

Progetto

**SINAPSI**

**asSistenza alla Navigazione per l'Accesso ai Porti in  
Sicurezza**



**ATTIVITA' T2.1: IMPLEMENTAZIONE MODELLISTICA NUMERICA**

**PRODOTTO T2.1.2: METODOLOGIA CONDIVISA DEFINIZIONE SCENARI DI  
PREVISIONE**

Partner responsabile: UNIGE

Partner contributori: UTLN, CNR-ISMAR, LaMMA

## **SOMMARIO**

<b><i>Descrizione del prodotto</i></b> .....	<b>1</b>
<b>1. INTRODUZIONE</b> .....	<b>1</b>
<b>2. CLUSTERIZZAZIONE DEI DATI METEO-MARINI</b> .....	<b>2</b>
<b>3. APPLICAZIONE AD UN CASO STUDIO</b> .....	<b>4</b>
<b>4. MODELLO DI CIRCOLAZIONE NELLA BAIA DI TOLONE</b> .....	<b>20</b>
<b><i>Bibliografia</i></b> .....	<b>22</b>

## **Descrizione del prodotto**

Il presente report sintetizza le tecniche di *data mining* utilizzate per la selezione di scenari meteo-marini, utili a rappresentare la climatologia delle aree di studio e a fornire il riferimento per previsioni di dati meteo-marini nel breve termine nell'ambito del progetto SINAPSI.

## **1. INTRODUZIONE**

Nel corso degli ultimi anni, l'analisi dei processi fisici nelle aree costiere si è avvalsa sempre più spesso di modelli numerici capaci di risolvere le complesse equazioni relative all'idrodinamica ed alla morfodinamica in prossimità delle coste. Tali modelli simulano processi di dispersione sia da un punto di vista euleriano (dinamica della concentrazione) che da un punto di vista lagrangiano (dispersione di massa e oggetti), permettendo dunque di riprodurre in dettaglio i processi fisici che si sviluppano sotto costa; l'utilizzo dei medesimi richiede tuttavia elevate risorse computazionali (i.e., supercomputers), che non sempre risultano essere disponibili.

Lo studio dell'idrodinamica costiera si basa infatti su cospicue quantità di dati che, nel caso di prodotti di rianalisi (i cosiddetti *hindcast*) si estendono su intervalli temporali estesi. In questi casi, vi è la necessità di ridurre il numero di condizioni ambientali da utilizzarsi a fini modellistici, per poter individuare i modi più significativi della variabilità del fenomeno studiato. Risolvere un numero limitato di condizioni ambientali, altresì dette "scenari", è vantaggioso perché non solo permette di selezionare gli stati meteo-marini più importanti per il processo analizzato, ma anche perché viene ridotto significativamente il carico computazionale necessario per risolvere l'intera catena modellistica.

A tal fine, si possono impiegare algoritmi di clusterizzazione dei dati, che permettono di ripartire le serie di dati di partenza in un numero ridotto di sottoinsiemi, ciascuno rappresentato da un centroide di riferimento (gli scenari di cui sopra). L'utilizzo di tali algoritmi, seppur solo di recente diffusosi nell'ambito dell'ingegneria marittima e costiera, ha visto una crescita repentina nella bibliografia di settore, poiché consente di descrivere dettagliatamente la variabilità delle condizioni meteo-marine di un sito concentrandosi su un numero ridotto di scenari.

Nell'ambito del progetto SINAPSI, il Dipartimento di Ingegneria Civile, Chimica e Ambientale dell'Università di Genova (di seguito DICCA) ha sviluppato una serie di scripts per la clusterizzazione di variabili meteo-marine da prodotti hindcast (dunque

estremamente ampi per quanto riguarda la numerosità delle variabili archiviate) e la definizione di scenari variabili nel tempo. Tali scenari saranno successivamente utilizzati per caratterizzare l'idrodinamica nelle aree pilota del progetto SINAPSI tramite opportuni modelli di calcolo. Gli algoritmi alla base della clusterizzazione sono spiegati nel capitolo 2, mentre il capitolo 3 riporta un caso studio esemplificativo della procedura di calcolo per la selezione degli scenari meteo-marini di riferimento per un sito pilota del progetto.

## 2. CLUSTERIZZAZIONE DEI DATI METEO-MARINI

Le tecniche di clusterizzazione sono atte a identificare gruppi di dati che siano fra loro omogenei, e tuttavia eterogenei rispetto agli elementi degli altri gruppi identificati. Nell'ambito del progetto SINAPSI, il fine ultimo di tali tecniche è la selezione di scenari in grado di rappresentare con ragionevole approssimazione la variabilità totale dei dati di partenza, gli stessi venendo successivamente utilizzati come input dei modelli numerici impiegati.

Si consideri un dataset formato da un numero  $V$  di variabili e da  $n$  elementi. Seguendo il diagramma di flusso proposto in (Camus 2011) Il primo passo per eseguire la clusterizzazione di tale dataset è disporlo in forma matriciale, normalizzando i diversi elementi in uno spazio comune colonna per colonna:

$$X_{n,V} = \begin{bmatrix} x_{1,1} & \dots & x_{1,V} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,V} \end{bmatrix}$$

2.1

Ciascun vettore colonna della matrice di cui all'equazione 2.1 rappresenta una variabile del dataset opportunamente normalizzata, ad esempio tramite centraggio dei dati rispetto alla media e successivo ridimensionamento sulla deviazione standard, come mostrato nell'equazione 2.2.

$$X'_{:,j} = \frac{x'_{:,j} - \mu(x'_{:,j})}{\sigma(x'_{:,j})}$$

2.2

dove  $\mu$  e  $\sigma$  rappresentano rispettivamente la media e la deviazione standard della  $i$ -esima colonna della matrice  $X_{n,v}$ . Tale operazione risulta necessaria dal momento che permette all'algoritmo di clustering di trattare sia dati negativi sia positivi, caratterizzati da differenti ordini di grandezza, evitando che le variabili più rilevanti alterino i successivi calcoli.

Nel caso si vogliano selezionare scenari variabili nel tempo, è dapprima necessario ristrutturare i dati di partenza (Bárcena 2015); selezionato un numero  $nt$  di istanti temporali, si modella il singolo vettore  $x$  di un'assegnata variabile in forma matriciale come segue:

$$x'_j = x[i\delta + 1 : nt + i\delta], i = 0, \dots, \frac{n - nt}{\delta}; j = i + 1$$

2.3

Dove  $\delta$  rappresenta lo shift tra due istanti temporali successivi (ovvero il numero di passi temporali tra gli istanti iniziali di due successivi subset  $x'$ ). Il clustering si esegue dunque sulla matrice risultante  $X'$ , nella quale ogni riga definisce una finestra di lunghezza  $nt$  per un'assegnata variabile.

L'equazione 2.4 riassume i procedimenti poc'anzi introdotti.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{nt} \\ \vdots \\ x_n \end{bmatrix} \quad X' = \begin{bmatrix} x_1 & \dots & x_{nt} \\ x_{i\delta+1} & \dots & x_{nt+i\delta} \\ \vdots & \dots & \vdots \\ x_{n-nt+1} & \dots & x_n \end{bmatrix}$$

2.4

Qualora si lavori con molteplici parametri, è sufficiente replicare i passaggi esposti nelle equazioni 2.4 e 2.4 a ciascuno di essi e successivamente concatenare le matrici risultanti in una matrice di dimensioni totali  $[n-nt+1] \times [nt \times V]$ . La selezione degli scenari viene fatta dunque su quest'ultima matrice tramite il Maximum Dissimilarity Algorithm (MDA). Per facilitare la comprensione di questo algoritmo, si fa riferimento alla matrice di cui all'equazione 2.1.

## PRODOTTO N. T2.1.2

L'obiettivo del MDA è quello di selezionare un sottoinsieme dalla matrice iniziale  $X_{n,V}$ , detto  $X_{m,V}^*$  (matrice degli stati significativi,  $m < n$ ) costituita dagli stati che meglio rappresentano la varianza complessiva dei dati. Dopo aver normalizzato i dati colonna per colonna, al fine di poter lavorare con variabili caratterizzate da diversi ordini di grandezza, si identifica un primo stato significativo selezionando lo stato più lontano dalla nuvola di punti iniziale (ovvero, quello più "dissimilare" dal resto del dataset). Successivamente, si procede con il calcolo della dissimilarità tra i vettori della matrice  $X_{n-1,V}$  ed il sottoinsieme  $X_{1,V}^*$  grazie al calcolo della distanza Euclidea così definita:

$$d_{i1} = ||x_i - x_1^*||, \quad i = 1, \dots, n - 1$$

2.5

dove  $x$  e  $x^*$  indicano rispettivamente le righe delle matrici di input e di output. Il nuovo elemento viene successivamente selezionato come quello caratterizzato dal valore massimo di  $d_{i1}$  e aggiunto alla matrice target  $X^*$ . Si applica la versione MaxMin dell'algoritmo: infatti, in corrispondenza della  $k^{esima}$  iterazione ( $k < m$ , dove  $k$  è il numero di elementi della matrice  $X^*$ ), la distanza da considerare per ogni elemento di  $X_{n-k,V}$  è la minima rispetto a tutti i  $k$  vettori di  $X_{k,V}^*$ ; tra tutte queste distanze viene selezionata quella massima e il corrispondente stato viene di conseguenza aggiunto a  $X^*$ . Il calcolo finisce quando  $k$  diventa uguale a  $m$ , ovvero una volta che viene raggiunto il numero di cluster (da stabilirsi in precedenza). Alla fine del procedimento di costruzione della matrice  $X^*$ , si ottiene un sottoinsieme di vettori che permette di riassumere il dataset di partenza con un numero minore di stati: gli elementi rimasti vengono assegnati al corrispondente vettore modello più vicino, formando delle vere e proprie classi di stati.

La stessa procedura di calcolo si applica a prescindere dalle dimensioni della matrice di partenza, dunque può essere estesa senza perdita di generalità alle matrici composte concatenando le sottomatrici di cui all'equazione 2.4

### 3. APPLICAZIONE AD UN CASO STUDIO

La metodologia introdotta nel precedente capitolo è stata utilizzata per la selezione di scenari meteomarini al largo di Genova, avvalendosi dei dati hindcast forniti dal DICCA (<http://www3.dicca.unige.it/meteocean/hindcast.html>). L'hindcast suddetto fornisce serie orarie di parametri ondametrici e anemometrici nell'intervallo 1979-2020 su una

**PRODOTTO N. T2.1.2**

griglia regolare estesa su tutto il Mar Mediterraneo a risoluzione pari a ca. 10 km sia in longitudine che in latitudine; a partire dal 2021, i dati sono aggiornati su una maglia non strutturata a risoluzione variabile. Il nodo utilizzato nel presente studio è il no. 000203, mostrato in Figura 1, mentre Tabella 1 ne riporta le coordinate.

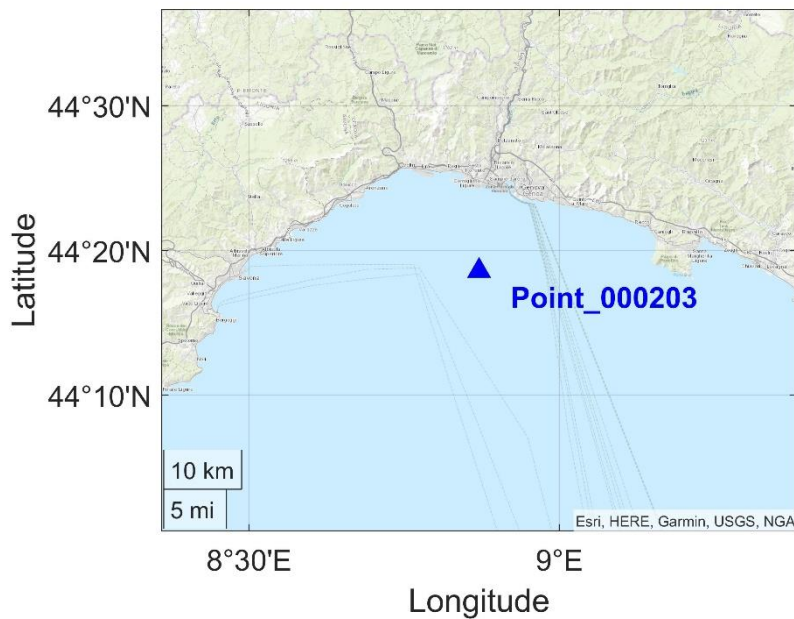


Figura 1. Nodo hindcast utilizzato per lo studio.

Tabella 1. Coordinate del nodo hindcast considerate (SR: WGS84).

PUNTO	LON	LAT
000203	8.8707	44.31

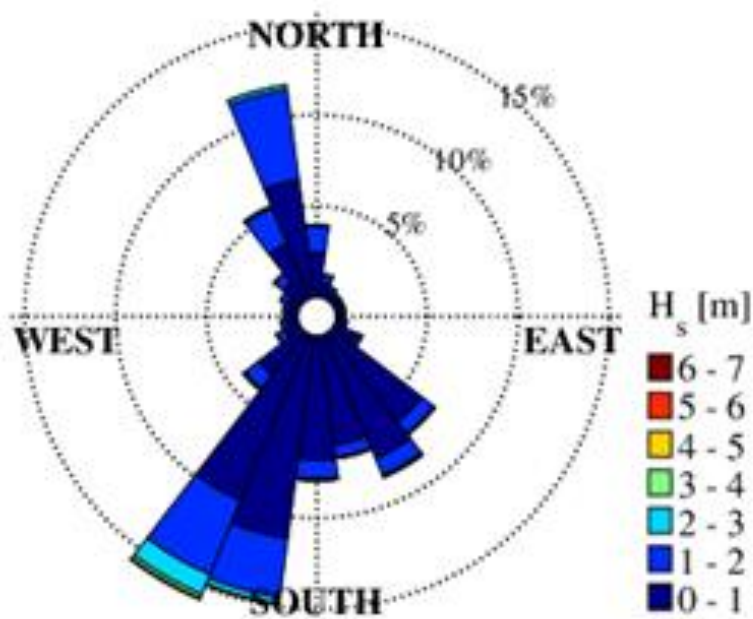
I parametri meteo-marini considerati sono altezza d'onda significativa ( $H_s$ ), periodo e direzione di picco ( $T_p$  e  $\theta_p$  rispettivamente; quest'ultima espressa in convenzione nautica, ovvero onde caratterizzate da  $\theta_p$  pari a  $0^\circ/360^\circ$  viaggiano da Nord a Sud; onde caratterizzate da  $\theta_p$  pari a  $180^\circ$  viaggiano da Sud a Nord), velocità longitudinale e latitudinale del vento ( $v_x/v_y$ ). Agli scenari selezionati, si sono associate le rispettive escursioni mareali ( $\Delta\eta$ ) tramite il software di calcolo TPXO (<https://www.tpxo.net/>).

Per la selezione degli scenari, sono state prese a riferimento finestre temporali di una settimana ( $nt = 168$ , si noti che i dati hanno risoluzione oraria), con uno shift  $\delta$  pari a 3 giorni. Gli scenari estratti secondo la metodologia introdotta nel capitolo 2 sono



**PRODOTTO N. T2.1.2**

riportati nelle figure da 3 a 27, e mostrano come gli stessi siano in grado di riprodurre i principali modi del clima ondoso locale. Questo risultato può essere apprezzato analizzando i diagrammi polari di  $H_s$  in funzione delle diverse classi di intensità (Figura 2), e confrontando le direzioni prevalenti con quelle associate agli scenari di cui sopra (i.e., Libeccio, Scirocco, Grecale).



*Figura 2. Digramma polare delle onde nel punto 000203. Le direzioni sono espresse in convenzione nautica.*



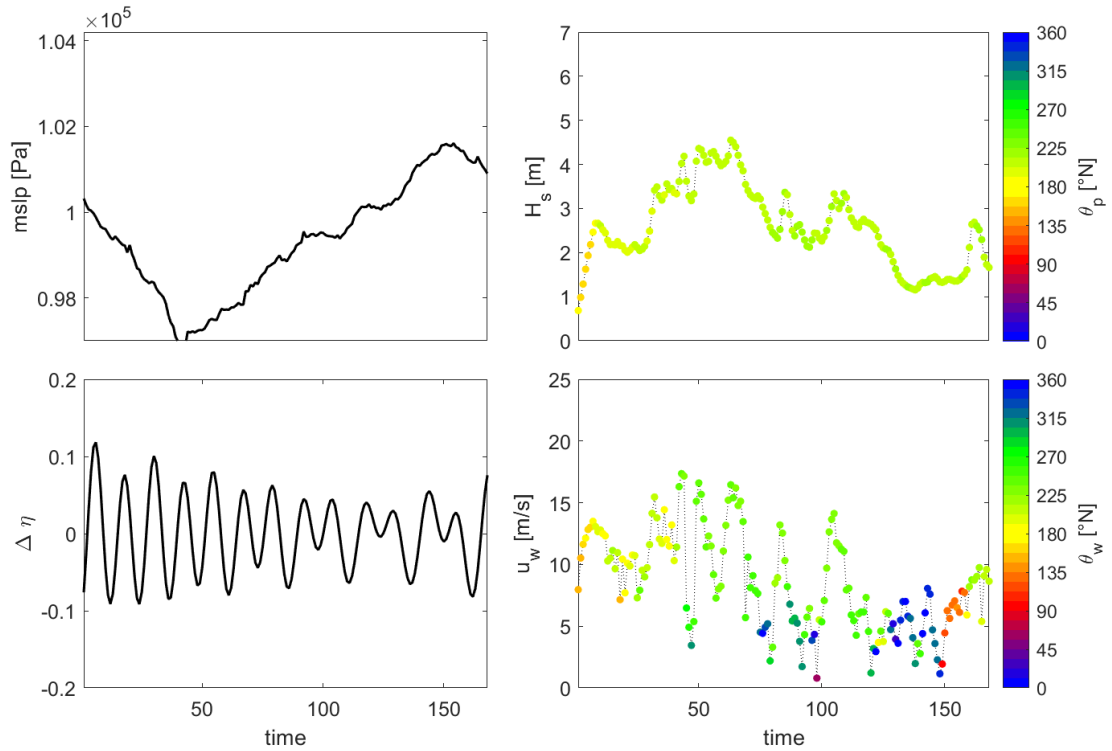


Figura 3 Scenario no. 1. La direzione del vento è indicata con  $\theta_w$  (notazione analoga nelle figure successive).

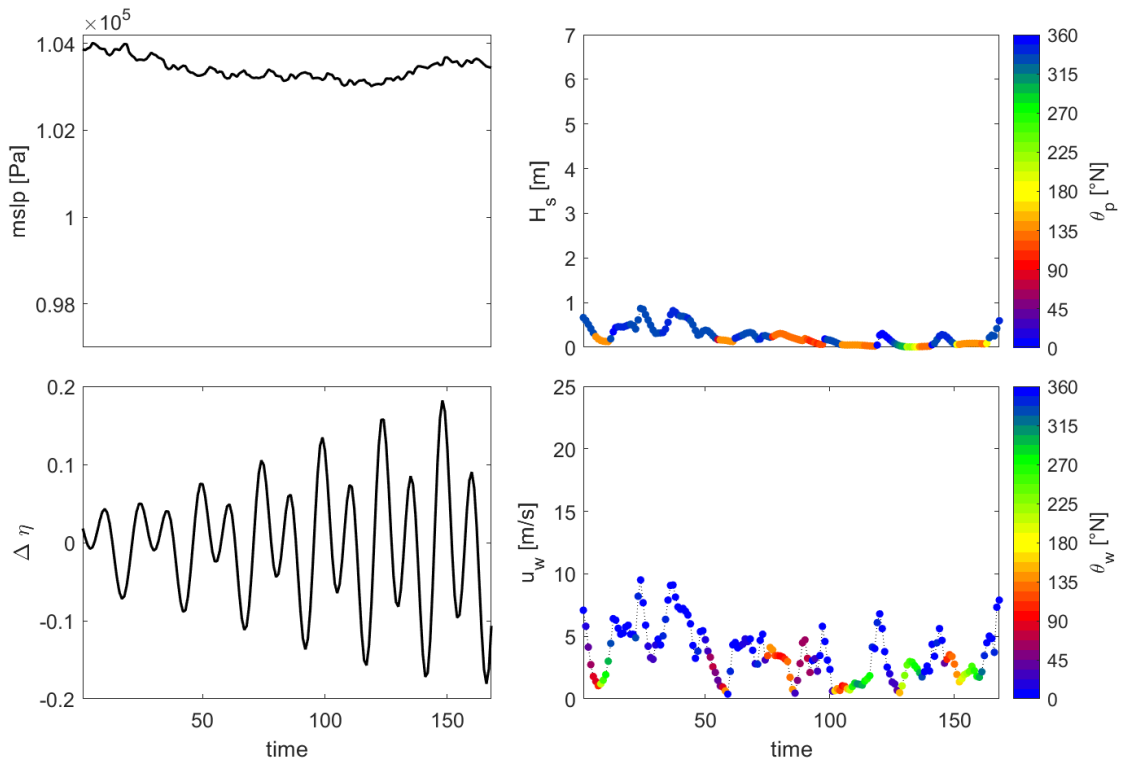


Figura 4. Scenario no. 2.

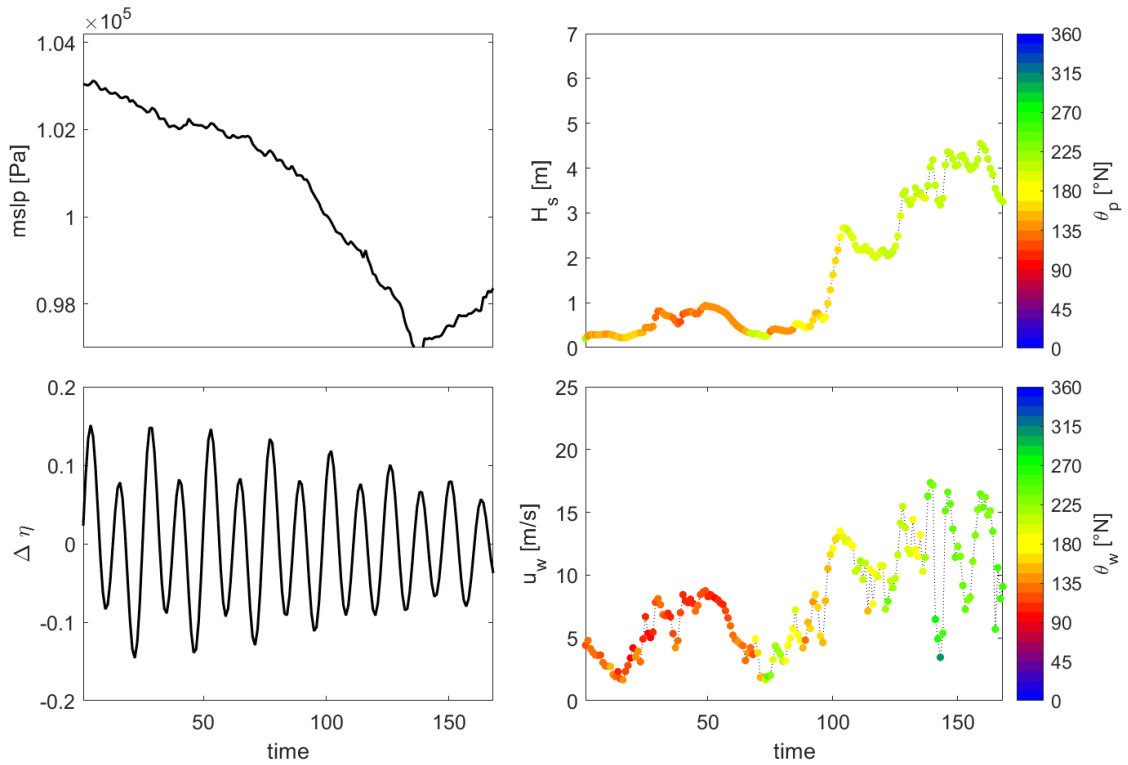


Figura 5. Scenario no. 3.

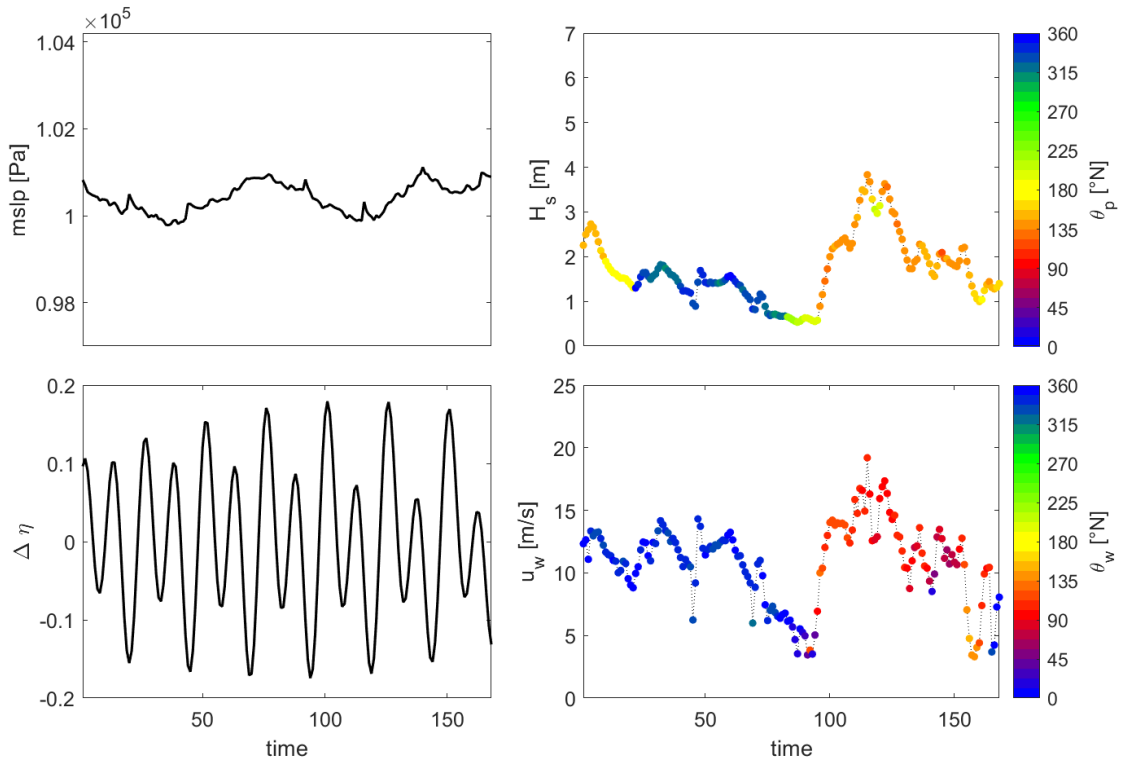


Figura 6. Scenario no. 4.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

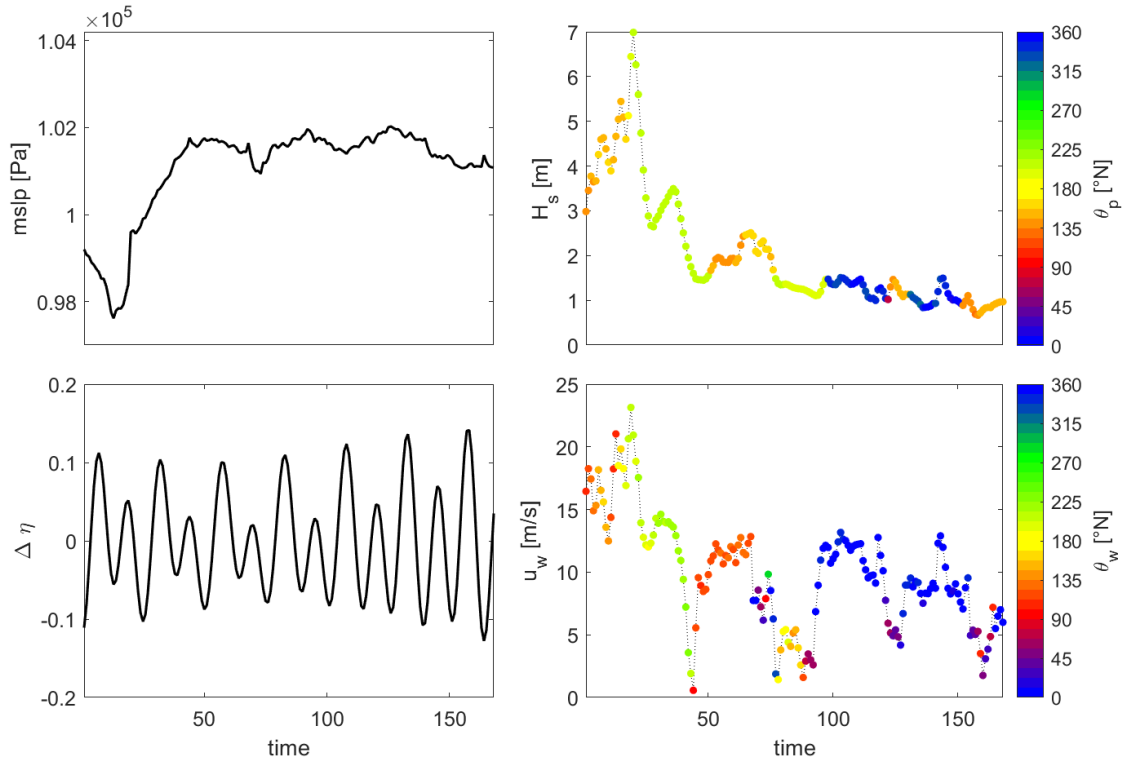


Figura 7. Scenario no. 5.

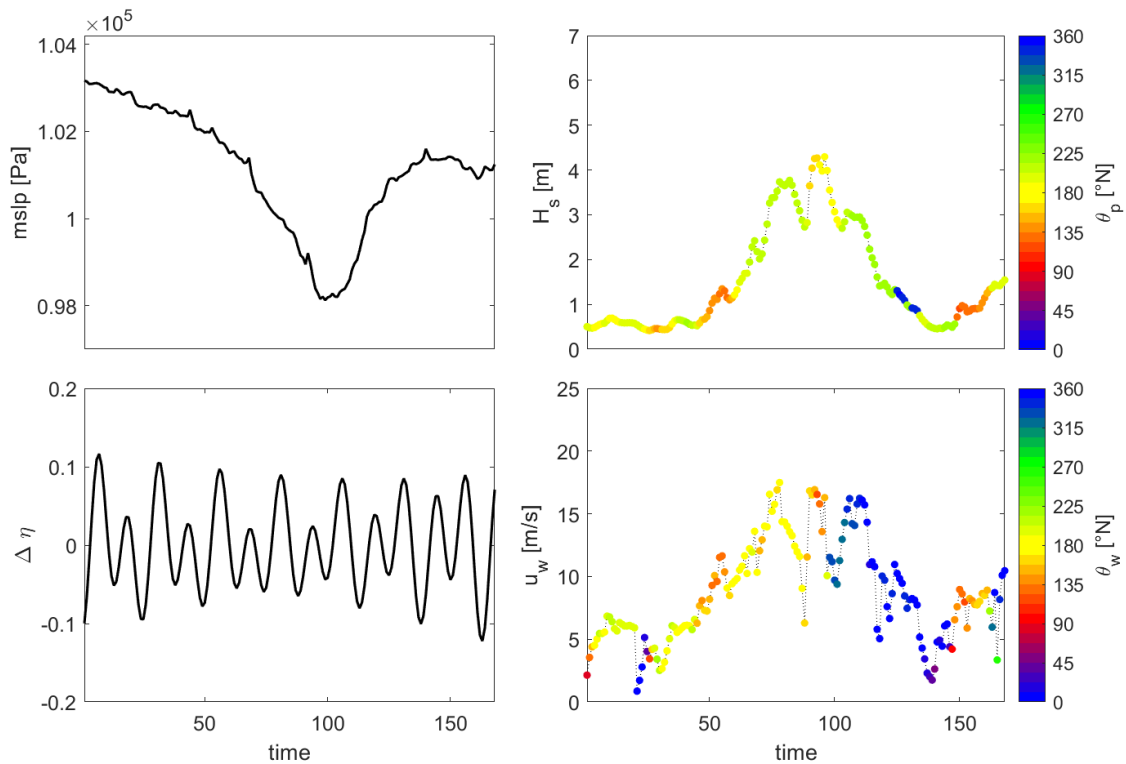


Figura 8. Scenario no. 6.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



SINAPSI

MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

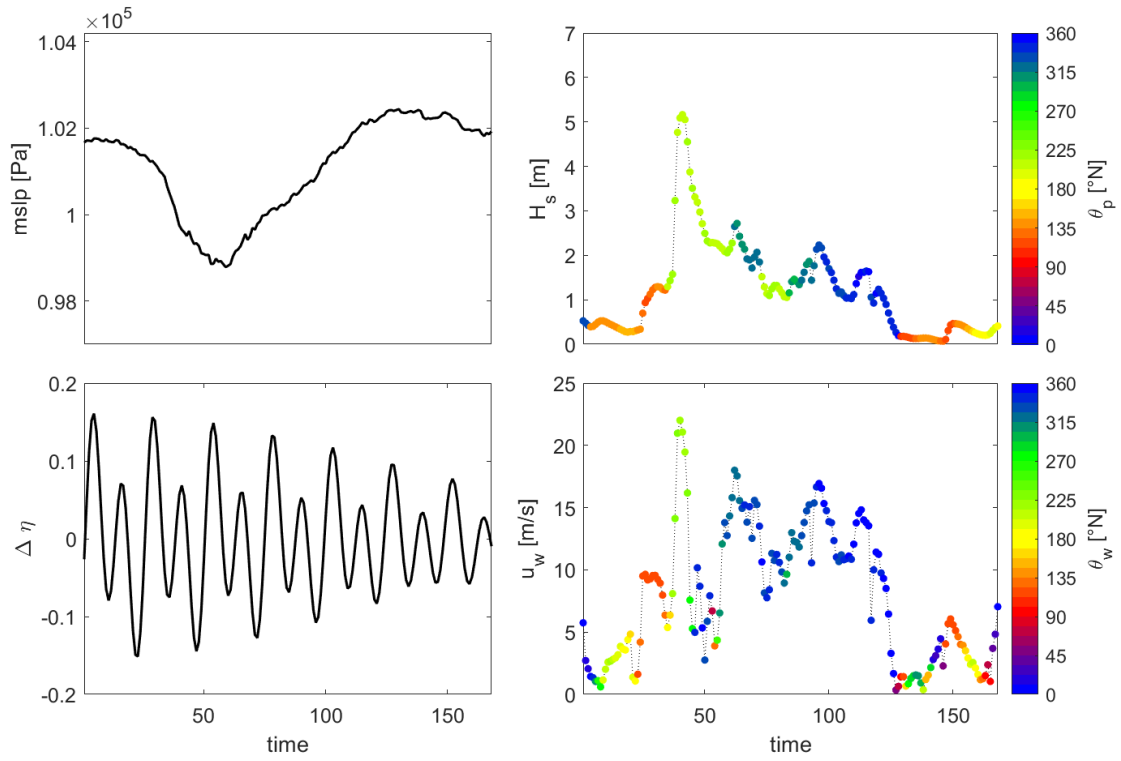


Figura 9. Scenario no. 7.

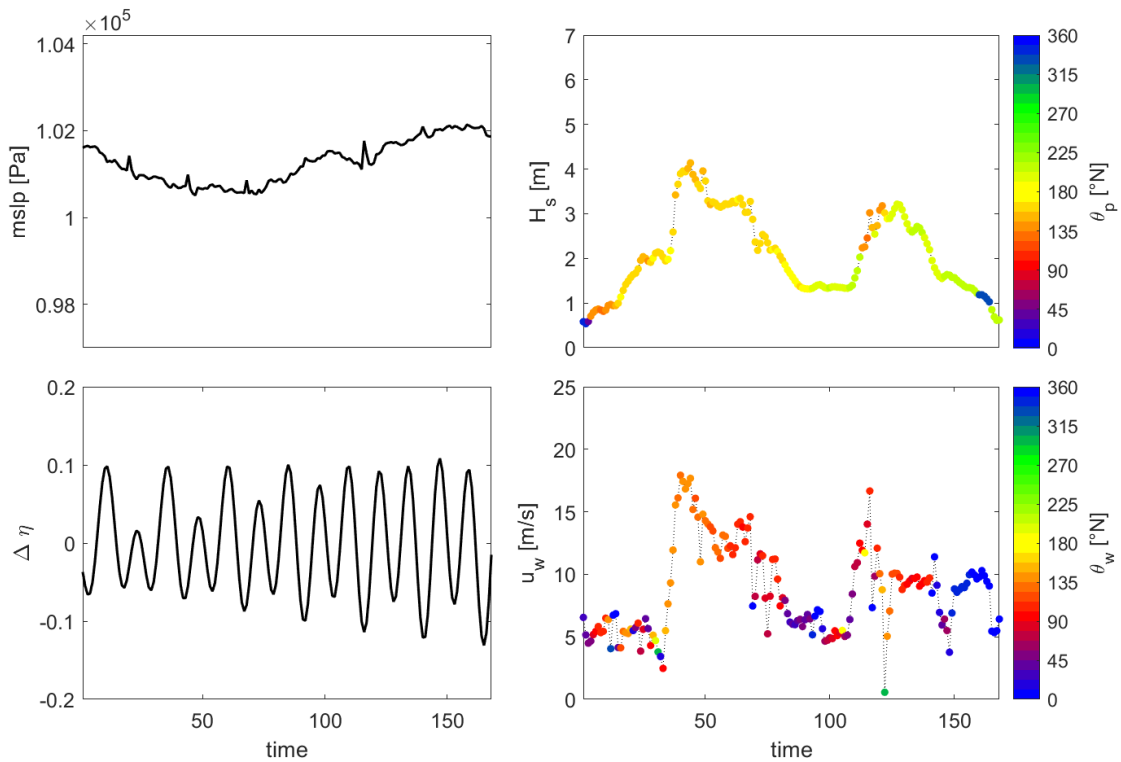


Figura 10. Scenario no. 8.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

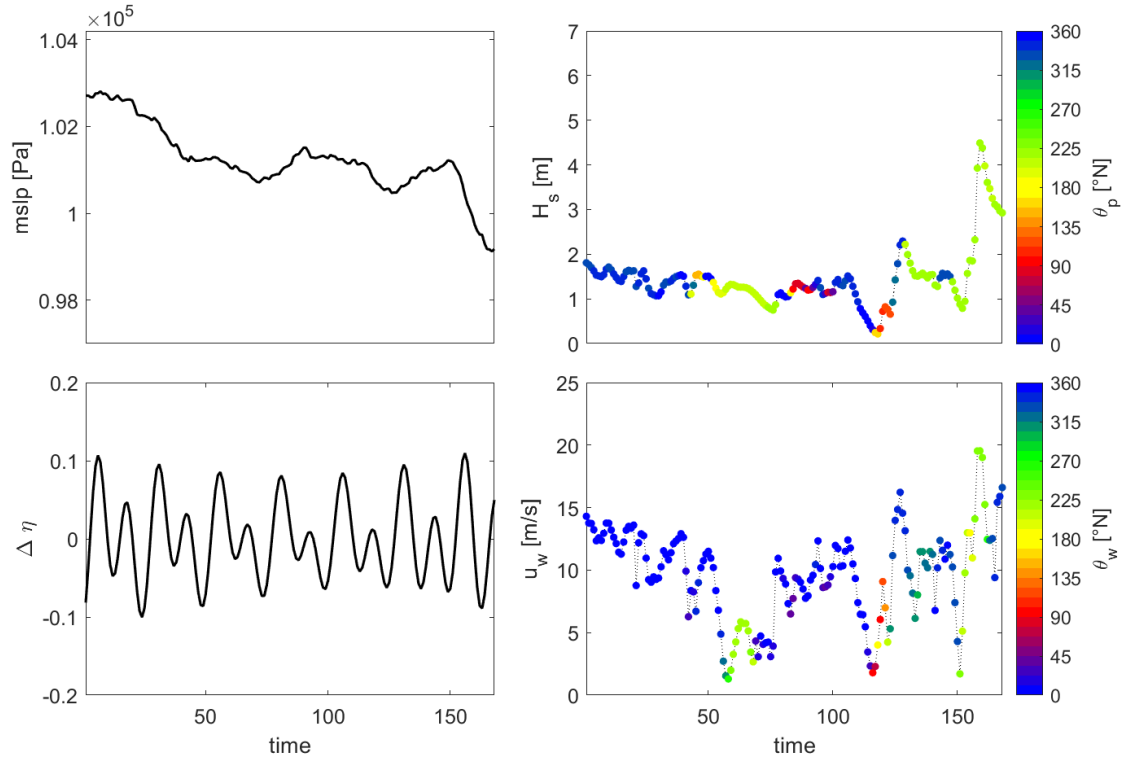


Figura 11. Scenario no. 9.

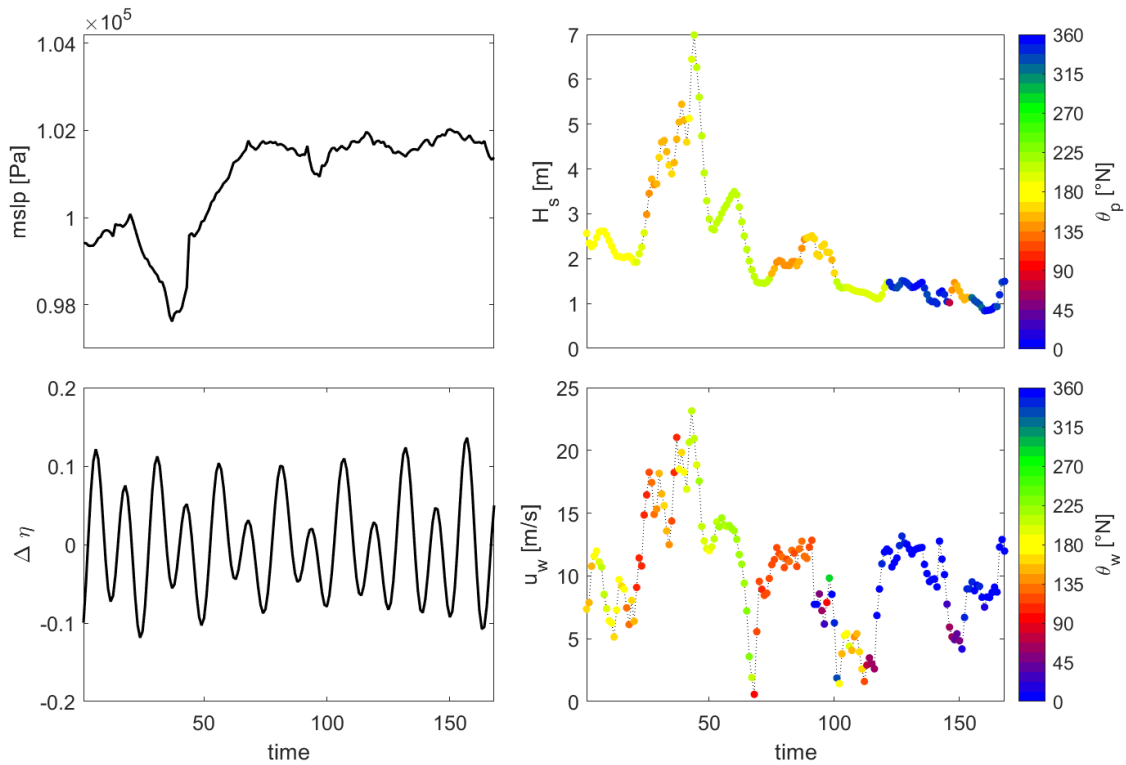


Figura 12. Scenario no. 10.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



SINAPSI

MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

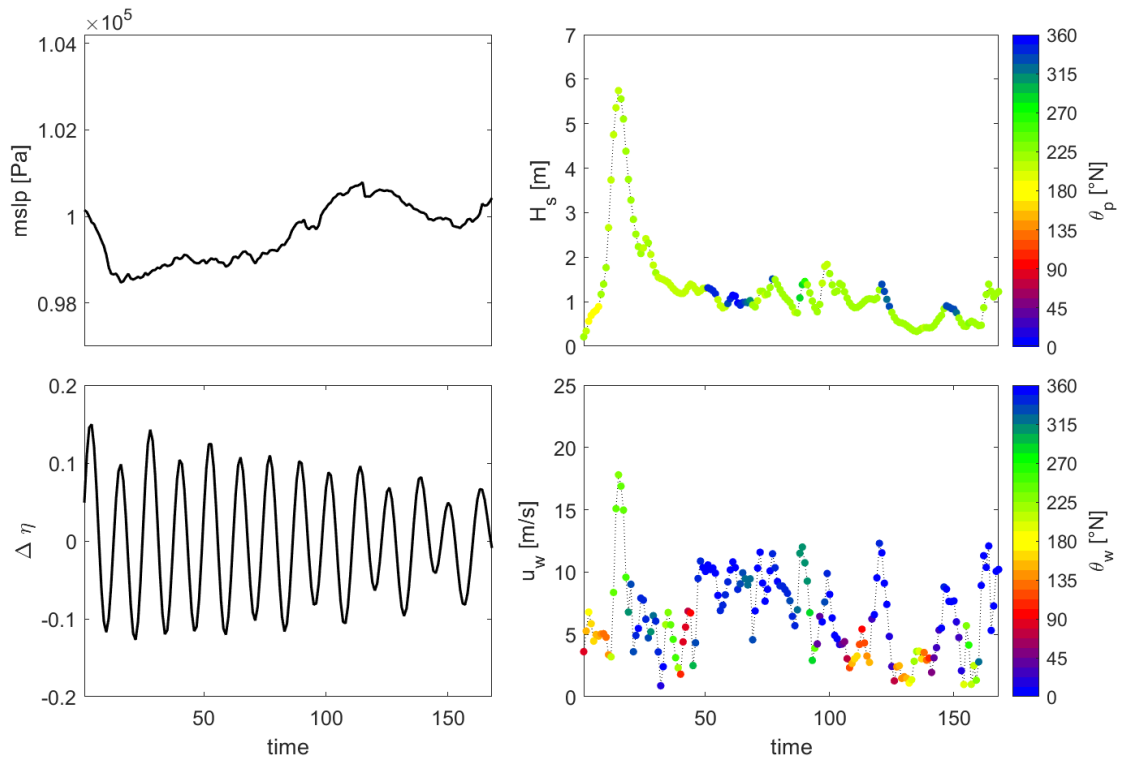


Figura 13. Scenario no. 11.

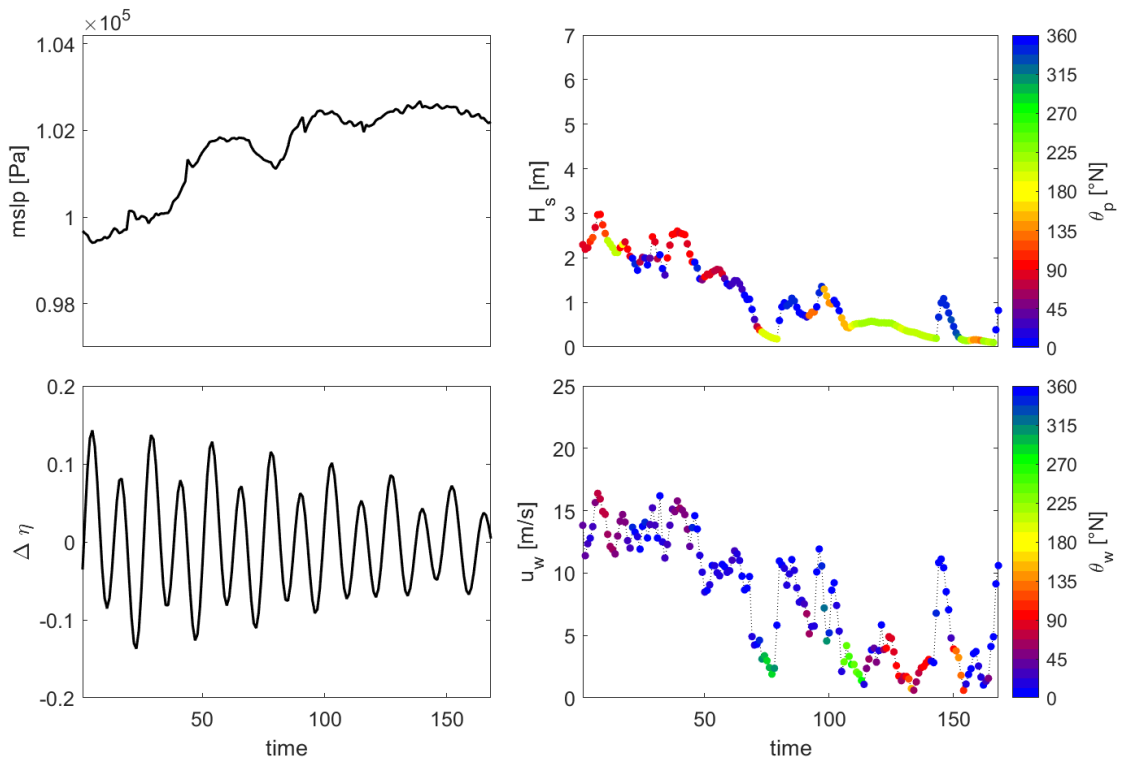


Figura 14. Scenario no. 12.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

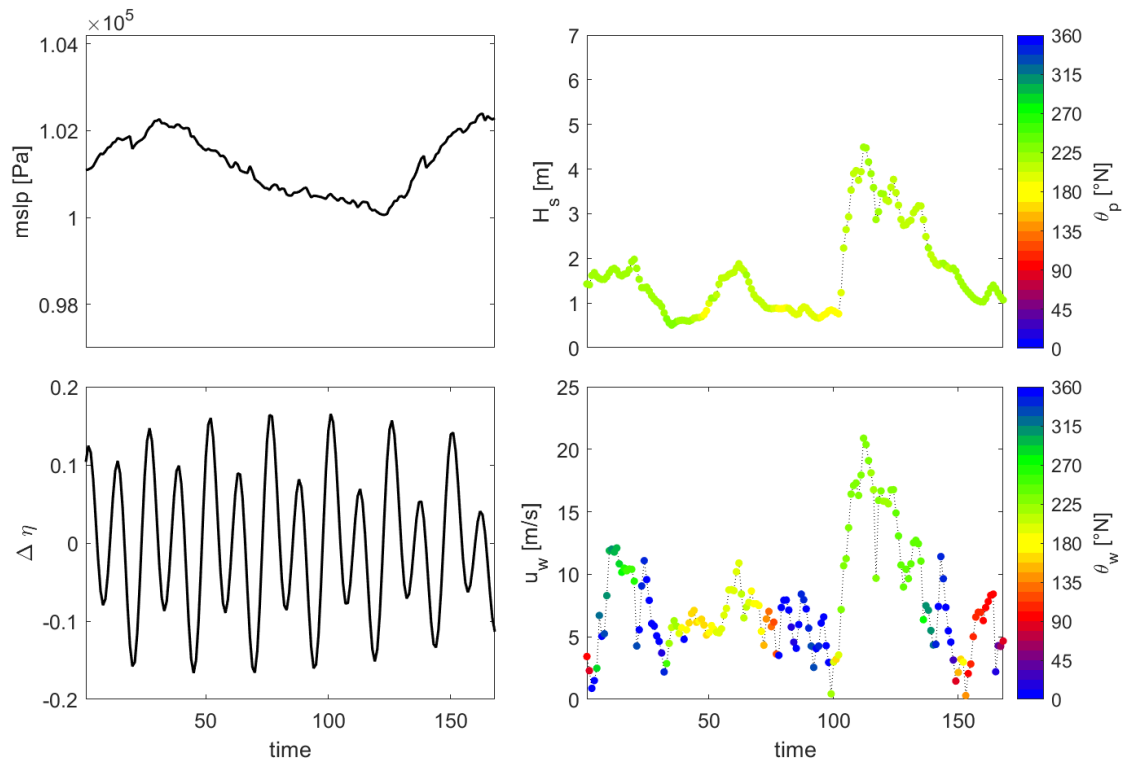


Figura 15. Scenario no. 13.

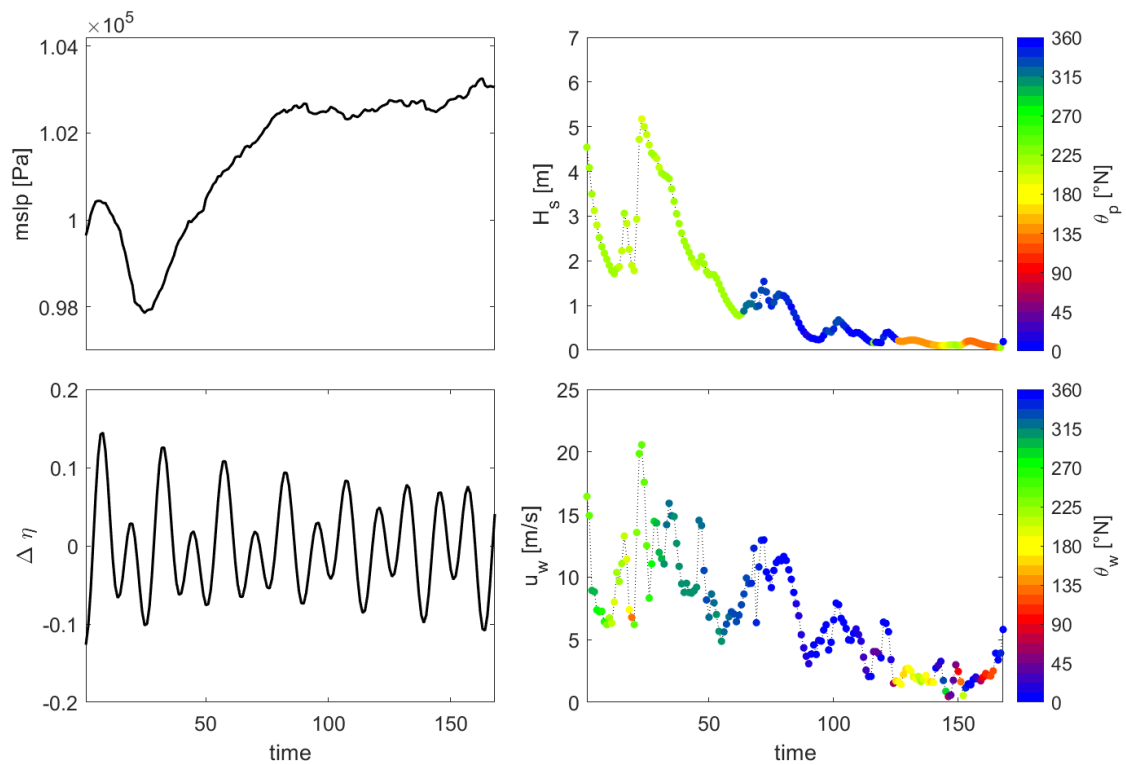


Figura 16. Scenario no. 14.





**Interreg**



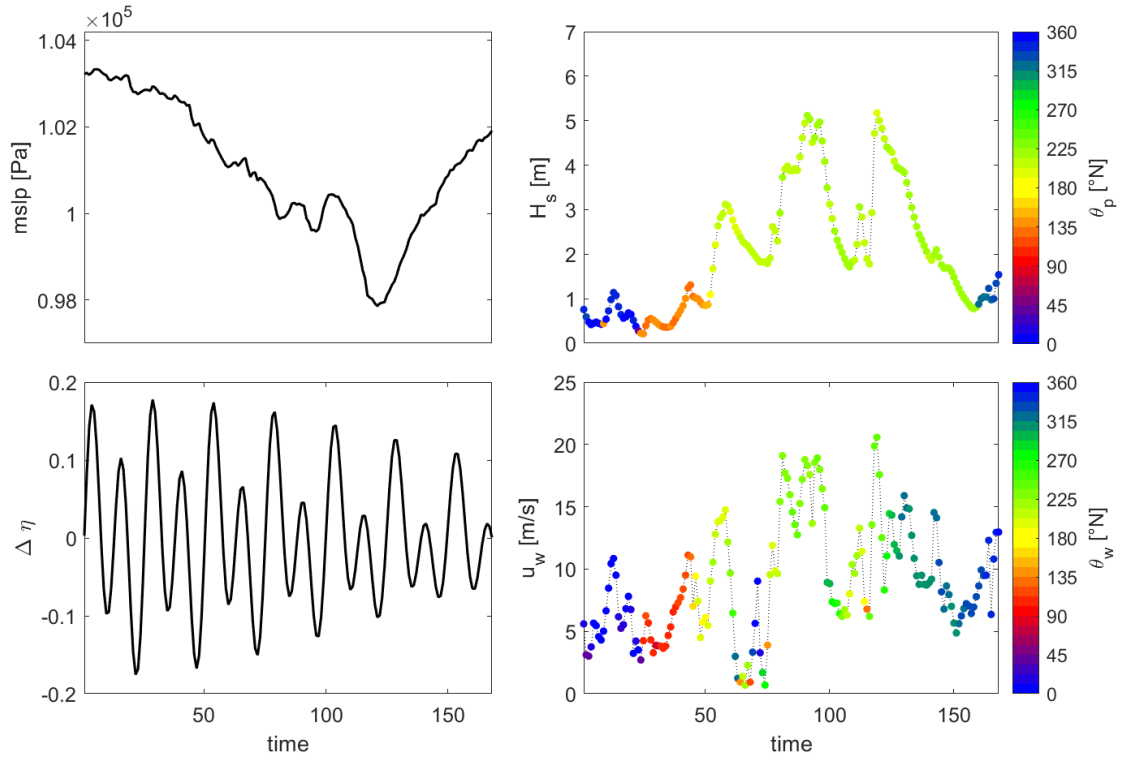
UNION EUROPÉENNE  
UNIONE EUROPEA



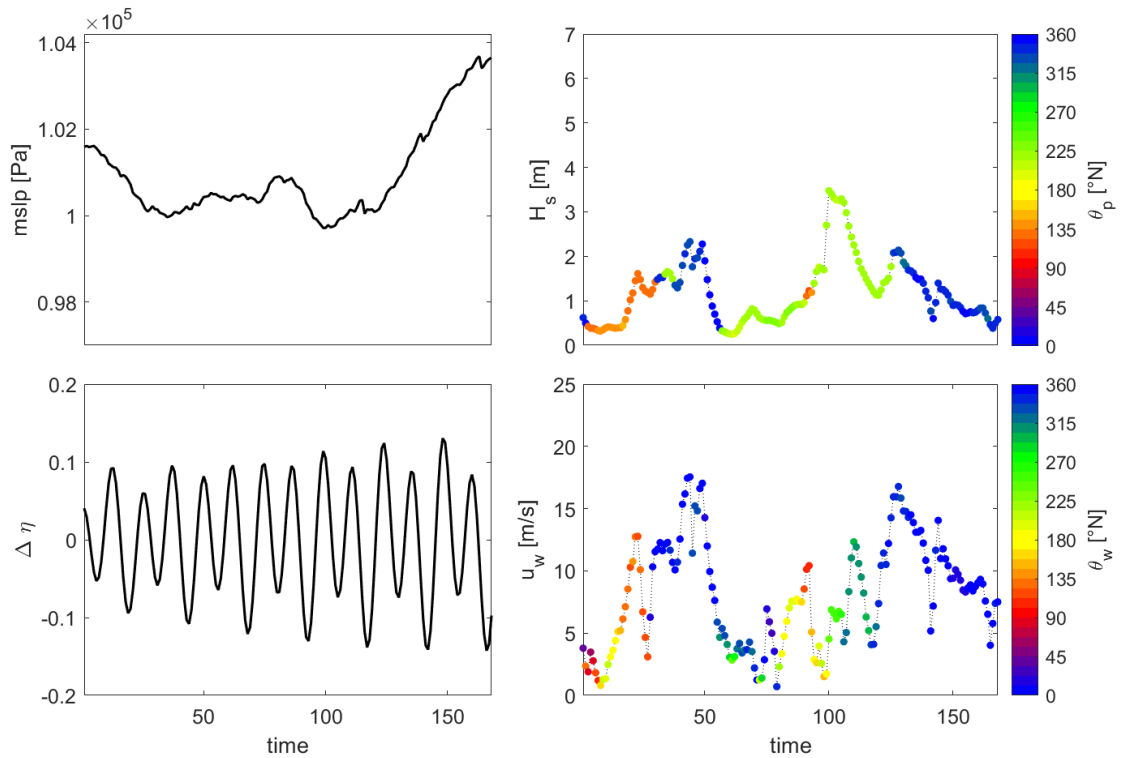
MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

**PRODOTTO N. T2.1.2**



*Figura 17. Scenario no. 15.*



*Figura 18. Scenario no. 16.*

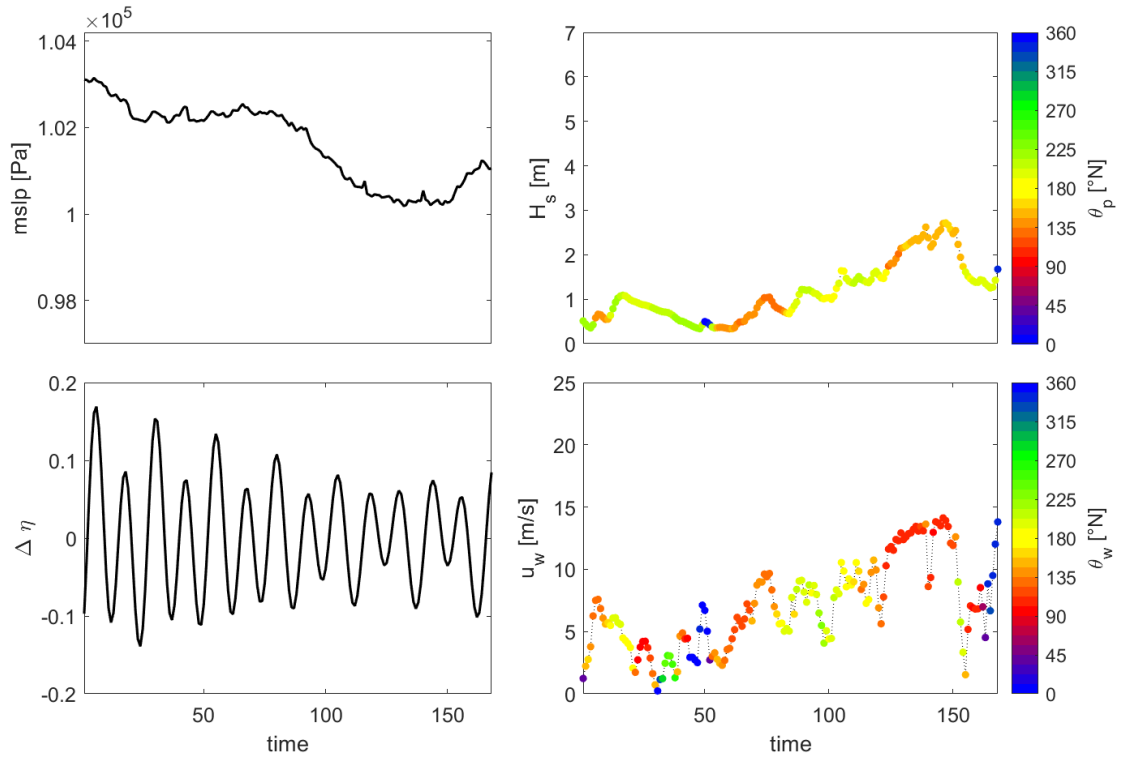


Figura 19. Scenario no. 17.

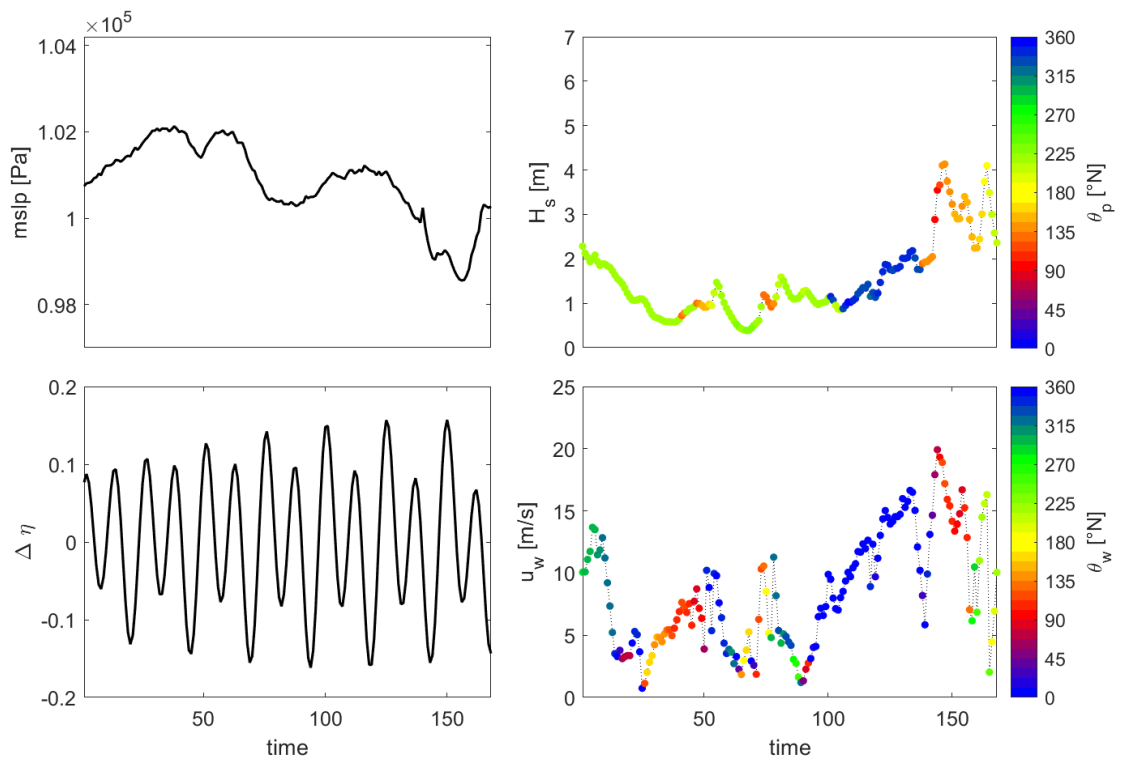


Figura 20. Scenario no. 18.



# Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



## MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

### PRODOTTO N. T2.1.2

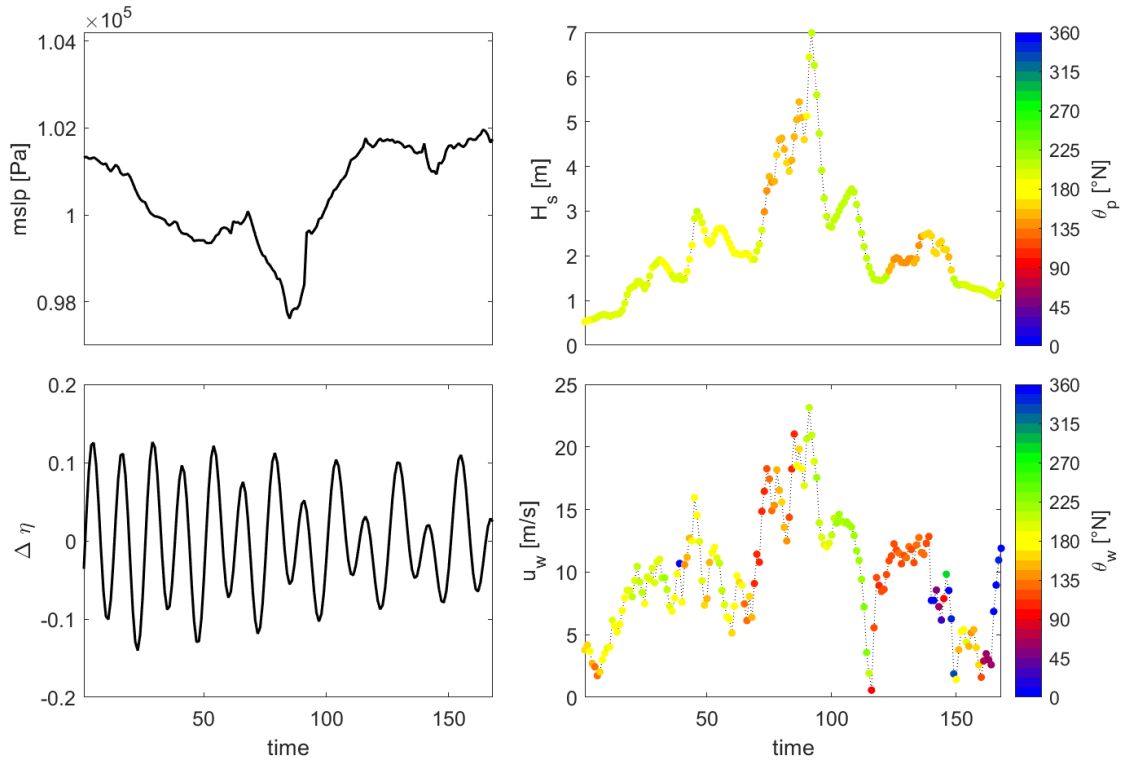


Figura 21. Scenario no. 19.

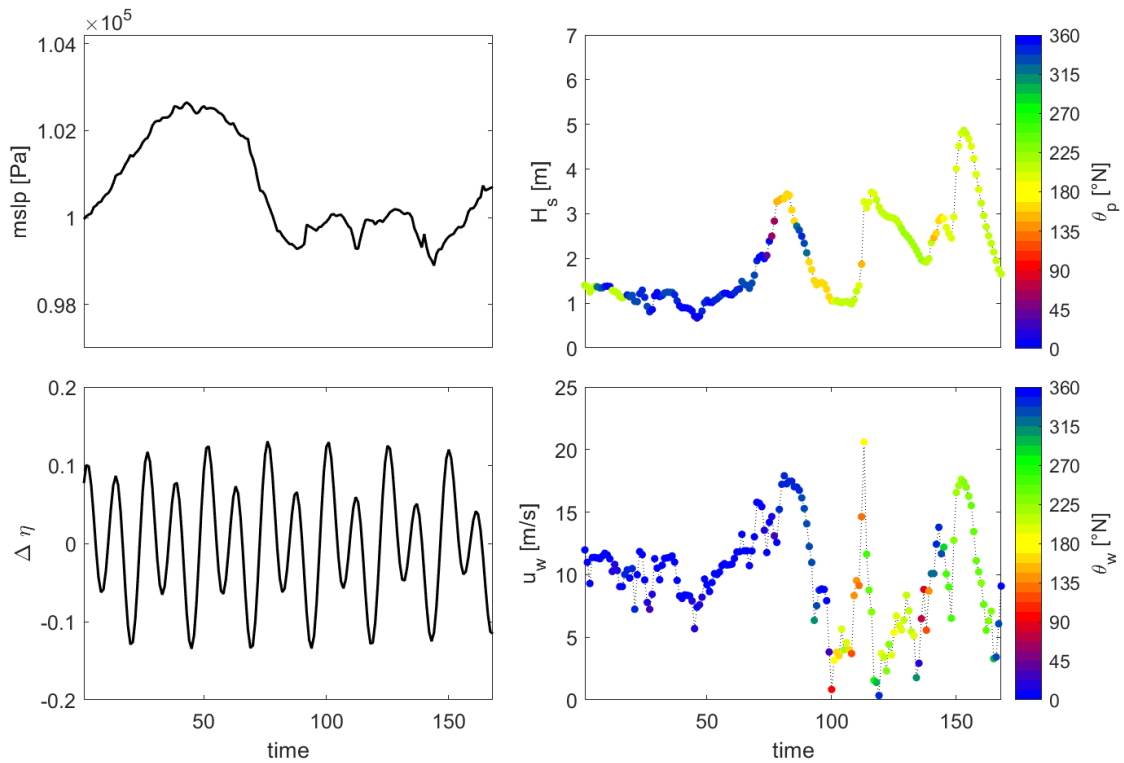


Figura 22. Scenario no. 20.



Interreg



UNION EUROPÉENNE  
UNIONE EUROPEA



SINAPSI

MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional  
Fondo Europeo di Sviluppo Regionale

PRODOTTO N. T2.1.2

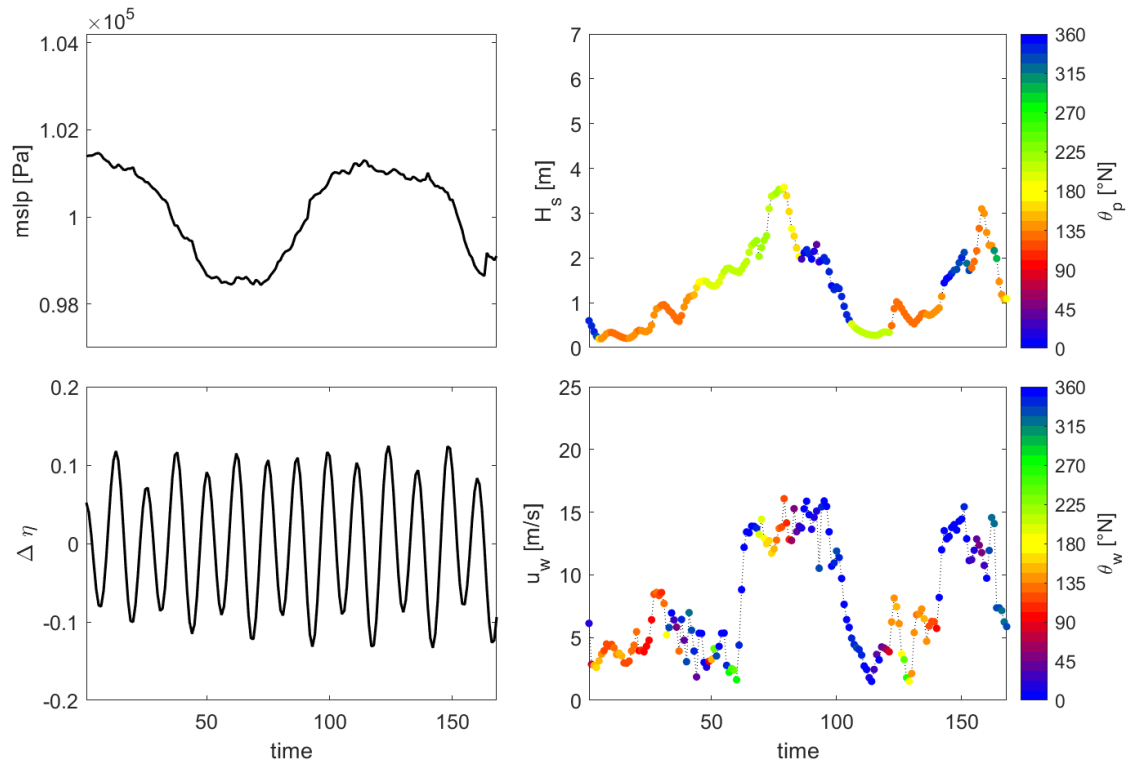


Figura 23. Scenario no. 21.

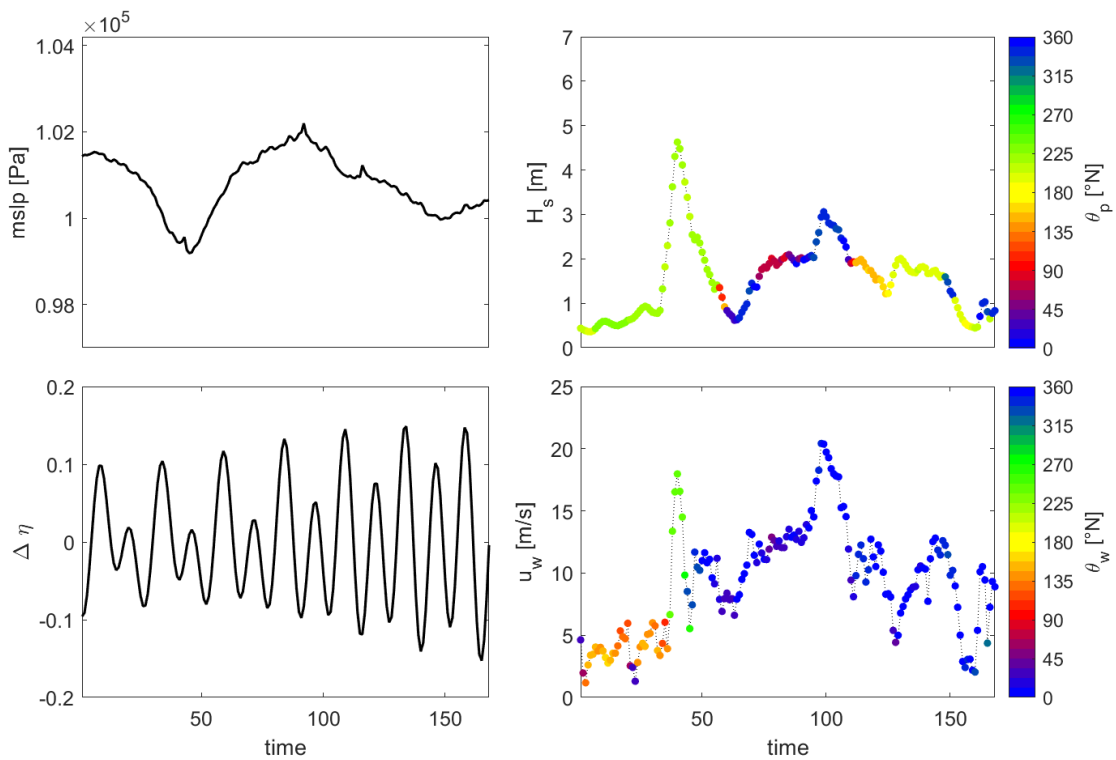


Figura 24. Scenario no. 22.

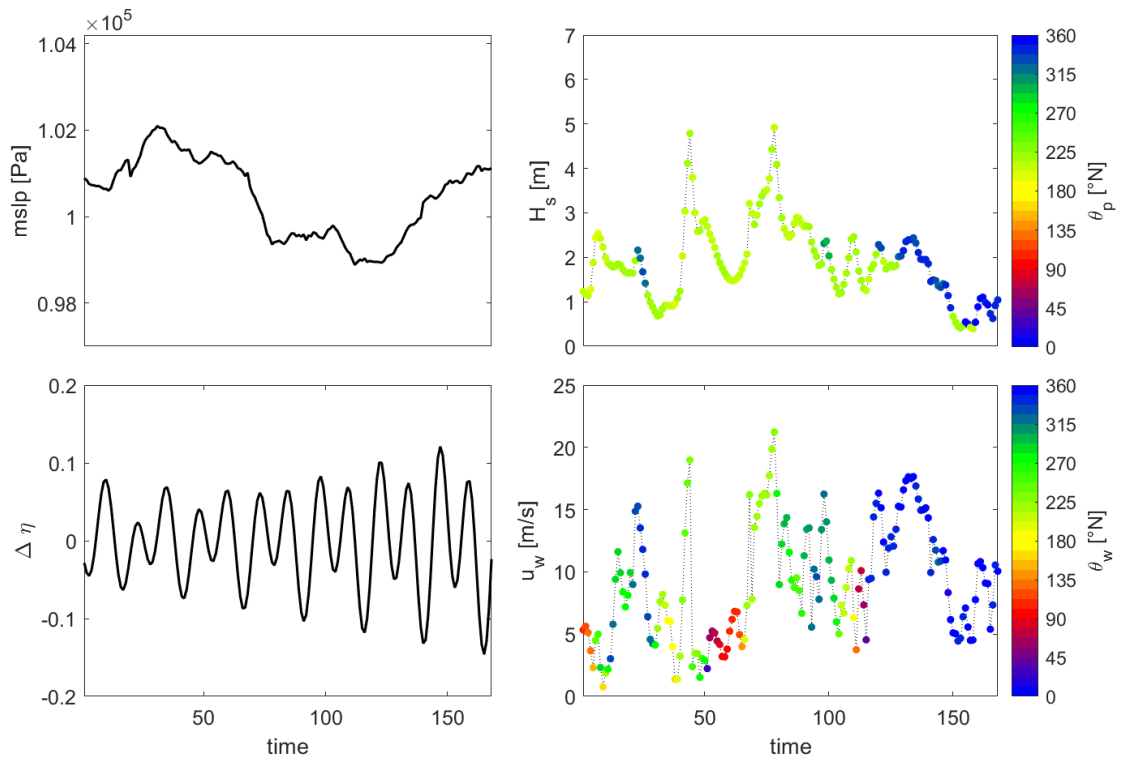


Figura 25. Scenario no. 23.

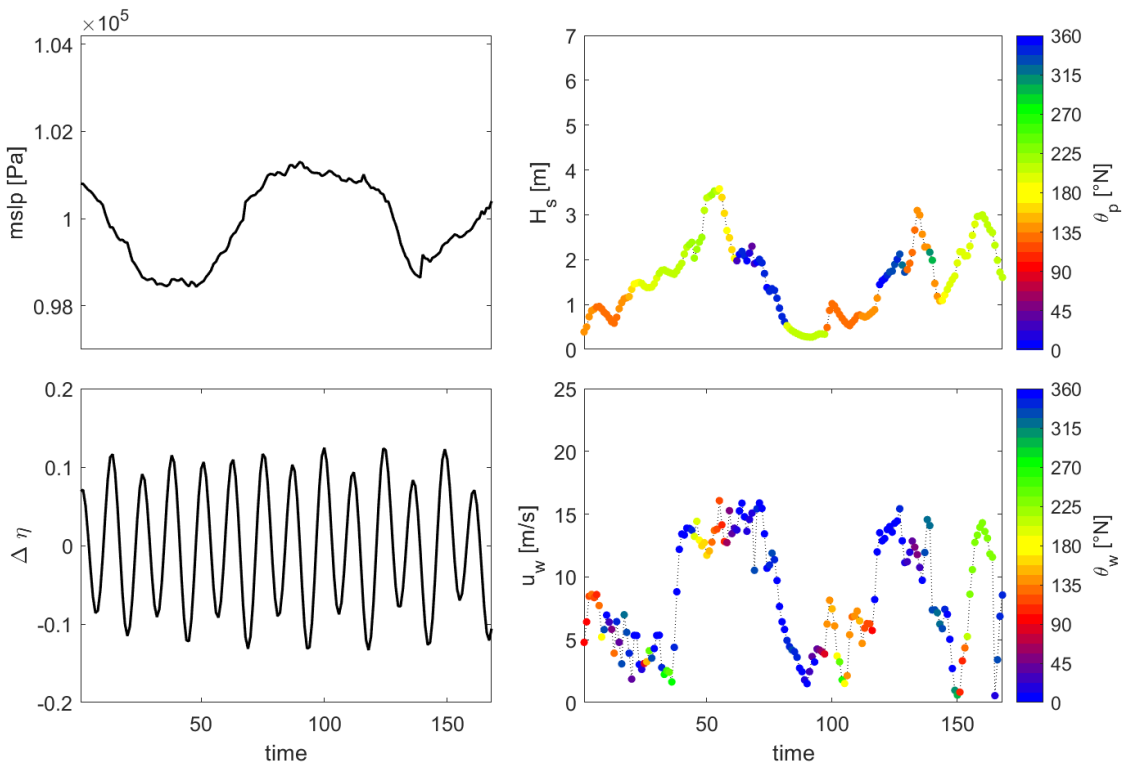


Figura 26. Scenario no. 24.

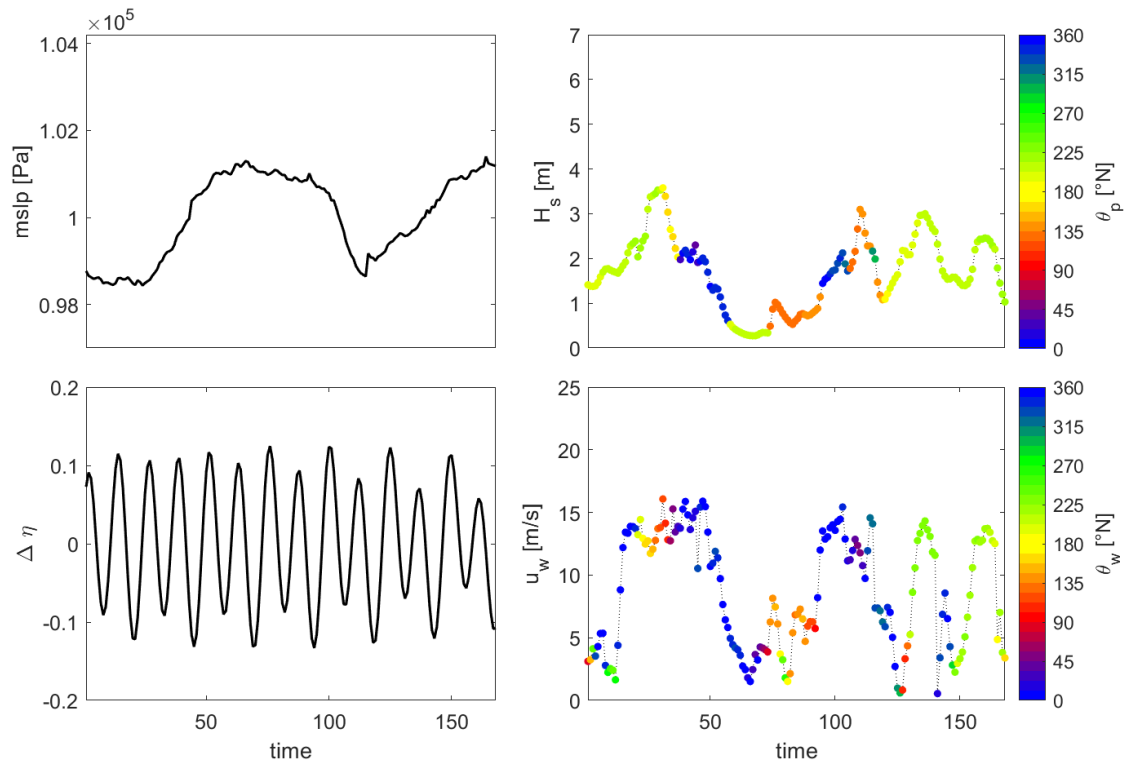


Figura 27. Scenario no. 25.

Le analisi sin qua introdotte permettono di analizzare subset caratterizzati da variabilità temporale all'interno delle finestre di riferimento selezionate, e sono dunque particolarmente efficaci qualora i processi da modellare siano fortemente non-stazionari. Tuttavia, gli algoritmi di selezione permettono di indentificare scenari significativi per singoli nodi, mentre i segnali estratti su molteplici nodi risultano essere generalmente "rumorosi".

Quando la variabilità spaziale del fenomeno investigato assuma particolare rilievo, occorre ricorrere a metodologie che permettano l'identificazione di pattern significativi su aree estese, pur non conservando in modo esaustivo le informazioni sulla variabilità temporale del fenomeno oggetto di Studio. Uno dei metodi maggiormente diffusi in tal senso è il cosiddetto SOM (*Self Organizing Maps*) di cui si riporta nel seguito un'applicazione nella baia di Tolone.

#### **4. MODELLO DI CIRCOLAZIONE NELLA BAIJA DI TOLONE**

La tecnica SOM è utilizzata per l'estrazione di pattern di circolazione nell'area del Golfo del Leone, avvalendosi dei dati prodotti con apposito modello numerico nestato su tre maglie computazionali, attualmente operativo con copertura dal Golfo del Leone alla Baia di Tolone.

Il metodo di apprendimento non supervisionato SOM consente di ridurre un database multidimensionale ad una rappresentazione (in genere 2D), mantenendone la struttura topologica di partenza. Il metodo prevede due fasi (addestramento e mappatura): in pratica, si sviluppa una mappa auto-adattiva al fine di rappresentare un insieme di dati tramite un determinato numero di neuroni; ogni neurone è deputato alla rappresentazione di un particolare gruppo di dati in base alle caratteristiche comuni. Tale approccio consente una visualizzazione multidimensionale dei dati incrociati.

Nell'ambito del progetto SINAPSI, è stato applicato il metodo SOM a diversi set di dati, con l'obiettivo di ottenere una classificazione delle circolazioni nella baia di Tolone e nella zona costiera circostante. In particolar modo, si è operato sulle mappe delle correnti superficiali ottenute da HFR ([hfradar.univ-tln.fr](http://hfradar.univ-tln.fr)), elaborate a partire dai dati di vento ottenuti con il modello atmosferico ARPEGE di Météo-France. Attraverso le SOM, si è potuto ridurre un set di dati orari definiti su un intervallo di 6 mesi (giugno 2000-novembre 2000) a 6 neuroni, che rappresentano ciascuno un determinato pattern di circolazione marina associata a una forzante del vento. I neuroni (ossia, le mappe) estratte sono mostrate a titolo di esempio in Figura 28, mentre la Figura 29 mostra le cosiddette Best Matching Units (BMU), ossia la sequenza di neuroni da utilizzare per riprodurre la serie di mappe in input con il massimo grado di verosimiglianza.

Come si evince dalle figure sopra riportate, questa rappresentazione schematica permette di visualizzare in modo rapido ed efficace la circolazione marina forzata dal vento. Nei neuroni N2, N5, N6 si trovano condizioni di vento di Maestrale (da medio a forte) che generano una corrente superficiale diretta verso Ovest, Sud-Ovest. Una situazione (N3) di forte vento da est che rafforza la corrente onshore Liguro-Provenzale. Una situazione con un vento molto moderato da sud-ovest che riduce la corrente di bordo. Infine, una situazione senza vento (N1) che permette di evidenziare la corrente geostrofica (Caceres-Euse, et al. 2022).



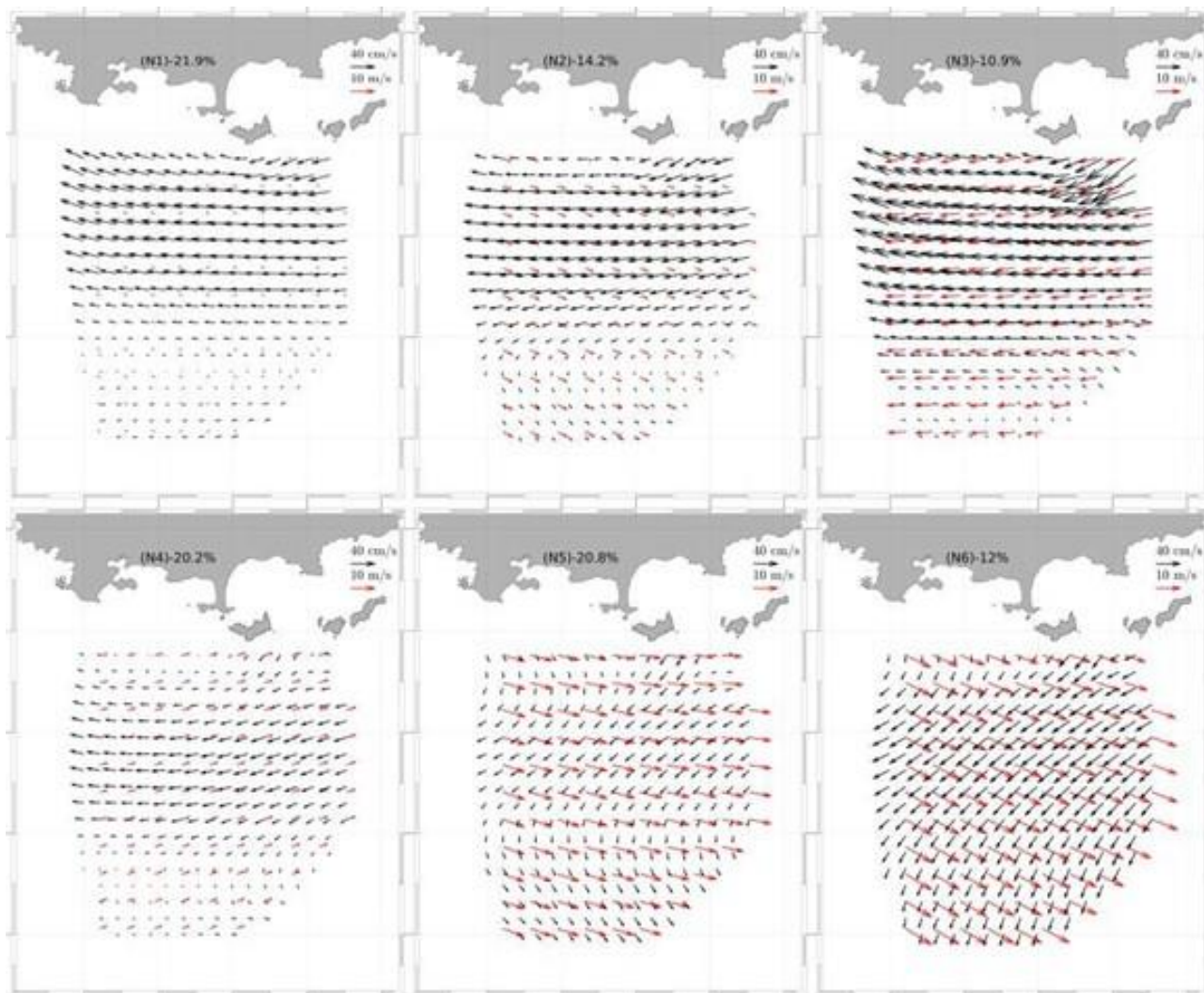


Figura 28. Risultato dell'analisi SOM 3 x 2 allegata delle correnti superficiali da HFR (nero) e del vento a 10 m (rosso) per il periodo da giugno a novembre 2020.

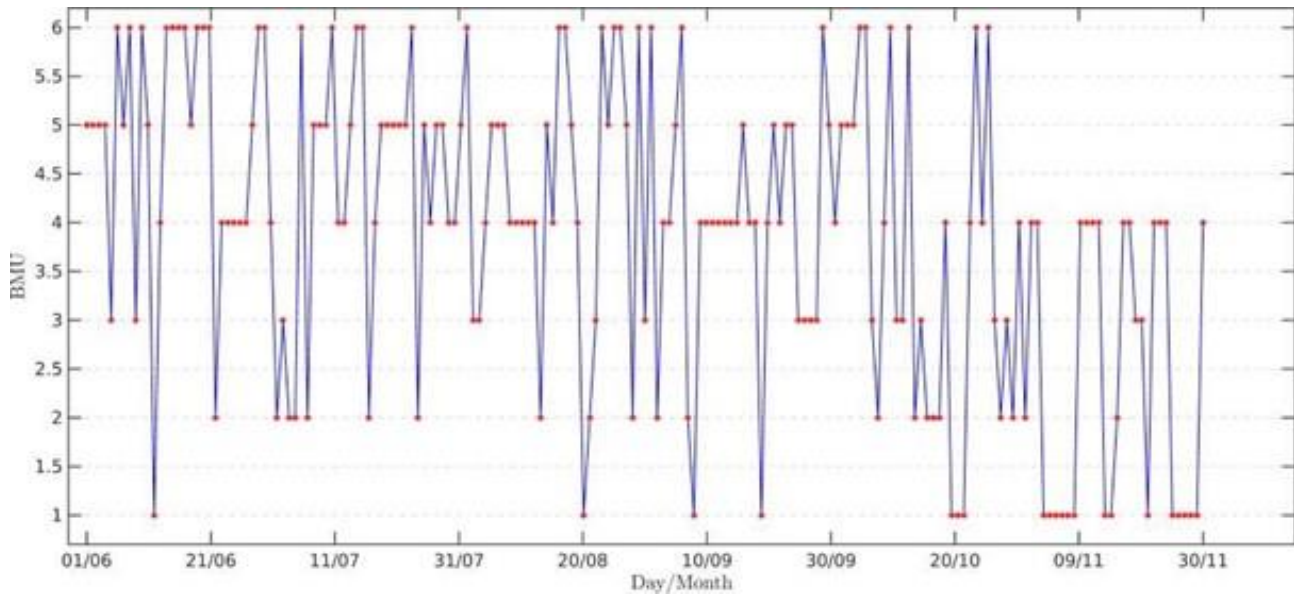


Figura 29. BMU corrispondente alla Figura 28. Ogni data corrisponde a un neurone.

## Bibliografia

Bárcena, J. F., Camus, P., García, A., Álvarez, C. "Selecting model scenarios of real hydrodynamic forcings on mesotidal and macrotidal estuaries influenced by river discharges using K-means clustering." *Environmental Modelling & Software* (68), 2015: 70-82.

Caceres-Euse, Alejandro, Anne Molcard, Natacha Bourg, Dylan Dumas, Charles-Antoine Guerin, and Giovanni Besio. "Breakdown of Near-Surface Sea Current from High-Frequency Radar Data." *Journal of Atmospheric and Oceanic Technology*, 2022: 1927-1942.

Camus, P., Mendez, F. J. , Medina, R. "A hybrid efficient method to downscale wave climate to coastal areas." *Coastal Engineering*, 2011: 851-862.