

# Le nouveau système d'échantillonnage des enquêtes ménages pour les départements d'Outre-Mer

---



---

## Partie 1 -

**Lâcher le dispositif historique pour utiliser le recensement : oui, mais comment ?**

# Combien d'EAR mobiliser pour construire la base de sondage ?

---

La question est simple pour les petites communes :  
*On ne peut pas ne pas utiliser le cycle entier*

La question l'est moins pour les grandes communes.  
*Seule chose sûre : on ne peut pas se contenter la dernière EAR*

# Combien d'EAR mobiliser pour construire la base de sondage en grande commune ?

---

La question peut s'envisager sous trois angles :

- **le volume** : quel est le volume *in fine* de la base quand on mobilise  $n$  EAR ?
- **la précision** : de combien améliore-t-on la précision quand on utilise  $n$  EAR plutôt que  $n-1$  ?
- **la structure de la base** : les EAR étant déséquilibrées en volume et structure, comment apprécier l'effet de « lissage » que produit l'empilement ?  
(*non abordé dans cet exposé, cf. article associé*)

## Combien d'EAR en GC : le volume

---

EAR : année n	EAR : année n-1		EAR : année n-2		
		ponctionné en n-1		ponctionné en n-1	ponctionné en n-2

Les hypothèses et l'exercice :

On va *raisonner en moyenne* dans la situation simplifiée où les EAR ont une taille constante  $X$  et où on doit mobiliser tous les ans  $Y$  fiches-adresses pour l'ensemble des enquêtes à mener.

On est amené à résoudre des problèmes de suites récurrentes.

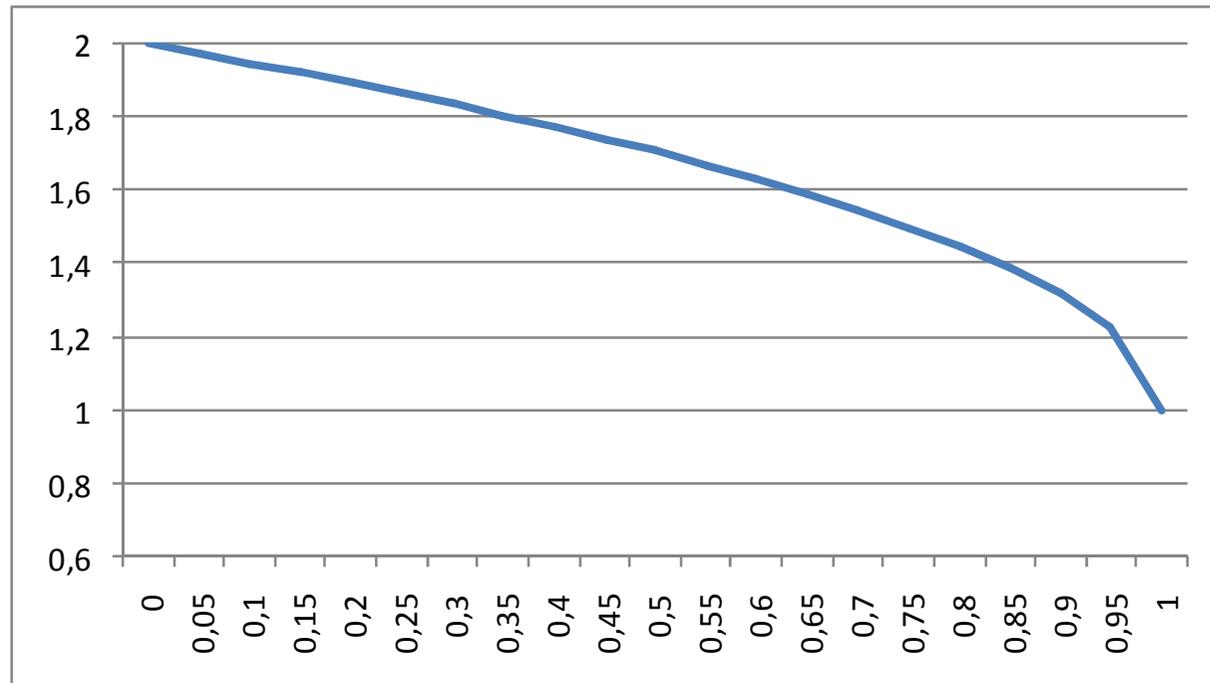
On va noter  $P_n$  la ponction de l'année  $n$  dans la dernière EAR constituée.

# Combien d'EAR en GC – le volume : cas 2 EAR

Le cas est aisé : une suite récurrente simple par une application contractante (d'où existence d'un point fixe attractif) :

$$P_n = Y.X / (2.X - P_{n-1})$$

On en déduit facilement la taille en régime constant de la base de sondage :  $X.(1 + \sqrt{1 - \varepsilon})$  en notant  $\varepsilon = Y/X$



## Combien d'EAR en GC – le volume : cas 3 EAR

---

Le cas est bien plus complexe.

La relation de récurrence est la suivante dans ce cas :

$$P_n = \frac{X^2 \cdot Y}{X^2 + (P_{n-1} - X)(P_{n-2} - 2X)}$$

.....relation d'ordre 2 et non linéaire 😞

# Combien d'EAR en GC – le volume : cas 3 EAR

On va considérer la suite  $(\mathcal{P}_n)_n$  telle que pour tout  $n$  on ait  $\mathcal{P}_n = (P_n; P_{n-1})$ .

Cela correspond à la suite définie par la relation de récurrence suivante :

$$\mathcal{P}_{n+1} = (P_{n+1}; P_n) = \phi(\mathcal{P}_n) = \phi((P_n; P_{n-1}))$$

$$\text{avec } \phi \left\{ \begin{array}{l} \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} f(x, y) = \frac{X^2 \cdot Y}{3 \cdot X^2 - 2X \cdot x - X \cdot y + x \cdot y} \\ g(x, y) = x \end{pmatrix} \end{array} \right\}'$$

$\phi$  restreinte au compact  $[0; Y]^2$  possède un unique point fixe :

$$(x^*, y^*) = (P; P) \quad \text{avec } P = X - \sqrt[3]{X^2 \times (X - Y)} = X \cdot (1 - \sqrt[3]{1 - \varepsilon}) \quad \text{avec } \varepsilon = Y/X$$

qui induirait une taille de base en régime constant de  $\frac{Y}{1 - \sqrt[3]{1 - \varepsilon}} = \frac{\varepsilon \cdot X}{1 - \sqrt[3]{1 - \varepsilon}}$

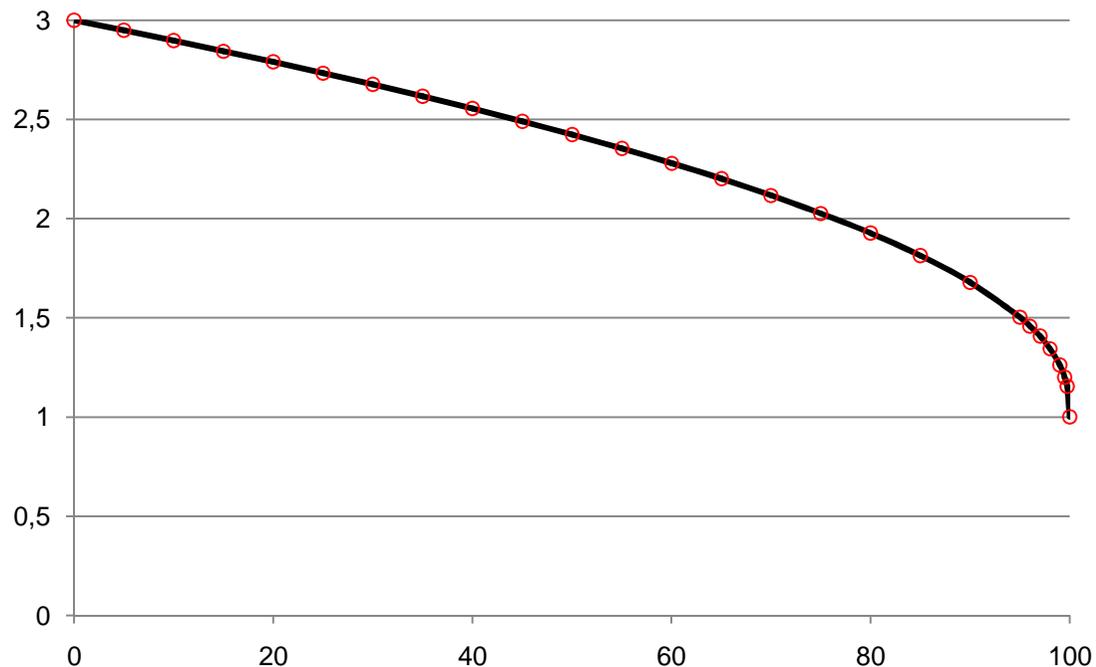
...Ce point fixe est-il attractif ?

# Combien d'EAR en GC – le volume : cas 3 EAR

Il faut calculer la matrice jacobienne de  $\Phi$  au point d'équilibre et montrer que son rayon spectral  $\gamma$  est strictement inférieur à 1.

La démonstration est technique (cf. actes) ; on montre que le point fixe est systématiquement (localement) attractif, sauf pour un cas dégénéré, quand  $Y = X$  (valeurs propres 0 et -1) ; il y a quand même une convergence, mais extrêmement lente.

Taille de la  
base fonction  
de  $\varepsilon$  :



## Combien d'EAR en GC : la précision

---

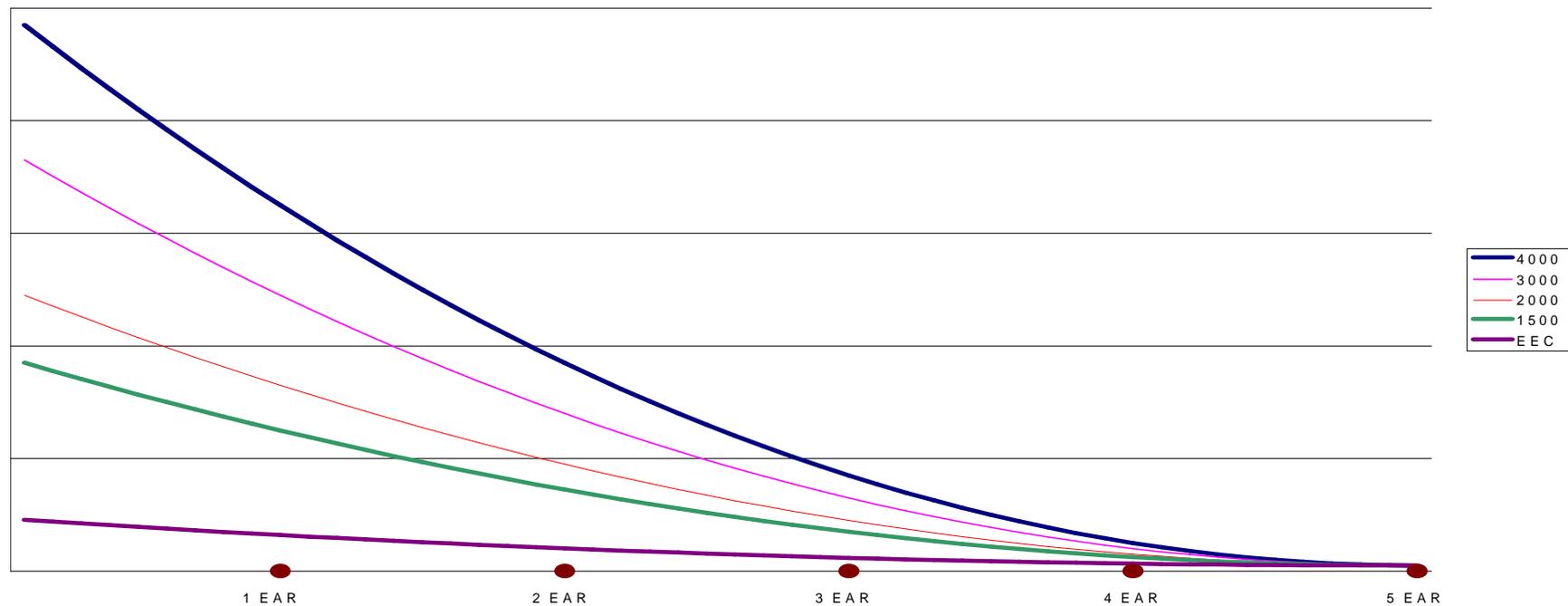
Pour le tirage d'enquêtes, les empilements possibles d'EAR s'interprètent comme *une première phase d'échantillonnage*

**Les intuitions, relativement à un tirage en une passe (i.e. tirage dans l'ensemble du cycle) :**

- A taille d'échantillon donnée : plus on utilise d'EAR, plus on doit être précis
- A nombre d'EAR donné : plus l'enquête est volumineuse, moins on doit être précis

## Combien d'EAR en GC : la précision

Si on pouvait rapporter les variances ou les écarts-type des situations croisées « à nombre d'EAR donné » et « à taille d'échantillon donnée » au tirage en une passe « à taille donnée d'échantillon », on devrait obtenir une représentation du type :



# Combien d'EAR en GC : la précision

---

Méthode par simulation (proposée par P. Ardilly, mise en œuvre au Criem)

Sur quelle population va-t-on raisonner ?

*On va « créer » une population à partir du recensement*

Sur « quel recensement » va-t-on raisonner ?

*On va créer des « pseudo-recensements » sur cette « pseudo-population » (création – par SAS – de 5 GR, et ponction de 60% de chacun)*

# Combien d'EAR en GC : la précision

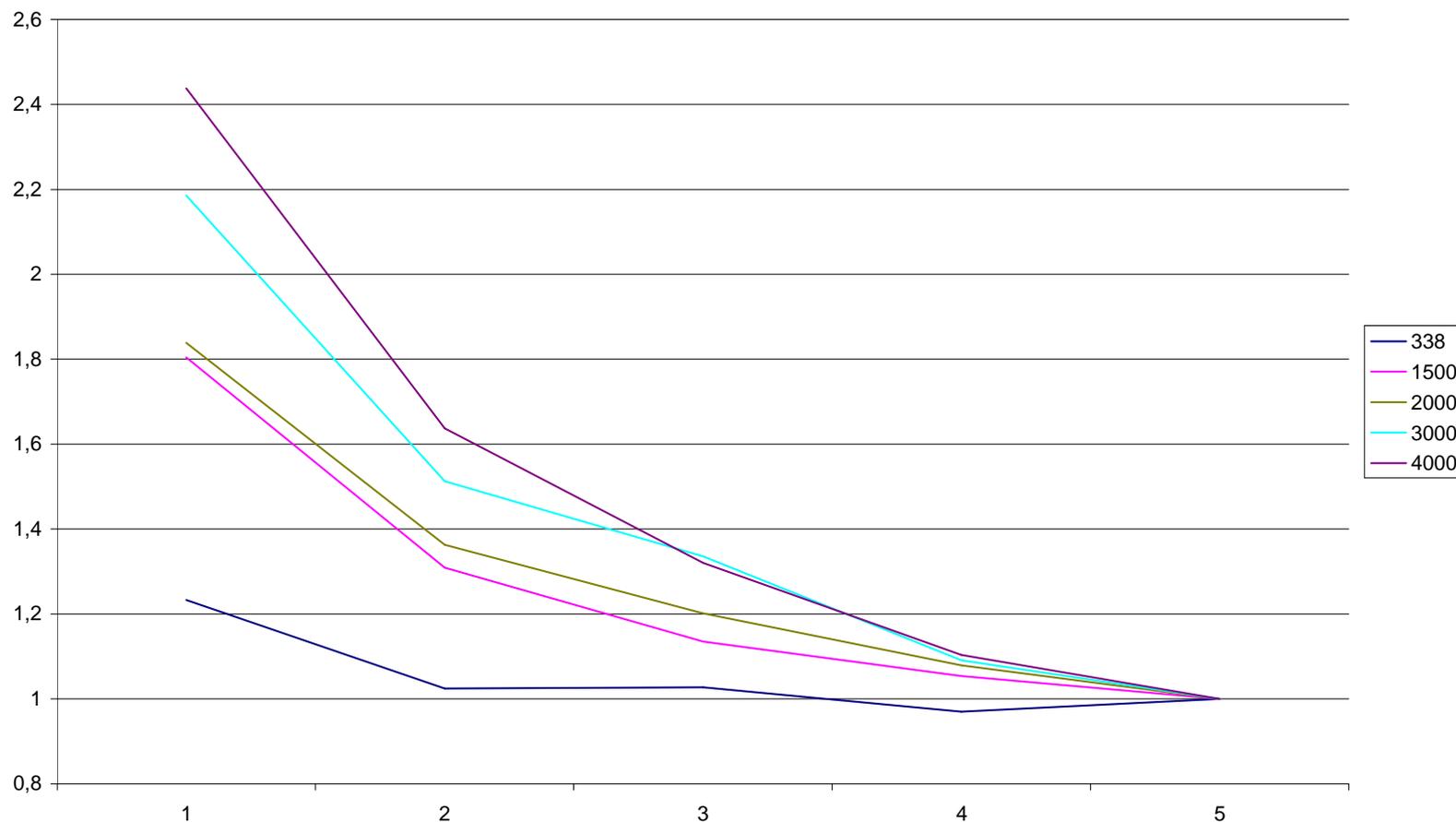
---

- pour chaque grande commune de chaque DOM, chaque valeur classique de taille d'échantillon (taille EEC, taille 1500, 2000, 3000, 4000), et chaque empilement de  $n$  EAR ( $n=1, 2, 3, 4$ ), on fait 1000 fois l'opération suivante :
  - on crée une pseudo-population et un pseudo-recensement
  - on tire deux échantillons :
    - un dans les caractéristiques du cas (nombre d'EAR, taille d'échantillon)
    - un de contrôle (on tire directement la même taille d'échantillon dans le pseudo-RP)
  - on estime pour chacun des deux échantillons le nombre de chômeurs et le nombre d'actifs occupés (au sens du recensement)
- dans chacun des 20 cas (1 à 4 EAR, cinq tailles d'échantillon), on rassemble les chiffres par région
- pour chaque cas (considéré régionalement cette fois), on a 2 fois 1000 valeurs (le cas et son contrôle) ; on calcule les variances empiriques du cas et du contrôle et on rapporte les *écarts-type*.

# Combien d'EAR en GC : la précision

## exemple graphique d'un résultat de la simulation

rapports des écarts-types de l'estimation du nombre d'actifs occupés sur les grandes communes de Guadeloupe, en fonction du nombre d'EAR mobilisées



# Combien d'EAR en Grande Commune ?

---

## Conclusion :

Le dispositif « standard » mobilisera deux EAR

La base contiendra cependant l'ensemble du cycle afin de garantir assez de flexibilité au dispositif, en pouvant mobiliser plus d'EAR si nécessaire pour éviter des situations « limites »

---

## Partie 2 -

# Intégrer les contraintes de collecte dans la nouvelle méthode d'échantillonnage

# Le paradoxe de l'EEC DOM

---

## Disperser dans l'espace

- Grappe de logements contigus en métropole
- Tirage dispersé sur tout le territoire dans les DOM

## Regrouper dans le temps

- Période de collecte très courte sur l'EEC

# Le paradoxe de l'EEC DOM

---

**1 vague = 338 logements dispersés sur tout le territoire**

**Chaque trimestre, 13 enquêtes de 2 semaines et demi**

**1 vague = 13 semaines \* 26 logements**

**Quelle charge enquêteur ?**

- soit 13 charges de 26 logements
- soit 26 charges de 13 logements

**1 vague = 26 secteurs \* 13 logements « regroupés »**  
**sous entendu : affecter 26 enquêteurs à l'EEC**

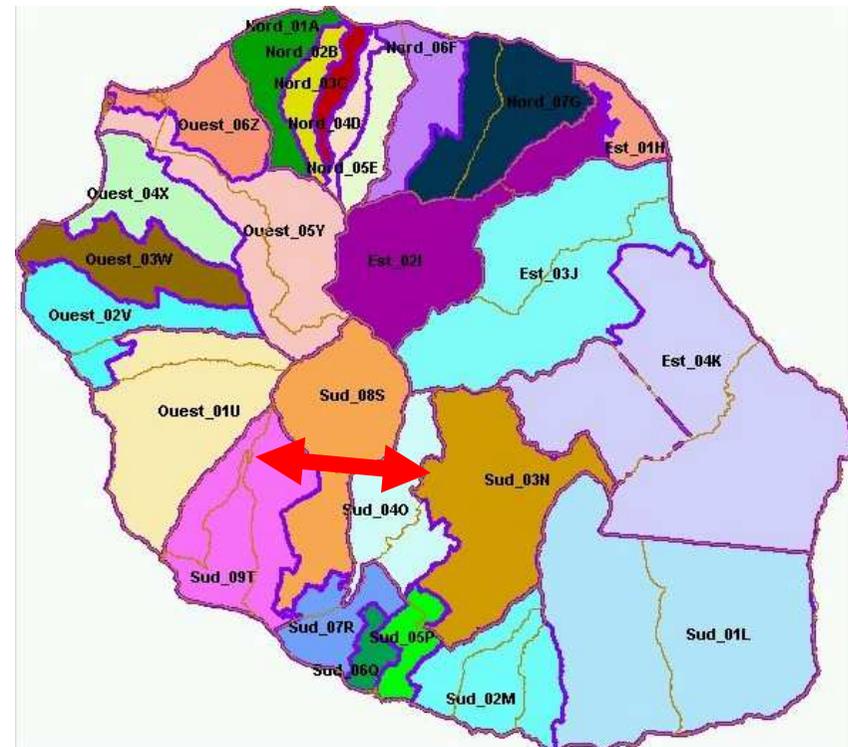
# Des SAE pour la collecte, Des strates pour le tirage

## Les 26 SAE

### Découper chaque territoire

- en 26 secteurs (SAE) de taille proche → logique de collecte
- en strates géographiques fines → logique de tirage

Pourquoi pas 1 strate = 1 SAE ?



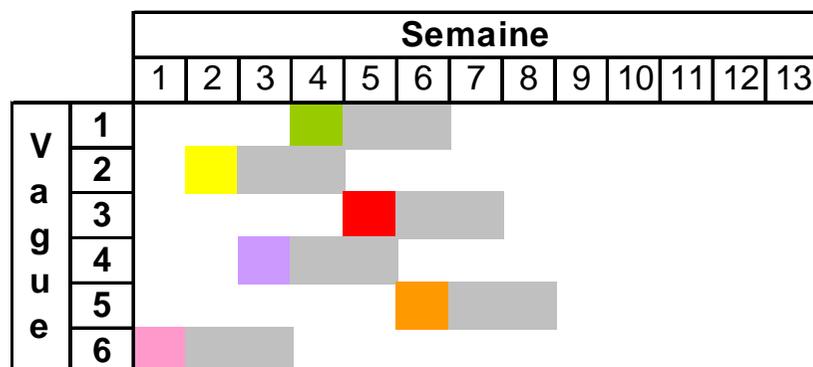
# L'articulation des SAE et de la collecte EEC

## Légende :

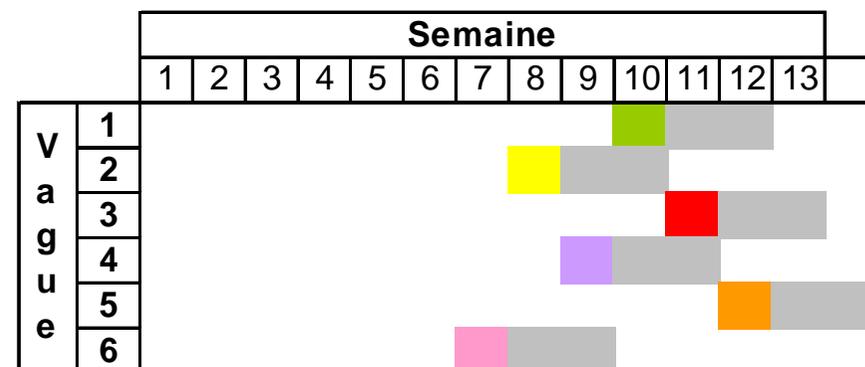
- 1ère interrogation face à face
- 2ème interrogation téléphonique
- 3ème interrogation téléphonique
- 4ème interrogation téléphonique
- 5ème interrogation téléphonique
- 6ème interrogation face à face

- Concentrer les 6 vagues
- Collecte espacée de 6 semaines entre 2 SAE voisins
- Collecte espacée de 2 ou 3 semaines entre 2 charges fàf

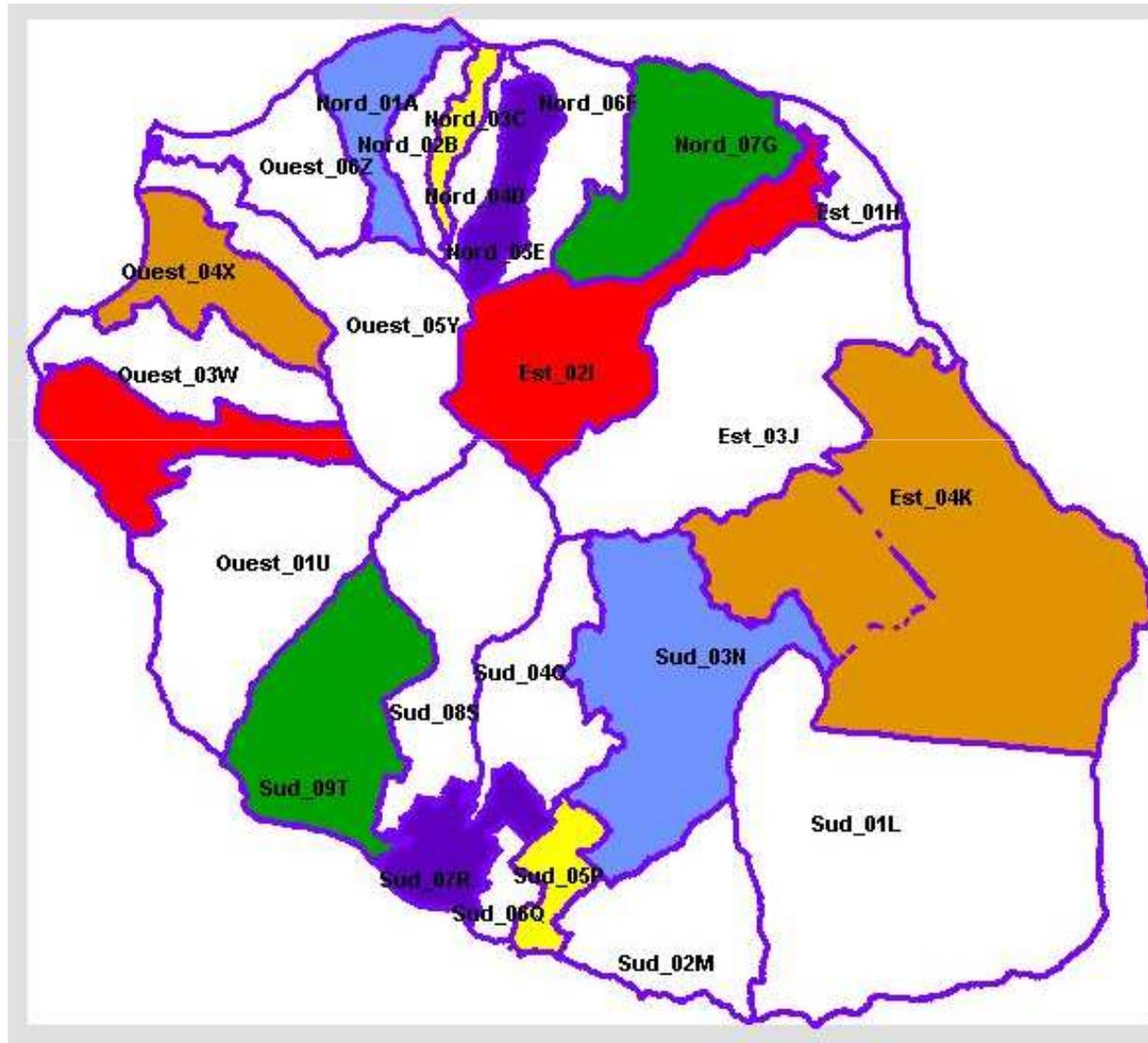
Secteur A



Secteur B



# Pour une semaine de référence donnée...



# Taux de réussite

3 DOM sur 4 avec un taux de réussite  $\geq$  métropole (75%)  
Guyane à 65% mais forte progression en 2014  
Réunion en tête (plus de 85%) !

