



LES MODÈLES D'ÉQUATIONS STRUCTURELLES À VARIABLES LATENTES SOUS DEUX APPROCHES : FRÉQUENTISTE ET BAYÉSIENNE



JAD ABOU-GHANTOUS

Bio-statistique

Irstea

jad.abou-ghantous@irstea.fr

PLAN DE LA PRESENTATION

*Avant-propos : mise en relation de variables comparaison glm
versus Confirmatory Path Analysis versus SEM*

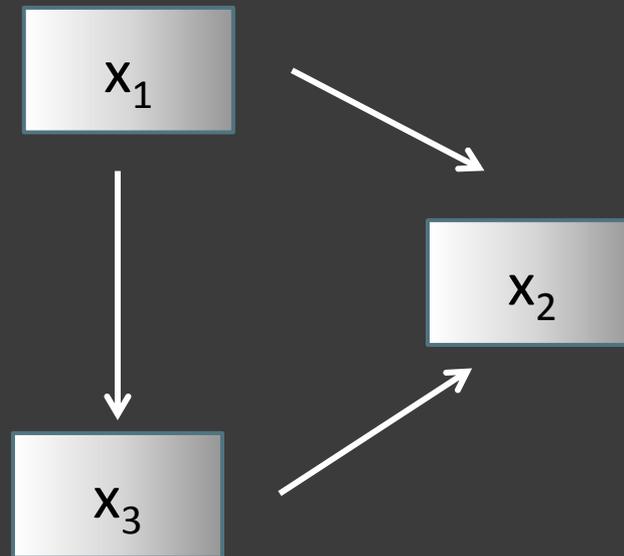
I) SEM sous l'approche Fréquentiste (LISREL)

II) SEM sous l'approche Bayésienne

III) Conclusion

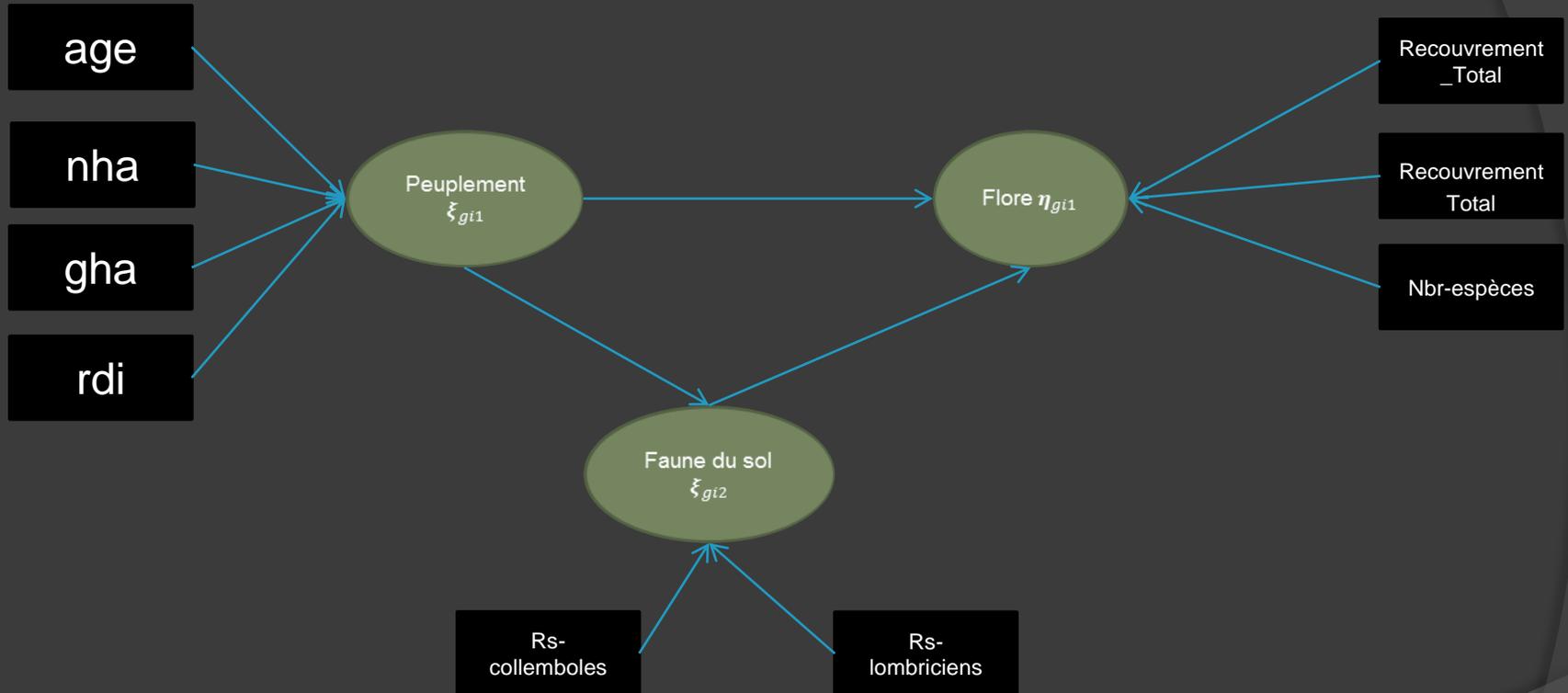
CONFIRMATORY PATH ANALYSIS

CPA Méthode de Bill Shipley : (lm, glm, lme ppm(prend en compte les coordonnées géographiques d'un point, nid..)...)...



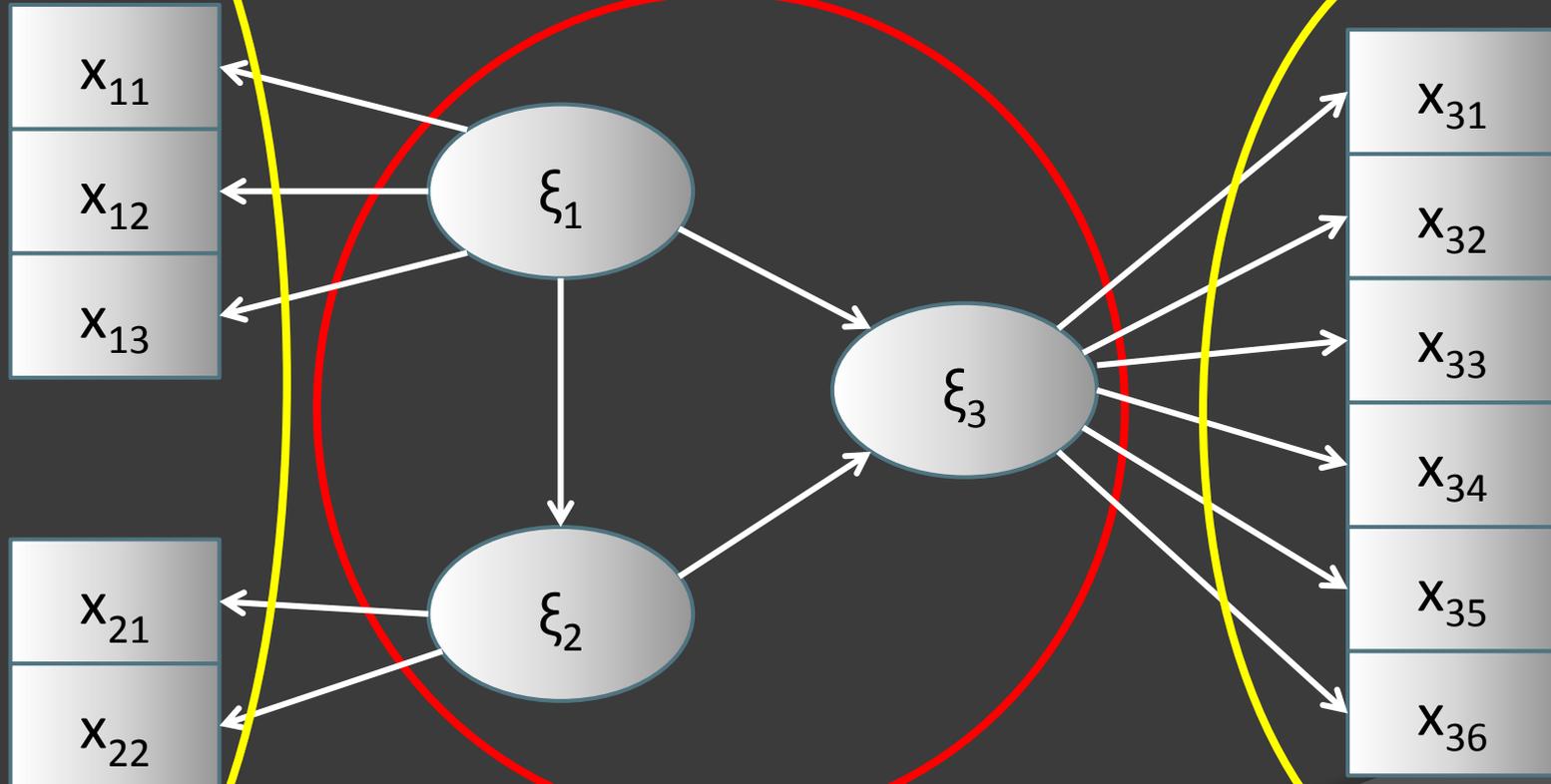
*Limites : complexité du modèle avec l'accroissement du nombre de variables
Impossibilité de rendre compte de méta-variables (type sol, peuplement, etc.)*

EXEMPLE EN ÉCOLOGIE



SEM: VARIABLES OBSERVÉES ET LATENTES

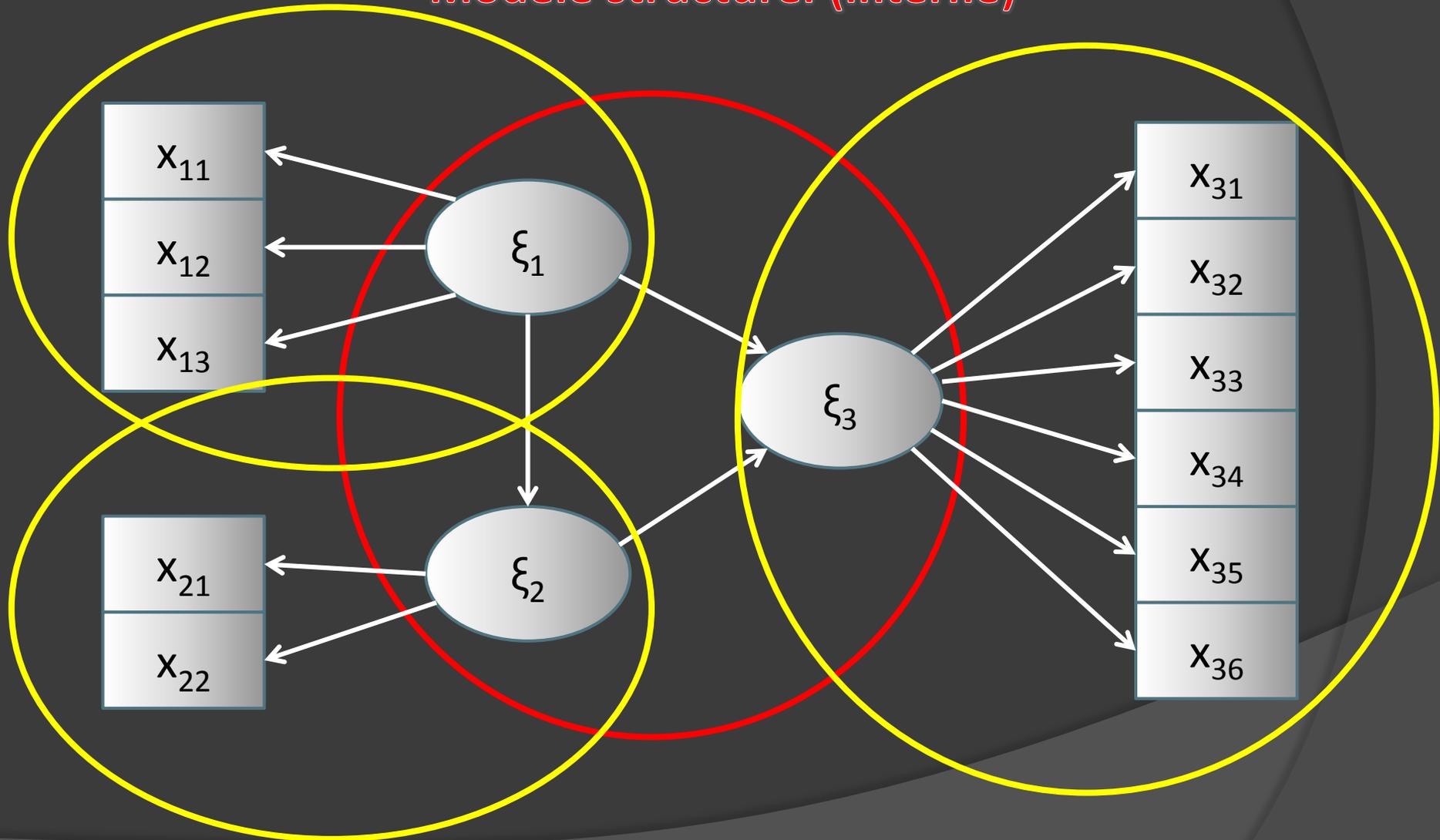
Variables observées/manifestes



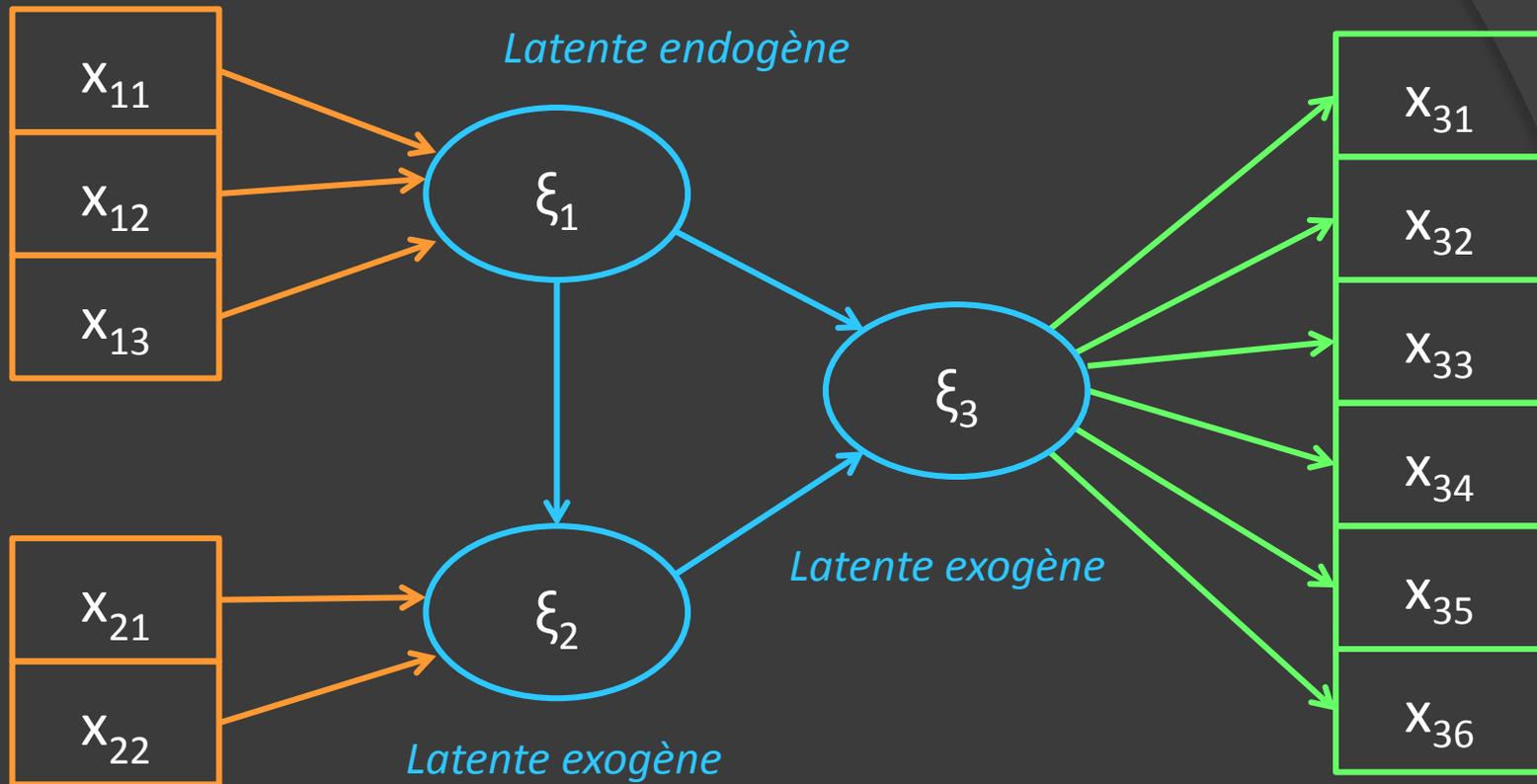
Variables latentes

SEM: MODÈLE STRUCTUREL ET DE MESURE

Modèle structurel (interne)



Modèle de mesure (externe)



Type formatif

Variable latente $\xi_j \rightarrow$ reflet des variables manifestes du bloc x_j

Type réflectif

Variables manifestes \rightarrow reflet de leur variable latente

LISREL (Linear Structural Relationship)

Caractéristiques :

- Formulation d'hypothèses a priori mises à l'épreuve de modèles
- Fondée sur *la covariance*

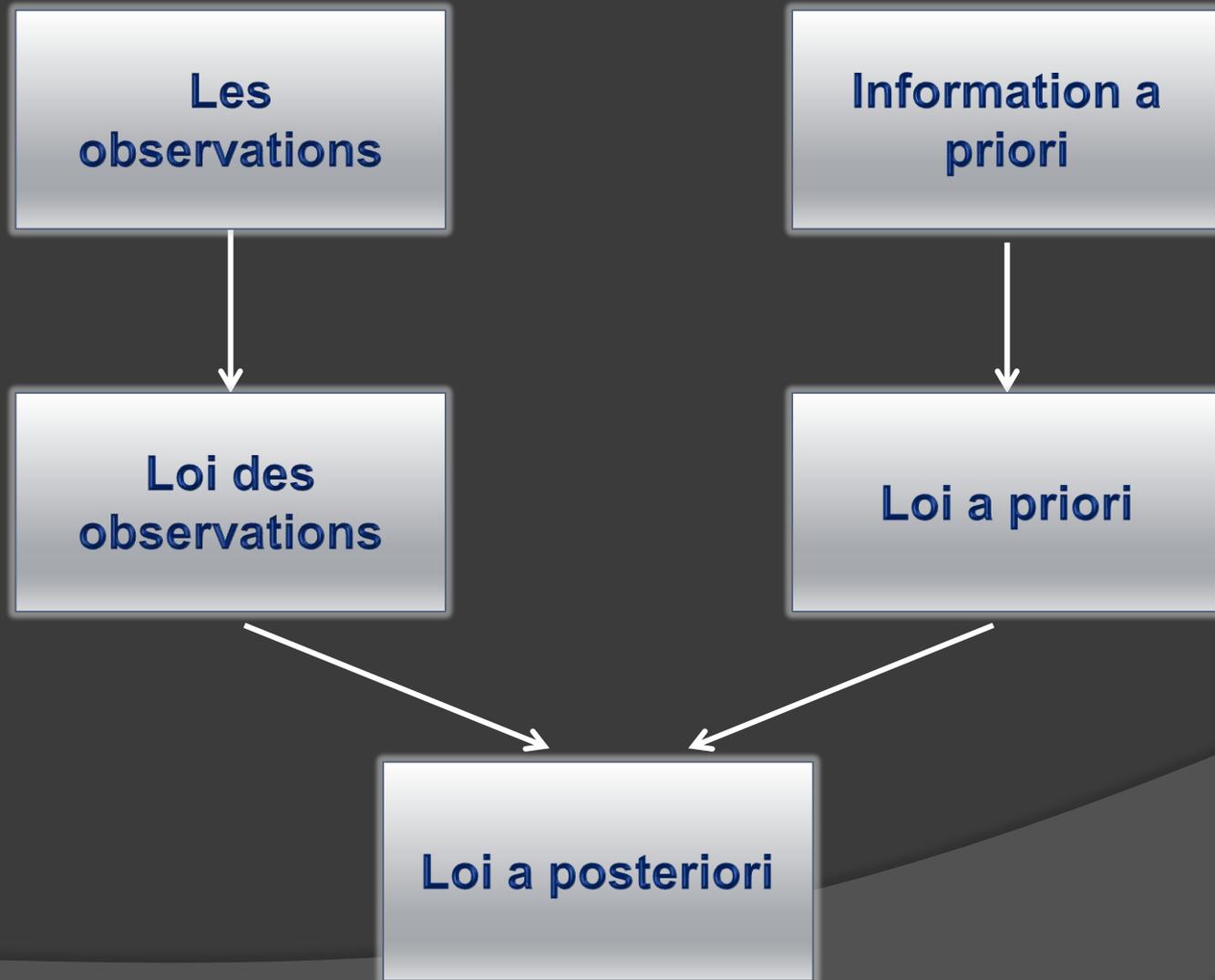
Le principe :

- *Spécification, estimation, comparaison* et *évaluation* des modèles de relations entre variables.
- Test d'une théorie

Procédure :

- Construction d'un modèle
- Collection des données pour tester le modèle
- Modèle comparé aux données et évalué
- Si nécessaire le modèle est modifié et testé avec de nouvelles données

MODELE STATISTIQUE PARAMETRIQUE BAYESIEN



EN QUOI CONSISTENT LES APPROCHES FREQUENTISTE ET BAYESIENNE ?

« Inférence statistique » : problème où il s'agit de retrouver une loi de probabilité inconnue à partir d'un ensemble de réalisations de cette loi.
Exemple de la pièce qu'on lance cent fois.

➤ Point de vue fréquentiste :

- Vraie valeur p : la probabilité d'obtenir pile à partir des résultats obtenus (ex : 53% si 53 x pile)
- Calcul des moyennes

➤ Approche bayésienne :

- Chaque pièce a sa propre valeur de p
- Seconde loi de probabilités, la loi *a priori* sur les pièces elles-mêmes
- On n'estime pas p mais sa loi de probabilité ou fonction de répartition (loi *a posteriori*)

Modèles Bayésiens : complètement probabilistes.

ARGUMENTS EN FAVEUR ET EN DEFAVEUR DU BAYESIEN

Quelques arguments en faveur du Bayésien :

- ✓ Distributions de probabilité aussi sur les paramètres du modèle
- ✓ Possibilité de concevoir des modèles complexes (MCMC)
- ✓ Paramètres basés sur la distribution autant que sur les observations

Quelques arguments en défaveur du Bayésien :

- ✓ Nécessité de définir des distributions sur les paramètres
- ✓ La p valeur (probabilité) postérieure prédictive est plus mauvaise que l'outil utilisant le maximum de vraisemblance chez les Fréquentistes.

CONCLUSION

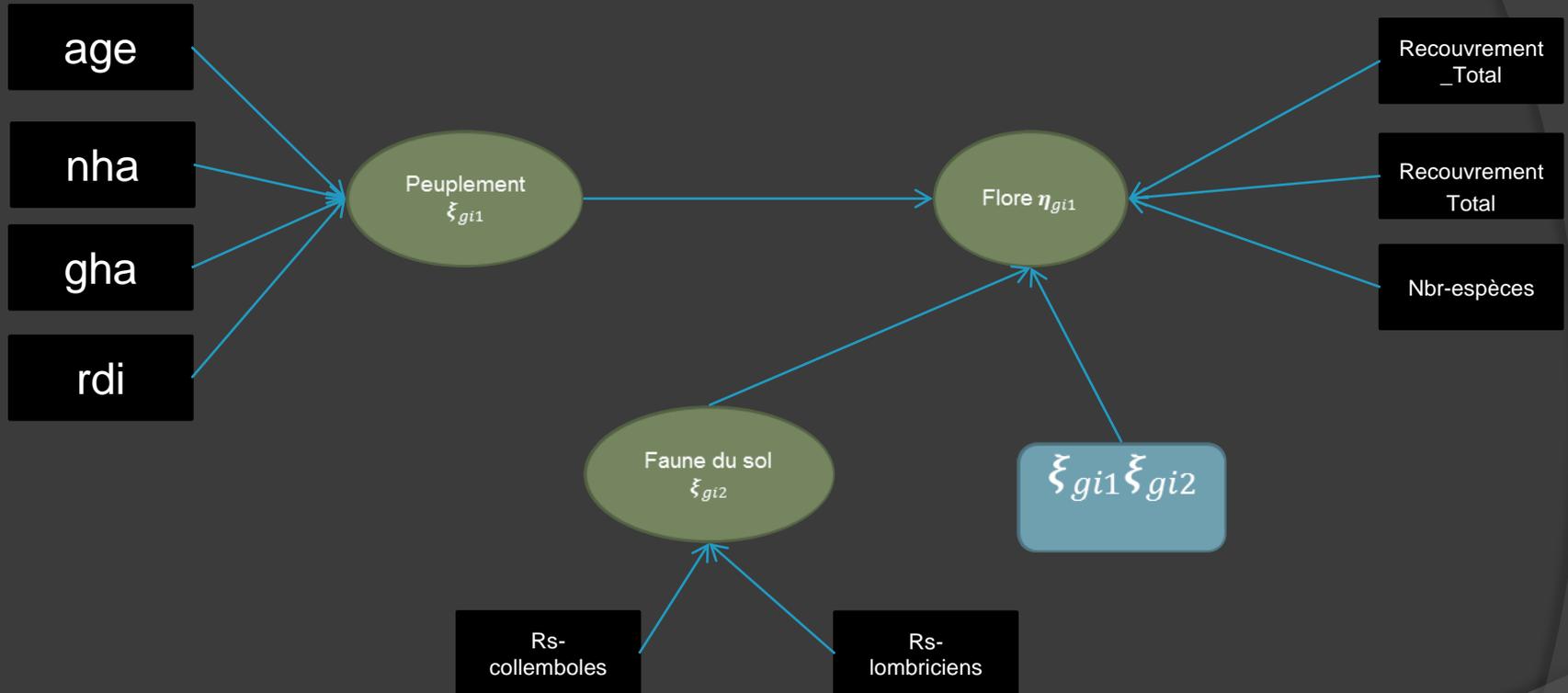
La statistique classique repose sur la loi des observations, la statistique bayésienne repose sur la loi *a posteriori*.

La loi *a posteriori* peut s'interpréter comme résumé (en un sens probabiliste) de l'information disponible sur θ , une fois x observé.

Ainsi les méthodes Bayésiennes semblent meilleures que leur équivalent Fréquentiste pour l'estimation et l'inférence sous un modèle donné.

EXEMPLE EN ÉCOLOGIE

Modèle à deux niveaux de structures hiérarchiques Non linéaire à effets croisés.



COMPARAISON LISREL-BAYÉSIEN

CRITERE	LISREL (Linear Structural Relationship)	SEM bayésien
Objectif	Estimation des paramètres	Estimation des paramètres
Méthodologie	Basée sur covariance	Loi a posteriori
Variables latentes (VL)	Combinaison linéaire de toutes les VM	Variables endogènes et exogènes
Relations entre les VM et les VL	Type réflectif	Tout types
Optimalité	Pour la précision des paramètres	Idem
Qualité des sous-modèles	VL sont estimées dans un espace non restreint	Bonne estimation des VL (s)

SUITE

CRITERE	LISREL (Linear Structural Relationship)	SEM bayésien
Hypothèses	Multinormalité des données + unidimensionnalité	TOUT
Complexité du modèle	Réduite ou modéré (<100VM)	Pas de contraintes
Taille échantillon	200 – 800 cas	50 suffit
Données manquantes	Prétraitement	Prétraitement
Identification	Dépend du modèle : au moins 3 VM par VL pour être correctement identifiée	Au moins 3
Consistance (cohérence)	Consistance des estimations	Consistance des estimations

SEM SOUS L'APPROCHE BAYESIENNE

SEM est composé d'une équation de mesure (1) et structurelle (2)

En paramétrant mes équations (1) et (2) et en combinant mes données → algorithme de Gibbs ou M-H -- simulation de v.a i.i.d selon la loi a posteriori et approcher la moyenne de cette loi selon LOI DES GRANDS NOMBRES.

Avantage :

Les distributions *a posteriori* des paramètres des modèles SEM sont calculées sous les hypothèses de normalité ou non, et de conjugaison. Ces distributions permettent d'étudier différents aspects du modèle comme la variabilité des paramètres et de fonctions de paramètres et aussi de réaliser des tests d'hypothèses. L'algorithme de Gibbs proposé pour prendre en compte les contraintes d'identifiabilité converge rapidement avec de faibles autocorrélations des simulations *a posteriori*, réduisant ainsi le nombre de simulations nécessaires.