

# EFFETS DU SOUS ÉCHANTILLONNAGE TEMPOREL : LE CAS DES DONNÉES ACOUSTIQUES DES GRANDES BALEINES DANS L'OCEAN INDIEN

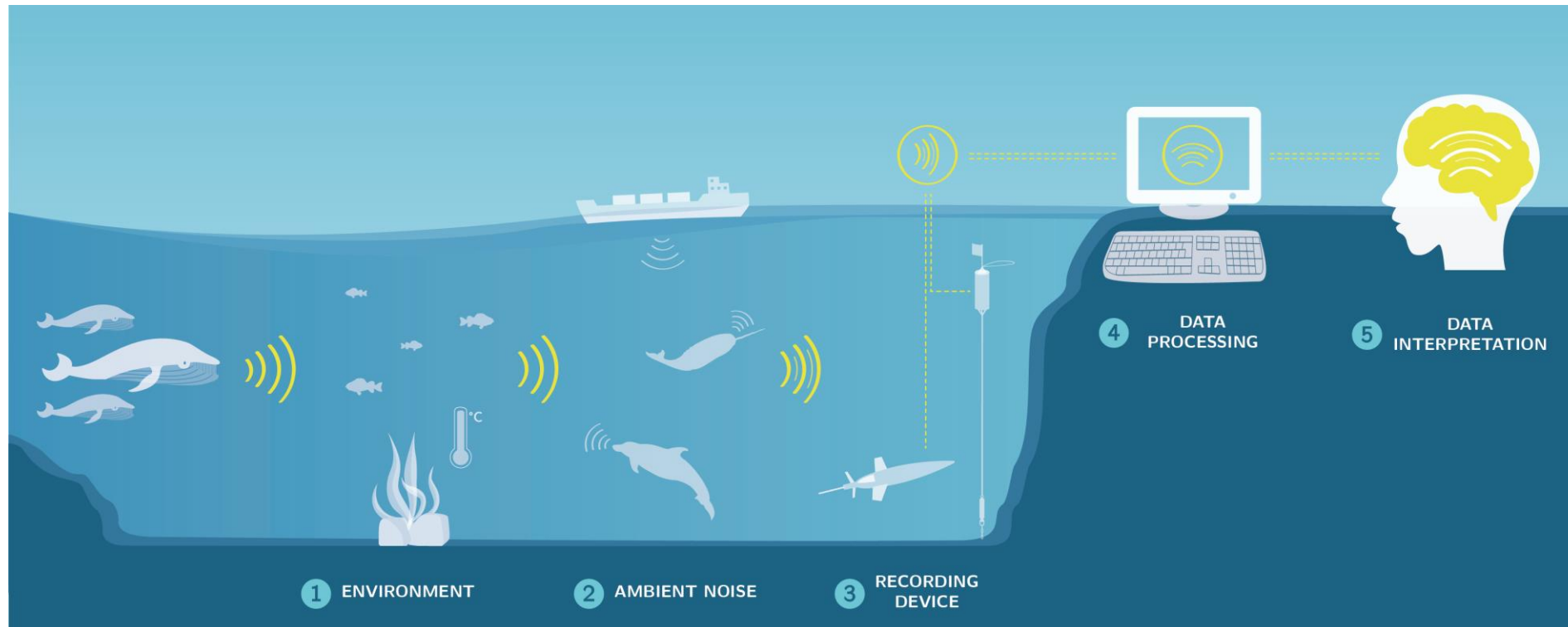
MATHILDE MICHEL, MAËLLE TORTEROTOT, FLORE SAMARAN

LAB-STICC

Workshop SERENADE: 27-30 Juin 2022



# CONTEXTE: SURVEILLANCE PAR ACOUSTIQUE PASSIVE



- Collecte des données acoustiques
- Traitement des données
- Interprétation

# CONTEXTE: COLLECTES DES DONNÉES

- Nécessite un effort d'enregistrement limité par :

Stockage des données

Capacité des batteries

Coûts logistiques et  
d'entretiens

- Solution : sous échantillonnage temporel avec un **duty cycle**
  - Collecter les données avec un pourcentage du temps, une durée définie et un cycle répétitif

# PROBLÉMATIQUE

Est-ce que des données sous échantillonnées changent l'interprétation du traitement des données ?  
Comment optimiser le choix d'un duty cycle ?

1

Patrons saisonniers

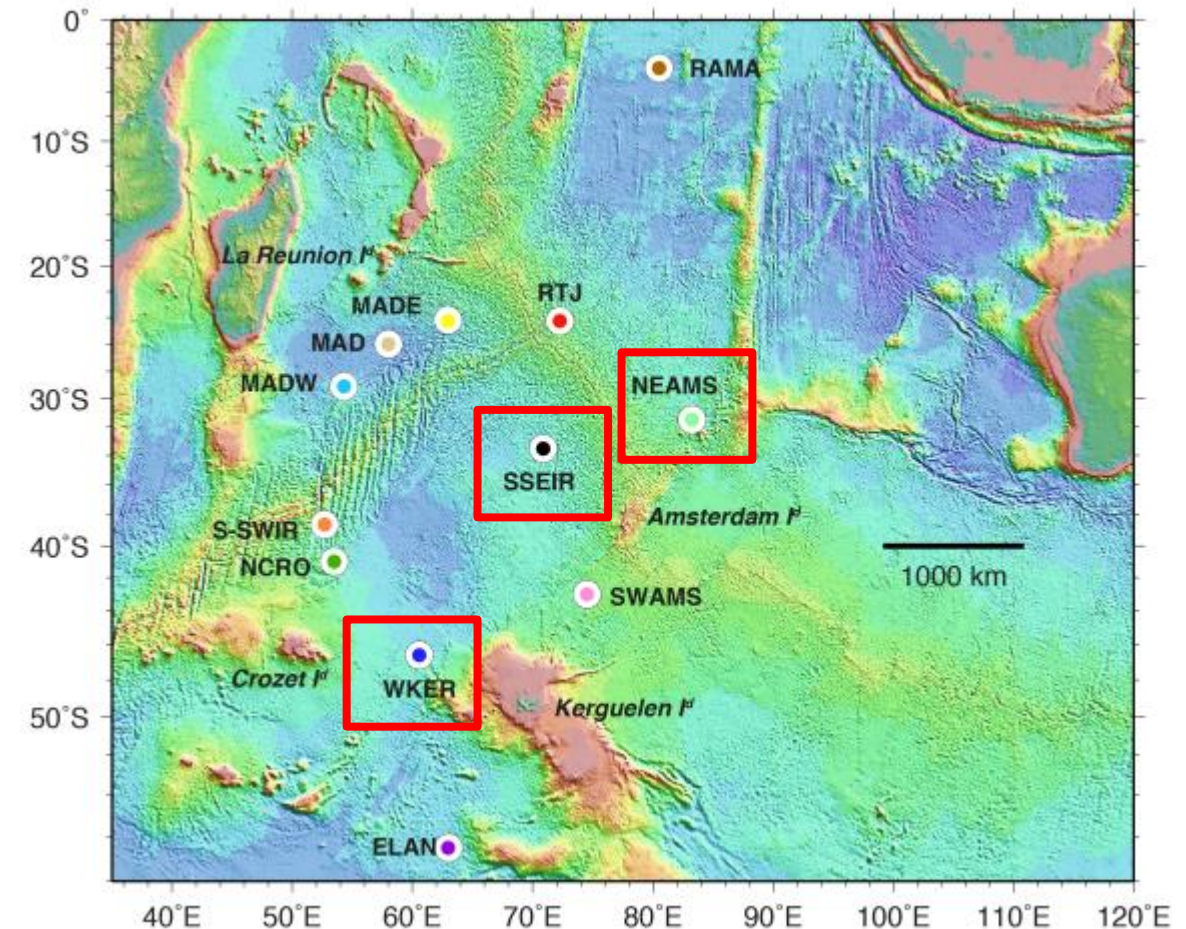
2

Richesse spécifique

Taux d'appel  
quotidien

# MATÉRIEL : BASES DE DONNÉES ÉTUDIÉES

- Réseau d'hydrophones OHASISBIO (J.Y. Royer, 2009)
  - Sud Ouest de l'Océan Indien
  - Enregistrement en continu entre 2010 et 2019
  - 3 sites étudiés sur les 9

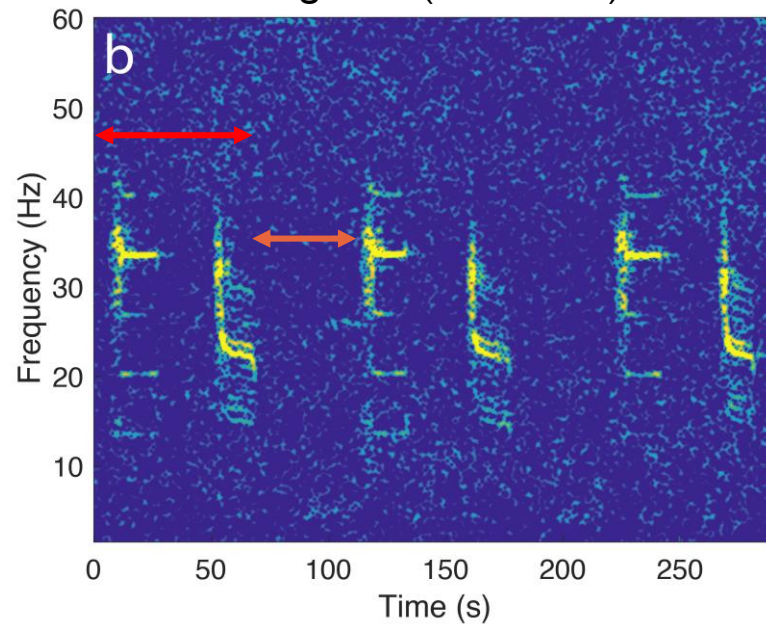




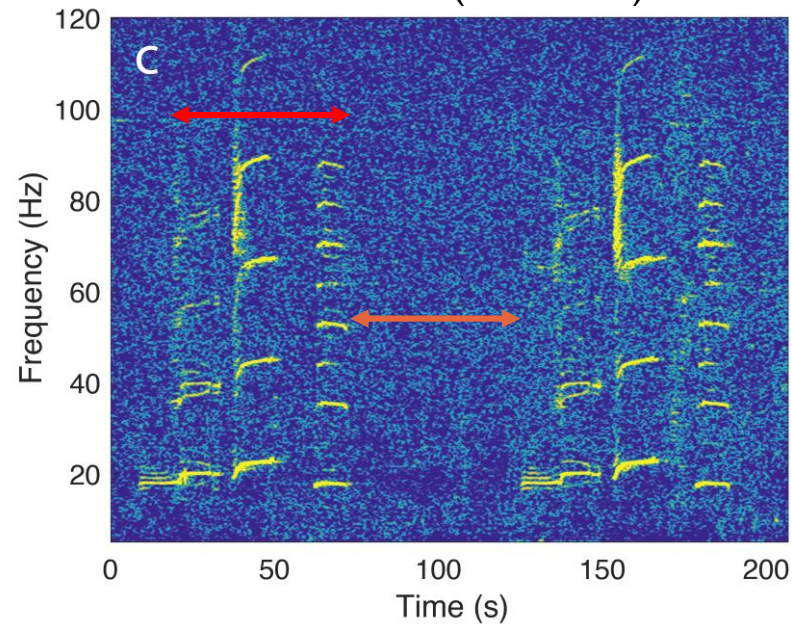
# MATÉRIEL : BASES DE DONNÉES ÉTUDIÉES

- Chants stéréotypés de baleines bleues : 3 populations acoustiques

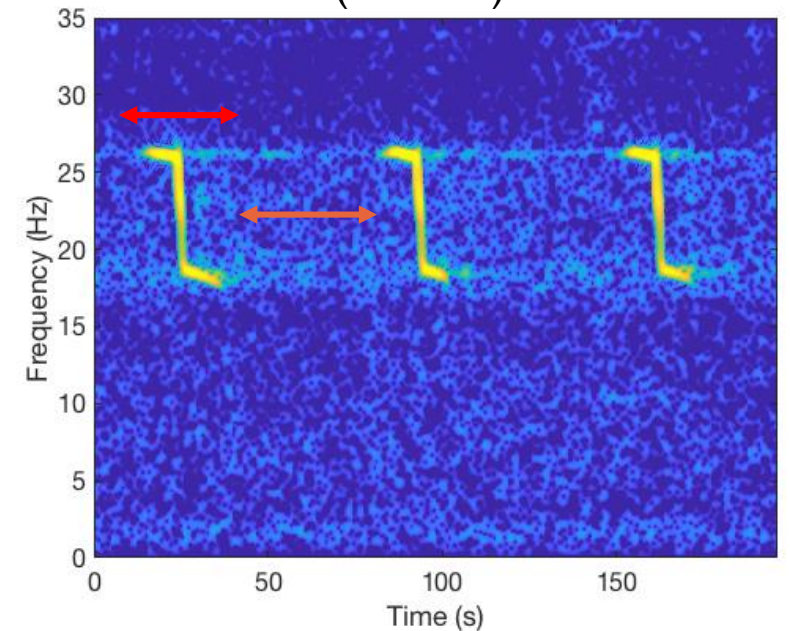
Baleine bleue pygmée de Madagascar (Mad PBV)



Baleine bleue pygmée Australienne (Aus PBV)



Baleine bleue Antarctique (Ant BW)

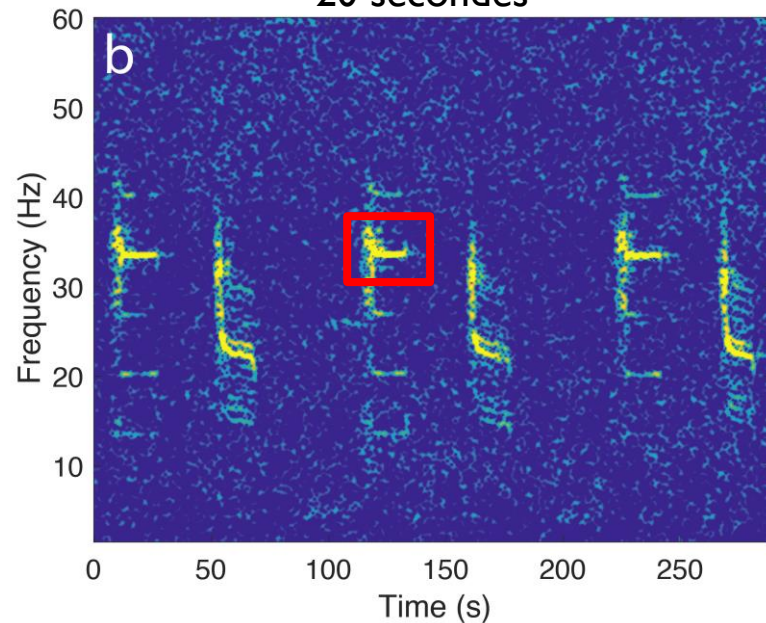




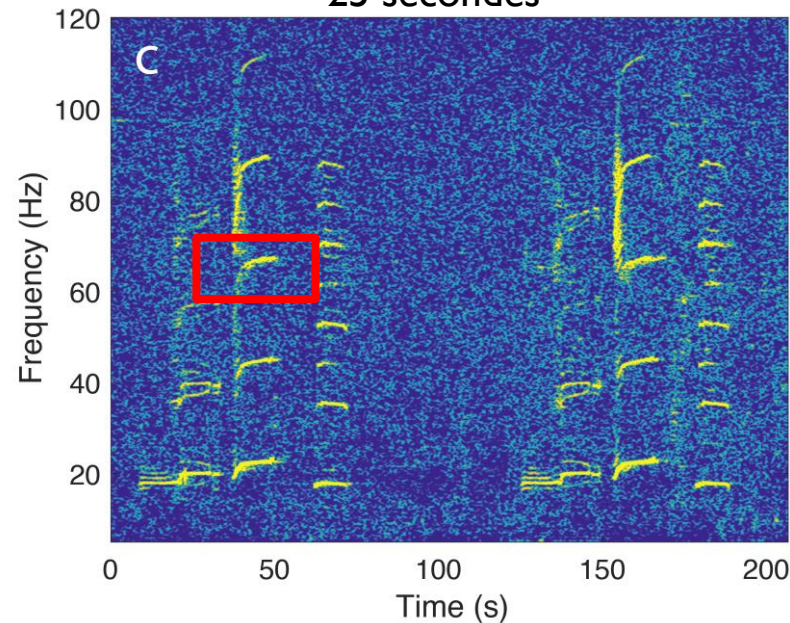
# MATÉRIEL : BASES DE DONNÉES ÉTUDIÉES

- Algorithme de détections automatiques utilise un dictionnaire et une fenêtre de détection égale à la durée de la vocalisation (Socheleau et Samaran, 2018)

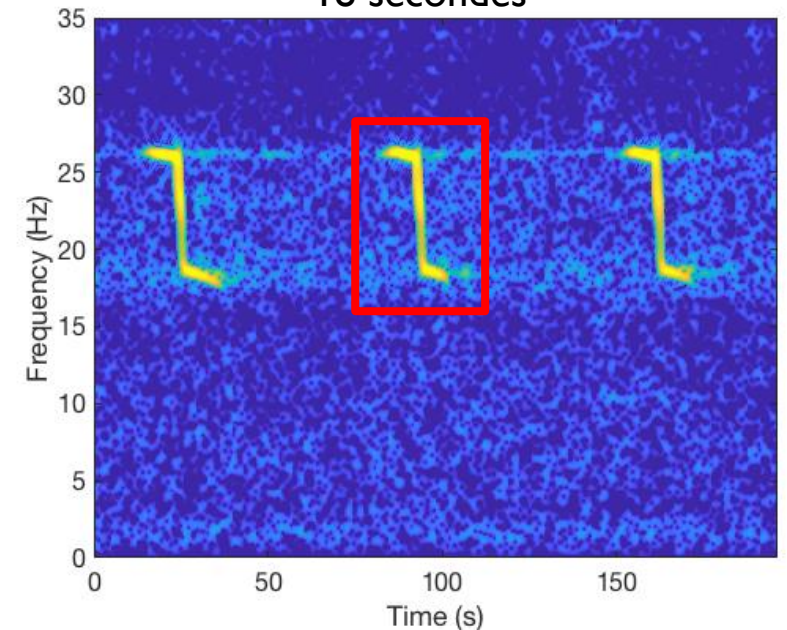
Baleine bleue pygmée de  
Madagascar (Mad PBW)  
20 secondes



Baleine bleue pygmée  
Australienne (Aus PBW)  
25 secondes



Baleine bleue Antarctique  
(Ant BW)  
18 secondes



# MÉTHODE : SCHÉMA DE SOUS ÉCHANTILLONNAGE

- Défini par une période d'écoute  $L$ , une période de cycle  $T_c$  et un duty cycle  $D$  (pourcentage d'écoute)

Enregistrement continu



Période de cycle  $T_c$



Période d'écoute  $L$

Enregistrement sous échantillonné : duty cycle de 50%

1 minute / 2 minutes





# MÉTHODE : SCHÉMA DE SOUS ÉCHANTILLONNAGE

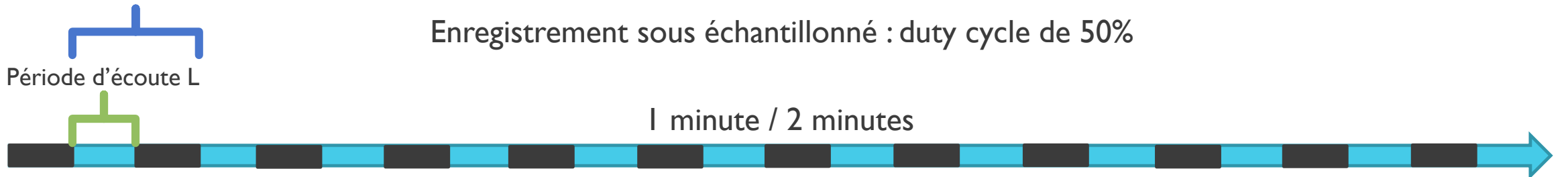
- Défini par une période d'écoute  $L$ , une période de cycle  $T_c$  et un duty cycle  $D$  (pourcentage d'écoute)

Enregistrement continu



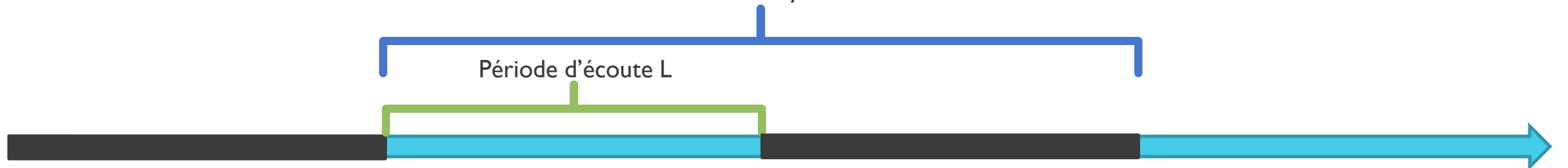
Période de cycle  $T_c$

Enregistrement sous échantillonné : duty cycle de 50%



Période de cycle  $T_c$

Période d'écoute  $L$



6 heures / 12 heures

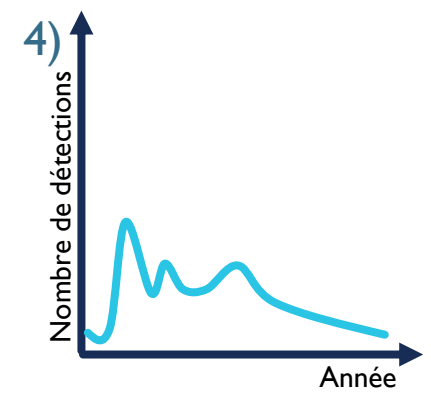
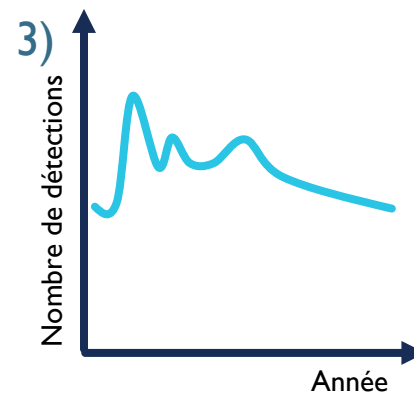
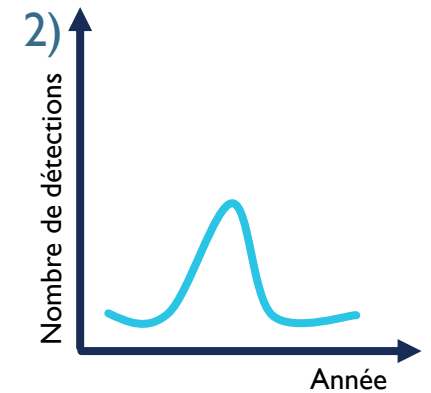
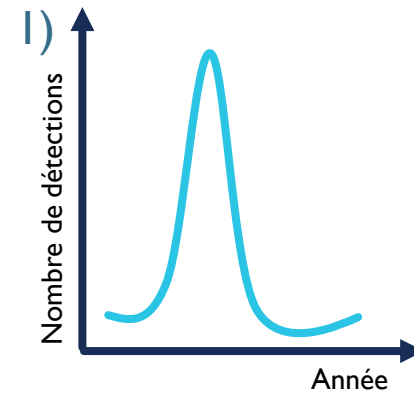
# MÉTHODE : SCHÉMA DE SOUS ÉCHANTILLONNAGE

- 3 pourcentages de duty cycle : 50%, 33% et 25%
- Périodes d'écoute de 1 minute à 8h
- 100 % de la durée de la vocalisation dans la période d'écoute

Listening period L	Cycle period Tc D=50%	Cycle period Tc D=33%	Cycle period Tc D=25%	Cycles/days N
1min	2min	3min	4min	1440
5min	10min	15min	20min	288
15min	30min	45min	1h	96
30min	1h	1h30	2h	48
1h	2h	3h	4h	24
3h	6h	9h	12h	8
6h	12h	18h	24h	4
8h	16h	24h		3

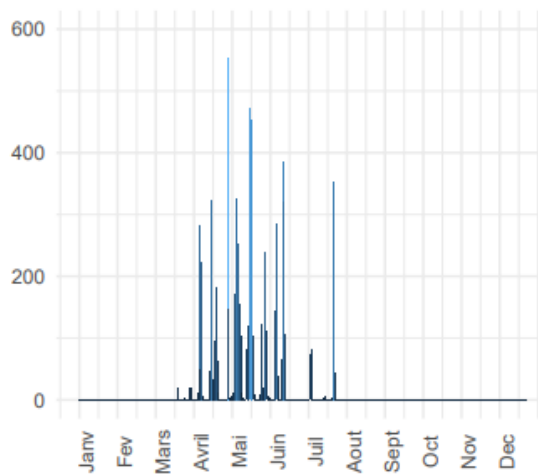
# MÉTHODE : PATRONS SAISONNIERS

- Défini par le nombre absolu de détections par jours sur une année
- Etude de 4 patrons saisonniers différents :
  - 1) Patron saisonnier fort avec un nombre élevé de détections
  - 2) Patron saisonnier fort avec un nombre faible de détections
  - 3) Patron saisonnier faible avec un nombre élevé de détections
  - 4) Patron saisonnier faible avec un nombre faible de détections

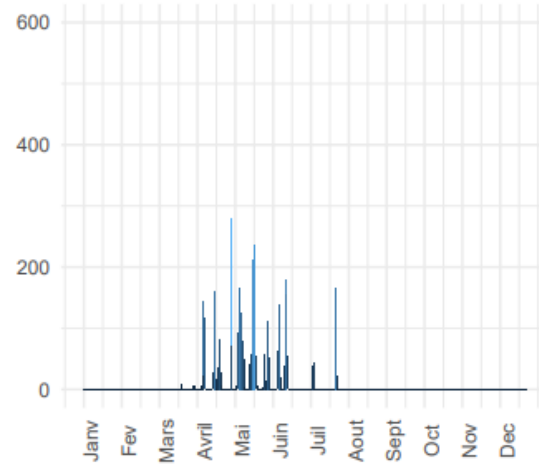


# MÉTHODE : PATRONS SAISONNIERS

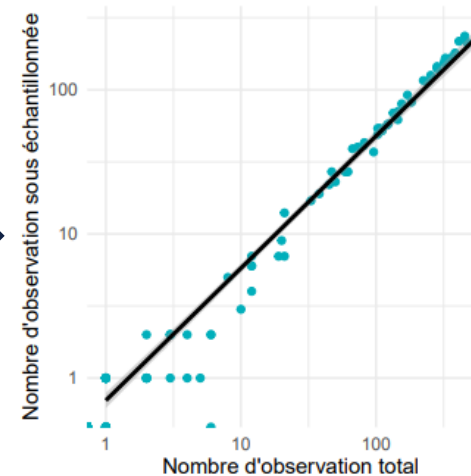
- Evaluation de la qualité de la représentation de la saisonnalité



Nombre de détections par jours  
des données continues



Nombre de détections par jours  
des données sous  
échantillonnées 1 minute / 2



Ajustement des données avec  
un modèle de régression  
linéaire



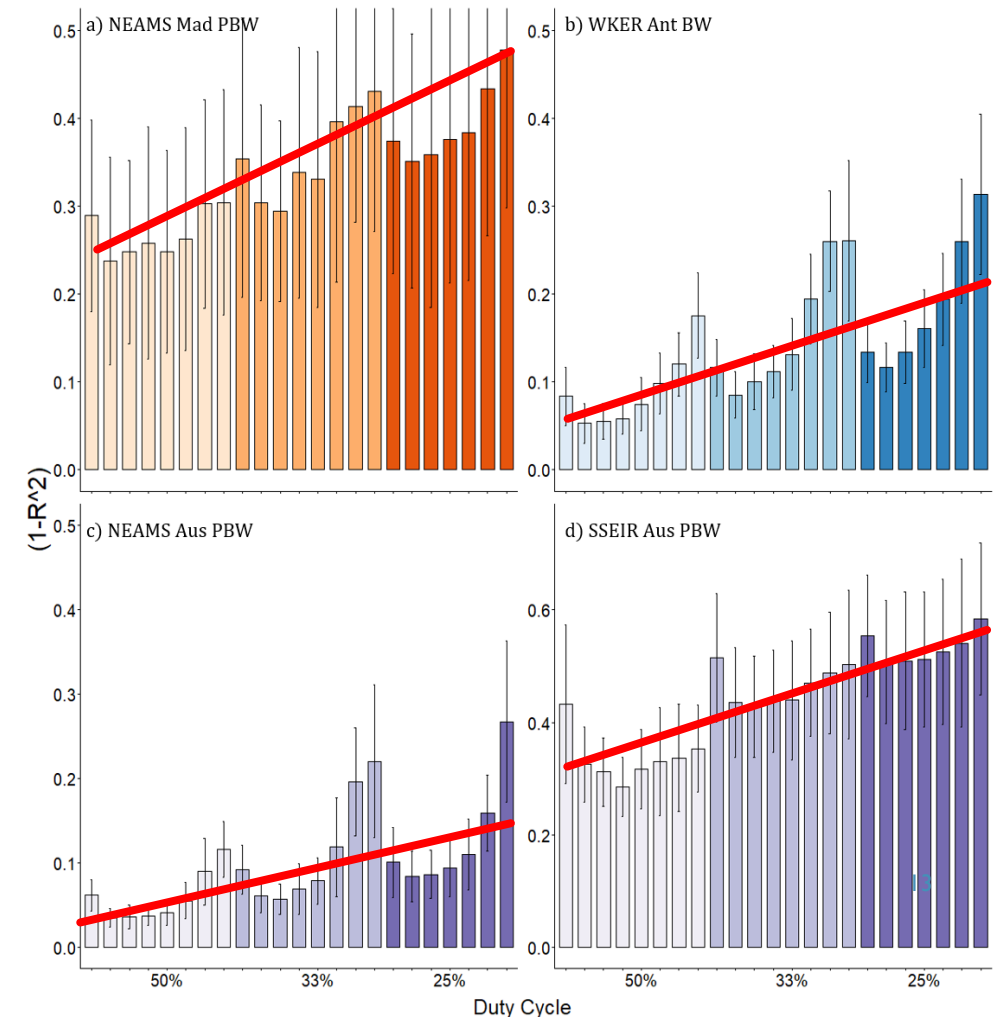
$$(1 - R^2) = 1 - \left( 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right),$$

Calcul du complément du  
coefficient de détermination  
(1-R<sup>2</sup>) compris entre 0 et 1



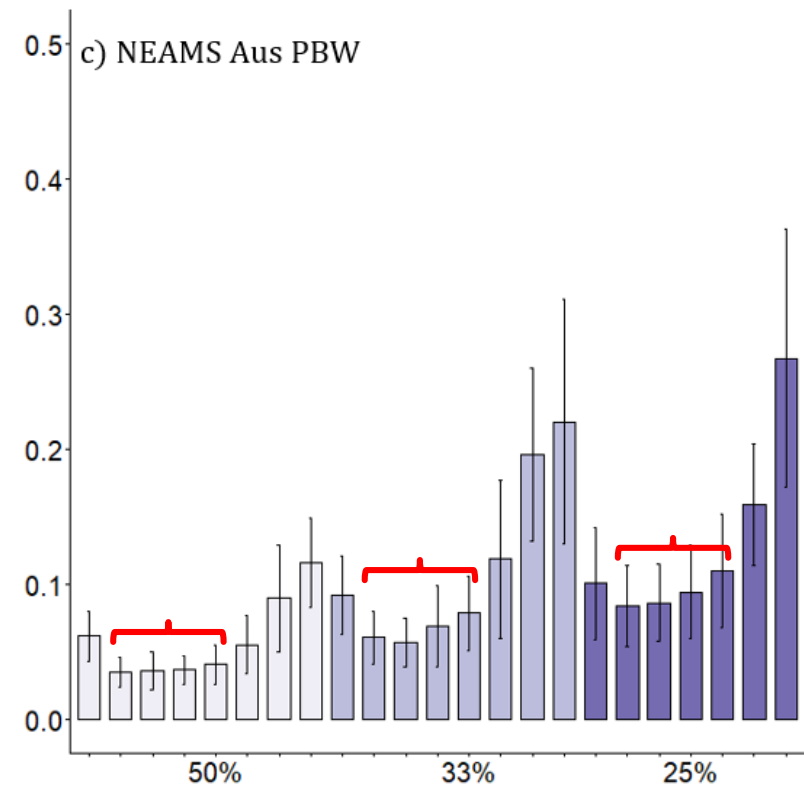
# RÉSULTATS: PATRONS SAISONNIERS

- $(1-R^2)$  plus faible pour un duty cycle élevé 50%



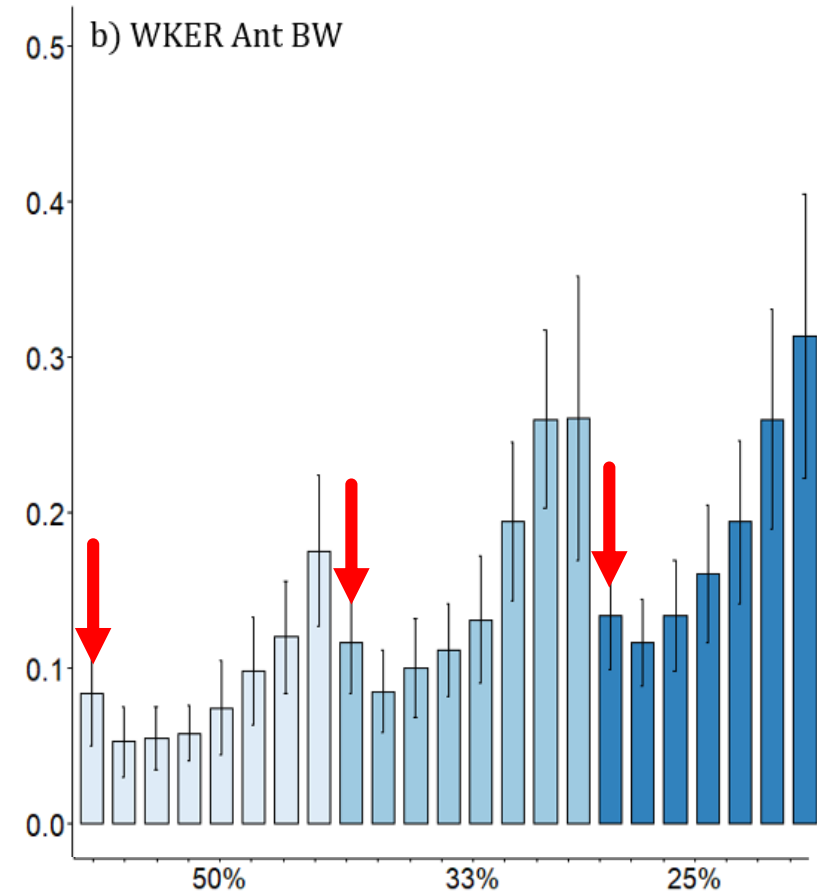
# RÉSULTATS: PATRONS SAISONNIERS

- $(1-R^2)$  plus faible pour un duty cycle élevé 50%
- Dans un duty cycle: courtes périodes d'écoute de 5 à 30 minutes donne une meilleure précision du modèle que les longues > 1h



# RÉSULTATS: PATRONS SAISONNIERS

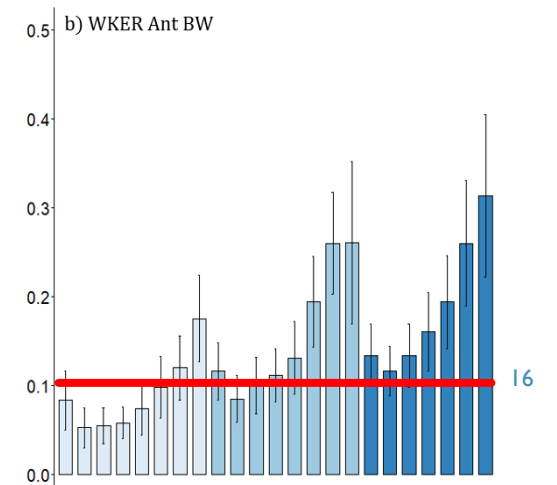
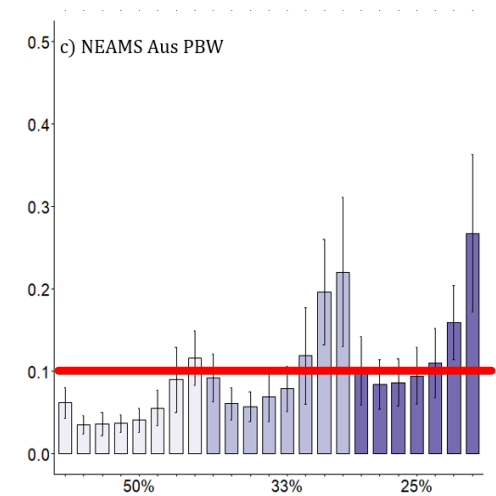
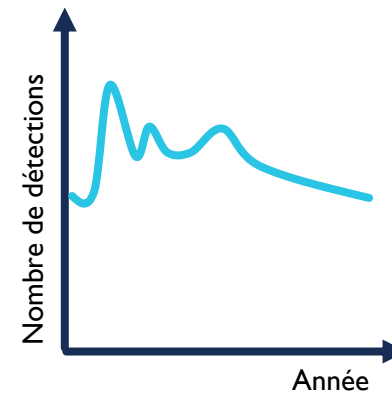
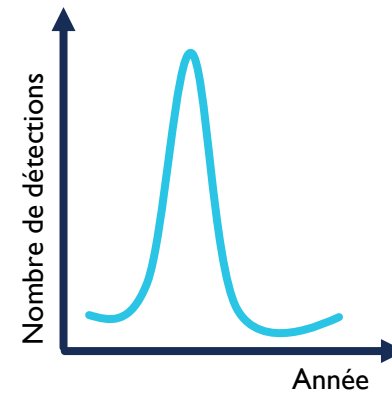
- $(1-R^2)$  plus faible pour un duty cycle élevé 50%
- Dans un duty cycle: courtes périodes d'écoute de 5 à 30 minutes donne une meilleure précision du modèle que les longues  $> 1h$
- Période trop courte: 1 minute affecte la précision



# RÉSULTATS: PATRONS SAISONNIERS

- Patron saisonnier fort + beaucoup de détections
- Patron saisonnier faible + beaucoup de détections

Bon ajustement des données entre le modèle sous échantillonné et les données continues  
Environ 10% de perte d'informations

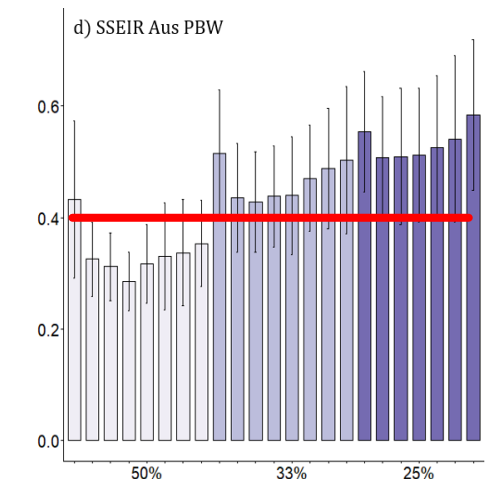
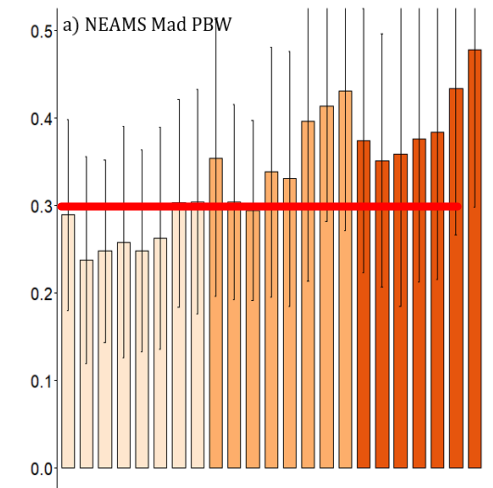
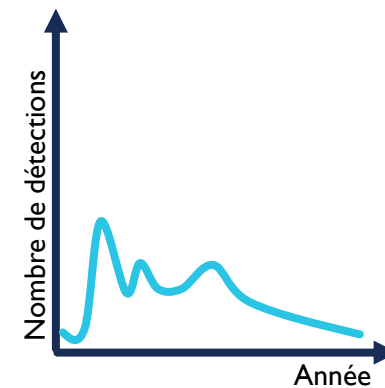
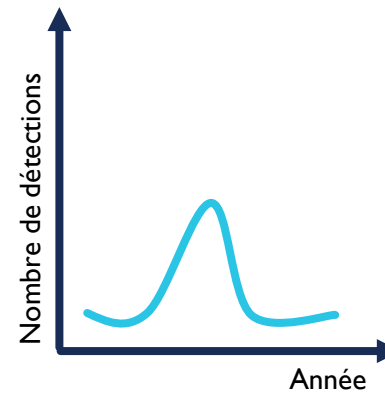




# RÉSULTATS: PATRONS SAISONNIERS

- Patron saisonnier fort + peu de détections
- Patron saisonnier faible + peu de détections

Mauvais ajustement des données entre le modèle  
sous échantillonné et les données continues  
Plus de 25 % de perte d'informations



## MÉTHODE : RICHESSE SPÉCIFIQUE

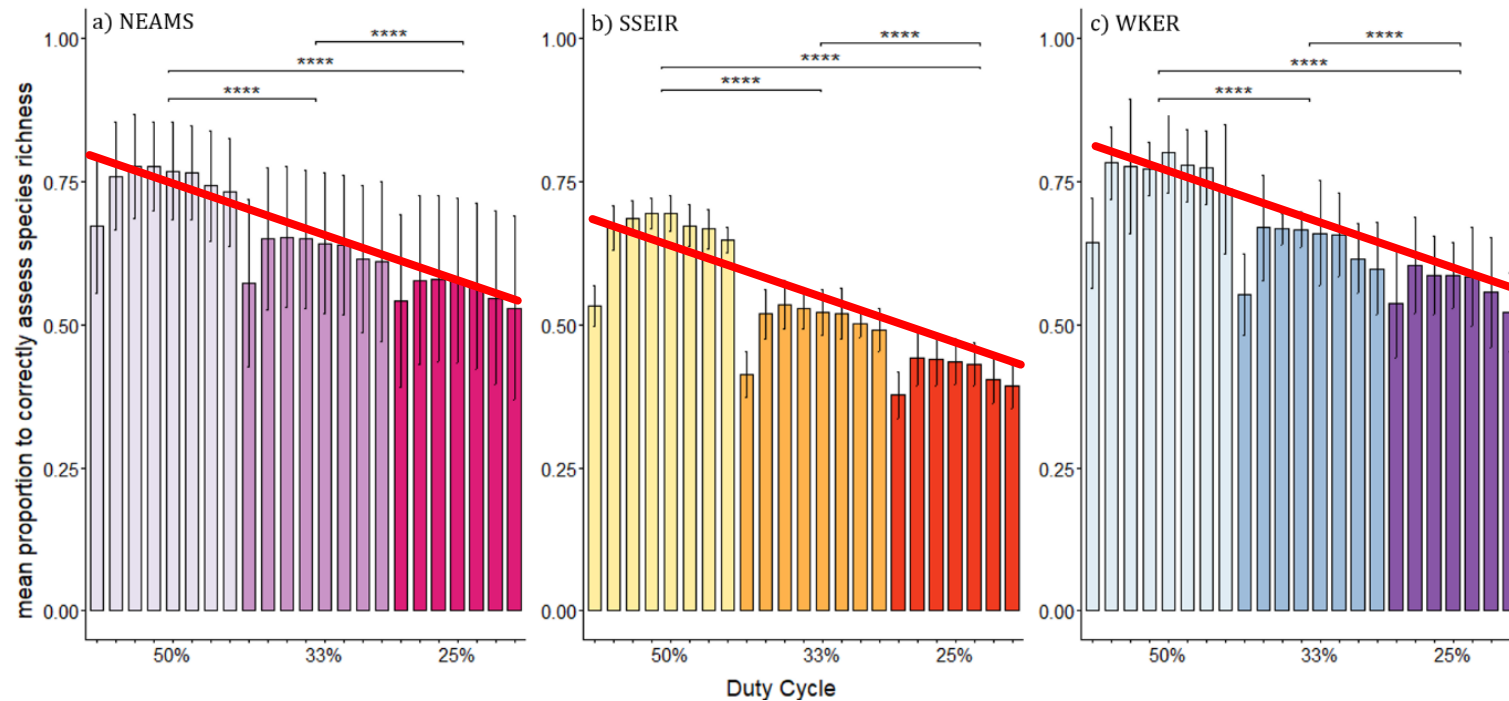
- Défini par le nombre d'espèces présentes acoustiquement dans une zone à un moment donné
- Présence acoustique de l'espèce: minimum une détection par jours
- Si présence acoustique des données continues et des données sous échantillonnées est la même alors  $d = 1$ , sinon  $d = 0$
- Calcul de la proportion de jours  $p_d$  où la richesse spécifique est correctement évaluée pour chaque site

$$\overline{p_d} = \frac{1}{np} \sum_{j=1}^p \sum_{i=1}^n d_{ij},$$

n nombre de jours avec détection  
p nombre de position de période d'écoute

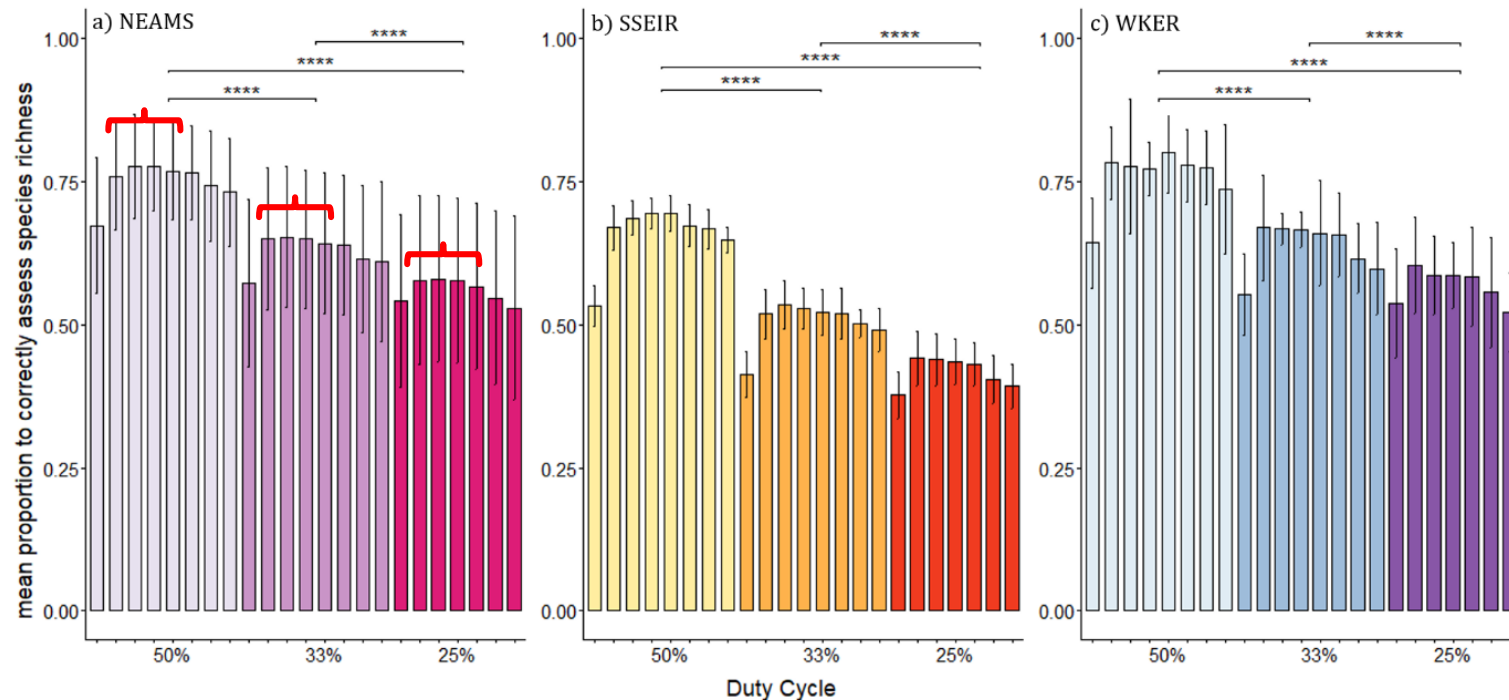
# RÉSULTATS : RICHESSE SPÉCIFIQUE

- Plus grande proportion de jours avec une richesse spécifique correctement évaluée pour duty cycle de 50%



# RÉSULTATS : RICHESSE SPÉCIFIQUE

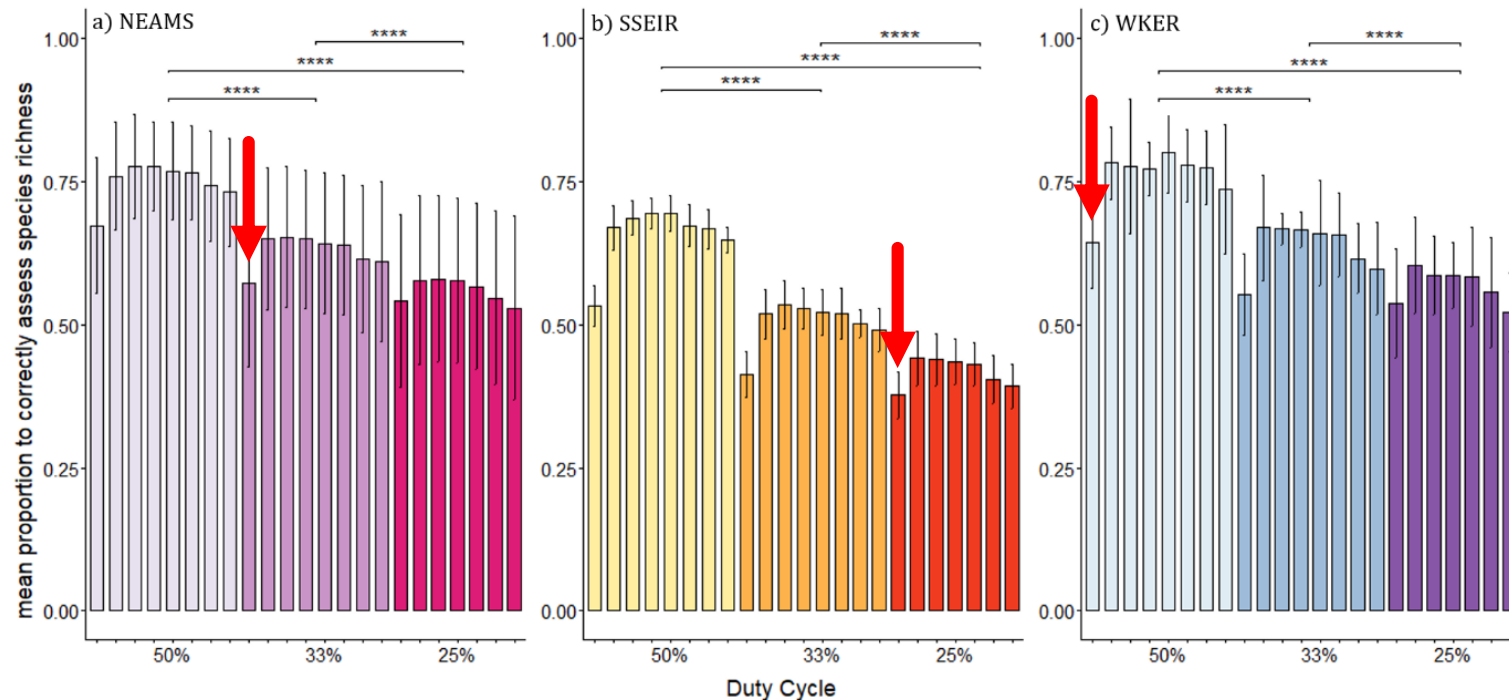
- Plus grande proportion de jours avec une richesse spécifique correctement évaluée pour duty cycle de 50%
- Pour chaque duty cycle : courtes périodes d'écoute (5min à 30 min) meilleures que les longues <1h





# RÉSULTATS : RICHESSE SPÉCIFIQUE

- Plus grande proportion de jours avec une richesse spécifique correctement évaluée pour duty cycle de 50%
- Pour chaque duty cycle : petite période d'écoute meilleure que les longues
- 1 minute: moins bonne proportion de jours correctement évalués



## CONCLUSION GÉNÉRALE

**Duty Cycle % :**

50% > 33% > 25%

**Période d'écoute:**

Courtes périodes  
d'écoute

Entre 5 et 30 minutes

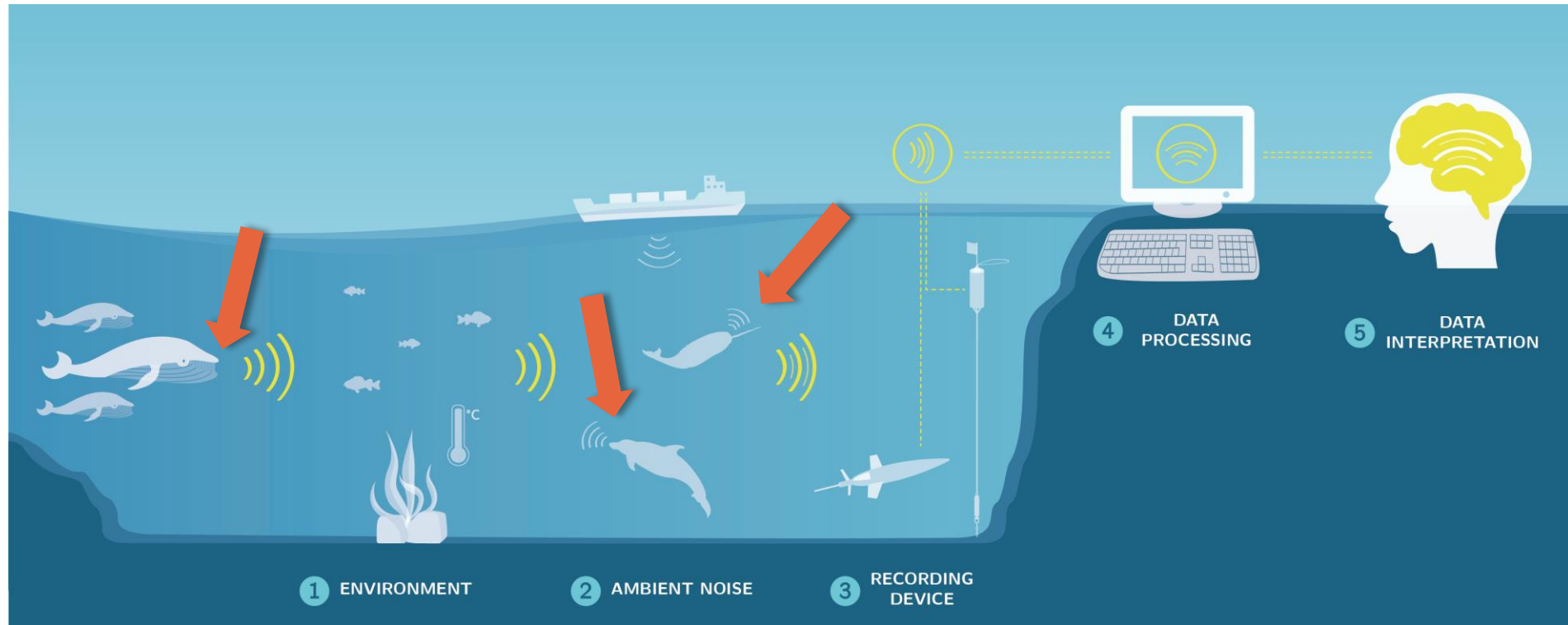
**Nombre de  
détections:**

Beaucoup de perte  
d'informations si une  
espèce est peu  
présente



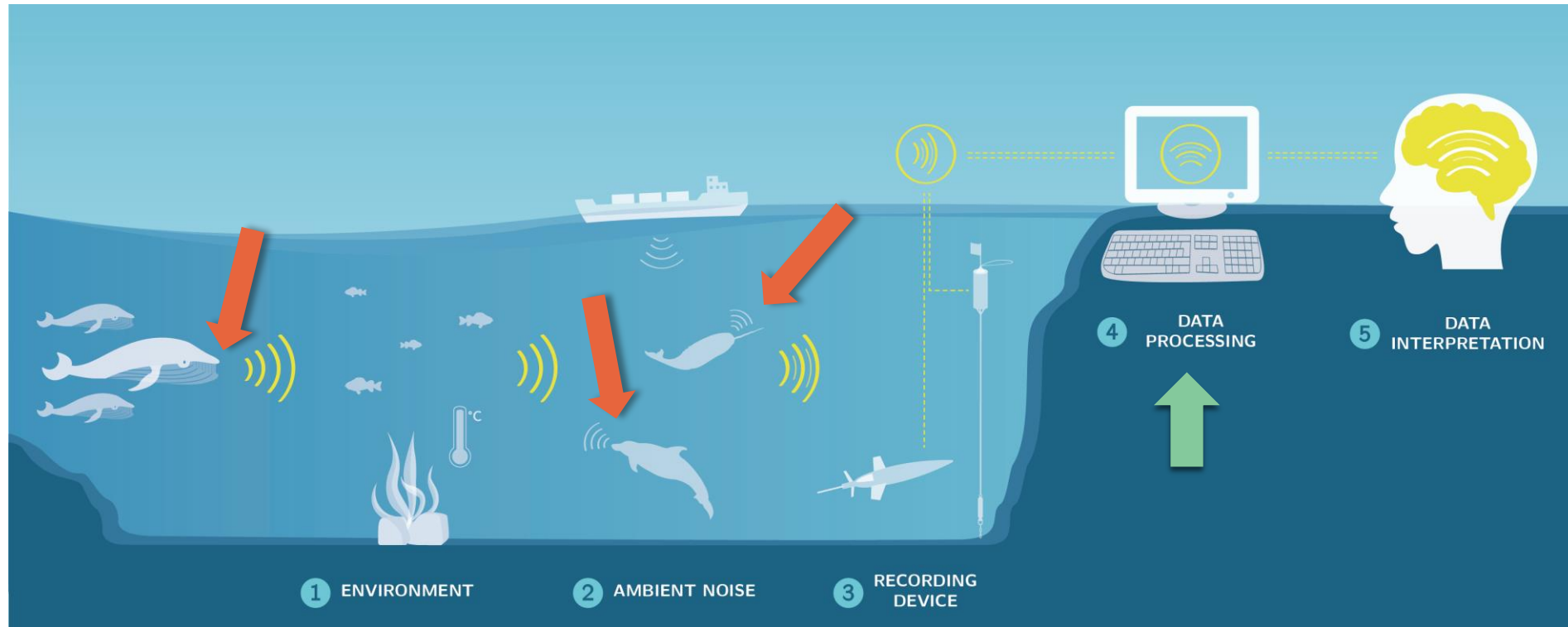
1 minute = trop de détections  
coupées

# RECOMMANDATIONS



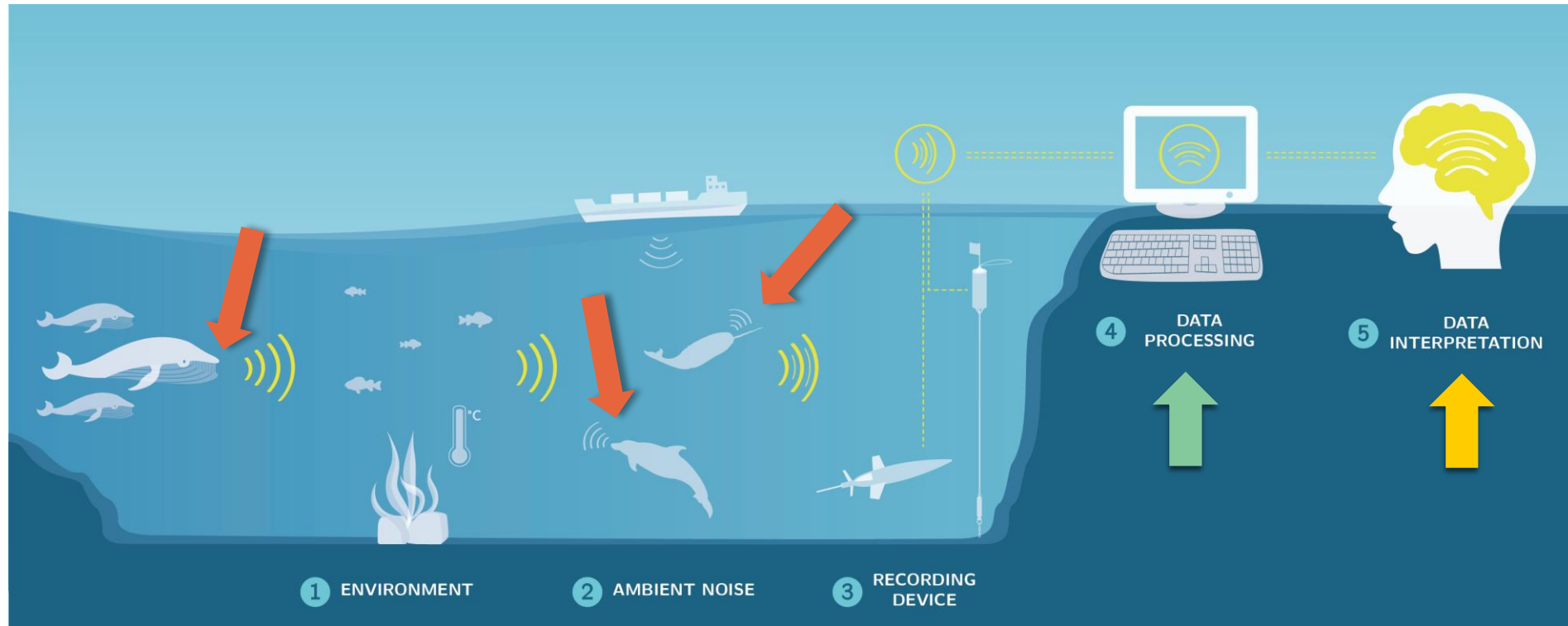
- Cibler en amont les espèces d'intérêts, leur utilisation de l'environnement et leur comportement vocal

# RECOMMANDATIONS



- Cibler en amont les espèces d'intérêts, leur utilisation de l'environnement et leur comportement vocal
- Définir le traitement des données: algorithme de détection

# RECOMMANDATIONS



- Cibler en amont les espèces d'intérêts, leur utilisation de l'environnement et leur comportement vocal
- Définir le traitement des données: algorithme de détection
- Définir les métriques de description

## ET APRÈS...

Autres espèces avec d'autres types de cris:  
clics / sifflements

Autres espèces avec des patrons de présence  
différents

## ET APRÈS...

Autres espèces avec d'autres types de cris:  
clics / sifflements

Autres espèces avec des patrons de présence  
différents

Autres types de traitement de données /  
Algorithme de détections

Sous échantillonner en pré détecteur

## ET APRÈS...

Autres espèces avec d'autres types de cris:  
clics / sifflements

Autres types de traitement de données /  
Algorithme de détections

Sous échantillonner en pré détecteur

Autres espèces avec des patrons de présence  
différents

Autres métriques de description de données en  
surveillance acoustique passive



## ET APRÈS...

Autres espèces avec d'autres types de cris:  
clics / sifflements

Autres types de traitement de données /  
Algorithme de détections

Sous échantillonner en pré détecteur

Autres espèces avec des patrons de présence  
différents

Autres métriques de description de données en  
surveillance acoustique passive

Algorithme de sous échantillonnage aléatoire

Sous échantillonnage en post traitement quand trop  
de données collectées

Merci pour votre attention !



# CONCLUSION : CHOIX D'UN SCHÉMA DE SOUS ÉCHANTILLONNAGE

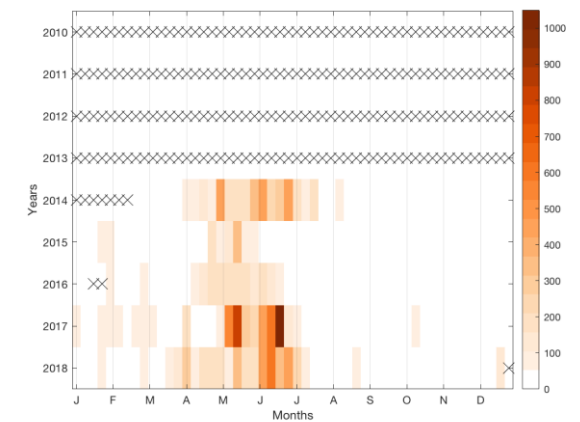
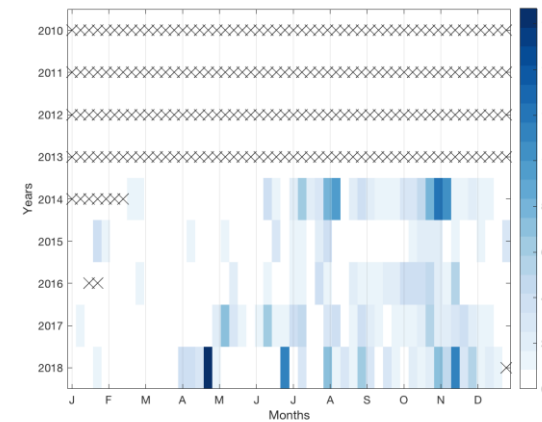
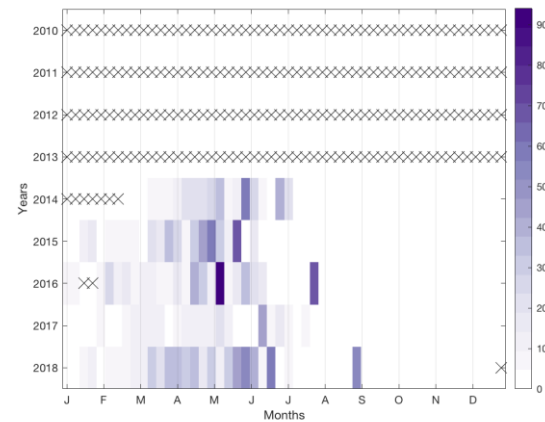
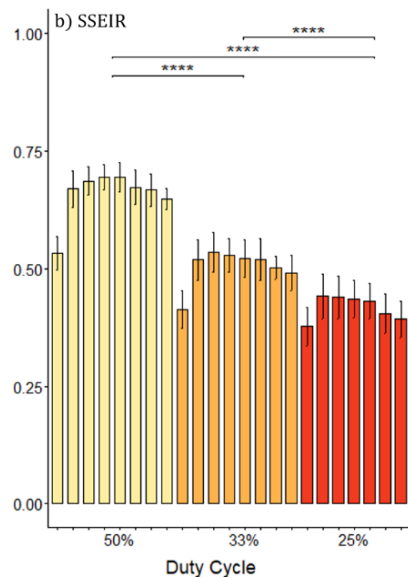
Connaissance de la présence acoustique d'une espèce: présence d'espèce rare nécessite enregistrements continues

Connaissance sur les caractéristiques de la vocalisation: impact de la durée et/ou de la régularité

Connaissance sur le choix du traitement des données / détecteurs

Connaissances des métriques utilisées pour décrire les enregistrements

# RÉSULTATS: RICHESSE SPÉCIFIQUE



- Moins bonne proportion de jours correctement évalués quand présence d'espèce rare dans la zone: plus forte probabilité de manquer ses vocalisations

# MÉTHODE : TAUX D'APPEL QUOTIDIEN

- Nombre moyen de vocalisation détectées par jours

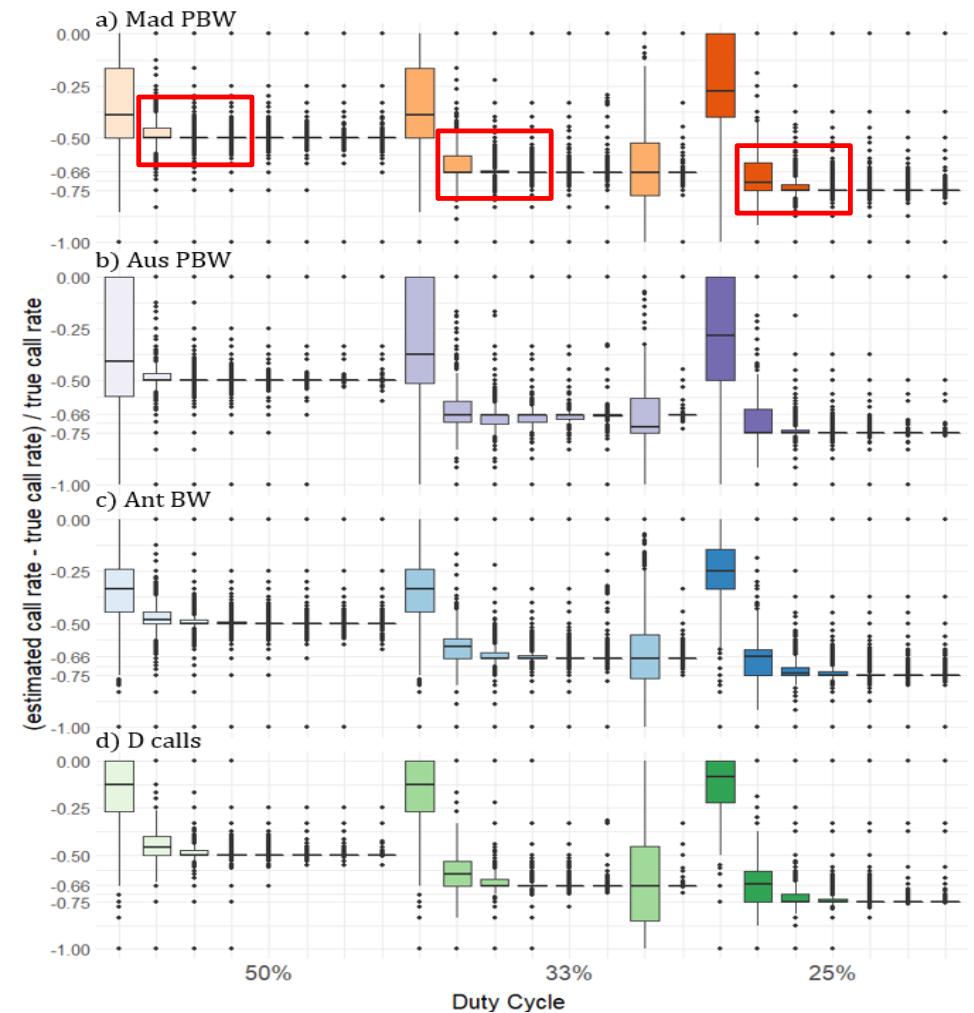
$$\bar{\gamma}_j = \frac{1}{N} \sum_{i=1}^N \gamma_{i,j},$$

- Calcul du ratio entre le taux d'appel estimée par les données sous échantillonnées et le taux réel des données continues

$$ratio_j = \frac{\bar{\gamma}_j - \gamma_{true}}{\gamma_{true}},$$

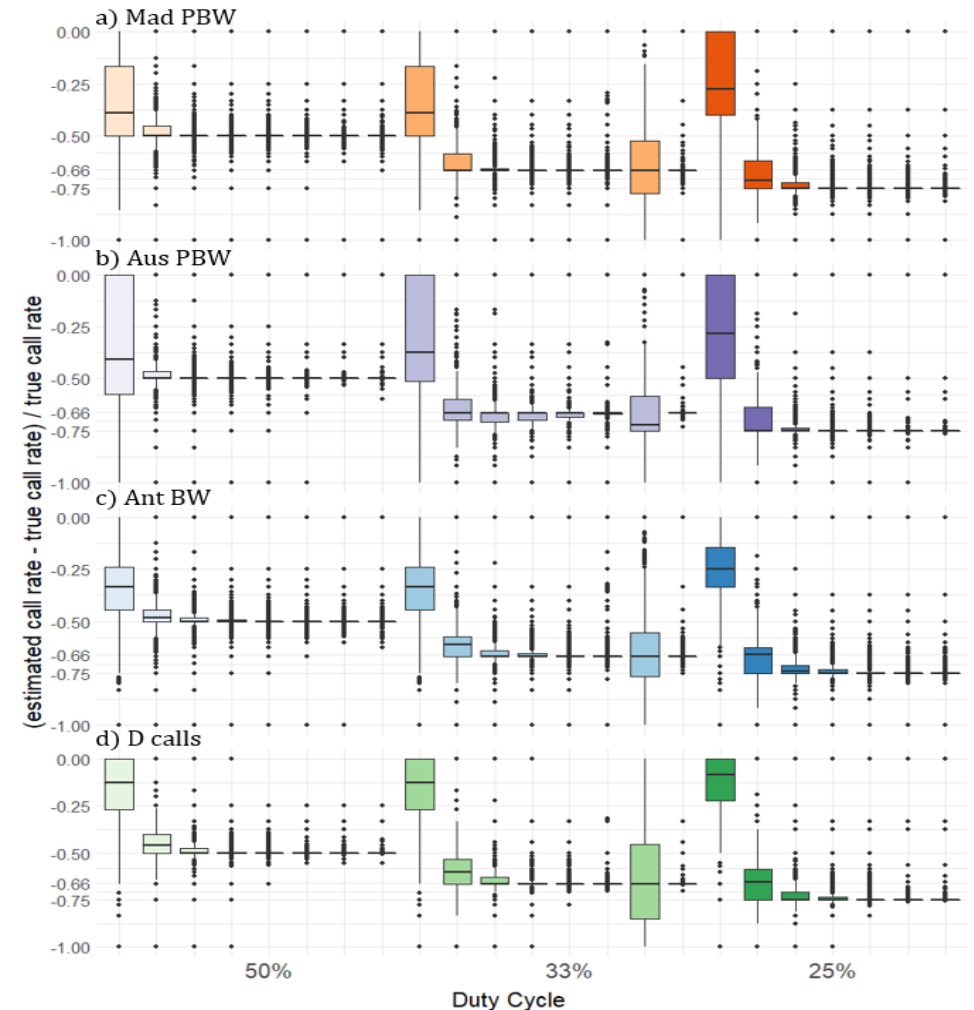
# RÉSULTATS : TAUX D'APPEL QUOTIDIEN

- Variabilité du taux d'appel quotidien estimé augmente avec la diminution du duty cycle



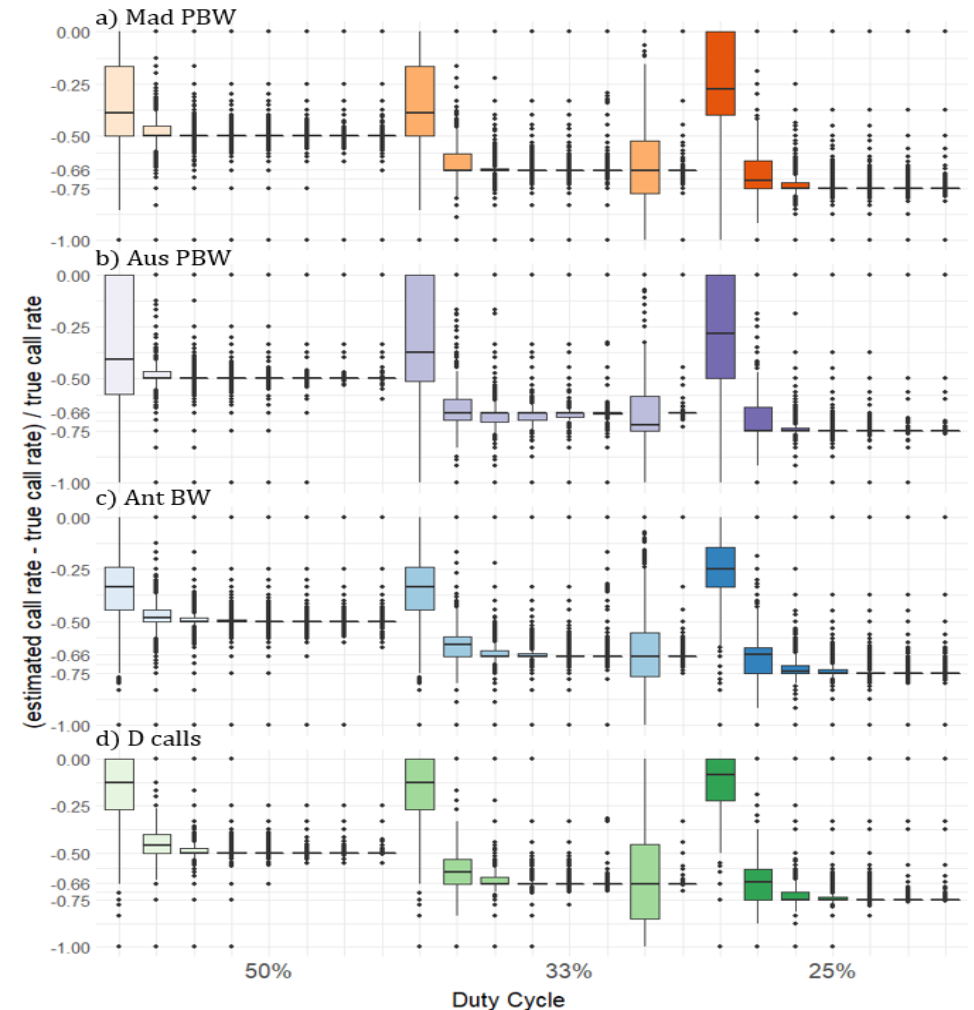
# RÉSULTATS : TAUX D'APPEL QUOTIDIEN

- Variabilité du taux d'appel quotidien estimé augmente avec la diminution du duty cycle
- Dans un même duty cycle: variabilité plus importante pour les courtes périodes d'écoute (1 à 30 minutes)



# RÉSULTATS : TAUX D'APPEL QUOTIDIEN

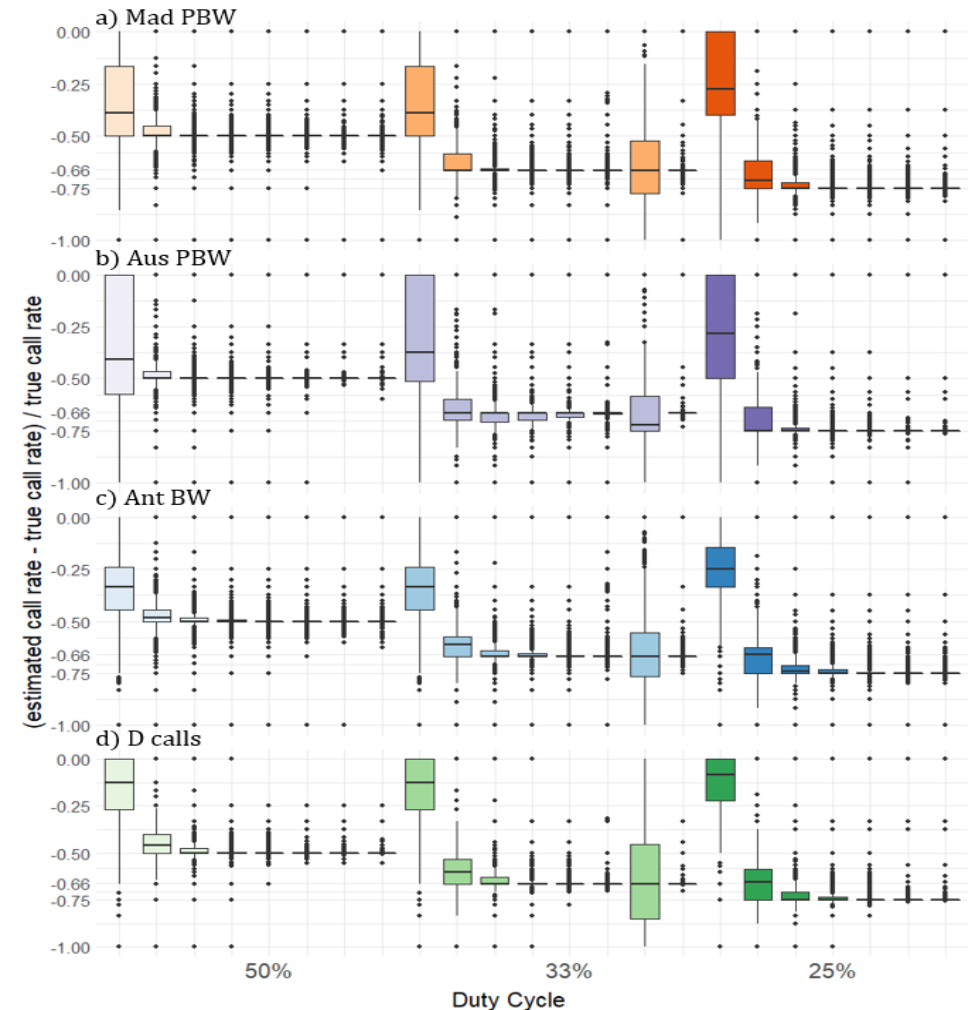
- Variabilité du taux d'appel quotidien estimé augmente avec la diminution du duty cycle
- Dans un même duty cycle: variabilité plus importante pour les courtes périodes d'écoute (1 à 30 minutes)
- Période d'écoute d'1 minute : s'écarte de la valeur attendue avec une très forte variabilité





# RÉSULTATS : TAUX D'APPEL QUOTIDIEN

- Variabilité du taux d'appel quotidien estimé augmente avec la diminution du duty cycle
- Dans un même duty cycle: variabilité plus importante pour les courtes périodes d'écoute (1 à 30 minutes)
- Période d'écoute d'1 minute : s'écarte de la valeur attendue avec une très forte variabilité
- Plus grande variabilité quand les vocalisations sont irrégulières: D calls ou lorsqu'il y a peu de détections



## DISCUSSION: EFFETS SUR LES PATRONS SAISONNIERS

- Meilleure représentation de la saisonnalité quand saisonnalité très marquée et beaucoup de détections
- Moins bonne représentation quand faible activité acoustique
- Collecte des données en continues quand espèce rare ou en transit
- Prendre en compte la durée de la vocalisation pour le choix de la période d'écoute si utilisation d'un détecteur utilisant une fenêtre de durée de vocalisation

## DISCUSSION: EFFET SUR LE TAUX D'APPEL QUOTIDIEN

- Plus grande variabilité pour les courtes périodes d'écoute
- En particulier 1 min qui s'éloigne même de la valeur attendue
- Plus grande variabilité lorsque les appels sont irréguliers Dcalls ou lorsqu'il y a moins de détections

## DISCUSSION: EFFET SUR LA RICHESSE SPÉCIFIQUE

- Richesse spécifique sous estimée plus le duty cycle est petit
- Petite période d'écoute entre 5 et 30 minutes sont meilleures que les grandes, 1 min encore moins bon
- Meilleurs résultats pour des sites très fréquentés par les espèces présentes, si une espèce rare alors moins bon résultats
- Connaissance de l'utilisation de la zone par les espèces ciblée