



**RÉPUBLIQUE
FRANÇAISE**

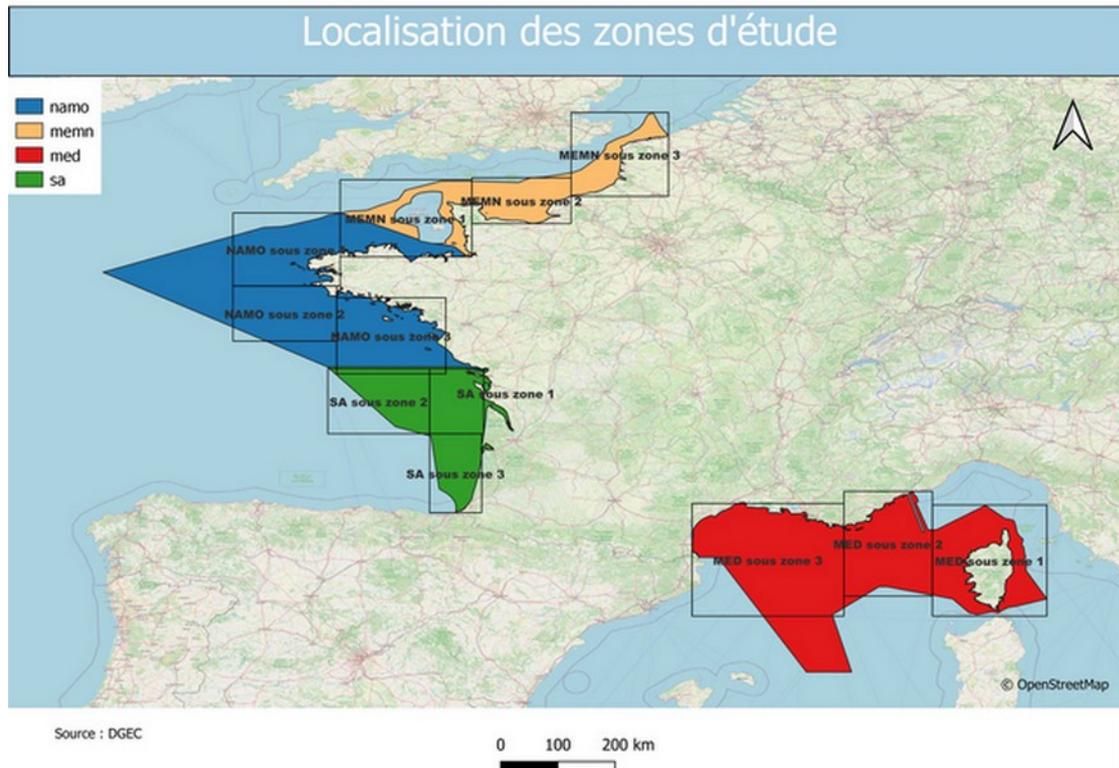
*Liberté
Égalité
Fraternité*



**METEO
FRANCE**

Dérisquage face aux risques atmosphériques

Lot 7 - Zones homogènes et points de mesure associés



<p>Ministère de la Transition écologique (MTE) Direction Générale de l'Énergie et du Climat (DGE)</p>	<p>@guillaume.rouviere@developpement-durable.gouv.fr 01 40 81 20 98</p>	
<p>Chargé d'affaires : M. Christophe Jacolin</p>	<p>@christophe.jacolin@meteo.fr 05 61 07 86 85</p>	
<p>Chef de projet : M. Raphaël Legrand</p>	<p>@raphael.legrand@meteo.fr 05 61 07 86 11</p>	

Table des matières

1. Expression de la demande.....	3
2. Méthodologie.....	4
2.1. Données utilisées.....	4
2.2. Analyse en composantes principales.....	5
2.3. Gaussian Mixture Model.....	6
2.4. Sensibilité à l’initialisation de l’algorithme.....	6
2.5. Choix du nombre de clusters.....	10
2.6. Sensibilité au niveau d’altitude.....	13
3. Résultats obtenus.....	14
3.1. Zones MEMN.....	14
3.2. Interface MEMN-NAMO.....	18
3.3. Zones NAMO.....	21
3.4. Interface NAMO-SA.....	27
3.5. Zones SA.....	29
3.6. Zones en Méditerranée.....	33
4. Livrables.....	36
5. Conclusions et perspectives.....	37
ANNEXE 1.....	39

1. Expression de la demande

Ce rapport est réalisé dans le cadre des fournitures demandées par la DGEC pour le lot préliminaire du lot 7 du Dérisquage face aux risques atmosphériques. Ce lot préliminaire a deux objectifs pour chacune des façades :

- Pour chacune des sous-zones définies dans la Figure 1 (MEMN sous zone 3 par exemple), identifier 3 à 6 zones homogènes en termes de climatologie de vent.
- Pour chacune de ces zones homogènes, un point de mesure “optimal” (ou une enveloppe) le plus représentatif de la zone sera identifié.

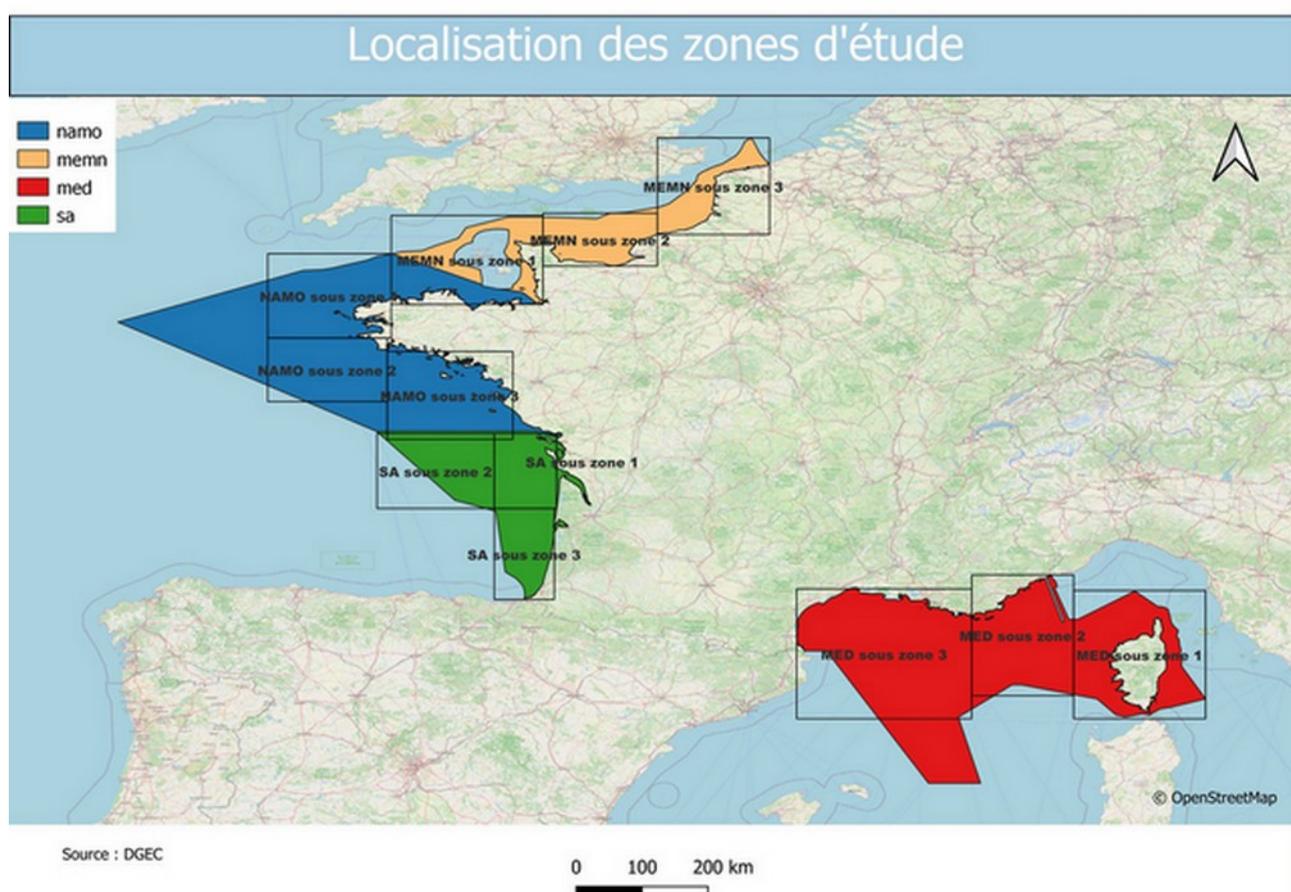
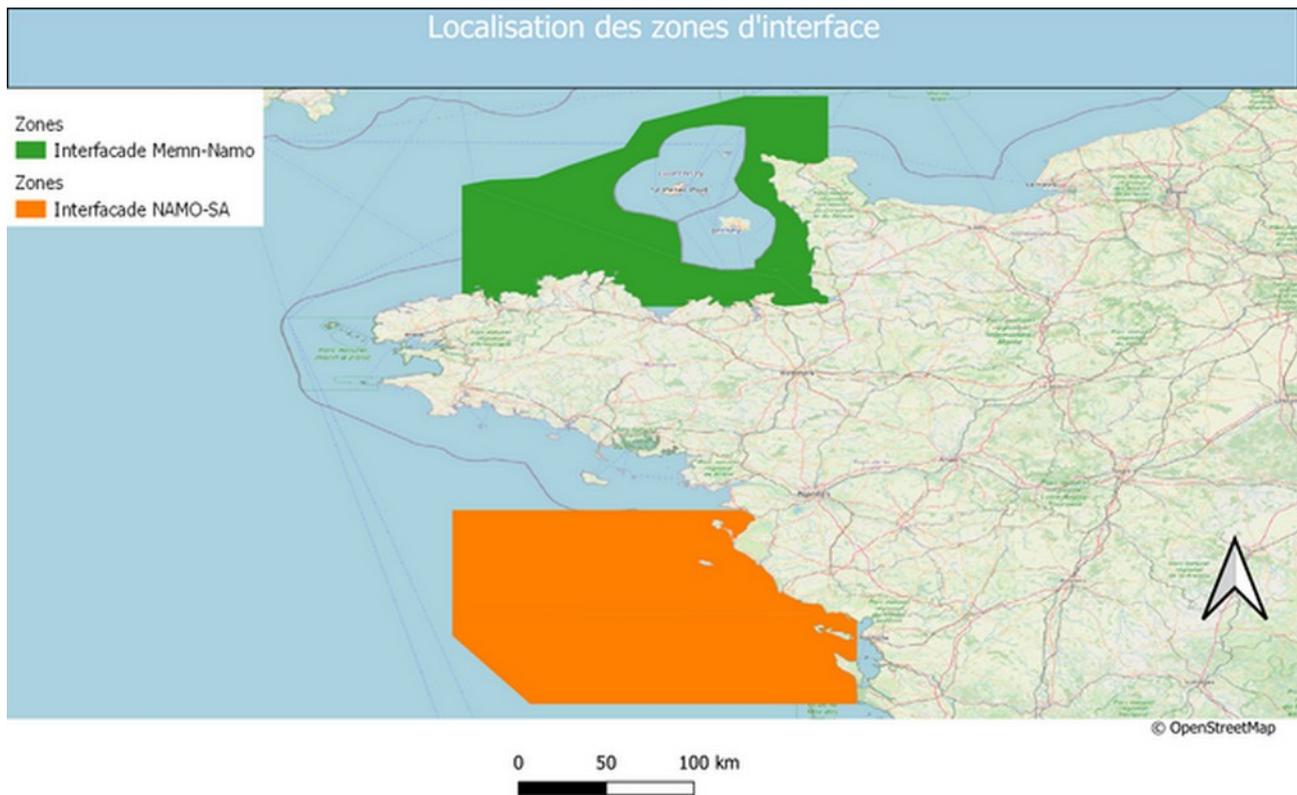


Figure 1: Localisation des sous-zones d'étude pour chacune des quatre façades métropolitaines.

En plus de ces sous-zones d'étude, l'étude porte à la demande de la DGEC sur deux sous-zones d'interface dont les contours sont définis dans la Figure 2 :

- une interfaçade « MEMN-NAMO »
- une interfaçade « NAMO-SA »



Source : DGEC

Figure 2: Localisation des zones d'interface

2. Méthodologie

Ce lot s'appuiera principalement sur les travaux de Marcille et al., 2023¹ en retenant la méthode GMM (Gaussian Mixture Model). Les grands principes de cette méthodologie, ainsi que les modifications apportées à cette dernière, seront décrites dans la suite de ce chapitre.

Les codes sources des travaux de Marcille et al., 2023 sont disponibles sur github : https://github.com/rmarcille/gmm_sparse_sampling/tree/main

2.1. Données utilisées

Les travaux de Marcille et al., 2023 s'appuient sur les données AROME² issues de Meteonet, au niveau 10 mètres sur une grille à résolution 0.01° et sur la période 2016 à 2018. De son côté, Météo-France dispose de séries horaires de vent plus conséquentes issues du modèle AROME. Les

1 Marcille, Robin, et al. "Gaussian Mixture Models for the Optimal Sparse Sampling of Offshore Wind Resource." *Wind Energy Science Discussions* (2022): 1-24

2 Y. Seity et al., 2010, The AROME-France Convective-Scale Operational Model <https://journals.ametsoc.org/view/journals/mwre/139/3/2010mwr3425.1.xml>

données sont disponibles sur la période 2018-2022, à différentes hauteurs entre 10 mètres et 200 mètres. En termes de résolution spatiale seuls les niveaux 10 et 100m sont disponibles à résolution 0.01°. Les niveaux intermédiaires et plus élevés ne sont disponibles qu'à la résolution 0.025° (exemple de grille en Figure 3 pour la sous-zone MEMN 2).

Dans la suite du rapport, à la demande de la DGEC pour être représentatif de la hauteur des turbines, ce sont les données AROME 2018-2022 au niveau 200 mètres qui seront analysées.

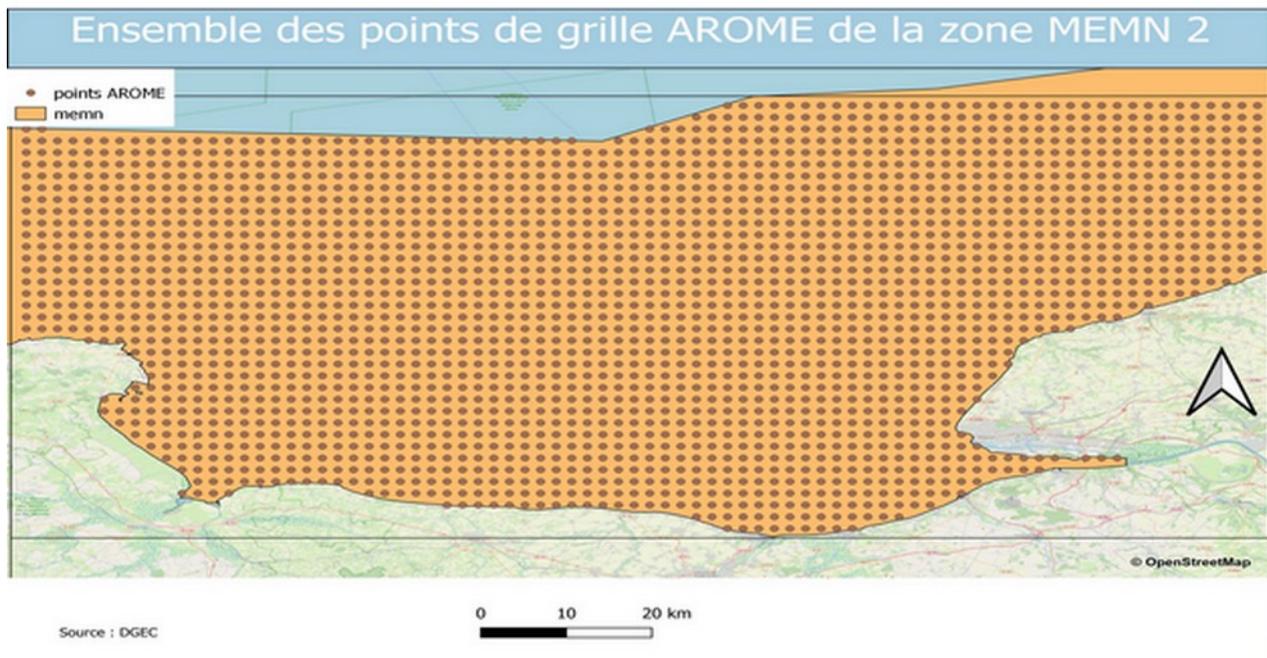


Figure 3: Ensemble des points de grille AROME de la sous-zone MEMN 2

2.2. Analyse en composantes principales

Le modèle d'ordre réduit utilisé pour diminuer la dimension des données d'entrée est l'analyse de la fonction orthogonale empirique (EOF), également connue sous le nom d'analyse en composantes principales (ACP). Les données sont décomposées sur une base orthogonale. En pratique cette décomposition est liée aux valeurs singulières d'une matrice X telle que

$$X = U\Sigma V^T,$$

avec Σ une matrice diagonale de σ_k valeurs singulières positives, U une matrice dont les colonnes sont les vecteurs de la base orthogonale et V les poids des vecteurs associés.

Les vecteurs singuliers sont des vecteurs orthogonaux sur lesquels la variance des données projetées est maximisée. Les éléments diagonaux de Σ sont triés par valeur et sont égaux au pourcentage de variance de l'ensemble de données expliqué par chaque composante principale.

Dans le modèle de prévision numérique du temps, les points de grille sont fortement corrélés spatialement ; par conséquent, seul un faible nombre d'EOF est nécessaire pour expliquer la grande majorité de la variance de l'ensemble des données. Nous retiendrons un nombre d'EOF de 10 pour chacune des composantes u (vent zonal) et v (vent méridien) du vent, en accord avec les travaux de Marcille et al, 2023.

2.3. Gaussian Mixture Model

Un modèle GMM est un modèle probabiliste permettant de représenter des sous-populations distribuées selon une loi normale au sein d'une population globale (Reynolds, 2009³). Chaque distribution gaussienne représente un ensemble de points, c'est-à-dire un cluster. Le modèle est un mélange, c'est-à-dire une superposition de composantes gaussiennes multivariées qui définissent alors une distribution de probabilité $p(x)$ sur les données :

$$\rho(x) = \sum_{j=1}^D \pi_j \mathcal{N}(x|\mu_j, \Sigma_j),$$

π_j étant le poids de la composante gaussienne j , avec $0 \leq \pi_j \leq 1$ et $\sum_{j=1}^D \pi_j = 1$. $\mathcal{N}(x|\mu)$ étant la distribution de densité gaussienne telle que :

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^r \det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right),$$

avec x étant le vecteur d'entrée de dimension r , μ le vecteur moyen de dimension r et Σ la matrice de covariance.

2.4. Sensibilité à l'initialisation de l'algorithme

L'algorithme GMM est particulièrement sensible aux conditions d'initialisation comme indiqué par Marcille et al., 2023 dans ses travaux. Différentes initialisations pourront amener à différentes solutions, avec une multitude de centroïdes et clusters possibles.

La Figure 4 montre bien l'incertitude résultant des conditions initiales, avec une dispersion importante des centroïdes obtenus lors de 100 tirages initiaux distincts pour la sous-zone Med3 avec 6 clusters distincts.

³ Reynolds, D. A.: Gaussian mixture models, Encyclopedia of biometrics, 741, 659–663, 2009.

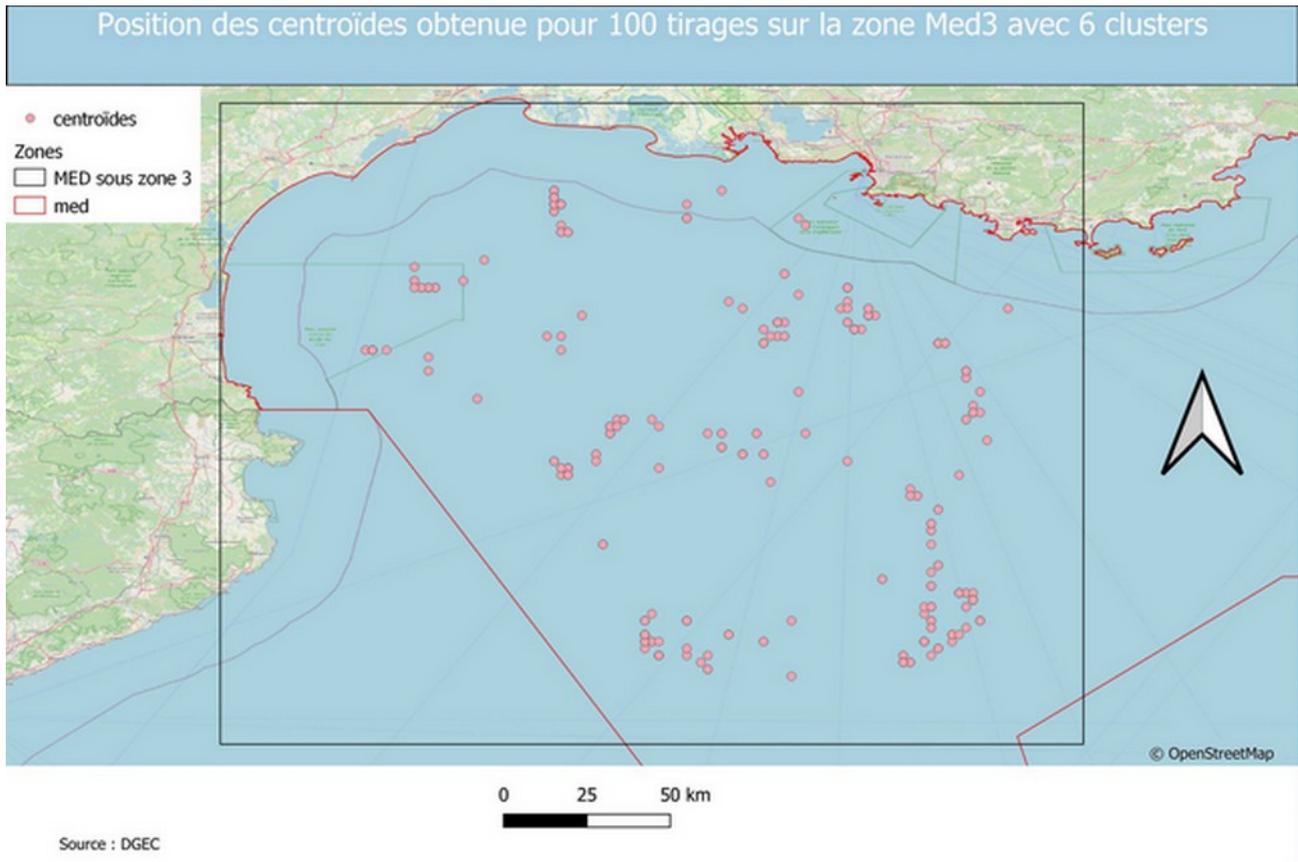


Figure 4: Position des centroïdes obtenus avec 100 tirages pour 6 clusters sur la sous-zone Med3

L'utilisation des densités spatiales permet toutefois de faire ressortir dans la Figure 5, des zones préférentielles des centroïdes (c'est-à-dire des zones préférentielles où installer les bouées afin qu'elles soient les plus représentatives de chacun des clusters).

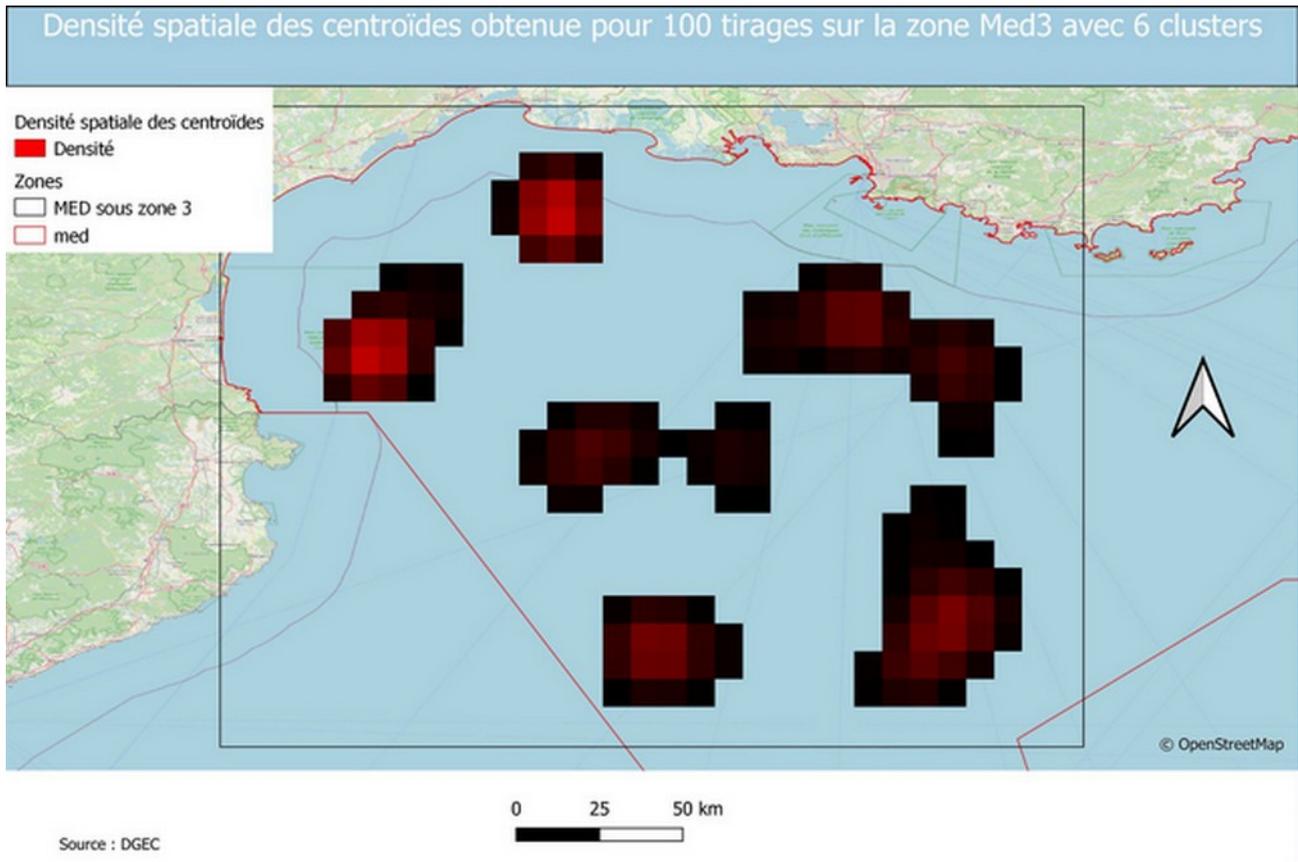


Figure 5: Densité spatiale des centroïdes obtenue pour 100 tirages sur la zone Med3 avec 6 clusters

Par la suite, nous retiendrons parmi les 100 tirages celui pour lequel le plus de centroïdes sont situés sur les maxima de densité spatiale afin de fournir un clustering possible.

La position des bouées obtenue pour le meilleur tirage est illustrée sur la Figure 6 : 4 bouées seraient bien placées au niveau ou à proximité immédiate des maxima de densité dans cette configuration, les 2 autres étant placées sur des maxima secondaires.

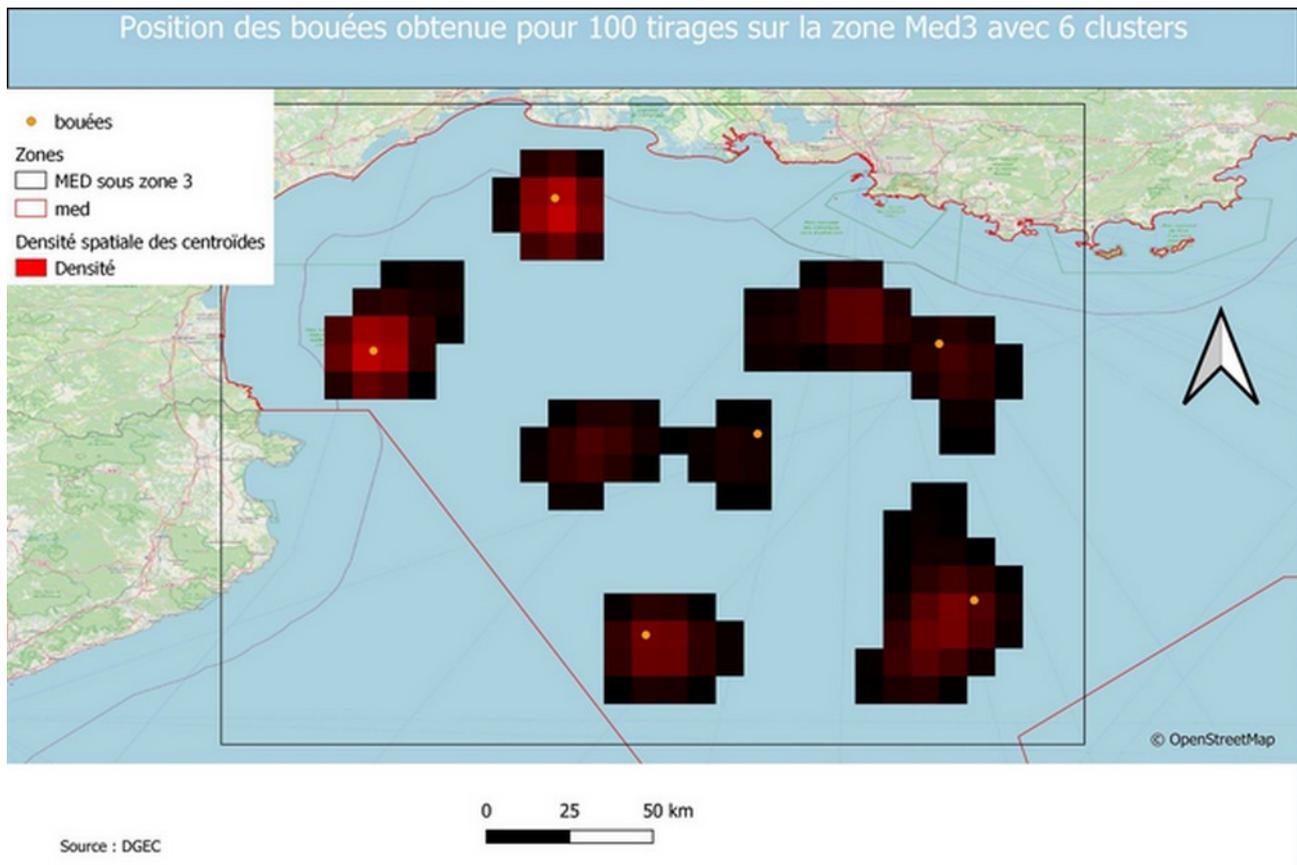


Figure 6: Position des bouées obtenue pour 100 tirages sur la zone Med3 avec 6 clusters

Le clustering obtenu pour la sous-zone Med3, illustré sur la Figure 7, est ainsi la solution retenue parmi les 100 tirages. Les contours des clusters pourraient être légèrement différents pour des centroïdes placés exactement sur les maxima de densité, mais à la marge puisque la majorité des bouées sont placées pour ce tirage au niveau ou à proximité immédiate des maxima de densité.

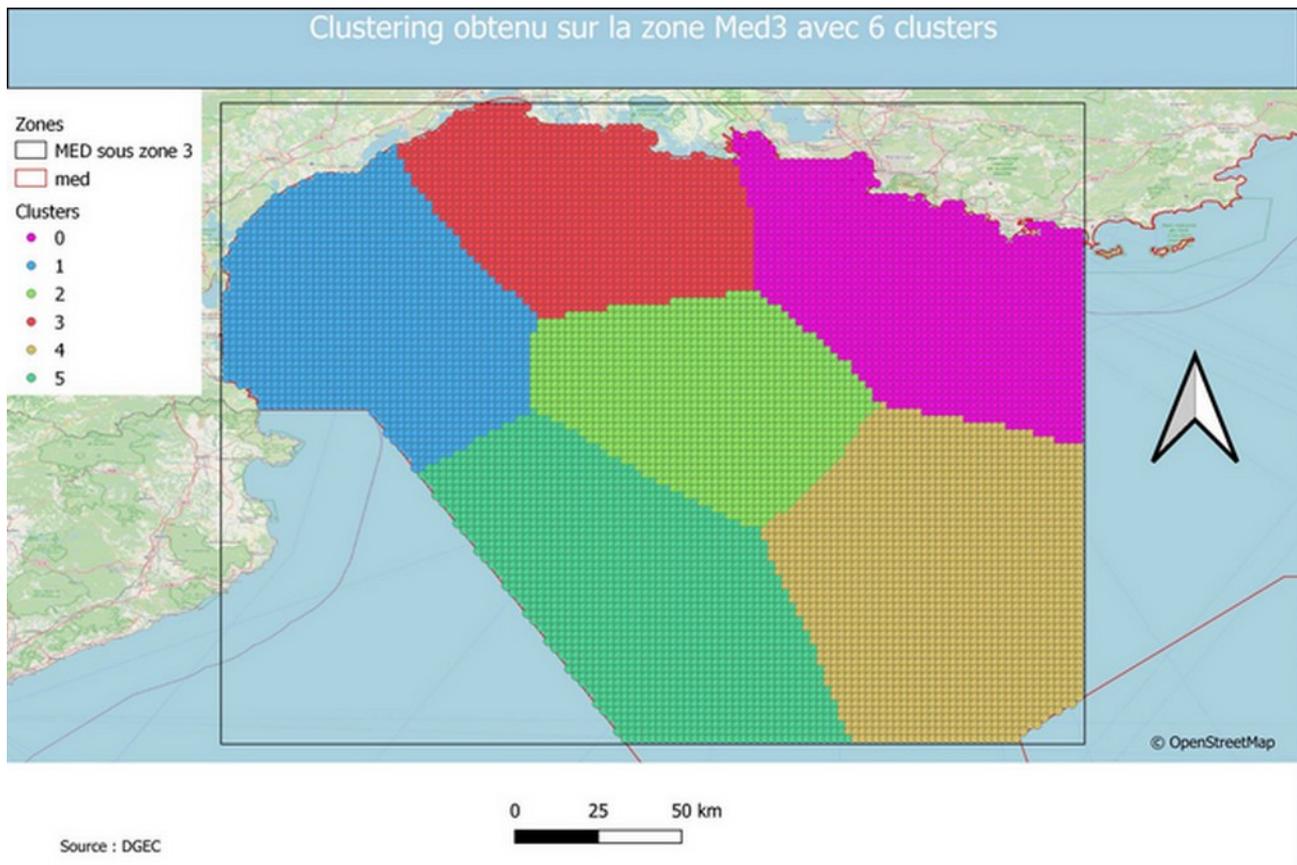


Figure 7: Clustering obtenu sur la sous-zone Med3 avec 6 clusters

2.5. Choix du nombre de clusters

GMM nécessite que le nombre de clusters dans le modèle soit imposé en entrée. Le nombre optimal de clusters peut être défini grâce au calcul du score du critère d'information bayésien (BIC) (Schwarz, 1978⁴)

$$\text{BIC} = -2\ln(\mathcal{L}) + G\ln(K)$$

avec L la valeur maximisée de la fonction de vraisemblance, G le nombre de paramètres dans les vecteurs moyens et les matrices de covariance des composantes gaussiennes, et K le nombre de points de données. Les modèles trop complexes, qui sur-ajusteraient les données, sont pénalisés par ce score BIC. Plus le BIC est bas, meilleur est le modèle.

Traditionnellement la pente du score BIC est utilisée afin d'identifier plus facilement un minimum, grâce à la recherche d'un coude dans la courbe du gradient du score BIC qui permet de faire ressortir un ensemble de valeurs possibles pour le nombre de clusters optimaux. Ainsi, on peut dire que le gradient du score BIC donne une indication sur l'étendue du nombre optimal de clusters.

⁴ Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 1st edition, <https://doi.org/10.1214/aos/1176344136>, 461–464, 1978.

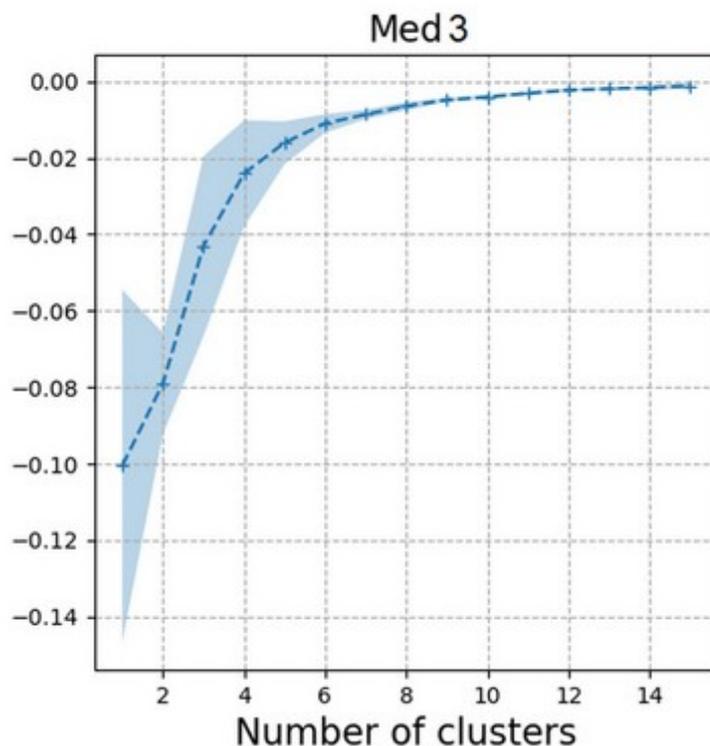


Figure 8: Gradient du score BIC normalisé sur la zone Med3 pour un nombre de clusters allant de 1 à 15. Les enveloppes sont l'intervalle de confiance obtenu à partir de 20 initialisations différentes de la méthode GMM. La détermination d'un nombre optimal de clusters à partir de ces courbes est incertaine, allant de trois à six pour cette sous-zone Med3.

En complément de cette méthode traditionnelle, nous nous appuyerons aussi sur la détermination d'un seuil de reconstruction d'erreur du champ de vent, sur la base des travaux de Marcille et al, 2023.

Pour comparer plusieurs zones qui ont des régimes de vent différents, le seuil d'erreur est défini comme l'erreur de reconstruction du vent normalisé (erreur quadratique moyenne normalisée ou N-RMSE). Le nombre de clusters associé au seuil d'erreur minimal choisi est alors considéré comme optimal.

Le seuil d'erreur choisi est somme toute assez arbitraire et choisi de façon empirique. Les zones les plus grandes et ayant la plus grande variabilité sont celles situées sur le pourtour méditerranéen, en accord avec les travaux de Marcille et al, 2023 : ce sont sur ces zones que le nombre de clusters minimal nécessaire pour atteindre 75 % de la zone sous le seuil est le plus important.

La sous-zone Med3 nous permet de déterminer un seuil N-RMSE de 0.24 pour lequel le nombre de clusters minimal nécessaire est de 5 (cf Figure 9). Nous utiliserons notamment ce seuil de N-RMSE pour déterminer un nombre minimal de capteurs nécessaires pour l'ensemble des zones. De façon tout à fait logique, ce seuil est plus élevé que celui utilisé dans les travaux de Marcille et al., du fait

de zones nettement plus grandes dans nos travaux et de la volonté de limiter le nombre de clusters à 5 ou 6 au plus par zones.

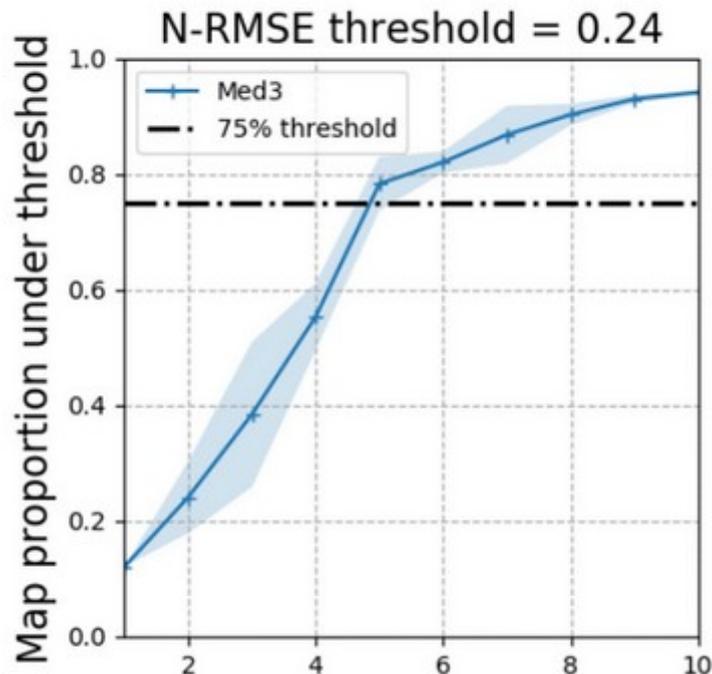


Figure 9: Proportion de la sous-zone Med3 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes. La courbe bleue représente la proportion moyenne du N-RMSE sous le seuil d'erreur de 0.24 pour les 5 initialisations GMM différentes, et la plage bleue l'intervalle de confiance à 95% associé

Au final, afin de déterminer le nombre de clusters, l'ensemble des critères suivants seront utilisés :

- le gradient de la courbe BIC qui permet de déterminer une plage de valeurs possibles
- le seuil de N-RMSE de 0.24, qui permet de déterminer un nombre minimal de capteurs nécessaires pour lequel la proportion moyenne du N-RMSE pour les 5 initialisations dépasse ce seuil sur au moins 75 % de la sous-zone.
- la densité spatiale des centroïdes (obtenue par Kernel Density Estimation⁵) et en particulier si un nombre adéquat de zones préférentielles se détachent.
- les centroïdes sélectionnés dans la solution retenue et leur cohérence avec la densité spatiale des centroïdes.
- les clusters obtenus et en particulier leur cohérence spatiale entre zones voisines

⁵ KDE, voir par exemple Murray Rosenblatt. "Remarks on Some Nonparametric Estimates of a Density Function." Ann. Math. Statist. 27 (3) 832 - 837, September, 1956. <https://doi.org/10.1214/aoms/1177728190>

Nous nous limiterons à la demande de la DGEC à un nombre maximal de 6 clusters par zones.

2.6. Sensibilité au niveau d'altitude

Les résultats obtenus selon le niveau d'altitude choisi peuvent être sensiblement différents, en particulier entre les niveaux d'altitude les plus éloignés (10 m et 200 m). Ainsi pour la sous-zone SA1, avec 4 clusters, les maxima de densité spatiale des centroïdes sont positionnés parfois sensiblement différemment entre 10 m et 200 m (cf Figure 10 et Figure 11). De plus, il est plus difficile de distinguer exactement 4 maxima de densité à 10 m contrairement au niveau 200 m.

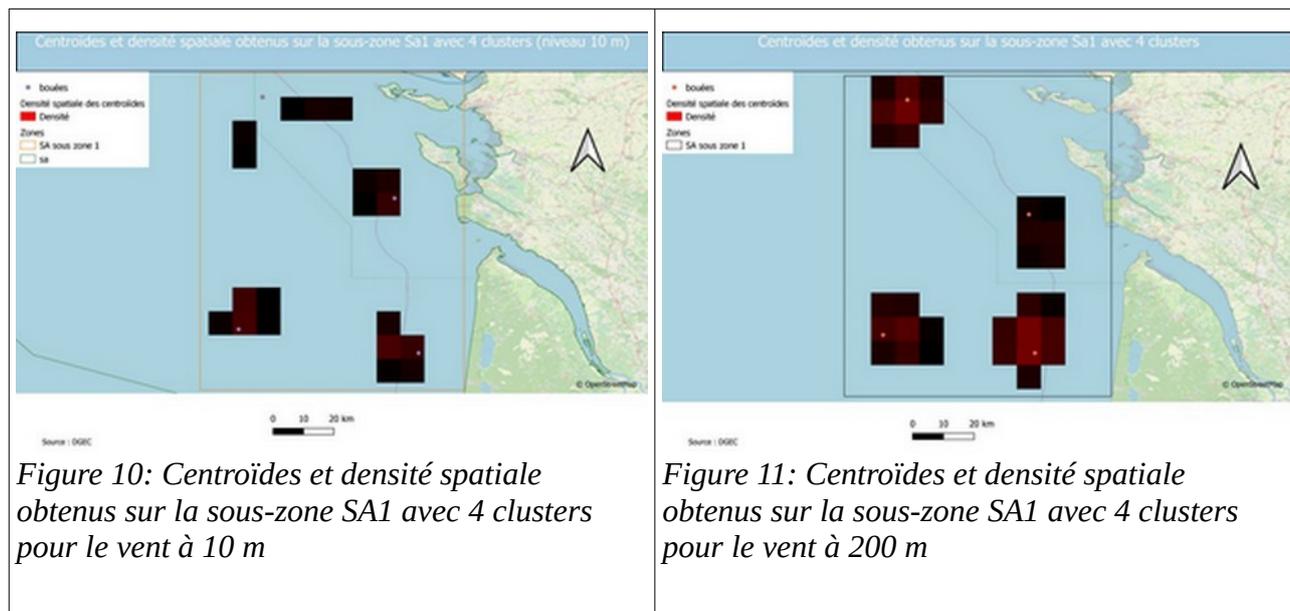


Figure 10: Centroides et densité spatiale obtenus sur la sous-zone SA1 avec 4 clusters pour le vent à 10 m

Figure 11: Centroides et densité spatiale obtenus sur la sous-zone SA1 avec 4 clusters pour le vent à 200 m

Au final les clusters obtenus ont une structure nettement différente pour la sous-zone SA1 entre le vent à 10 m et celui à 200 m. Dans la suite du rapport, à la demande de la DGEC pour être représentatif de la hauteur des turbines, ce sont les données AROME 2018-2022 au niveau 200 mètres qui seront analysées.

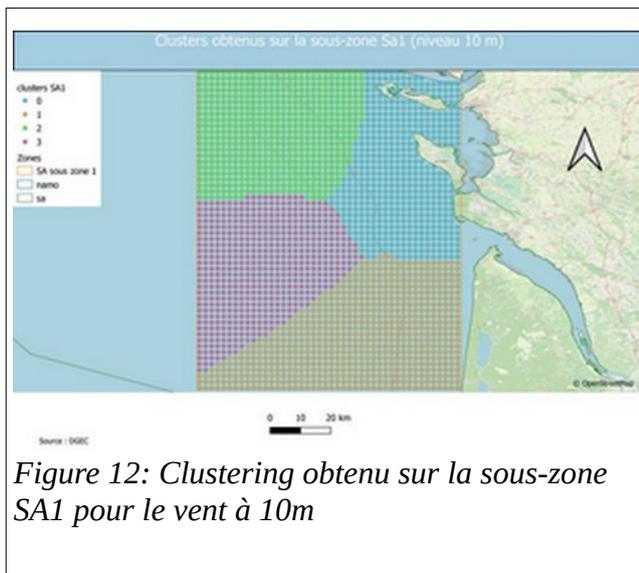


Figure 12: Clustering obtenu sur la sous-zone SA1 pour le vent à 10m

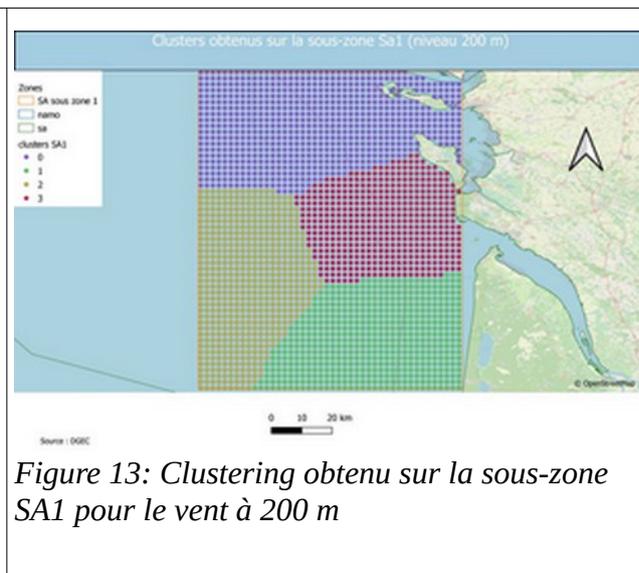


Figure 13: Clustering obtenu sur la sous-zone SA1 pour le vent à 200 m

3. Résultats obtenus

3.1. Zones MEMN

Sur la sous-zone MEMN1, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 14).

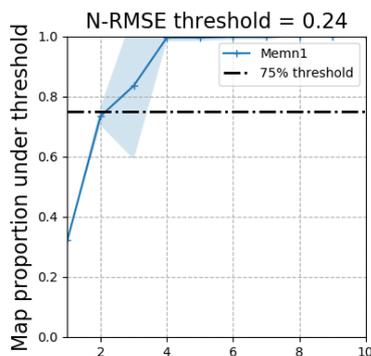


Figure 14: Proportion de la sous-zone MEMN1 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 3 zones préférentielles (cf Figure 15). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 3 clusters semblent satisfaisants sur la sous-zone MEMN1.

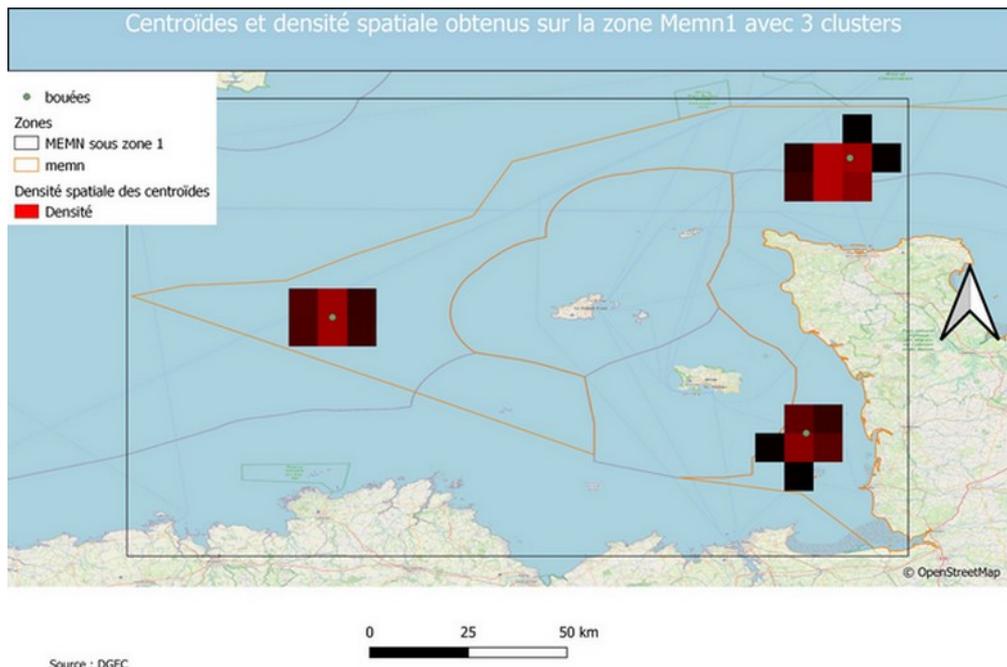


Figure 15: Centroides et densité spatiale obtenus sur la zone MEMN1 avec 3 clusters

Sur la sous-zone MEMN2, le nombre de clusters nécessaires afin d'atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 2 (cf courbe bleue de la Figure 16). Afin d'assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroides, nous retenons finalement 4 clusters sur cette sous-zone.

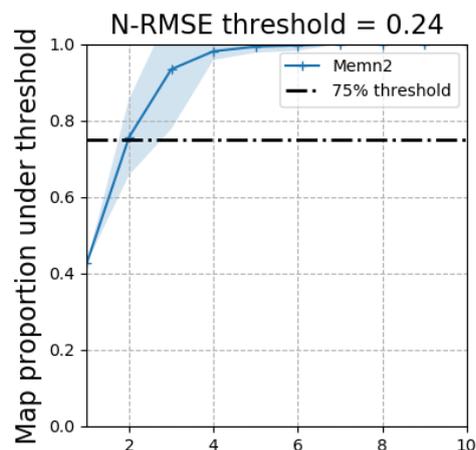


Figure 16: Proportion de la sous-zone MEMN2 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroides montre une bonne concentration autour de 4 zones préférentielles (cf Figure 17). La position retenue des bouées pour la construction des clusters est aussi proche de

chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone MEMN2.

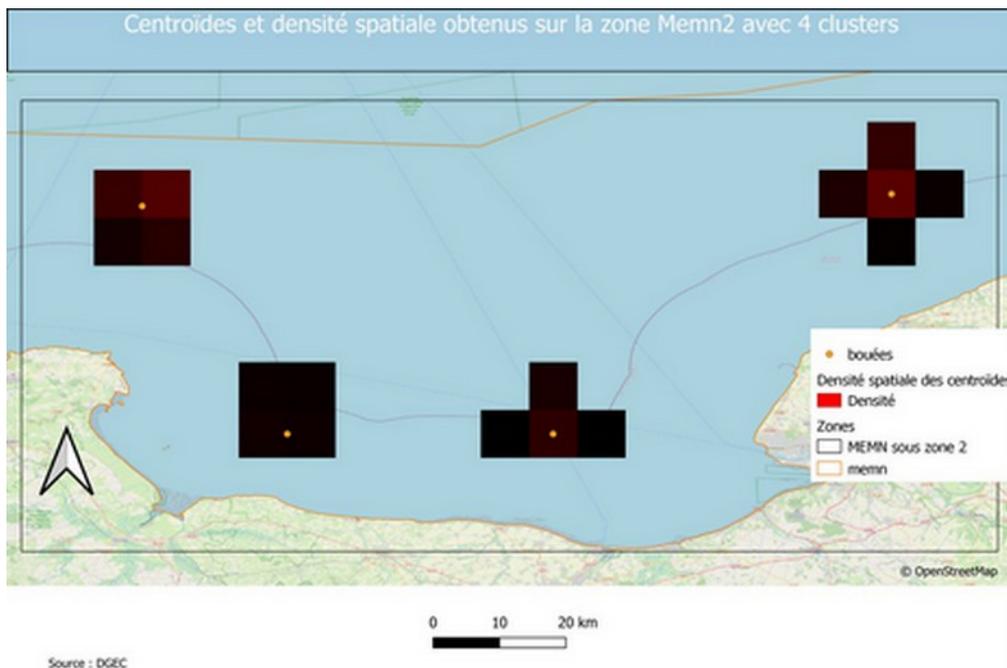


Figure 17: Centroides et densité spatiale obtenus sur la zone MEMN2 avec 4 clusters

Sur la sous-zone MEMN3, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 18).

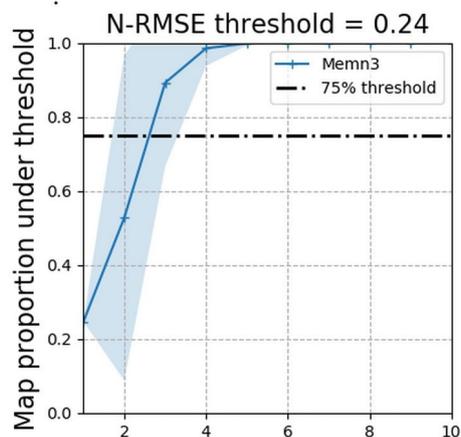


Figure 18: Proportion de la sous-zone MEMN3 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroides montre une bonne concentration autour de 3 zones préférentielles (cf Figure 19). La position retenue des bouées pour la construction des clusters est aussi proche de

chacun des maxima de densité. Ainsi les résultats obtenus avec 3 clusters semblent satisfaisants sur la sous-zone MEMN3.

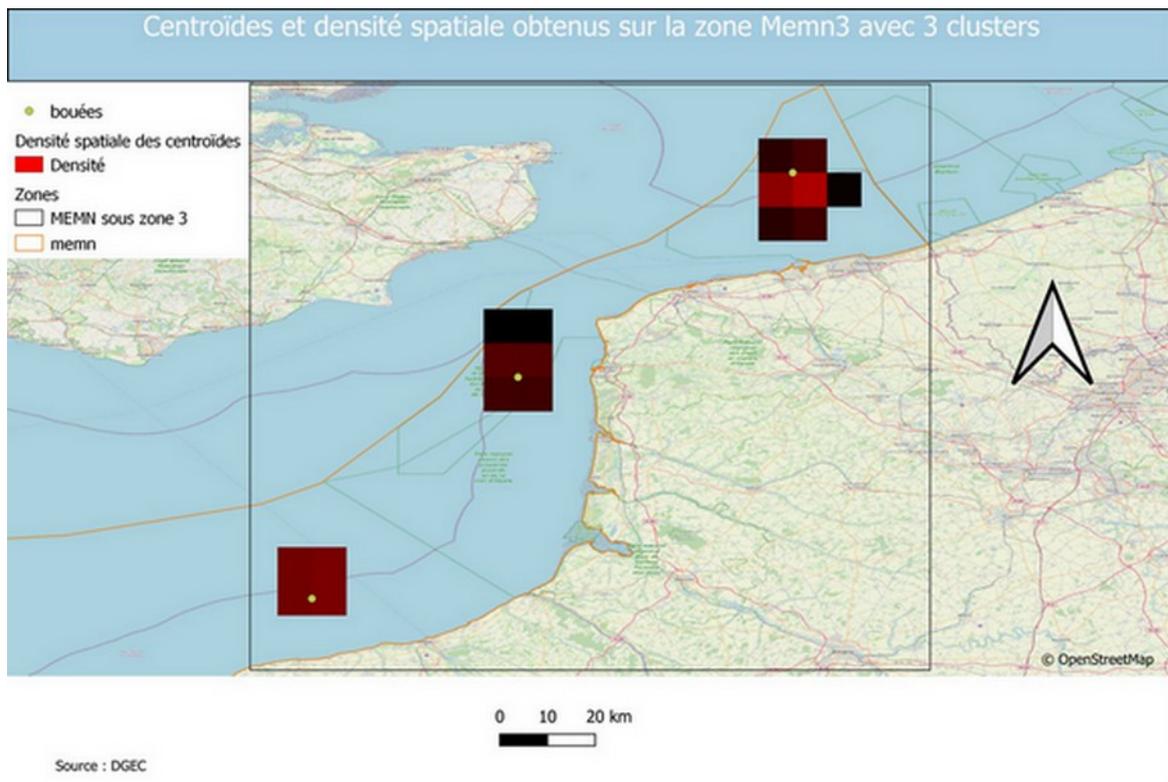


Figure 19: Centroides et densité spatiale obtenus sur la sous-zone MEMN3 avec 3 clusters

Au final, le clustering obtenu sur la zone MEMN (cf Figure 20) montre une bonne cohérence spatiale.

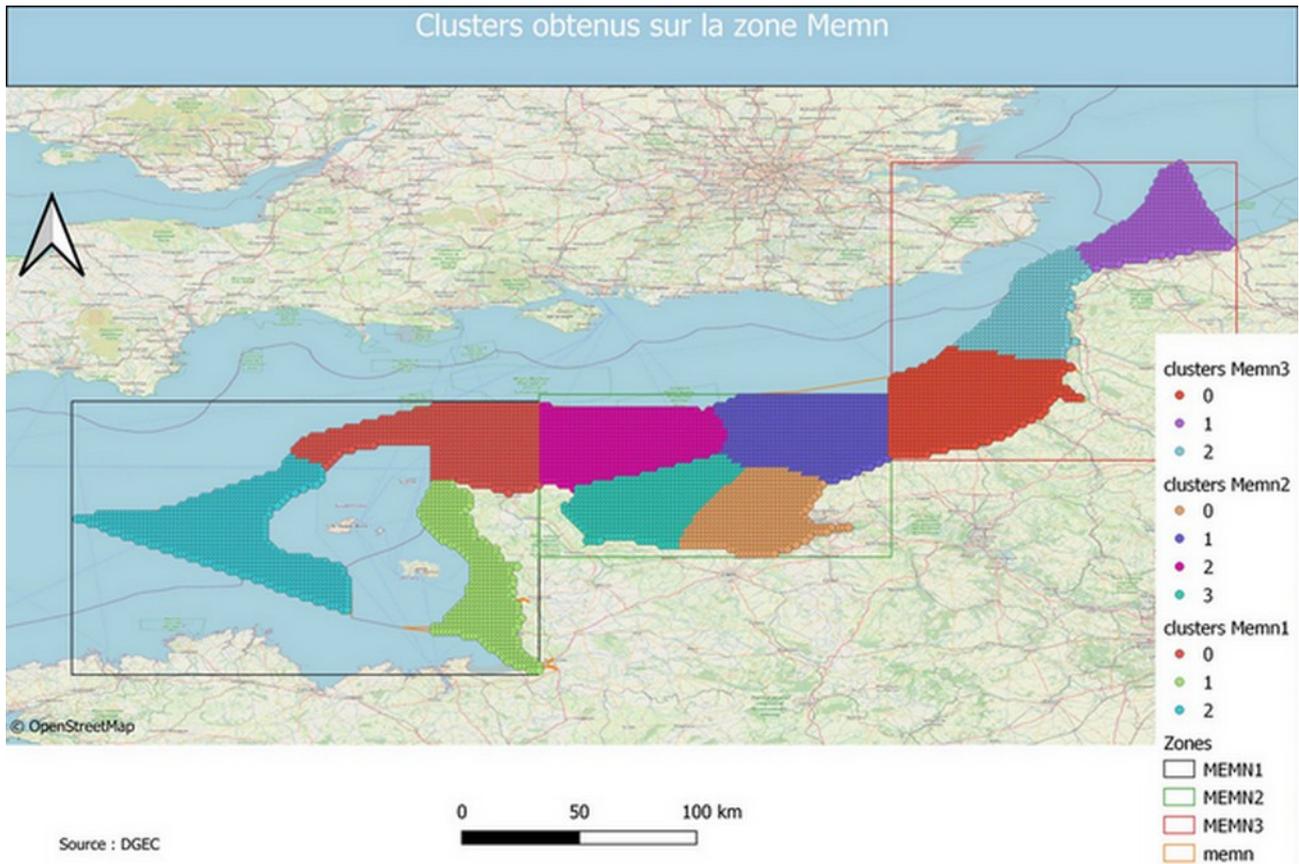


Figure 20: Clustering obtenu sur la zone MEMN

3.2. Interface MEMN-NAMO

Sur la sous-zone d'interface MEMN-NAMO, le nombre de clusters nécessaires afin d'atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 21). Afin d'assurer une cohérence spatiale des clusters obtenus, un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes et une cohérence avec le nombre de clusters retenus pour les sous-zones Memn1 et Nam04 (6 clusters au total), nous retenons finalement 6 clusters sur cette sous-zone.

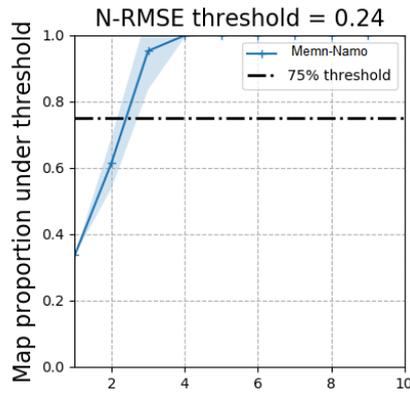
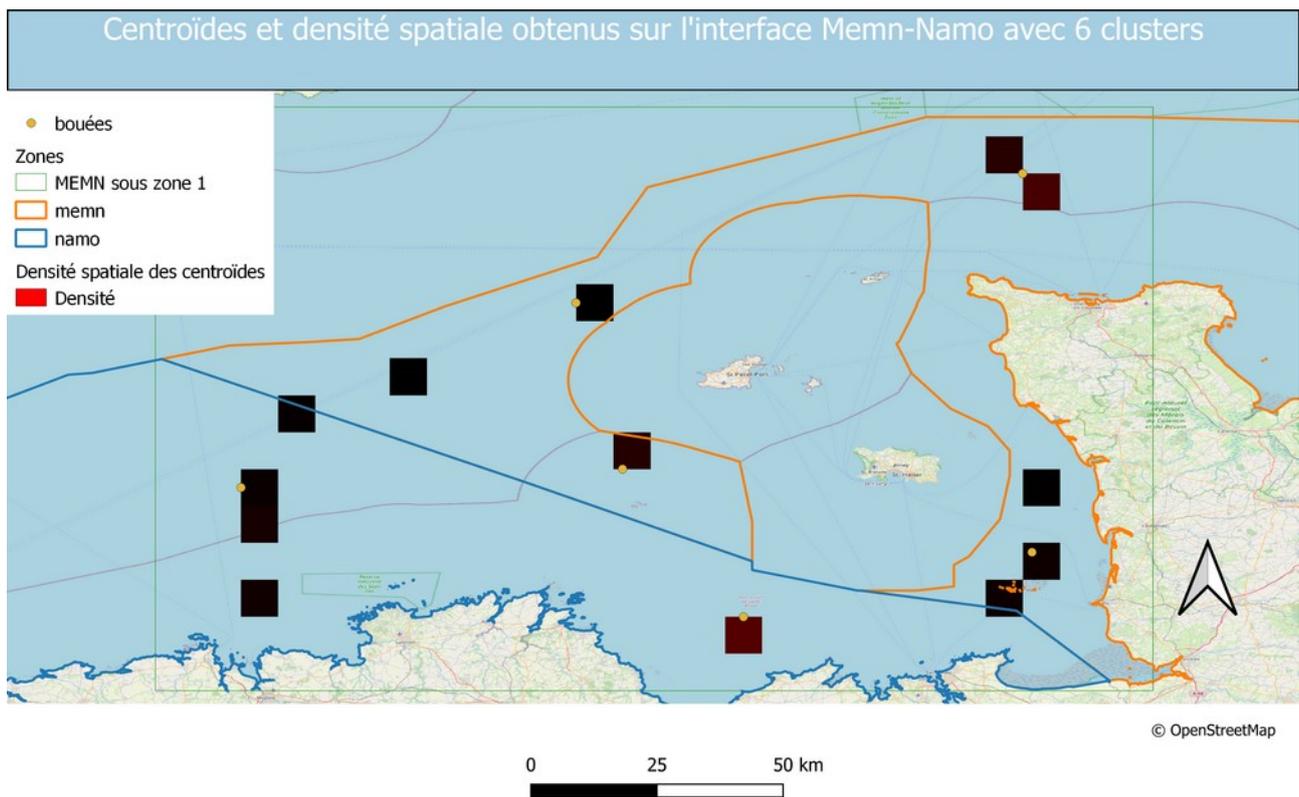


Figure 21: Proportion de la sous-zone d'interface MEMN-NAMO en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes

La densité spatiale des centroïdes montre une bonne concentration autour de 6 zones préférentielles (cf Figure 22). La position retenue des bouées pour la construction des clusters est aussi proche de la majorité des maxima de densité. Ainsi les résultats obtenus avec 6 clusters semblent satisfaisants sur la sous-zone d'interface MEMN-NAMO.



Source : DGEC

Figure 22: Centroïdes et densité spatiale obtenus sur la sous-zone d'interface MEMN-NAMO avec 6 clusters

Au final, le clustering obtenu sur la sous-zone d'interface MEMN-NAMO est décrit dans la Figure 23.

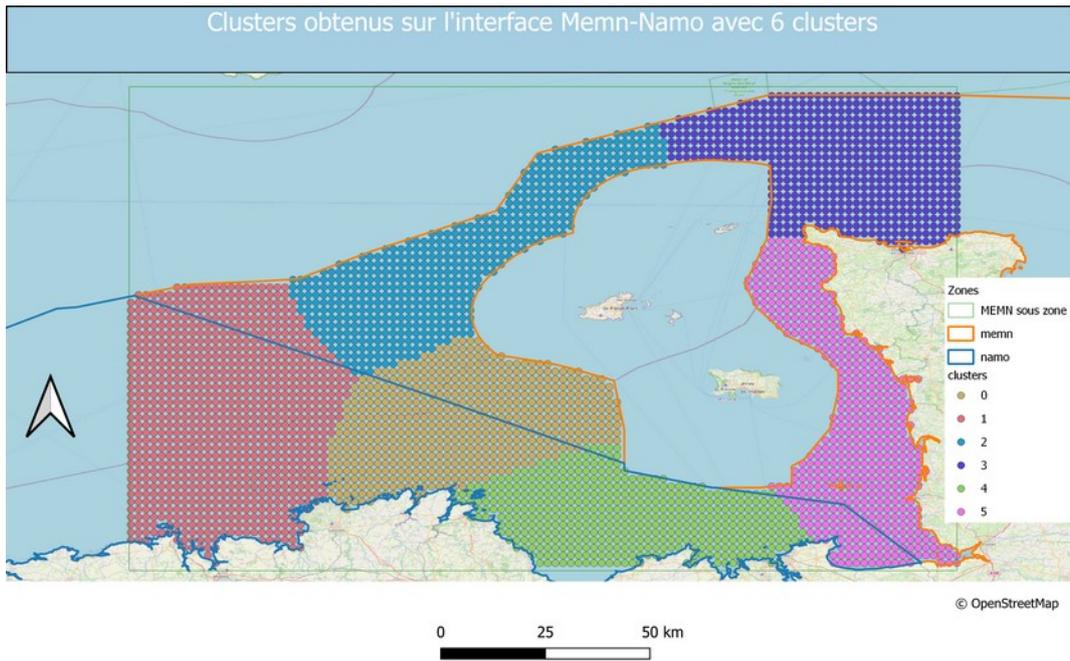


Figure 23: Clustering obtenu sur la sous-zone d'interface MEMN-NAMO

Le clustering obtenu est assez cohérent avec les clusters obtenus pour les sous-zones NAMO4 et MEMN1 (cf Figure 24). Toutefois des différences existent sur l'ouest de la sous-zone MEMN1 : le découpage entre les zones MEMN et NAMO n'est pas forcément cohérent au regard des clusters obtenus sur l'interface MENM/NAMO.

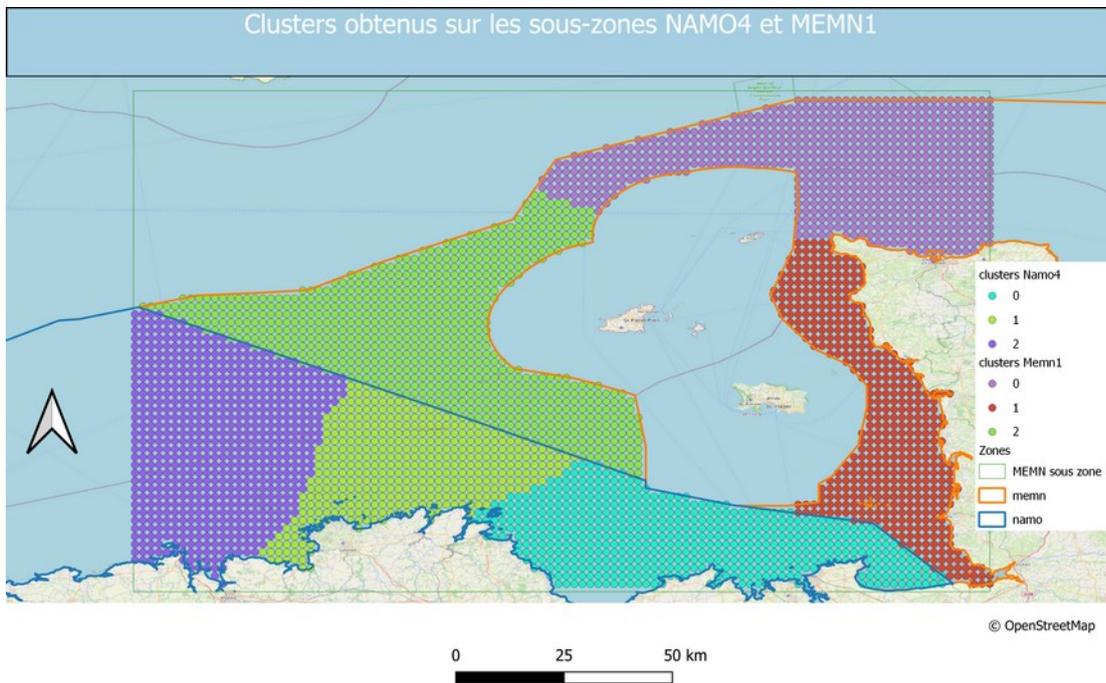


Figure 24: Clusters obtenus sur les sous-zones NAMO4 et MEMN1

3.3. Zones NAMO

Sur la sous-zone NAMO1, le nombre de clusters nécessaires afin d'atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 25). Afin d'assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons finalement 4 clusters sur cette sous-zone.

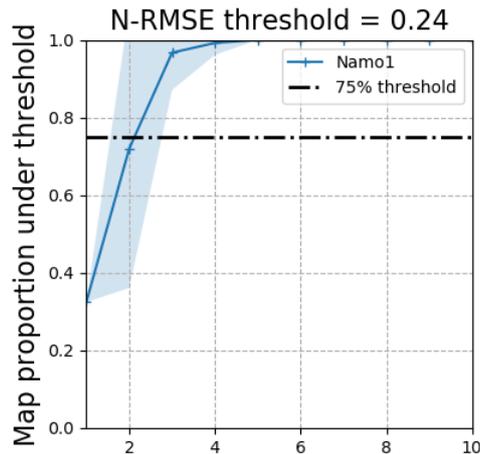


Figure 25: Proportion de la sous-zone NAMO1 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 26). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone NAMO1.

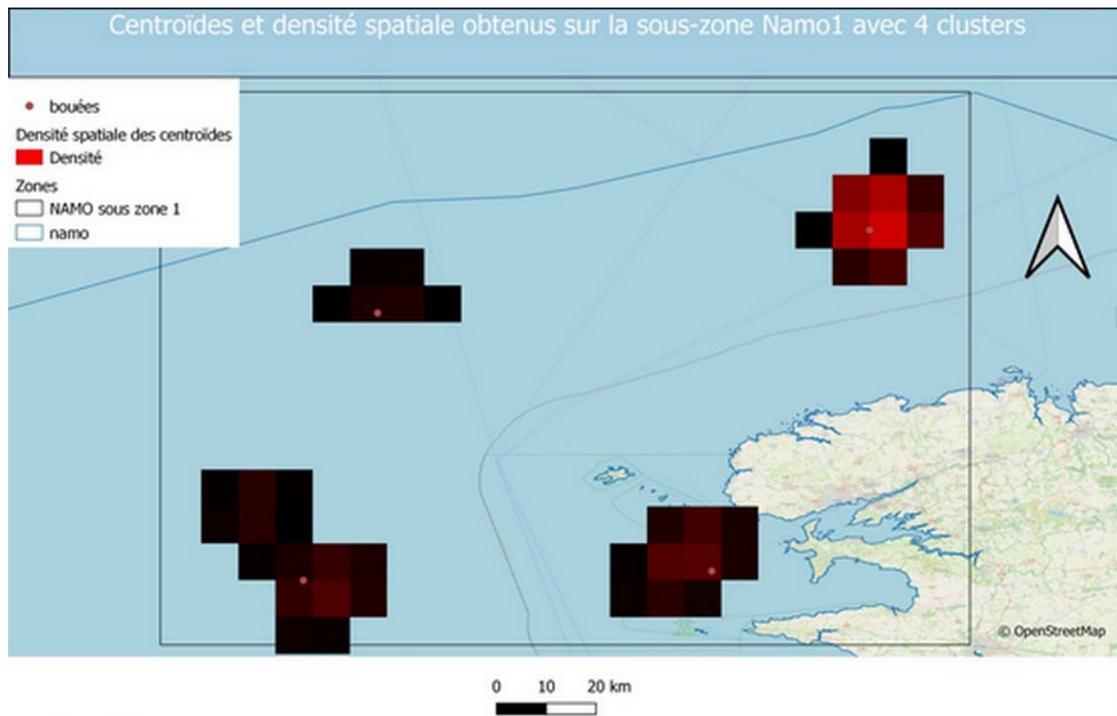


Figure 26: Centroides et densité spatiale obtenus sur la sous-zone NAMO1 avec 4 clusters

Sur la sous-zone NAMO2, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 2 (cf courbe bleue de la Figure 27). Afin d’assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroides, nous retenons finalement 4 clusters sur cette sous-zone.

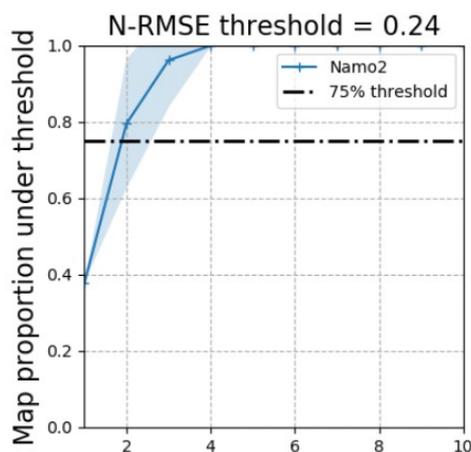


Figure 27: Proportion de la sous-zone NAMO2 en dessous d’un seuil d’erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 28). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone NAMO2.

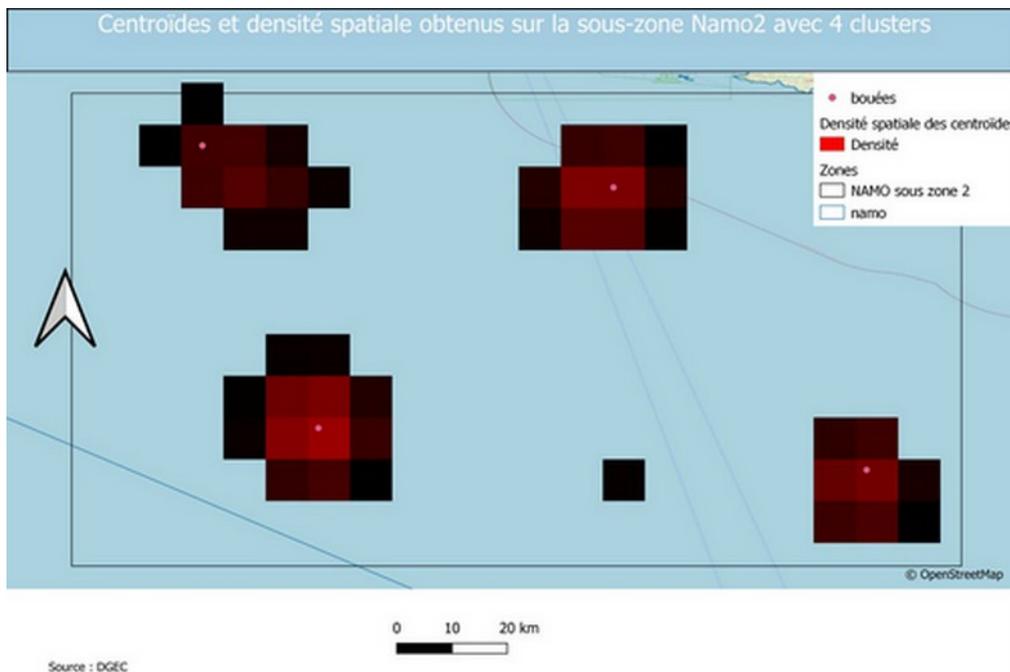


Figure 28: Centroïdes et densité spatiale obtenus sur la sous-zone NAMO2 avec 4 clusters

Sur la sous-zone NAMO3, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 29). Afin d’assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons finalement 4 clusters sur cette sous-zone.

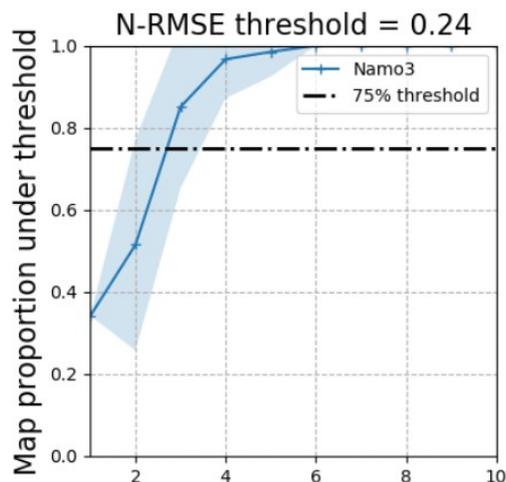


Figure 29: Proportion de la sous-zone NAMO3 en dessous d’un seuil d’erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 30). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone NAMO3.

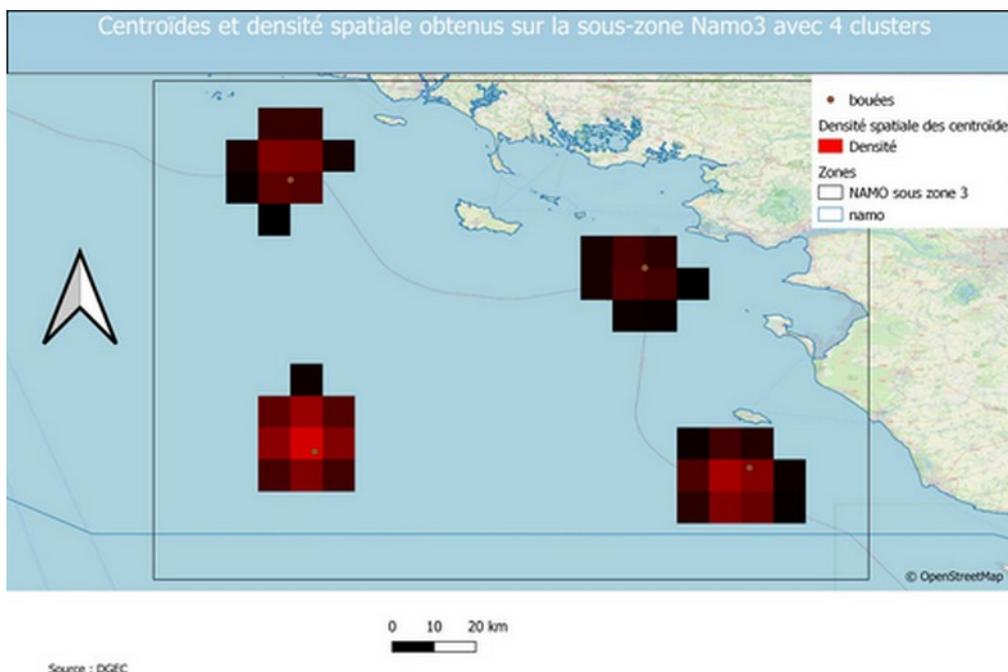


Figure 30: Centroïdes et densité spatiale obtenus sur la sous-zone NAMO3 avec 4 clusters

Sur la sous-zone NAMO4, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 2 (cf courbe bleue de la Figure 31). Afin d’assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons finalement 3 clusters sur cette sous-zone.

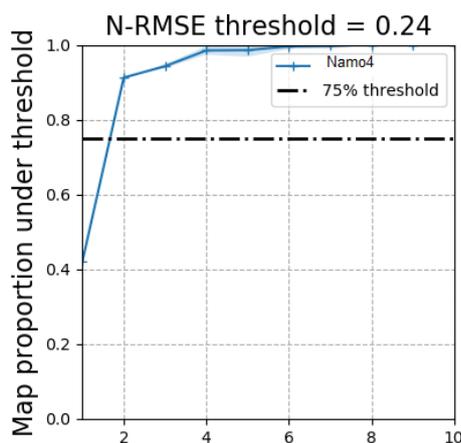


Figure 31: Proportion de la sous-zone d’interface NAMO4 en dessous d’un seuil d’erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 3 zones préférentielles (cf Figure 32). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 3 clusters semblent satisfaisants sur la sous-zone NAMO4.

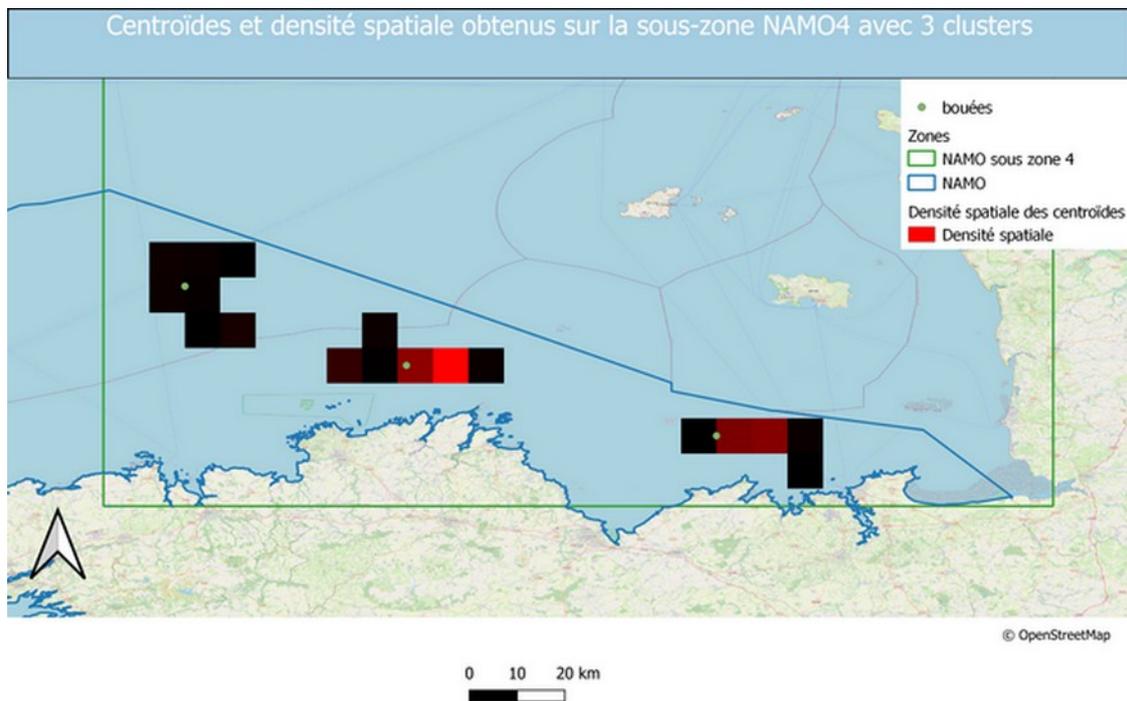


Figure 32: Centroïdes et densité spatiale obtenus sur la sous-zone NAMO4 avec 3 clusters

Au final, le clustering obtenu sur la sous-zone NAMO4 est décrit dans la Figure 33.

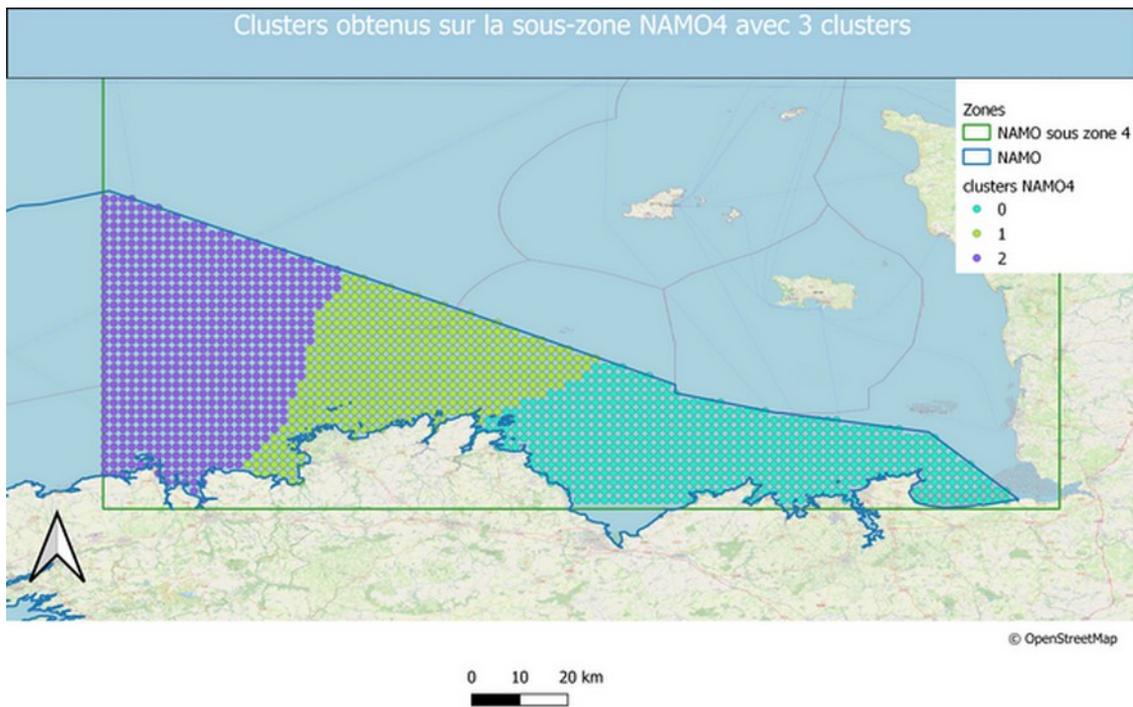


Figure 33: Clustering obtenu sur la sous-zone NAMO4

Le clustering obtenu sur la zone NAMO (cf Figure 34) montre une assez bonne cohérence spatiale, même si de légers hiatus sont parfois présents entre sous-zones.

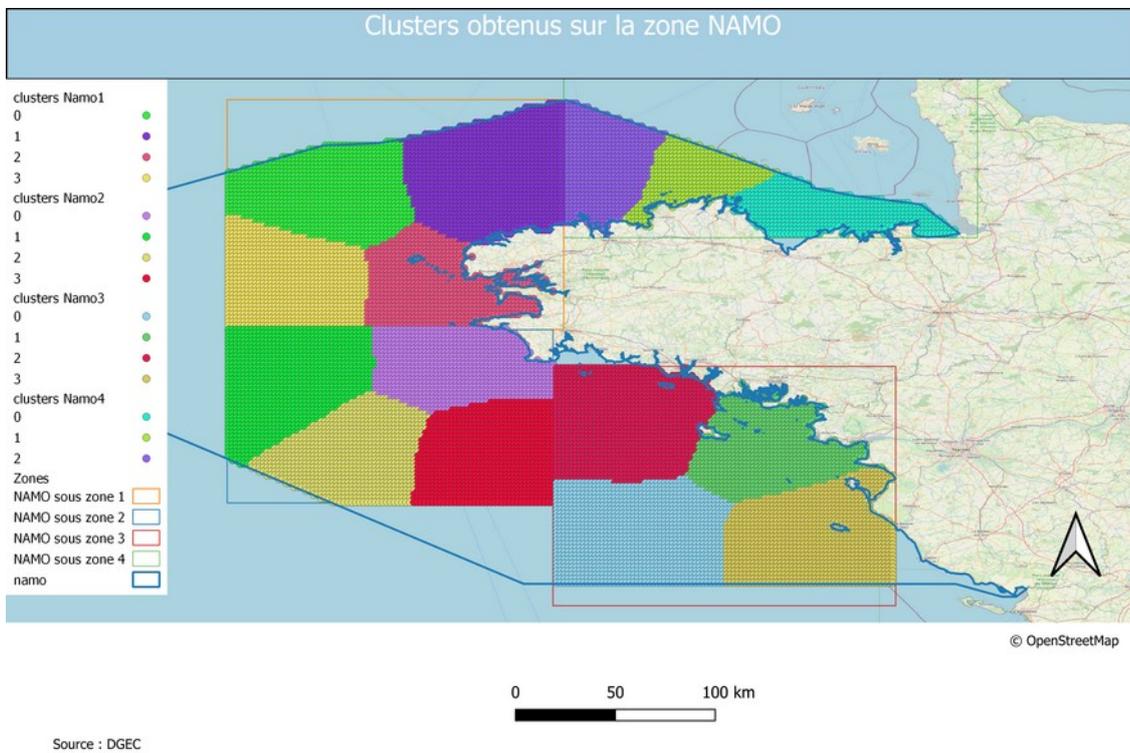


Figure 34: Clustering obtenu sur la zone NAMO

3.4. Interface NAMO-SA

Sur la sous-zone d'interface MEMN-NAMO, le nombre de clusters nécessaires afin d'atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24 est de 3 (cf courbe bleue de la Figure 35). Afin d'assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons finalement 4 clusters sur cette sous-zone.

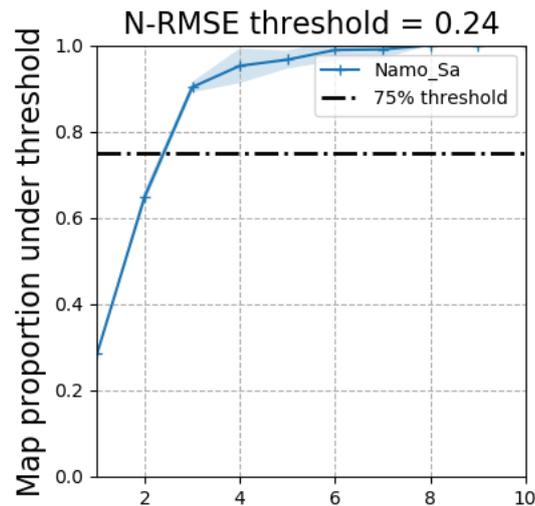


Figure 35: Proportion de la sous-zone d'interface NAMO-SA en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 36). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 3 clusters semblent satisfaisants sur la sous-zone d'interface MEMN-NAMO

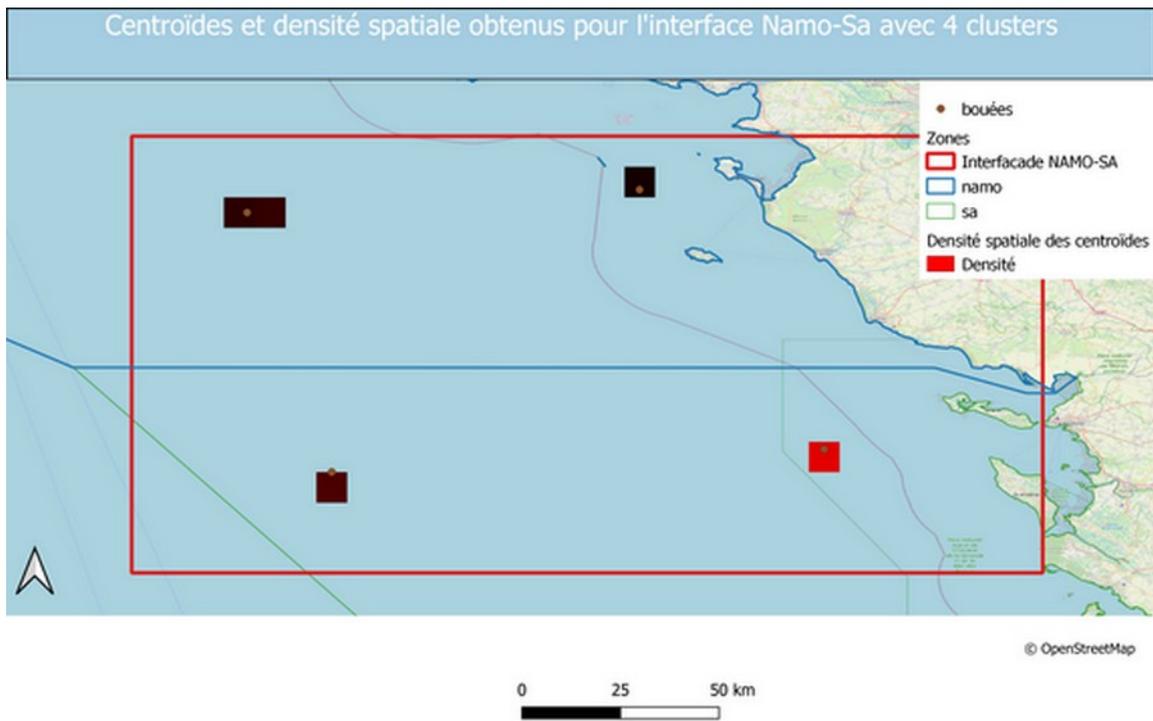


Figure 36: Centroides et densité spatiale obtenus sur la sous-zone d'interface NAMO-SA avec 4 clusters

Au final, le clustering obtenu sur la sous-zone d'interface NAMO-SA est décrit dans la Figure 37.

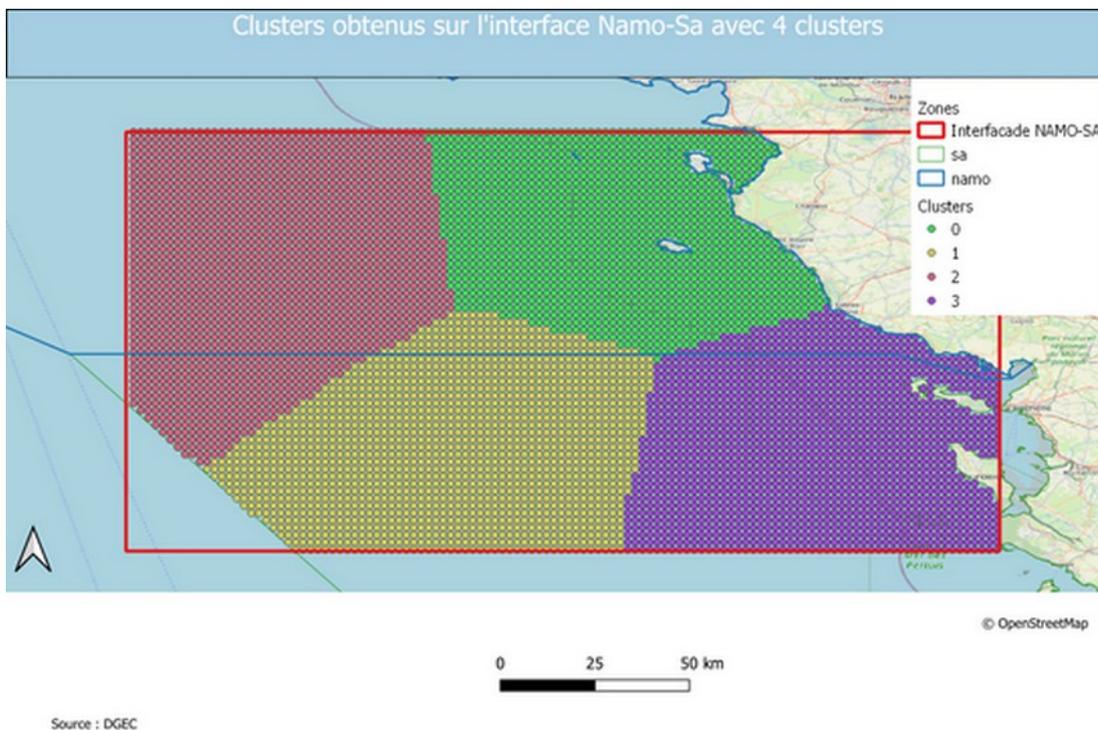


Figure 37: Clusters obtenus sur l'interface NAMO-SA

3.5. Zones SA

Le nombre de clusters nécessaires pour atteindre, en moyenne sur les 5 initialisations GMM différentes, 75 % des sous-zones avec un seuil d'erreur N-RMSE inférieur à 0.24 est, en accord avec les travaux de Marcille et al., 2023, pour la façade Atlantique, moindre que la façade méditerranéenne. Sur la sous-zone SA1, le nombre de clusters nécessaires est ainsi seulement de 2 (cf courbe bleue de la Figure 38). Afin d'assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons finalement 4 clusters sur cette sous-zone.

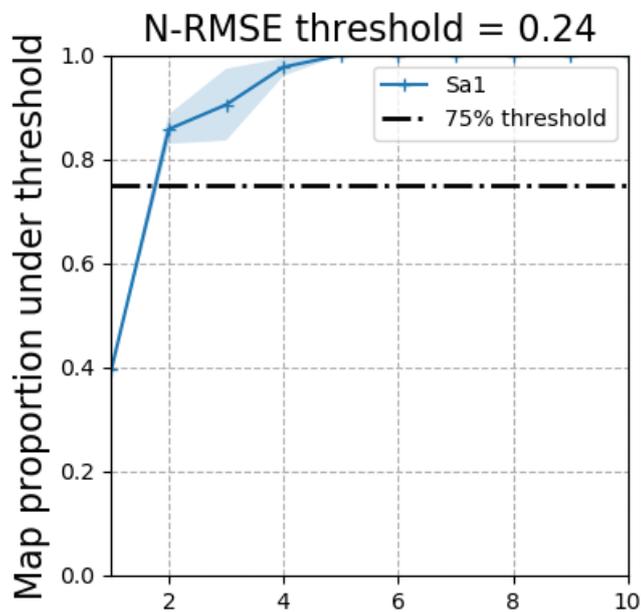


Figure 38: Proportion de la sous-zone SA1 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 39). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone SA1.

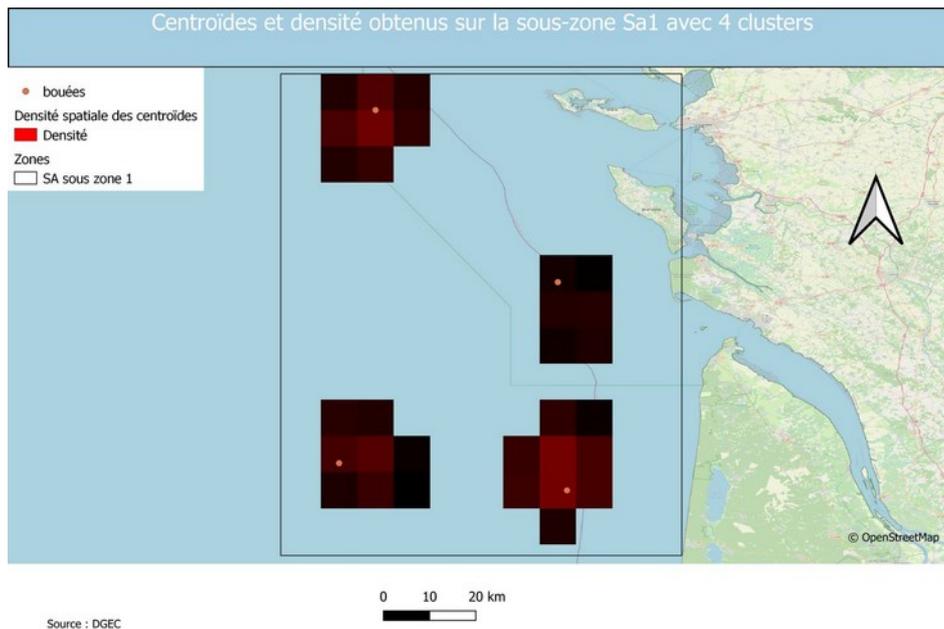


Figure 39: Centroïdes et densité obtenus sur la sous-zone SA1 avec 4 clusters

Sur la sous-zone SA2, le nombre de clusters nécessaires afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24, est aussi seulement de 2 (cf courbe bleue de la Figure 40). Afin d’assurer une cohérence spatiale des clusters obtenus et un meilleur positionnement des bouées par rapport aux maxima de densité des centroïdes, nous retenons aussi finalement 4 clusters sur cette sous-zone.

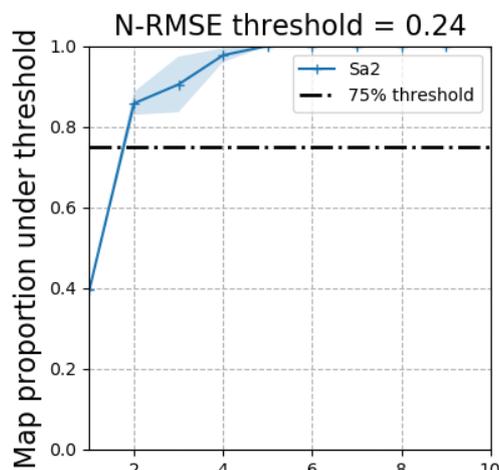


Figure 40: Proportion de la sous-zone SA2 en dessous d’un seuil d’erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 4 zones préférentielles (cf Figure 41). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 4 clusters semblent satisfaisants sur la sous-zone SA2.

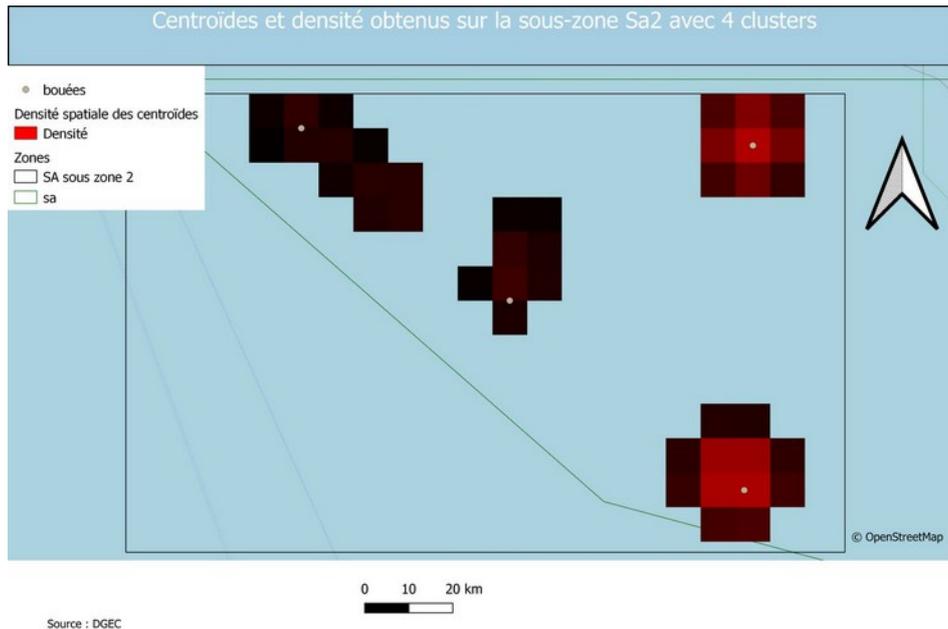


Figure 41: Centroïdes et densité obtenus sur la sous-zone SA2 avec 4 clusters

Enfin sur la sous-zone SA3, le nombre de clusters nécessaires afin d'atteindre, en moyenne sur les 5 initialisations GMM différentes, au moins 75 % de points avec un N-RMSE inférieur à 0.24, est plus élevé et atteint 4 (cf courbe bleue de la Figure 42).

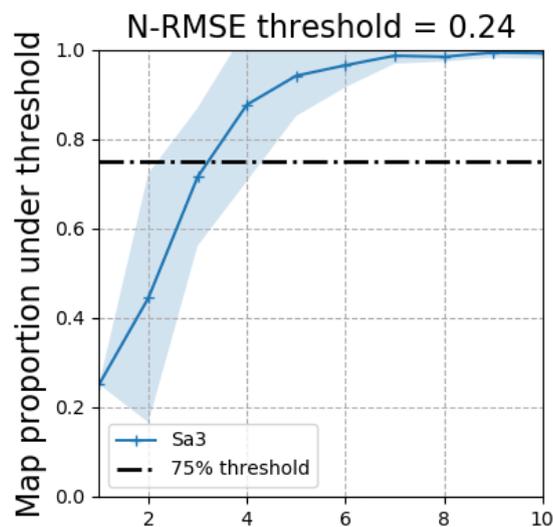


Figure 42: Proportion de la sous-zone SA3 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La majorité des bouées sont bien placées par rapport aux maxima de densité (cf Figure 43). Le clustering obtenu peut donc être jugé satisfaisant est proche de celui qui serait obtenu avec des centroïdes placés exactement au niveau de chacun des maxima de densité.

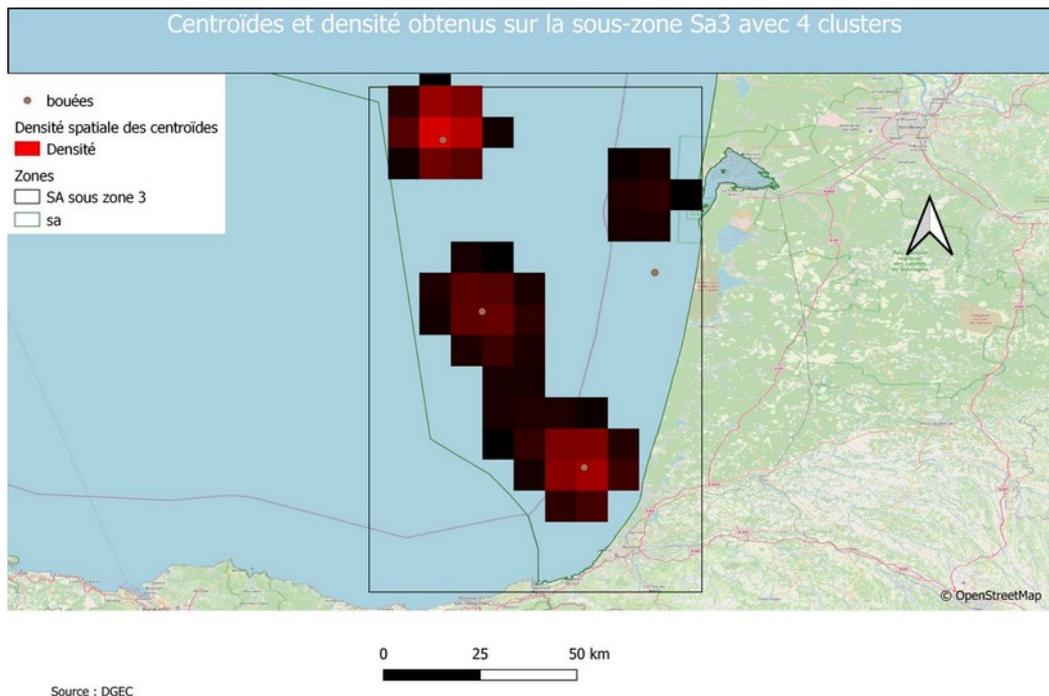


Figure 43: Centroides et densité spatiale obtenus sur la sous-zone SA3 avec 4 clusters

Au final, le clustering obtenu sur la zone SA (cf Figure 44) montre une assez bonne cohérence spatiale, même si des hiatus sont souvent présents à la jonction entre sous-zones.

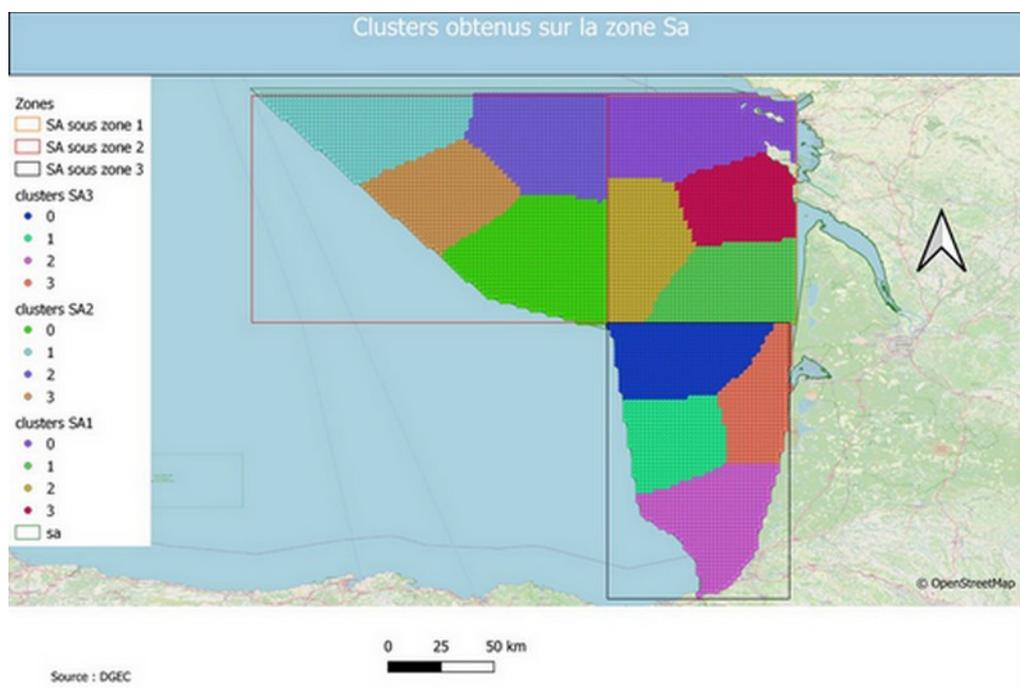


Figure 44: Clustering obtenu sur la zone SA

3.6. Zones en Méditerranée

Sur la sous-zone Med1, autour de la Corse, le seuil de N-RMSE de 0.24 nécessiterait un nombre de capteurs de 10 (cf courbe bleue de la Figure 45). Cela n'est guère surprenant vu la complexité des flux autour de l'île en raison de la multitude des régimes de vent (libeccio, mistral, sirroco, grécale et brises pour les vents principaux) et de l'influence du relief.

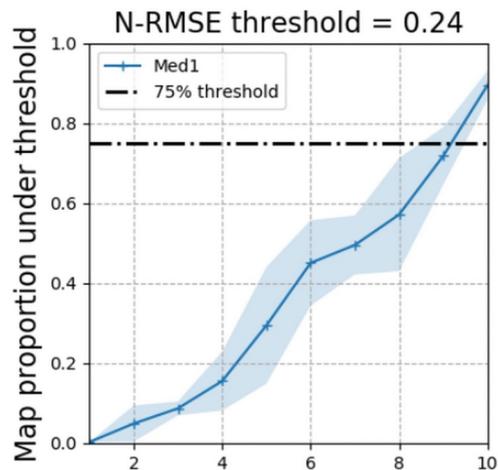


Figure 45: Proportion de la sous-zone Med1 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

Nous nous limiterons à la demande de la DGEC au clustering obtenu avec 6 capteurs. La densité spatiale des centroïdes montre une bonne concentration autour de 6 zones préférentielles (cf Figure 46). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 6 clusters semblent satisfaisants même si, pour ce nombre de clusters, une proportion de 75 % de points au-dessous du seuil de N-RMSE de 0.24 n'est pas atteinte.

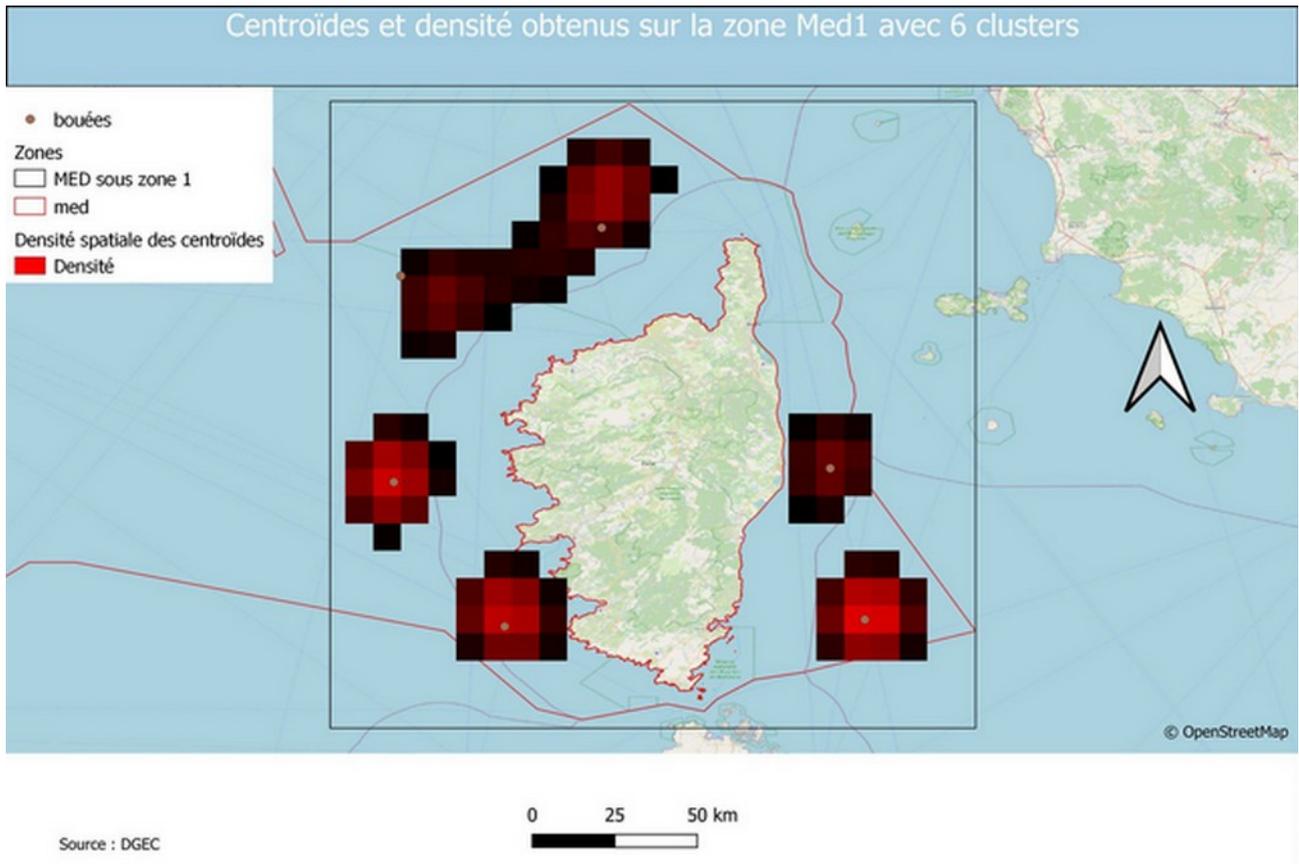


Figure 46: Centroïdes et densité spatiale obtenus sur la zone Med1 avec 6 clusters

Sur la sous-zone Med2, le seuil de N-RMSE de 0.24 nécessiterait un nombre de capteurs de 5 (cf courbe bleue de la Figure 47).

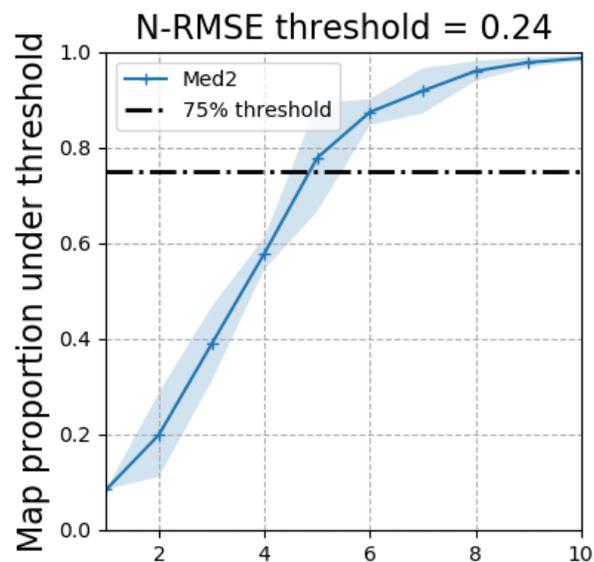


Figure 47: Proportion de la sous-zone Med2 en dessous d'un seuil d'erreur N-RMSE de 0.24 avec 5 initialisations GMM différentes.

La densité spatiale des centroïdes montre une bonne concentration autour de 5 zones préférentielles (cf Figure 48). La position retenue des bouées pour la construction des clusters est aussi proche de chacun des maxima de densité. Ainsi les résultats obtenus avec 5 clusters semblent satisfaisants.

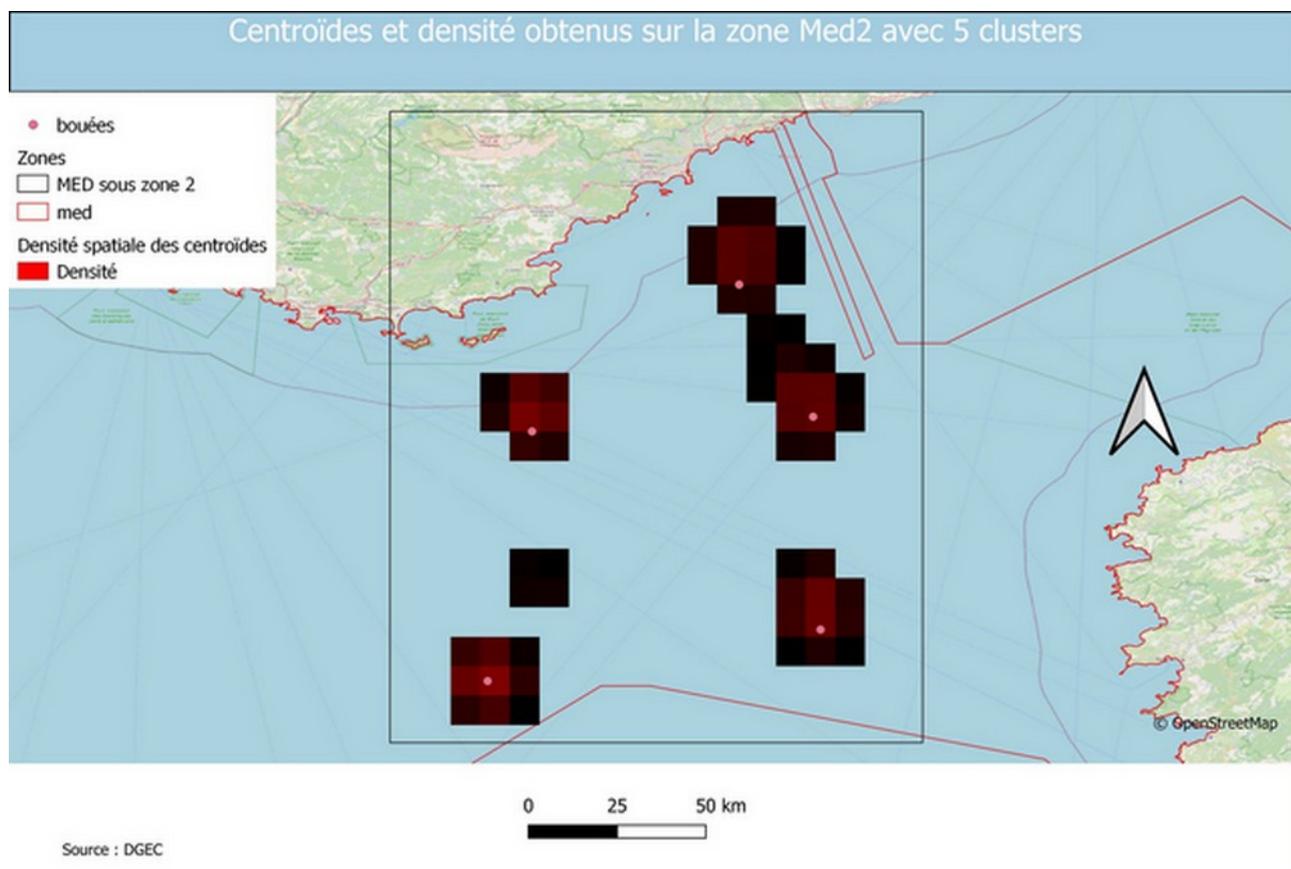


Figure 48: Centroïdes et densité obtenus sur la sous-zone Med2 avec 5 clusters

Pour la sous-zone Med3, afin d'assurer une cohérence spatiale des clusters obtenus, nous retenons un nombre de centroïdes de 6. Au final, les clusters obtenus sont assez cohérents spatialement entre sous-zones sur la zone Med, en particulier entre les sous-zones Med1 et Med2 (cf Figure 49).

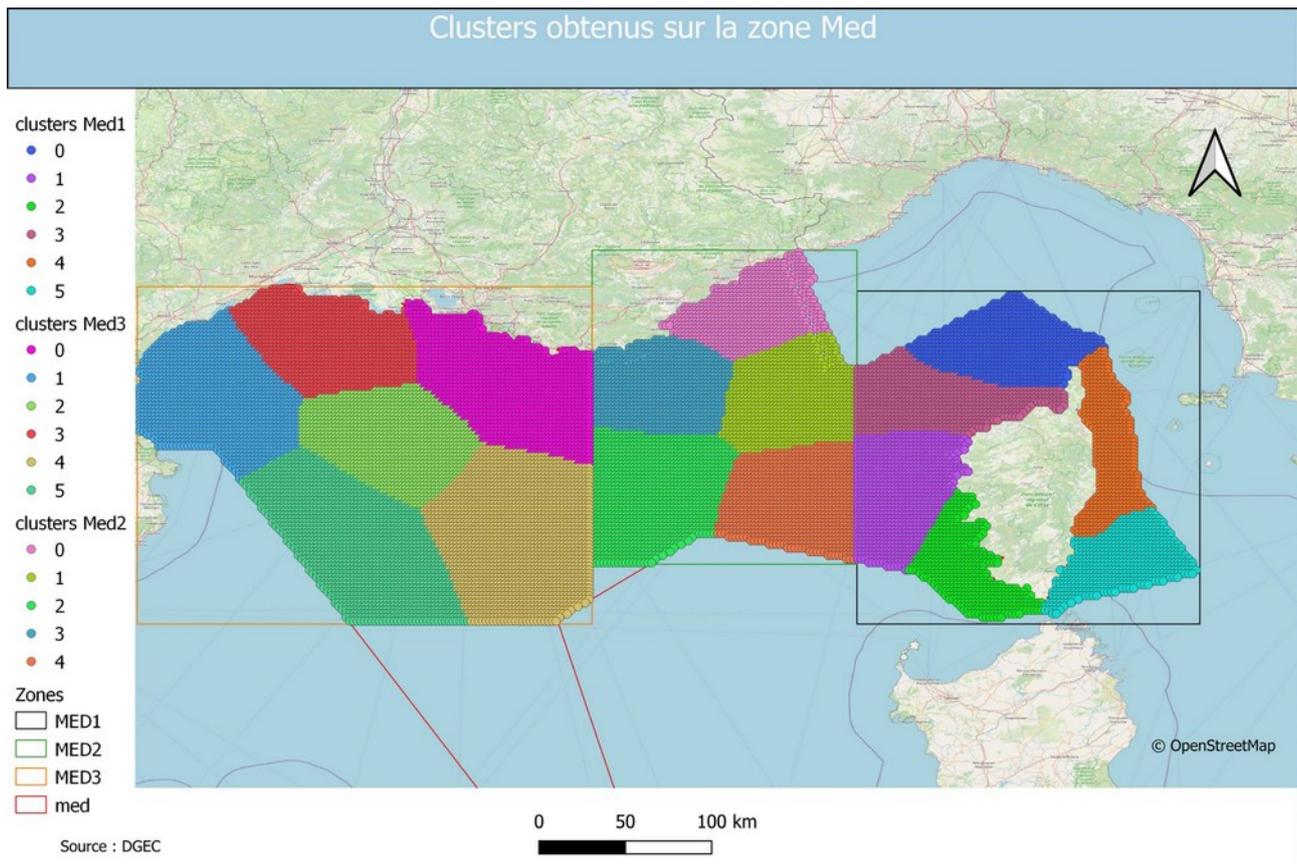


Figure 49: Clustering obtenu sur la zone Med

4. Livrables

Pour chaque sous-zone les livrables fournis à la DGEC sont les suivants :

- un fichier csv avec la position des centroïdes retenue pour les clusters
- un shapefile de contour des clusters
- un fichier csv contenant la latitude et la longitude des points de grille ainsi que leur cluster d'appartenance
- un fichier raster de densité spatiale des centroïdes

5. Conclusions et perspectives

Ce lot préliminaire avait deux objectifs pour chacune des façades :

- Pour chacune des sous-zones (MEMN sous zone 3 par exemple), identifier 3 à 6 zones homogènes en termes de climatologie de vent à 200 m.
- Pour chacune de ces zones homogènes, un point de mesure “optimal” (ou une enveloppe) le plus représentatif de la zone a été identifié.

Pour cela, cette étude s’est appuyée sur les travaux de Marcille et al., qui repose sur l’utilisation de la méthode GMM (Gaussian Mixture Model) afin d’identifier des clusters dont les centroïdes sont proposés comme emplacement des capteurs. En entrée, les données de vent à une hauteur de 200 m du modèle haute résolution AROME de Météo-France ont été utilisées sur la période 2018-2022 et une grille de résolution 0.025°.

Afin de tenir compte de la sensibilité importante de cette méthode aux conditions initiales, 100 tirages ont été effectués pour chacune des sous-zones : ces tirages permettent alors de définir une densité spatiale des centroïdes par Kernel Density Estimation (KDE). Les enveloppes des maxima de densité permettent alors de déterminer des zones préférentielles d’installation des capteurs.

Le choix du nombre de clusters pour chaque sous-zone, nécessaire pour appliquer la méthode GMM, a été effectué à l’aide de différents critères :

- la recherche d’un coude dans la courbe de gradient du score BIC. Cette méthode permet de déterminer une plage de valeurs possibles pour le nombre de clusters
- le nombre minimal de clusters afin d’atteindre, en moyenne sur les 5 initialisations GMM différentes, un seuil de N-RMSE inférieur à 0.24 sur au moins 75 % des points de grille de la sous-zone. Ce seuil de N-RMSE a été déterminé sur la sous-zone Med3, l’une des plus complexes, afin qu’un nombre de clusters de 5 ou 6 soit suffisant sur cette sous-zone.
- la densité spatiale des centroïdes et en particulier la présence d’un nombre adéquat de zones de maxima
- la cohérence spatiale des clusters obtenus entre sous-zones

En cohérence avec les travaux de Marcille et al., 2023, les résultats obtenus font ressortir un nombre de clusters plus important en Méditerranée (avec 5 ou 6 clusters par sous-zones) par rapport aux façades Atlantique et Manche (3 à 4 clusters par sous-zones). Les régimes de vent particulièrement complexes autour de la Corse, nécessiteraient même au moins 10 capteurs sur la sous-zone Med1 afin d’en apprécier toute la variabilité, mais les résultats obtenus avec 6 capteurs sont déjà assez satisfaisants même si la reconstruction du champ de vent serait d’une qualité moindre que pour les autres sous-zones.

Un clustering est fourni en choisissant parmi les 100 tirages celui pour lequel le plus de centroïdes obtenus sont proches des maxima de densité. Pour la majorité des sous-zones, une grande partie des centroïdes sont proches des maxima, indiquant une bonne qualité du clustering choisi.

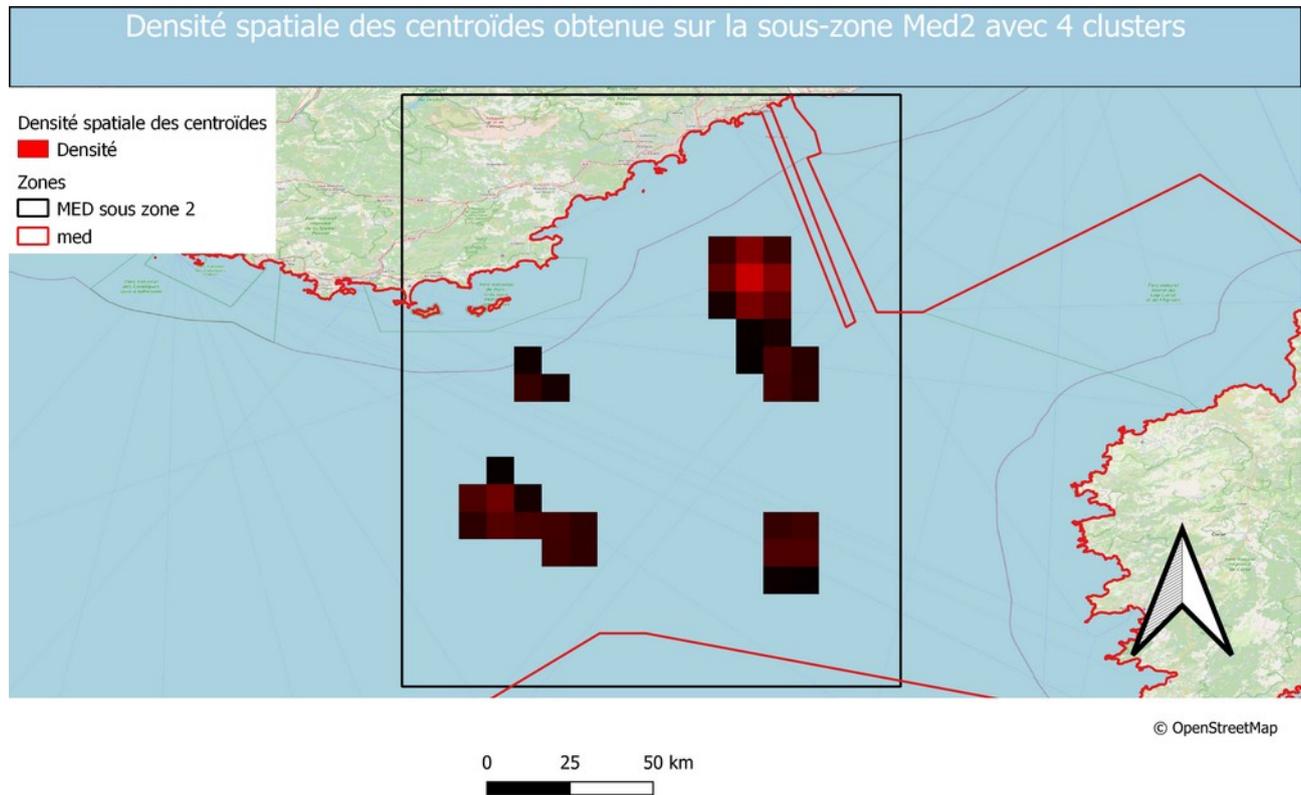
La sensibilité au choix des niveaux de vent a été aussi étudiée. Les résultats obtenus sont sensibles au choix du niveau, en particulier les clusters obtenus et les maxima de densité sont différents entre les niveaux les plus proches du sol (10 m) et ceux plus élevés (200 m). À la demande de la DGEC

pour être représentatif de la hauteur des turbines, ce sont les données AROME 2018-2022 au niveau 200 mètres qui ont été analysées pour le clustering.

La méthode de clustering GMM semble être une solution simple mais efficace pour le placement de capteurs clairsemés. Appliqué à la reconstruction de l'énergie éolienne offshore, il permet un placement optimal des capteurs. Les contraintes économiques, notamment liées à la bathymétrie, ainsi que les contraintes de trafic maritime, ne sont cependant pas prises en compte dans ces travaux. L'utilisation des cartes de densité des centroïdes permet cependant une certaine latitude dans le positionnement des capteurs.

Toutefois, dans certains cas, cette latitude dans le positionnement des capteurs pourrait ne pas être suffisante au regard de la multitude des contraintes. Il pourrait alors être nécessaire, afin d'obtenir une solution satisfaisante de positionnement des bouées, de tester la méthode de clustering GMM avec un nombre différent de clusters sur certaines zones. L'ANNEXE 1, qui fournit un exemple de densité spatiale des centroïdes obtenue avec 4 clusters sur la zone Med2, montre des évolutions parfois importantes de la position de certains maxima par rapport à ceux obtenus avec 5 clusters (cf Figure 48). Il faut toutefois garder en tête que cette méthode pourrait amener à des placements moins optimaux des capteurs, surtout lorsque le nombre de clusters serait moindre que la solution présentée dans ce rapport.

ANNEXE 1



Source : DGEC

Figure 50: Densité spatiale des centroïdes obtenue sur la sous-zone Med2 avec 4 clusters

Bibliographie

- Marcille, Robin, et al. "Gaussian Mixture Models for the Optimal Sparse Sampling of Offshore Wind Resource." *Wind Energy Science Discussions* (2022): 1-24
- Murray Rosenblatt. "Remarks on Some Nonparametric Estimates of a Density Function." *Ann. Math. Statist.* 27 (3) 832 - 837, September, 1956. <https://doi.org/10.1214/aoms/1177728190>
- Reynolds, D. A.: Gaussian mixture models, *Encyclopedia of biometrics*, 741, 659–663, 2009.
- Schwarz, G.: Estimating the dimension of a model, *Ann. Stat.*, 1st edition, <https://doi.org/10.1214/aos/1176344136>, 461–464, 1978
- Seity, Y. et al., 2010, The AROME-France Convective-Scale Operational Model <https://journals.ametsoc.org/view/journals/mwre/139/3/2010mwr3425.1.xml>

FIN DE DOCUMENT