

# Génération et validation d'une population synthétique

## Application à la région Auvergne

Maxime Lenormand & Floriana Gargiulo

Laboratoire d'Ingénierie pour les Systèmes Complexes - IRSTEA  
Clermont-Ferrand

*Microsimulation sociale : théorie, applications, défis*  
*naXys – 18 Avril 2012*

Pour mieux  
affirmer  
ses missions,  
le Cemagref  
devient Irstea



# Plan

- 1 Etat de l'art
- 2 La région Auvergne
- 3 Génération des ménages
- 4 Génération des activités
- 5 Validation à l'échelle de la région
- 6 Conclusion et perspectives

# Etat de l'art

## Synthetic reconstruction method

- IPFSR *Wilson et al. 1974*
- Iterative Proportional Fitting (IPF) *Beckman et al. 1996*
- Iterative Proportional Updating (IPU) *Ye et al. 2009*

# Etat de l'art

## Synthetic reconstruction method

- IPFSR *Wilson et al. 1974*
- Iterative Proportional Fitting (IPF) *Beckman et al. 1996*
- Iterative Proportional Updating (IPU) *Ye et al. 2009*

## Combinatorial optimization

- *Voas et Williamson 2000*

# Etat de l'art

## Synthetic reconstruction method

- IPFSR *Wilson et al. 1974*
- Iterative Proportional Fitting (IPF) *Beckman et al. 1996*
- Iterative Proportional Updating (IPU) *Ye et al. 2009*

## Combinatorial optimization

- *Voas et Williamson 2000*

⇒ **Nécessite un échantillon détaillé !!!**

# Etat de l'art

## Synthetic reconstruction method

- IPFSR *Wilson et al. 1974*
- Iterative Proportional Fitting (IPF) *Beckman et al. 1996*
- Iterative Proportional Updating (IPU) *Ye et al. 2009*

## Combinatorial optimization

- *Voas et Williamson 2000*

⇒ **Nécessite un échantillon détaillé !!!**

**Gargiulo, F., Ternes, S., Huet, S. & Deffuant, G.** *An iterative approach for generating statistically realistic populations of households.* PLoS ONE (2010)

# Plan

- 1 Etat de l'art
- 2 La région Auvergne**
- 3 Génération des ménages
- 4 Génération des activités
- 5 Validation à l'échelle de la région
- 6 Conclusion et perspectives

# La région Auvergne en 1990

- 1310 municipalités
- 1 321 719 habitants
- 515 736 ménages





# Plan

- 1 Etat de l'art
- 2 La région Auvergne
- 3 Génération des ménages**
- 4 Génération des activités
- 5 Validation à l'échelle de la région
- 6 Conclusion et perspectives

# Génération des ménages

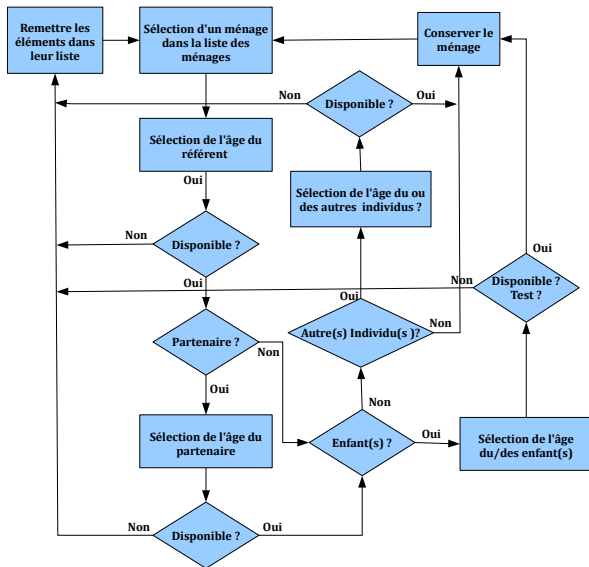
## Données de base

### Pour chaque municipalité :

- Nombre de ménages "vides" par type et taille :  
*15 couples avec enfants de taille 3*
- Nombre d'habitants par classe d'âge :  
*6 individus entre 11 et 15 ans*
- Des distributions de probabilités :  
*Distribution de l'âge du chef de famille par type de ménages*

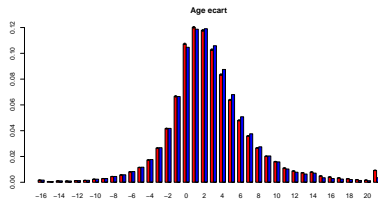
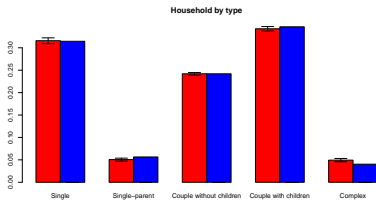
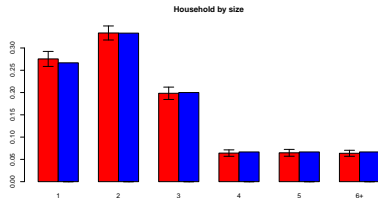
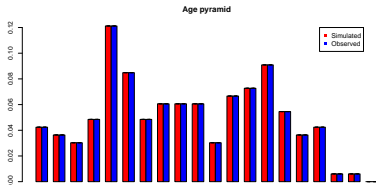
# Génération des ménages

## Diagramme de flux



# Génération des ménages

## Résultats



# Plan

- 1 Etat de l'art
- 2 La région Auvergne
- 3 Génération des ménages
- 4 Génération des activités**
- 5 Validation à l'échelle de la région
- 6 Conclusion et perspectives

# Génération des activités

## Données de base

### Pour chaque municipalité :

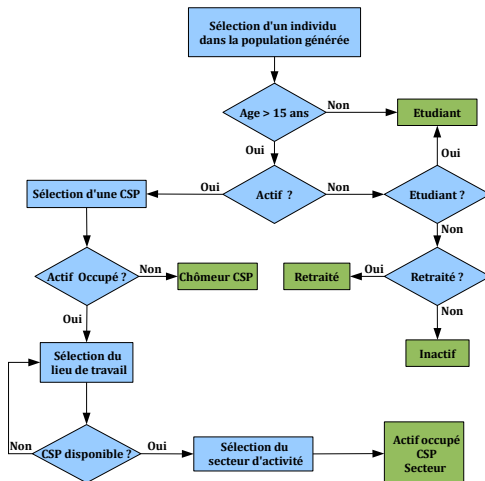
- Les individus de la population artificielle.
- Des distributions de probabilités :  
*Probabilité d'être étudiant en fonction de l'âge*

### Au niveau de la région :

- Réseau de navettage simulé.

# Génération des activités

## Diagramme de flux



# Génération des activités

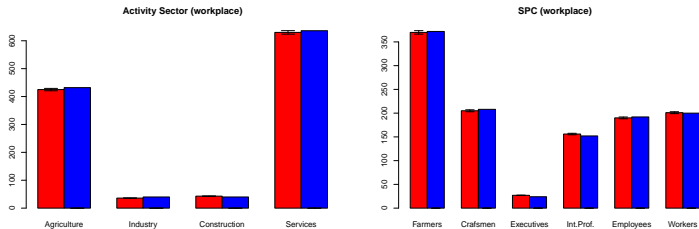
## Génération d'un réseau de navettage

RP \ WP	$u_1$	...	$u_j$	...	$u_n$	Out.	Total
$u_1$	0	...	$Y_{1j}$	...	$Y_{1n}$	$Y_{1out}$	$s_1^{in}$
...	...	...	...	...	...	...	...
$u_j$	$Y_{j1}$	...	$Y_{jj}$	...	$Y_{jn}$	$Y_{jout}$	$s_j^{out}$
...	...	...	...	...	...	...	...
$u_n$	$Y_{n1}$	...	$Y_{nj}$	...	0	$Y_{nout}$	$s_n^{out}$
Out.	$Y_{out1}$	...	$Y_{outj}$	...	$Y_{outn}$		
Total	$s_1^{in}$	...	$s_j^{in}$	...	$s_n^{in}$		



# Génération des activités

## Résultats



# Plan

- 1 Etat de l'art
- 2 La région Auvergne
- 3 Génération des ménages
- 4 Génération des activités
- 5 Validation à l'échelle de la région**
- 6 Conclusion et perspectives

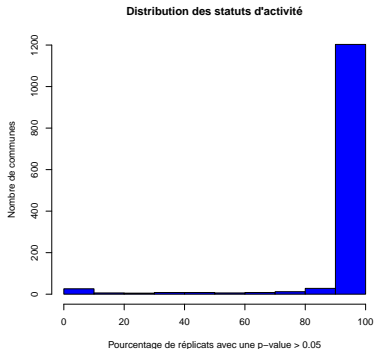
# Validation à l'échelle de la région

## Tests statistiques

- Génération de ( $r \times 1310$ ) populations artificielles
- Calcul de ( $r \times \text{Nb d'agrégats} \times \text{Nb Test Stat}$ )  $p$  – *values*. Le type d'agrégats et le type de tests dépendent des données observées
- Calcul du pourcentage de réplicat "passant" le test par agrégat et par test statistique
- Calcul d'un score global ou représentation graphique par résultat

# Validation à l'échelle de la région

## Tests statistiques



# Plan

- 1 Etat de l'art
- 2 La région Auvergne
- 3 Génération des ménages
- 4 Génération des activités
- 5 Validation à l'échelle de la région
- 6 Conclusion et perspectives**

# Conclusion et perspectives

## Conclusion

- Ménage : 2h57 min pour 50000 x 2 itérations et de nombreux tests à passer
- Activité : 1 min + 15 min pour la génération du réseau
- Pas besoin de données détaillées

## Perspectives

- Comparaison avec d'autres algorithmes
- Etudier l'effet du paramètre "nombre d'itérations"