

Projet

asSistance à la Navigation pour l'Accès aux Ports en Sécurité



ACTIVITÉ T2.1: MISE EN ŒUVRE DE LA MODÉLISATION NUMÉRIQUE

PRODUIT T2.1.2: MÉTHODOLOGIE PARTAGÉE DÉFINITION DES SCÉNARIOS
DE PRÉVISION

Partner responsable: UNIGE

Partenaires contributeurs: UTLN, CNR-ISMAR, LaMMA

SOMMAIRE

| | |
|--|-----------|
| <i>Description du produit</i> | 1 |
| 1. INTRODUCTION | 1 |
| 2. REGROUPEMENT DE DONNÉES MÉTÉOROLOGIQUES ET MARINES | 2 |
| 3. APPLICATION À UNE ÉTUDE DE CAS | 5 |
| 4. MODÈLE DE CIRCULATION DANS LA BAIE DE TOULON | 20 |
| Bibliographie | 22 |

Description du produit

Ce rapport résume les techniques d'exploration de données utilisées pour la sélection des scénarios météo-marins, utiles pour représenter la climatologie des zones d'étude et fournir la référence pour les prévisions de données météo-marines à court terme dans le cadre du projet SINAPSI.

1. INTRODUCTION

Ces dernières années, l'analyse des processus physiques dans les zones côtières a fait de plus en plus appel à des modèles numériques capables de résoudre les équations complexes liées à l'hydrodynamique et à la morphodynamique près de la côte. Ces modèles simulent les processus de dispersion à la fois d'un point de vue eulérien (dynamique de la concentration) et lagrangien (dispersion de la masse et des objets), permettant ainsi la reproduction détaillée des processus physiques qui se déroulent sur le littoral ; cependant, leur utilisation nécessite des ressources informatiques élevées (c'est-à-dire des supercalculateurs), qui ne sont pas toujours disponibles.

L'étude de l'hydrodynamique côtière repose en effet sur de grandes quantités de données qui, dans le cas des produits de réanalyse (les *hindcasts*), s'étendent sur de longs intervalles de temps. Dans ce cas, il est nécessaire de réduire le nombre de conditions environnementales à utiliser pour la modélisation afin d'être en mesure d'identifier les modes de variabilité les plus significatifs du phénomène étudié. La résolution d'un nombre limité de conditions environnementales, également appelées "scénarios", est avantageuse car elle permet non seulement de sélectionner les états météorologiques les plus importants pour le processus analysé, mais aussi de réduire considérablement la charge de calcul nécessaire pour résoudre l'ensemble de la chaîne de modélisation.

À cette fin, il est possible d'utiliser des algorithmes de clutérisation des données, qui permettent de décomposer les ensembles de données de départ en un petit nombre de sous-ensembles, chacun représenté par un centroïde de référence (les scénarios mentionnés ci-dessus). Bien que l'utilisation de tels algorithmes ne se soit répandue que récemment dans le domaine de l'ingénierie maritime et côtière, elle a connu une croissance soudaine dans la bibliographie du secteur, car elle permet une description détaillée de la variabilité des conditions météorologiques-marines d'un site en se concentrant sur un nombre réduit de scénarios.

PRODUIT N. T2.1.2

Dans le cadre du projet SINAPSI, le Département d'ingénierie civile, chimique et environnementale de l'Université de Gênes (ci-après DICCA) a développé une série de scripts pour le regroupement de variables météorologiques-marines à partir de produits hindcast (donc extrêmement importants en termes de nombre de variables stockées) et la définition de scénarios variant dans le temps. Ces scénarios seront ensuite utilisés pour caractériser l'hydrodynamique dans les zones pilotes du projet SINAPSI à l'aide de modèles de calcul appropriés. Les algorithmes qui sous-tendent le regroupement sont expliqués au chapitre 2, tandis que le chapitre 3 présente une étude de cas illustrant la procédure de calcul pour la sélection des scénarios météo-maritimes de référence pour un site pilote du projet.

2. REGROUPEMENT DE DONNÉES MÉTÉOROLOGIQUES ET MARINES

Les techniques de clustering sont conçues pour identifier des groupes de données qui sont homogènes entre eux, mais hétérogènes par rapport aux éléments des autres groupes identifiés. Dans le contexte du projet SINAPSI, le but ultime de ces techniques est la sélection de scénarios capables de représenter avec une approximation raisonnable la variabilité totale des données sources, ceux-ci étant ensuite utilisés comme entrée pour les modèles numériques employés.

Considérons un ensemble de données composé d'un nombre V de variables et de n éléments. Suivant l'organigramme proposé dans (Camus 2011) la première étape du clustering d'un tel jeu de données est de l'organiser sous forme matricielle, en normalisant les différents éléments dans un espace commun colonne par colonne :

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

2.1

Chaque vecteur colonne de la matrice de l'équation 2.1 représente une variable de l'ensemble de données qui a été normalisée de manière appropriée, par exemple en centrant les données par rapport à la moyenne puis en les mettant à l'échelle de l'écart type, comme indiqué dans l'équation 2.2.

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

2.2

où μ et σ représentent respectivement la moyenne et l'écart-type de la jème colonne de la matrice $X_{n,V}$. Cette opération est nécessaire car elle permet à l'algorithme de clustering de traiter des données négatives et positives, caractérisées par des ordres de grandeur différents, en évitant que les variables les plus pertinentes n'altèrent les calculs ultérieurs.

Si des scénarios variant dans le temps doivent être sélectionnés, il est d'abord nécessaire de restructurer les données sources (Bárcena 2015) ; après avoir sélectionné un nombre nt d'instant, le vecteur unique x d'une variable assignée est modélisé sous forme de matrice comme suit :

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

2.3

Où δ représente le décalage entre deux instants successifs (c'est-à-dire le nombre de pas de temps entre les instants initiaux de deux sous-ensembles successifs x'). Le clustering est donc effectué sur la matrice résultante X' , dans laquelle chaque ligne définit une fenêtre de longueur nt pour une variable assignée.

L'équation 2.4 résume les procédures introduites ci-dessus.

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

2.4

Lorsque l'on travaille avec plusieurs paramètres, il suffit de reproduire les étapes décrites dans les équations 2.3 et 2.4 pour chacun d'eux, puis de concaténer les matrices résultantes en une matrice de taille totale $[n-nt+1] \times [nt \times V]$. La sélection des

PRODUIT N. T2.1.2

scénarios se fait ensuite sur cette dernière matrice en utilisant l'algorithme de dissimilarité maximale (MDA). Pour faciliter la compréhension de cet algorithme, il est fait référence à la matrice de l'équation 2.1.

L'objectif du MDA est de sélectionner un sous-ensemble de la matrice initiale $X_{n,V}$, appelé $X_{m,V}^*$ (matrice des états significatifs, $m < n$) composé des états qui représentent le mieux la variance globale des données. Après avoir normalisé les données colonne par colonne, afin de pouvoir travailler avec des variables caractérisées par différents ordres de grandeur, un premier état significatif est identifié en sélectionnant l'état le plus éloigné du nuage de points initial (c'est-à-dire celui qui est le plus "dissemblable" du reste de l'ensemble de données). On procède ensuite au calcul de la dissimilarité entre les vecteurs de la matrice $X_{n-1,V}$ et le sous-ensemble $X_{1,V}^*$ en calculant la distance euclidienne ainsi définie :

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

2. 5

où x et x^* indiquent les lignes des matrices d'entrée et de sortie respectivement. Le nouvel élément est alors sélectionné comme étant celui caractérisé par la valeur maximale de d_{i1} et ajouté à la matrice cible X^* . La version MaxMin de l'algorithme est appliquée : en effet, à la k^e itération ($k < m$, où k est le nombre d'éléments de la matrice X^*), la distance à considérer pour chaque élément de $X_{(n-k),V}$ est le minimum par rapport à tous les k vecteurs de $X_{k,V}^*$; parmi toutes ces distances, celle qui est maximale est sélectionnée et l'état correspondant est par conséquent ajouté à X^* . Le calcul se termine lorsque k devient égal à m , c'est-à-dire lorsque le nombre de clusters (à déterminer au préalable) est atteint. A la fin du processus de construction de la matrice X^* , on obtient un sous-ensemble de vecteurs qui permet de résumer l'ensemble de données de départ avec un plus petit nombre d'états : les éléments restants sont affectés au vecteur modèle correspondant le plus proche, formant ainsi de véritables classes d'états.

La même procédure de calcul s'applique quelle que soit la taille de la matrice de départ, elle peut donc être étendue sans perte de généralité aux matrices composites en concaténant les sous-matrices dans l'équation 2.4

3. APPLICATION À UNE ÉTUDE DE CAS

La méthodologie introduite dans le chapitre précédent a été utilisée pour sélectionner les scénarios météorologiques au large de la côte de Gênes, en utilisant les données hindcast fournies par DICCA (<http://www3.dicca.unige.it/meteocean/hindcast.html>). Le hindcast susmentionné fournit des séries temporelles horaires de paramètres de vagues et anémométriques sur la période 1979-2020 sur une grille régulière étendue à toute la mer Méditerranée à une résolution d'environ 10 km en longitude et en latitude ; à partir de 2021, les données sont mises à jour sur une grille non structurée à résolution variable. Le nœud utilisé dans cette étude est le point 000203, représenté sur la figure 1, tandis que le tableau 1 indique ses coordonnées.

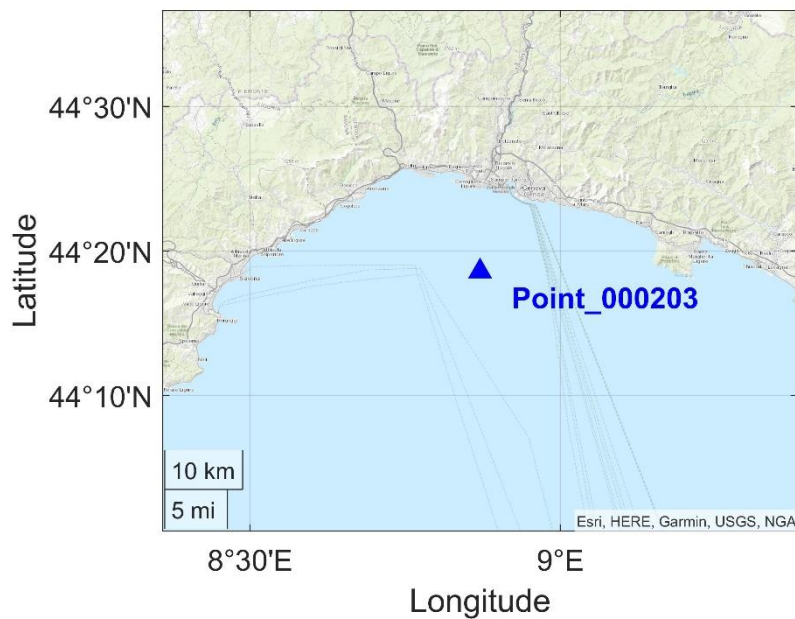


Fig 1. Nœud de Hindcast utilisé pour l'étude.

Tableau 1. Coordonnées du nœud de Hindcast considérées (SR: WGS84).

| POINT | LON | LAT |
|--------|--------|-------|
| 000203 | 8.8707 | 44.31 |

Les paramètres météorologiques-marins considérés sont la hauteur significative des vagues (H_s), la période et la direction du pic (T_p et θ_p respectivement ; ces derniers étant exprimés en convention nautique, c'est-à-dire que les vagues caractérisées par θ_p égal

PRODUIT N. T2.1.2

à 0°/360° se déplacent du nord au sud ; les vagues caractérisées par θ_p égal à 180° se déplacent du sud au nord), la vitesse du vent longitudinal et latitudinal (v_x/v_y). Aux scénarios sélectionnés, les excursions de marée respectives ($\Delta\eta$) ont été associées à l'aide du logiciel de calcul TPXO (<https://www.tpxo.net/>).

Pour la sélection des scénarios, des fenêtres temporelles d'une semaine ($nt = 168$, à noter que les données ont une résolution horaire), avec un décalage δ de 3 jours, ont été prises comme référence. Les scénarios extraits selon la méthodologie introduite au chapitre 2 sont présentés dans les figures de 3 à 27 , et montrent comment ils sont capables de reproduire les principaux modes du climat local des vagues. Ce résultat peut être apprécié en analysant les diagrammes polaires de H_s en fonction des différentes classes d'intensité (figure 2), et en comparant les directions dominantes avec celles associées aux scénarios ci-dessus (c'est-à-dire Libeccio, Scirocco, Grecale).

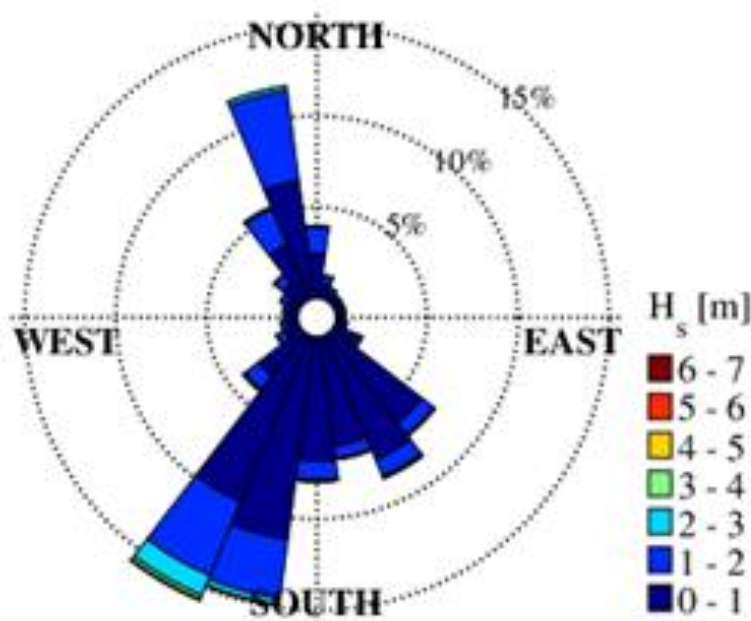


Fig. 2. Digramme polaire des ondes au point 000203. Les directions sont exprimées en convention nautique.

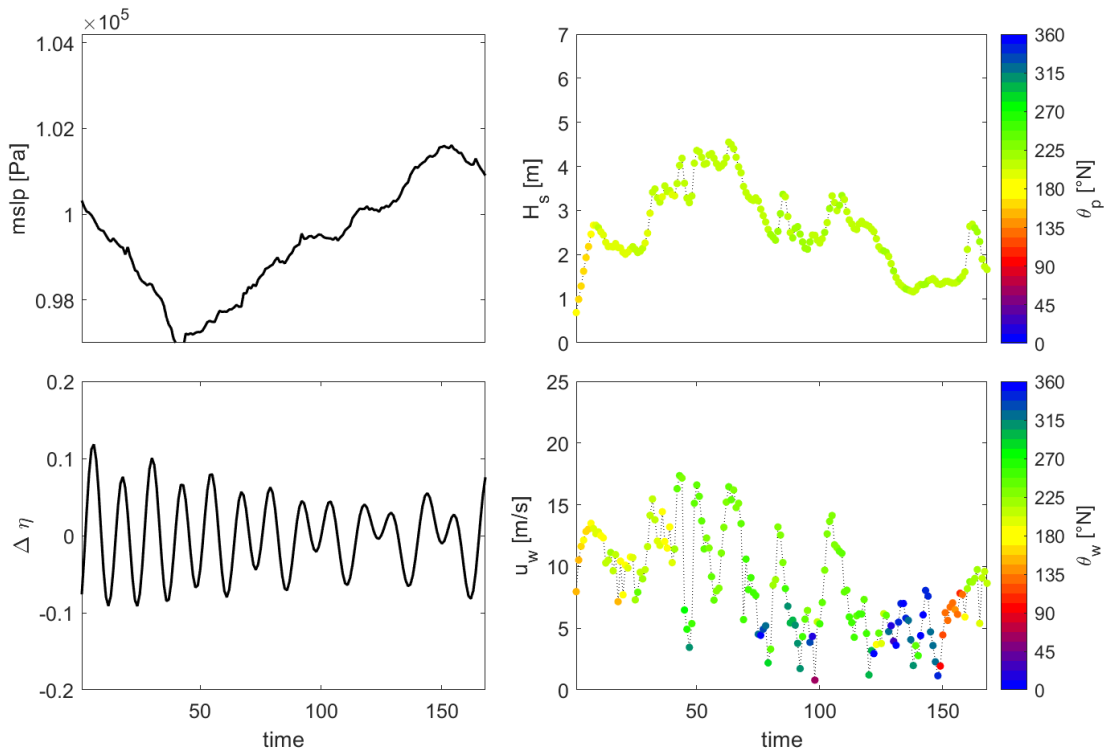


Fig. 3 Scénario no. 1. La direction du vent est indiquée par θ_w (notation similaire dans les figures suivantes.).

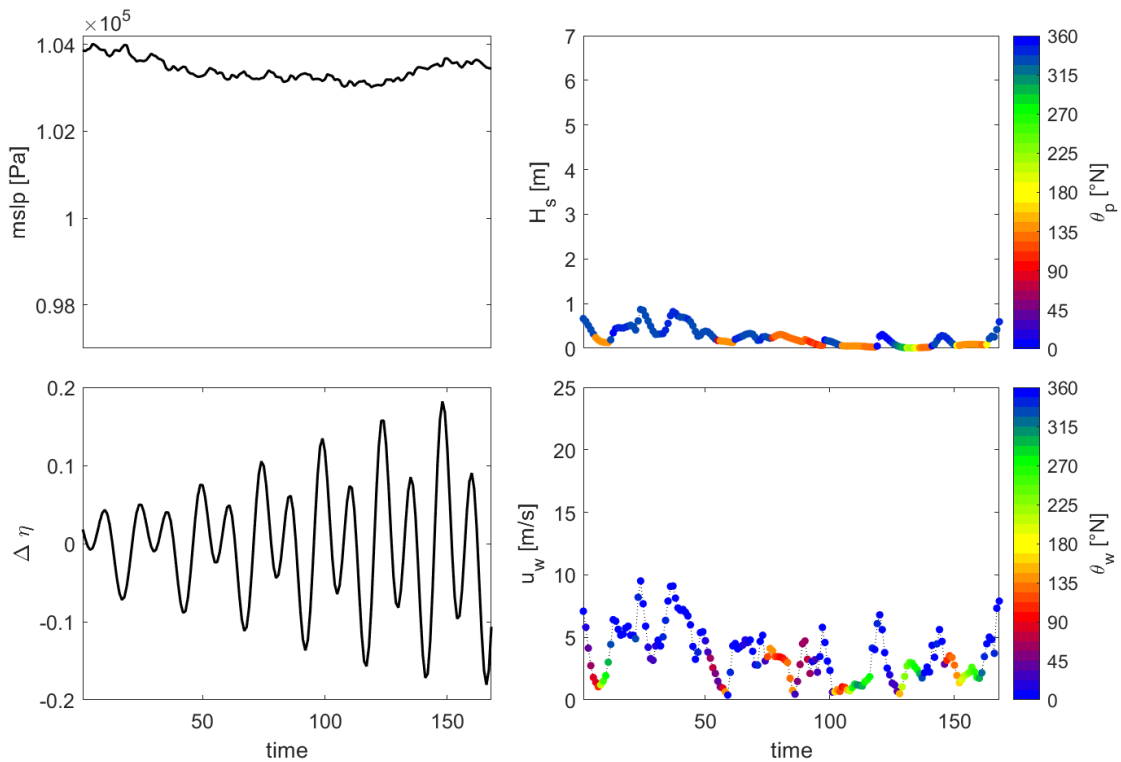


Fig. 4. Scénario no. 2.

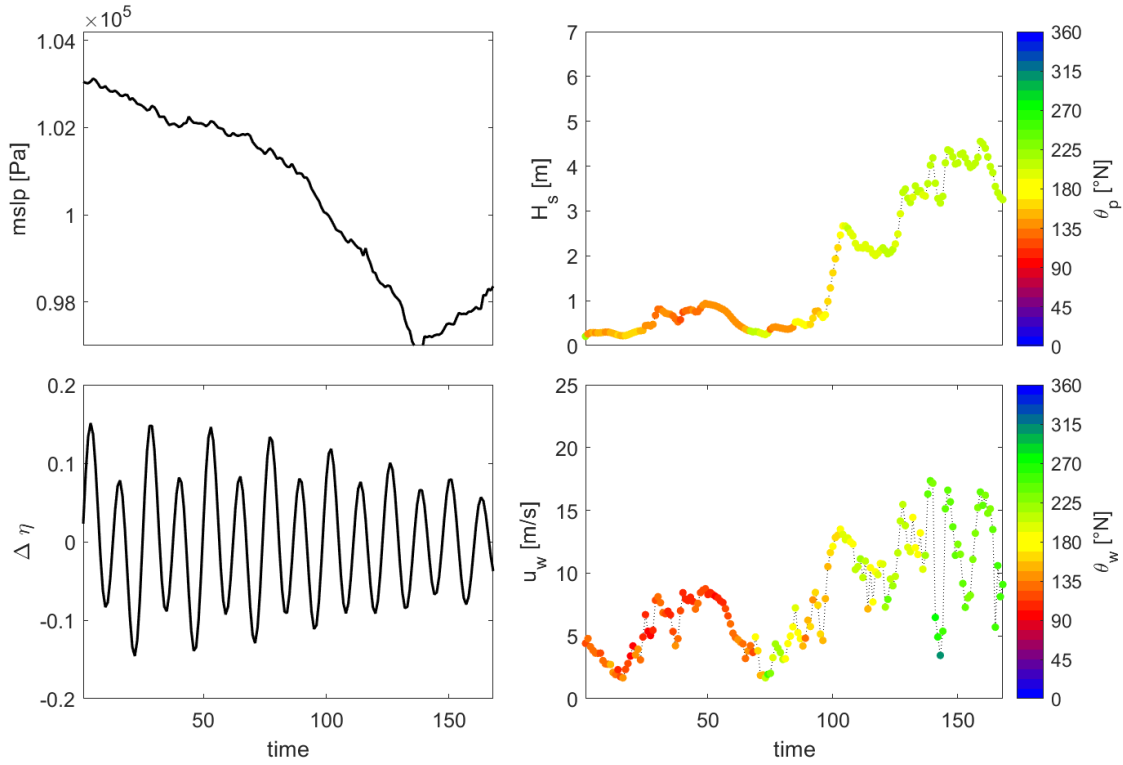


Fig. 5. Scénario no. 3.

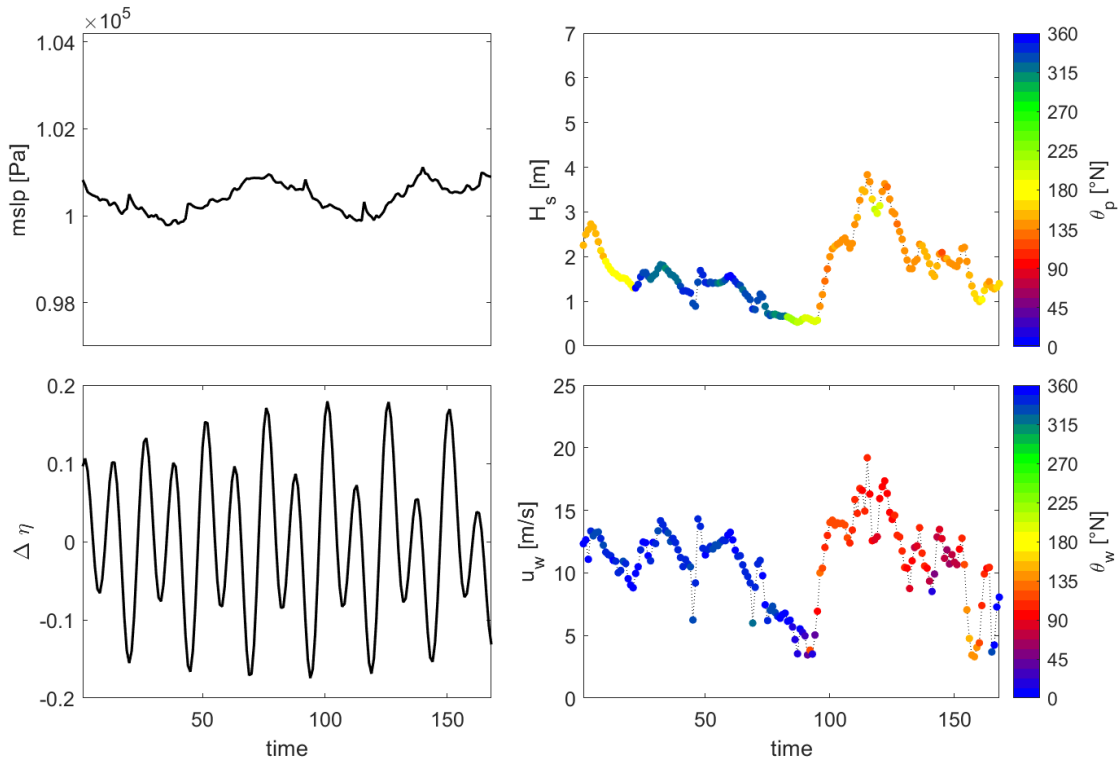


Fig. 6. Scénario no. 4.



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

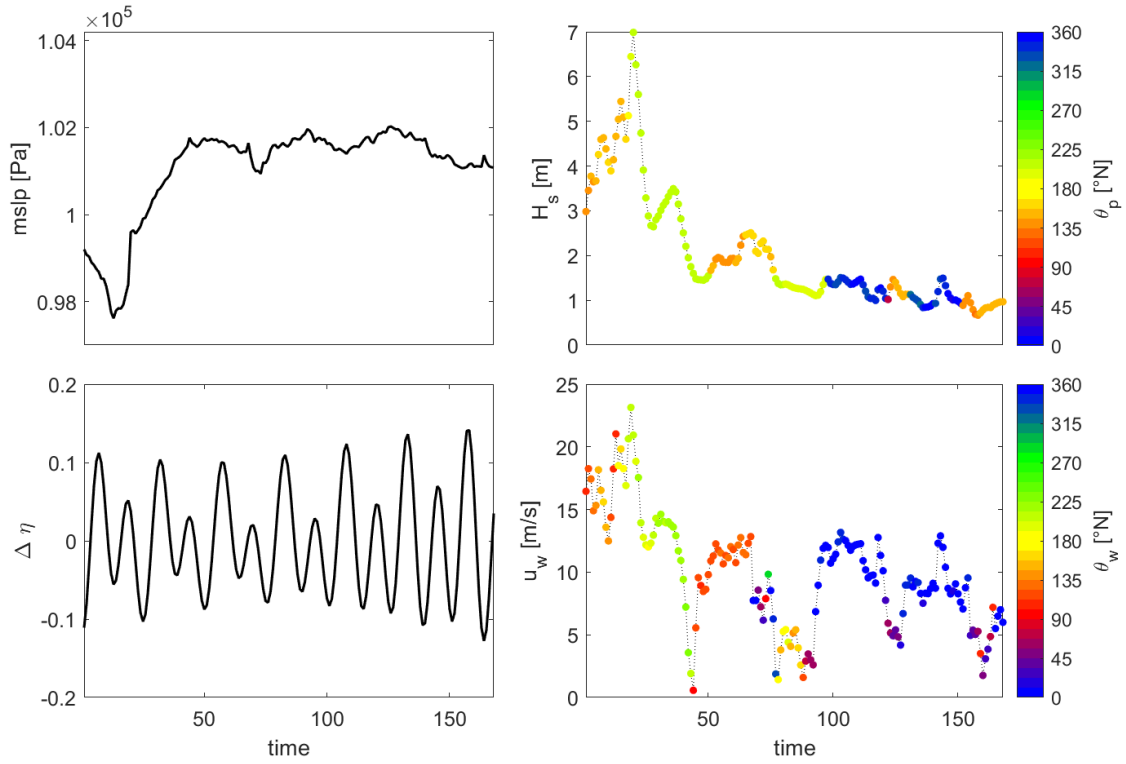


Fig. 7. Scénario no. 5.

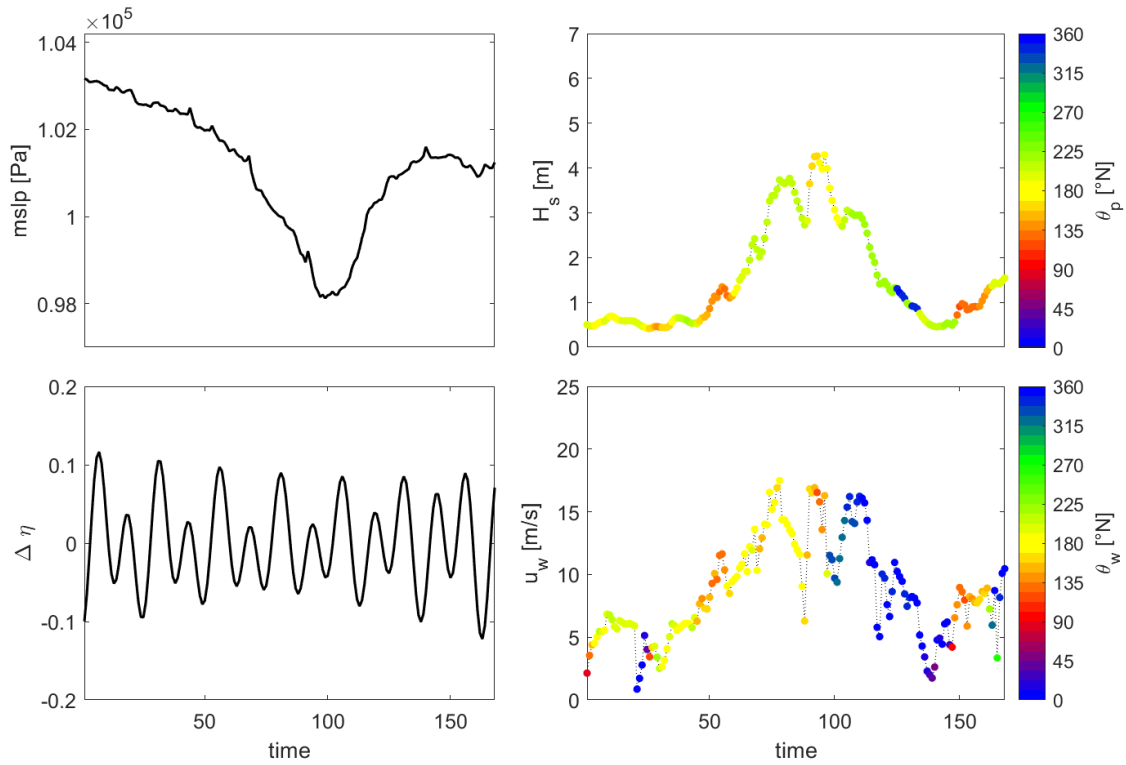


Fig. 8. Scénario no. 6.



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

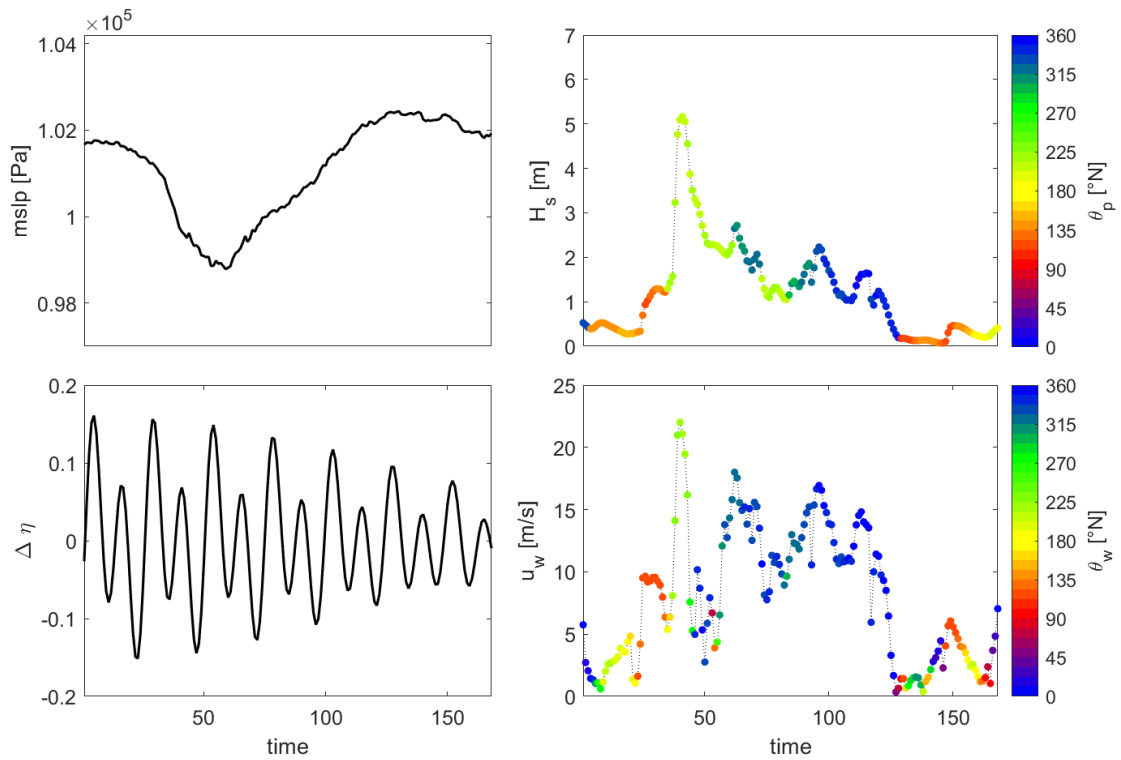


Fig. 9. Scénario no. 7.

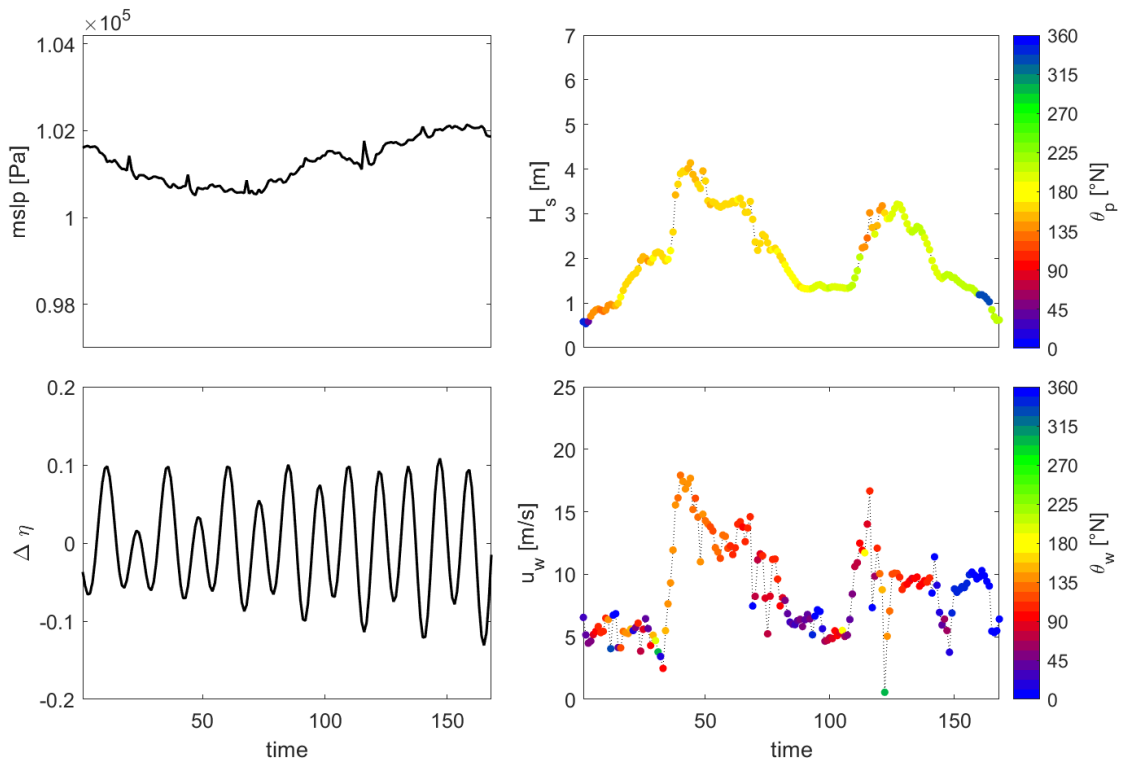


Fig. 10. Scénario no. 8.



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

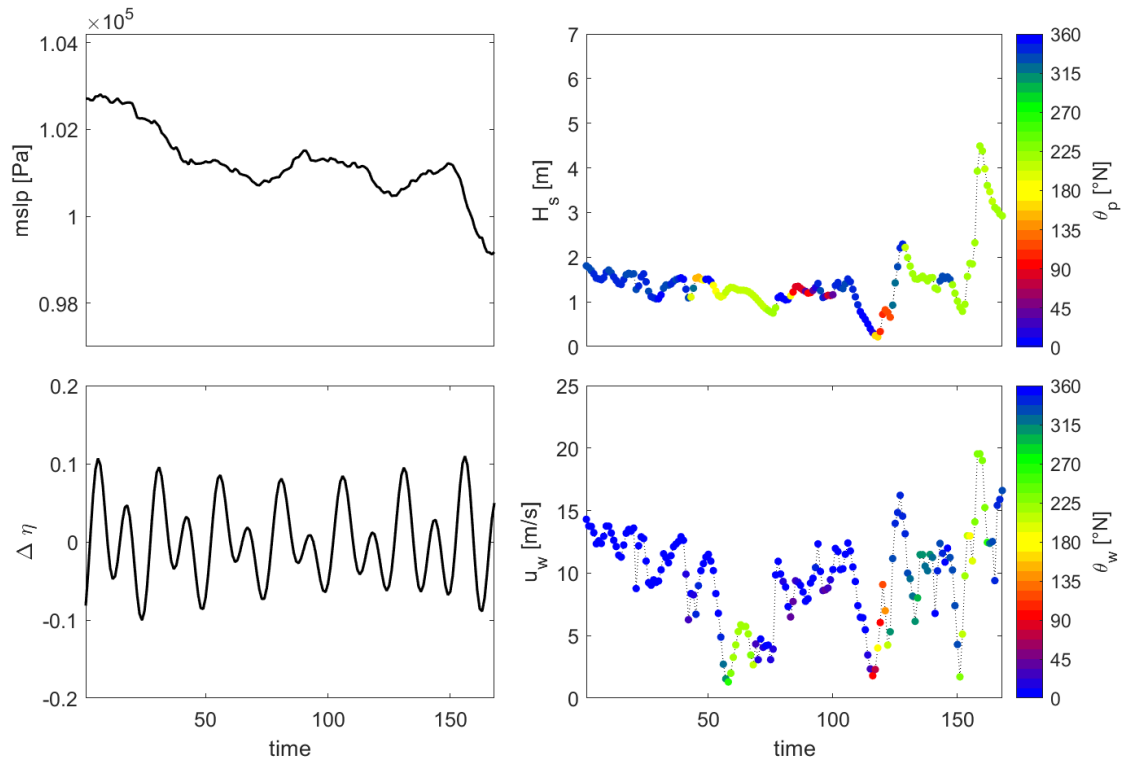


Fig. 11. Scénario no. 9.

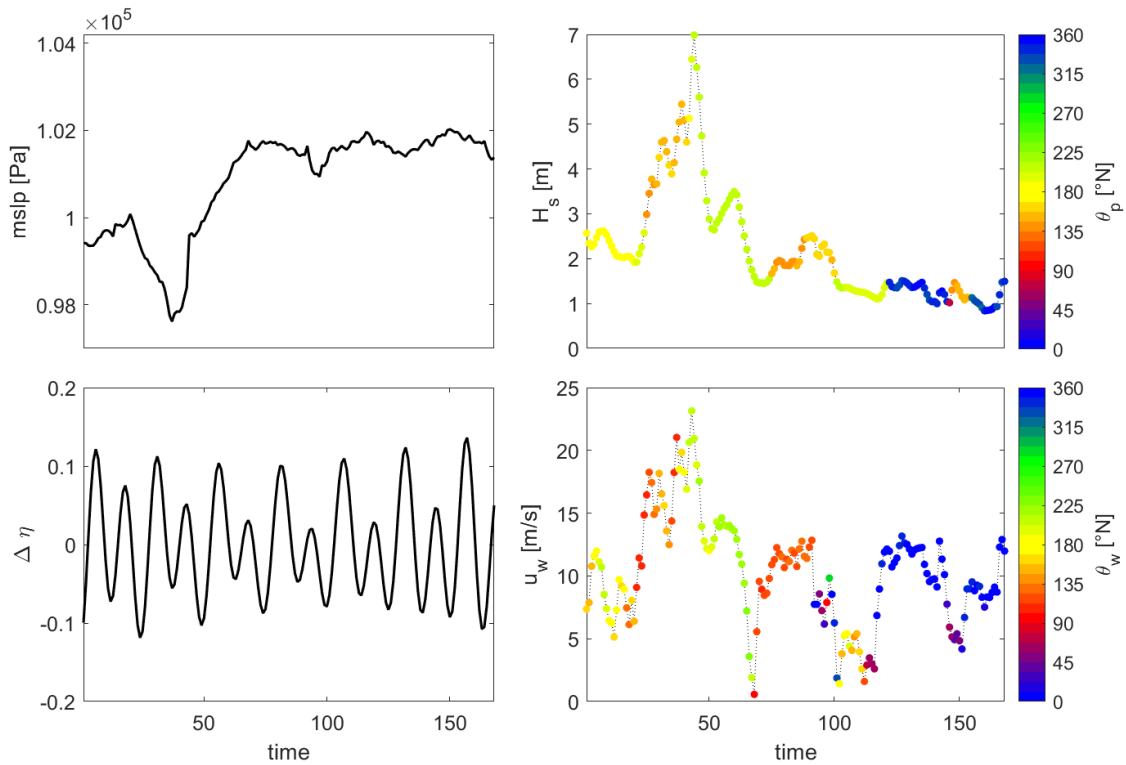


Fig. 12. Scénario no. 10.



PRODOTTO N. T2.1.2

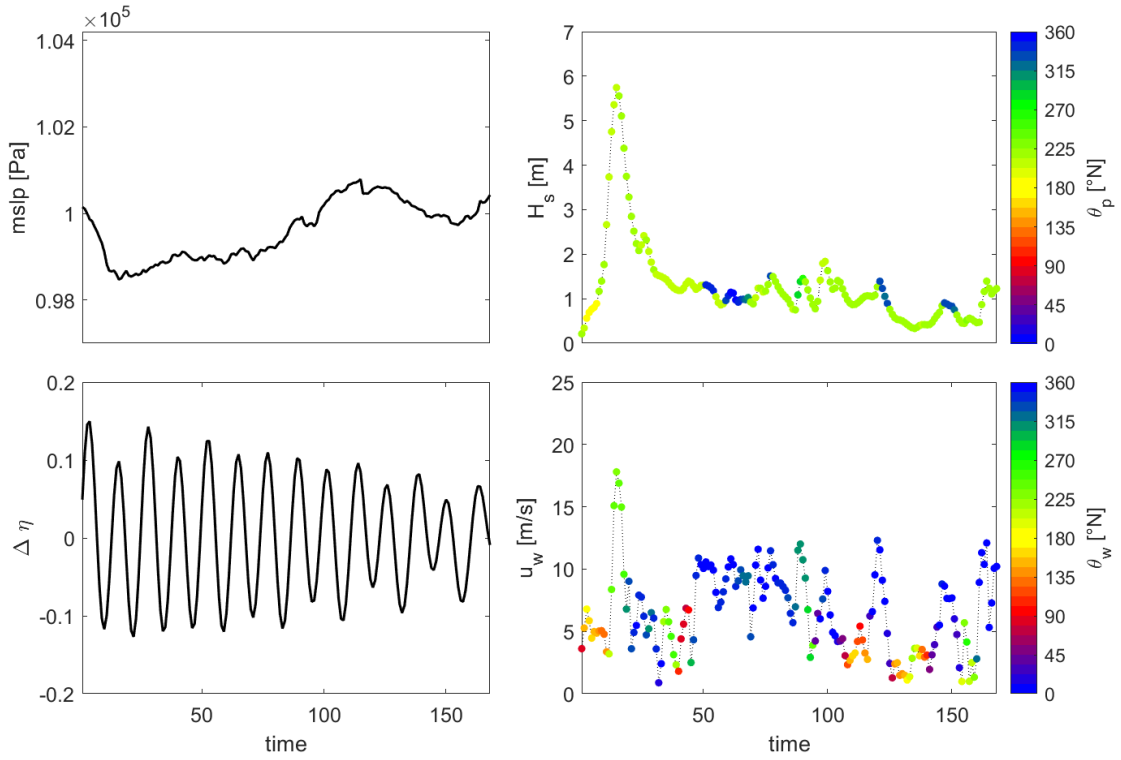


Fig. 13. Scénario no. 11.

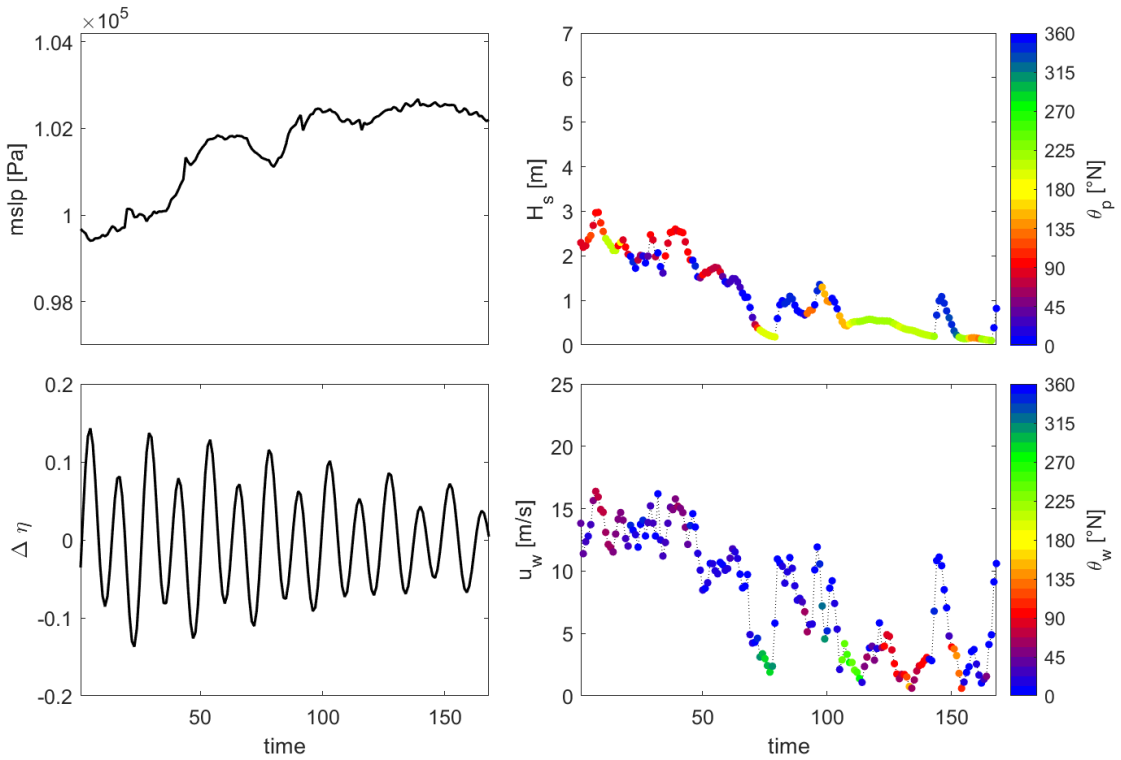


Fig. 14. Scénario no. 12.

PRODOTTO N. T2.1.2

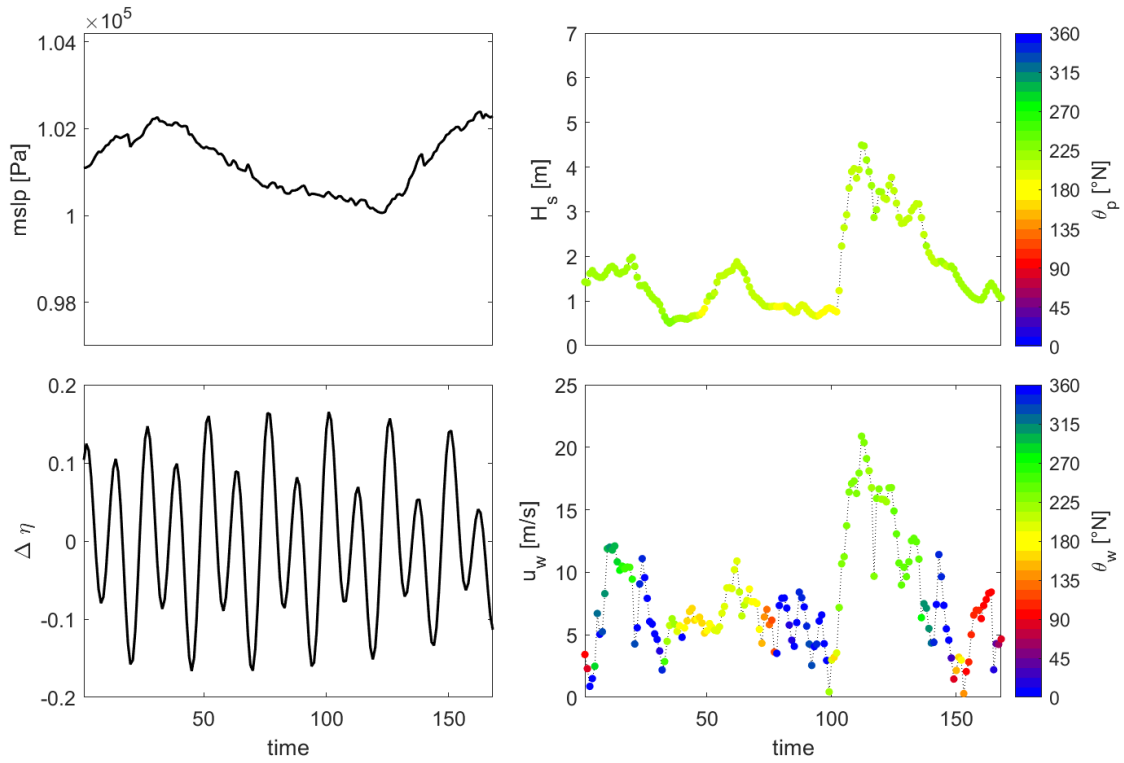


Fig. 15. Scénario no. 13.

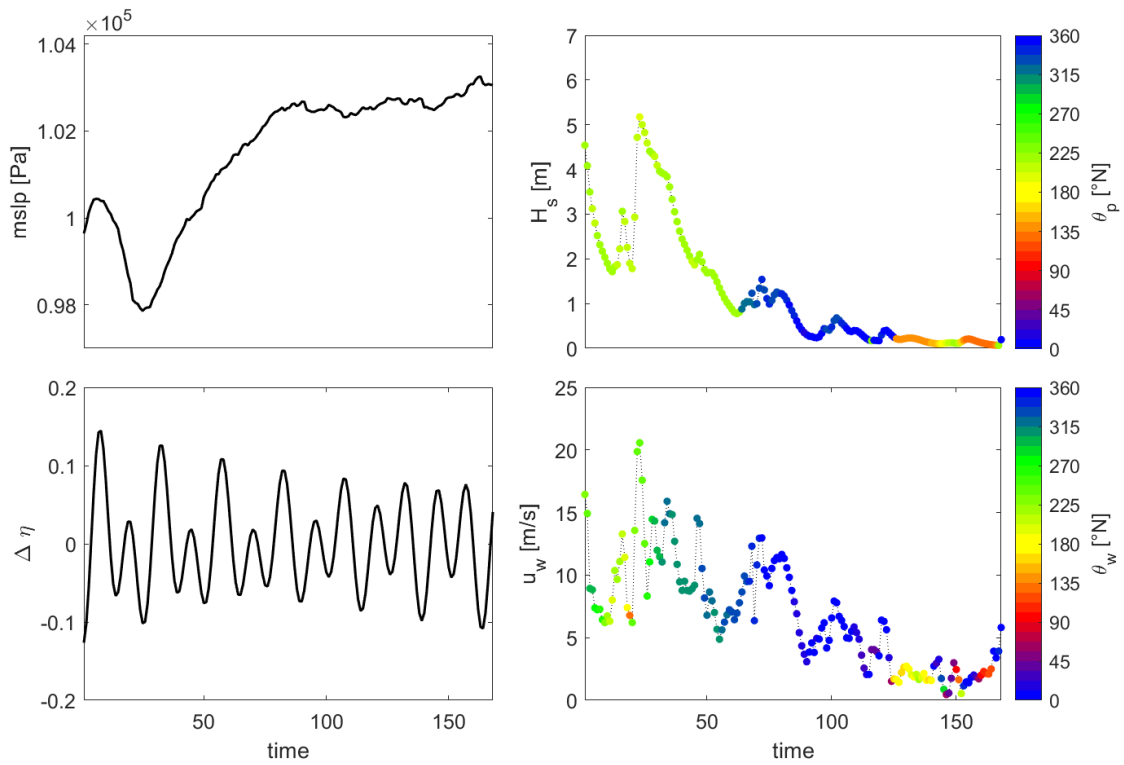


Fig. 16. Scénario no. 14.



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

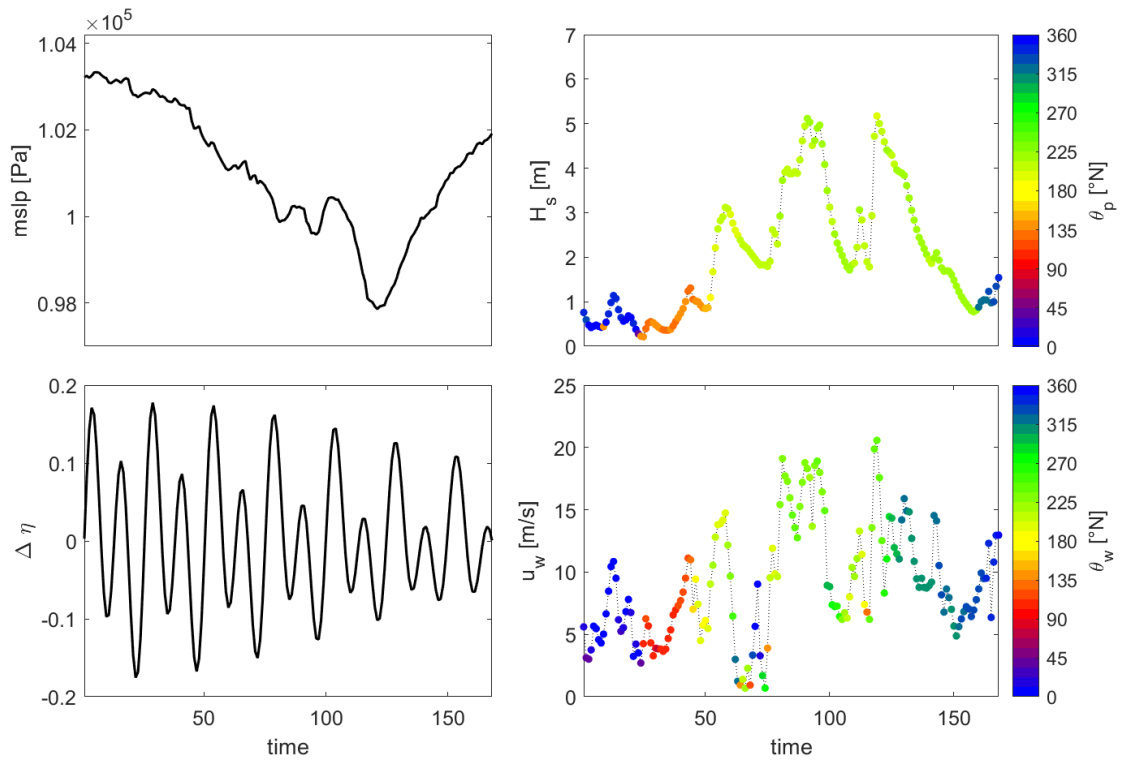


Fig. 17. Scénario no. 15.

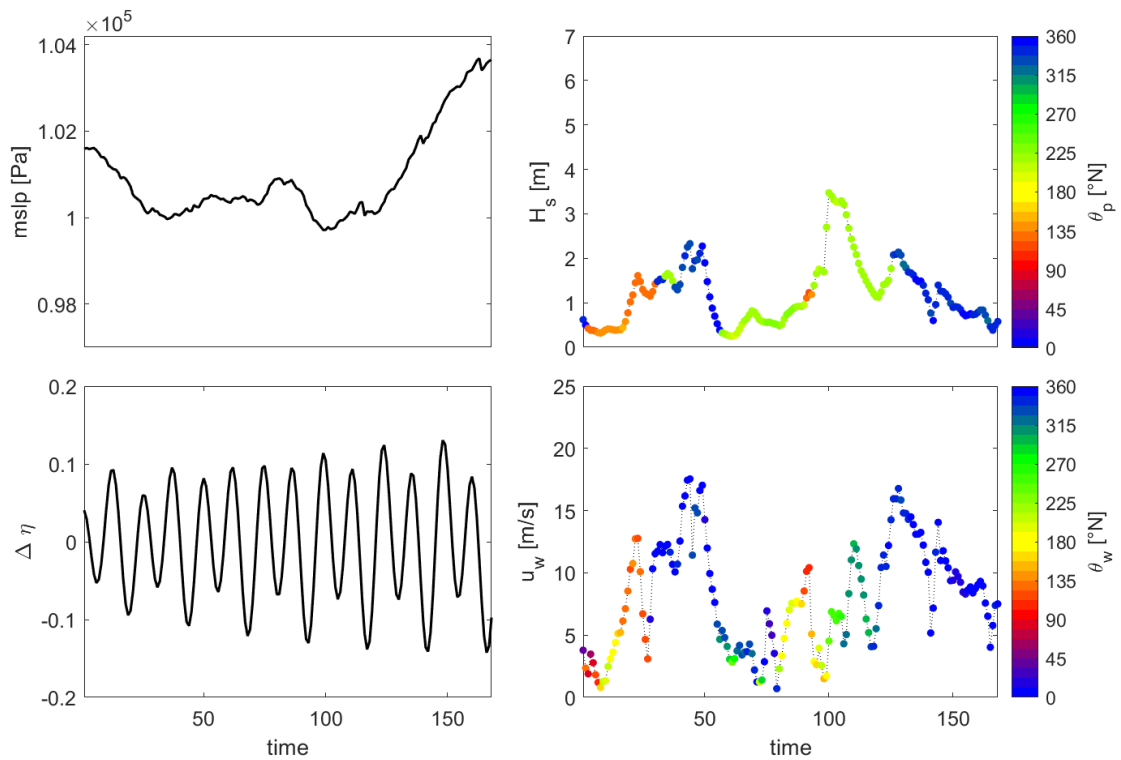


Fig. 18. Scénario no. 16.



PRODOTTO N. T2.1.2

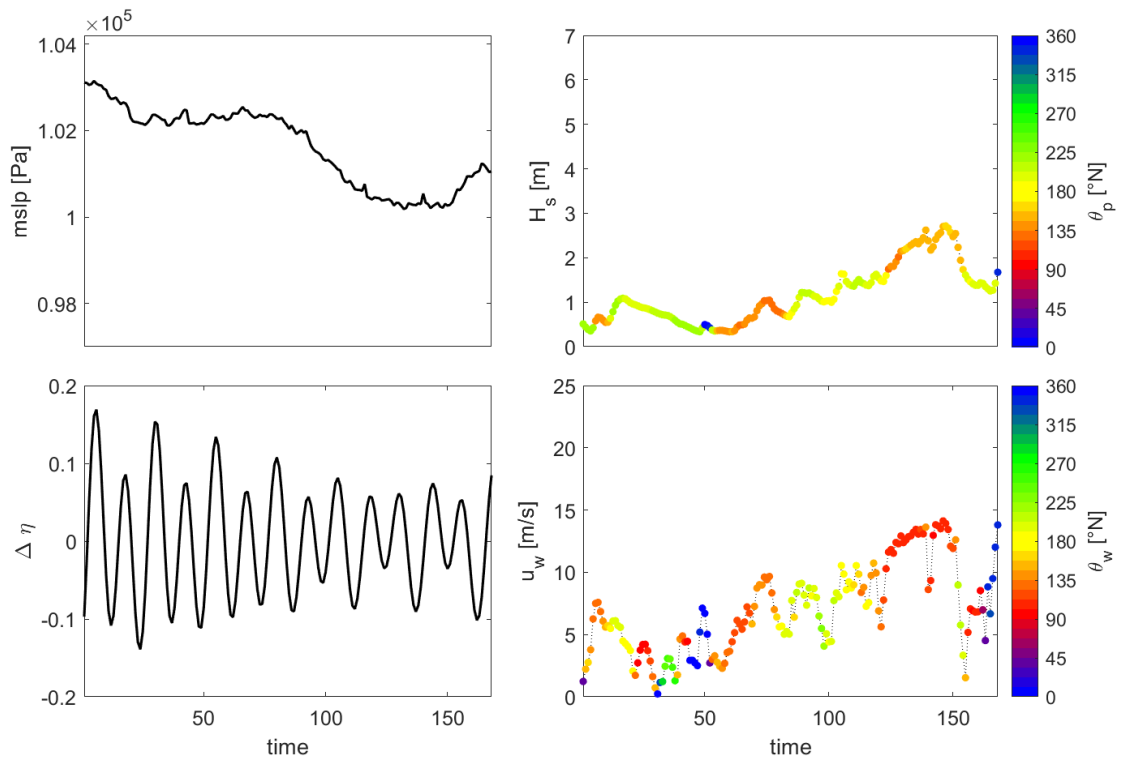


Fig. 19. Scénario no. 17.

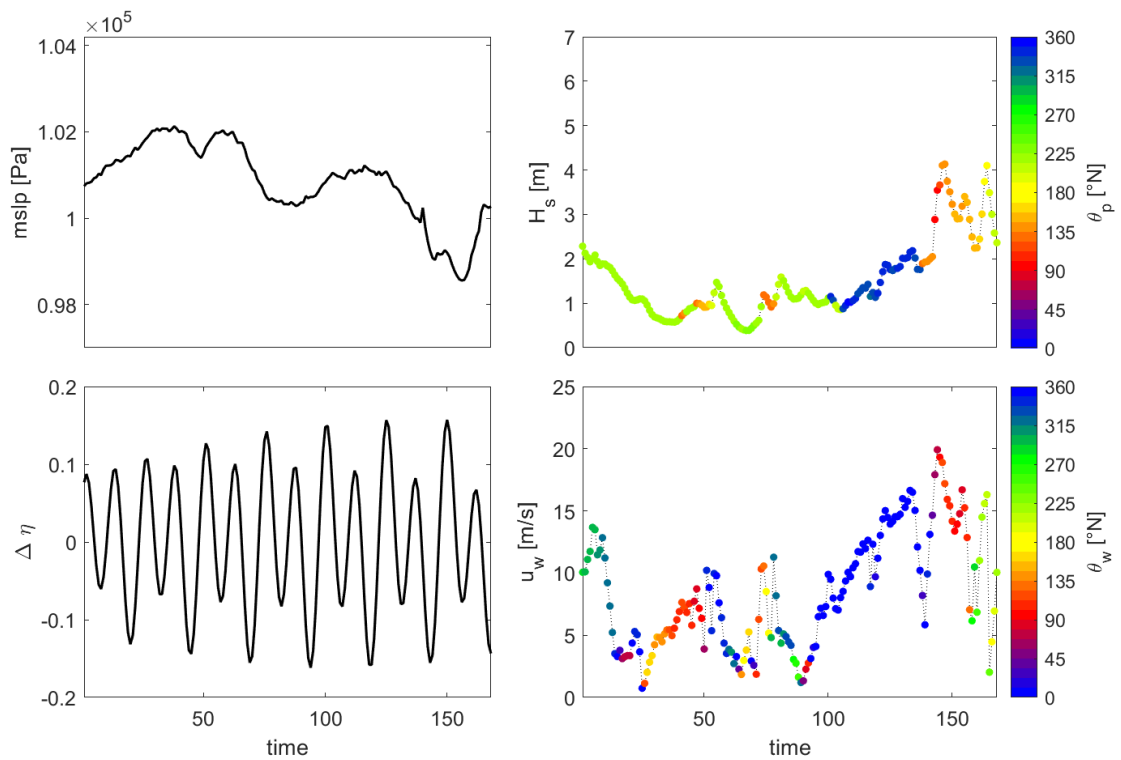


Fig. 20. Scénario no. 18.

PRODOTTO N. T2.1.2

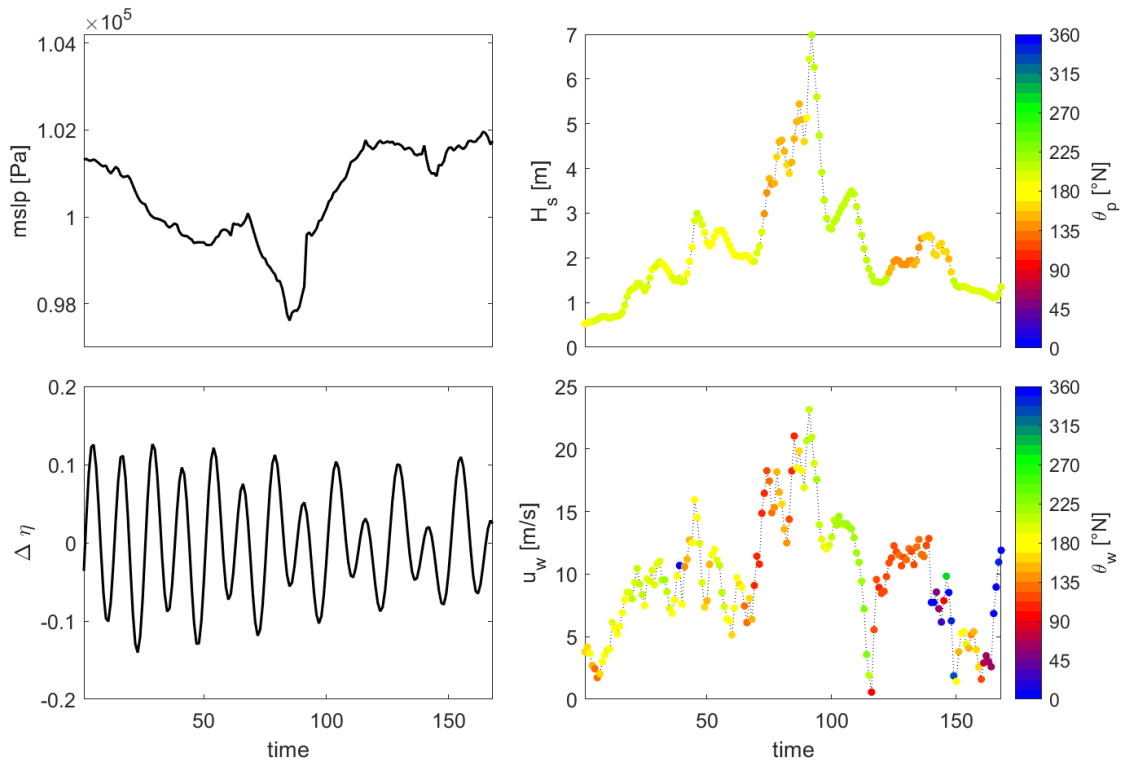


Fig. 21. Scénario no. 19.

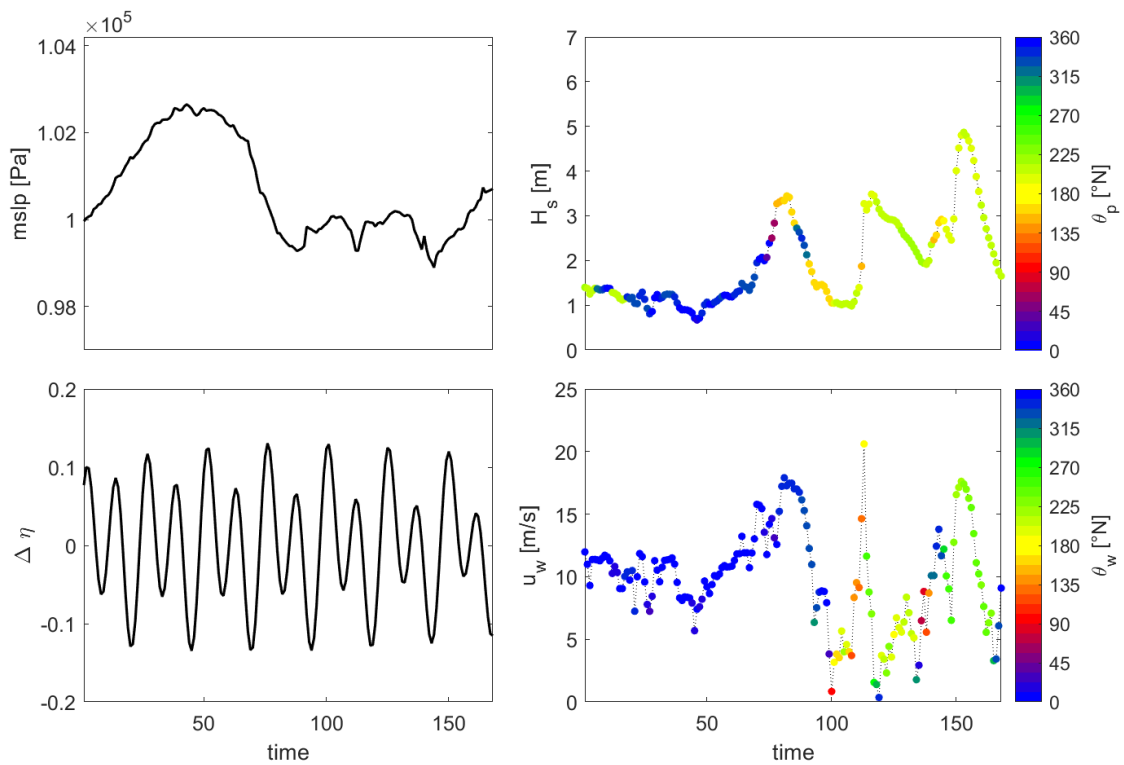


Fig. 22. Scénario no. 20.



Interreg



UNION EUROPÉENNE
UNIONE EUROPEA



SINAPSI

MARITTIMO-IT FR-MARITIME

Fonds européen de développement régional
Fondo Europeo di Sviluppo Regionale

PRODUIT N. T2.1.2

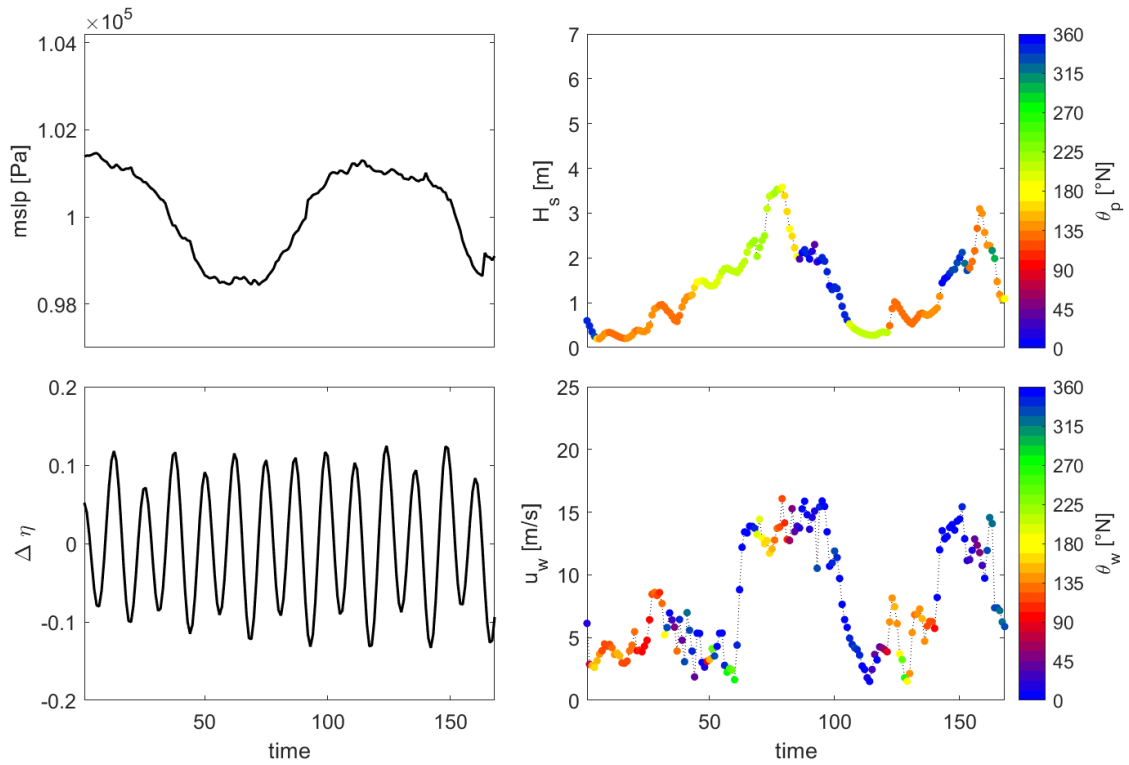


Fig. 23. Scénario no. 21.

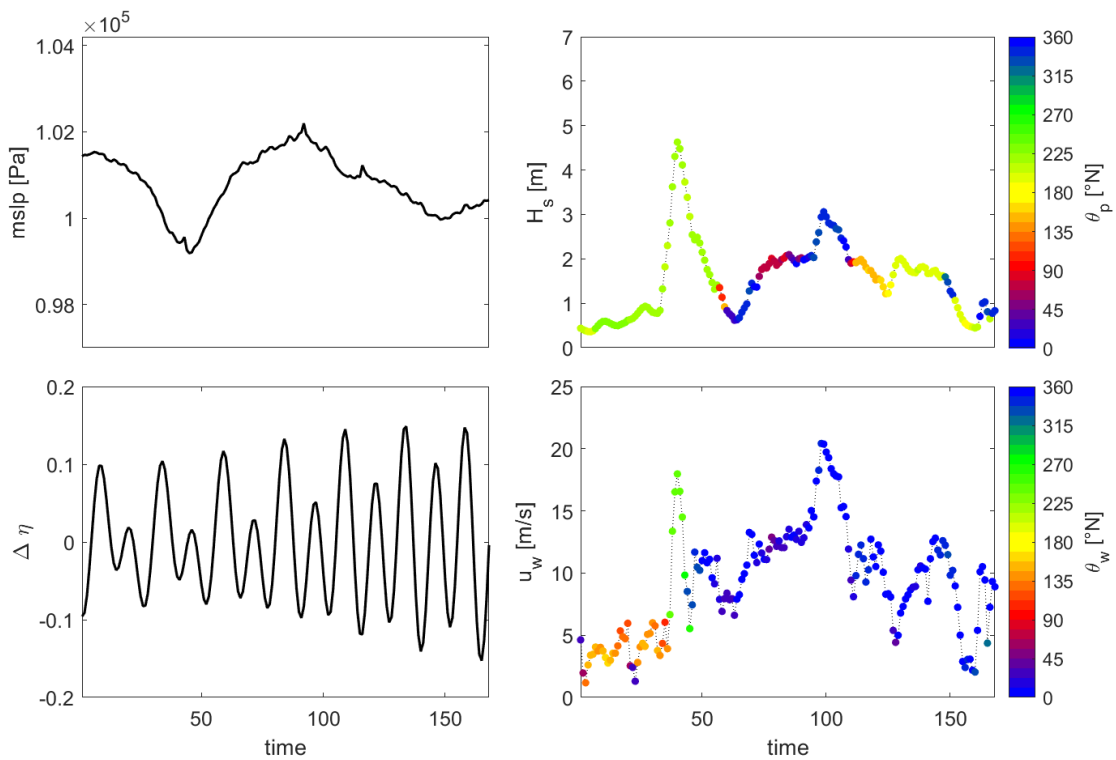


Fig. 24. Scénario no. 22.



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

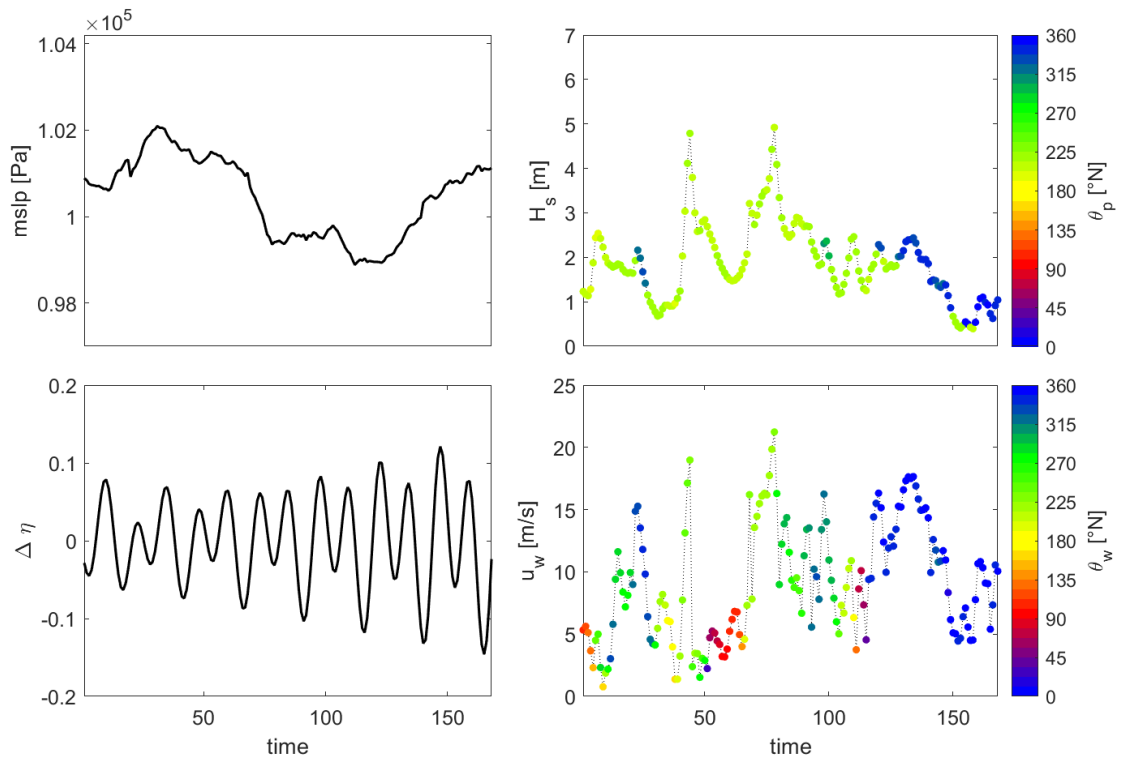


Fig. 25. Scénario no. 23.

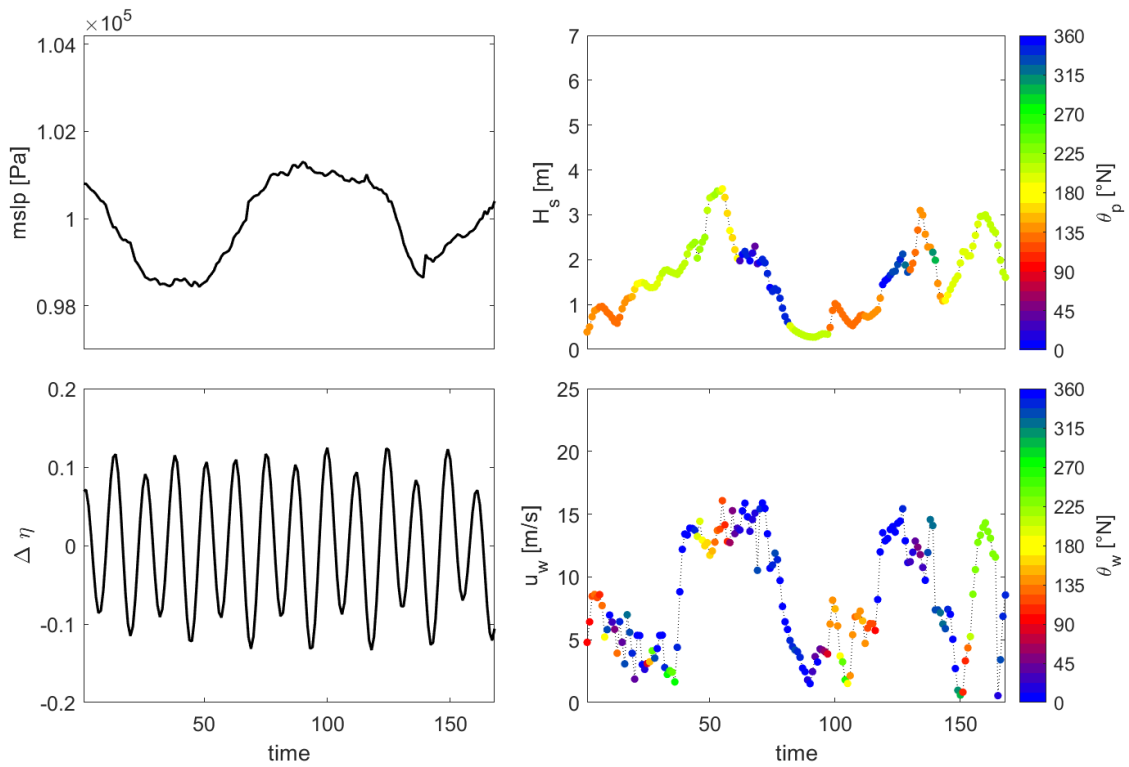


Fig. 26. Scénario no. 24.

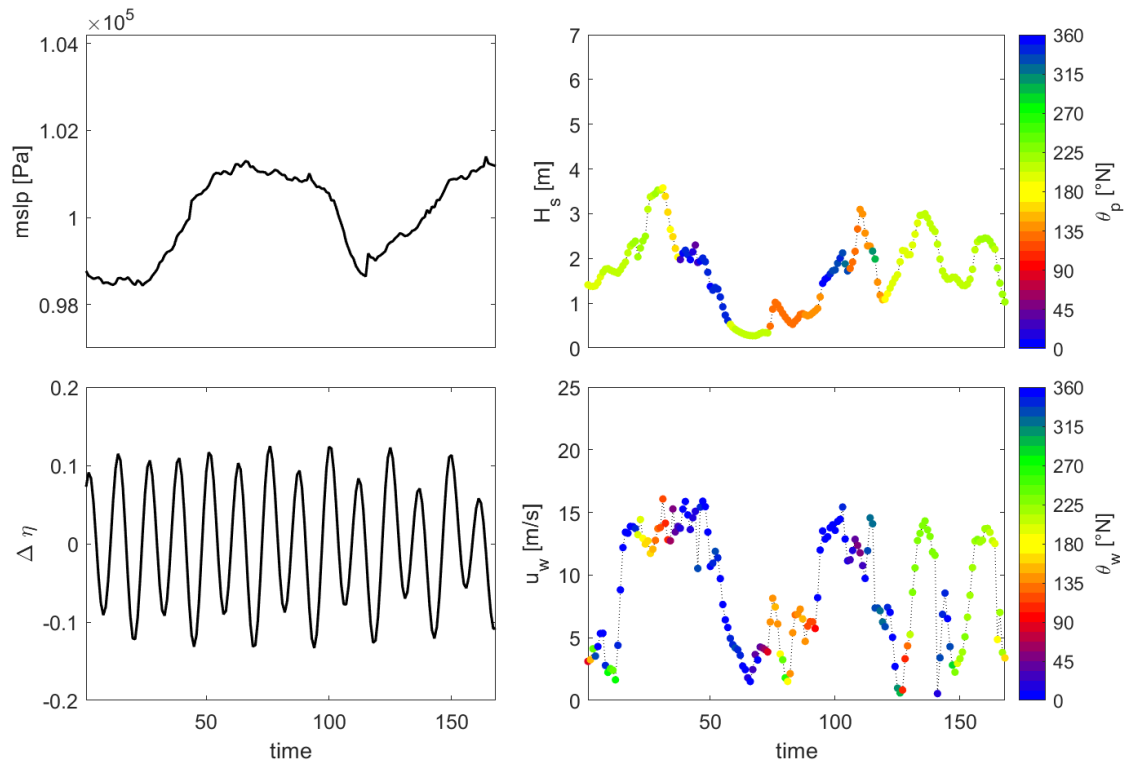


Fig. 27. Scénario no. 25.

Les analyses présentées jusqu'à présent permettent d'analyser des sous-ensembles caractérisés par une variabilité temporelle dans les fenêtres de référence sélectionnées, et sont donc particulièrement efficaces si les processus à modéliser sont fortement non stationnaires. Cependant, les algorithmes de sélection permettent d'identifier des scénarios significatifs pour des nœuds uniques, alors que les signaux extraits sur des nœuds multiples sont généralement "bruyants".

Lorsque la variabilité spatiale du phénomène étudié est particulièrement importante, il est nécessaire de recourir à des méthodologies qui permettent d'identifier des schémas significatifs sur de grandes surfaces, bien qu'ils ne contiennent pas de manière exhaustive toutes les informations sur la variabilité temporelle du phénomène étudié. L'une des méthodes les plus répandues dans ce sens est la méthode SOM (*Self Organising Maps*) dont une application est présentée ci-dessous dans la baie de Toulon.

4. MODÈLE DE CIRCULATION DANS LA BAIE DE TOULON

La technique SOM est utilisée pour l'extraction des schémas de circulation dans la zone du Golfe de Lion, en utilisant les données produites par un modèle numérique imbriqué sur trois mailles de calcul, fonctionnant actuellement avec une couverture allant du Golfe du Lion à la Baie de Toulon.

La méthode d'apprentissage non supervisé SOM permet de réduire une base de données multidimensionnelle à une représentation (généralement en 2D), tout en conservant sa structure topologique originale. La méthode comporte deux phases (formation et cartographie) : en pratique, une carte auto-adaptative est développée afin de représenter un ensemble de données au moyen d'un certain nombre de neurones ; chaque neurone est affecté à la représentation d'un ensemble de données particulier sur la base de caractéristiques communes. Cette approche permet une visualisation multidimensionnelle des données croisées.

Dans le cadre du projet SINAPSI, la méthode SOM a été appliquée à plusieurs jeux de données, dans le but d'obtenir une classification des circulations dans la baie de Toulon et la zone côtière environnante. En particulier, des travaux ont été réalisés sur des cartes de courants de surface obtenues à partir de HFR (hfradar.univ-tln.fr), traitées à partir de données de vent obtenues avec le modèle atmosphérique ARPEGE de Météo-France. Grâce aux SOM, il a été possible de réduire un ensemble de données horaires définies sur un intervalle de 6 mois (juin 2000-novembre 2000) à 6 neurones, chacun représentant un schéma de circulation marine spécifique associé à un forçage du vent. Les neurones (c'est-à-dire les cartes) extraits sont illustrés à titre d'exemple à la figure 28, tandis que la figure 29 montre les unités dites de meilleure correspondance (BMU), c'est-à-dire la séquence de neurones à utiliser pour reproduire l'ensemble des cartes d'entrée avec le plus haut degré de vraisemblance.

Comme on peut le voir sur les figures ci-dessus, cette représentation schématique permet de visualiser rapidement et efficacement la circulation marine due au vent. Dans les neurones N2, N5, N6 on trouve des conditions de vent de Mistral (moyen à fort) générant un courant de surface dirigé vers l'Ouest, Sud-Ouest. Une situation (N3) de fort vent d'est qui renforce le courant côtier Liguro-Provençal. Une situation avec un vent très modéré de sud-ouest qui réduit le courant de terre. Enfin, une situation sans vent (N1) qui met en évidence le courant géostrophique (Caceres-Euse, et al. 2022).

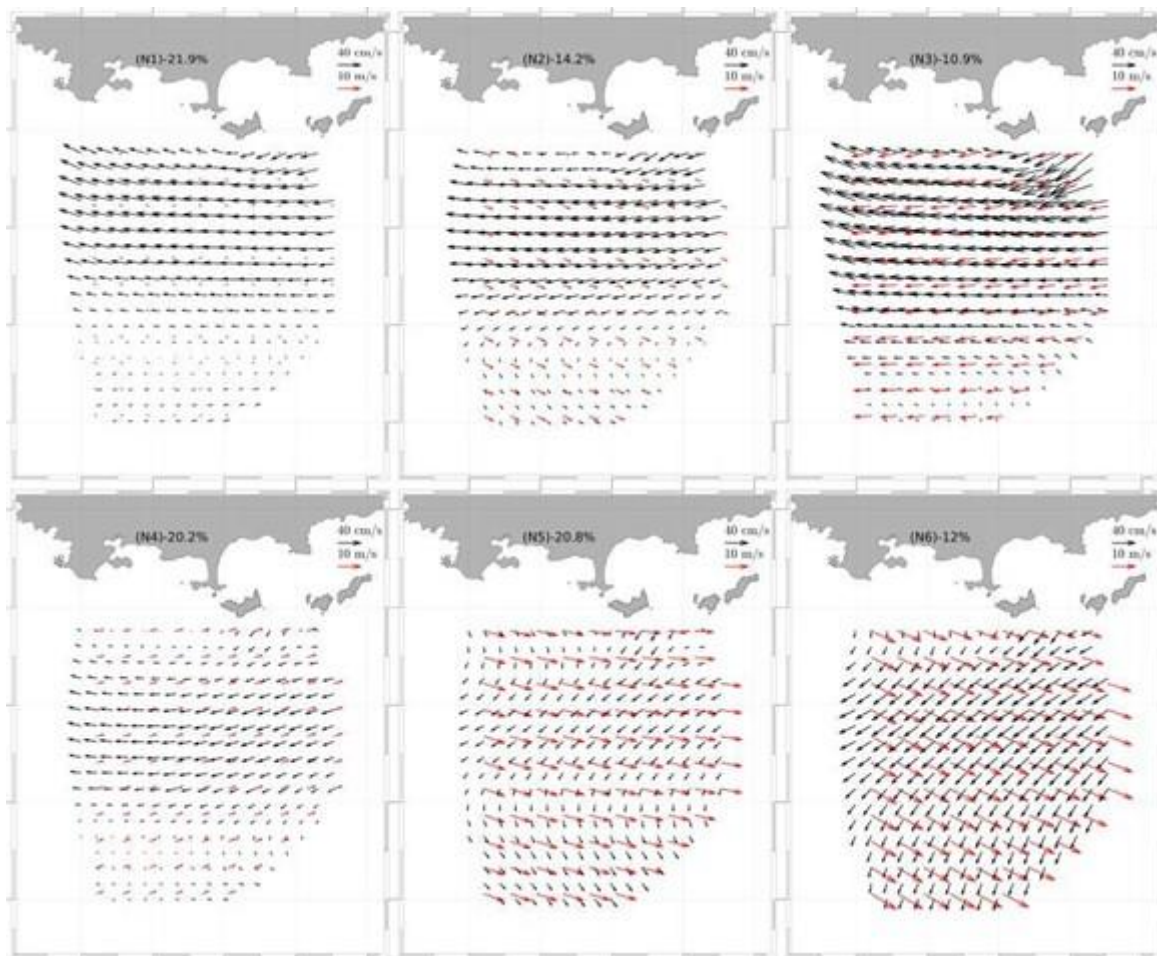


Fig. 28. Résultat de l'analyse jointe SOM 3 x 2 des courants de surface par HFR (en noir) et du vent à 10m (en rouge) pour la période allant de Juin à Novembre 2020..

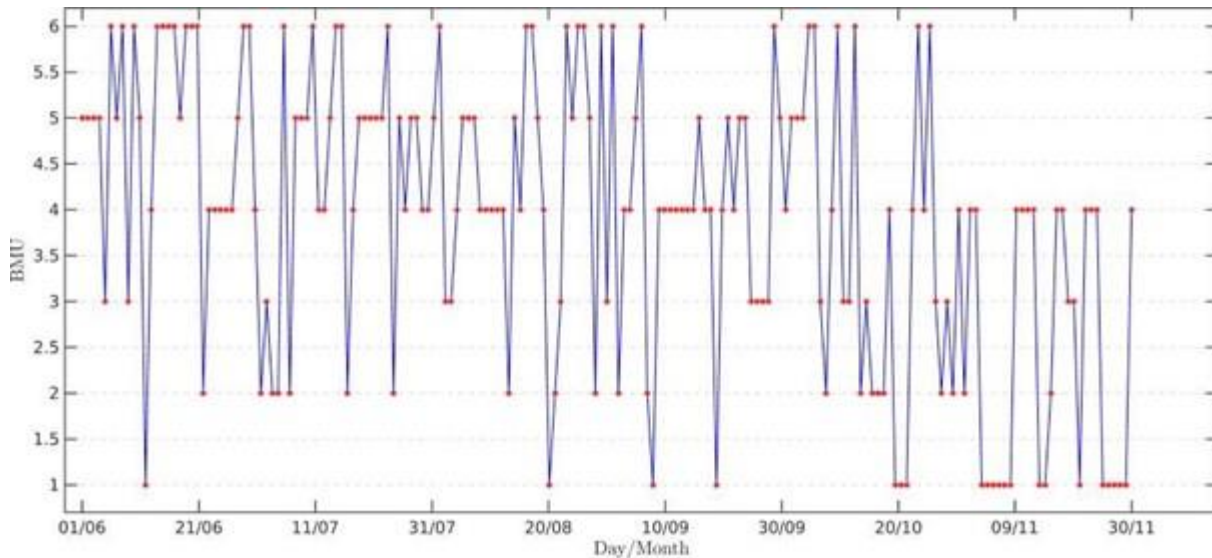


Figura 29. BMU corrispondants à la Fig. 28. Chaque date correspond à un neurone.

Bibliographie

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.