesi di normalità: i valori in risposta alle variabili esplicative si distribuiscono secondo la normale gaussiana; i valori individuati corrispondono ai valori medi;
- Legame lineari tra le variabili: le funzioni lineari ben approssimano le eventuali relazioni non lineari presenti tra le variabili.

Il modello di imputazione multipla è stato implementato grazie all'utilizzo del software statistico R attraverso la libreria MICE (*Multivariate Imputation by Chained Equations*). Il codice in R consente di:

- chiamare in input i valori di temperatura presenti nei dataset RON e RMN relativi ad ogni stazione presente nel bacino considerato;
- riempire il dataset con eventuali campi mancanti;

- assegnare ad ogni intervallo di tempo ($\Delta t = 1$ ora) nel periodo di riferimento (anni 2002-2006) un valore di temperatura dell'acqua (valore disponibile o non disponibile) per ogni strumento di misura;
- ordinare il dataset in funzione della numerosità dei missing values;
- eseguire l'algoritmo di imputazione multipla con assegnato numero di imputazioni e di iterazioni;
- analizzare e combinare i risultati;
- stampare i risultati in termini di serie storica ricostruita e di standard error.

Il dataset viene ordinato in funzione del numero di dati mancanti (la prima colonna è caratterizzata da un minor numero di dati mancanti, l'ultima dal maggior numero di dati mancanti) in modo da massimizzare l'effetto dell'imputazione multipla [35]. L'algoritmo MI viene eseguito con 10 imputazioni e 5 cicli. Anche se, come dimostrato dalla Fig. 3. 8, all'aumentare del numero di imputazioni aumenta la precisione della ricostruzione (diminuiscono gli standard error), per questioni legate ai tempi di esecuzione dell'algoritmo si preferisce limitare il numero degli step di calcolo. E' importante, infatti, definire il modello e valutarne la robustezza. In un secondo momento, quando verranno inserite anche le dipendenze spazio-temporali si potrà procedere ad un miglioramento della precisione.

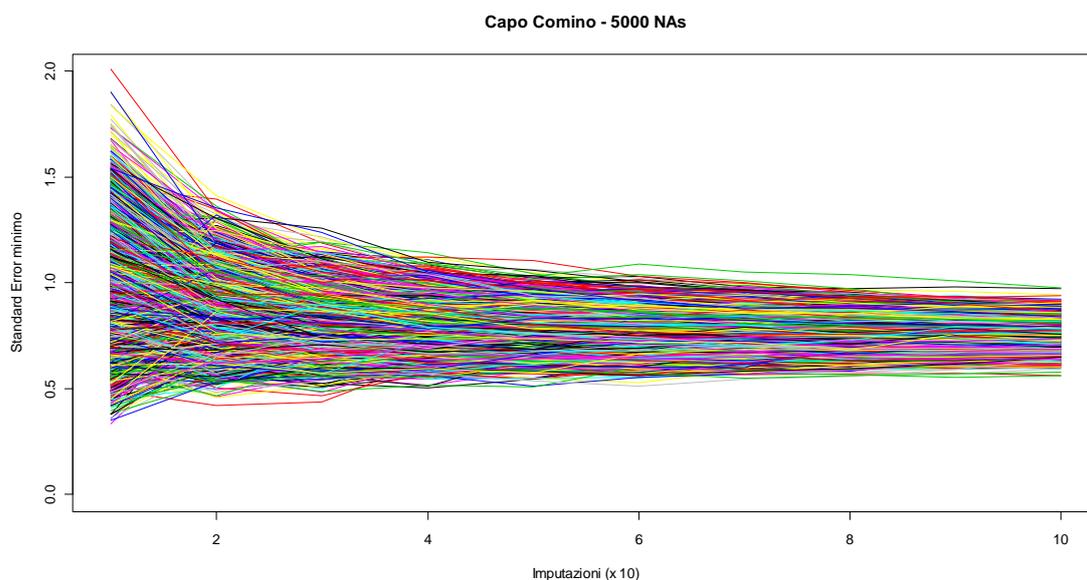


Fig. 3. 8 - Variazione dello standard error dei primi 5000 dati mancanti ricostruiti nel dataset della boa di Capo Comino in funzione del numero di imputazioni (da 10 a 100)

Per lo stesso motivo non sono state aggiunte nell'analisi altre covariate come ad esempio dati di temperatura dell'aria misurata dai mareografi o valori di altezza d'onda significativa (indice del grado di rimescolamento delle acque) ottenuti dalla rete ondometrica che darebbero sicuramente una maggiore robustezza al modello.

Nelle sezioni successive verranno illustrati i diagrammi generati per la ricostruzione dei dati mancanti per ogni boa utilizzando come covariate i dati di temperatura dell'acqua relativi ad ondometri e mareografi presenti nello stesso bacino della stazione in esame.

I grafici riportati evidenziano:

- l'andamento dei dati reali e ricostruiti (Fig. 3. 9a): in tutte le boe si può notare come la ricostruzione riproduca fedelmente le diverse componenti stagionali e risulta perfettamente sovrapponibile alla serie storica realmente osservata;
- l'andamento dello standard error espresso in termini di semiampiezza dell'intervallo di confidenza (Fig. 3. 9b): l'errore è strettamente connesso con il numero di variabili presenti. Esso risulta quindi più basso in corrispondenza delle osservazioni che presentano il minor numero di campioni non disponibili (tipicamente nel 2004). L'intervallo calcolato ha semiampiezza pari a $2 \cdot \text{standard error}$: il valore stimato ha quindi un livello di significatività del 95%;
- il boxplot degli errori (Fig. 3. 9c): descrive in modo compatto e grafico la distribuzione della funzione degli standard error tramite cinque punti rappresentativi. Su un piano cartesiano viene disegnato un rettangolo, i cui estremi sono il primo e terzo quartile, tagliato da una linea all'altezza della mediana. Vengono aggiunte due righe (detti anche baffi) corrispondenti ai valori distanti 1,5 volte la distanza interquartile a partire rispettivamente dal primo dal terzo quartile. Sono rappresentati nel grafico anche i valori che fuoriescono dall'intervallo delimitato dalle due righe come punti isolati [37]. In tutti i casi si verifica che l'errore assume valori decisamente bassi (a volte anche minore dell'unità), ma con punte (1 o 2 valori per ogni stazione) anche di 7 °C, legati alla presenza di dati anomali nell'intero dataset.

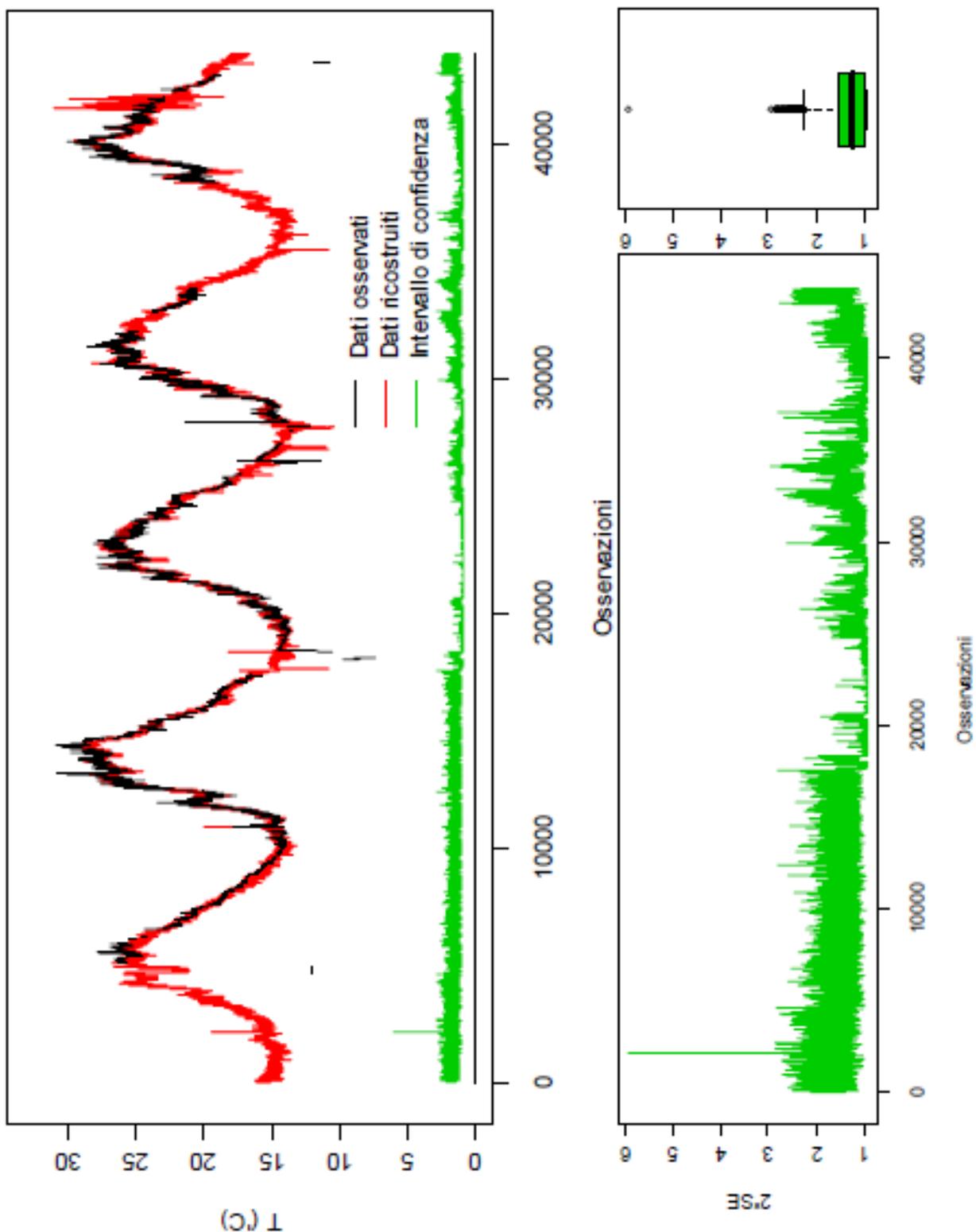


Fig. 3. 9 - Grafici tipo - (a) Andamento della serie osservata e della serie ricostruita; (b) andamento degli standard error; (c) boxplot degli standard error

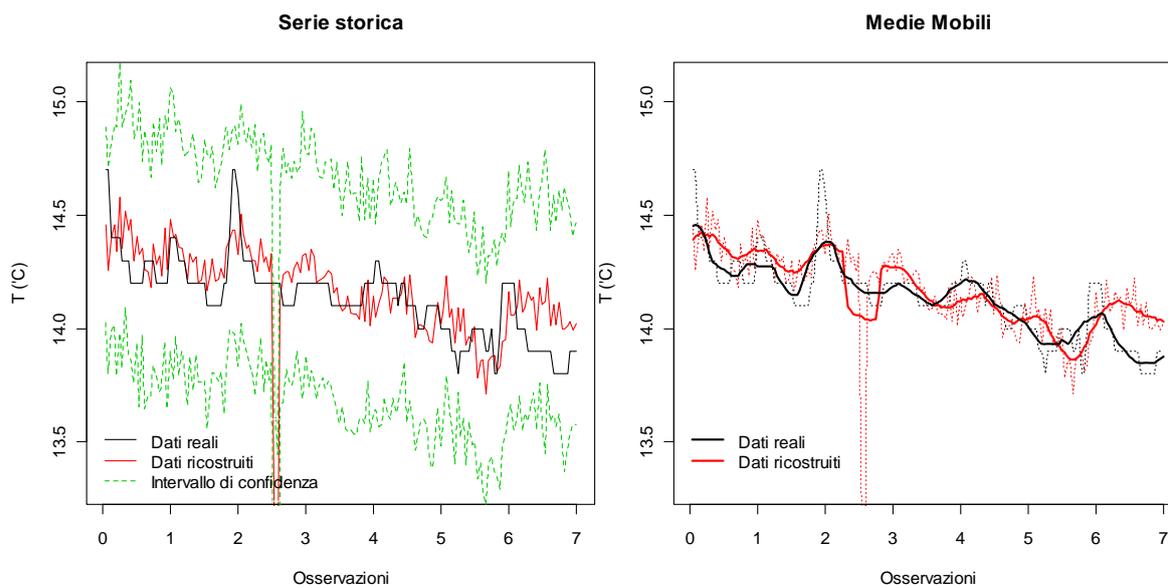


Fig. 3. 10 - Ingrandimento delle serie storiche reali e ricostruite in una settimana (Boa di Cetraro): a- Sovrapposizione dei dati reali e ricostruiti; b- Andamento delle medie mobili.

La Fig. 3. 10 consente di effettuare uno zoom sulle serie storiche osservate e ricostruite ed evidenzia sia l'ottima sovrapposizione dei dati che la precisione nella stima dell'errore: tutti i valori reali ricadono infatti all'interno dell'intervallo di confidenza.

L'ipotesi forte di indipendenza dei valori non consente però di ottenere un dataset di valori ricostruiti tale che la temperatura vari gradualmente nel tempo: la figura infatti evidenzia un andamento notevolmente altalenante della serie anche se con escursioni estremamente limitate. In particolare, la mancanza di un modello capace di correlare i dati di una singola osservazione ai valori precedenti e successivi, genera l'imputazione di pochi dati anomali (spike) derivanti dalla regressione lineari a pesi fissi effettuata nel modello.

Il diagramma della media mobile (centrata di ordine 12) dimostra come la serie dei valori imputati riesca a riprodurre le oscillazioni giornaliere tramite una sequenza di picchi e creste che ben approssimano le serie delle medie mobili ricavate dai dati effettivamente osservati dalla boa.

3.3.1. - Mar Tirreno

Il Tirreno è il sottobacino del Mediterraneo che contiene il maggior numero di stazioni di misura (ben 16 con 10 mareografi e 6 boe, Tab. 3. 4). Dalle figure 3.10-15 si può notare come la ricostruzione dei dati abbia dato dei buoni risultati soprattutto in stazioni in cui la numerosità del dato osservato è elevato (Boa di Cetraro Fig. 3. 14).

BOE		MAREOGRAFI	
Nome	Dati mancanti	Nome	Dati mancanti
Capo Comino	32383	Cagliari	554
Capo Gallo	23289	Civitavecchia	233
Capo Linaro	28102	Genova	1274
Cetraro	7091	Imperia	8092
La Spezia	21219	Livorno	51
Ponza	15753	Messina	483
		Napoli	544
		Palermo	78
		Palinuro	556
		Salerno	30

Tab. 3. 4 - Stazioni di misura nel Mar Tirreno

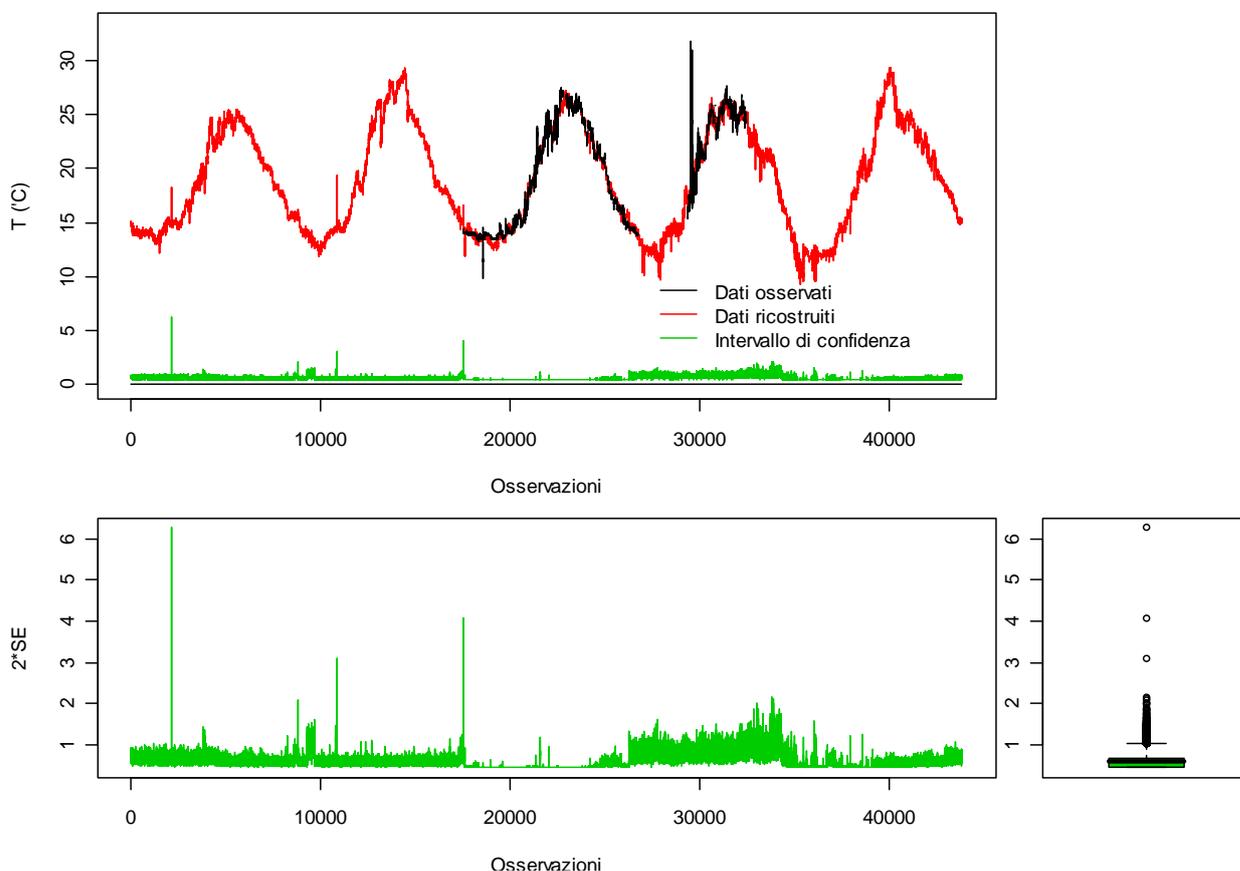


Fig. 3. 11 - Capo Comino - Ricostruzione dei Dati e Standard error

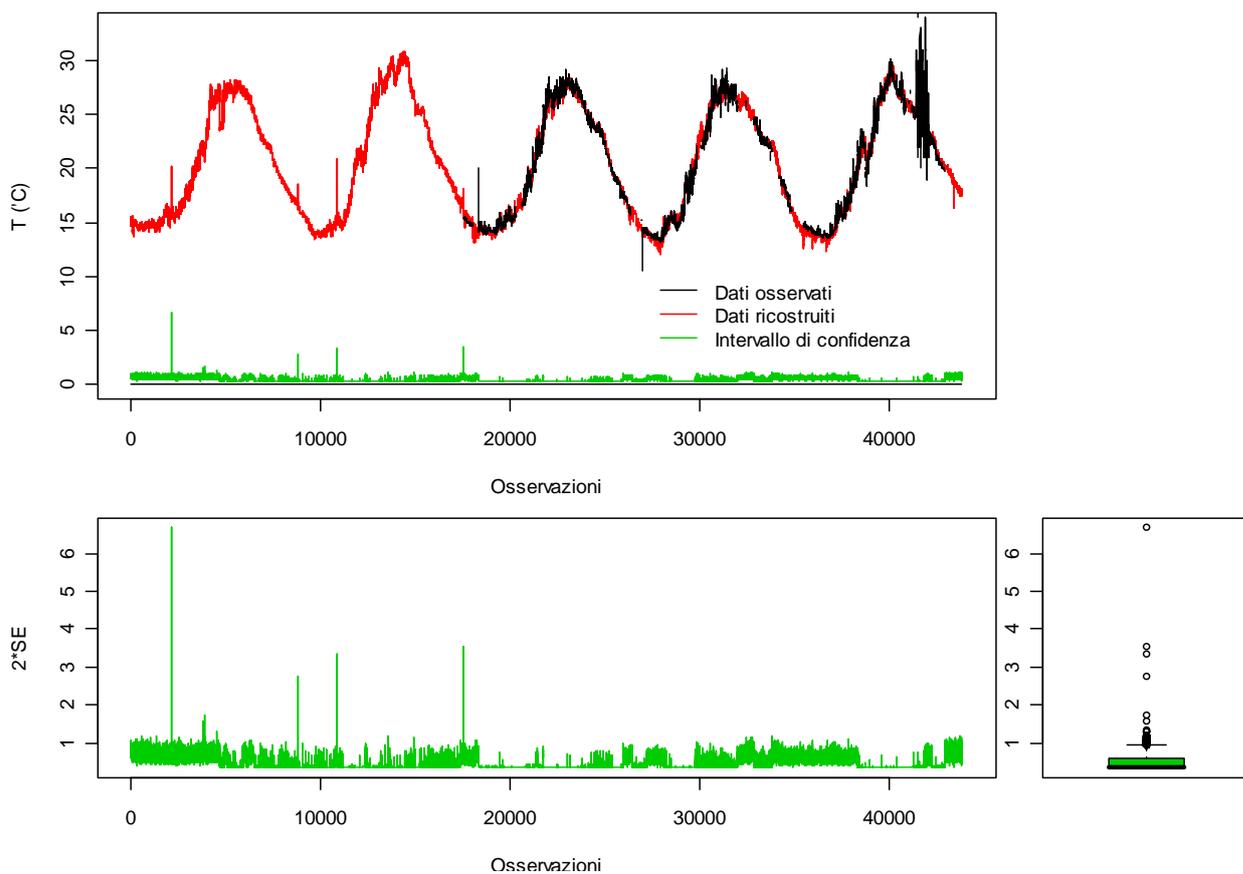


Fig. 3.12 - Capo Gallo - Ricostruzione dei Dati e Standard error

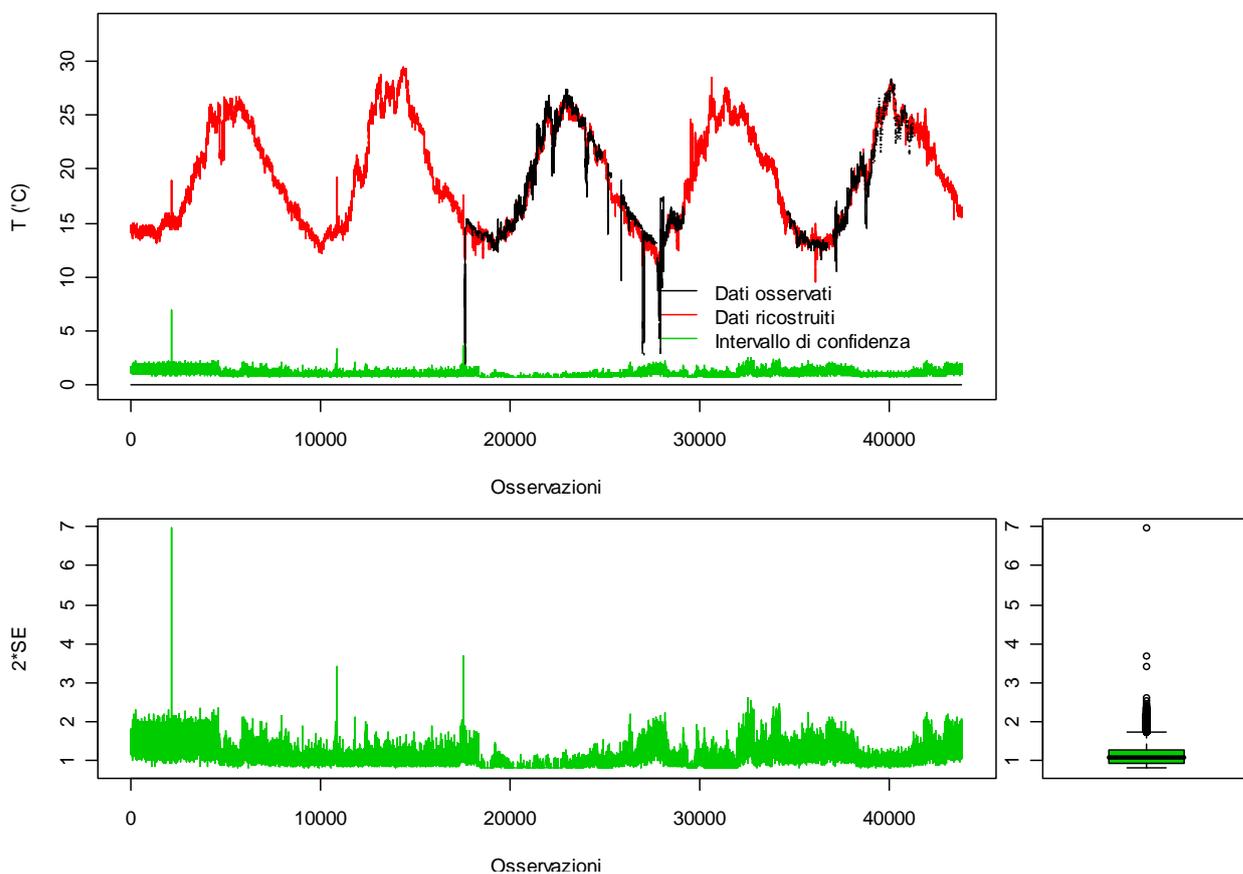


Fig. 3.13 - Capo Linaro - Ricostruzione dei Dati e Standard error

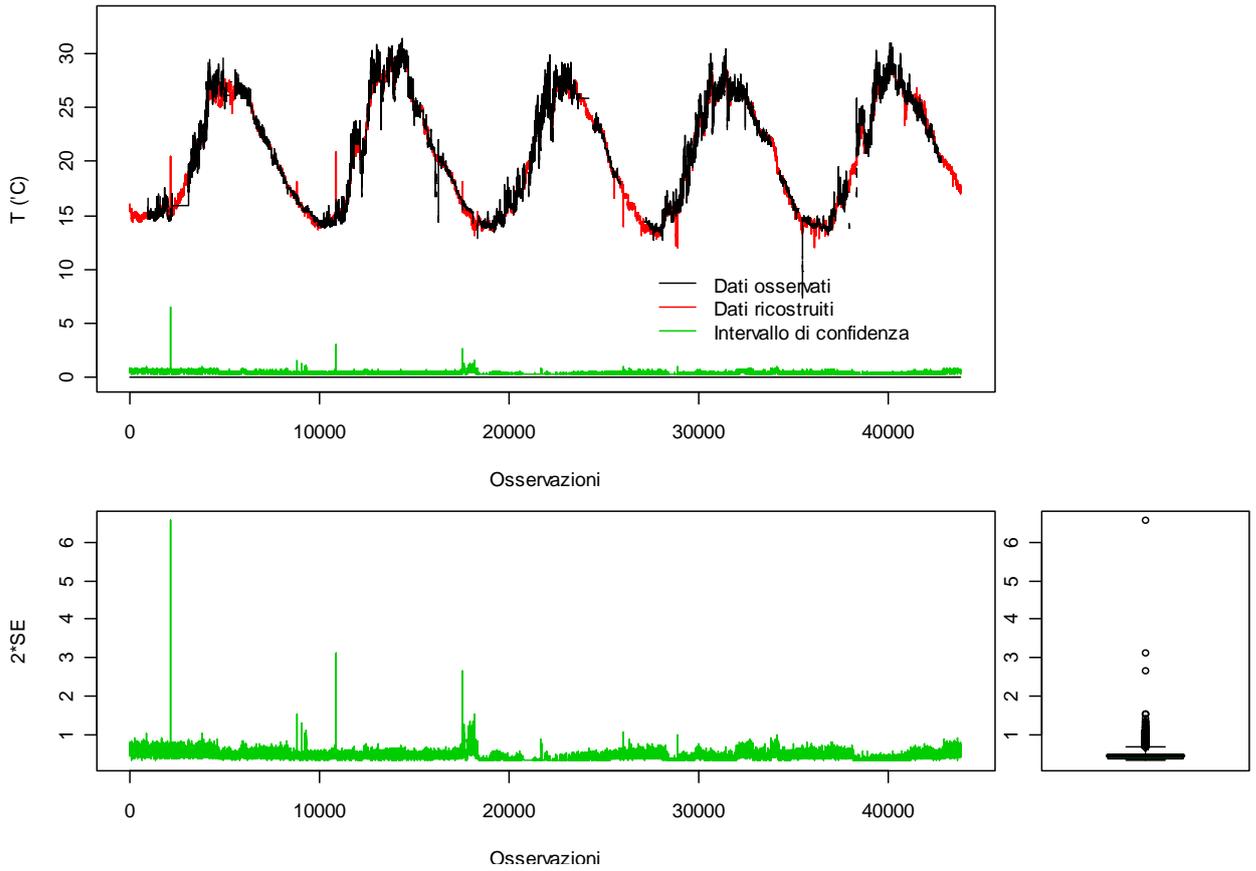


Fig. 3.14 - Cetraro - Ricostruzione dei Dati e Standard error

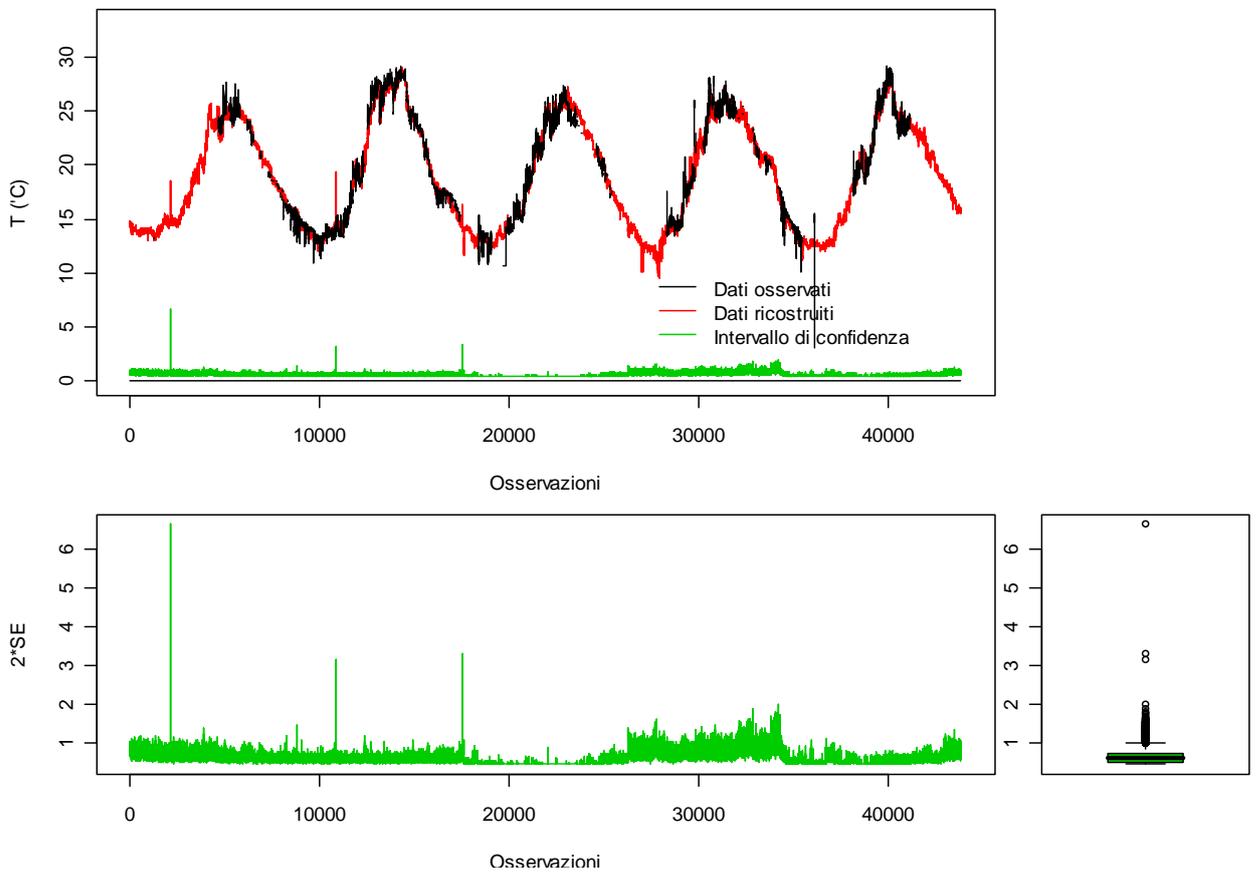


Fig. 3.15 - La Spezia - Ricostruzione dei Dati e Standard error

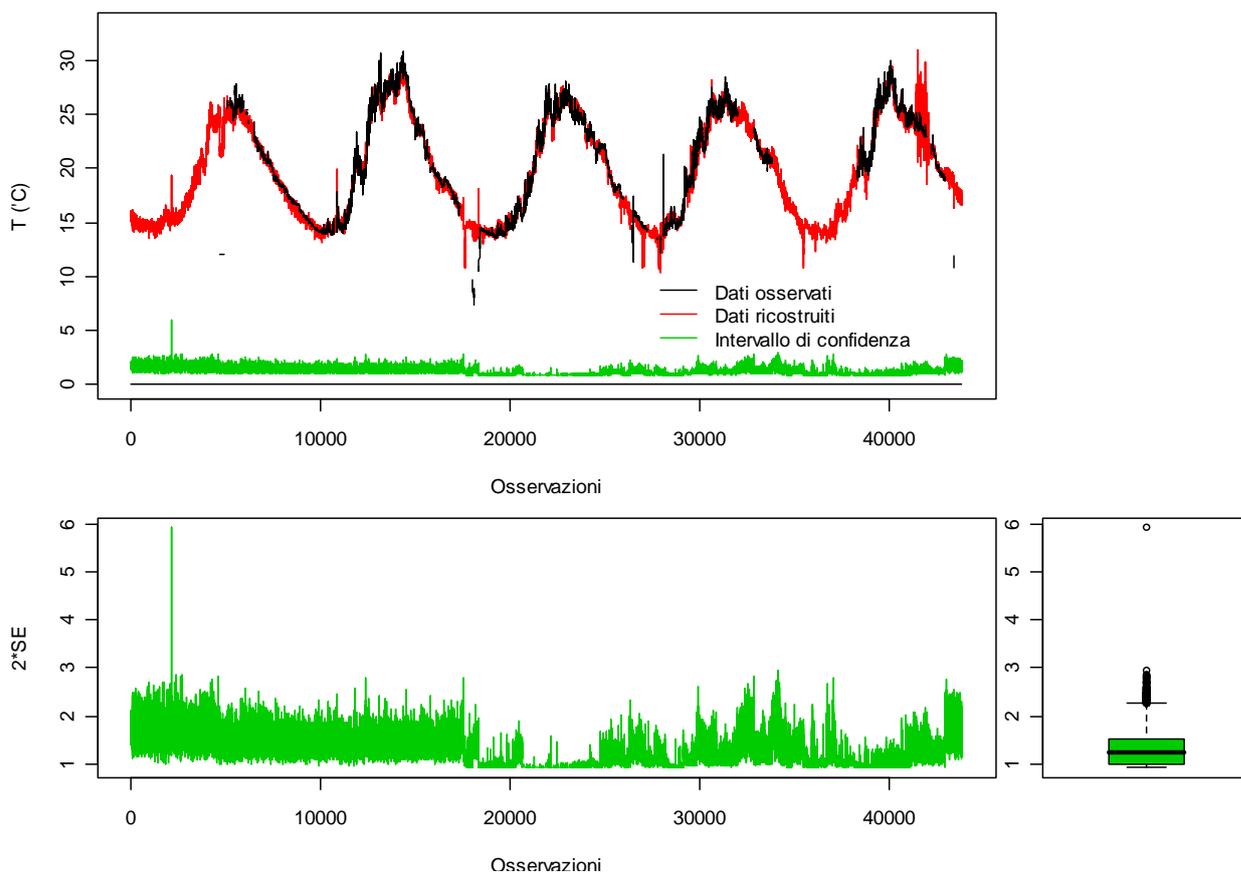


Fig. 3. 16 - Ponza - Ricostruzione dei Dati e Standard error

3.3.2. - *Mar Adriatico*

L'Adriatico presenta un buon numero di stazioni di misura uniformemente distribuite su tutto il bacino (12 stazioni di cui 8 mareografi e 4 boe, Tab. 3. 5). Anche in questo caso la ricostruzione dei dati ha dato buoni risultati con un errore medio superiore al caso del Tirreno a causa della minore numerosità delle stazioni presenti.

BOE		MAREOGRAFI	
Nome	Dati mancanti	Nome	Dati mancanti
Ancona	25360	Ancona	24
Monopoli	14850	Bari	122
Ortona	24664	Ortona	275
Punta della Maestra	39116	Otranto	934
		Ravenna	82
		Trieste	701
		Venezia	917
		Vieste	276

Tab. 3. 5 - Stazioni di misura nel Mar Adriatico

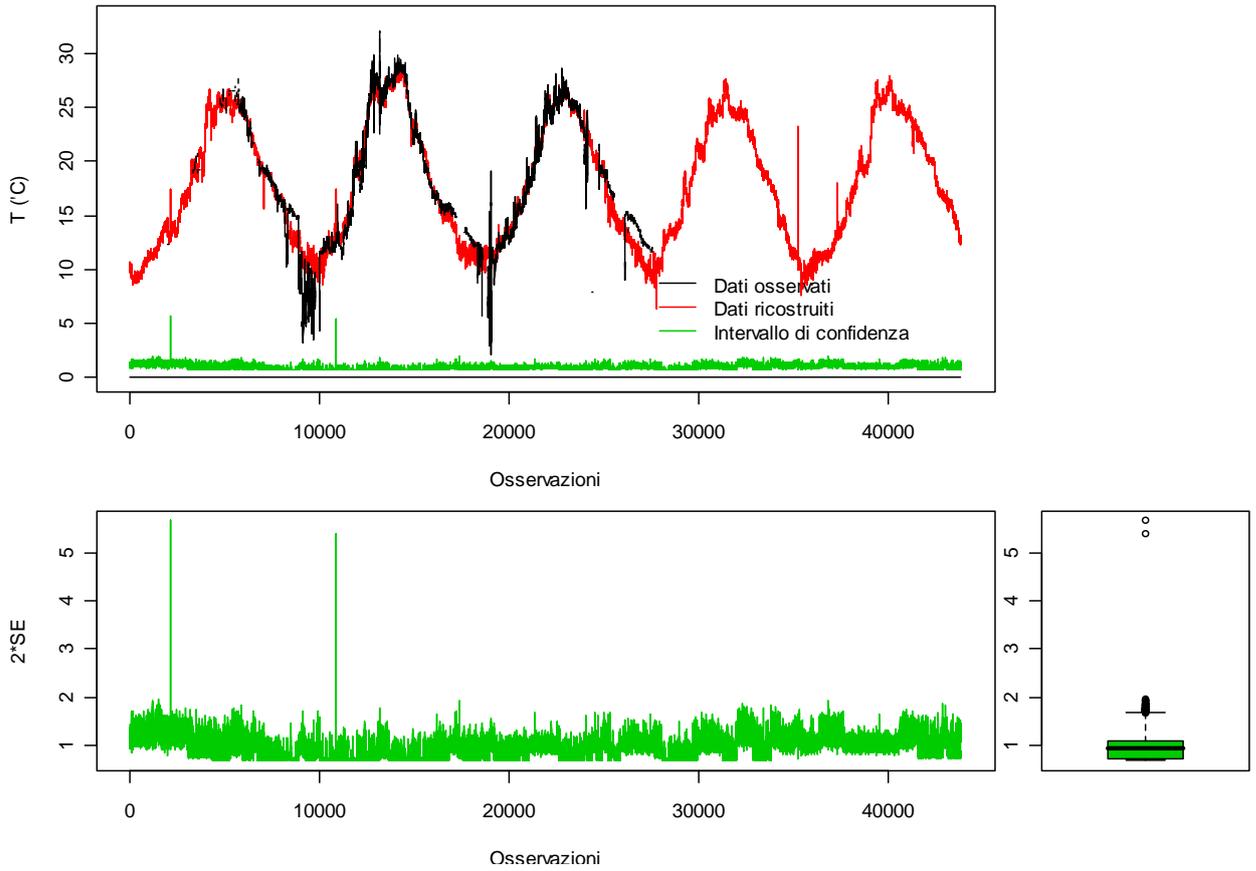


Fig. 3.17 - Ancona - Ricostruzione dei Dati e Standard error

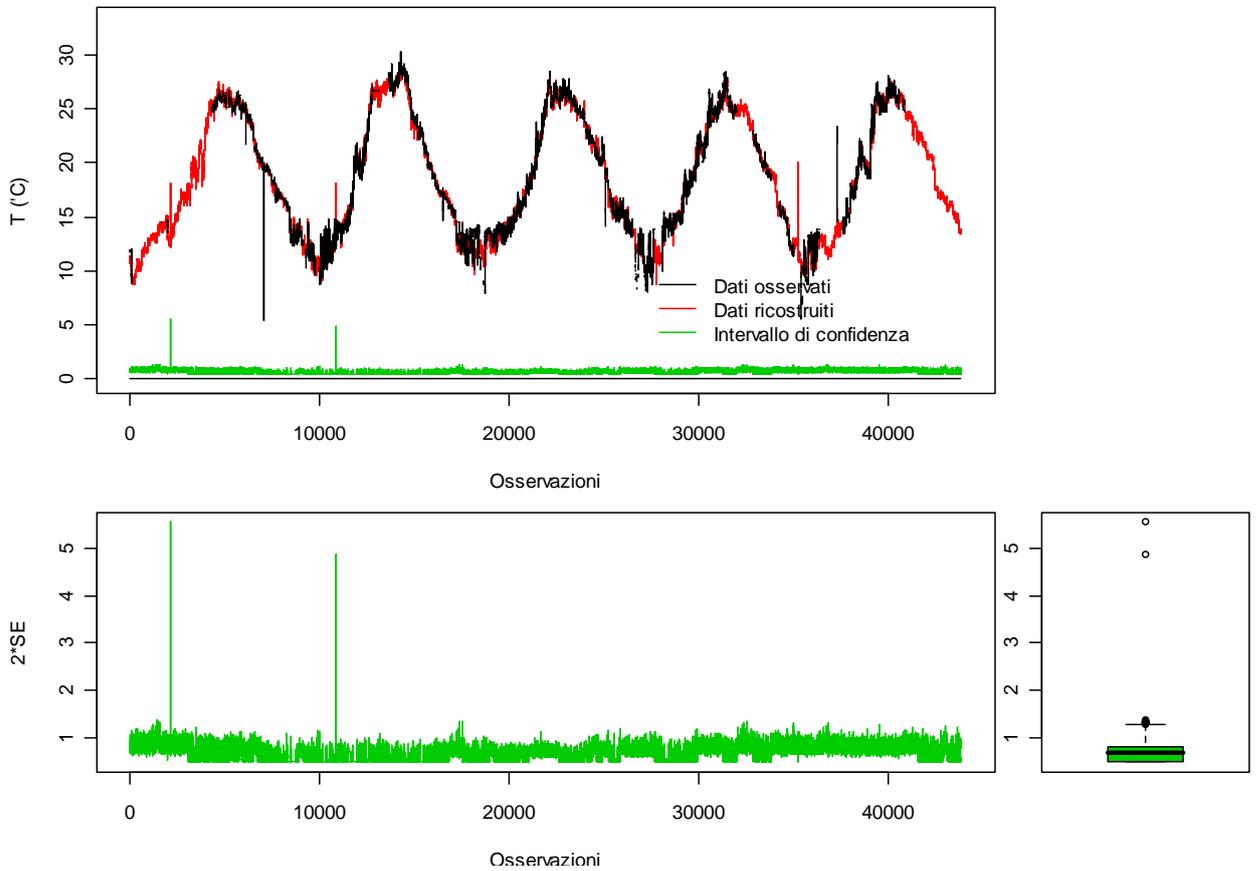


Fig. 3.18 - Monopoli - Ricostruzione dei Dati e Standard error

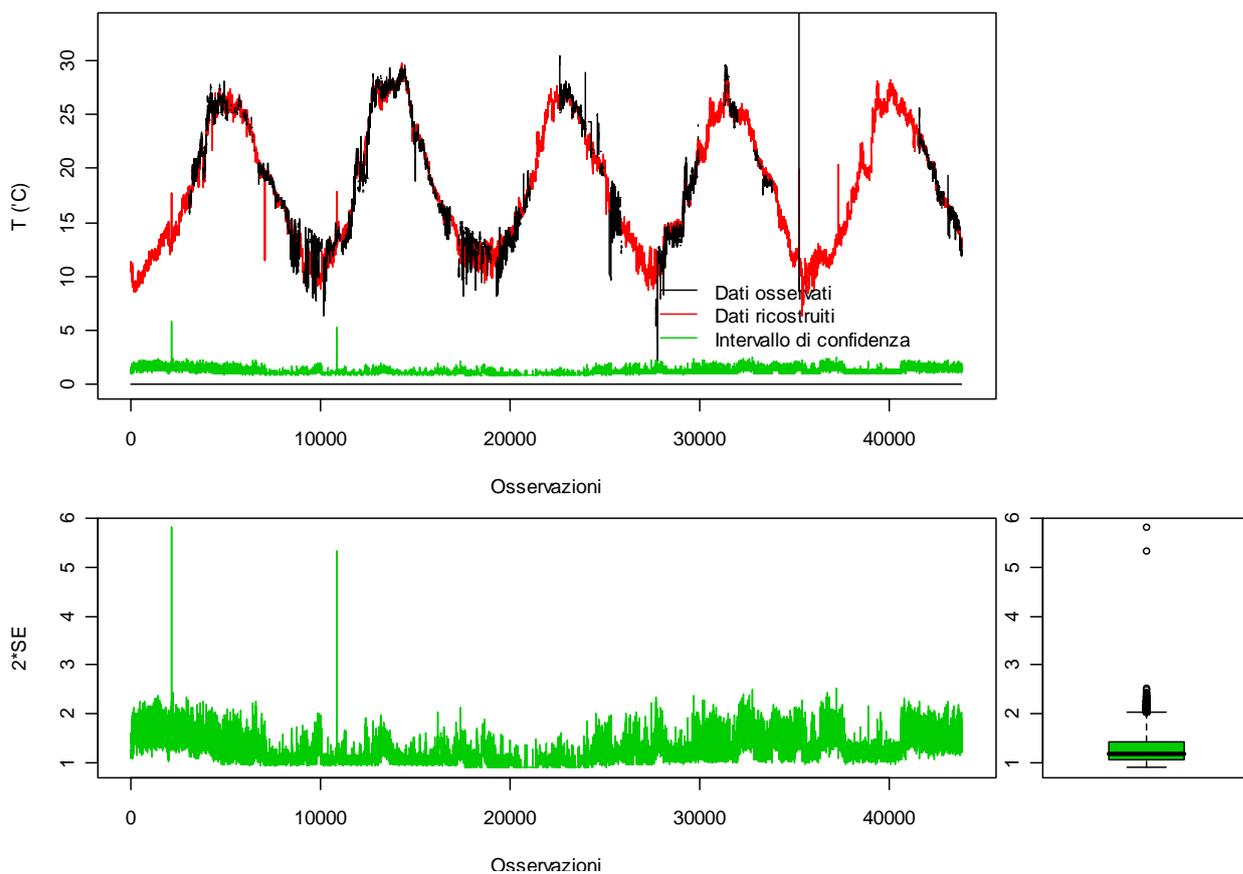


Fig. 3. 19 - Ortona - Ricostruzione dei Dati e Standard error

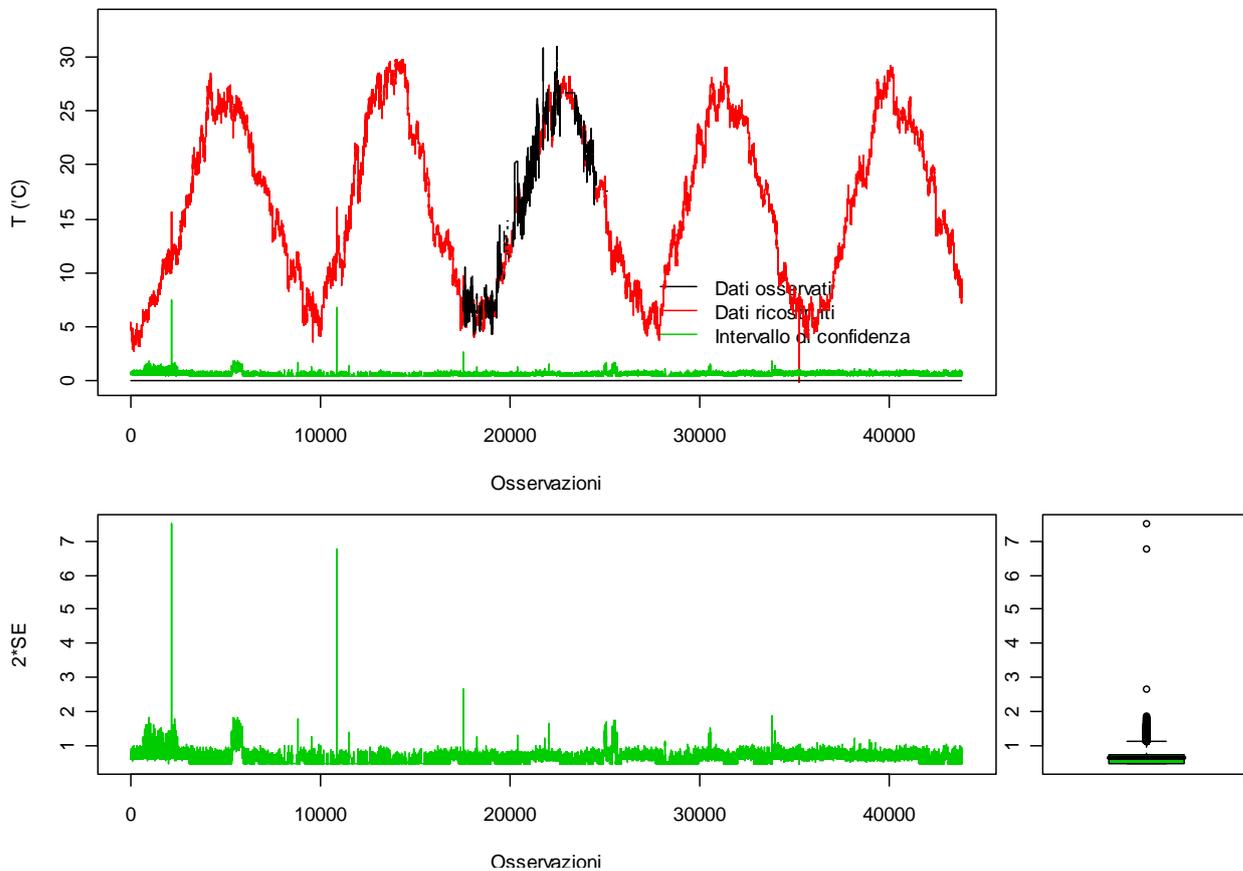


Fig. 3. 20 – Punta della Maestra - Ricostruzione dei Dati e Standard error

3.3.3. - *Mar Ionio*

Lo Ionio Tirreno è il sottobacino del Mediterraneo che contiene il minor numero di stazioni di misura (8 stazioni con 6 mareografi e 2 boe, Tab. 3. 6). Il dato ricostruito, seppur di “buona qualità”, risulta notevolmente influenzato dalla presenza di spike nelle serie dei dati osservati. Nella serie storica di Crotone sono stati eliminati tutti i dati dalla osservazione 27000 in poi poiché affetti da un evidente errore di misura.

BOE		MAREOGRAFI	
Nome	Dati mancanti	Nome	Dati mancanti
Catania	14570	Catania	377
Crotone	21059	Crotone	747
		Lampedusa	151
		Otranto	934
		Reggio Calabria	35
		Taranto	59

Tab. 3. 6 - Stazioni di misura nel Mar Ionio

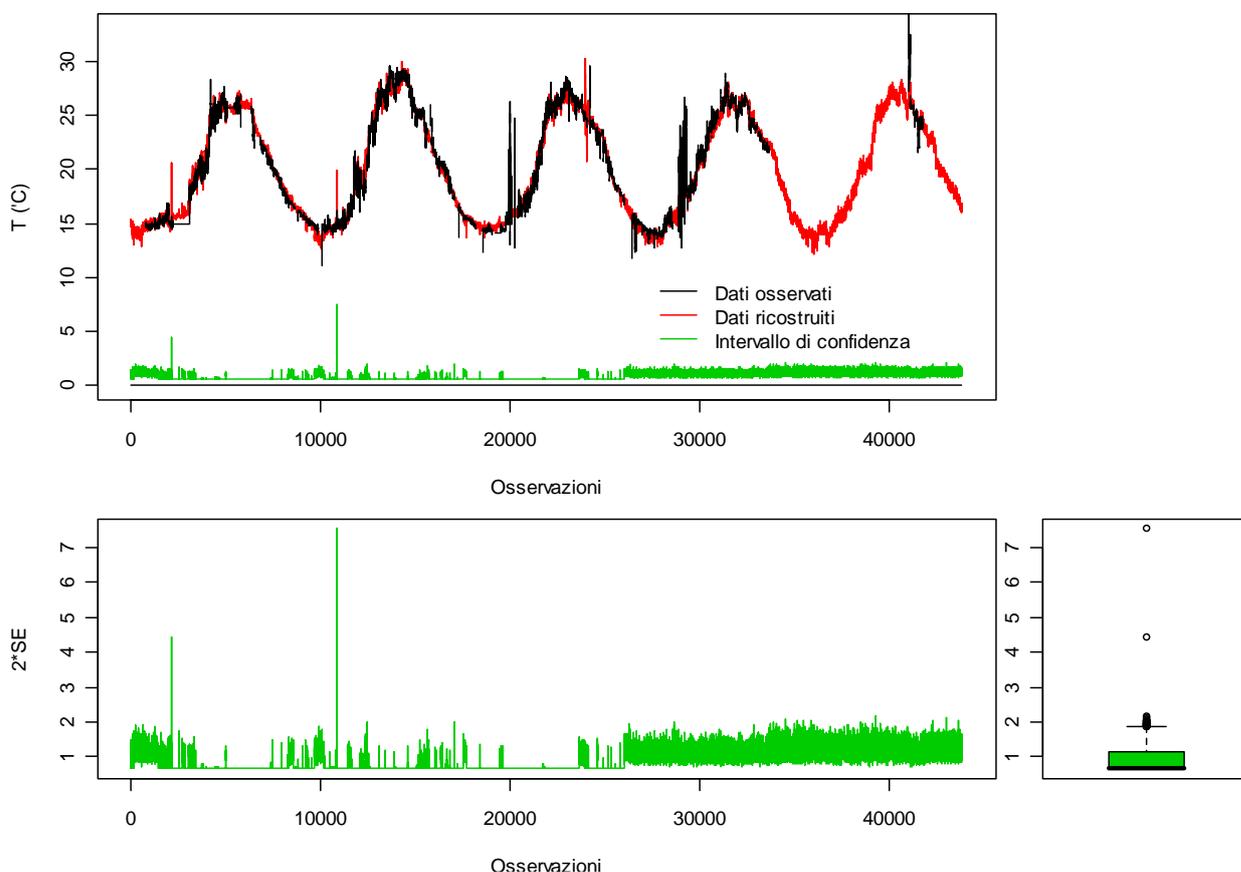


Fig. 3. 21 - Catania - Ricostruzione dei Dati e Standard error

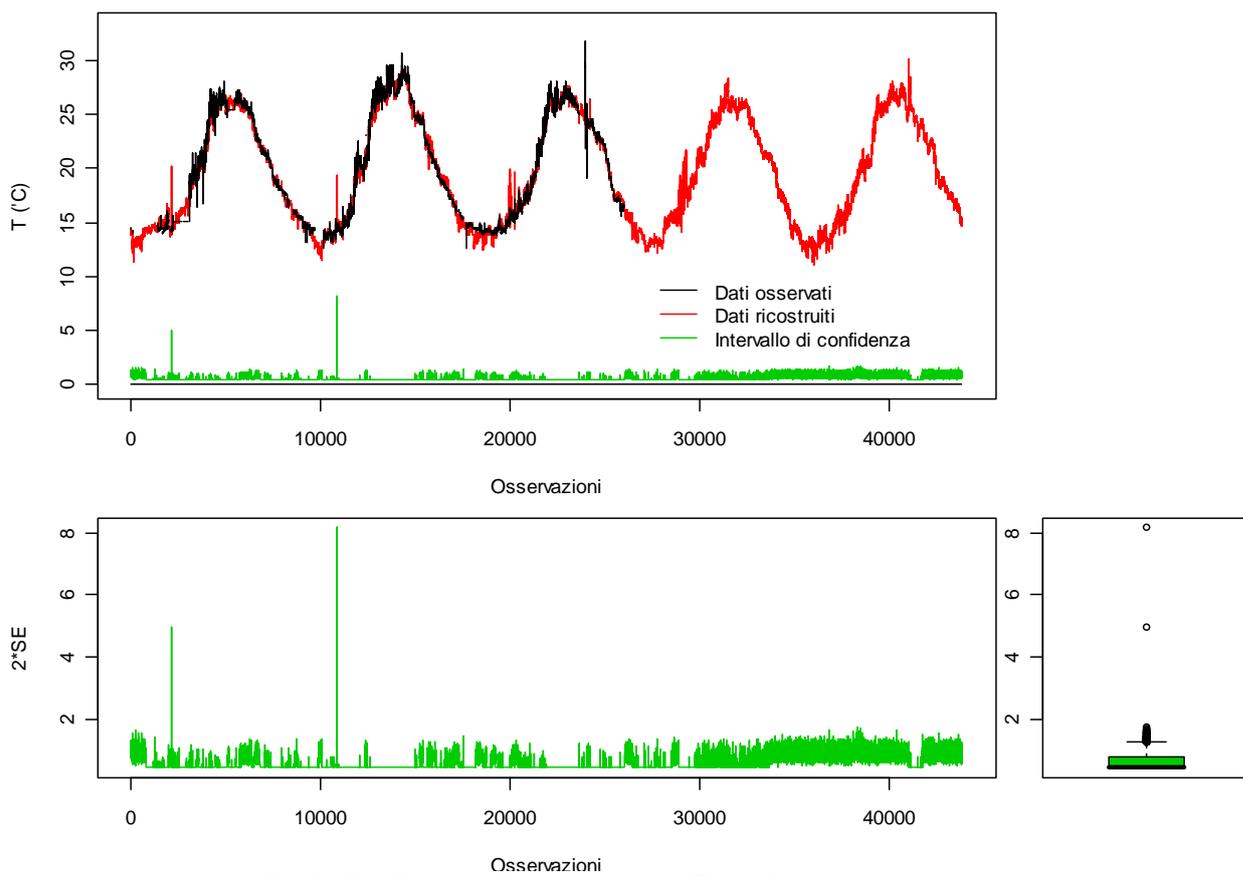


Fig. 3. 22 - Crotone - Ricostruzione dei Dati e Standard error

Ricostruzione dei missing values dell'altezza d'onda significativa

I dati di altezza d'onda significativa (H_s) vengono elaborati a partire dai dati di spostamento (in elevazione, in direzione Nord e in direzione Est) ottenuti dalla doppia integrazione delle misure dei sensori accelerometrici presenti sulle boe.

A differenza dei dati di temperatura dell'acqua, l'altezza d'onda non presenta delle marcate componenti stagionali. Il fenomeno fisico alla base della definizione del moto ondoso è il vento, che può essere estremamente variabile durante l'anno sia in termini di intensità che di direzione. Inoltre, lo stato di mare è influenzato da una serie di fattori difficilmente modellabili, quali la profondità del mare, il fetch efficace, le condizioni di pressione atmosferica, l'influenza delle coste, etc. Per tutti questi motivi non sempre esiste una coincidenza tra le direzioni delle onde e dei venti e le relative intensità. Risulta quindi particolarmente complicato applicare il metodo di imputazione multipla senza inserire un modello fisicamente basato che sia in grado di sintetizzare tutte le informazioni relative alle variabili da cui dipende lo stato di mare. In alternativa, si potrebbe utilizzare il metodo dell'imputazione multipla considerando anche la componente direzionale e includendo nell'elaborazione le misurazioni di intensità e direzione del vento registrate dai mareografi. Anche in questo caso, però, occorre valutare attentamente le correlazioni, effettuare delle opportune trasposizioni dei dati di vento dai mareografi alle boe e valutare in maniera corretta le condizioni di esposizione delle stazioni mareografiche.

In conclusione, la ricostruzione dei dati di altezza d'onda significativa appare alquanto complessa.

Il database ISPRA contiene un elevatissimo numero di file HNE, contenenti informazioni circa gli spostamenti della boa in elevazione, direzione Nord e direzione Est (fig. 4.1).

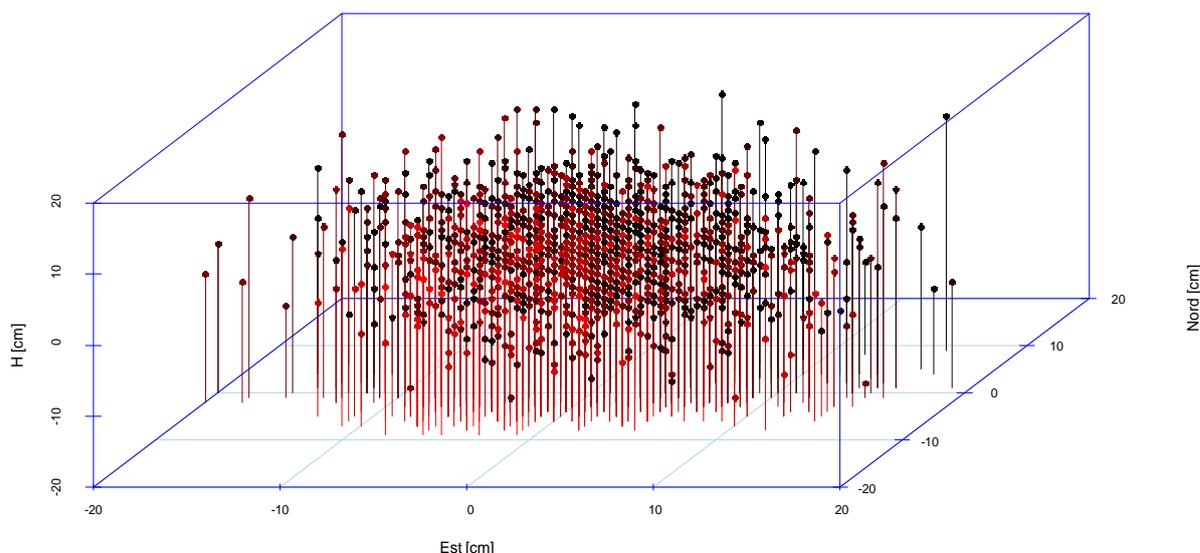


Fig. 4. 1 - Esempio di dati contenuti in un file HNE – il piano orizzontale individua la posizione della boa in coordinate Nord-Est; l'asse verticale è relativo alle elevazioni (altezze d'onda).

Non tutti questi dati sono stati elaborati dal modulo di post-processing del sistema: una parte è stata infatti scartata perché incompleta, altri ancora sono stati eliminati poiché presentano pochi dati anomali. E' quindi possibile esaminare questi files e provvedere, tramite controlli sulla qualità del dato e/o tecniche di riempimento dei dataset, alla rielaborazione dei valori di altezza d'onda significativa tramite l'analisi spettrale.

In questo capitolo verrà effettuata un'analisi degli HNE individuando i files che, soggetti ad una ulteriore fase di controllo e validazione, potranno essere successivamente rielaborati al fine di fornire valori attendibili di altezza d'onda significativa. Il numero di dati recuperabili con questa procedura risulta comunque limitato poiché presuppone che la boa sia rimasta in funzione nel periodo considerato (anni 2002-2005) anche quando il valore di H_s non è stato elaborato. Per i periodi di malfunzionamento della stazione (disormeggi, boa a terra, gaps nella trasmissione) occorrerà invece ricostruire dei modelli statistici piuttosto evoluti.

4.1 - Dati HNE (anni 2002-2006)

L'analisi dei file HNE nel periodo che va dal 1 gennaio 2002 al 31 dicembre 2006 è stata effettuata tramite l'ausilio di un software appositamente elaborato in codice FORTRAN. Il codice, in base ai dati in input (codice della stazione ondometrica, numero di osservazioni giornaliere, etc.) valuta la presenza (o l'assenza) degli HNE, li classifica in base alla qualità e alla numerosità dei dati e crea in output una serie di files sintetici rappresentativi dell'analisi (tabelle e grafici). La qualità viene valutata in base alla dimensione del file che

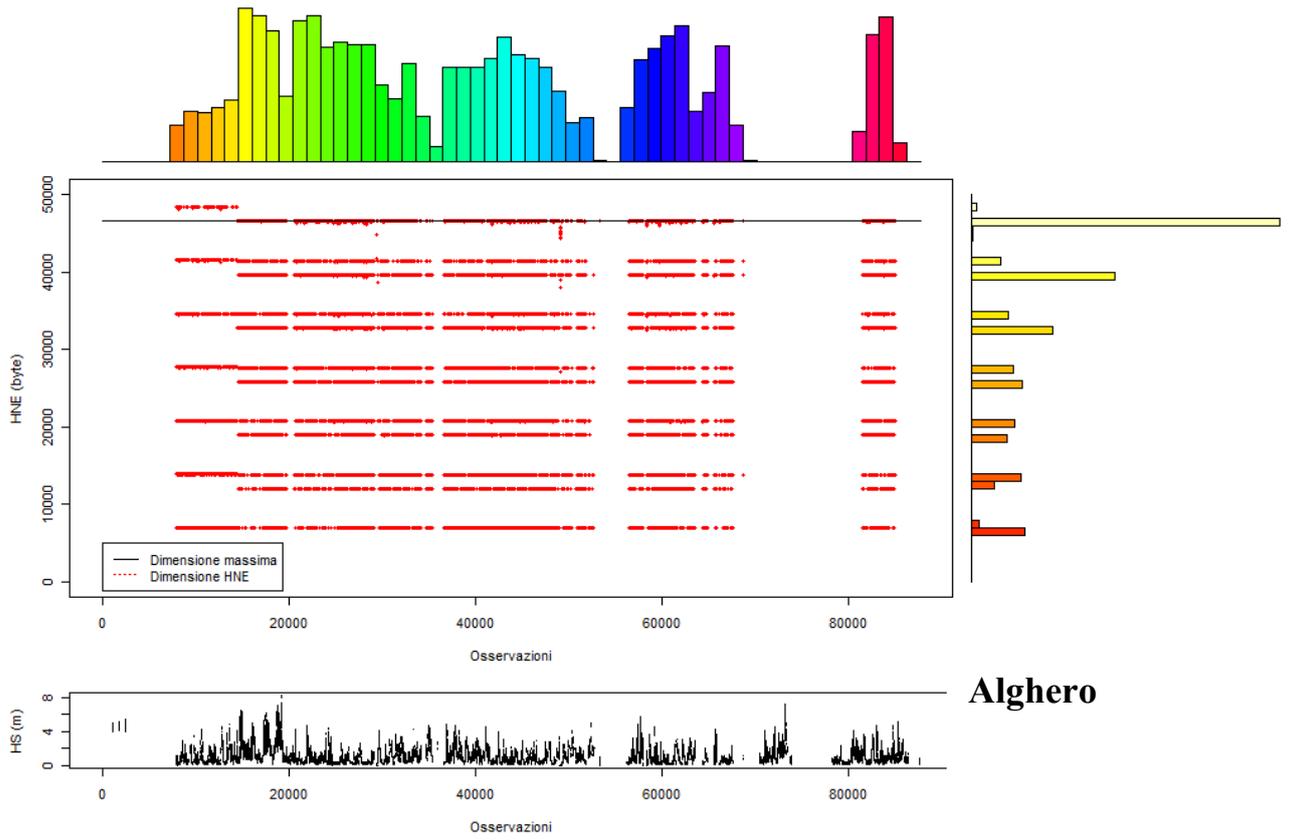
è strettamente legata al numero di righe (e quindi di osservazioni) presenti per ogni singolo campionamento. La numerosità è data dal rapporto tra il numero di file HNE presenti e il numero complessivo di osservazioni semiorarie nel periodo di riferimento (87648 misurazioni).

Numero HNE							
Alghero	Ancona	Capo Comino	Catania	Certaro	Civitavecchia	Crotone	Totale
35162	29838	35043	47516	60239	44190	58738	604633
Mazara	Monopoli	Ortona	Palermo	Ponza	Punta della Maestra	La Spezia	
47991	47572	36315	49001	47931	24390	40707	

Tab. 4. 1 - Numero di file HNE presenti nel database

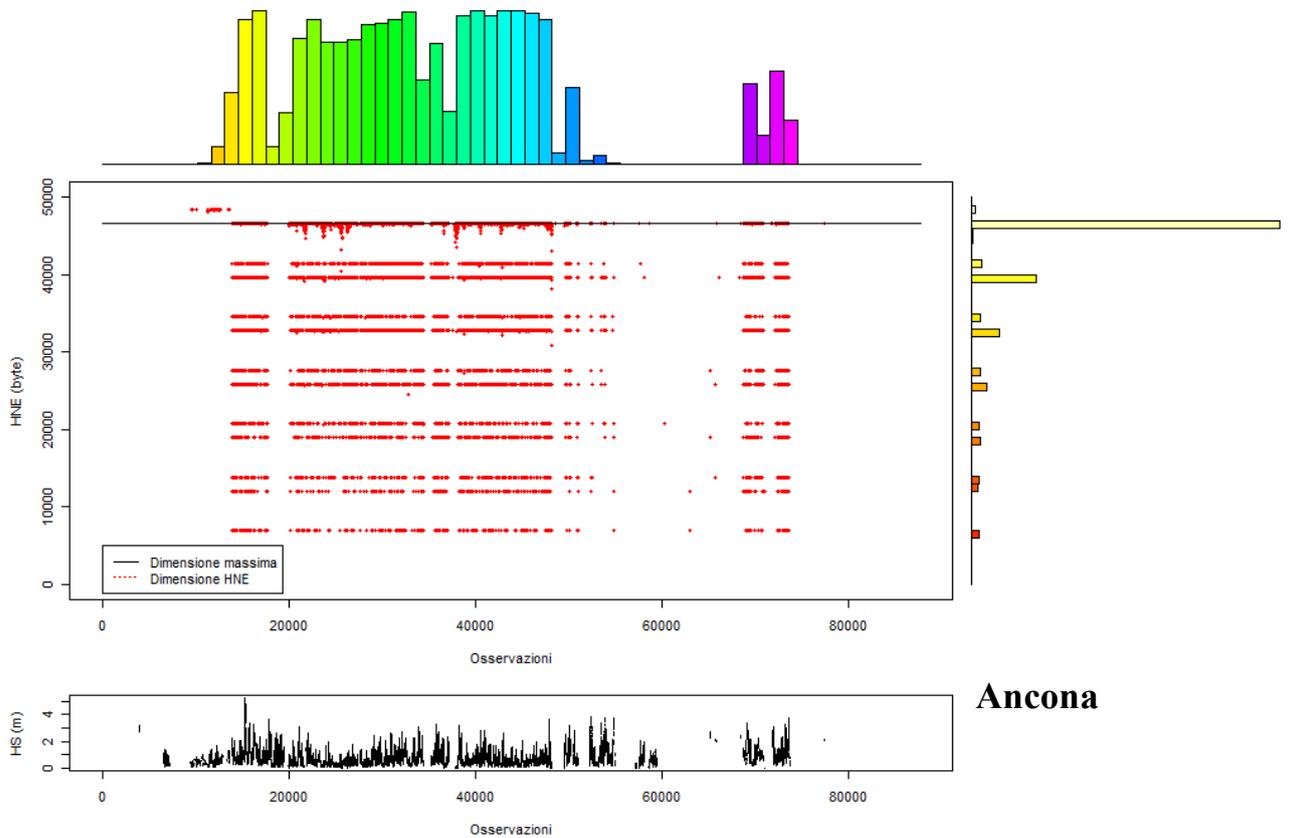
I diagrammi (figg. 4.2-15) analizzano le caratteristiche dei file HNE relativi ad ogni boa descrivono:

- La presenza nel database dei file HNE (diagramma centrale puntinato): al variare delle osservazioni, vengono individuati i files presenti in funzione della loro dimensione. La dimensione dei file non varia con continuità: le informazioni contenute sono registrate in pacchetti da 256 righe. In particolare, un file completo presenta 6 pacchetti da 256 righe ed un pacchetto finale incompleto. Inoltre i files relativi ai primi mesi del 2002 sono caratterizzati da un formato diverso con un numero maggiore di righe e quindi da un diverso valore della dimensione massima;
- La numerosità mensile del dato (istogramma orizzontale): consente di valutare la quantità dei file HNE per ogni mese nel periodo 2002-2006;
- La qualità del dato (istogramma verticale): consente di enumerare il numero di file HNE completi e incompleti. Risulta evidente come la maggior parte dei files presenti siano completi.
- L'andamento delle altezze d'onda significative calcolate nel periodo di riferimento (diagramma inferiore): è possibile individuare periodi lunghi di funzionamento intervallati da periodi di interruzione del servizio o periodi (brevi) in cui il file HNE è presente, ma non elaborato;



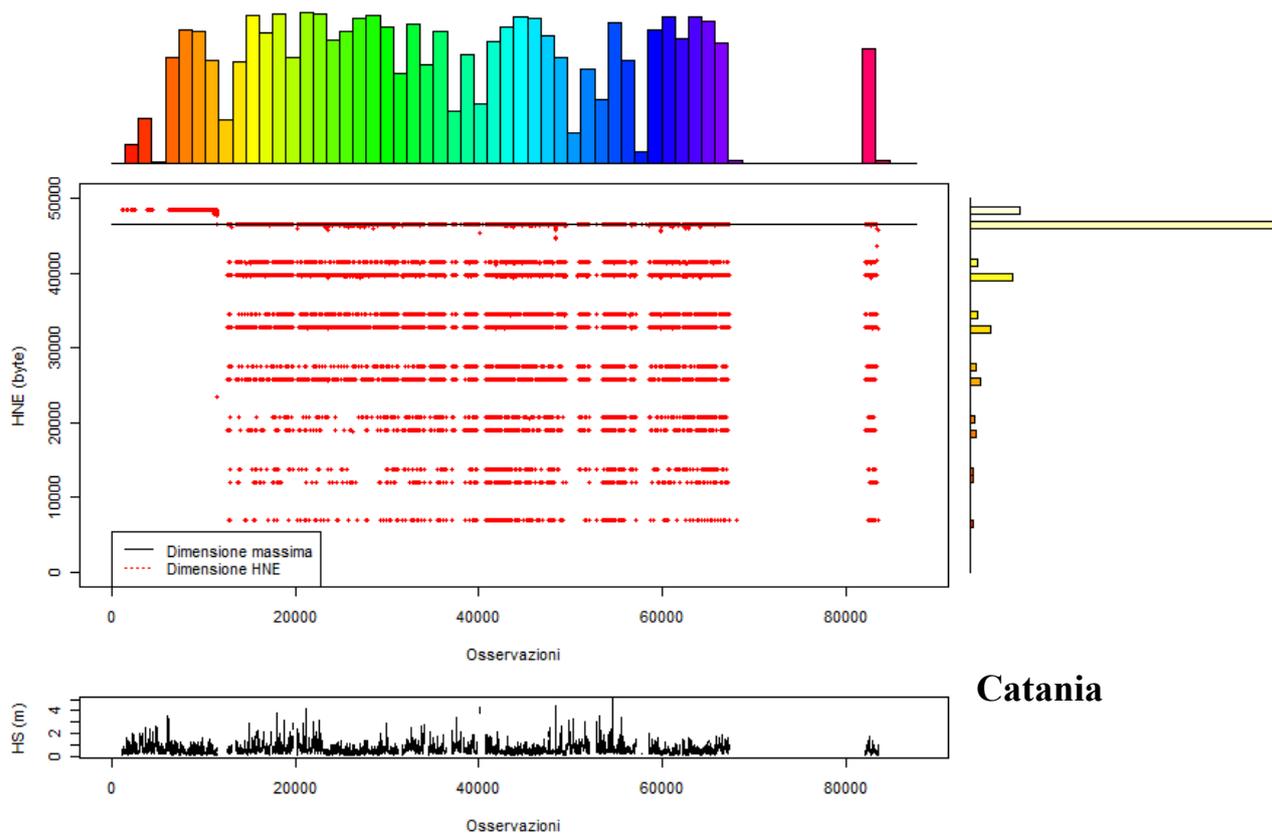
Alghero

Fig. 4. 2 - Diagrammi HNE - Alghero



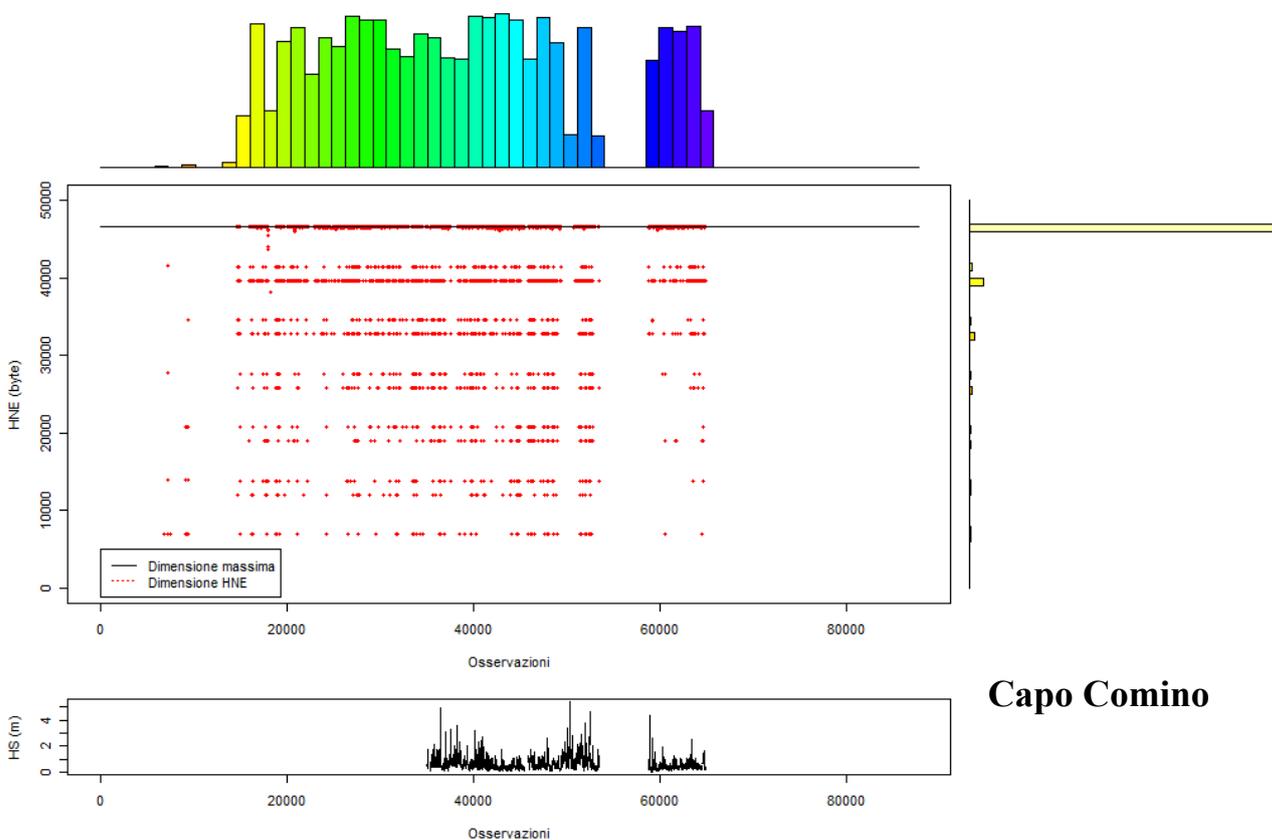
Ancona

Fig. 4. 3 - Diagrammi HNE - Ancona



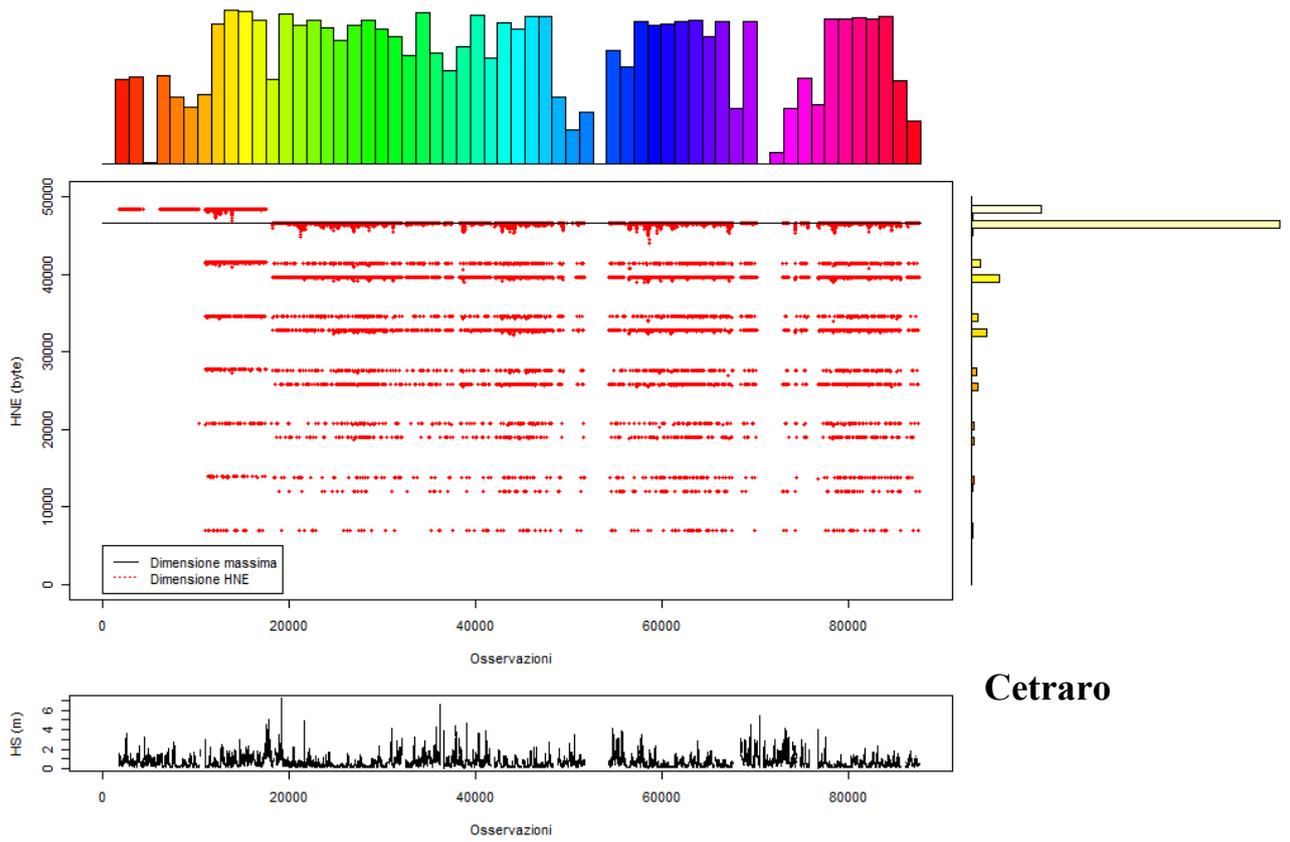
Catania

Fig. 4. 4 - Diagrammi HNE - Catania



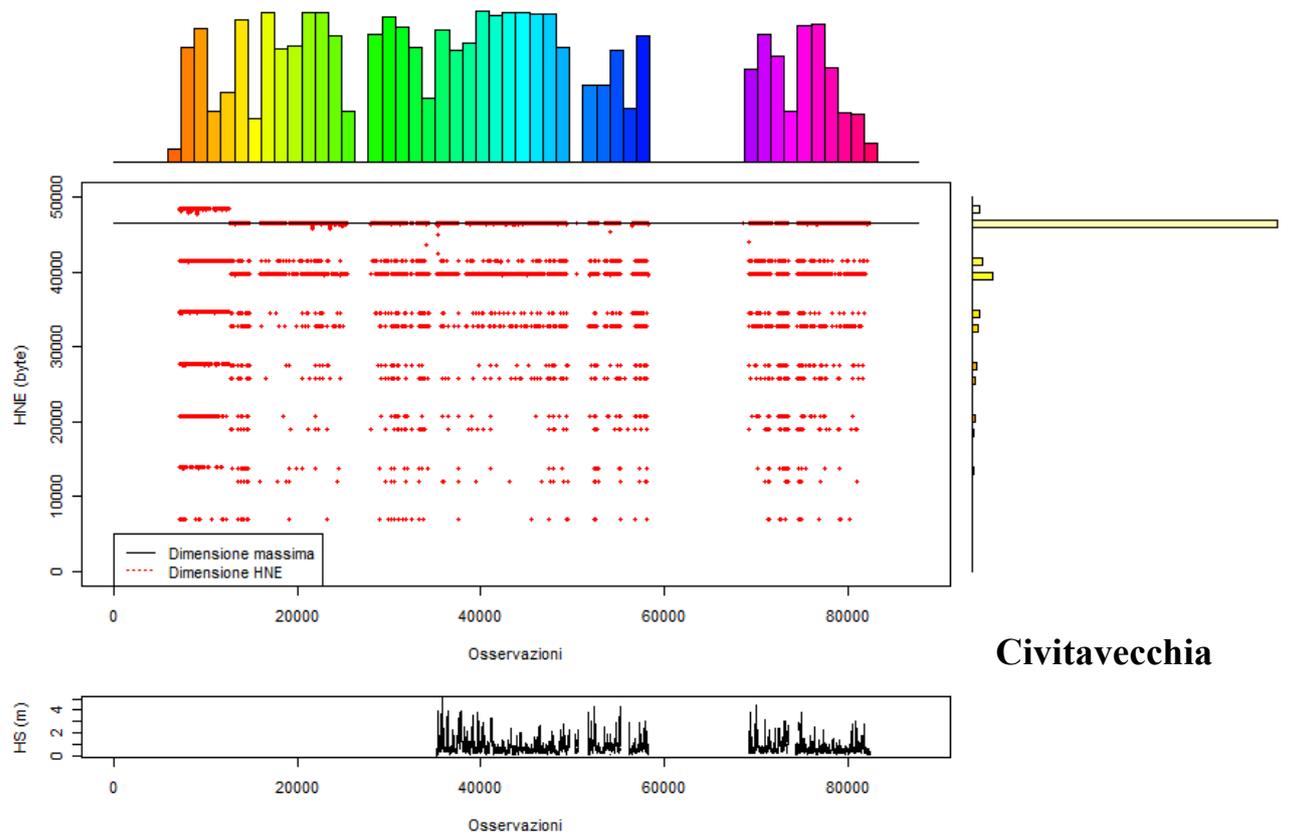
Capo Comino

Fig. 4. 5 - Diagrammi HNE – Capo Comino



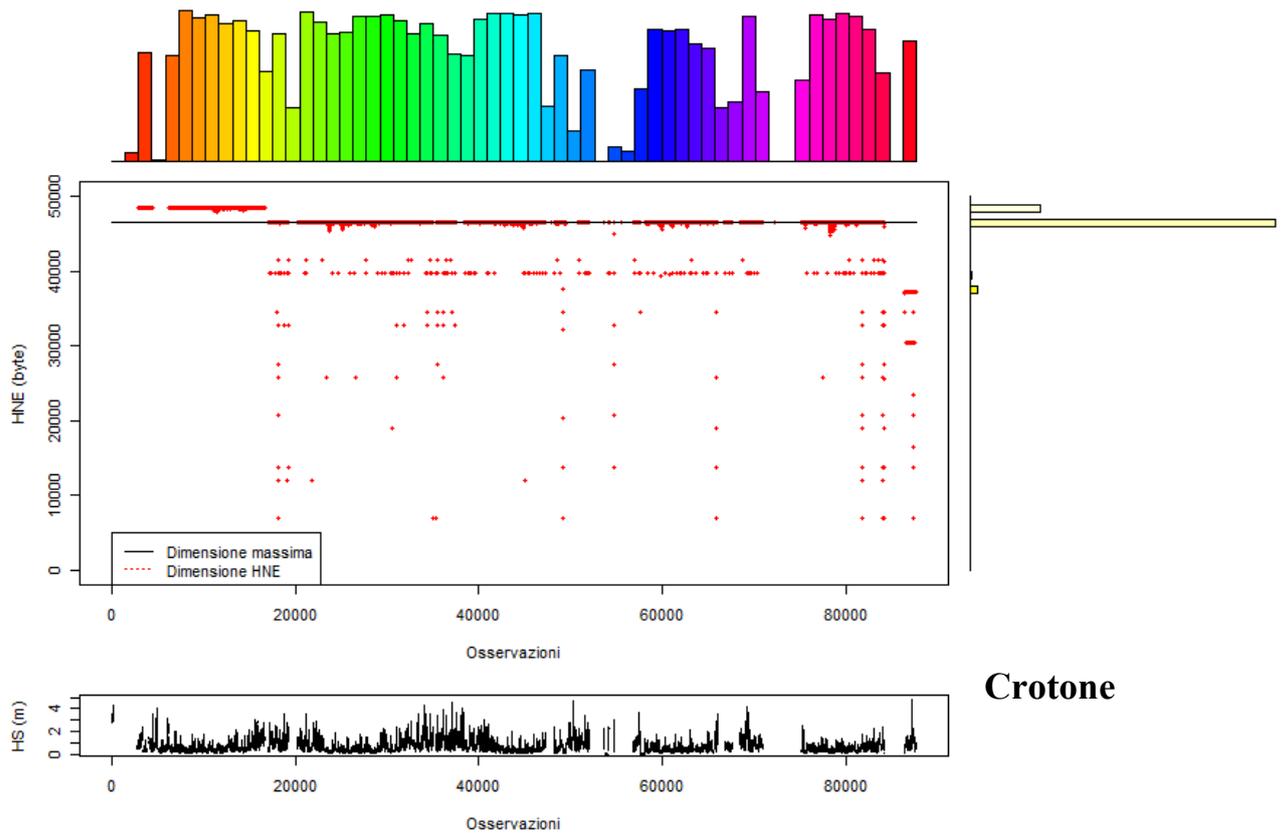
Cetraro

Fig. 4. 6 - Diagrammi HNE - Cetraro



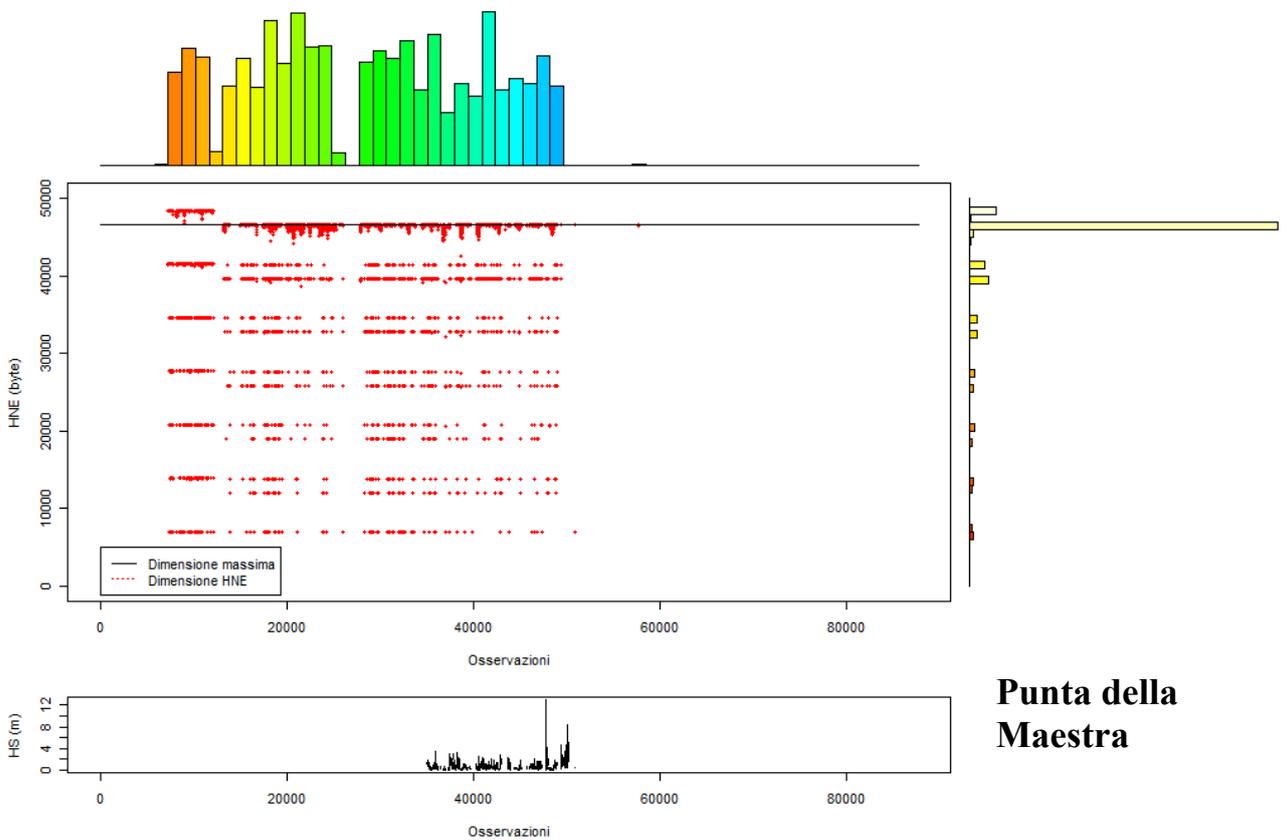
Civitavecchia

Fig. 4. 7 - Diagrammi HNE - Civitavecchia



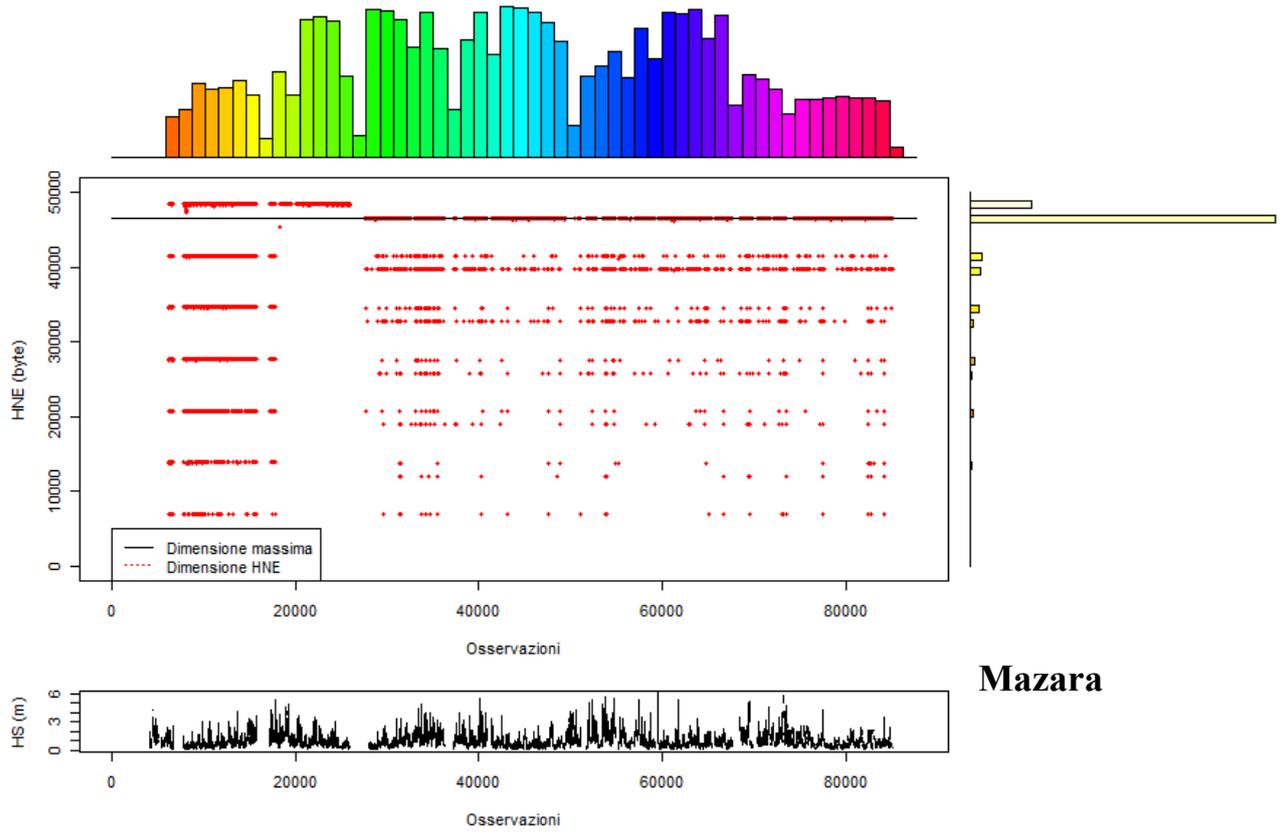
Crotona

Fig. 4. 8 - Diagrammi HNE - Crotona



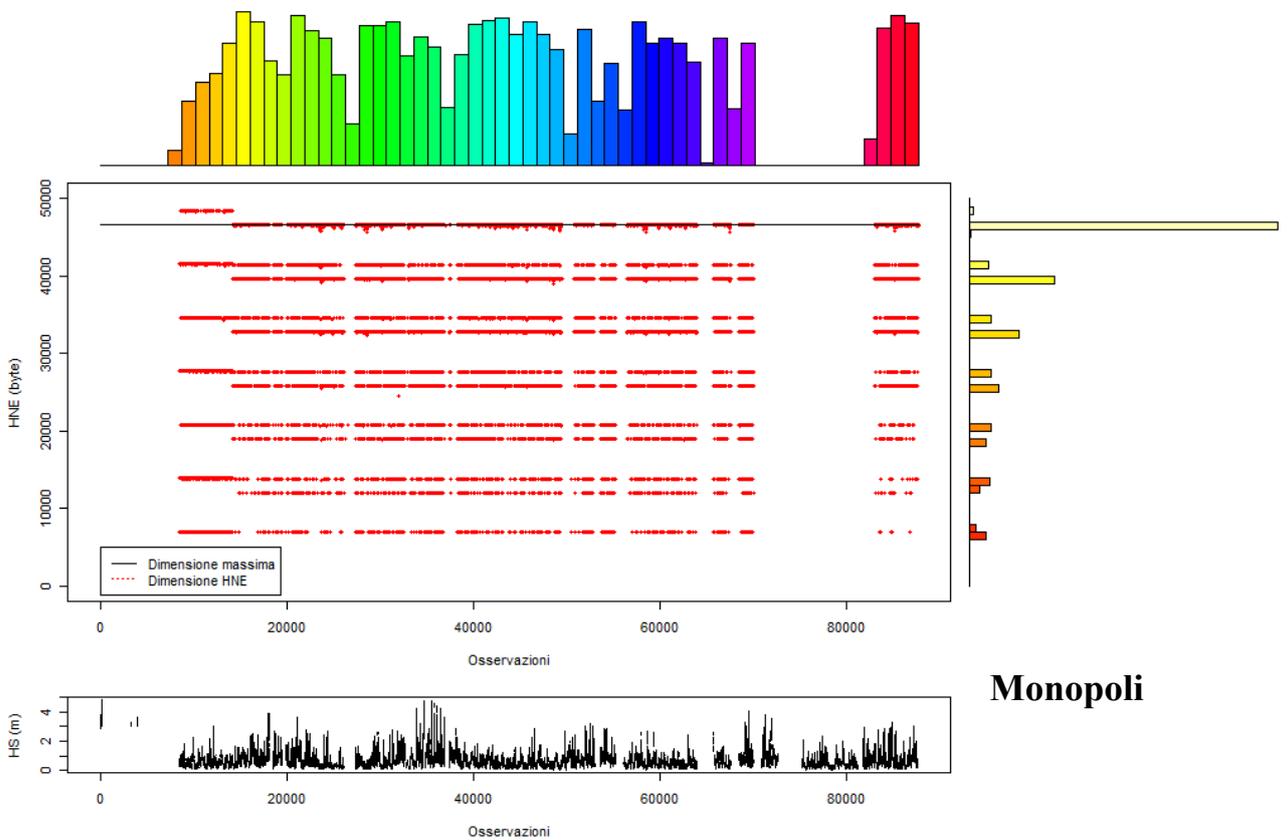
Punta della Maestra

Fig. 4. 9 - Diagrammi HNE – Punta della Maestra



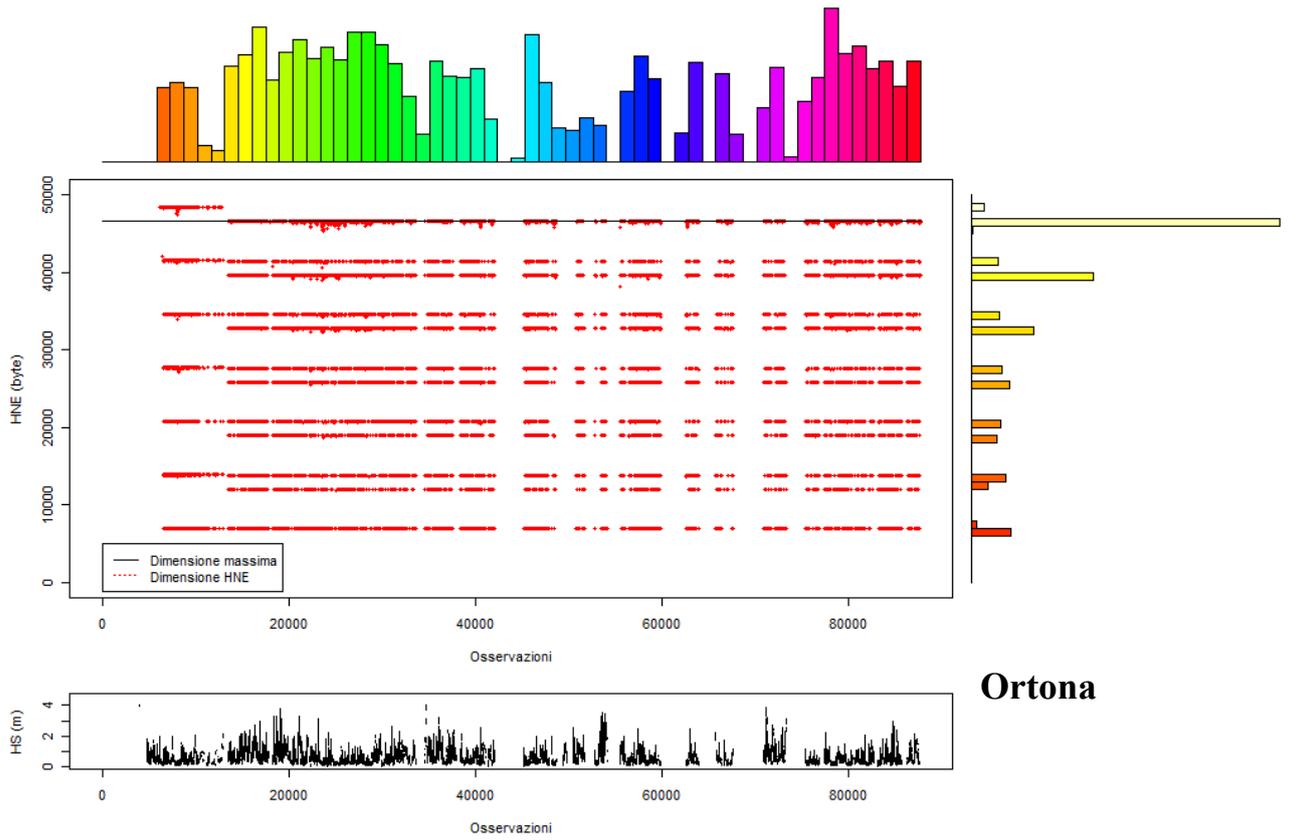
Mazara

Fig. 4. 10 - Diagrammi HNE - Mazara



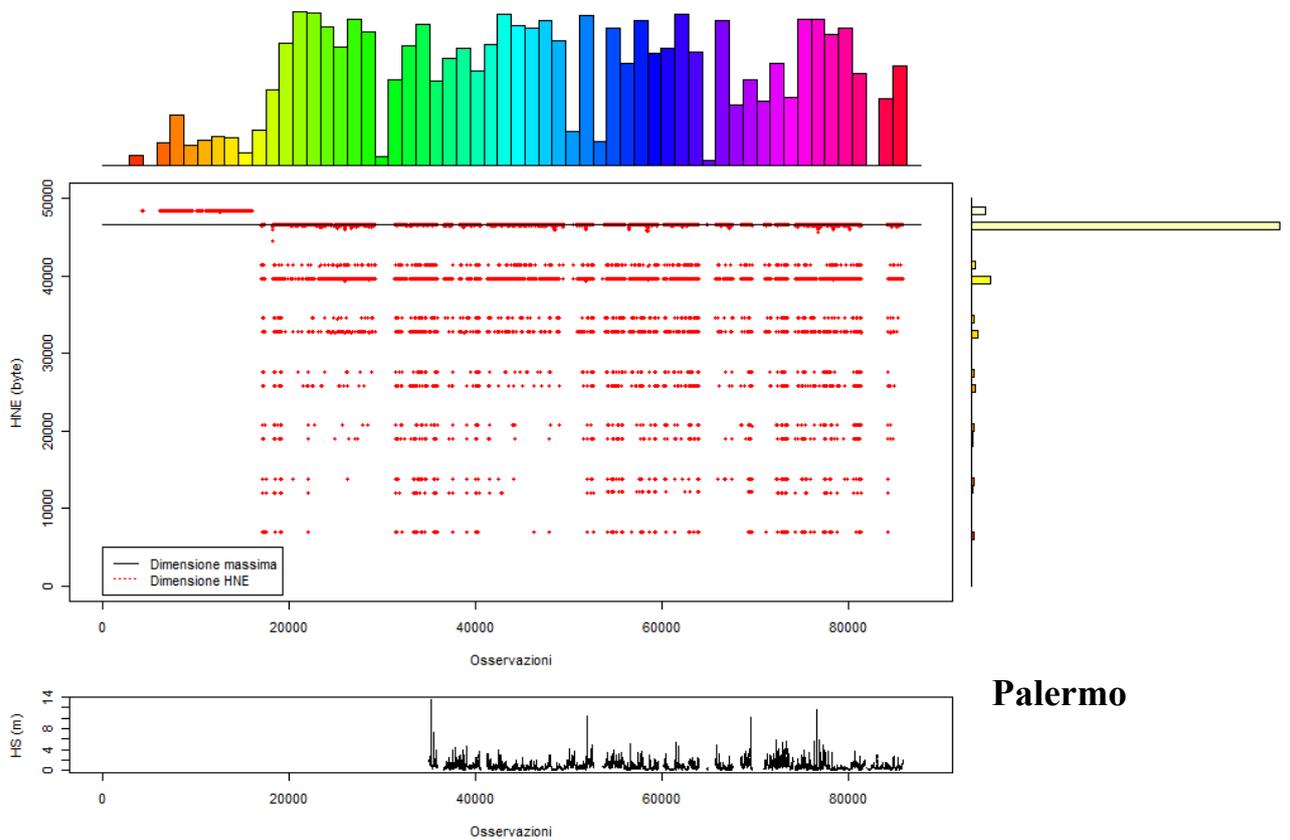
Monopoli

Fig. 4. 11 - Diagrammi HNE - Monopoli



Ortona

Fig. 4. 12 - Diagrammi HNE - Ortona



Palermo

Fig. 4. 13 - Diagrammi HNE - Palermo

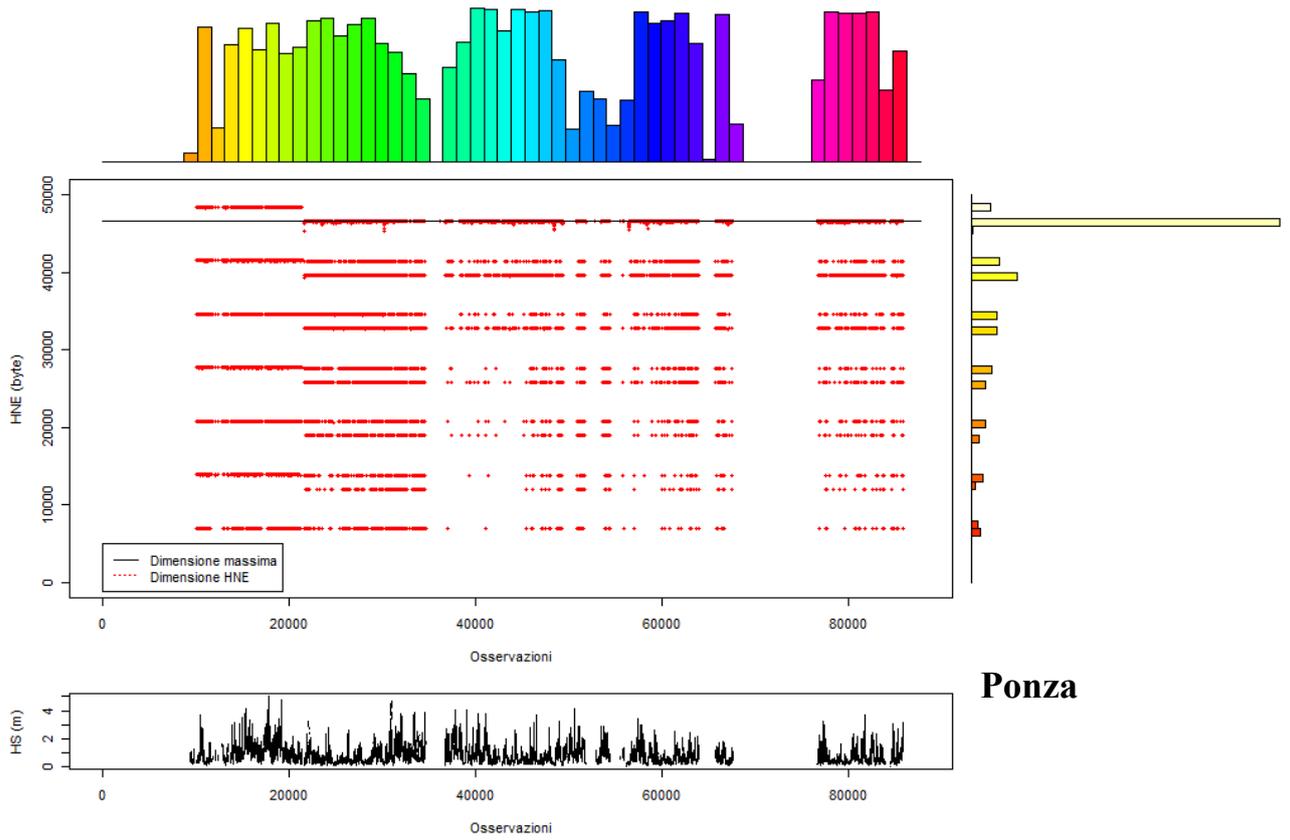


Fig. 4. 14 - Diagrammi HNE - Ponza

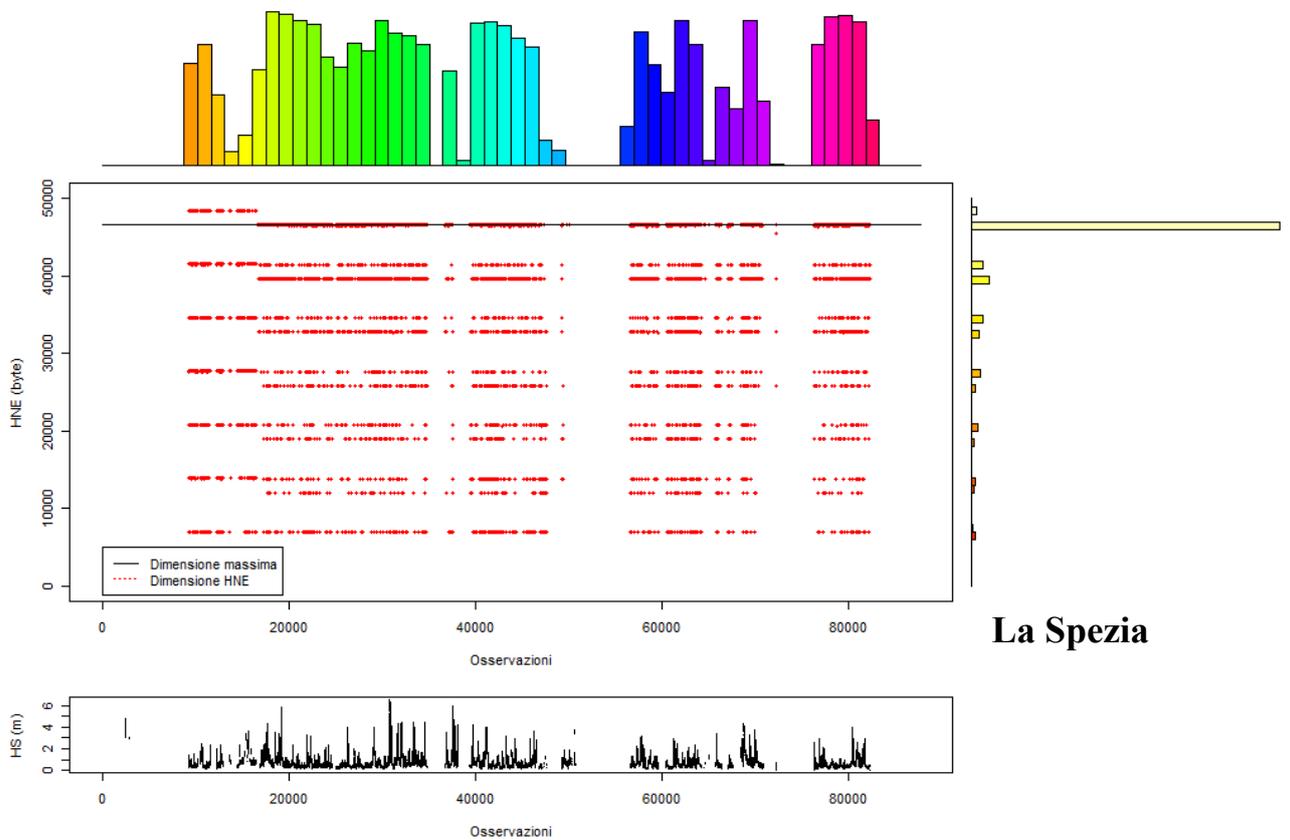


Fig. 4. 15 - Diagrammi HNE – La Spezia

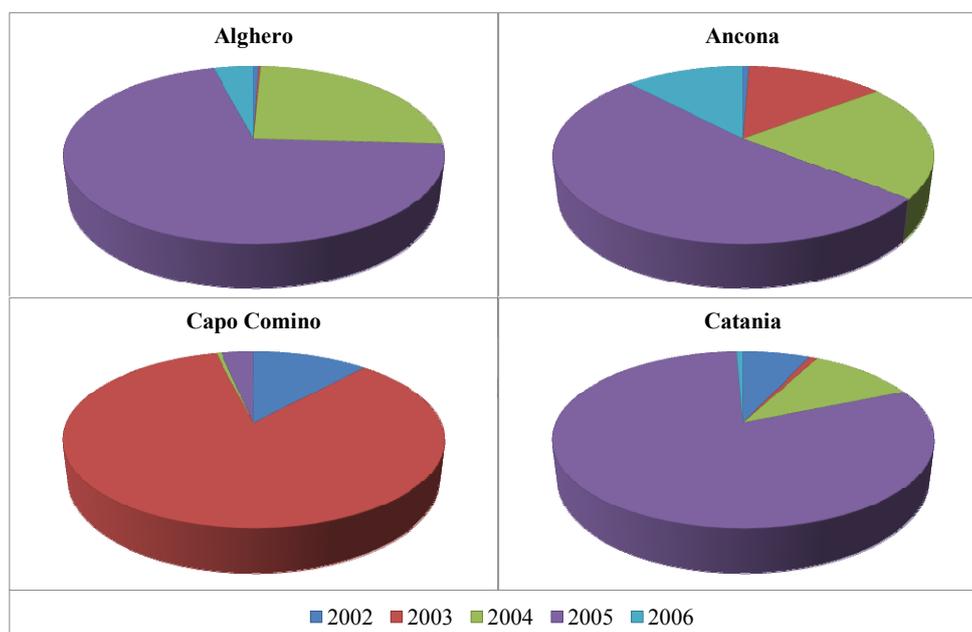
4.2 - Ricostruzione da file HNE

La ricostruzione dei dati H_s avviene tramite la semplice ricerca dei files HNE presenti a cui corrispondono campi mancanti nel dataset RON. I dati grezzi contenuti in questi files dovranno essere ri-processati dal software già utilizzato dall'ISPRA per l'elaborazione dell'analisi spettrale. Il risultato della ricerca è riassunto in tabella 4.2 in cui viene riportato anche il grado di affidabilità inteso come rapporto tra la dimensione del file recuperabile e la dimensione massima di un HNE.

	Dati mancanti	Dati recuperabili (affidabilità %)					
		Totale	2002	2003	2004	2005	2006
Alghero	34031	956 (60,3%)	5 (65,6%)	2 (100,0%)	241 (39,0%)	669 (68,9%)	39 (41,8%)
Ancona	44919	322 (79,1%)	2 (100,0%)	45 (96,8%)	68 (70,5%)	168 (79,4%)	39 (71,7%)
Capo Comino	64865	16093 (97,5%)	1889 (96,7%)	13590 (97,8%)	87 (76,2%)	527 (98,7%)	0 (-)
Catania	30101	628 (56,2%)	43 (38,9%)	5 (46,2%)	69 (42,2%)	507 (59,3%)	4 (96,0%)
Certaro	13572	1029 (94,4%)	42 (102,0%)	220 (98,3%)	76 (96,5%)	653 (92,9%)	38 (84,8%)
Civitavecchia	57161	20362 (94,8%)	6761 (89,2%)	12165 (97,8%)	61 (91,3%)	19 (92,0%)	1356 (96,5%)
Crotone	20892	1084 (98,5%)	5 (103,0%)	13 (98,8%)	37 (93,2%)	950 (99,8%)	79 (85,2%)
Mazara	23811	1164 (98,1%)	3 (63,7%)	420 (99,5%)	67 (98,0%)	517 (98,6%)	157 (93,4%)
Monopoli	23462	949 (70,6%)	3 (81,3%)	13 (79,2%)	145 (59,3%)	715 (70,1%)	73 (95,4%)
Ortona	38572	643 (63,6%)	5 (100,4%)	11 (82,6%)	170 (52,4%)	319 (67,0%)	138 (66,8%)
Palermo	46541	15749 (97,4%)	2192 (101,5%)	12915 (97,2%)	88 (78,8%)	494 (92,8%)	60 (68,0%)
Ponza	30425	522 (89,3%)	0 (-)	8 (89,1%)	27 (66,7%)	466 (91,6%)	21 (68,0%)
P. d. Maestra	77810	16123 (92,9%)	5401 (91,1%)	10608 (94,0%)	102 (83,6%)	12 (99,9%)	0 (-)
La Spezia	41172	772 (85,5%)	147 (98,4%)	2 (100,0%)	85 (62,4%)	525 (85,9%)	13 (72,4%)
Totale	547334	76396	16498	50017	1323	6541	2017

Tab. 4. 2 - Numero dei dati recuperabili con grado di affidabilità medio per ogni boa e ogni anno di osservazione

E' possibile notare come sia possibile recuperare un gran numero di dati nelle boe in cui minore è la quantità dei valori elaborati (Capo Comino, Civitavecchia, Palermo, Punta della Maestra). Inoltre, dai diagrammi successivi, si può notare la distribuzione negli anni dei valori recuperabili (concentrati soprattutto nel 2003 e nel 2005)



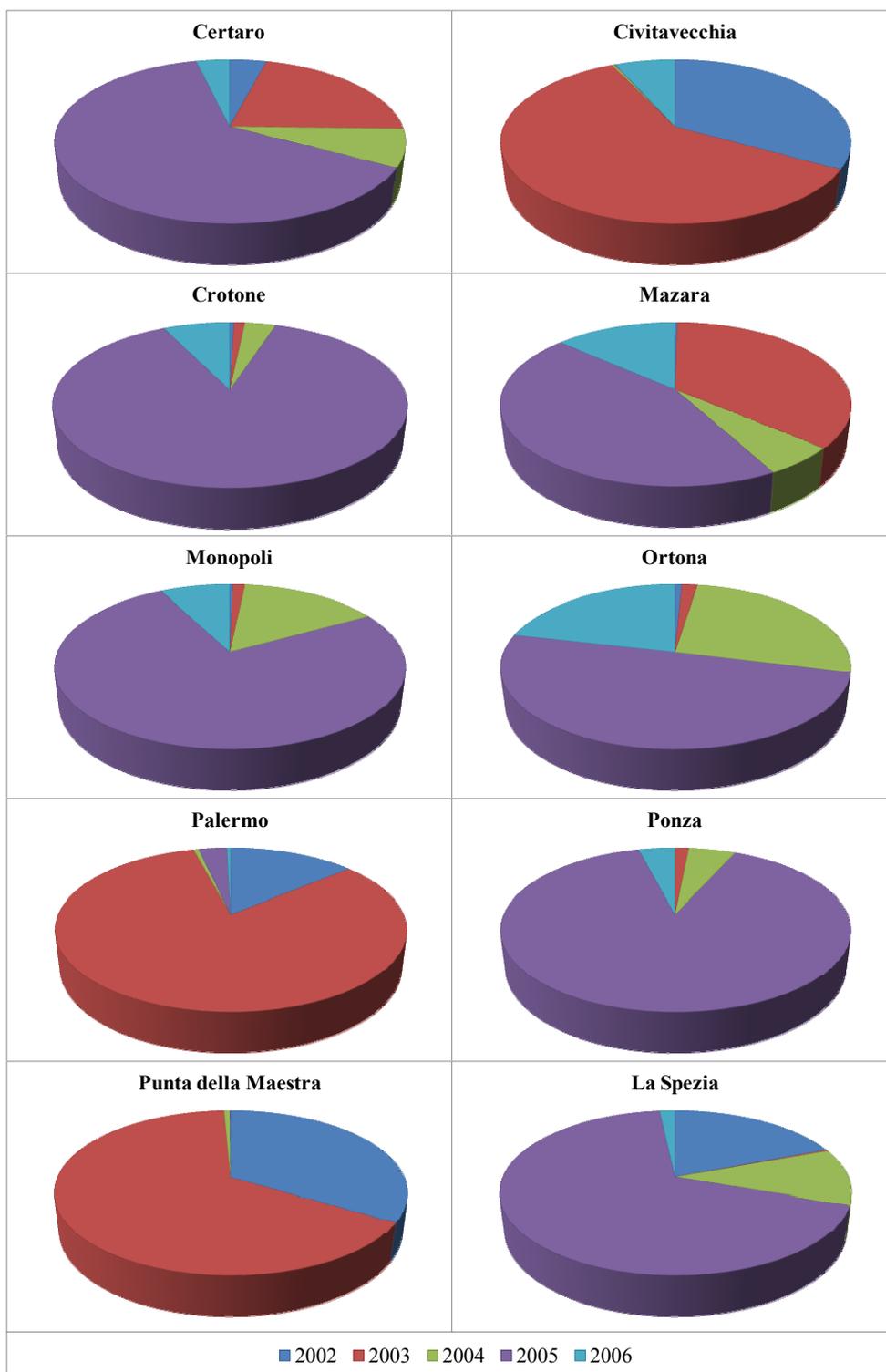


Fig. 4. 16 - Distribuzione temporale dei valori recuperabili

Una ulteriore considerazione va fatta circa la qualità del dato ricostruibile: in alcuni casi, la maggior parte dei file HNE idonei allo scopo sono in realtà incompleti (fig. 4.17, Ortona). Per poter effettuare l'analisi spettrale occorre riempire i campi vuoti con valori rappresentativi del periodo da analizzare. Le tecniche di riempimento dei dati prevedono la ripetizione del dato esistente n volte fino al completamento dei campi, oppure l'introduzione di dati relativi a campi vicini (es. i dati relativi a misurazioni semiorarie precedenti e/o successivi) o ancora l'utilizzo di valori neutri ai fini dell'analisi. Si pongono quindi dei problemi circa l'attendibilità del dato ricostruito che potrebbe risultare non rappresentativo della mezz'ora di analisi.

Infine è opportuno sottolineare che questo metodo, seppur estremamente affidabile poiché basato sui dati reali registrati dalle boe, consente di recuperare solo una percentuale ridotta dei dati mancanti nel dataset RON (13,96%).

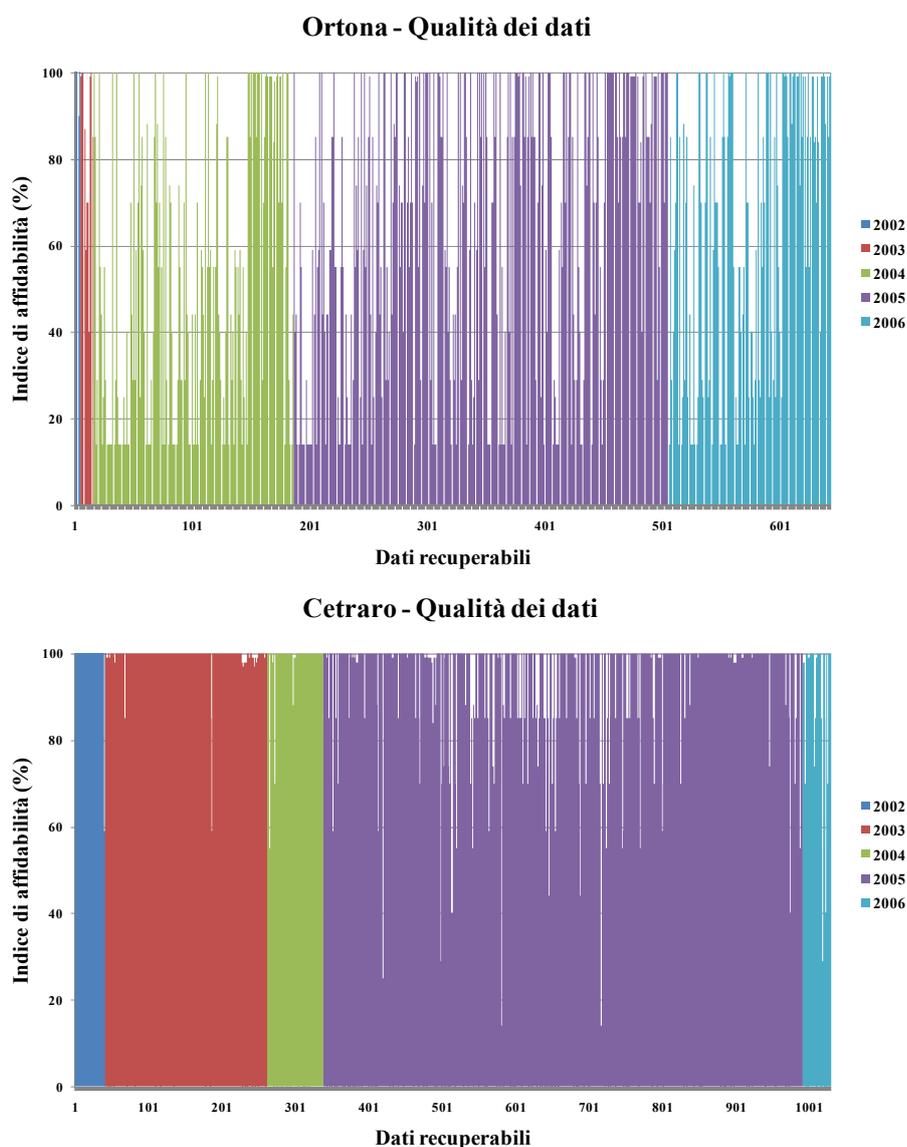


Fig. 4. 17 - Qualità dei dati ricostruibili ad Ortona e Cetraro

Conclusioni

Il Servizio Mareografico dell'ISPRA svolge funzioni che non si limitano alla sola raccolta e catalogazione dei dati. Le misurazioni delle reti RON e RMN, elaborate in diverse forme, contribuiscono a definire il clima ondoso e sono il punto di partenza per la progettazione di porti, la realizzazione di mezzi di trasporto marino, l'individuazione delle opere di difesa dei litorali, la valutazione degli effetti su strutture off-shore, la calibrazione di strumenti satellitari, la previsione di fenomeni di mareggiata e soprattutto per lo studio di fenomeni oceanografici ed atmosferici.

Lo studio di fenomeni su scala decennale, come le maree, o pluridecennale, come i cambiamenti climatici, necessitano quindi di misurazioni continue ed affidabili. Risulta quindi prioritario avere un database completo, caratterizzato da un'elevata qualità dei dati sia in termini di precisione che di affidabilità.

Dall'analisi svolta sul database RON si evince che il dato non sempre può essere misurato con continuità. Problemi di natura tecnica (malfunzionamenti o manutenzioni) o legati alle caratteristiche del mare (mareggiate che causano interferenze nelle trasmissioni o disormeggi) o ancora cause accidentali (urti di scafi) possono causare soluzioni di continuità nelle serie storiche anche di lunga durata. Cercare di ridurre questi gap sul "campo", tramite ad esempio l'utilizzo di sistemi di trasmissione più efficienti, comporterebbe interventi che, oltre ad essere economicamente onerosi, non garantirebbero in ogni caso la copertura totale delle misurazioni nel tempo.

La presenza di dati mancanti genera quindi:

- perdita di efficienza;
- problemi nella manipolazione e analisi dei dati;
- un "uso improprio" dei dati osservati, spesso considerati come rappresentativi dell'intero campione.

I metodi di imputazione multipla possono ovviare a tutti questi problemi. La possibilità di applicare un modello statistico robusto capace di ricostruire i dati ondometrici a partire da una serie numerosa di misurazioni avvenute, permetterebbe di ottenere valori affidabili che, pur rimanendo "non reali", riempiono il database consentendo di effettuare qualsiasi tipo di inferenza statistica.

Il presente lavoro ha permesso di dimostrare la bontà del modello di imputazione multipla applicato ai dati delle reti RON e RMN. L'utilizzo di questo metodo, tuttora innovativo e dalle capacità ancora inesplorate, ha permesso la ricostruzione dei dati di temperatura dell'acqua in tutto il periodo di riferimento. Nell'ambito delle ipotesi effettuate la stima dei parametri risulta relativamente precisa con valori standard error dell'ordine di 1°C. Inoltre lo studio ha dimostrato la stretta correlazione sia spaziale che temporale tra i valori misurati in mare aperto e i dati ricavati dalle stazioni ondametriche: sono sufficienti le sole misurazioni di temperatura dell'acqua delle due reti per avere una buona stima dei dati. È stato anche dimostrato che al variare dei parametri inseriti nel modello, cambia la precisione della stima. In particolare l'intervallo di confidenza relativo al singolo valore ricostruito decresce all'aumentare del numero di imputazioni e delle covariate inserite nel modello.

Tuttavia, le ipotesi semplificative introdotte per la definizione del modello non consentono di modellare in maniera precisa il fenomeno: le serie imputate infatti presentano una marcata componente stagionale sia annuale che giornaliera, ma l'andamento delle temperature in poche ore di osservazione risulta troppo altalenante. È evidente che l'indipendenza delle variabili risulta essere un'ipotesi troppo forte: è necessario quindi introdurre nuove dipendenze spazio-temporali tra i dati osservati sulla base delle correlazioni ricavate tra boe e mareografi.

La suddivisione del Mediterraneo in sottobacini (Tirreno, Adriatico e Ionio) risulta particolarmente efficace ai fini della ricostruzione dei dati. Ogni singolo bacino presenta infatti una notevole uniformità nei dati di temperatura dell'acqua. La qualità del dato ricostruito risulta elevata soprattutto nel Tirreno dove si evidenzia l'elevato numero di stazioni (6 boe e 10 mareografi) e la bassa numerosità dei missing values in determinati predittori. Lo Ionio, caratterizzato da un ridotto numero di stazioni, presenta valori dello standard error mediamente più elevati rispetto agli altri bacini.

L'analisi, inoltre consente di valutare in prima approssimazione la presenza di spike nelle serie osservate tramite il confronto tra dati ricostruiti e dati reali.

La ricostruzione dei dati di altezza d'onda risulta alquanto complessa a causa della presenza della componente direzionale difficilmente modellabile. L'analisi si è quindi limitata alla valutazione di quei dati grezzi che, opportunamente rielaborati, possono permettere la ricostruzione di un cospicuo numero di dati (pari a circa il 14% dei dati mancanti) soprattutto nelle stazioni ondametriche che presentano un elevato numero di missing values.

Sviluppi futuri

Benché i risultati ottenuti siano caratterizzati da un buon livello di precisione, è possibile ancora migliorare il modello attraverso:

- l'eliminazione dell'ipotesi di indipendenza statistica delle variabili: è possibile introdurre le correlazione spazio-temporali nel modello accostando i metodi di implementazione multipla a modelli autoregressivi non solo per la ricostruzione di dati, ma anche per la previsione di breve durata;
- introduzione di ulteriori covariate: nell'analisi possono essere inserite variabili quali la temperatura dell'aria (da fonti mareografiche) o l'altezza d'onda (dalla rete ondometrica), o ancora l'irraggiamento solare. Il numero di variabili è strettamente correlato con la precisione del modello: aumentando le covariate la ricostruzione potrà dare solo risultati maggiormente affidabili;
- analisi per longitudine. La temperatura varia con la longitudine: è quindi possibile ipotizzare una diversa suddivisione in settori delle stazioni di misura in modo da migliorare anche la ricostruzione dei valori dello Ionio incrementando il numero di variabili;
- test di validazione dei risultati ottenuti.

Lo stesso metodo può essere applicato all'altezza d'onda significativa. In questo caso è possibile utilizzare come covariate i valori di intensità e direzione del vento misurati dalle stazioni mareografiche. Dato che il fenomeno fisico alla base della generazione delle onde è estremamente complesso con forte variabilità sia nello spazio che nel tempo, occorre costruire dei modelli fisicamente basati a monte del metodo di imputazione multipla capaci di riprodurre le condizioni atmosferiche ed oceanografiche nell'intorno della boa.

Benché manchi una letteratura circa l'imputazione multipla applicata su massive dataset, a conclusione di questo lavoro si può affermare che il metodo risulta estremamente valido e consente di garantire dati di elevata qualità per analisi future. E' però opportuno sottolineare che questa applicazione non si può sostituire alla misurazione diretta delle grandezze in esame. Essa infatti contribuisce a ricostruire dati sulla base di misurazioni esistenti, ma non è in grado da sola di rappresentare il fenomeno generatore della variabile. Inoltre, come ogni metodologia statistica, l'imputazione multipla deve essere utilizzata con prudenza, altrimenti si corre il rischio di generare dati non coerenti con la realtà osservata.

Molta attenzione deve essere data quindi alla verifica delle assunzioni fatte e dei risultati ottenuti, e alle variabili che influenzano il modello. Un modello robusto e affidabile si può costruire solo se tutti i fattori in esso inclusi sono rappresentativi del fenomeno effettivamente osservato.

Bibliografia

- [1]. Accadia, C., Arena, G., Barbano, A., Bencivenga, M., Casaioli, M., Colagrossi, A., et al. (2003). *Il Sistema Idro-Meteo-Mare e le reti di monitoraggio marino dell'APAT*.
- [2]. Anselmi, A., Chiodini, P., & Verrecchia, F. (2008). *ESeC-Rubin missing value interpretation for a regional bottom-up hierarchical forecasting*.
- [3]. APAT. (1997). *Bollettino delle rilevazioni ondametrische. Trimestre Luglio - Settembre 1989*. Istituto Poligrafico e Zecca dello Stato.
- [4]. APAT. (2002). *Progetto Atlante Costiero - Il moto ondoso lungo le coste italiane*.
- [5]. APAT, Università degli Studi Roma Tre. (2005). *Atlante delle onde nei mari italiani*.
- [6]. Arena, F., & Barbaro, G. (1999). *Il rischio ondoso nei mari italiani*. Editoriale Bios.
- [7]. AxyS Technologies. (s.d.). <http://www.axystechnologies.com/products/triaxys.asp>.
- [8]. Buongiorno Nardelli, B., Santoleri, R., Marullo, S., & Guarracino, M. (2007). *La temperatura superficiale del Mar Mediterraneo negli ultimi 21 anni: analisi delle misure satellitari*.
- [9]. Cressie, N. A. (1993). Statistics for spatial data. In N. A. Cressie, *Statistics for spatial data*. Wiley Interscience.
- [10]. Envirtech. (s.d.). <http://www.envirtech.org/ron.htm>.
- [11]. Envirtech, & Ruggiero, F. (2008). *RON - Rete Ondametrisca Nazionale - Interventi straordinari per la riorganizzazione e normalizzazione dei dati acquisiti tra il 1999 ed il 2008*.
- [12]. Franco, L. (2006). Dispense del corso di Ingegneria Costiera - *Fonti di dati ondametrischi*.
- [13]. Franco, L. (2006). Dispense del corso di Ingegneria Costiera - *Il vento*.
- [14]. Guarnera, U. (2004). *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi. Il software QUIS*.
- [15]. Horton, N. J., & Lipsitz, S. R. (Vol. 55, No 3, 2001). *Multiple Imputation in practice: comparison of software packages for regression model with missing values*. Journal of American Statistical Association, (p. pp. 244-254).
- [16]. ISPRA. (s.d.). www.idromare.it.
- [17]. ISPRA. (s.d.). www.mareografico.it.
- [18]. Kohn, R., & Ansley, C. F. (1986). *Estimation, prediction and interpolation for ARIMA models with missing values*. Journal of the American Statistical Association, Vol. 81, No 395, pp. 751-761.
- [19]. Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*.
- [20]. Longford, N. T. (2002). *Examples of multiple imputation*. Statistics Canada Symposium.
- [21]. Melani, S., Orlandi, A., Brandini, C., & Ortolani, A. (2007). *Temperatura superficiale del mare da dati satellitari*.
- [22]. NOAA. (2003). *Triaxys directional wave buoy for nearshore wave measurements - Test and evaluation plan*. NOAA Technical Report .
- [23]. Piscopia, R., Inglesini, R., Panizzo, A., Corsini, S., & Franco, L. (2002). *Analysis of 12 years wave measurements by the italian wave network*. Coastal Engineering Conference .

-
- [24]. Puca, S., Tirozzi, B., Arena, G., Corsini, S., & Inghilesi, R. (2001). *A neural network approach to the problem of recovering lot data in a network of marine buoys*. Conference of ISOPE, (p. pp. 615-620). Norway.
- [25]. Reiter, J. P., & Raghunathan, T. E. (2007). *The multiple adaptations of multiple imputation*.
- [26]. Rubin, D. (1996). *Multiple Imputation After 18+ Years*. Journal of the American Statistical Association , Vol. 91, No. 434, pp. 473-489.
- [27]. Sarle, W. S. (1998). *Prediction with missing inputs*.
- [28]. Schafer, J. (1997). *Analysis of incomplete multivariate data*.
- [29]. Schafer, J. L. (s.d.). Tratto da http://www.stat.psu.edu/~jls/missing_data/index.html.
- [30]. Schafer, J., & Graham, J. (2002). *Missing Data: our view of the state of the art*.
- [31]. Schafer, J., & Olsen, M. (1998). *Multiple Imputation for multivariate missing-data problems: a data analyst's perspective*.
- [32]. Soldani, M., Regazzoni, C., Bozzano, R., Faggioni, O., & Musso, M. (2008). *Misure ambientali in mare aperto: sviluppo di tecnologie per l'acquisizione e l'analisi di dati meteo-mareografici misurati da una boa oceanografica in mar ligure*.
- [33]. U.S. Army Coastal Engineering Research. (2002). *Shore Protection Manual*.
- [34]. United nations statistical commission. (2000). *Glossary of terms on statistical data editing*.
- [35]. Van Buuren, S., & Oudshoorn, C. (2005). *Multivariate imputation by chained equations - MICE v1.14 User's manual*.
- [36]. Van Buuren, S., Brand, J., Groothuis-Oudshoorn, C., & Rubin, D.B. (2005). *Fully conditional specification in Multivariate Imputation*.
- [37]. *Wikipedia*. (s.d.). Tratto da <http://it.wikipedia.org/wiki/Box-plot>.
- [38]. www.multiple-imputation.com. (s.d.).
- [39]. Zhang, P. (2003). *Multiple Imputation: Theory and Method*.