

SAE 3.EMS.01 Recueil et analyse de données par échantillonnage ou plan d'expérience

Echantillonnage stratifié - Données EVHOE

Réaliser par Enora Aveline, Rose Nerriec, Ariane Souchard

IUT Vannes – Année universitaire 2024/2025

BUT Science des Données – Semestre 3

SAE 301 : Recueil et analyse de données par échantillonnage ou plan d'expérience

Partie 1. Analyse Exploratoire des Données

Objectif de l'analyse

L'objectif principal de cette analyse exploratoire est d'étudier les captures de *Merluccius merluccius* (*merlu*) à partir des données collectées lors de la campagne **EVHOE** entre 2008 et 2019 dans le Golfe de Gascogne et en Mer Celtique. Cette étude permettra de comprendre les schémas de capture des merlus selon différentes variables telles que l'année, la localisation géographique, la profondeur et les caractéristiques des chalutiers.

Jeux de données et préparation

Deux jeux de données principaux ont été utilisés :

- **Save_Datras\$datras_HH.full** : Contient les informations générales sur les traits de chalut, incluant l'année (Year), la position géographique (long, lati), la profondeur (Depth) et la distance parcourue (Distance).
- **Save_Datras\$datras_sp.HL.full** : Données détaillées sur les captures, avec le nombre d'individus par espèce (TotalNo), la taille des individus (LngtClass) et le sexe.

Ces deux jeux ont été reliés grâce aux colonnes communes Year, StNo et HaulNo. Cela nous a permis d'obtenir un tableau final combinant les informations générales sur les traits de chalut et les données spécifiques des captures.

Traitement des données manquantes

Une vérification des données a révélé :

- Dans **Haul_df** : Des valeurs manquantes dans les variables Depth (5 valeurs) et Area_swept (702 valeurs). Les données manquantes ont été imputées par la **moyenne des valeurs existantes**.
- Dans **Catch_df** : Aucune donnée manquante n'a été détectée.

Visualisation des données

a) Nombre de merlus capturés par année

Nous avons calculé le nombre total de merlus capturés pour chaque année de la période étudiée. Le graphique montre des fluctuations importantes, avec des captures maximales en **2012** et **2016**. Ces années se distinguent nettement par rapport aux autres périodes. [\(Figure 1\)](#)

b) Carte des strates et positions des captures

Une carte a été réalisée pour représenter les positions des traits de chalut (points rouges) en superposition aux strates géographiques du Golfe de Gascogne et de la Mer Celtique. Cette carte met en évidence la répartition spatiale des échantillonnages selon les strates de profondeur. [\(Figure 2\)](#)

c) Nombre de merlus en fonction de la profondeur

Nous avons analysé la relation entre la profondeur des traits de chalut et le nombre de merlus capturés. Il apparaît que les captures sont principalement concentrées dans des profondeurs comprises entre **0 et 100 mètres**, ce qui indique une préférence des merlus pour les zones peu profondes. [\(Figure 3\)](#)

d) Distribution des sexes des merlus

Nous avons comparé les captures selon le sexe des individus :

- La majorité des captures correspond à des individus de sexe **inconnu (U)**.
- Les individus mâles (**M**) et femelles (**F**) sont présents en proportions beaucoup plus faibles. Cette distribution permet de mieux comprendre la démographie des captures. *(Figure 4)*

e) Top 10 des chalutiers ayant capturé le plus de merlus

Nous avons regroupé les captures par chalutier afin d'identifier les 10 plus performants sur l'ensemble de la période. Le chalutier **73** se distingue nettement avec environ **11 537 captures**, suivi des chalutiers 62 et 59. *(Figure 5)*

Conclusion de l'analyse exploratoire

L'analyse exploratoire des données nous a permis de dégager plusieurs tendances :

1. Les captures de merlus fluctuent considérablement d'une année à l'autre, avec des pics notables en 2012 et 2016.
2. La répartition géographique des traits de chalut confirme une couverture efficace des zones d'étude dans le Golfe de Gascogne et la Mer Celtique.
3. Les captures sont principalement concentrées dans des zones de faible profondeur (0-100 mètres).
4. La distribution des sexes montre une prédominance des individus de sexe inconnu, ce qui peut s'expliquer par des limitations techniques lors des campagnes.
5. Certains chalutiers, comme le 73, ont contribué de manière significative aux données collectées.

Ces analyses constituent une base solide pour la suite de l'étude, notamment pour l'estimation stratifiée de l'abondance des merlus.

Partie 2. Estimation Stratifiée du Total et Analyse de sa Variance

Dans le contexte de l'échantillonnage stratifié, les formules présentées servent à **estimer le total d'une variable d'intérêt** (par exemple, le **nombre total de merlus dans une population marine**) ainsi que **l'incertitude associée à cette estimation**. Ces **outils statistiques** sont **fondamentaux** pour **obtenir des estimations** et pour **comprendre leurs précisions**. Dans notre cas d'étude, la **population est divisée en 28 strates**.

Pour l'ensemble des formules :

N_h (taille de la population dans la strate) :

- Représentée par la surface de la strate donnée avec la variable **area_strata**
- Exprime le nombre total d'unités possibles dans cette strate

nh (taille de l'échantillon dans la strate) :

- Correspond à la surface chalutée dans la strate donnée avec la variable **Area_swept**
- Exprime le nombre d'unités observées dans cette strate

Y^h (moyenne des observations dans la strate) :

- Représente une estimation moyenne du nombre de merlus capturés par unité de surface chalutée dans la strate

Sh² (variance dans la strate) :

- Mesure la variabilité des captures de merlus dans les traits de chalut au sein de chaque strate

$$\hat{T}_s = \sum_{h=1}^H N_h \cdot \bar{Y}_h$$

Estimateur du total stratifié :

Cette formule **calcule la contribution de chaque strate** en multipliant N_h (taille totale de la strate) par \bar{Y}_h (moyenne de la variable d'intérêt pour cette strate). Ensuite, on **additionne ces contributions** pour obtenir une **estimation du total pour toute la population**.

Comme **chaque strate n'a pas la même taille**, cette formule **ajuste la contribution de chaque strate** en fonction de son **poids dans la population**. Cela permet d'obtenir une **estimation représentative** de l'ensemble de la population, **sans surreprésentation ou sous-représentation de certaines strates**.

$$\text{Var}(\hat{T}_s) = \sum_{h=1}^H N_h^2 \cdot \left(\frac{S_h^2}{n_h} \cdot \frac{N_h - n_h}{N_h - 1} \right)$$

Variance de l'Estimateur du Total Stratifié :

La **variance** prend en compte la **variabilité dans chaque strate**, en **pondérant cette variabilité** par N_h^2 (taille relative de la strate au carré).

Le terme $\frac{S_h^2}{n_h}$ mesure la **précision de l'estimation** à l'intérieur de chaque strate, en fonction de la **variance observée** (S_h^2) et de la **taille de l'échantillon** (n_h).

Le facteur $\frac{N_h - n_h}{N_h - 1}$ ajuste cette **précision** en tenant compte de la **proportion de la population échantillonnée**. Plus l'échantillon est **grand par rapport à la population**, plus l'incertitude diminue. Cette formule mesure l'**incertitude liée à l'estimation** du total stratifié. En **intégrant la variabilité** dans chaque strate et la **taille des échantillons**, elle fournit une **indication précise de la fiabilité** de l'estimation \hat{T}_s . En résumé, elle **indique la confiance** que l'on peut avoir dans les résultats.

Partie 3. Estimation de l'Abondance du Merlu pour l'Année 2012 : Méthodes HTstrata et varest

L'**objectif** de cette étude est **d'estimer l'abondance annuelle des merlus** en utilisant les fonctions **HTstrata** et **varest**, tout en **quantifiant l'incertitude** grâce aux **intervalles de confiance à 95 %**. L'analyse repose sur une **boucle** qui parcourt chaque année des données disponibles, agrège les résultats par strate et calcule **l'aire balayée totale** pour estimer l'abondance. La variance associée est ensuite obtenue avec **varest** pour déterminer les intervalles de confiance. Les résultats (**abondance, variance et intervalles**) sont synthétisés dans un tableau nommé **estimations_annuels**.

Les résultats sont représentés sous quatre graphiques pour observer les tendances et variations annuelles :

- Graphique en barres : Il simplifie la comparaison visuelle des abondances annuelles (2008-2019). L'année **2012** se démarque avec une **abondance élevée**, tandis que les **autres années** sont plus **stables** avec des **fluctuations en 2016 et 2018**. (*Figure 6*)
- Graphique linéaire avec intervalles de confiance : Ce graphique principal présente des points qui relie les estimations annuelles, révélant des tendances marquées comme les **pics en 2012 et 2017** et des valeurs minimales en 2010 et 2018. De plus, nous avons ajouté barres d'erreur autour des estimations (intervalle de confiance à 95%). Les années comme 2012 présentent des

intervalles de confiance larges, traduisant **une forte incertitude**, tandis que des années comme **2009 et 2016** montrent **des estimations plus fiables avec des intervalles étroits**. (*Figure 7*)

Les résultats révèlent une **variabilité interannuelle importante** dans l'abondance des merlus, particulièrement en **2012**, où **l'abondance est élevée** mais accompagnée d'une **incertitude marquée**. À l'inverse, des années comme **2009 et 2016** présentent des valeurs **plus faibles mais plus précises**. Ces variations pourraient **être liées** à des facteurs **environnementaux** (conditions climatiques) ou à des fluctuations **de l'effort de pêche**.

Pour conclure, la méthode utilisant **Hstrata** et **varest** permet **d'estimer l'abondance** des merlus tout en tenant compte de **l'incertitude** grâce aux **intervalles de confiance**. Les **graphiques** produits offrent une vision claire des **tendances temporelles et de la précision des résultats**.

Pour **améliorer l'analyse**, il serait utile **d'ajouter des étiquettes** aux graphiques pour afficher **les valeurs exactes et/ou de comparer** les estimations avec des variables comme **la température de l'eau ou l'effort de pêche** pour mieux observer les **tendances globales**.

Partie 4. Evaluer l'impact de la stratification sur l'estimation de l'abondance du merlu

Cette étude vise à **évaluer l'impact de la stratification géographique** sur l'estimation de l'abondance des merlus. Pour cela, les **strates** ont été regroupées selon leurs **deux premières lettres** (Cc, Cn, Cs, Gn, Gs). **L'abondance annuelle** a été recalculée pour cette **nouvelle stratification**, et les résultats ont été **comparés aux estimations précédentes**.

Le regroupement a été effectué pour les strates suivantes :

- Cc : Centre côtier
- Cn : Centre nord
- Cs : Centre sud
- Gn : Grande nord
- Gs : Grande sud

Cette **nouvelle classification** simplifie l'analyse en **réduisant le nombre de strates** et en **facilitant l'interprétation spatiale** des résultats.

Les résultats sont de nouveaux représentés sous **quatre graphiques** pour observer les **tendances et variations annuelles** :

Les résultats sont de nouveaux représentés sous quatre graphiques pour observer les tendances et variations annuelles :

- Graphique en barre (noir) : Ce graphique en barres montre **l'abondance totale des merlus par strate regroupée**. On observe que la **strate Gn domine largement** en termes d'abondance, tandis que les **autres strates présentent des niveaux nettement plus faibles et homogènes**. (*Figure 8*)
- Graphique linéaire par strate : Ce graphique linéaire montre les **estimations annuelles d'abondance** pour chaque **strate regroupée**. La **strate Gn affiche des pics marqués en 2012 et 2017**, tandis que les **autres strates restent globalement stables** avec des valeurs beaucoup plus faibles. (*Figure 9*)

- Graphique linéaire par estimation classique et regroupées : Ce graphique linéaire **compare les estimations classiques** (sans stratification) et les **estimations stratifiées**. Les résultats montrent que **l'abondance totale suit des tendances similaires** dans les deux approches, mais les **valeurs stratifiées sont généralement plus élevées**. Cela s'explique par une **meilleure prise en compte des disparités géographiques** grâce à la stratification. Les **pics en 2012 et 2017** restent visibles, confirmant leur importance. *(Figure 10)*

La stratification par zones géographiques améliore la précision des estimations en prenant en compte les différences régionales d'abondance.

La strate Gn contribue de manière disproportionnée à l'abondance totale, soulignant l'importance de cette région pour l'espèce. Les autres strates montrent des **niveaux d'abondance plus faibles et stables**, mais leur contribution reste **essentielle pour une vision globale**. La comparaison avec les estimations classiques révèle des **écarts positifs avec la stratification**, montrant que la méthode permet une **meilleure représentativité des données**.

La stratification par zones géographiques améliore la **fiabilité** des estimations d'abondance des merlus en considérant les **variations spatiales**. Les résultats mettent en évidence des **différences marquées entre les strates**. Cette approche permet une **interprétation plus précise des dynamiques spatiales et temporelles** de l'abondance, tout en confirmant les **tendances observées dans les estimations globales**.

Pour **approfondir l'analyse**, il serait pertinent d'**explorer les facteurs environnementaux** propres à chaque zone (**température, effort de pêche**) afin de mieux comprendre les **variations d'abondance observées**.

Partie 5. Analyse de l'espèce : Lophius budegassa

Nous avons réalisé une analyse en **quatre étapes** pour l'espèce **Lophius budegassa**, puis comparé les résultats avec **Merluccius merluccius** :

1. **Analyse exploratoire des données** pour comprendre la structure et la distribution des captures.
2. **Estimation stratifiée du total et de sa variance**, en tenant compte des strates écologiques.
3. **Estimation de l'abondance pour l'année 2012** à l'aide des méthodes **HTstrata** et **varest**.
4. **Impact de la stratification** sur l'estimation de l'abondance totale.

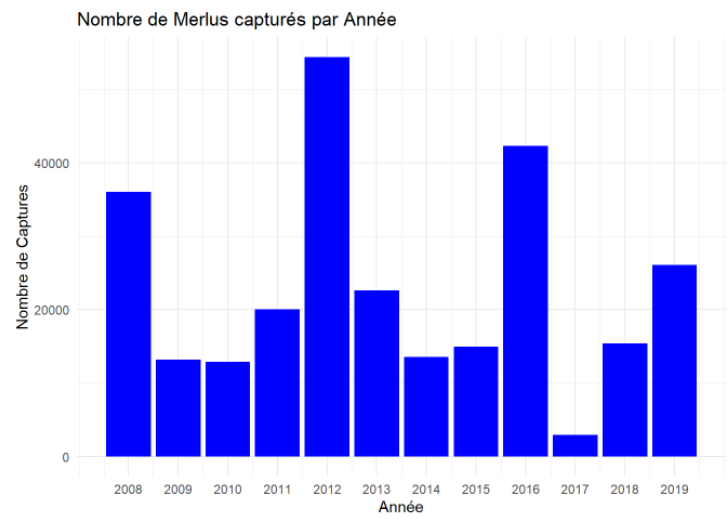
Le graphique présente une **comparaison des abondances totales** des deux espèces par année. *(Figure 11)*

- **Bleu foncé : Merluccius merluccius** montre des pics significatifs, notamment en **2013** et **2018**, avec une abondance nettement supérieure.
- **Bleu clair : Lophius budegassa** présente des abondances plus faibles et stables sur la même période.

Le graphique met en évidence une **abondance globalement plus élevée** pour **Merluccius merluccius** par rapport à **Lophius budegassa**, soulignant des dynamiques d'abondance différentes entre ces deux espèces.

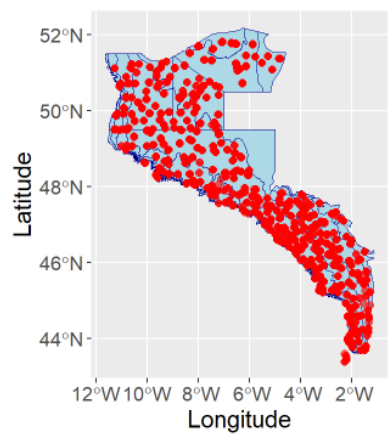
Annexes

1. Figure 1

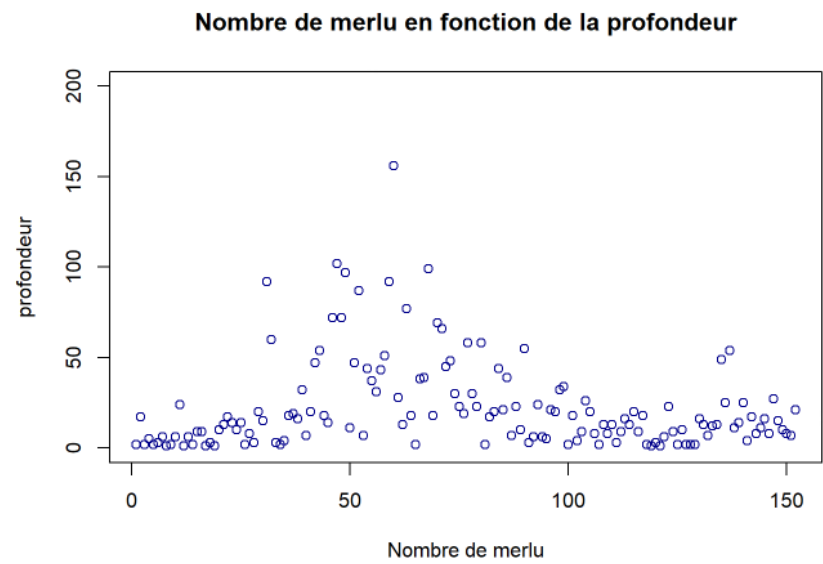


2. Figure 2

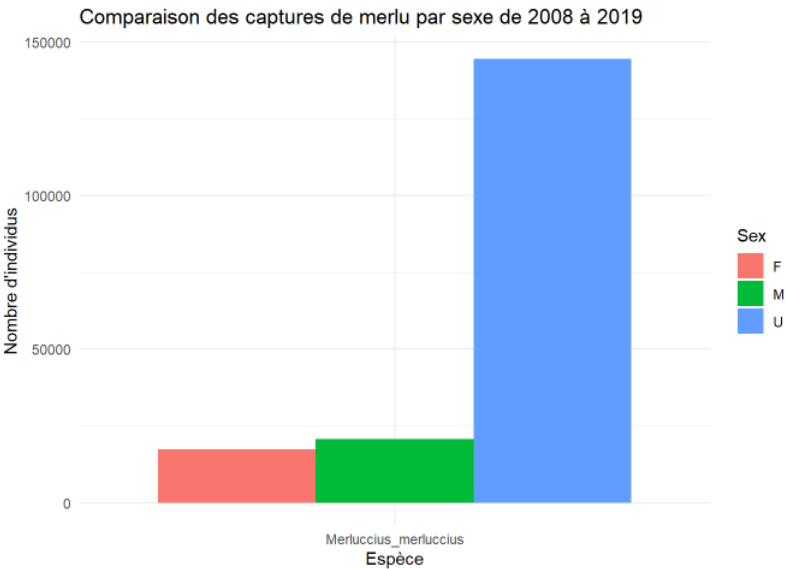
Carte des Strates de la Campagne avec Merluccius merluccius



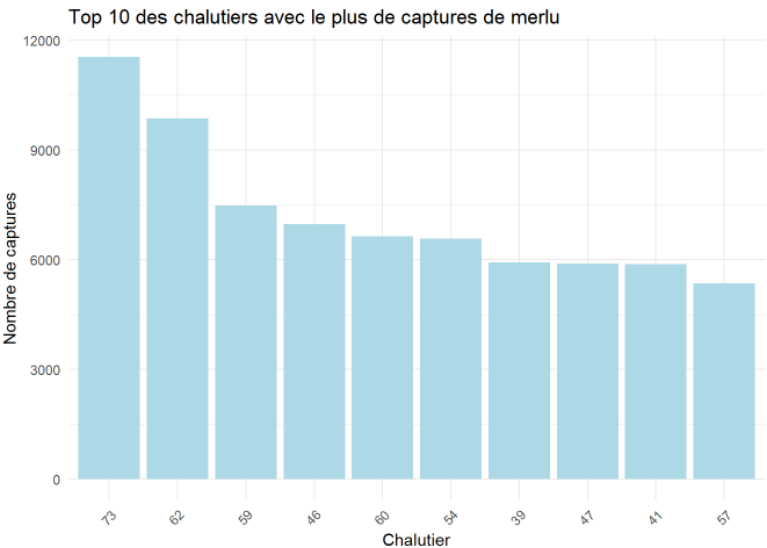
3. Figure 3



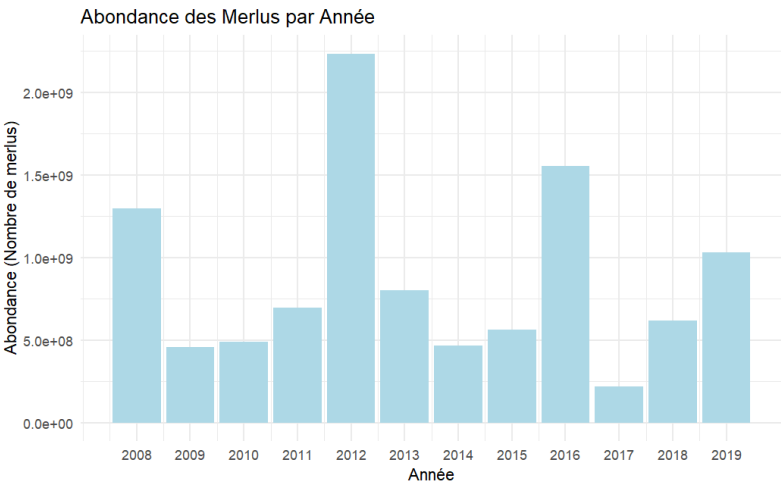
4. Figure 4



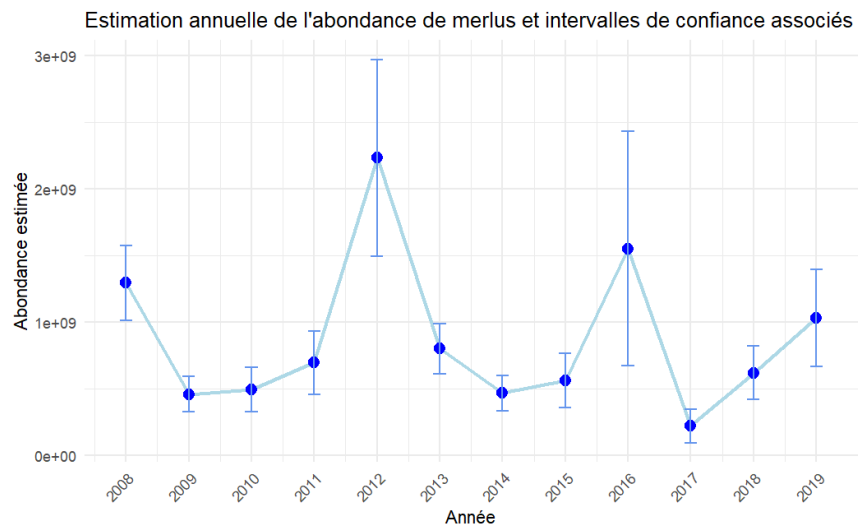
5. Figure 5



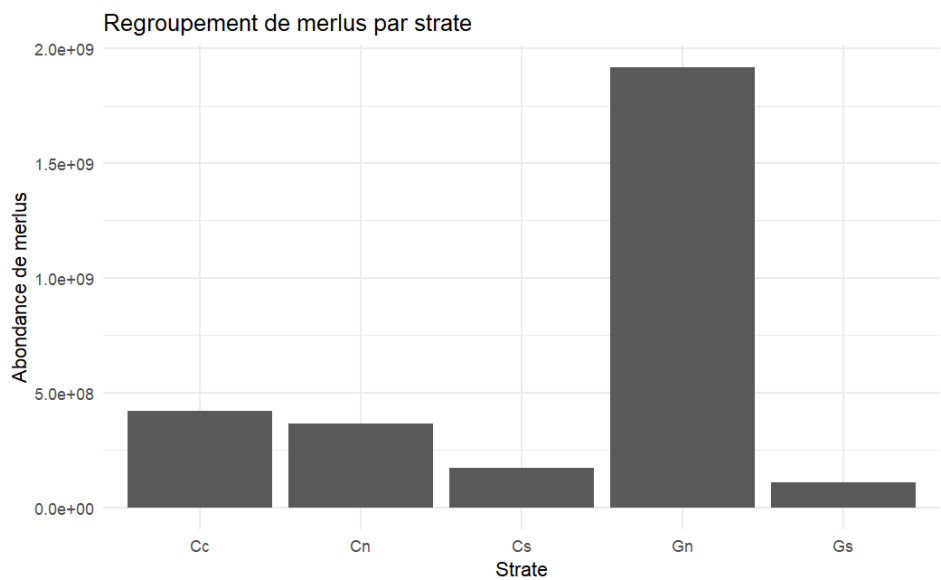
6. Figure 6



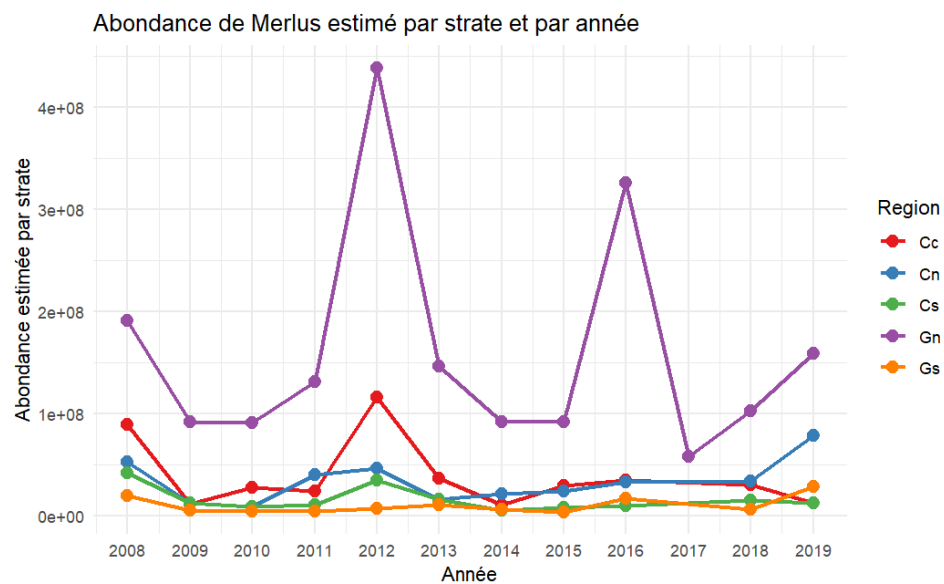
7. Figure 7



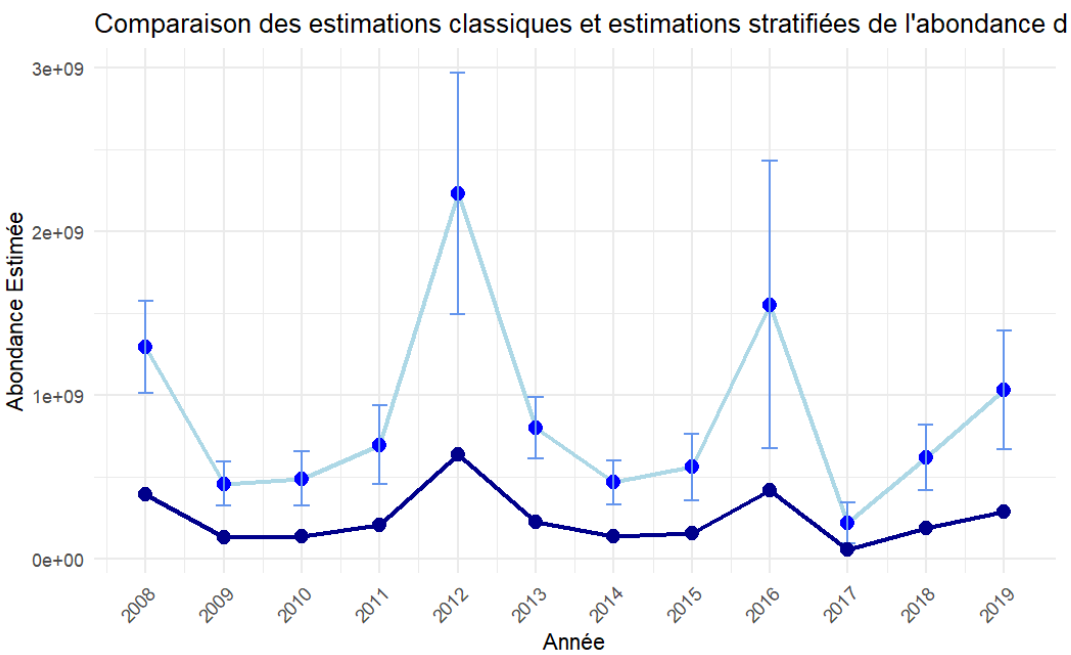
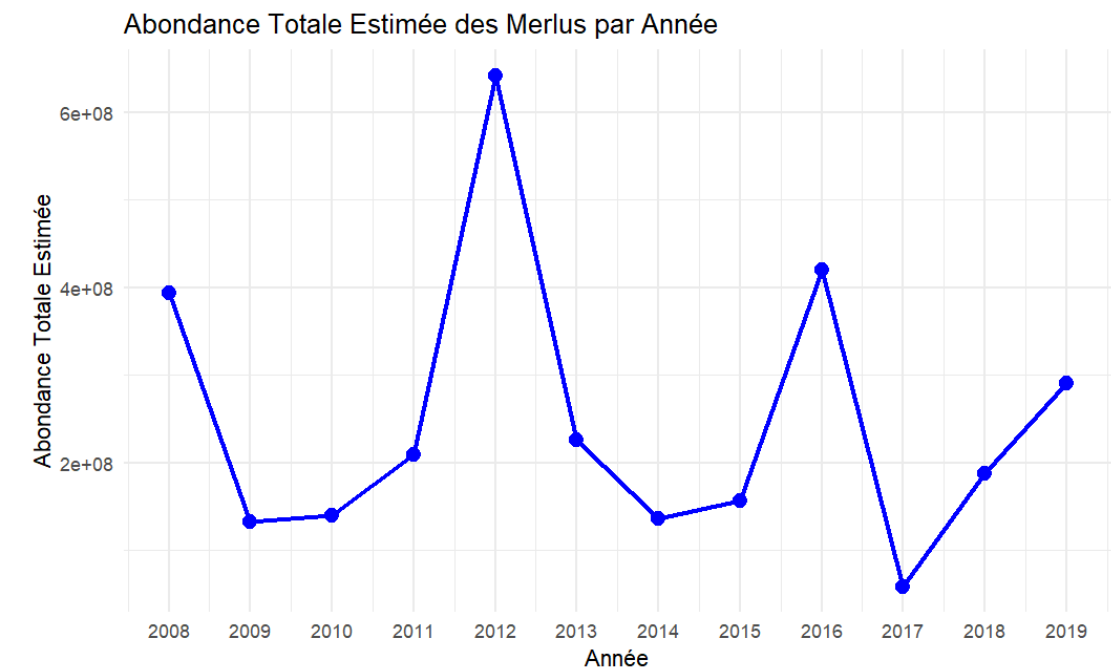
8. Figure 8



9. Figure 9



10. Figure 10



11. Figure 11

