

Ré-ingénierie de la chaîne de traitements T3 : T3+
(basée sur les travaux de A. & V. Fonteneau, R. Pianet)
GT T3 2011, La Laguna

Pascal CAUQUIL, Emmanuel CHASSOT

IRD, UMR 212 EME, Observatoire Thonier - prenom.nom@ird.fr

25-27 avril 2011



- 1 Buts de T3
- 2 Etapes du projet
- 3 Nouvelle architecture
- 4 Algorithmes

Plan

1 Buts de T3

2 Etapes du projet

3 Nouvelle architecture

4 Algorithmes

Correction de composition spécifique

Biais dans les livres de bord

L'identification des juvéniles de thons n'est généralement pas faite à bord car :

- Les petits patudos et albacores se ressemblent beaucoup
- L'ensemble des petits thons ont une même valeur marchande

Ils sont généralement considérés sous l'appellation de l'espèce dominante (le plus souvent le listao). Les prises d'albacores (YFT) et de patudos (BET) sont ainsi sous-estimées, les prises de listaos (SKJ) surestimées.

Echantillonnage multi-spécifique

L'échantillonnage multispécifique réalisé au débarquement fournit, pour les cuves échantillonnées, la composition spécifique et les structures de taille

Correction de composition spécifique

En partant du principe que l'échantillonnage est réalisé selon un protocole adapté, on applique des extrapolations pour construire une information de taille et de composition spécifique couvrant l'ensemble des captures.

Autres traitements

Correction des poids capturés des livres de bord et des efforts

Dans tous les calculs, le poids capturé pris comme référence est celui des livres de bord. Cette information doit d'abord être corrigée car :

- C'est une estimation visuelle (par les patrons)
- Des livres de bord peuvent manquer dans les bases de données

Extrapolation des échantillons

Les individus échantillonnés ne représentent qu'une petite partie des captures totales. Pour obtenir des structures de taille à l'échelle des captures totales, Les échantillons sont d'abord extrapolés au poids des calées échantillonnées, puis au poids de toutes les calées.

Calcul des efforts

- Temps de mer
- Temps de pêche
- Temps de recherche

Synthèses

Sorties générées

- Captures totales corrigées en poids et en composition spécifique
- Echantillons extrapolés au poids des captures totales
- Efforts

Le tout peut être agrégé selon différents types de zones : carrés CWP, ZEE, ZFAO, etc.

Dans T3+ les sorties seront découplées des calculs et générées par requêtes sur la base de données PostGIS

Plan

1 Buts de T3

2 Etapes du projet

3 Nouvelle architecture

4 Algorithmes

Conception

- Analyse des programmes existants (documentation, réunions)
- Conception de la nouvelle architecture
- Rédaction du cahier des charges (future documentation de référence)
- Multiples compléments et ajustements

Développement

- Transmission de compétences au développeur
- Conception du nouveau modèle de données
- Mise en place du socle technique
- Correction des jeux de données AVDTH
- Développement des algorithmes
- Tests, comparaisons, ajustements

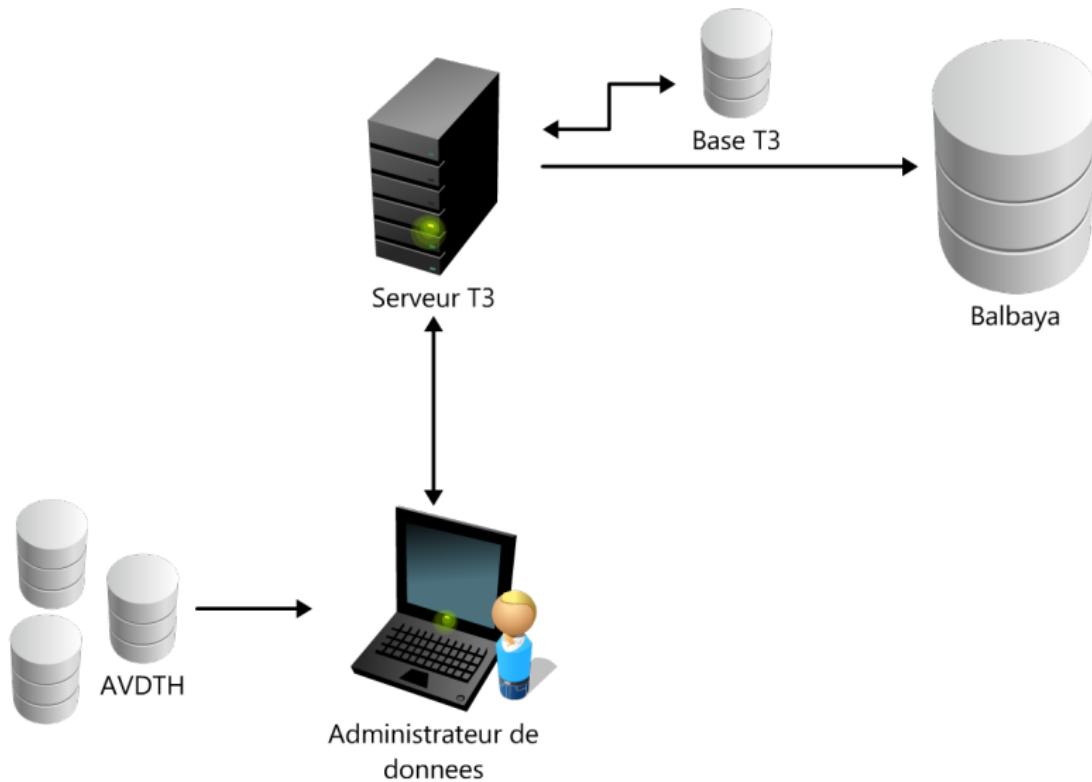
Deuxième cycle

- Amélioration des algorithmes
- Amélioration des interfaces
- Extension du modèle de données pour gestion de :
 - plusieurs versions des données
 - plusieurs versions des zones
 - l'historisation à différents niveaux
- Fonctions de gestion de données :
 - mise à jour
 - remplacement
 - suppression
- Développement des formats de sortie :
 - données agrégées
 - cartes, graphiques
 - rapports
- Tests, comparaisons, ajustements

Plan

- 1 Buts de T3
- 2 Etapes du projet
- 3 Nouvelle architecture
- 4 Algorithmes

Principe général



Vers une gestion intégrée des données thonières

Gestion intégrée des données thonières

- Base de données qui contiendra :
 - les données de captures, débarquements, plans de cuves et échantillons
 - dans leurs versions primaire, intermédiaires et finale
- Procédure d'importation automatique depuis AVDTH, avec contrôles d'intégrité
- Traitements T3 appliqués directement sur la base de données
- Fonctions d'administration : référentiel, modification, recalcul, suppression
- Outils de traçabilité

Application serveur

- Application hébergée sur le serveur de BD pour optimiser les performances
- Affranchir l'utilisateur de l'installation
- Maîtriser la version du logiciel utilisée
- Installation locale toujours possible

PostGIS et traçabilité

Fonctions d'analyse spatiale pré-existantes

T3 se prête bien à l'exploitation des fonctions d'analyse spatiales de PostGIS, notamment pour la correction de composition spécifique et l'extrapolation des échantillons aux captures

Fonctions de journalisation

- Génération de journaux détaillés sur les données et les calculs
- Historisation

Plan

1 Buts de T3

2 Etapes du projet

3 Nouvelle architecture

4 Algorithmes

Entrée : Centralisation des données et calcul des paramètres

Vérifications d'intégrité

- Reconstruction de la fk MAREE.C_FLOTTE -> PAYS
- Reconstruction de la fk ECH_CALEE -> ACTIVITE
- Reconstruction de la fk CUVE_CALEE -> ACTIVITE (si plan de cuves)
- Reconstruction de la fk ECHANTILLON -> MAREE
- Reconstruction de la fk ECHANTILLON -> CUVE

Calcul et application des facteurs de correction du poids capturé

- $rf1$: Ajustement des captures déclarées aux lots commerciaux. Application aux captures uniquement.
- $rf2$: Compensation des livres de bord manquants aux lots com. Application aux captures et efforts.

Facteurs non pris en charge à l'heure actuelle

- $rf3$: Compensation des marées absentes (captures et lots com.)

Niveau 1 : Standardisation des échantillons et extrapolation aux calées échantillonnées

Définitions

- Où i est l'échantillon, e l'espèce
- $rf4_{(i,e)}$ le ratio qui permet d'extrapoler l'effectif mesuré à l'effectif complet de l'échantillon
- $N_{mes(i,e)}$ le nombre d'individus mesurés
- $N_{tot(i,e)}$ le nombre d'individus comptés
- $N_{base(i,e)}$ l'effectif mesuré de (i,e)
- $N_{rf4(i,e)}$ l'effectif mesuré extrapolé à l'échantillon complet

Extrapolation des individus comptés au nombre mesuré

- Pour chaque (i,e) , calcul du ratio $rf4_{(i,e)} = \frac{N_{mes(i,e)}}{N_{tot(i,e)}}$
- Notification du $rf4_{(i,e)}$ calculé
- Application du $rf4_{(i,e)}$: $N_{rf4(i,e)} = N_{base(i,e)} * rf4_{(i,e)}$

Niveau 1 : Standardisation des échantillons et extrapolation aux calées échantillonnées

Standardisation des mesures en tout LF

- Classes de taille AVDTH : LD = 0,5 cm ; LF = 1 cm
- Conversion LD AVDTH (0,5cm) -> LF (2cm) via les clés de conversion
- Recodage des classes de taille LF AVDTH (1 cm) -> LF T3 (2 cm) pour les espèces YFT, BET, ALB
- Finalement : espèces SKJ, LTA, FRI : classes de 1cm ; espèces YFT, BET, ALB : classes de 2 cm
- Algorithme générique

Niveau 1 : Standardisation des échantillons et extrapolation aux calées échantillonnées

Calcul capture par catégorie +10/-10kg (pour chaque calée d'échantillon)

- Si le plan de cuves existe
 - Obtention directe de $P_{(i,j,k)}$, poids capturé de catégorie k dans la calée j ; seules les calées de l'échantillon-cuve i sont considérées
- Sinon, si $(\text{ECH.V_POIDS_M10} <> 0 \text{ || } \text{ECH.V_POIDS_P10}) \text{ && } \text{ECH.V_POIDS_ECH} = 0$
 - Calcul de $Pr_{(i,j)}$, proportion de la calée j dans la cuve i (à partir de ECH_CALEE.V_POND)
 - $P_{(i,j,k)} = Pr_{(i,j)} * P_{(j,k)}$ à partir de $\text{ECH.V_POIDS_M10}/\text{V_POIDS_P10}$
 - où i est l'échantillon-cuve, j la calée de l'échantillon-cuve i , k la catégorie de poids (+10/-10kg), $P_{(i,j,k)}$ le poids de la catégorie k dans la calée j de la cuve i , $Pr_{(i,j)}$ la proportion de la calée j dans la cuve i , $P_{(j,k)}$ le poids de la catégorie k dans la cuve i
- Sinon, si $(\text{ECH.V_POIDS_M10} = 0 \text{ && } \text{ECH.V_POIDS_P10} = 0 \text{ && } \text{ECH.V_POIDS_ECH} <> 0)$
 - Calcul différencié +10/-10kg impossible. Calcul de $P_{(i,j,k=0..\infty)}$
- Sinon, si $(\text{ECH.V_POIDS_M10} = 0 \text{ && } \text{ECH.V_POIDS_P10} = 0 \text{ && } \text{ECH.V_POIDS_ECH} = 0)$
 - Echantillon inexploitable

Niveau 1 : Standardisation des échantillons et extrapolation aux calées échantillonnées

Distribution des effectifs (ECH_FRQT.V_EFF) $E_{cuve} \rightarrow E_{calee}$

- $N_{P_{ech}(i,j,l)} = N_{P_{ech}(i,l)} * Pr_{(i,j)}$
- Où i est l'échantillon-cuve, j l'échantillon-calée, l la classe de taille, $N_{P_{ech}(i,j,l)}$ l'effectif d'individus de la classe de taille l dans (i,j) , $N_{P_{ech}(i,l)}$ l'effectif d'individus de la classe de taille l dans i , $Pr_{(i,j)}$ la proportion de la calée i dans la cuve j

Extrapolation des effectifs distribués $P_{ech} \rightarrow P_{calee}$

- Calcul des paramètres d'extrapolation $rf_{(i,j,k)}$ (i.e. rf_{m10} , rf_{p10} ou rf_{tot})
- $N_{P_{calee}(i,j,k)} = \sum_{l \in k} (N_{P_{ech}(i,j,l)} * rf_{(i,j,k)})$
- Où i est l'échantillon-cuve, j l'échantillon-calée, k la catégorie $\{+10 ; -10\text{kg}\}$, l la classe de taille, $N_{P_{calee}(i,j,k)}$ l'effectif d'individus de catégorie k extrapolé au poids de la calée j , $N_{P_{ech}(i,j,l)}$ l'effectif d'individus de classe de taille l dans la calée j de la cuve échantillonnée i

Niveau 1 : Standardisation des échantillons et extrapolation aux calées échantillonnées

Regroupement par catégories de poids

- Conversion des classes de taille en classes de poids
- $N_{(i,j,m)} = \sum_{(l \in m)} (N_{(i,j,l)})$
- Où i est l'échantillon-cuve, j l'échantillon-calée, m la catégorie $\{-10\text{kg} ; 10-30\text{kg} ; +30\text{kg}\}$ de la nouvelle catégorisation, $N_{(i,j,l)}$ l'effectif d'individus de la classe de poids l où l est inclus dans m , $N_{(i,j,m)}$ l'effectif d'individus de la classe de poids m selon la nouvelle catégorisation

Niveau 2 : Correction de la composition spécifique

Conversion des catégories de poids : Catégories AVDTH (1/2)

	YFT	BET	ALB	SKJ	LTA	FRI
< 3 kg	x	x	x			
3-10 kg	x	x	x			
10-30kg	x	x	x			
6-20 kg	x	x	x			
30-50 kg	x	x	x			
20-40 kg	x	x	x			
50 kg	x	x	x			
40-60 kg	x	x	x			
< 10 kg	x	x	x			
> 10 kg	x	x	x			
> 30 kg	x	x	x			
? kg	x	x	x	x	x	x

Niveau 2 : Correction de la composition spécifique

Conversion des catégories de poids : Catégories AVDTH (2/2)

	YFT	BET	ALB	SKJ	LTA	FRI
< 1.8 kg				x	x	x
> 1.8 kg				x	x	x
10-30kg				x	x	x
1.8-4 kg				x	x	x
1.8-6 kg				x	x	x
4-6 kg				x	x	x
4-8 kg				x	x	x
6-8 kg				x	x	x
> 8 kg				x	x	x
? kg	x	x	x	x	x	x

Niveau 2 : Correction de la composition spécifique

Conversion des catégories de poids : Catégories T3

	Océan Atlantique	Océan Indien
Banc objet	-10 kg, +10 kg,	-10 kg +10 kg
Banc libre	-10 kg 10-30 kg +10 kg	-10 kg +10 kg

Conversion des catégories de poids : Exemples de conversions

Cat. T3	OA / BL / YFT
< 10 kg(j)	$1_j + 2_j + 4_j * 0.2 + 01_j + 9_j * Pr_{(b,t,e,c,z)}$
10-30 kg(j)	$3_j + 4_j * 0.8 + 6_j * 0.5 + 12_j + 11_j * 0.1 + 9_j * Pr_{(b,t,e,c,z)}$
> 30 kg(j)	$5_j + 7_j + 8_j + 6_j * 0.5 + 13_j + 11_j * 0.9 + 9_j * Pr_{(b,t,e,c,z)}$
? kg	inutilisée

où j est la calée, b le type de banc, t le trimestre, e l'espèce, c la catégorie de poids T3, z la zone ET, $Pr(b,t,e,c,z)$ la proportion de c dans la strate

Niveau 2 : Correction de la composition spécifique

Définitions

- Soient e l'espèce $\{YFT; SKJ; BET\}$, c la catég. $\{<10 \text{ kg}; 10-30 \text{ kg}; >30 \text{ kg}\}$, b le type de banc, t le trimestre, n l'année, z la zone ET, i la calée
- Soit $Sc_{(b,t,n,z)}$ une strate-capture
- Soit $Se_{(b,t,n,z)}$ sa strate-échantillon déterminée par cumul de strates
- Soit $P_{(i,e,c)}$ le poids corrigé de (e,c) dans i
- Soit $Pr_{(e,c)}$ la proportion de (e,c) dans le cumul des échantillons de $Se_{(b,t,n,z)}$
- Soit $P_{(i,c)}$ le poids de c dans i , toutes e confondues

Noyau de l'algorithme

```

Pour chaque  $Sc_{(b,t,n,z)}$  {
  Pour chaque  $i$  de  $Sc_{(b,t,n,z)}$  {
    Pour chaque  $c$  {
      Pour chaque  $e \in \{YFT; SKJ; BET\}$  {
         $P_{(i,e,c)} = Pr_{(e,c)} * P_{(i,c)}$ 
      }
    }
  }
}

```



Niveau 3 : Extrapolation des échantillons aux captures totales

Définitions

- Soient e l'espèce $\{\text{YFT ;SKJ ;BET ;ALB ;LTA ;FRI}\}$, I la classe de taille, b le type de banc, t le trimestre, n l'année, z la zone ET
- Soit $Sc_{(b,t,n,z)}$ une strate-capture
- Soit $Se_{(b,t,n,z)}$ sa strate-échantillon déterminée par cumul de strates
- Soient a_e et b_e les paramètres de la relation taille-poids de e
- Soit $P_{e,I}$ le poids de classe I pour e
- Soit $N_{(e,I,Se_{(b,t,n,z)})}$ l'effectif de (e,I) dans la strate-échantillon $Se_{(b,t,n,z)}$
- Soit $P_{N_{(e,I,Se_{(b,t,n,z)})}}$ le poids de l'effectif de (e,I) dans $Se_{(e,b,t,n,z)}$
- Soit $Pr_{(e,I,Se_{(b,t,n,z)})}$ le pourcentage du poids de l'effectif $P_{N_{(e,I,Se_{(b,t,n,z)})}}$ dans le poids total de l'échantillon de $Se_{(b,t,n,z)}$
- Soit $P_{N_{(e,I,Sc_{(b,t,n,z)})}}$ le poids de l'effectif de (e,I) dans la strate-capture $Sc_{(b,t,n,z)}$
- Soit $P_{(Sc_{(b,t,n,z)})}$ le poids total de la strate-capture $Sc_{(b,t,n,z)}$
- Soit $P_{e,I}$ le poids de (e,I)
- Soit $N_{(e,I,Sc_{(b,t,n,z)})}$ l'effectif de (e,I) dans la strate-capture $Sc_{(b,t,n,z)}$

Niveau 3 : Extrapolation des échantillons aux captures totales

Noyau de l'algorithme

Pour chaque $Sc_{(b, t, n, z)}$ {

 Pour chaque e {

 Pour chaque I {

 Convertir I en poids : $P_{e, I} = a_e * I^{b_e}$

 Calculer le poids de l'effectif : $P_{N_{(e, I, Sc_{(b, t, n, z)})}} = N_{(e, I, Sc_{(b, t, n, z)})} * P_I$

 }

$$Pr_{(e, I, Sc_{(b, t, n, z)})} = \frac{P_{N_{(e, I, Sc_{(b, t, n, z)})}}}{\sum_{I=0}^n (P_{N_{(e, I, Sc_{(b, t, n, z)})}})}$$

On connaît à présent la composition de l'échantillon en poids

 Pour chaque I {

$$P_{N_{(e, I, Sc_{(b, t, n, z)})}} = Pr_{(e, I, Sc_{(b, t, n, z)})} * P_{(Sc_{(b, t, n, z)})}$$

$$N_{(e, I, Sc_{(b, t, n, z)})} = \frac{P_{N_{(e, I, Sc_{(b, t, n, z)})}}}{P_{e, I}}$$

 }

}

Résultats

Format des données résultat

- Tous résultats intermédiaires et finaux sauvés dans la base T3+ (PostgreSQL)
- Dans des tables dédiées

Mode d'interrogation

- Extraction de toute donnée par requête SQL
- La réalisation des sorties synthétiques sera découpée de T3
- Possibilité de requêter la base avec n'importe quel langage :
 - SQL via requêteur (pgAdmin, SquirrelSQL, etc.)
 - R, Sweave
 - PHP
 - etc.
- Traçage de cartes par connexion directe d'un SIG à la base