

UNIVERSITÀ DEGLI STUDI DI TORINO

SCUOLA DI SCIENZE DELLA NATURA

Corso di Laurea Magistrale Interateneo in Fisica dei Sistemi Complessi



Tesi di Laurea Magistrale

Studio di dati di rianalisi di venti e onde a
largo di Alghero per il periodo 1979-2019

Relatore:
Prof. Miguel Onorato

Controrelatrice:
Prof. ssa Elisa Palazzi

Candidato:
Gennaro Martone

Anno Accademico 2021/2022

*Science is not only a discipline of reason but, also,
one of romance and passion*

Stephen Hawking

Indice

1	Introduzione - La rianalisi climatica	1
2	Teoria	3
2.1	Assimilazione dei dati	3
2.1.1	BLUE scalare	4
2.1.2	4D-Var	8
2.2	Equazione di bilancio delle onde	13
3	Downscaling dinamico e modello delle onde WAVEWATCH	19
3.1	Downscaling dinamico	19
3.2	WAVEWATCH III e setup del modello delle onde	22
3.3	Calibrazione e validazione del modello	23
4	Analisi dei dati	24
4.1	Rose di onde e venti	24
4.1.1	Rosa delle onde	25
4.1.2	Rose dei venti	29
4.2	Scatter plot e coefficiente di correlazione dei ranghi di Spearman	32
4.3	Analisi aggregata per stagione	36
4.4	Fit dei dati di venti e onde	41
4.5	Analisi per anno	44
5	Conclusioni	47
	Bibliografia	49

Capitolo 1

Introduzione - La rianalisi climatica

Lo studio del clima e l'affinamento di modelli predittivi sono sempre stati di importanza vitale per l'uomo. Tali studi ci permettono di poter predire quanta pioggia farà in futuro, quali regioni saranno più colpite da venti estremi, di quanto si alzeranno le maree a causa dell'innalzamento della temperatura, quali specie sono più a rischio a causa di cambiamenti nel clima del loro habitat. Questi sono solo alcuni dei risultati dell'analisi climatica e il cambiamento climatico ha reso più evidente la necessità di avere modelli affidabili sul clima.

Appare importante quindi avere uno storico di dati ricco e quanto più preciso possibile del clima del passato. I dati del passato però non sono sempre completi (soprattutto se paragonati all'abbondanza di misurazioni moderne) e presentano, in molte zone geografiche, lacune che generalmente aumentano andando indietro nel tempo.

La rianalisi climatica si propone come obiettivo di risolvere queste lacune combinando osservazioni del passato ottenute tramite stazioni meteo, palloni sonda, aerei, navi e satelliti, con modelli meteorologici del presente. Si può immaginare il clima del passato come un puzzle con pezzi mancanti, la rianalisi si occupa di colmare quei pezzi mancanti (figura 1.1).

I dati di rianalisi forniscono molte informazioni non solo sull'atmosfera (temperatura, venti, precipitazioni), ma anche sulla superficie terrestre e sugli oceani.

Lo scopo del presente lavoro è quello di analizzare i dati di downscaling dinamico, ricavati dalla rianalisi climatica ERA5 per la zona di Alghero. È stata analizzata la composizione dei dati di venti e di onde e sono stati ricercati eventuali trend in atto nelle serie storiche.

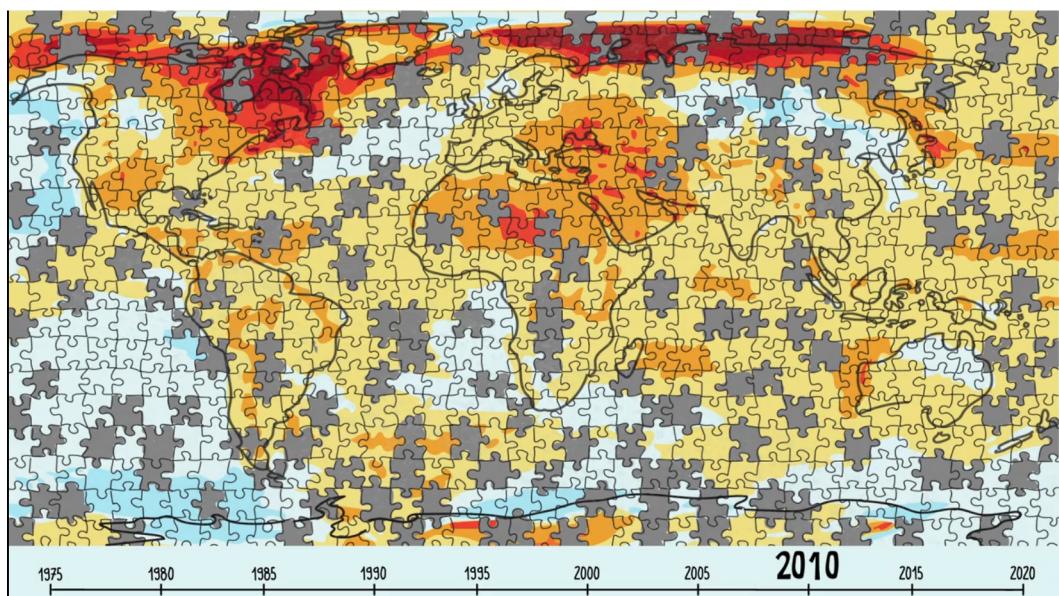


Fig. 1.1: Carta geografica terrestre. I pezzi del puzzle rappresentano osservazioni di grandezze fisiche, i pezzi del puzzle mancanti rappresentano invece porzioni geografiche in cui sono assenti osservazioni.

Capitolo 2

Teoria

In questo capitolo verranno esposte le conoscenze teoriche necessarie per comprendere l'assimilazione dei dati, necessaria per lo svolgimento della rianalisi climatica. Inoltre, verrà introdotta l'equazione spettrale di bilancio dell'azione per le onde, sulla quale si basano i modelli di onde.

2.1 Assimilazione dei dati

Il cuore della rianalisi climatica risiede nell'assimilazione dei dati.

L'assimilazione dei dati è il processo di stima dello stato di un sistema dinamico combinando dati osservazionali con una stima a priori dello stato (tipicamente un modello di forecast numerico). In più è necessario utilizzare la conoscenza che si ha della dinamica del sistema, le proprietà fisiche dell'atmosfera e l'incertezza sui dati osservazionali.

I dati osservazionali non sono sufficienti né per la comprensione dello stato del sistema né per inizializzare un modello di forecast qualsiasi. Per inizializzare, ad esempio, un modello di forecast atmosferico o oceanico bisognerebbe inizializzare per ogni punto della griglia il valore della temperatura, dell'umidità, della velocità del vento etc. Il numero dei gradi di libertà del problema eccede di molto il numero di osservazioni a disposizione, che spesso sono parziali e non descrivono interamente lo stato del modello. La mancanza di dati dipende significativamente dalla posizione spaziale, nell'atmosfera le misurazioni sono molto più rade rispetto alle misurazioni in superficie, e persino in superficie ci sono zone con grosse lacune di dati (per esempio l'Africa).

Una stima a priori dello stato del modello è quindi necessaria per approcciare un problema di enorme dimensionalità, ed è dato da un modello di forecast precedente.

Un esempio di efficacia del metodo di assimilazione dati è quella del buco dell'ozono nel 2002 intorno al polo sud in figura 2.1. L'assimilazione dei dati svolta mostra la divisione in due parti distinte del buco dell'ozono nell'arco dei giorni fra il 22 e il 26 Settembre del 2002. Questo effetto sarebbe stato impossibile da ottenere col solo modello di forecast e risulta difficilmente intuibile dalla sola osservazione dei dati, poiché parziali.

Inoltre le sole osservazioni sono insufficienti perché non forniscono sempre le grandezze da inizializzare, ma funzioni di queste quantità, che non sono sempre lineari.

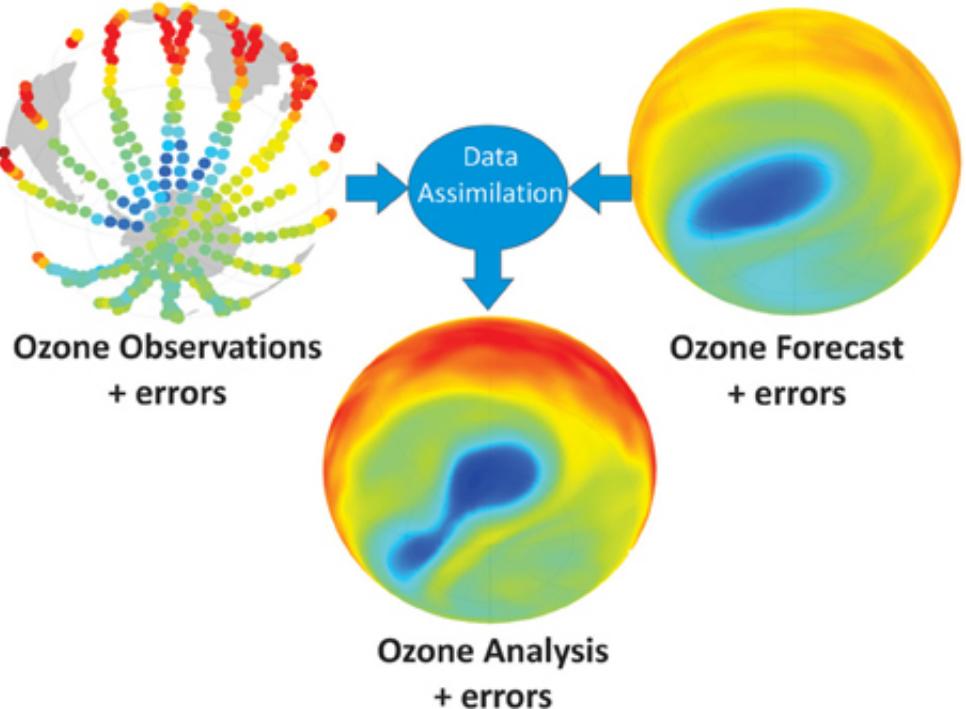


Fig. 2.1: Assimilazione di dati per quanto riguarda il buco nell’ozono nella zona del polo sud. A sinistra vi sono le osservazioni ottenute dal satellite nel tempo, a destra il modello di forecast a priori. Il risultato della combinazione dei dati osservazionali con lo stato a priori del modello fa emergere il fenomeno dello split in due parti del buco dell’ozono. [1]

Si può definire, per esempio, il model state $\mathbf{x} = \begin{pmatrix} u \\ \nu \end{pmatrix}$ dove u è la componente Est-Ovest del vento e ν è la componente Nord-Sud. Quello che si osserva non sono le due componenti separate, ma $y = w_s$, la velocità del vento totale. Il modello di forecast lavora invece sulle componenti in maniera separata e quindi bisogna definire l’operatore osservazione $H(x)$ che computa:

$$H(x) = \sqrt{u^2 + \nu^2} \quad (2.1)$$

Solo a questo punto si possono comparare le osservazioni y con le predizioni del modello $H(x)$.

Qual è il modo per tenere conto di tutte le sorgenti di informazioni e delle loro incertezze? Un metodo è il **BLUE** (Best linear unbiased estimate).

2.1.1 BLUE scalare

Per approcciare il problema vettoriale si partirà con un esempio scalare [2].

Si suppone di avere una stima di background e un’osservazione di una quantità fisica scalare come la temperatura, T_b e T_o . Si assume che le stime siano non correlate fra di loro e che siano unbiased. La migliore stima dell’analisi è una combinazione lineare

CAPITOLO 2. TEORIA

delle due quantità:

$$T_a = \alpha_b T_b + \alpha_o T_o \quad (2.2)$$

La scelta dei pesi α_b e α_o dipende ovviamente dalle incertezze delle due quantità. Se l'osservazione e la stima di background hanno la stessa incertezza, ci si aspetta che l'analisi sia la media delle due.

Per trovare i valori esatti bisogna imporre due condizioni:

1. la T_a deve essere unbiased
2. la T_a deve minimizzare l'errore

Sapendo che le grandezze possono essere scritte in funzione del valore vero di T :

$$\begin{aligned} T_o &= T_t + \epsilon_o \\ T_b &= T_t + \epsilon_b \\ T_a &= T_t + \epsilon_a \end{aligned}$$

allora, con qualche riarrangiamento matematico, imponendo che l'errore di analisi sia unbiased:

$$\begin{aligned} \langle \epsilon_a \rangle &= \langle T_a - T_t \rangle = \langle \alpha_b T_b + \alpha_o T_o - T_t \rangle = \langle \alpha_b(T_b - T_t) + \alpha_o(T_o - T_t) + (\alpha_b + \alpha_o - 1)T_t \rangle = \alpha_b \langle \epsilon_b \rangle + \alpha_o \langle \epsilon_o \rangle + (\alpha_b + \alpha_o - 1) \langle T_t \rangle = 0. \end{aligned}$$

L'osservazione e la stima di background sono unbiased, quindi le due quantità in rosso sono nulle. La condizione di analisi si riduce quindi alla seguente equazione:

$$\alpha_b + \alpha_o = 1 \quad (2.3)$$

Per questo motivo l'equazione 2.2 diventa:

$$T_a = \alpha_b T_b + (1 - \alpha_b) T_o \quad (2.4)$$

Rimane dunque un solo parametro da stimare. Si deve quindi imporre la seconda condizione, quindi trovare il parametro α_b che minimizzi l'incertezza. Per minimizzare l'errore si minimizza la varianza. Assumendo di conoscere la varianza dell'osservazione e del background, si ha:

$$\begin{aligned} \langle \epsilon_b^2 \rangle &= \sigma_b^2 \\ \langle \epsilon_o^2 \rangle &= \sigma_o^2 \\ \langle \epsilon_a^2 \rangle &= \sigma_a^2 \end{aligned}$$

Allora, la varianza della T_a è:

$$\begin{aligned} \langle \sigma_a^2 \rangle &= \langle (T_a - T_t)^2 \rangle = \langle (\alpha_b T_b + (1 - \alpha_b) T_o - T_t)^2 \rangle = \langle (\alpha_b(T_b - T_t) + (1 - \alpha_b)(T_o - T_t))^2 \rangle = \langle (\alpha_b \epsilon_b + (1 - \alpha_b) \epsilon_o)^2 \rangle = \alpha_b^2 \sigma_b^2 + (1 - \alpha_b)^2 \sigma_o^2. \end{aligned}$$

CAPITOLO 2. TEORIA

Per ottenere il seguente risultato è stata sfruttata l'indipendenza fra gli errori dell'osservazione e del background (covarianza nulla). A questo punto, si trova il minimo della varianza, imponendo la derivata della varianza rispetto al parametro α_b nulla, si ottiene:

$$\alpha_b = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} \quad (2.5)$$

Per questo motivo l'equazione 2.4 diventa:

$$T_a = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} T_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} T_o \quad (2.6)$$

con:

$$\sigma_a^2 = \frac{\sigma_o^2 \sigma_b^2}{\sigma_o^2 + \sigma_b^2} < \min(\sigma_o^2, \sigma_b^2) \quad (2.7)$$

La T_a viene ottenuta pesando le incertezze delle quantità di background e di osservazione, quella con minore incertezza contribuisce di più alla stima. Inoltre si ha la garanzia che la varianza della quantità di analisi sia minore sia delle varianze di osservazione e di background (equazione 2.7).

L'analisi compiuta si chiama **BLUE** (best linear unbiased error) ed è stata applicata ad un problema scalare.

Per generalizzare il problema ad un model state con molte variabili bisogna utilizzare il concetto di **funzione di densità di probabilità** per rappresentare l'incertezza.

Assumendo di avere una distribuzione a priori dello stato x dato da $p(x)$ e di avere delle osservazioni con una probabilità condizionata $p(y|x)$, il teorema di Bayes afferma che:

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} \quad (2.8)$$

La stima a posteriori è dunque uguale al prodotto della stima a priori per la probabilità delle osservazioni (likelihood) diviso la probabilità delle osservazioni a priori, la quale è una costante di x e può essere trattata come una costante di normalizzazione.

Il model state con probabilità massima date le osservazioni dipende dalla distribuzione a priori e dalla likelihood, la curva più piccata intorno ad un valore di model state contribuisce con un peso maggiore (figura 2.2).

Nella pratica le funzioni densità di probabilità sono altamente dimensionali (nell'ordine di 10^9). Per questo motivo non è possibile calcolare analiticamente la loro forma completa, piuttosto si cerca di calcolare solo la media o la moda della funzione (uno stimatore) oppure di generare campioni da essa.

Per approcciare il problema si assume la gaussianità delle funzioni densità di probabilità. In molte dimensioni, oltre al valore medio bisogna tenere in conto della matrice di covarianza. Formalmente, queste assunzioni portano alle seguenti equazioni:

$$p(x) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} \exp\left\{-\frac{1}{2}(x - x_b)^T P^{-1}(x - x_b)\right\} \quad (2.9)$$

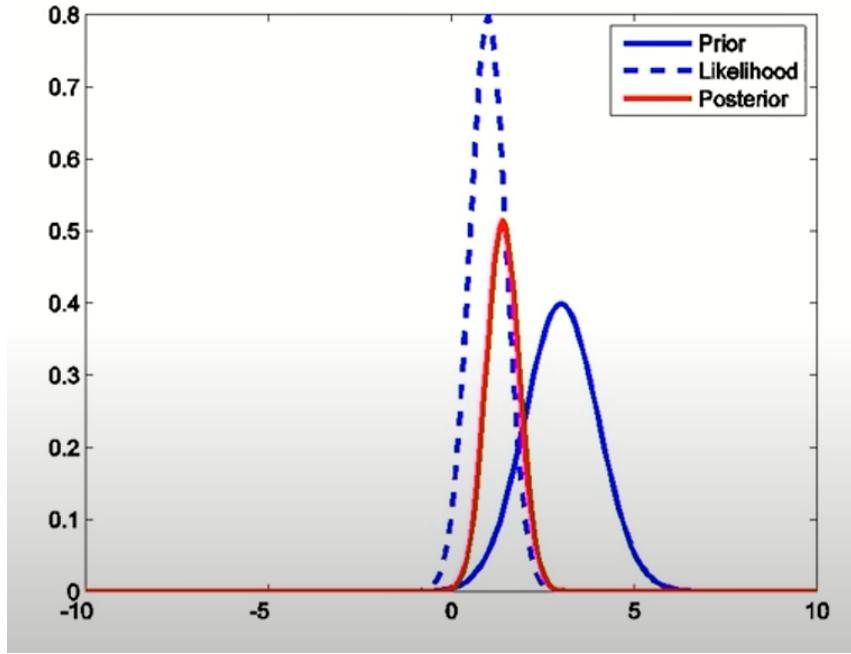


Fig. 2.2: Esempio di funzioni densità di probabilità. La curva blu rappresenta la distribuzione a priori, la blu tratteggiata la likelihood, in rosso la distribuzione a posteriori. La distribuzione a posteriori si trova fra le altre due curve, più vicina alla distribuzione più piccata, nell'esempio, la likelihood.

$$p(y|x) = \frac{1}{(2\pi)^{n/2}|E|^{1/2}} \exp\left\{-\frac{1}{2}(y - H(x))^T E^{-1}(y - H(x))\right\} \quad (2.10)$$

della funzione densità di probabilità a priori e della likelihood. Le matrici di covarianza sono E e P per la funzione a priori e la likelihood rispettivamente. Di conseguenza:

$$p(x|y) \propto \exp\left\{-\frac{1}{2}((y - H(x))^T E^{-1}(y - H(x)) + (x - x_b)^T P^{-1}(x - x_b))\right\} \quad (2.11)$$

Lo scopo di questa analisi è trovare il model state per cui si ha la **MAP** (Maximum a Posteriori Probability). La probabilità massima si ha quando il model state x minimizza la seguente funzione, detta di **costo**:

$$J(x) = (x - x_b)^T P^{-1}(x - x_b) + (y - H(x))^T E^{-1}(y - H(x)) \quad (2.12)$$

Le matrici P e R di covarianza sono dell'ordine di $10^9 \times 10^9$, impossibili da tenere in memoria, per cui bisogna trovare un modo approssimato per minimizzare tale funzione.

Uno dei metodi per approcciare questo problema in maniera approssimata è il metodo **variazionale**, di cui si considerano due algoritmi:

- Assimilazione tridimensionale (3D-Var), in cui si considerano le tre dimensioni spaziali.

- Assimilazione quadridimensionale, dove si considerano le tre dimensioni spaziali più il tempo. In questo caso l'operatore osservazione deve includere il modello dinamico.

2.1.2 4D-Var

Lo scopo del 4D-Var è di stimare il vero stato del sistema (analisi), consistente sia con le osservazioni distribuite nel tempo che con la dinamica del sistema. La funzione di costo definita precedentemente deve essere aggiornata, per tenere conto delle differenti osservazioni ad istanti di tempo diversi[3]:

$$J(x) = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2} \sum_{t=0}^{\Delta t} (H_t^o[x(t)] - y(t))^T E_t^{-1}(H_t^o[x(t)] - y(t)) = J_b + J_o \quad (2.13)$$

Al posto di P è stato inserito B, assumendo che la covarianza degli errori di background non cambi col tempo. La matrice B contiene sulla diagonale la varianza delle componenti del background state, e gli elementi fuori dalla diagonale misurano la correlazione non solo tra stesse grandezze in differenti posizioni nello spazio, ma anche tra differenti grandezze.

H_t^o rappresenta il modello osservazionale *forward* (diretto) che appare nel termine J_o , ed agisce direttamente sul model state $x(t)$. L'operatore di evoluzione temporale, M_t , produce questo stato a partire da x:

$$x(t) = M_t M_{t-\delta t} \dots M_{\delta t} x \quad (2.14)$$

A causa della complessità di M_t e H_t^o , J_o è molto più complicata rispetto al termine di background. Date le condizioni iniziali, x, la predizione del modello delle osservazioni al tempo t è:

$$y^{model}(t) = H_t^o[x(t)] = H_t^o[M_t M_{t-\delta t} \dots M_{\delta t} x] \quad (2.15)$$

come illustrato in figura 2.3, per tutti gli istanti di tempo nella finestra temporale considerata.

Vanno dunque trovate le condizioni iniziali x , tali da minimizzare la funzione di costo. Per fare ciò si calcola il gradiente di J . Inizialmente si espande l'equazione dalla sua forma matriciale alla sua forma in serie:

$$J(x) = \frac{1}{2} \sum_{ij} (x^j - x_b^j) B_{ij}^{-1} (x^j - x_b^j) + \frac{1}{2} \sum_{t=0}^{\Delta t} \sum_{mn} (H_t^{o,m}[x(t)] - y^m(t)) E_{mn}^{-1} (H_t^{o,n}[x(t)] - y^n(t)) \quad (2.16)$$

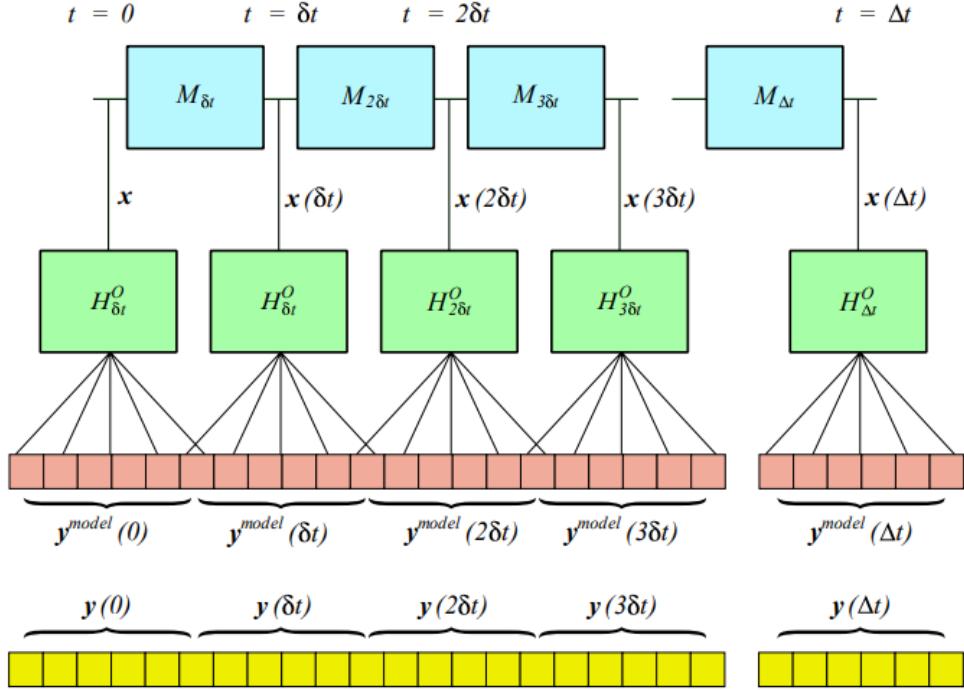


Fig. 2.3: Illustrazione schematica del modello diretto. L’operatore dell’evoluzione temporale del modello, l’ M_t (quadrati blu) muove il modello in avanti nel tempo, l’operatore osservazione H_t^o usa ogni model state per produrre le sue previsioni. Tali previsioni (nella figura, in rosa) vengono confrontate con le vere osservazioni (in giallo).[3]

dove gli indici i,j scorrono sulle componenti del model state, m,n scorrono sulle componenti delle osservazioni, e t scorre su tutti i time steps, di lunghezza δt . Il gradiente rispetto alla componente individuale x^k è:

$$\frac{\partial J}{\partial x^k} = - \sum_i B_{ki}^{-1} (x^j - x_b^j) + \sum_{t=0}^{\Delta t} \sum_{mn} (H_t^{o,n}[x(t)] - y^n(t)) E_{mn}^{-1} \frac{\partial H_t^{o,m}[x(t)]}{\partial x^k} \quad (2.17)$$

dove x^k è la k-esima componente dello stato iniziale del modello (nell’ipotesi corrente). L’espressione è stata semplificata sfruttando la simmetria delle matrici di covarianza ($B_{ij}^{-1} = B_{ji}^{-1}$ e $E^{-1}_{mn} = E^{-1}_{nm}$). Tutte le componenti del gradiente possono essere unite in un vettore colonna, così facendo si ottiene la forma compatta:

$$\nabla_x J = -B^{-1}(x_b - x) - \sum_t \left(\frac{dH_t^o[x(t)]}{dx} \right)^T E^{-1}(H_t^o[x(t)] - y(t)) \quad (2.18)$$

dove $\frac{dH_t^o[x(t)]}{dx}$ è lo Jacobiano della matrice $H_t^o[x(t)]$. Dato che le derivate sono calcolate rispetto alle coordinate del model state iniziale, uno Jacobiano più naturale da calcolare sarebbe $\frac{dH_t^o[x(t)]}{dx(t)}$ piuttosto di $\frac{dH_t^o[x(t)]}{dx}$. Conoscendo la relazione tra model state al tempo t e al tempo 0 (equazione 2.14), si può riscrivere l’equazione 2.18 nel seguente modo:

$$\nabla_x J \approx -B^{-1}(x_b - x) - \sum_t M_{\delta t}^{\prime T} \dots M_{t-\delta t}^{\prime T} M_t^{\prime T} H_t'^o E^{-1} (H_t^o[x(t)] - y(t)) \quad (2.19)$$

In questa equazione gli operatori del model state diretto sono accentati per distinguere dai simboli finora utilizzati, poiché sono stati linearizzati per poter essere scritti come aggiunti (non esiste l'aggiunto di un operatore non lineare). Allo stesso modo $H_t'^o = \frac{dH_t^o[x(t)]}{dx(t)}$ è effettivamente l'operatore osservazione linearizzato. Nel contesto di un gradiente approssimato, dunque, gli operatori H e M agiscono come operatori di *perturbazione di forecast*. Inoltre, non esistendo una soluzione esatta e chiusa del problema, bisogna ricorrere a tecniche di minimizzazione numerica, come la **discesa del gradiente**.

Questa tecnica è iterativa e consiste nel aggiornare l'ipotesi iniziale con un passo di lunghezza finita, ad ogni iterazione, in direzione opposta al gradiente di una data funzione valutata nell'ipotesi, nel caso in esame quella di costo:

$$x^{(k+1)} = x^{(k)} - \alpha \phi(x^k) \quad (2.20)$$

dove α è la lunghezza dello step, k è il numero di iterazione e ϕ è la direzione che dipende da $J(x^k)$ e dal suo gradiente (valutato rispetto alle condizioni iniziali).

A questo punto la scelta di valutare per ogni step il gradiente della funzione di costo, utilizzando la formula 2.19 appare estremamente inefficiente. Ad ogni step infatti, bisognerebbe applicare la sequenza di operatori $M_{\delta t}^{\prime T} \dots M_{t-\delta t}^{\prime T} M_t^{\prime T} H_t'^o dx(t)^T E^{-1}$ al residuo, definito dalla seguente equazione: $r(t) = [y(t) - H_t^o[x(t)]]$. Gli stessi operatori vengono più volte usati in termini diversi della sommatoria.

Per risolvere questo problema di efficienza si ricorre al **modello aggiunto**.

Il modello aggiunto

L'equazione 2.19 può essere riscritta come in figura 2.4.

Ogni riga della somma in figura è un termine nella somma dell'equazione 2.19, con il tempo corrispondente. Le colonne colorate servono ad evidenziare il fatto che specifici operatori di evoluzione temporale vengono ripetuti in righe diverse (gli operatori evidenziati con lo stesso colore sono uguali). Avendo scritto l'equazione in questo modo è facile fattorizzare i termini in comune con un algoritmo che segua questi step:

1. Si fa girare il modello diretto di forecast utilizzando come model state iniziale x uno stato di prova. Questo stato sarà aggiornato con la discesa del gradiente. Ad ogni step si tiene in memoria il model state $x(t)$.
2. Lo step successivo dell'algoritmo parte da $t = \Delta t$ e stima il vettore $\lambda(\Delta t) = H_{\Delta t}^{\prime o T} E^{-1} r(\Delta t)$.¹ Le variabili λ vengono definite *variabili aggiunte*.
3. Si utilizzano le informazioni immagazzinate in memoria nello step 1 dei model state per calcolare i residui e poi si definisce la variabile aggiunta al tempo immediatamente precedente con la seguente relazione:

¹In figura l'operatore H , come M , non appaiono accentati, ma sono i modelli linearizzati

$t = \Delta t$	$-M_{\delta t}^T$	$M_{2\delta t}^T$	\square	$M_{\Delta t + \delta t}^T$	$M_{\Delta t}^T$	$H_{\Delta t}^{o^T} E^{-I} r(\Delta t) +$
$t = \Delta t - \delta t$	$-M_{\delta t}^T$	$M_{2\delta t}^T$	\square	$M_{\Delta t - \delta t}^T$	$H_{\Delta t - \delta t}^{o^T} E^{-I} r(\Delta t - \delta t) +$	
\square	\square	\square	\square	$H_{\Delta t - 2\delta t}^{o^T} E^{-I} r(\Delta t - 2\delta t) +$		
\square	\square	\square	\square	\square		
$t = 2\delta t$	$-M_{\delta t}^T$	$M_{2\delta t}^T$				
$t = \delta t$	$-M_{\delta t}^T$	$H_{\delta t}^{o^T} E^{-I} r(\delta t) +$				
$t = 0$	$-H_0^{o^T} E^{-I} r(0) +$					
b/g	$-B^{-I}(x_B - x)$					

Fig. 2.4: [3]

$$\lambda(t) = H_t^{o^T} E^{-1} r(t) + M_{t+\delta t}^T \lambda(t + \delta t) \quad (2.21)$$

Guardando alla figura 2.4 si nota che il termine $\lambda(t + \delta t)$ contiene tutto quello che è a destra dei termini al tempo t .

4. Facendo girare il modello aggiunto all'indietro nel tempo, iterativamente, fino ad ottenere $t = 0$, si ottiene la variabile aggiunta $\lambda(0)$, che è uguale all'opposto del gradiente di J_o rispetto a x .
5. Si calcola il gradiente completo, includendo il termine di background:

$$\nabla_x \approx -B^{-1}(x_b - x) - \lambda(0) \quad (2.22)$$

Il gradiente ottenuto da questo procedimento (schematizzato in figura 2.5) sarà utilizzato nella tecnica della discesa del gradiente per rendere più accurata l'ipotesi di model state x , dopo ogni iterazione. L'algoritmo si ferma quando si è raggiunti il livello di accuratezza voluto (i.e $\nabla_x \approx 0$).

Il modello aggiunto migliora e rende di fatto pratico l'approccio della 4D-Var.

Nonostante ciò, per fare una singola iterazione del modello aggiunto, c'è bisogno di fare una run completa del modello diretto, ed una del modello aggiunto.

Un ulteriore miglioramento consiste nell'approccio **incrementale**[4]. L'idea dietro l'approccio consiste nel linearizzare il modello intorno all'ipotesi di modello e cercare una perturbazione che ne migliori la stima. Anche questo approccio si basa su più iterazioni fino al raggiungimento di un model state soddisfacente.

Dal punto di vista matematico si tratta di risolvere una serie di problemi linearizzati [5]:

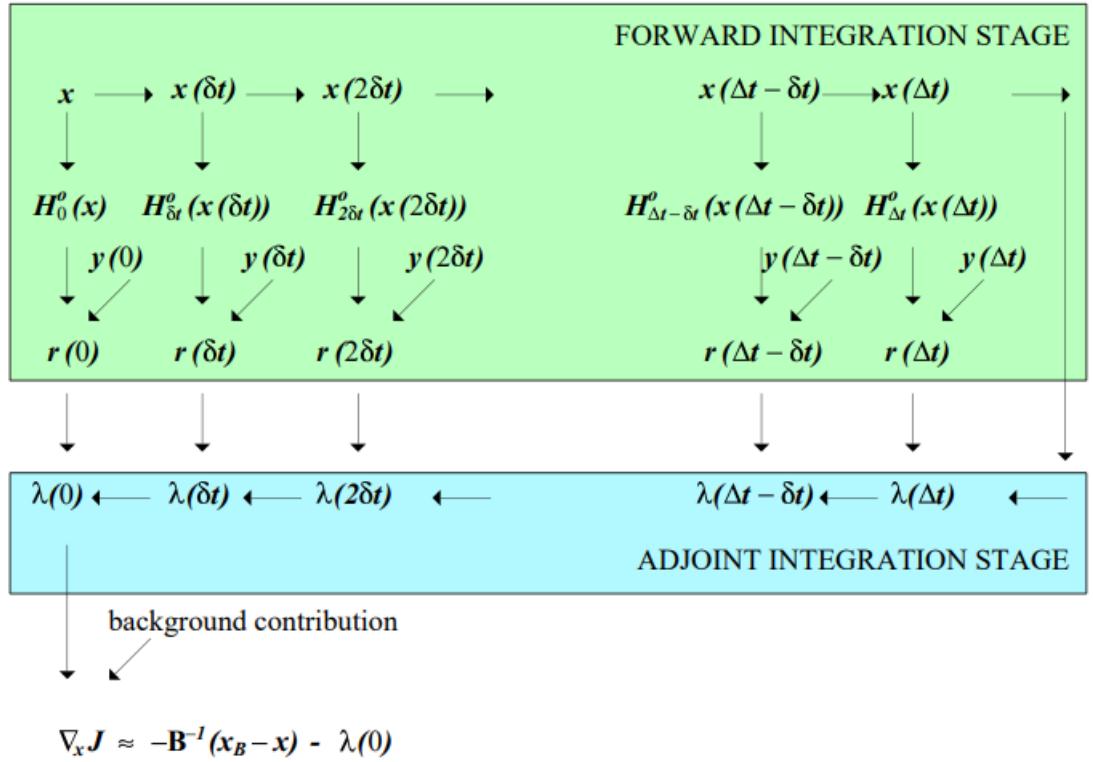


Fig. 2.5: Illustrazione schematica dell’algoritmo 4D-Var aggiunto. Il modello diretto, in verde, è composto dalle stesse parti evidenziate in figura 2.3 con l’aggiunta del calcolo dei residui. La stima delle variabili aggiunte è svolta all’indietro nel tempo (in blu). La combinazione del vettore $\lambda(0)$ con il contributo di background fornisce il gradiente valutato nel model state x , che è l’ipotesi originaria.[3]

$$J^{(k)}(\delta x_o^{(k)}) = \frac{1}{2}[\delta x_o^{(k)} - (x^b - x_o^{(k)})]^T B^{-1}[\delta x_o^{(k)} - (x^b - x_o^{(k)})] + \frac{1}{2} \sum_{t=0}^{\Delta t} [H_t' \delta x(t)^{(k)} - d(t)]^T E_t^{-1} [H_t' \delta x(t)^{(k)} - d(t)] \quad (2.23)$$

con $d(t) = y(t) - H_t[x(t)^{(k)}]$. Le approssimazioni vengono fatte a partire dalle seguenti quantità, espanso in serie di Taylor:

$$\begin{cases} M[x(t) + \delta x(t)] \approx M[x(t)] + M'[\delta x(t)] \\ H[x(t) + \delta x(t)] \approx H[x(t)] + H'[\delta x(t)] \end{cases} \quad (2.24)$$

dove gli operatori accentati sono gli operatori linearizzati. Tale assunzione è tanto più valida, quanto δx è piccolo rispetto al model state x . Una volta trovato il δx si effettua l’update del modello:

$$x(t)^{(k+1)} = x(t)^{(k)} + \delta x(t)^{(k)} \quad (2.25)$$

La soluzione della funzione di costo è unica, poiché è una funzione lineare quadratica. La 4D-Var incrementale ha inoltre il vantaggio, consistendo in un modello approssimato, di poter utilizzare modelli fisici semplificati e a risoluzione più bassa (diminuendo di fatto la dimensionalità degli operatori H e M), riuscendo a raggiungere una buona accuratezza grazie al numero di iterazioni utilizzate.

2.2 Equazione di bilancio delle onde

In questa sezione verrà esposta la fisica alla base dei modelli utilizzati per le onde, partendo col ricavare l'equazione di bilancio della densità di energia e di azione.

L'equazione del tasso di cambio della densità di energia è[6]:

$$\frac{dE(f, \theta; x, y, t)}{dt} = S(f, \theta; x, y, t) \quad (2.26)$$

dove $\frac{dx}{dt} = c_{g,x}$ e $\frac{dy}{dt} = c_{g,y}$ sono le componenti della velocità di gruppo della componente dell'onda presa in considerazione. In acque profonde non sono considerati altri termini, perché la direzione θ e la frequenza f sono costanti. Nelle regioni costiere i termini $\frac{d\theta}{dt} = c_\theta$ e $\frac{df}{dt} = c_f$ devono invece essere considerati. Il termine a destra nell'equazione 2.26 (chiamato termine sorgente) tiene conto di tutti gli effetti generativi, interazione onda-onda, dissipazione e termini non lineari vari.

Un approccio per la risoluzione di questo problema viene detto euleriano e si basa sul computo dello spettro non in un singolo punto, ma su un gran numero di posizioni nel mare simultaneamente, utilizzando l'equazione di bilancio energetico localmente per ogni posizione. Tale approccio è usato in genere nella modellizzazione avanzata delle onde.². La formulazione euleriana tratta il bilancio dell'energia in una griglia regolare, cartesiana o attraverso coordinate di latitudine e longitudine.

Per derivare il bilancio locale si consideri una cella della griglia geografica, utilizzando le coordinate cartesiane. Il bilancio energetico di una cella è, per una componente arbitraria dell'onda, il saldo di quanta energia viene creata localmente più l'importo netto di energia in un intervallo di tempo:

$$\text{variazione di energia nella cella} = \text{importo netto di energia} + \text{generazione locale di energia} \quad (2.27)$$

Il termine a sinistra dell'equazione di bilancio è uguale all'energia nella cella, meno l'energia della cella all'inizio dell'intervallo temporale. Ignorando la dipendenza rispetto

²anche nel modello WAVEWATCH3

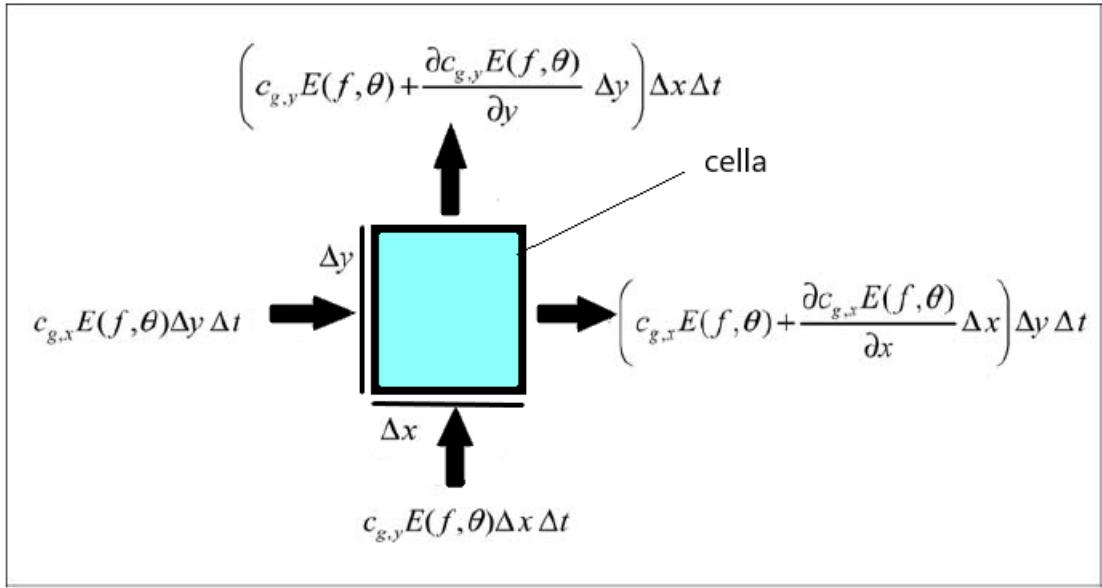


Fig. 2.6: Propagazione dell’energia attraverso una cella nell’approccio euleriano. La figura tiene conto solo dei contributi spaziali i quali, in caso di acque profonde, sono gli unici contributi esistenti.

a x , y e t nella notazione, l’equazione 2.27 può essere riscritta:

$$\begin{aligned} \text{variazione di energia nella cella} &= \left(E(f, \theta) \Delta x \Delta y + \frac{\partial E(f, \theta)}{\partial t} \Delta x \Delta y \Delta t \right) \\ &\quad - E(f, \theta) \Delta x \Delta y \\ &= \frac{\partial E(f, \theta)}{\partial t} \Delta x \Delta y \Delta t \end{aligned} \quad (2.28)$$

Il primo termine a destra dell’equazione è l’importo netto di energia durante l’intervallo Δt . Per quanto riguarda la direzione- x l’importo netto è l’energia che entra dall’estremo sinistro della cella (con velocità di propagazione $c_{g,x}$), meno quello che esce dalla parte destra, con un termine di trasporto dell’energia che evolve nell’intervallo Δx (figura 2.6):

$$\begin{aligned} \text{importo netto di energia nella direzione } x &= c_{g,x}E(f, \theta) \Delta y \Delta t \\ &\quad - \left(c_{g,x}E(f, \theta) \Delta y + \frac{\partial E(f, \theta)}{\partial x} \Delta x \right) \Delta y \Delta t \\ &= - \frac{\partial c_{g,x}E(f, \theta)}{\partial x} \Delta x \Delta y \Delta t \end{aligned} \quad (2.29)$$

Allo stesso modo, l’importo netto dell’energia per la direzione y durante l’intervallo Δx è:

$$\text{importo netto di energia nella direzione } x = - \frac{\partial c_{g,y}E(f, \theta)}{\partial x} \Delta x \Delta y \Delta t \quad (2.30)$$

Il secondo termine nella parte destra dell’equazione di bilancio dell’energia 2.27 rappresenta l’energia generata localmente nella cella, durante l’intervallo Δt :

$$\text{energia generata localmente} = S(f, \theta) \Delta x \Delta y \Delta t \quad (2.31)$$

dove $S(f, \theta)$ è il termine di sorgente, che rappresenta tutti gli effetti di generazione (sorgenti), dissipazione (pozzi), interazioni onda-onda per unità di tempo e di superficie. Sostituendo tutti i contributi trovati nell'equazione 2.27 e dividendo per tutti i fattori comuni si ottiene:

$$\frac{\partial E(f, \theta; x, y, t)}{\partial t} + \frac{\partial c_{g,x} E(f, \theta; x, y, t)}{\partial x} + \frac{\partial c_{g,y} E(f, \theta; x, y, t)}{\partial y} = S(f, \theta; x, y, t) \quad (2.32)$$

l'equazione di bilancio energetico in acque oceaniche o profonde.

Il bilancio di energia spettrale viene formulato con l'approccio euleriano in maniera molto simile per le zone costiere. Le differenze principali sono dovute al fatto che la formulazione deve tenere conto di alcuni effetti che si verificano in acque costiere, inoltre i termini sorgenti sono più numerosi e complessi.

Uno di questi effetti, lo **shoaling**, si ha quando il fondale marino diminuisce di profondità, causando una diminuzione della velocità di gruppo per le onde. Per la conservazione dell'energia la diminuzione della velocità deve essere bilanciata con l'aumento dell'ampiezza delle onde (figura 2.7).

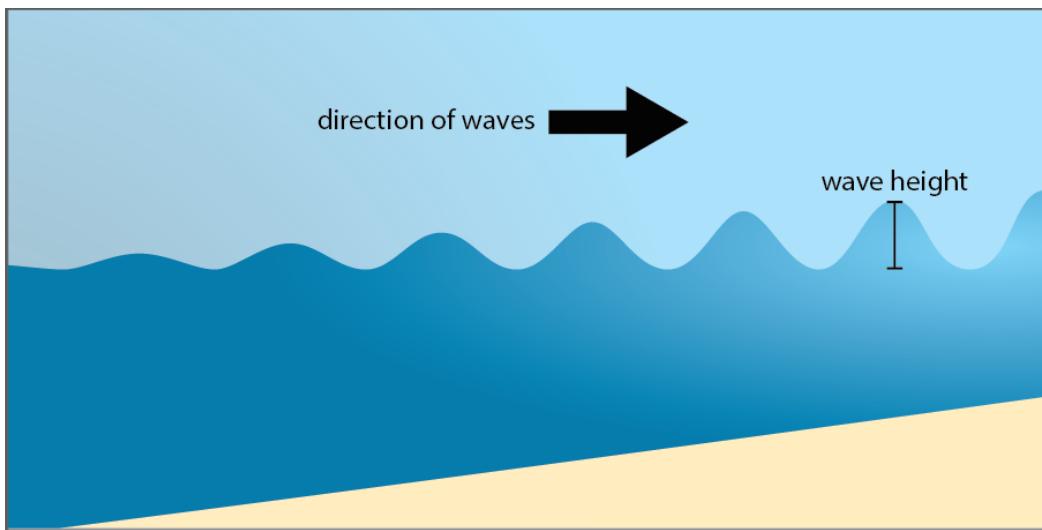


Fig. 2.7: Illustrazione del fenomeno dello shoaling. Le onde che approcciano la costa rallentano e la loro ampiezza aumenta.

Un altro effetto, la **rifrazione**, si ha perché la cresta dell'onda si muove più velocemente in acque profonde rispetto alle acque basse, quindi, a parità di intervallo temporale, la cresta percorre una distanza maggiore in acque profonde, con il risultato di girare verso la regione ad acque basse. Questa è una caratteristica universale delle onde: muoversi verso la regione con una velocità di propagazione più bassa.

Infine, il fenomeno della **diffrazione** si ha quando, ad esempio, un'onda armonica con lunga cresta viaggia attorno ad un promontorio o ad un frangiflutti. In assenza

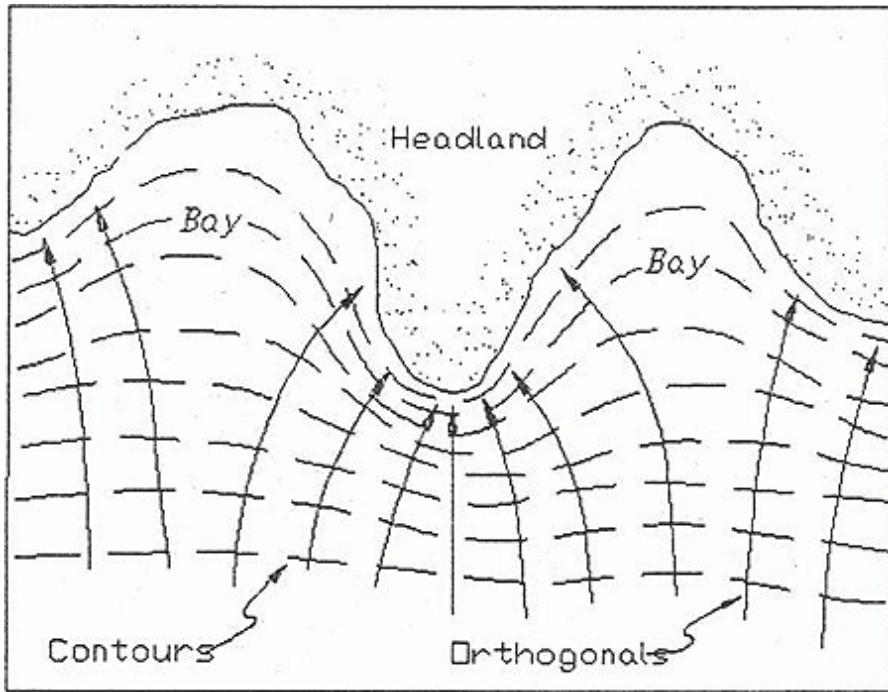


Fig. 2.8: Illustrazione del fenomeno della rifrazione in acque costiere. Le linee tratteggiate sono isolinee, le linee ortogonali rappresentano la direzione di movimento della cresta dell'onda. Il punto in cui vanno ad accumularsi le onde sarà il punto più energetico della costa e sarà soggetto ad una maggiore erosione.

di diffrazione l'onda, incontrando l'ostacolo, dovrebbe proseguire lungo linee rette e non entrerebbe nella parte d'ombra dell'ostacolo. Considerando invece la diffrazione, l'onda viaggerà attorno all'ostacolo formando pattern di creste quasi circolari (figura 2.9). Per il fenomeno della diffrazione la variazione di profondità dell'acqua non gioca un ruolo determinante.

Nell'equazione di bilancio ricavata per le onde oceaniche (2.32) si tiene conto dello shoaling usando velocità di gruppo dipendenti dalla profondità. Per la rifrazione e la diffrazione invece c'è bisogno di aggiungere un termine di propagazione dell'energia. Mentre la densità di energia di una componente dell'onda viaggia nello spazio x, y , essa si propaga anche nello spazio θ , poiché cambia direzione. Dunque, la propagazione avviene in uno spazio tridimensionale x, y, θ . Per una situazione non stazionaria ovviamente bisogna tenere conto anche della componente temporale.

Si procede quindi ragionando analogamente a quanto fatto per lo spazio bidimensionale x, y . Si discretizza l'energia in bin direzionali, ognuno con larghezza $\Delta\theta$. Nel bilancio energetico il cambiamento di direzione delle onde viene rappresentato come densità di energia che si muove da un bin all'altro: l'importo netto di energia è dunque, per un dato bin, l'importo netto che si propaga dal bin a sinistra meno l'importo di energia che esce ed entra nel bin successivo (a destra) nello stesso intervallo di tempo $\Delta\theta$. Questo ragionamento è illustrato in figura 2.10.

Il contributo netto di energia direzionale risulta quindi:

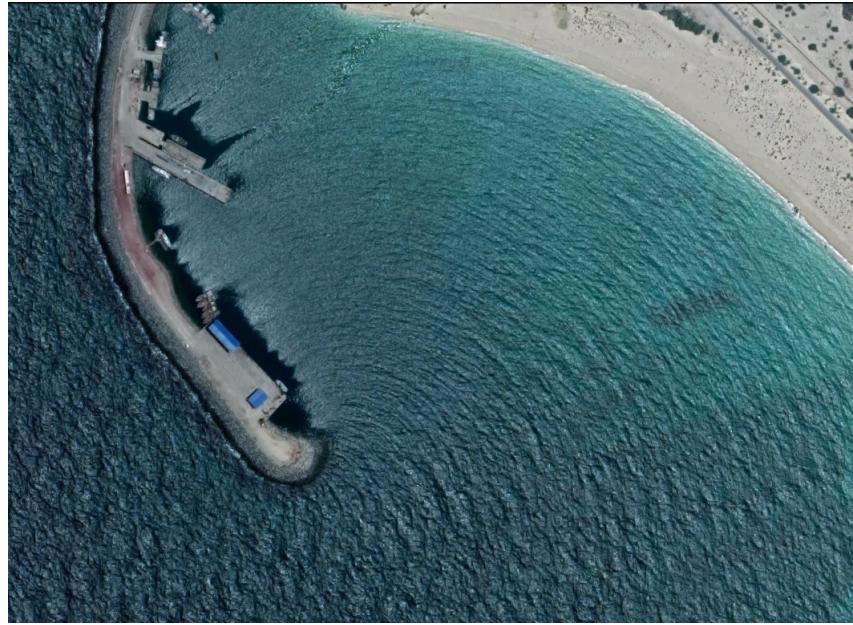


Fig. 2.9: Illustrazione del fenomeno della diffrazione. I raggi dell'onda curvano intorno al frangiflutti.

$$\begin{aligned}
 & \text{importo netto di energia nella direzione } \theta = c_\theta E(f, \theta) \Delta x \Delta y \Delta t \\
 & - \left(c_\theta E(f, \theta) + \frac{\partial E(f, \theta)}{\partial x} \Delta \theta \right) \Delta x \Delta y \Delta t \\
 & = - \frac{\partial c_\theta E(f, \theta)}{\partial x} \Delta \theta \Delta x \Delta y \Delta t
 \end{aligned} \tag{2.33}$$

Aggiungendo questo termine a quelli trovati in precedenza per le acque profonde e dividendo per i fattori comuni, l'equazione 2.32 diventa:

$$\begin{aligned}
 & \frac{\partial E(f, \theta; x, y, t)}{\partial t} + \frac{\partial c_{g,x} E(f, \theta; x, y, t)}{\partial x} + \frac{\partial c_{g,y} E(f, \theta; x, y, t)}{\partial y} \\
 & + \frac{\partial c_\theta E(f, \theta; x, y, t)}{\partial \theta} = S(f, \theta; x, y, t)
 \end{aligned} \tag{2.34}$$

Se sono presenti correnti ambientali il bilancio energetico deve essere corretto con termini che rappresentano energia trasferita fra onde e correnti e gli effetti sulla propagazione delle onde stesse. Alcuni effetti sulla propagazione sono, ad esempio, rifrazione e spostamenti di frequenza relativi all'effetto Doppler. Lo spostamento di frequenza può essere tenuto in conto aggiungendo un termine di propagazione nello spazio delle frequenze. La derivazione può essere compiuta similmente a quanto fatto con il contributo sullo spazio θ . Il trasferimento di energia tra onde e correnti non è facilmente rappresentabile e per questo motivo si utilizza un altro approccio.

L'approccio consiste nel considerare l'equazione di bilancio per l'azione piuttosto che per l'energia delle onde. L'equazione corrispondente per l'azione è identica all'equazione

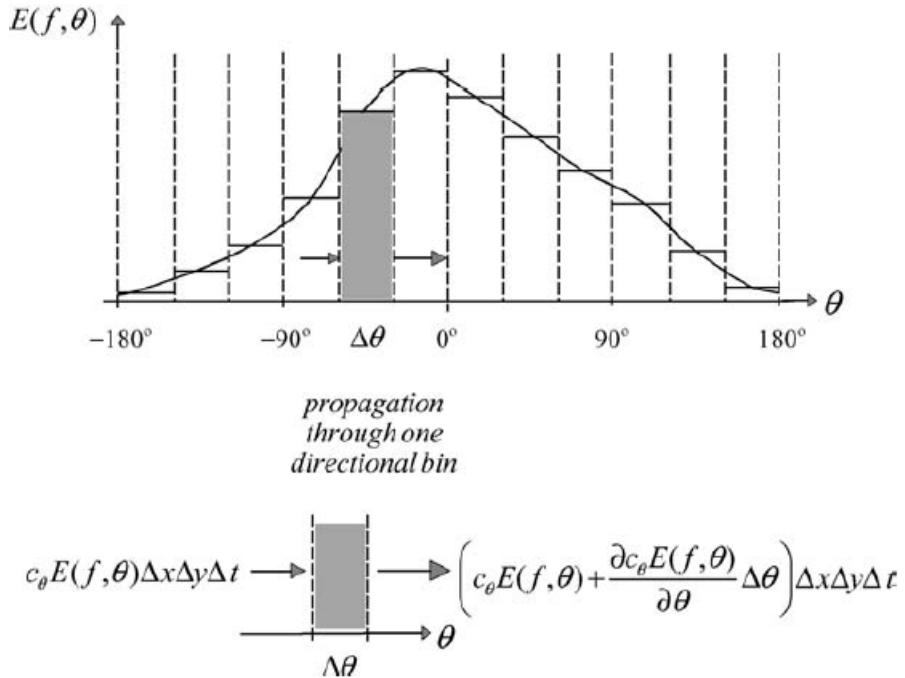


Fig. 2.10: Propagazione dell'energia da un bin direzionale all'altro, nell'approccio euleriano[6].

per la densità di energia $E(\sigma, \theta)$ sostituendo l'azione $N(\sigma, \theta) = \frac{E(\sigma, \theta)}{\sigma}$, dove σ è la frequenza angolare relativa (cioé la frequenza angolare di un sistema in movimento con la corrente). La ragione di questa sostituzione è dovuta al fatto che l'azione, a differenza dell'energia, è conservata in presenza di correnti.

L'equazione di bilancio dell'azione diventa dunque:

$$\begin{aligned} & \frac{\partial N(\sigma, \theta; x, y, t)}{\partial t} + \frac{\partial c_{g,x} N(\sigma, \theta; x, y, t)}{\partial x} + \frac{\partial c_{g,y} N(\sigma, \theta; x, y, t)}{\partial y} \\ & + \frac{\partial c_\theta N(\sigma, \theta; x, y, t)}{\partial \theta} + \frac{\partial c_\sigma N(\sigma, \theta; x, y, t)}{\partial \sigma} = \frac{S(\sigma, \theta; x, y, t)}{\sigma} \end{aligned} \quad (2.35)$$

dove l'ultimo termine al lato sinistro dell'equazione tiene conto delle onde che cambiano frequenza. Questa equazione di bilancio è importante poiché rappresenta la base di partenza di modelli come SWAN[7] e WAVEWATCH 3[8].

Capitolo 3

Downscaling dinamico e modello delle onde

Nei pressi della costa, quando i dati di misurazione diretta non sono sufficienti, vengono utilizzate serie storiche lunghe e affidabili di dati di wave hindcast: è il caso dell'analisi dei valori estremi e delle valutazioni sulla vulnerabilità costiera. Inoltre, le serie storiche fungono da base per studi di processi morfologici dinamici.

I dati di wave hindcast vengono anche utilizzati per la valutazione di disponibilità spaziale e temporale di energia del moto ondoso, allo scopo di sfruttarne le potenzialità.

I modelli di onde possono essere forzati da dati di rianalisi di venti, ma per passare da una risoluzione globale (per ERA5 31 km) ad una risoluzione a scala regionale si rende necessario l'utilizzo di tecniche di nesting e di procedure di downscaling. Tali tecniche sono soprattutto necessarie per la generazione di dati di wave hindcast regionali in aree costiere o bacini chiusi come il Mar Mediterraneo, il Mar Nero e il Mar Baltico.

Esistono due procedure di downscaling, quello dinamico e quello statistico.

Il downscaling statistico deriva relazioni statistiche tra le variabili di piccola scala osservate e le grandezze ottenute dai modelli climatologici globali a grande scala (GCM). I futuri valori delle variabili di grande scala ottenute dalle proiezioni GCM del clima futuro sono quindi utilizzati per guidare le relazioni statistiche e quindi stimare i dettagli su scala ridotta del clima futuro[9].

Il downscaling dinamico invece utilizza un modello climatico regionale utilizza un modello climatico regionale (RCM) con risoluzione spaziale più elevata (o anche più di un modello regionale annidato) su un'area limitata che viene alimentato con condizioni al contorno da modelli climatici globali[10].

Nelle prossime sezioni verranno descritte le procedure di downscaling dinamico e di implementazione del modello WAVEWATCH III svolte nel lavoro (Vannucchi et al. 2021)[11] per ottenere i dati utilizzati in questo lavoro.

3.1 Downscaling dinamico

Per produrre le forzanti di vento è stata implementata una configurazione annidata di due modelli climatici regionali a mesoscala: BOLAM e MOLOCH.

CAPITOLO 3. DOWNSCALING DINAMICO E MODELLO DELLE ONDE WAVEWATCH

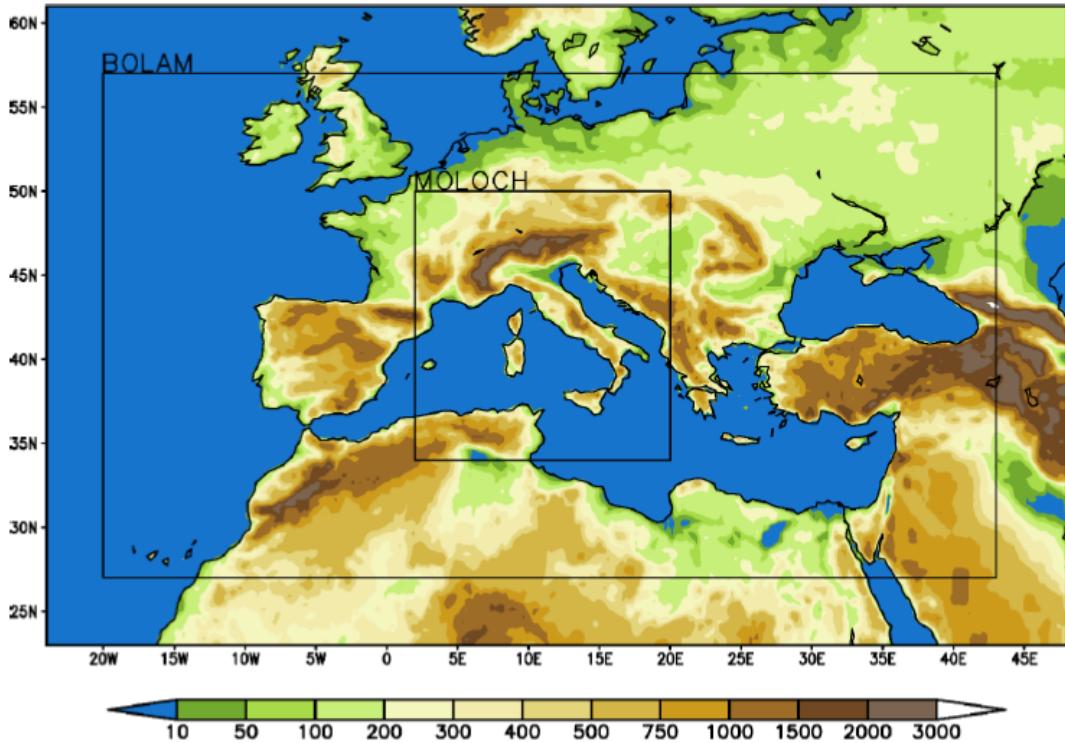


Fig. 3.1: Domini di integrazione dei modelli BOLAM e MOLOCH[11]

BOLAM è un modello idrostatico che integra le equazioni primitive e tratta la convezione in maniera parametrizzata. È stato utilizzato con uno passo di griglia di 7 km per fornire condizioni al contorno al modello MOLOCH.

MOLOCH invece è un modello non idrostatico e tratta la convezione in maniera esplicita (senza la parametrizzazione dei cumuli), ciò lo rende particolare funzionale per fare simulazioni ad altissima definizione. È stato utilizzato con un passo di griglia di 2.5 km.

Il dominio di integrazione dei due modelli è mostrato in figura 3.1.

I dati giornalieri ad alta risoluzione sono stati così prodotti: ogni giorno alle ore 18 UTC viene svolta una simulazione BOLAM utilizzando come condizioni iniziali ERA5 e le condizioni al contorno vengono fornite ogni 6 ore per le successive 30. Gli output orari della simulazione BOLAM forniscono condizioni iniziali e al contorno alla simulazione MOLOCH, che inizia ogni giorno alle ore 21 UTC, con un intervallo temporale di forecast di 27 ore. Il modello MOLOCH produce anch'esso output orari.

I dati giornalieri sono stati prodotti a partire dalle ultime 24 ore delle simulazioni dei due modelli, mentre le prime 6 e 3 ore d'integrazione dei modelli BOLAM e MOLOCH sono state considerate come tempi di spin-up e dunque non prese in considerazione. La rappresentazione schematica del funzionamento dei due modelli annidati è presente in figura 3.2.

I risultati orari di downscaling del vento atmosferico sono stati usati per forzare il modello delle onde WAVEWATCH III (da qui in poi WW3). Per avere la forzante del vento su una griglia non strutturata lungo tutto il Mar Mediterraneo, con la migliore

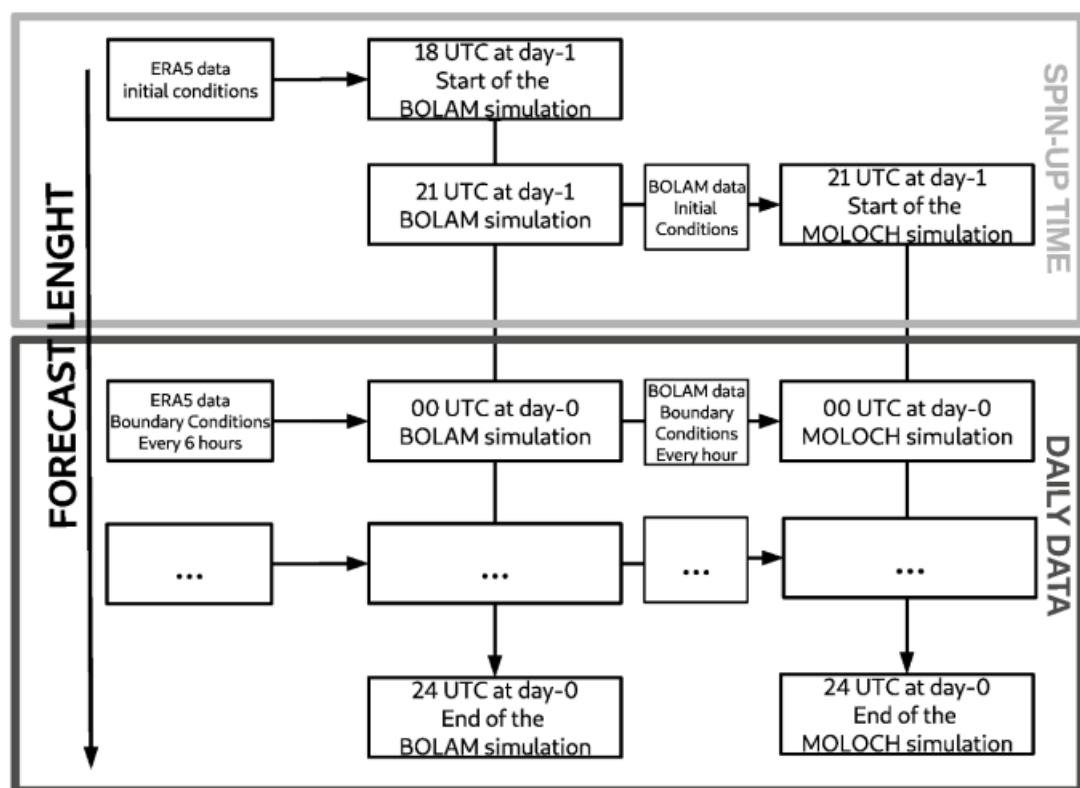


Fig. 3.2: Setup del modello per un singolo giorno di dati[11].

CAPITOLO 3. DOWNSCALING DINAMICO E MODELLO DELLE ONDE WAVEWATCH

risoluzione possibile, i dati di BOLAM e MOLOCH sono stati combinati insieme. Più precisamente, è stata costruita una griglia di 2,5 km su tutto il dominio, tale griglia è stata riempita con dati MOLOCH all'interno ed interpolata con dati BOLAM all'esterno. Per avere una transizione più graduale i dati sono stati mediati usando pesi lineari in un'apposita zona cuscinetto di circa 150 km nei pressi dei confini tra le zone ad alta e bassa risoluzione.

3.2 WAVEWATCH III e setup del modello delle onde

WW3 è un framework che ingloba le conoscenze allo stato dell'arte di modellizzazione delle onde. WW3 è stato sviluppato dall'US National Centers for Environmental prediction ispirandosi al modello WAM [12]. Il modello è utilizzabile in acque profonde e in acque basse. La propagazione di onde può essere risolta utilizzando griglie strutturate (curve o regolari) e non strutturate.

L'equazione di propagazione usata nel capitolo 2 è quella utilizzata da WW3 come equazione di bilancio. L'unica differenza sta nel fatto che l'equazione di bilancio di WW3 è espressa nello spazio k invece che σ . Per passare al nuovo set di variabili basta computare lo Jacobiano[13]:

$$N(\sigma, \theta) = N(k, \theta) \frac{\partial k}{\partial \sigma} = \frac{1}{c_g} N(k, \theta) \quad (3.1)$$

L'equazione finale di bilancio, presa in considerazione nel modello WW3, è dunque:

$$\begin{aligned} & \frac{\partial N(k, \theta; x, y, t)}{\partial t} + \frac{\partial c_{g,x} N(k, \theta; x, y, t)}{\partial x} + \frac{\partial c_{g,y} N(k, \theta; x, y, t)}{\partial y} \\ & + \frac{\partial c_\theta N(k, \theta; x, y, t)}{\partial \theta} + \frac{\partial c_k N(k, \theta; x, y, t)}{\partial k} = \frac{S(k, \theta; x, y, t)}{\sigma} \end{aligned} \quad (3.2)$$

Nel lavoro di Vannucchi et al. 2021 [11] è stata utilizzata una griglia non strutturata lungo tutto il dominio di analisi, il quale comprende l'intero bacino del Mediterraneo e un'area 150 km ad ovest dello stretto di Gibilterra. L'utilizzo di una maglia non strutturata ha permesso di considerare regioni ad altissima risoluzione nelle zone costiere, in modo da ridurre il costo computazionale ed evitare errori di condizioni al contorno[14]. Infatti la griglia non strutturata che è stata utilizzata raggiunge la risoluzione di circa 500 metri lungo le coste Nord-Ovest del mar Mediterraneo.

Per le interazioni non lineari onda-onda è stata usata la DIA (Discrete Interaction Approximation) [15]. Per quanto riguarda il termine di interazione S_{in} tra onde e vento e il termine dissipativo dovuto alle breaking waves S_{ds} è stato utilizzato il pacchetto ST4 descritto nel lavoro Ardhiun et al. [16] aggiornato da Leckler et al. [17]. È stato scelto lo schema numerico esplicito al primo ordine, nonostante causi diffusione in eccesso. La diffusione in eccesso però, nel caso di onde che viaggiano attorno ad isole in acque profonde, compensa l'effetto Garden Sprinkler. L'effetto Garden Sprinkler, infatti, consiste nella propagazione in direzioni discrete che causa una disintegrazione

Calibration Setting	Selected	Evaluated	Constant
Maximum CFL time step for x-y and k-theta	60 s	60–120 s	N scheme, ST4
Explicit numerical scheme	N	N-PSI-FCT	60 s, ST4
Source term parameterization	ST4	ST2-ST3-ST4	60 s, N scheme

Fig. 3.3: Schema riassuntivo della calibrazione del modello[11].

dell’onda, la quale dovrebbe essere continua. La causa è ovviamente una risoluzione spettrale grossolana [18].

L’output del modello di onde è orario e consiste nelle seguenti grandezze: altezza dell’onda significativa Hs_m , periodo dell’onda media Tm , periodo dell’onda di picco Tp , direzione dell’onda media $Dirm$ e direzione dell’onda di picco $Dirp$.

3.3 Calibrazione e validazione del modello

Per validare e calibrare gli output di venti e onde sono state utilizzate osservazioni da varie stazioni di misuramento nella zona Nord-Ovest del mediterraneo. I parametri che sono stati valutati sono: il time step, differenti tipi di schemi numerici (anche superiori al primo ordine) e differenti pacchetti per i termini non lineari sorgenti (figura 3.3). Gli indicatori statistici usati per le grandezze fisiche scelte per la calibrazione del modello (Hs , Tm , $Dirm$) sono stati: MBE (Mean Bias Error), MAE (Mean Absolute Errror), RMSE (Root Mean Square Error), r (coefficiente di correlazione di Pearson), $circ - r$ (coefficiente di correlazione di Pearson circolare).

Una volta calibrato il modello sono stati prodotti l’hindcast (con i modelli BOLAM e MOLOCH) e il wave hindcast dei dati (con il modello WAVEWATCH III). Tali hindcast sono stati confrontati anche con i dati di rianalisi climatica a bassa risoluzione ERA5. Dalla validazione dei dati è emerso che i modelli migliori di forecast sono MOLOCH per i venti e WW3 per le onde.

Capitolo 4

Analisi dei dati

Nel presente capitolo si esporranno i passaggi dell'analisi dei dati di downscaling dinamico a partire dall'acquisizione dei dati stessi.

I dati sono stati generati dal consorzio LAMMA, attraverso il downscaling dinamico descritto nel lavoro *Dynamical Downscaling of ERA5 Data on the North-Western Mediterranean Sea: From Atmosphere to High-Resolution Coastal Wave Climate* (Vannucchi et al. 2021).

Per l'analisi sono stati presi in considerazione le serie storiche di venti e onde al largo di Alghero. La scelta di Alghero è dovuta alla sua particolare posizione nel Mediterraneo, che la rende soggetta a venti forti, specialmente il Maestrale.

Il termine Maestrale si riferisce al forte vento che si sviluppa lungo la valle del Rodano e influenza tutta la costa Sud-Est della Francia estendendosi lungo le Bocche di Bonifacio (tra la Corsica e la Sardegna). Il Maestrale è considerato un vento potenzialmente pericoloso per la sua forza e persistenza, inoltre è in grado di generare onde estreme ha un riconosciuto effetto di raffreddamento della superficie marina [19].

4.1 Rose di onde e venti

Scaricati i dati riguardanti la serie storica di onde, in file *txt*, è stata effettuata la lettura tramite il metodo python *fwf* (fixed-width formatted lines) e sono stati manualmente inseriti i nomi delle colonne, come indicati dal lavoro Vannucchi et al 2021 (figura 4.1).

La lettura trasforma i dati dal formato *txt* in *Dataframe*, un oggetto della libreria *Pandas* ottimo per gestire serie storiche di dati. Il database prescelto per la realizzazione di rose di venti e onde comprende 40 anni di eventi dal 2/01/1979 al 1/01/2019 (compresi). Sono stati esclusi tutti gli altri dati.

I dati di serie storiche sono costituiti da varie grandezze fisiche:

l'altezza dell'onda significativa $Hs(m)$ è definita come quattro volte la radice della varianza nello spettro di densità di energia.

il periodo dell'onda media $Tm(s)$ è la media di tutti i periodi delle onde nell'intervallo temporale di riferimento

il periodo di onda di picco $Tp(s)$ è il periodo dell'onda più energetica nello spettro di densità spettrale di energia

CAPITOLO 4. ANALISI DEI DATI

	YYMMDD	HH	Hs_m	Tm_s	Dirm_N	Tp_s	Dirp_N
0	19790102	0	2.438	6.07	321.0	6.831	321.0
1	19790102	1	3.006	6.86	314.6	7.628	313.6
2	19790102	2	3.444	7.50	315.2	8.453	317.7
3	19790102	3	3.789	7.91	315.4	8.929	317.7
4	19790102	4	4.197	8.25	315.2	9.470	317.1
...
350563	20190101	19	0.617	6.11	299.8	6.689	304.2
350564	20190101	20	0.607	6.02	299.0	6.623	304.2
350565	20190101	21	0.598	5.84	297.6	6.549	304.3
350566	20190101	22	0.602	5.48	294.6	6.439	305.1
350567	20190101	23	0.611	5.12	290.9	6.309	305.3

Fig. 4.1: Dataframe con serie storica di dati sulle onde. Da sinistra a destra: YYMMDD indica la data in formato anno/mese/giorno, HH l'ora in formato 0-23, $Hs(m)$ l'altezza dell'onda significativa, $Tm(s)$ il periodo dell'onda media, $Dirm(N)$ la direzione dell'onda media (partendo da Nord, in senso antiorario), $Tp(s)$ il periodo dell'onda di picco, $Dirp(N)$ la direzione dell'onda di picco (partendo da Nord, in senso orario).

la direzione dell'onda di picco $Dirp(N)$ è la direzione dell'onda più energetica nello spettro

la direzione dell'onda media $Dirm(s)$ è la media delle direzioni delle onde nell'intervalle temporale di riferimento

4.1.1 Rosa delle onde

Per poter costruire rose delle onde è stato necessario trattare prima i dati.

Partendo dal Dataframe originale, è stato costruito un nuovo Dataframe, facendo un *bining* sia per la direzione dell'onda media, sia per l'altezza dell'onda significativa. Per il binning della direzione dell'onda media, l'angolo giro è stato diviso in intervalli di 10° partendo da 0° . Tutte le direzioni multiple perfette di 5° sono state assegnate all'intervallo di riferimento inferiore, applicando un'approssimazione per difetto¹. Si tratta di una convenzione che risolve l'ambiguità nell'assegnazione degli intervalli di riferimento.

L'intensità della Hs_m invece è stata divisa nella seguente lista di intervalli ($< 0.5, 0.5 - 1.5, 1.5 - 2.5, 2.5 - 3.5, 3.5 - 4.5, 4.5 - 5.5, > 5.5$ metri). A questo scopo sono

¹Esempio: 335° è stata inserita nell'intervalllo 325° - 335° , chiamato con il valore centrale 330° , non nell'intervalllo 335° - 345° (340°)

CAPITOLO 4. ANALISI DEI DATI

state applicate alle colonne interessate ($Dirm(N)$ e $Hs(m)$ rispettivamente, figura 4.2 e figura 4.3) due funzioni che sfruttano un ciclo for.

```
def direzione(x):
    direz=[]
    for i in range (0, 360, 10):
        direz.append(i)
    for j in direz:
        if x> 355.01:
            return 0
        if abs(x-j)< 5.01:
            return j
    data["direction"] = data.Dirm_N.apply(direzione)
```

Fig. 4.2: Funzione usata per il binning della direzione dell'onda media sfruttando 2 cicli for. L'ultima riga consiste nella creazione di una nuova colonna (direction) i cui valori corrispondono ai valori sottoposti a binning della colonna $Dirm(N)$.

```
def binning_Hs(x):
    lista=[0.5, 1.5, 2.5, 3.5, 4.5, 5.5]
    for j in range(len(lista)):
        if x< 0.5:
            return "0-0.5"
        if x<lista[j] :
            return str(lista[j-1]) + "-" + str(lista[j])
    return "5.5+"
data["Significant_wave_height"] = data.Hs_m.apply(binning_Hs)
```

Fig. 4.3: Funzione usata per il binning dell'altezza d'onda significativa che impiega un ciclo for e la lista dei valori da cui ricavare gli intervalli. L'ultima riga consiste nella creazione di una nuova colonna (Significant wave height) i cui valori corrispondono ai valori sottoposti a binning della colonna $Hs(m)$.

È stato infine creato un dizionario python, che ha per chiavi la coppia direzione-altezza dell'onda significativa e per valori la frequenza della coppia stessa sul totale degli eventi (figura 4.4).

```
frequencies={}
for s in data.Significant_wave_height.unique():
    for d in data.direction.unique():
        key=(s,d)
        val=round((len(data[data.Significant_wave_height==s][data.direction==d])/len(data))*100, 2)
        frequencies[key]=val
frequencies
```

Fig. 4.4: Algoritmo usato per la creazione del dizionario *frequencies*. I due cicli for concatenati scorrono su tutti gli intervalli di direzione e altezza dell'onda significativa, e con questi ultimi costruiscono delle chiavi sotto forma di tuple. Ad ogni coppia di chiavi si associa la loro frequenza sul totale degli eventi

CAPITOLO 4. ANALISI DEI DATI

Da questo dizionario è stato creato un dataframe con tre sole colonne: altezza dell'onda significativa (intervalli), direzione dell'onda media (intervalli) e frequenza della coppia di intervalli sull'intero dataset (figura 4.6) Con il dataframe così ottenuto si costruisce una rosa delle onde utilizzando la funzione *px.bar polar* della libreria *plotly express* (figura 4.5). La rosa delle onde può essere visualizzata selezionando ogni combinazione di intervalli di altezza dell'onda significativa (figure 4.7 e 4.8).

```
import plotly.express as px
fig = px.bar_polar(original_data, r="frequency", theta="direction",
                    color="Significant_wave_height", template="plotly_dark",
                    color_discrete_sequence= px.colors.sequential.Plasma_r, category_orders=ordine)
fig.show()
```

Fig. 4.5: Codice *plotly express* utilizzato per la realizzazione delle rose delle onde e dei venti. Px.bar polar restituisce una rosa delle onde (o venti) prendendo in input: il dataframe (figura 5.4), la colonna "frequency" per la lunghezza del raggio, la colonna "direction" per l'angolo dei raggi, la colonna "Significant wave height" ("Wind velocity" per quanto riguarda i venti) per determinare il colore dei raggi e un dizionario (category orders) per determinare l'ordine negli intervalli di altezza dell'onda significativa (o velocità del vento).

direction	Significant_wave_height	frequency
320	1.5-2.5	0.92
310	2.5-3.5	2.14
320	2.5-3.5	0.31
320	3.5-4.5	0.11
310	4.5-5.5	0.33
...
230	>5.5	0
240	>5.5	0
110	1.5-2.5	0
170	3.5-4.5	0
210	>5.5	0

Fig. 4.6: Dataframe finale ottenuto dopo il binning e il calcolo delle frequenze per i dati di onde.

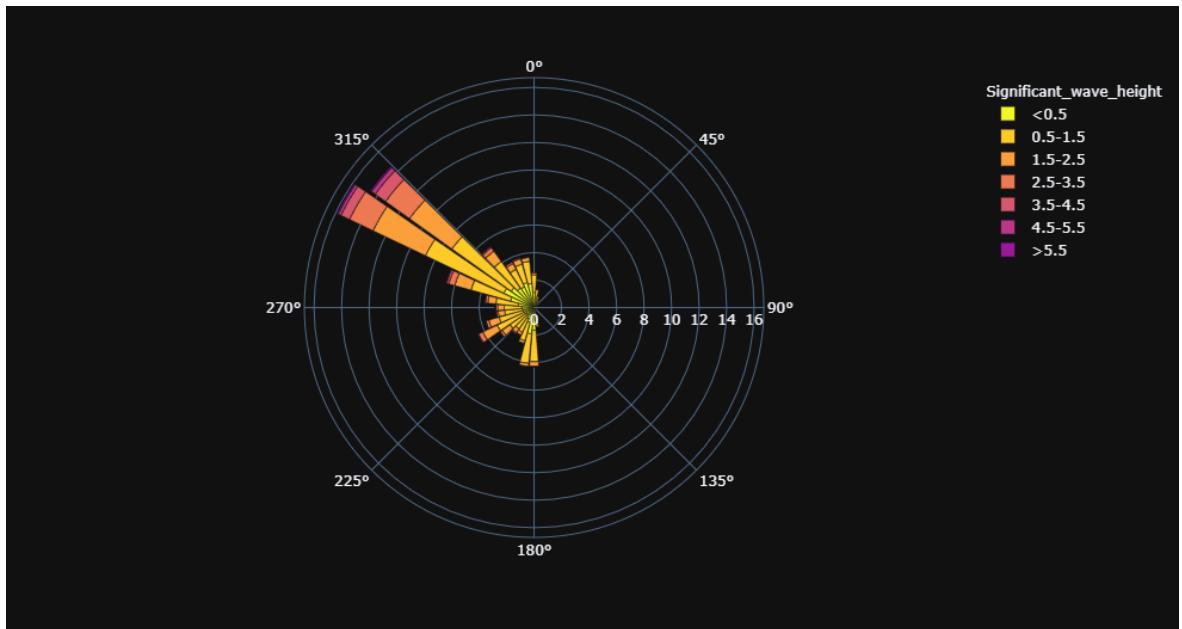


Fig. 4.7: Rosa delle onde ottenuto per l'intero periodo '79-'19

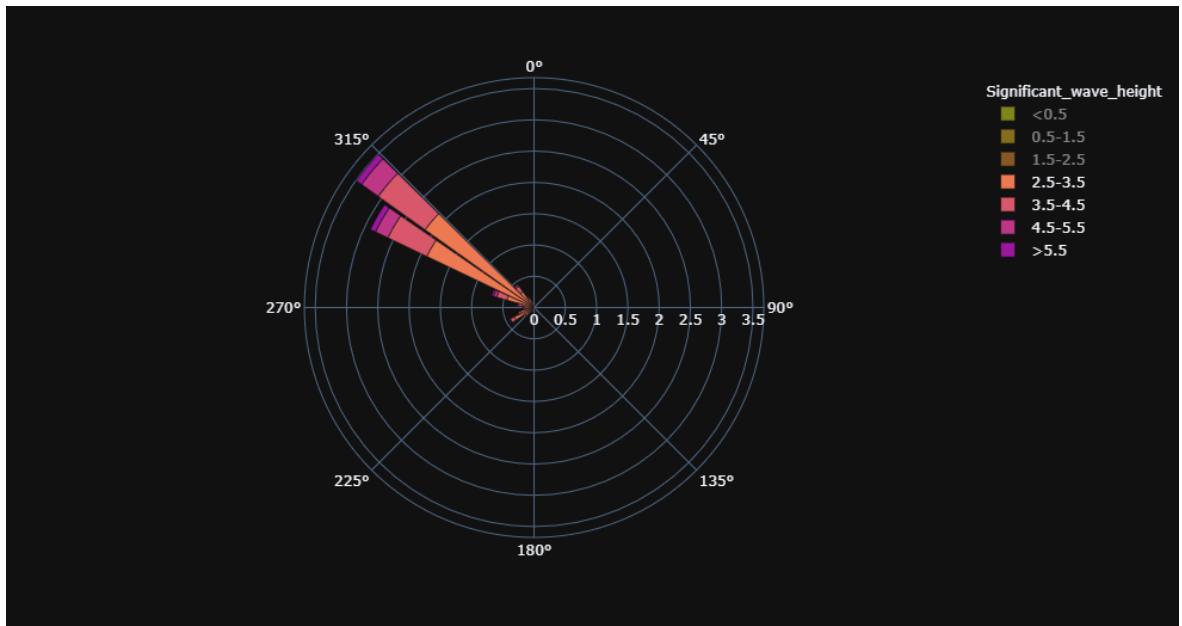


Fig. 4.8: Rosa delle onde per altezze significative superiori a 2.5 metri

4.1.2 Rose dei venti

A questo punto sono stati importati i dati di downscaling dinamico dei venti, in formato *txt*, letti tramite il metodo python *fwf* (fixed-width formatted lines).

Il dataframe così ottenuto (4.9) contiene, oltre alla data e all'ora, la velocità del vento e la direzione. Il binning per la direzione del vento è stato applicato con lo stesso metodo utilizzato per le onde (figura 4.2), mentre per il binning della velocità del vento sono stati cambiati soltanto gli intervalli di riferimento. Tali intervalli sono: $< 2.5, 2.5 - 5, 5 - 10, 10 - 15, > 15$ (m/s).

	YYYYMMDD	Hour	Vel	Dir
0	19790102	0	13.099	321.005
1	19790102	1	12.214	317.918
2	19790102	2	12.606	310.955
3	19790102	3	13.466	310.201
4	19790102	4	14.820	314.361
...
350635	20190101	19	2.522	263.226
350636	20190101	20	3.030	238.598
350637	20190101	21	3.647	244.391
350638	20190101	22	4.811	252.084
350639	20190101	23	4.974	237.272

Fig. 4.9: Dataframe con i dati di venti importati dal file *txt*

È stato poi creato un dizionario python che ha per chiavi la coppia direzione-velocità del vento e per valori la frequenza della coppia stessa sul totale degli eventi, similmente a quanto fatto per la rosa delle onde (figura 4.4).

Dal dizionario è stato ricavato il dataframe con cui realizzare la rosa dei venti (figura 4.10) analogamente a quanto fatto con la rosa delle onde, sia per tutti i tipi di eventi, sia per gli eventi estremi (figure 4.11 e 4.12).

Dalle rose dei venti si evince che a largo di Alghero colpiscono venti principalmente da tre direzioni: il Maestrale da Nord-Ovest, il Grecale da Nord-Est e l'Ostro da sud (confuso spesso con lo Scirocco). Di questi venti il più potente è sicuramente il Maestrale, ed è anche il più presente dal punto di vista quantitativo. Il Maestrale inoltre è quasi l'unica forzante delle onde, a causa della conformazione della costa Nord-Ovest della Sardegna e genera onde significative più alte di 5.5 metri.

direction	Wind_velocity	frequency	
0	320	10-15	0.66
1	310	10-15	1.04
2	310	>15	0.17
3	330	>15	0.05
4	0	10-15	0.15
...
174	20	>15	0.00
175	110	>15	0.00
176	100	>15	0.00
177	90	>15	0.00
178	130	>15	0.00

Fig. 4.10: Dataframe per i venti ottenuto dopo il binning e il calcolo delle frequenze.

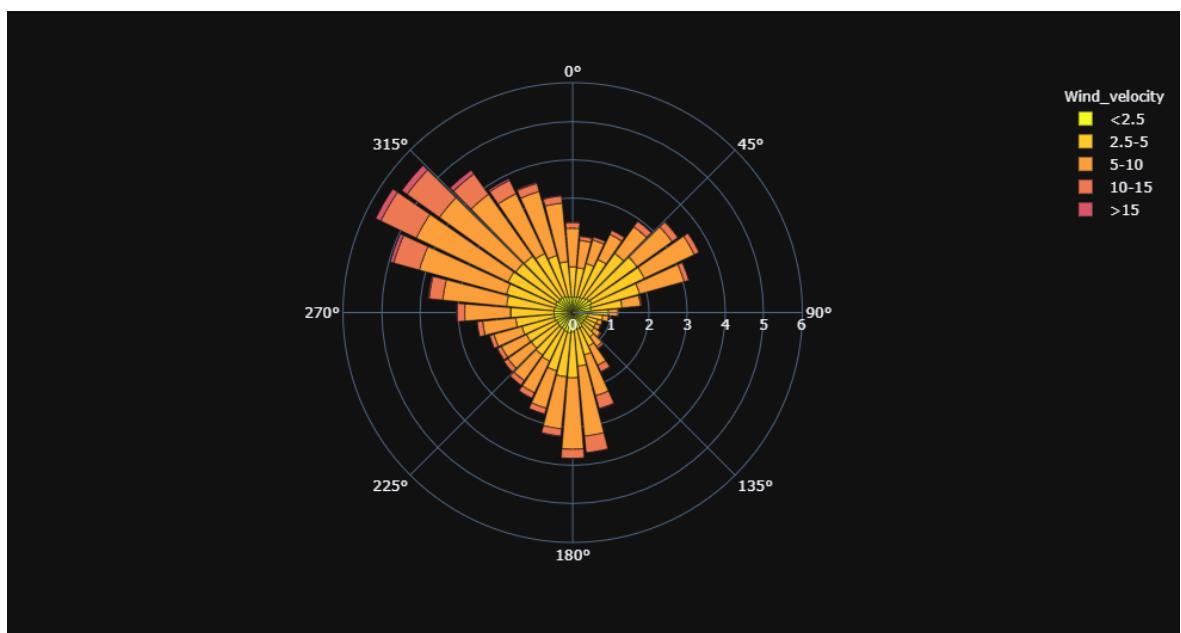


Fig. 4.11: Rosa dei venti ottenuta per l'intero periodo '70-'19

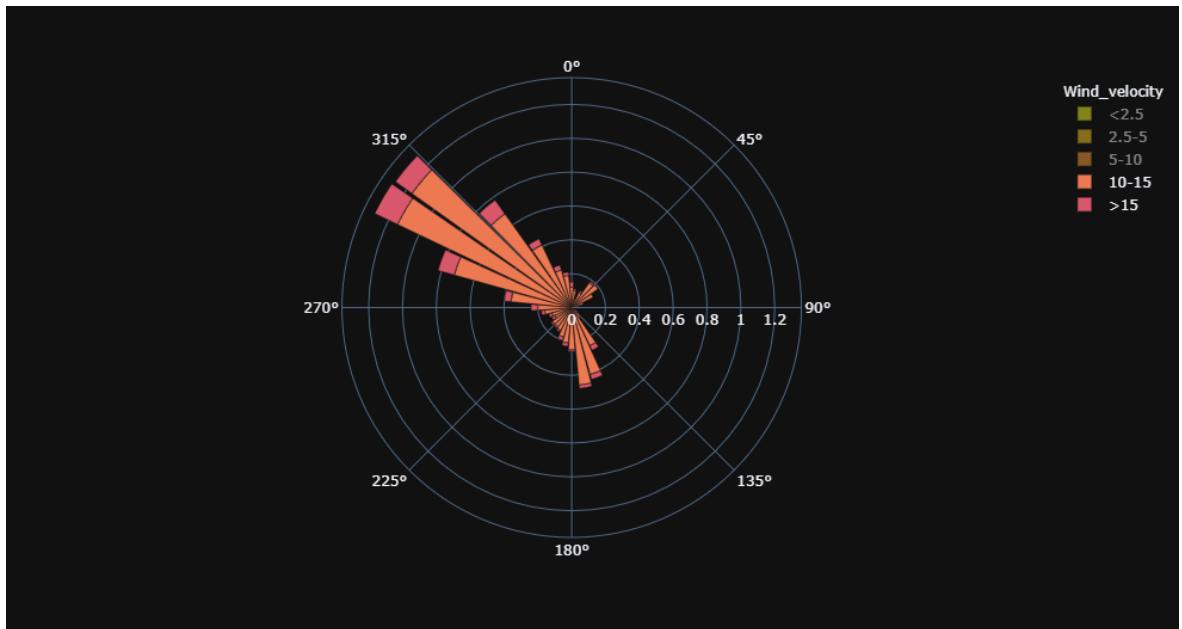


Fig. 4.12: Rosa dei venti per velocità superiori a 10 m/s

		Vel	Dir		Hs_m	Dirm_N	
YYYYMMDD	Hour			YYYYMMDD	Hour		
19790102	0	13.099	321.005	19790102	0	2.438	321.0
	1	12.214	317.918		1	3.006	314.6
	2	12.606	310.955		2	3.444	315.2
	3	13.466	310.201		3	3.789	315.4
	4	14.820	314.361		4	4.197	315.2
...	
20200701	19	0.260	119.255	20200701	20	0.229	315.8
	20	1.876	150.859		21	0.225	307.2
	21	4.025	181.290		22	0.287	227.4
	22	5.000	170.929		23	0.380	195.3
	23	5.464	169.709	20200702	0	0.473	192.6

((a))

((b))

Fig. 4.13: Dataframe con i dati importati di onde (b) e di venti (a). Gli indici sono stati rimpiazzati dai multi-indici formate da coppie data-ora dell'evento e tutte le colonne superflue sono state eliminate

4.2 Scatter plot e coefficiente di correlazione dei ranghi di Spearman

Si è proceduto poi alla realizzazione di un grafico di dispersione, per apprezzare visivamente la relazione tra il vento e le onde che esso genera.

Dunque sono stati importati i dati di vento e onde in due differenti Dataframe. Per i dati di onde sono state cancellate tutte le colonne e tenute in considerazione soltanto quelle di altezza di onda significativa e di direzione dell'onda media. Dato che in entrambi i Dataframe mancano dati in diversi punti, gli stessi indici corrispondono a eventi diversi (data e ora) nei due dataframe. Per ovviare al problema sono stati sostituiti i vecchi indici con dei multi-indici, costituiti dalla coppia data-ora dell'evento per entrambi i dataframe di riferimento (figura 4.13).

Partendo dal dataframe dei venti, sono state create due colonne: $Hs(m)$ e $Dirm(N)$, ed inseriti i valori delle omonime colonne del dataframe delle onde. Il dataframe finale ottenuto contiene dei *Nan* (not-a-number), cioè punti in cui manca almeno un dato. I Nan si creano in concomitanza delle coppie data-ora per cui sono presenti dati di venti e non di onde. Non si tratta di un numero rilevante di dati (meno dell'1% rispetto al totale del numero di eventi nei database iniziali).

A questo punto il dataframe è pronto per la realizzazione del grafico di dispersione di densità tramite l'utilizzo della libreria matplotlib e mpl-scatter-density (figure 4.14 e 4.15).

```
import mpl_scatter_density # adds projection='scatter_density'
from matplotlib.colors import LinearSegmentedColormap

# "Viridis-like" colormap with white background
white_viridis = LinearSegmentedColormap.from_list('white_viridis', [
    (0, '#ffffff'),
    (1e-20, '#440053'),
    (0.2, '#404388'),
    (0.4, '#2a788e'),
    (0.6, '#21a784'),
    (0.8, '#78d151'),
    (1, '#fde624'),
], N=256)

def using_mpl_scatter_density(fig, x, y):
    ax = fig.add_subplot(1, 1, 1, projection='scatter_density')
    density = ax.scatter_density(x, y, cmap=white_viridis)
    fig.colorbar(density, label='Number of points per pixel')
    plt.ylabel('Significant wave height')
    plt.xlabel('Wind velocity')

fig = plt.figure()
fig.set_size_inches(13.5, 7.5)
using_mpl_scatter_density(fig, Dataframe.Vel, Dataframe.Hs_m)
plt.show()
```

Fig. 4.14: Codice usato per le realizzazioni del grafico di dispersione di densità colorato.

Il vantaggio di utilizzare un grafico di dispersione di densità consiste nell'abbattere drasticamente i tempi di computazione, nonché di evidenziare le zone di grafico con una concentrazione di punti maggiore.

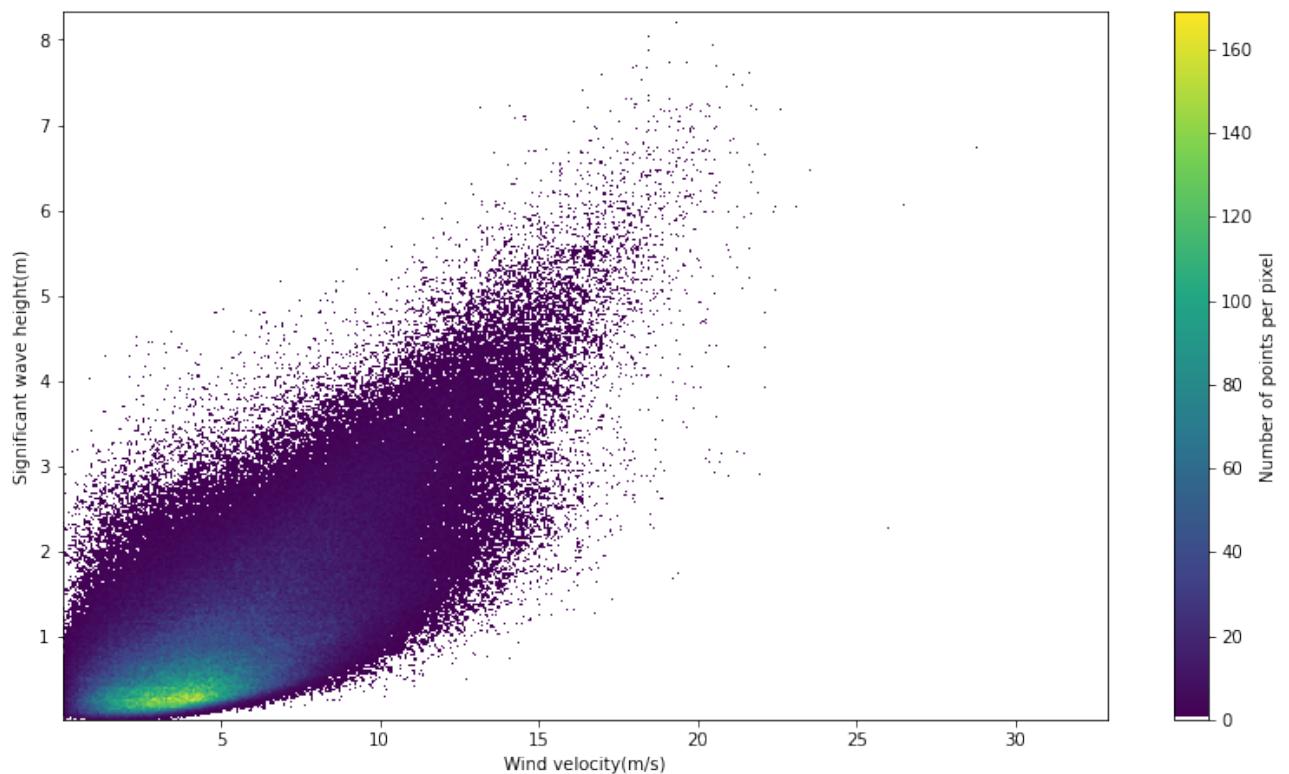


Fig. 4.15: Grafico di dispersione per i dati di velocità di vento e altezza dell’onda significativa. L’altezza dell’onda significativa è rappresentata sull’asse y, la velocità del vento sull’asse x. La legenda a destra del grafico mostra le sfumature di colore nel grafico legate alla densità di punti per pixel. La maggiore concentrazione di punti si ha nella zona ad onde calme.

Coefficiente di correlazione per ranghi di Spearman

Con i dati incrociati (figura 4.16) è stato calcolato il coefficiente di correlazione per ranghi di Spearman, per avere una misura di quanto i dati siano "rumorosi" e di come sono correlati i dati di vento e onde.

		Vel	Dir	Hs_m	Dirm_N
YYYYMMDD	Hour				
19790102	0	13.099	321.005	2.438	321.0
	1	12.214	317.918	3.006	314.6
	2	12.606	310.955	3.444	315.2
	3	13.466	310.201	3.789	315.4
	4	14.820	314.361	4.197	315.2
...
20200701	19	0.260	119.255	0.243	316.2
	20	1.876	150.859	0.229	315.8
	21	4.025	181.290	0.225	307.2
	22	5.000	170.929	0.287	227.4
	23	5.464	169.709	0.380	195.3

Fig. 4.16: Dataframe con i dati incrociati di venti e onde ottenuto a partire dai dataframe in figura 4.13.

Il coefficiente di correlazione per ranghi di Spearman è così definito:

$$r_s = \rho_{R(X), R(Y)} = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}} \quad (4.1)$$

Dove:

- ρ denota il coefficiente di correlazione di Pearson, applicato ai ranghi
- $\text{cov}(R(X), R(Y))$ è la covarianza dei ranghi delle variabili
- $\sigma_{R(X)}$ e $\sigma_{R(Y)}$ sono le deviazioni standard dei ranghi delle variabili

In sostanza, il coefficiente di correlazione di Spearman consiste nel coefficiente di correlazione di Pearson applicato ai ranghi delle variabili. Per definizione, il coefficiente di correlazione di Spearman misura quanto è possibile per una funzione monotona qualsiasi descrivere la dipendenza fra le due variabili e va da -1 (perfetta funzione monotona decrescente) a +1 (perfetta funzione monotona crescente).

Per calcolarlo è stata importata la libreria *scipy*, in particolare il pacchetto *stats* (figura 4.17). Del dataframe in figura 4.16 sono state conservate solo le due colonne *Hs(m)* e *Vel*; a partire da queste è stata creata una lista che ha per valori le coppie

```
lista_venti_onde= data.values.tolist()
Spearman_coeff= scipy.stats.spearmanr(lista_venti_onde)
```

Fig. 4.17: Codice usato per il calcolo del coefficiente di correlazione per ranghi di Spearman

di valori incrociati di altezza dell'onda significativa e velocità del vento. Questa lista è stata inserita come argomento del metodo *scipy.stats.spearmanr* che stima il coefficiente di Spearman e il relativo *p value*.

Nel caso in esame è stato ottenuto un $r_s = 0.66$ con *pvalue* = 0.

4.3 Analisi aggregata per stagione

L'analisi aggregata per stagione, realizzata con dati incrociati di venti e onde, è stata realizzata con l'intenzione di indagare come sono distribuiti gli eventi "calmi" e quelli più "estremi" a seconda della stagione.

Si è partiti dal dataframe incrociato (figura 4.18) che inizia nella primavera del 1979 e termina con l'inverno 2020.

	YYYYMMDD	Hour	Vel	Dir	Hs_m	Dirm_N
1392	19790301	0	13.242	161.441	1.763	186.9
1393	19790301	1	13.769	157.853	1.880	184.7
1394	19790301	2	14.339	156.161	2.005	181.7
1395	19790301	3	15.447	153.492	2.171	178.1
1396	19790301	4	15.551	152.655	2.273	176.5
...
360739	20200229	19	8.879	193.409	0.994	206.7
360740	20200229	20	9.146	186.287	1.112	204.4
360741	20200229	21	9.476	180.611	1.186	199.9
360742	20200229	22	9.439	179.173	1.223	195.3
360743	20200229	23	7.533	191.069	1.145	194.6

Fig. 4.18: Dataframe utilizzato per l'analisi di dati stagionali aggregati. Il database inizia con la primavera del 1979 e finisce con l'inverno del 2020.

Le stagioni sono state considerate nel seguente modo:

la **primavera** costituita dai mesi di marzo, aprile e maggio.

l'**estate** costituita dai mesi di giugno, luglio e agosto.

l'**autunno** costituito dai mesi di settembre, ottobre e novembre.

l'**inverno** costituito dai mesi di dicembre, gennaio e febbraio.

Per dividere il dataframe in stagioni è stata applicata tale procedura:

1. inizialmente sono state selezionate tutte le date di inizio di una stagione tramite la creazione di array (figura 4.19), sfruttando il fatto che le date sono composte da un numero intero nel formato YYYYMMDD multiplo perfetto di numeri diversi per ogni stagione (per la primavera, ad esempio 301). Questo procedimento è stato attuato per ogni stagione.
2. successivamente è stato selezionato un sottodataframe denominato "giorni" contenente l'evento al tempo 0 di tutti i giorni diversi nell'arco dell'intero dataframe originario.
3. partendo dalla lista con i giorni iniziali di tutte le primavere (figura 4.19) è stato utilizzato il metodo *append* per inserire nell'ultima posizione della lista, la lista con i giorni iniziali di tutte le estati. Questo procedimento è stato ripetuto per ogni stagione. In questo modo ogni lista indica una stagione, contenendo come suoi elementi tutti i giorni per anno che aprono la stagione, e come elemento finale una lista contenente tutti i giorni per anno che la chiudono.
4. con le liste così costruite sono stati selezionati dal dataframe originario tutti gli indici che corrispondono agli eventi consecutivi che aprono e chiudono una stagione (figura 4.20)².
5. infine i dataframe delle singole stagioni sono stati formati, sfruttando le ultime quattro liste di indici che sono state create, attraverso l'utilizzo di un ciclo for (figura 4.21).

```

1 inizio_primavera=data.YYYYMMDD[(new_data.YYYYMMDD-301) % 10000 == 0].unique()
2 inizio_primavera

array([19790301, 19800301, 19810301, 19820301, 19830301, 19840301,
       19850301, 19860301, 19870301, 19880301, 19890301, 19900301,
       19910301, 19920301, 19930301, 19940301, 19950301, 19960301,
       19970301, 19980301, 19990301, 20000301, 20010301, 20020301,
       20030301, 20040301, 20050301, 20060301, 20070301, 20080301,
       20090301, 20100301, 20110301, 20120301, 20130301, 20140301,
       20150301, 20160301, 20170301, 20180301, 20190301], dtype=int64)

```

Fig. 4.19: Codice usato per la formazione di un array contenente tutte le date di inizio della primavera.

²Per questi ultimi due passaggi ci sono piccole eccezioni che riguardano la formazione della lista di indici per la stagione invernale. Nel terzo passaggio è stato eliminato il primo elemento della lista "inizio primavera" e aggiunta nell'ultimo posto la data del primo marzo 2020 (assente nel dataframe originario). Nel quarto passaggio è stato utilizzato il dataframe "giorni" modificato con l'aggiunta dell'evento del primo marzo 2020, ore 00, con lo scopo di ottenere anche l'indice di fine dell'ultimo inverno.

CAPITOLO 4. ANALISI DEI DATI

```
indici_primavera=[]
for i in range(0,41):
    indici_primavera.append(giorni[giorni.YYYYMMDD == inizio_primavera[i]].index)
    indici_primavera.append(giorni[giorni.YYYYMMDD == inizio_primavera[41][i]].index)
```

Fig. 4.20: Codice usato per la creazione di una lista di indici che corrispondono nel dataframe originale agli eventi consecutivi che iniziano e terminano una stagione.

```
primavera_df=pd.DataFrame(columns=["YYYYMMDD", "Hour", "Vel", "Dir", "Hs_m", "Dirm_N"])
for i in range(0,82,2):
    primavera_df=primavera_df.append(data[(data.index > indici_primavera[i][0]) & (data.index < indici_primavera[i+1][0])])
primavera_df
```

Fig. 4.21: Codice usato per creazione del dataframe contenente tutti gli eventi primaverili (così come definiti in questo paragrafo) contenuti nel dataframe incrociato (figura 4.18), dal periodo della primavera del 1979, all'inverno 2020. Il dataframe si forma facendo l'*append* di pezzi di dataframe originario, definiti a partire dagli indici di inizio e fine, per ogni anno, della stagione primaverile.

Con i dataframe così ottenuti sono stati realizzati grafici di distribuzione in frequenza in scala logaritmica di tutte le stagioni, sia per i dati di vento, che per quelli di onde. Sono stati anche confrontati in un unico grafico sia di distribuzione in frequenza in scala logaritmica, sia come distribuzioni cumulativa (figure 4.22 e 4.23).

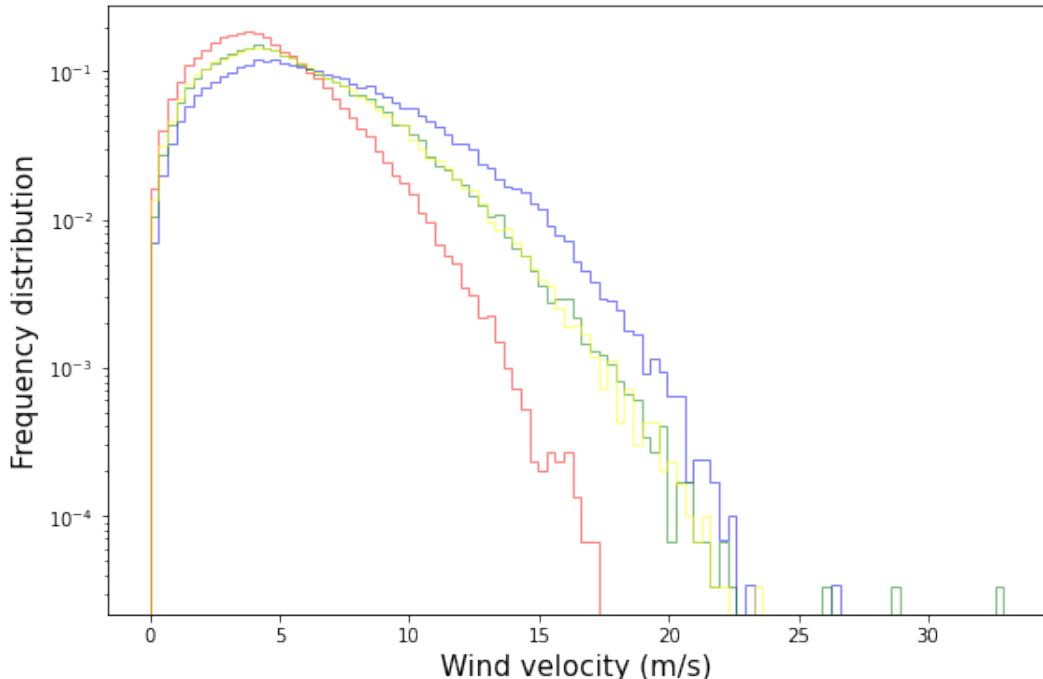


Fig. 4.22: Distribuzione in frequenza in scala logaritmica dei dati relativi alla velocità del vento per le quattro stagioni. In giallo la primavera, in rosso l'estate, in verde l'autunno, in blu l'inverno.

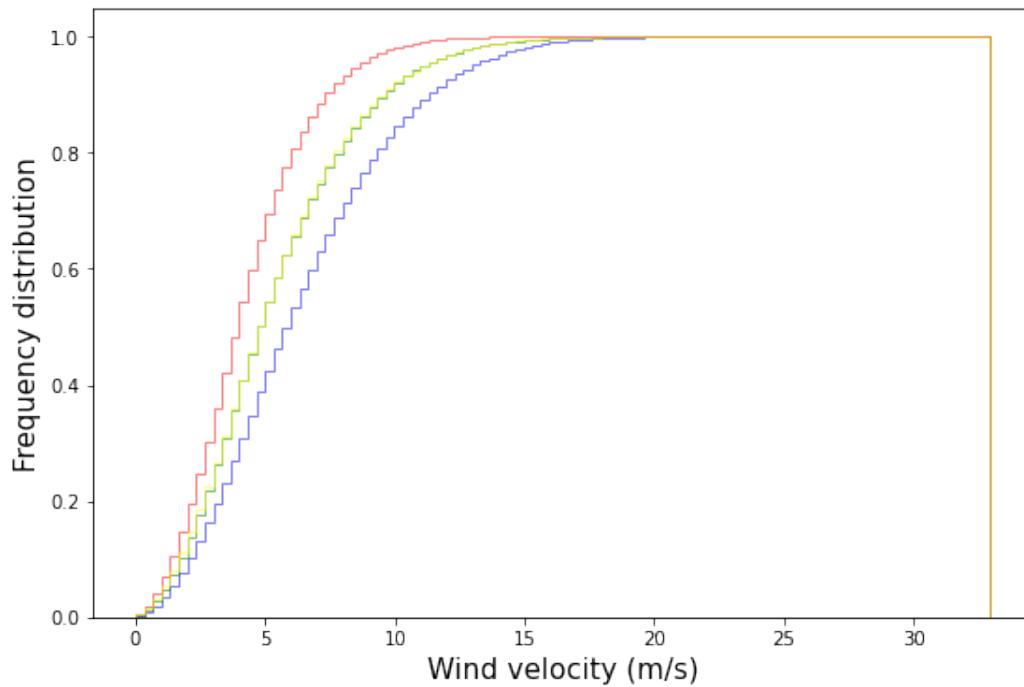


Fig. 4.23: Distribuzione cumulativa dei dati relativi alla velocità del vento per le quattro stagioni.

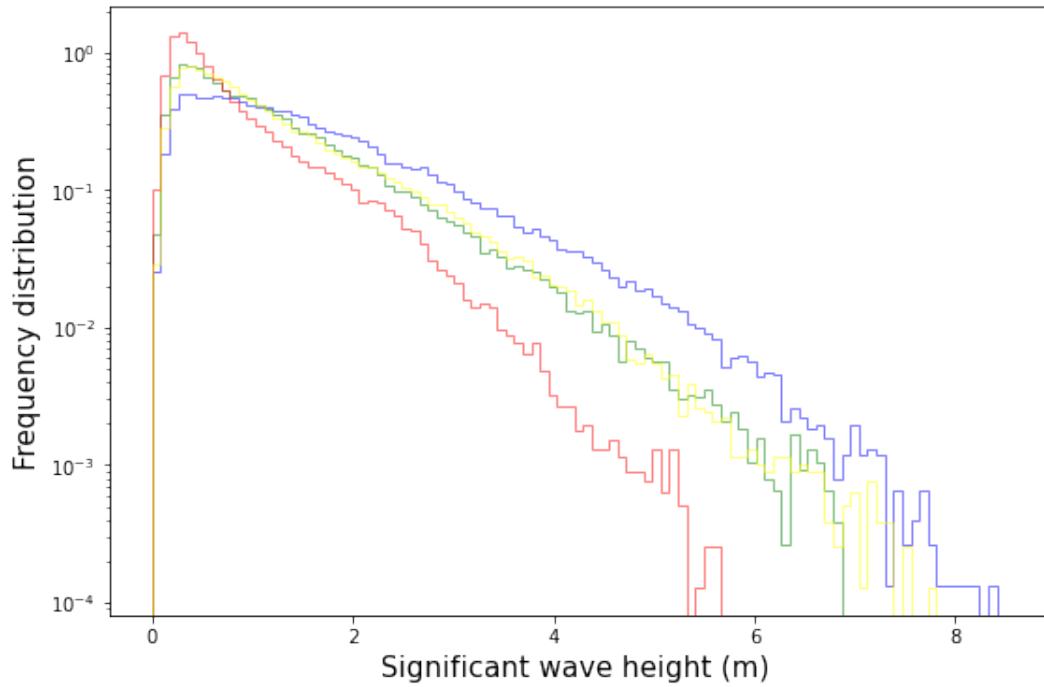


Fig. 4.24: Distribuzione in frequenza in scala logaritmica dei dati relativi alla lunghezza dell'onda significativa per le quattro stagioni.

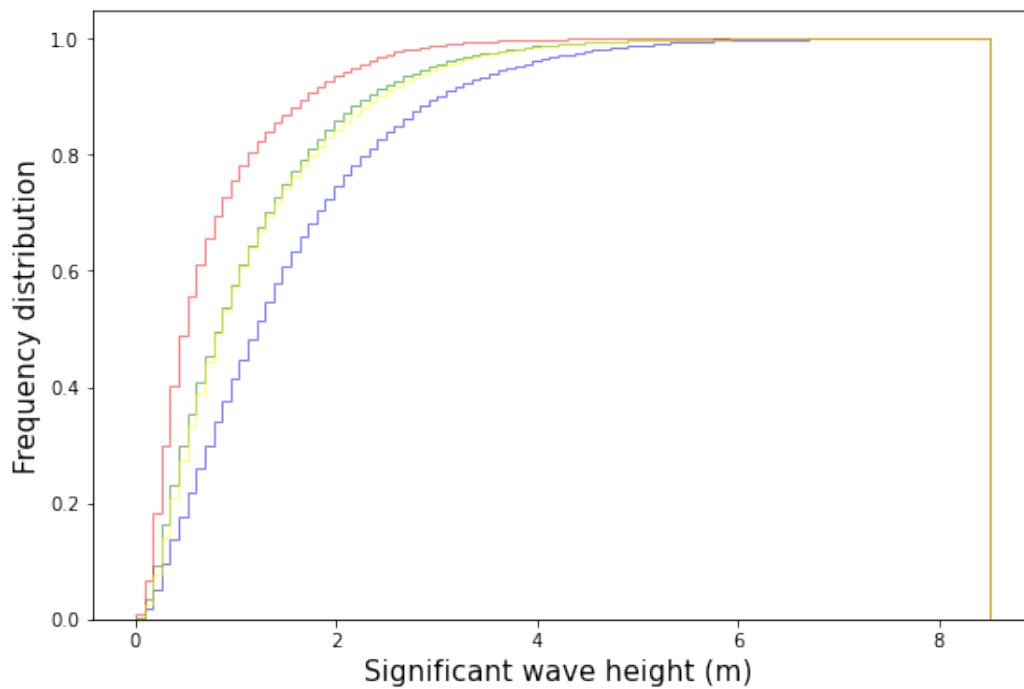


Fig. 4.25: Distribuzione cumulativa dei dati relativi alla lunghezza dell'onda significativa per le quattro stagioni.

Quello che risulta è la netta prevalenza della frequenza di venti deboli in estate e di venti forti in inverno. La primavera e l'autunno presentano distribuzioni molto simili sia per i dati di venti che per quelli di onde, con differenze per eventi estremi non statisticamente rilevanti.

4.4 Fit dei dati di venti e di onde

In letteratura scientifica la distribuzione lognormale e la distribuzione di Weibull sono comunemente considerate le migliori per fittare dati di velocità dei venti e di altezza di onda significativa, sia per quanto riguarda dati osservazionali puri, sia per quanto riguarda quelli di hind/wave cast.

La distribuzione Weibull a due parametri è riconosciuta come un modello appropriato e il più utilizzato nel settore dell'industria eolica. La sua forma generalizzata, a quattro parametri è:

$$p(\nu) = \alpha \frac{k}{\lambda} \left(\frac{\nu - \nu_0}{\lambda} \right)^{k-1} [1 - e^{-[(\nu - \nu_0)/\lambda]^k}]^{\alpha-1} e^{-[(\nu - \nu_0)/\lambda]^k} \quad (4.2)$$

ed è chiamata la funzione di Weibull esponenziata. La formula 4.2 si riduce alla formula di Weibull a tre parametri se $\alpha = 1$, a due parametri se anche $x_0 = 0$. I fattori adimensionali k e α sono rispettivamente l'exponentiation parameter e lo shape factor e determinano la forma della curva. Il parametro λ in m/s, è il fattore di scala, ν_0 è il location parameter (cioè il valore di aspettazione della distribuzione).

In alcuni casi, in cui il la probabilità di venti nulli è significativa, la distribuzione di Weibull a due parametri non si dimostra la più appropriata. Wais nella sua review [20] nota che per le stazioni con la media più bassa di velocità del vento (inferiore a 2 m/s), la distribuzione lognormale si adatta meglio ai dati rispetto alla distribuzione Weibull.

Avere una descrizione funzionale della distribuzione dei venti è importante per l'industria eolica perché permette il calcolo della potenza disponibile generata dal vento con il seguente integrale:

$$P_{disponibile} = \int_0^\infty P_\nu p(\nu) d\nu \quad (4.3)$$

dove P_ν è la potenza generata dal flusso d'aria con velocità ν .

In generale, la distribuzione che meglio approssima i dati può dipendere da vari fattori, principalmente la posizione geografica in cui i dati sono stati ottenuti (osservati o di scaling).

In questo lavoro è stato cercato il modello che meglio può fittare i dati a disposizione tra la distribuzione lognormale e Weibull a quattro, tre e due parametri. Per procedere al fit sono state utilizzati i metodi *.fit* delle rispettive funzioni, presenti nella libreria python *scipy*. Questi metodi utilizzano la MLE (Maximum Likelihood Estimation) per le loro stime di fit.

Il binning dei dati è stato effettuato tramite la libreria *binned statistic* di *scipy.stat*. La statistica “mean” (che computa la media di x dei punti di ogni bin) è stata utilizzata per ottenere i punti sugli assi x, nel caso di mancanza di punti è stato sostituito il punto centrale del bin. La statistica “count” è stata utilizzata per stimare la frequenza della distribuzione per ogni bin.

I fit ottenuti sono in figura 4.26 e 4.28. Per ogni funzione sono stati stimati il coefficiente di determinazione R^2 e l'errore standard del fit SE . Lo SE del fit può essere trovato con la seguente formula:

$$SE = \sqrt{\frac{SSR}{n - 2}} \quad (4.4)$$

dove **SSR** sta per la somma quadratica dei residui. I risultati ottenuti sono mostrati in figura 4.29 e 4.27.

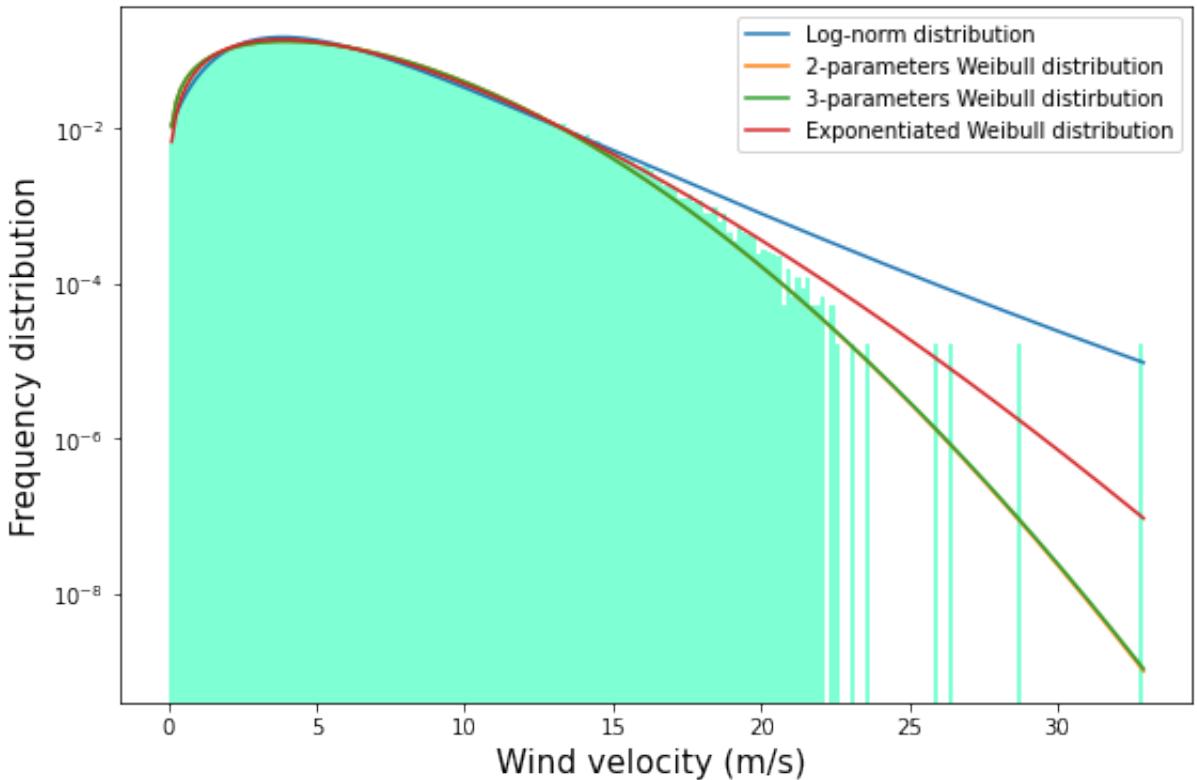


Fig. 4.26: Fit dei dati in scala semilogaritmica di velocità del vento, sfruttando la MLE, delle quattro distribuzioni: lognormale, Weibull esponenziata, Weibull a due e a tre parametri.

	e_Weibull	Weibull_2p	Weibull_3p	Lognormal
r2	0.997	0.991	0.991	0.996
Standard_error	0.003	0.004	0.004	0.002

Fig. 4.27: Tabella con i risultati di coefficiente di determinazione e di errore standard del fit sui dati di vento, per tipo di distribuzione usata.

I dati in esame non mostrano significative differenze tra il fit con la distribuzione di Weibull a due parametri o a tre parametri (i.e. $\nu_0 \approx 0$ nel caso di Weibull a tre parametri) ed entrambe sottoperfornano le altre distribuzioni usate.

Per il fit dei dati di vento i valori di R^2 sono molto alti, specialmente per Weibull esponenziato e per la distribuzione lognormale (superiori entrambi al 95%).

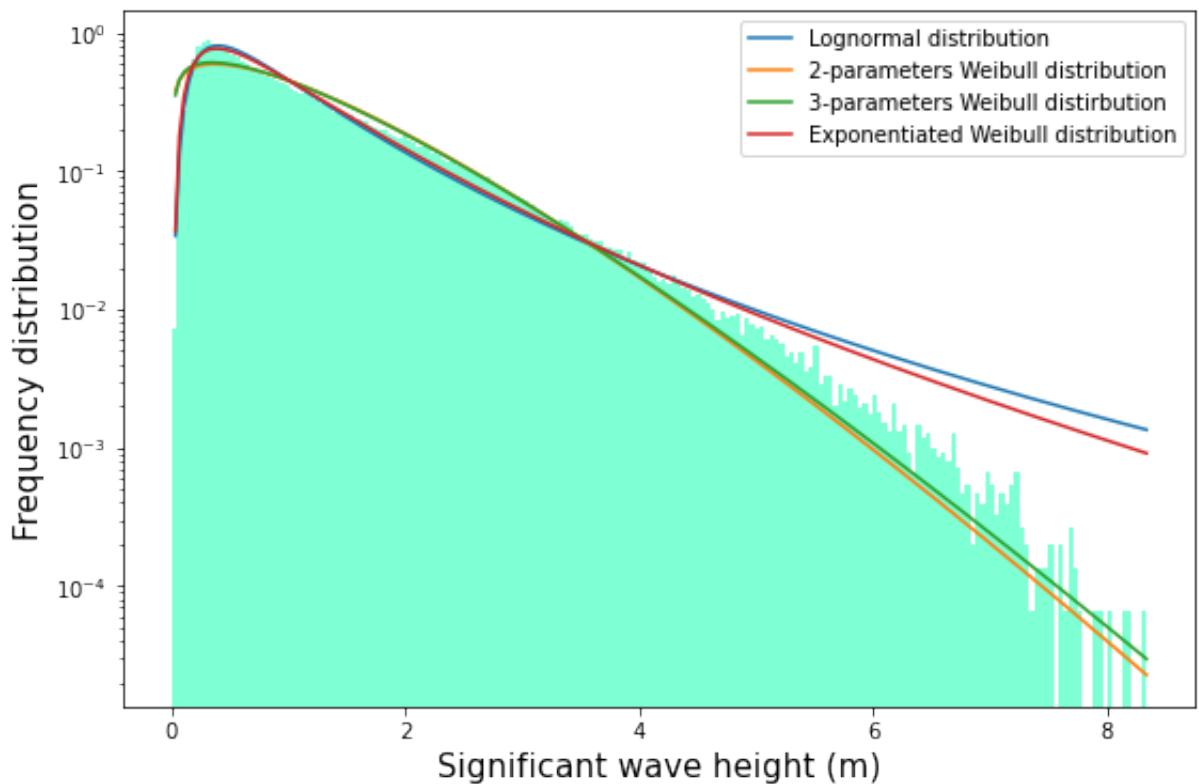


Fig. 4.28: Fit dei dati in scala semilogaritmica di altezza dell'onda significativa, sfruttando la MLE, delle quattro distribuzioni: lognormale, Weibull esponenziata, Weibull a due e a tre parametri.

	e_Weibull	Weibull_2p	Weibull_3p	Lognormal
r2	0.982	0.904	0.912	0.98
Standard_error	0.03	0.06	0.06	0.03

Fig. 4.29: Tabella con i risultati di coefficiente di determinazione e di errore standard del fit sui dati di onde, per tipo di distribuzione usata.

La distribuzione di Weibull esponenziata mostra risultati migliori per entrambi i tipi di dati e per i dati di venti, rispetto alla distribuzione lognormale, performa meglio sugli eventi estremi e peggio su quelli calmi. La migliore performance è certamente dovuta al fatto che Weibull esponenziata utilizza quattro parametri liberi per il fit, contro i tre della distribuzione lognormale.

4.5 Analisi per anno

Infine il dataset completo è stato diviso per anno. Questo tipo di analisi è volto a scoprire se ci sono state variazioni nell'intensità dei fenomeni ondosi e ventosi nel corso del tempo a causa del cambiamento climatico.

Il dataset è stato diviso per anno semplicemente selezionando le parti del dataframe originale per data, come in figura 4.30.

```
data80=data[(data.YYYYMMDD<19810101) & (data.YYYYMMDD>19791231)].copy()
```

Fig. 4.30: Codice usato per la creazione di dataframe divisi per anno. Nell'esempio il dataframe con tutti i dati relativi all'anno 1980.

Per ogni anno sono state selezionate le medie dei venti, delle onde, e il novantacinquesimo percentile sfruttando i metodi numpy “mean” e “percentile”. Poi sono stati eseguiti i plot (figure 4.31 e 4.32).

I dati di scaling analizzati non mostrano chiari trend nell'arco dei 40 anni analizzati, né per l'intensità degli eventi medi (velocità del vento media e altezza dell'onda significativa media), né per come sono distribuiti gli eventi estremi (nella fattispecie il novantacinquesimo percentile).

Se ci sono stati cambiamenti nella frequenza di eventi estremi nella zona di Alghero, essi devono necessariamente riguardare eventi estremi con frequenze molto rare, tali da non essere individuabili con un'analisi di questo tipo [21].

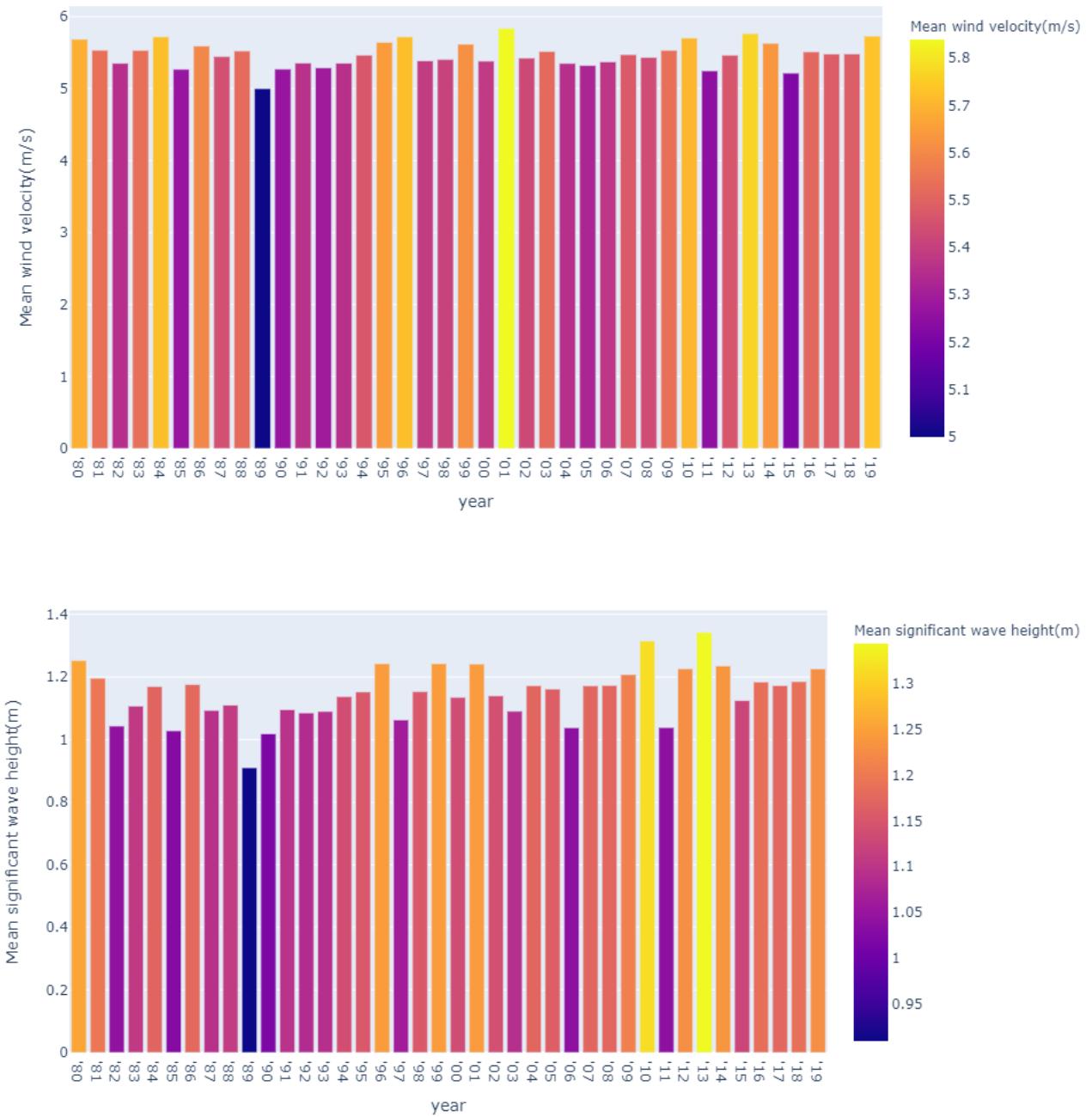


Fig. 4.31: Istogrammi relativi al valore medio della velocità del vento (in alto) e dell'altezza dell'onda significativa per anno (in basso). Il colore delle barre parte dal blu per valori inferiori al valor medio, fino ad arrivare al giallo per valori superiori al valor medio.

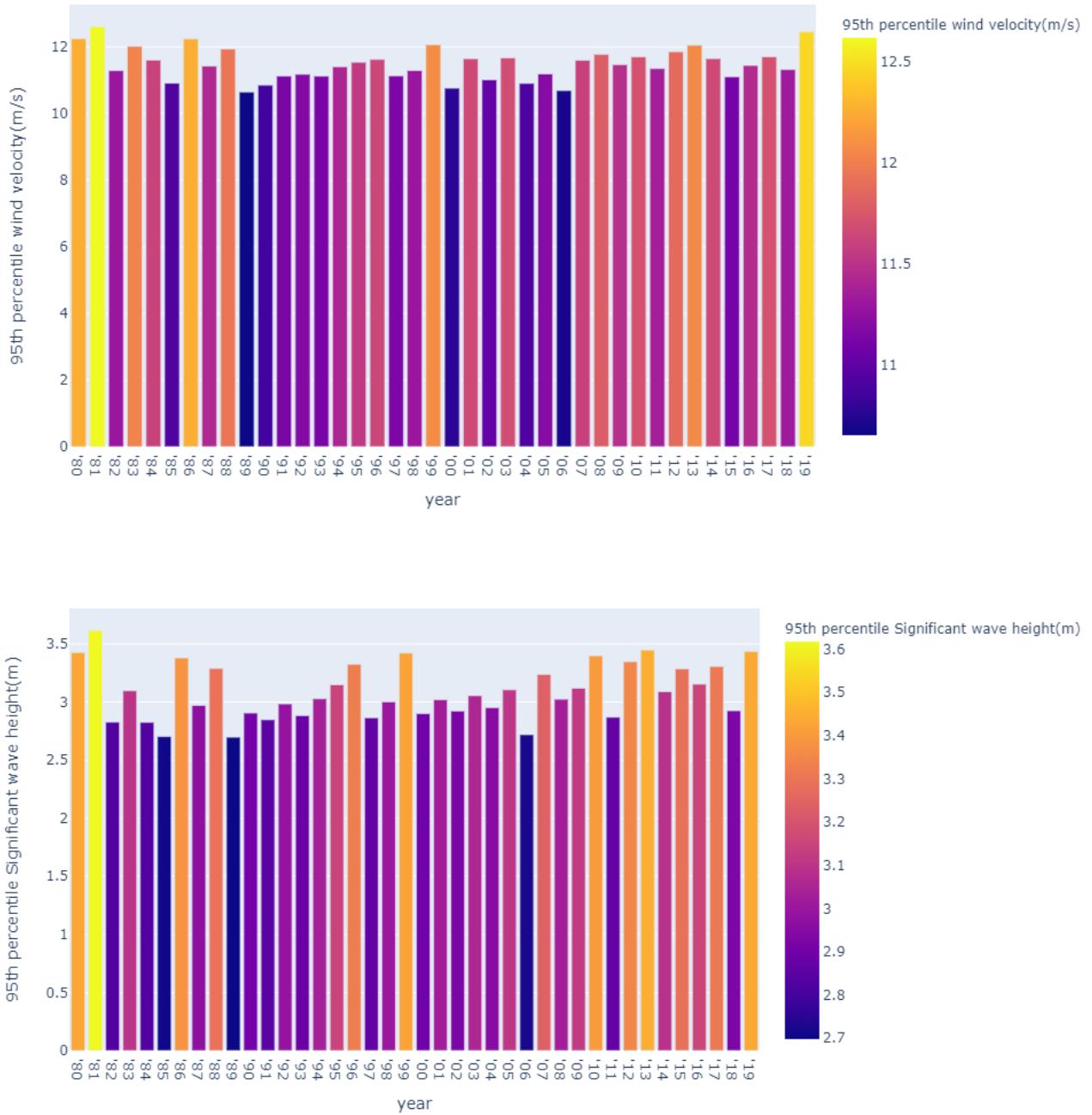


Fig. 4.32: Istogrammi relativi al novantacinquesimo percentile per la velocità del vento (in alto) e per l'altezza dell'onda significativa per anno (in basso). Il colore delle barre parte dal blu per valori inferiori al valor medio, fino ad arrivare al giallo per valori superiori al valor medio.

Capitolo 5

Conclusioni

La rianalisi climatica è un mezzo importante per comprendere il clima del passato e per formulare ipotesi per il futuro. Il cuore della rianalisi climatica risiede nell'assimilazione dei dati, una tecnica che ci permette di integrare dati osservazionali, seppur scarsi, con modelli predittivi per riuscire ad avere un quadro quanto più preciso possibile del clima del passato e dei trend in atto. Attualmente vi sono ricercatori che dedicano gran parte della loro carriera allo sviluppo di tecniche di risoluzione approssimate al problema di assimilazione dei dati sempre più efficienti dal punto di vista computazionale e precise.

Il cambiamento climatico ha reso ancora più evidente la necessità di avere serie storiche affidabili e chiare per poter predire in quali zone geografiche e con quali frequenze si verificheranno eventi estremi nel futuro, in modo da poter prendere le adeguate contromisure.

Nel presente studio sono stati utilizzati i dati di downscaling dinamico compiuto nel lavoro di Vannucchi et al. 2021, poiché più accurati sia per i venti che per le onde rispetto ai dati di rianalisi disponibili con ERA5.

La zona al largo di Alghero è stata scelta per la sua interessante posizione, che la rende soggetta al Maestrale, il quale nasce nella valle del Rodano in Francia.

L'analisi ha mostrato che il Maestrale non è l'unico vento presente nella zona al largo di Alghero, ma è sicuramente il più presente dal punto di vista quantitativo ed il più forte dal punto di vista dell'intensità. Esso è inoltre l'unico responsabile per la generazione di onde con altezze dell'onda significativa elevata.

I dati aggregati per stagione mostrano come previsto, che la stagione con numero maggiore di eventi estremi sia di venti che di onde è l'inverno, mentre quella con minori eventi estremi è l'estate. La primavera e l'autunno mostrano invece distribuzioni in frequenza quasi del tutto identiche sia per i dati di venti che per quelle di onde.

I dati di serie storiche analizzate per anno non hanno mostrato nessun chiaro trend né per l'altezza dell'onda significativa media, né per il 95-esimo percentile. Ciò significa non solo che in media le onde sono rimaste pressoché costanti nel tempo, ma che non vi è un trend nemmeno per gli eventi estremi.

I risultati sono positivi per i residenti nelle zone del Nord-Ovest della Sardegna, poiché il Maestrale nelle sue espressioni più violente causa spesso ingenti danni, accresce la probabilità di incendi e contribuisce ad alimentare quelli che si verificano.

Tuttavia l'analisi ha il limite di avere un intervallo temporale ristretto, dunque non

CAPITOLO 5. CONCLUSIONI

si possono escludere trend su eventi estremamente rari (frequenze nell'ordine di un evento ogni 100 anni).

Bibliografia

- [1] W. A. Lahoz and P. Schneider, “Data assimilation: making sense of earth observation,” *Frontiers in Environmental Science*, vol. 2, 2014.
- [2] A. S. Lawless, “An introduction to data assimilation.” https://eo4society.esa.int/wp-content/uploads/2021/05/lawless_1.pdf, https://www.youtube.com/watch?v=TmBEryh20XY&t=3094s&ab_channel=EuropeanSpaceAgency%2CESA, 2018. Accessed: 2022-11-03.
- [3] R. N. Bannister, “Elementary 4d-var.” <http://darc.nerc.ac.uk/~ross/Documents/Var4d.pdf>, 2007. Accessed: 2022-11-03.
- [4] P. Courtier, J.-N. Thépaut, and A. Hollingsworth, “A strategy for operational implementation of 4d-var, using an incremental approach,” *Quarterly Journal of the Royal Meteorological Society*, vol. 120, no. 519, pp. 1367–1387, 1994.
- [5] A. S. Lawless, “Variational data assimilation and the ensemble kalman filter.” http://eoscience4society.esa.int/EOSS18/files/lawless_2.pdf, https://www.youtube.com/watch?v=8niWYsM2akg&t=1388s&ab_channel=EuropeanSpaceAgency%2CESA, 2018. Accessed: 2022-11-03.
- [6] L. H. Holthuijsen, *Waves in Oceanic and Coastal Waters*. Cambridge University Press, 2007.
- [7] N. Booij, L. Holthuijsen, and R. Ris, “The quot;swan38;quot; wave model for shallow water,” *Coastal Engineering Proceedings*, vol. 1, Jan. 1996.
- [8] H. L. Tolman, *User manual and system documentation of WAVEWATCH III version 3.14*. Environmental Modeling Center Marine Modeling and Analysis Branch, May 2009.
- [9] F. Giorgi, “Simulation of regional climate using a limited area model nested in a general circulation model,” *Journal of Climate*, vol. 3, no. 9, pp. 941 – 963, 1990.
- [10] R. L. Wilby, T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, “Statistical downscaling of general circulation model output: A comparison of methods,” *Water Resources Research*, vol. 34, no. 11, pp. 2995–3008, 1998.

BIBLIOGRAFIA

- [11] V. Vannucchi, S. Taddei, V. Capecci, M. Bendoni, and C. Brandini, “Dynamical downscaling of era5 data on the north-western mediterranean sea: From atmosphere to high-resolution coastal wave climate,” *Journal of Marine Science and Engineering*, vol. 9, no. 2, 2021.
- [12] K. Hasselmann, K. Hasselmann, E. Bauer, P. Janssen, G. Komen, L. Bertotti, P. Lionello, A. Guillaume, V. Cardone, J. Greenwood, M. Reistad, L. Zambresky, and J. Ewing, “The wam model - a third generation ocean wave prediction model.,” *Journal of Physical Oceanography*, vol. 18, pp. 1775–1810, 01 1988.
- [13] H. L. Tolman and N. Booij, “Modeling wind waves using wavenumber-direction spectra and a variable wavenumber grid,” *The Global Atmosphere and Ocean System*, vol. 6, pp. 295–309, 1998.
- [14] E. Pallares, J. Lopez, M. Espino, and A. Sánchez-Arcilla, “Comparison between nested grids and unstructured grids for a high-resolution wave forecasting system in the western mediterranean sea,” *Journal of Operational Oceanography*, vol. 10, no. 1, pp. 45–58, 2017.
- [15] S. Hasselmann, K. Hasselmann, J. H. Allender, and T. P. Barnett, “Computations and parameterizations of the nonlinear energy transfer in a gravity-wave spectrum. part ii: Parameterizations of the nonlinear energy transfer for application in wave models,” *Journal of Physical Oceanography*, vol. 15, no. 11, pp. 1378 – 1391, 1985.
- [16] F. Ardhuin, E. Rogers, A. Babanin, O. Filipot, R. Magne, A. Roland, A. Van, D. Westhuyzen, P. Queffeulou, J.-M. Lefevre, L. Aouf, and F. Collard, “Semiempirical dissipation source functions for ocean waves. part i: Definition, calibration, and validation,” *Journal of Physical Oceanography*, vol. 40, p. 1917, 09 2010.
- [17] F. Leckler, F. Ardhuin, J.-F. Filipot, and A. Mironov, “Dissipation source terms and whitecap statistics,” *Ocean Modelling*, vol. 70, pp. 62–74, 2013. Ocean Surface Waves.
- [18] N. Booij and L. H. Holthuijsen, “Propagation of Ocean Waves in Discrete Spectral Wave Models,” *Journal of Computational Physics*, vol. 68, pp. 307–326, Feb. 1987.
- [19] F. Schott, M. Visbeck, U. Send, J. Fischer, L. Stramma, and Y. Desaubies, “Observations of deep convection in the gulf of lions, northern mediterranean, during the winter of 1991/92,” *Journal of Physical Oceanography*, vol. 26, no. 4, pp. 505 – 524, 1996.
- [20] P. Wais, “A review of weibull functions in wind sector,” *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1099–1107, 2017.
- [21] A. Meucci, I. R. Young, M. Hemer, E. Kirezci, and R. Ranasinghe, “Projected 21st century changes in extreme wind-wave events,” *Science Advances*, vol. 6, no. 24, p. eaaz7295, 2020.