Estimation des paramètres d'un modèle de dynamique de population structuré en longueur: maximum de vraisemblance et algorithme génétique

DROUINEAU Hilaire
MAHEVAS Stéphanie
BERTIGNAC Michel

Ifremer

Introduction

- En halieutique, les modèles de dynamique des pop ont trois principaux objectifs:
  - Améliorer la connaissance sur les espèces et leurs exploitations
  - Réaliser des diagnostics de l'état des stocks et de leur exploitation
  - Prédire leur évolution sous différentes hypothèses

Introduction

- Modèles traditionnels sont:
  - Structurés en âge: population divisée en classes d'âge considérées homogènes + besoin de données structurées en âge
  - Non spatialisés: la population et l'activité de pêche sont considérés homogènement répartis dans l'espace
  - Souvent un pas de temps annuel
- Or:
  - De nombreux processus sont plus liés à la taille du poisson qu'à son âge + données disponibles en taille
  - Des concentrations saisonnières de la population (reproduction, nourricerie…) et de la pêche
  - Les processus (croissance, mortalité, pêche…) sont continus dans le temps

Introduction

- Donc:
  - Structurés la population en taille en longueur plutôt qu’en âge
  - Spatialisé le modèle

- Mais:
  - Nécessité d’évaluer l’impact de la discrétisation des longueurs et du temps
  - Trouver des hypothèses rendant le modèle plus robuste

Introduction

- Modèle matriciel décrit les sous-processus guidant la dynamique de la population et de son exploitation
- Mais de nombreux paramètres inconnus
  - Doivent être estimés
  - Ajustement du modèle par maximum de vraisemblance sur des données observées (captures professionnelles, données de campagnes…)
  - Optimisation d’une fonction multi-dimensionnelle =>vers des algorithmes génétiques

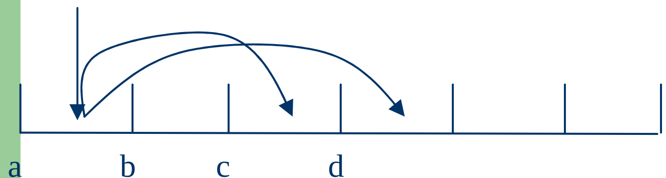
Impact de la discrétisation du temps et de la longueur

Un modèle matriciel de croissance

- N(t+1)=G(t)N(t)
  - G(t): matrice des probabilités de transition
  - La croissance moyenne suit une loi de Von Bertalanffy:

    Incrément Δl=(Linf-L(t))(1-exp(-K. Δt))
  - Les incréments de croissance suivent une loi de distribution autour de cette moyenne

- Quelles hypothèses?
  - Quel pas de temps? 1 an, 1 trimestre, 1 mois
  - Quelle taille de classe? 5cm, 1 cm
  - Position des individus dans les classes?



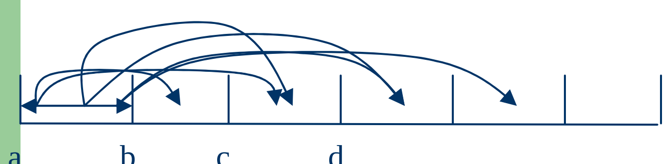
$$P_{1\to 3} = \int_{c-\frac{a+b}{2}}^{d-\frac{a+b}{2}} f(x)\,dx \quad avec\ E(X) = \left(L_\infty - \frac{a+b}{2}\right)\left[1-\exp\left(-K\cdot\Delta t\right)\right]$$

Un modèle matriciel de croissance

- N(t+1)=G(t)N(t)
  - G(t): matrice des probabilités de transition
  - La croissance moyenne suit une loi de Von Bertalanffy:

    Incrément Δl=(Linf-L(t))(1-exp(-K. Δt))
  - Les incréments de croissance suivent une loi de distribution autour de cette moyenne

- Quelles hypothèses?
  - Quel pas de temps? 1 an, 1 trimestre, 1 mois
  - Quelle taille de classe? 5cm, 1 cm
  - Position des individus dans les classes?



$$P_{1\to 3}=\int_a^b \int_{c-y}^{d-y} f(x)\,dx\,dy \quad avec\ E(X)=\left(L_\infty - y\right)\left[1-\exp\left(-K\cdot\Delta t\right)\right]$$

Un modèle matriciel de croissance

- $N(t+1)=G(t)N(t)$
  - G(t): matrice des probabilités de transition
  - La croissance moyenne suit une loi de Von Bertalanffy:

    Incrément $\Delta l=(Linf-L(t))(1-\exp(-K.\ \Delta t))$
  - Les incréments de croissance suivent une loi de distribution autour de cette moyenne

- Quelles hypothèses?
  - Quel pas de temps? 1 an, 1 trimestre, 1 mois
  - Quelle taille de classe? 5cm, 1 cm
  - Position des individus dans les classes?
  - Quelle loi de distribution des incréments? Normale, gamma, lognormale

Un modèle matriciel de croissance

- N(t+1)=G(t)N(t)
  - G(t): matrice des probabilités de transition
  - La croissance moyenne suit une loi de Von Bertalanffy:

    Incrément Δl=(Linf-L(t))(1-exp(-K. Δt))
  - Les incréments de croissance suivent une loi de distribution autour de cette moyenne

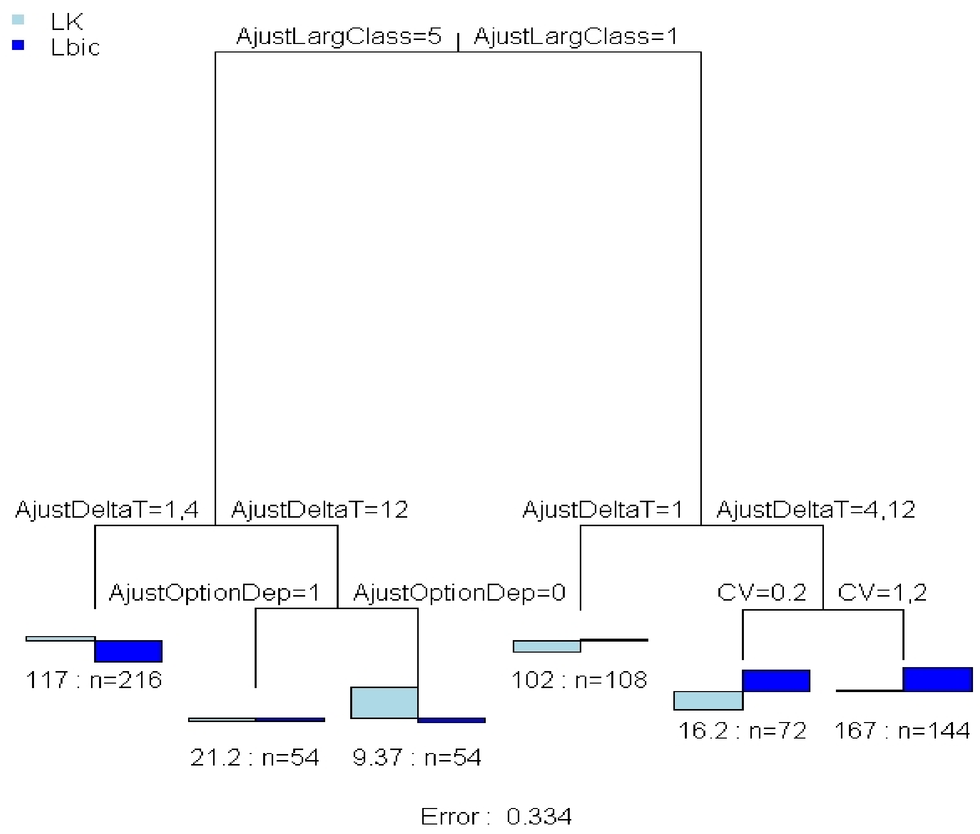
- Quelles hypothèses?
  - Quel pas de temps? 1 an, 1 trimestre, 1 mois
  - Quelle taille de classe? 5cm, 1 cm
  - Position des individus dans les classes?
  - Quelle loi de distribution des incréments? Normale, gamma, lognormale
  - Relation entre moyenne et variance des incréments? V=CE, V=CE^2

Ajusté sur des données quasi-continues

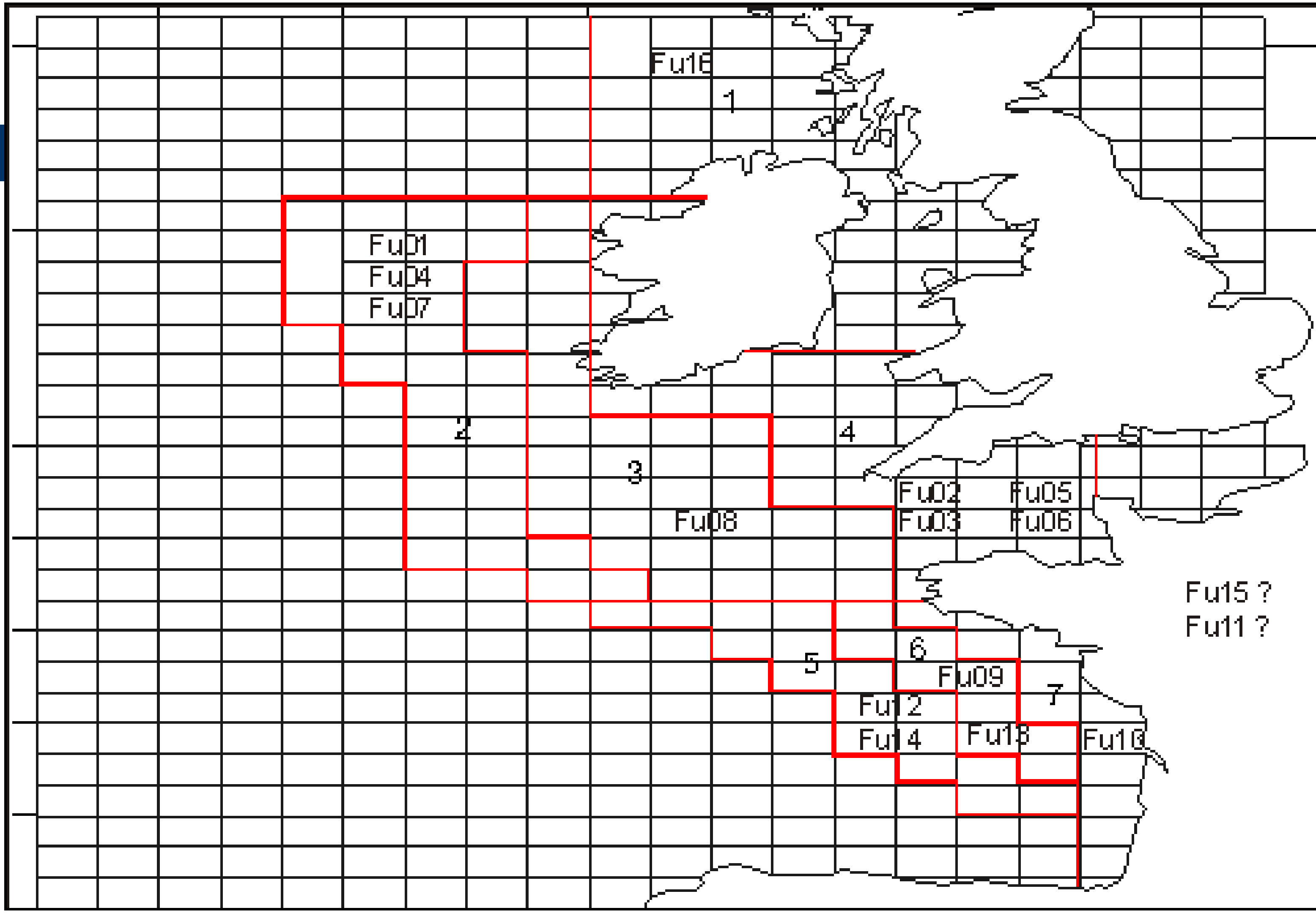
- Pour quantifier l'impact des différentes hypothèses:
  - On simule la croissance de 10000 individus avec un pas de temps mensuel => données quasi-continues
  - Hypothèses sur la loi des incréments, la variance, et le taux de croissance
- On ajuste le modèle matriciel à ces données
  - On estime K et la variance des incréments
  - Maximum de vraisemblance des fréquences « observées » et fréquences estimées
- Analyse statistique (modèles linéaires et arbre de régressions multiples) pour expliquer par les différentes hypothèses:
  - L'écart entre fréquences observées et estimées: BIC
  - L'écart entre taux de croissance estimé et observé: | Kobs-Kest |

Résultats

- Pas de … ueur améliо… ıeurs
- L'estim… mplexe
  - un p… ır de class… ance
  - Des … tribution des … unifo…



Ajustement du modèle aux données observées


Fu16
1
Fu01
Fu04
Fu07
2
3
4
Fu08
Fu02
Fu03
Fu05
Fu06
Fu15 ?
Fu11 ?
6
5
Fu09
7
Fu12
Fu14
Fu13
Fu10

Présentation du modèle matriciel

- N(t+1)=Sr(t)[Mig(t)G(t) (R(t)+ N(t))]

effectifs survie migration croissance recrutement

- Quatre processus biologiques:
  - Recrutement: un recrutement moyen et des effets aléatoires

$$N_1(t) = N(t) + \overline{R}\exp\left(\varepsilon_{recru} + \frac{\sigma_{recru}^{\ 2}}{2}\right) \quad with \quad \varepsilon_{recru} \sim N(0,\sigma)$$

Présentation du modèle matriciel

- N(t+1)=Sr(t)[Mig(t)G(t) (R(t)+ N(t))]

effectifs survie migration croissance recrutement

- Quatre processus biologiques:
  - Recrutement: un recrutement moyen et des effets aléatoires
  - Croissance: (cf avant) on doit estimer un taux de croissance et une variance

$$N_2(classe, zone) = \sum_c p_{c \to classe} N_1(c, zone)$$

Présentation du modèle matriciel

- $N(t+1)=Sr(t)[Mig(t)G(t)\ (R(t)+N(t))]$

effectifs survie migration croissance recrutement

- Quatre processus biologiques:
  - Recrutement: un recrutement moyen et des effets aléatoires
  - Croissance: (cf avant) on doit estimer un taux de croissance et une variance
  - Migration: proportion d'individus matures et immatures changeant de zones par saison

$$N(t+\varepsilon,classe,zone)=\sum_{z}\left[mat_{class}N_2(classe,z)mig_{mature}(z,zone)+(1-mat_{classe})N_2(class,z)mig_{immature}(z,zone)\right]$$

Présentation du modèle matriciel

- $\underbrace{N(t+1)}_{\text{effectifs}}=\underbrace{Sr(t)}_{\text{survie}}[\underbrace{Mig(t)}_{\text{migration}}\underbrace{G(t)}_{\text{croissance}}\ (\underbrace{R(t)}_{\text{recrutement}}+N(t))]$

- Quatre processus biologiques:
  - Recrutement: un recrutement moyen et des effets aléatoires
  - Croissance: (cf avant) on doit estimer un taux de croissance et une variance
  - Migration: proportion d'individus matures et immatures changeant de zones par saison
  - Mortalité naturelle des matures et immatures à estimer
  - +effectifs initiaux par classes et par zones à estimer

Présentation du modèle matriciel

- Dynamique de l'activité de pêche permet de calculer la mortalité par pêche par classes, zone et pas de temps
- Repose sur des Fishery Units:
  - 1 zone de pêche
  - 1 sélectivité (fonction de la longeur): à estimer
- Composées de différentes flottilles dans chaque FU
  - Effort (intensité de pêche) par pas de temps (connu)
  - Capturabilité (proba qu'un poisson soit capturé par une unité d'effort) à estimer

$$N(t+1,classe,zone) = mat_{classe} \times N(t+\varepsilon,classe,zone) \times \exp\left[-\left(M_{mature} + F_{tot}(t,classe,zone)\right) \times \Delta t\right]$$

$$+\left(1 - mat_{classe}\right) \times N(t+\varepsilon,classe,zone) \times \mathrm{e}\left[-\left(M_{immature} + F_{tot}(t,classe,zone)\right) \times \Delta t\right]$$

Présentation du modèle matriciel

- Bilan: environ 1 millier des paramètres inconnus! (dont 7*120= 840 pour les effectifs initiaux)

- Quelles données disponibles?
  - Captures par classes flottilles, classe de longueur et pas de temps
  - Données de campagnes scientifiques

- On suppose que les observations suivent une loi de distribution de moyenne la valeur estimée par le

$$L(an, flot) = \prod_{t \in an} \prod_{c \in CLASSE} f\left(C_{obs,\, flot}(t,c)\right)$$

$$avec \;\; f \; une \; densité \;\; de \;\; probabilit \; é \; de \; moyenne \;\; C_{est,\, fleet}(t,c) \; et \; de \; variance \;\; \sigma^2_{fleet}$$

Comment optimiser la fonction?

- La fonction à maximiser est multi-dimensionnelle, et ne peut être maximiser analytiquement

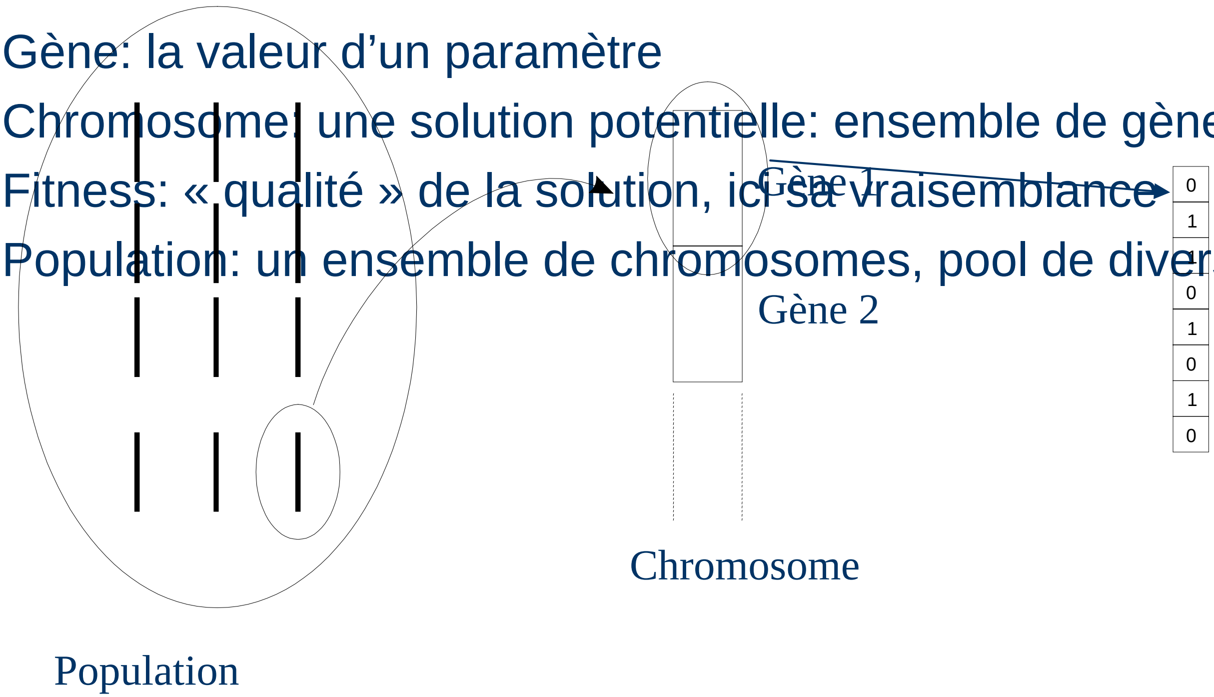
- Besoin d'une optimisation numérique

- Au vu du nombre de paramètres inconnus, il faut
  - Éviter les optima locaux
  - Converger en un temps « raisonnable vers la solution »

=>Essai d'utilisation d'algorithme génétique

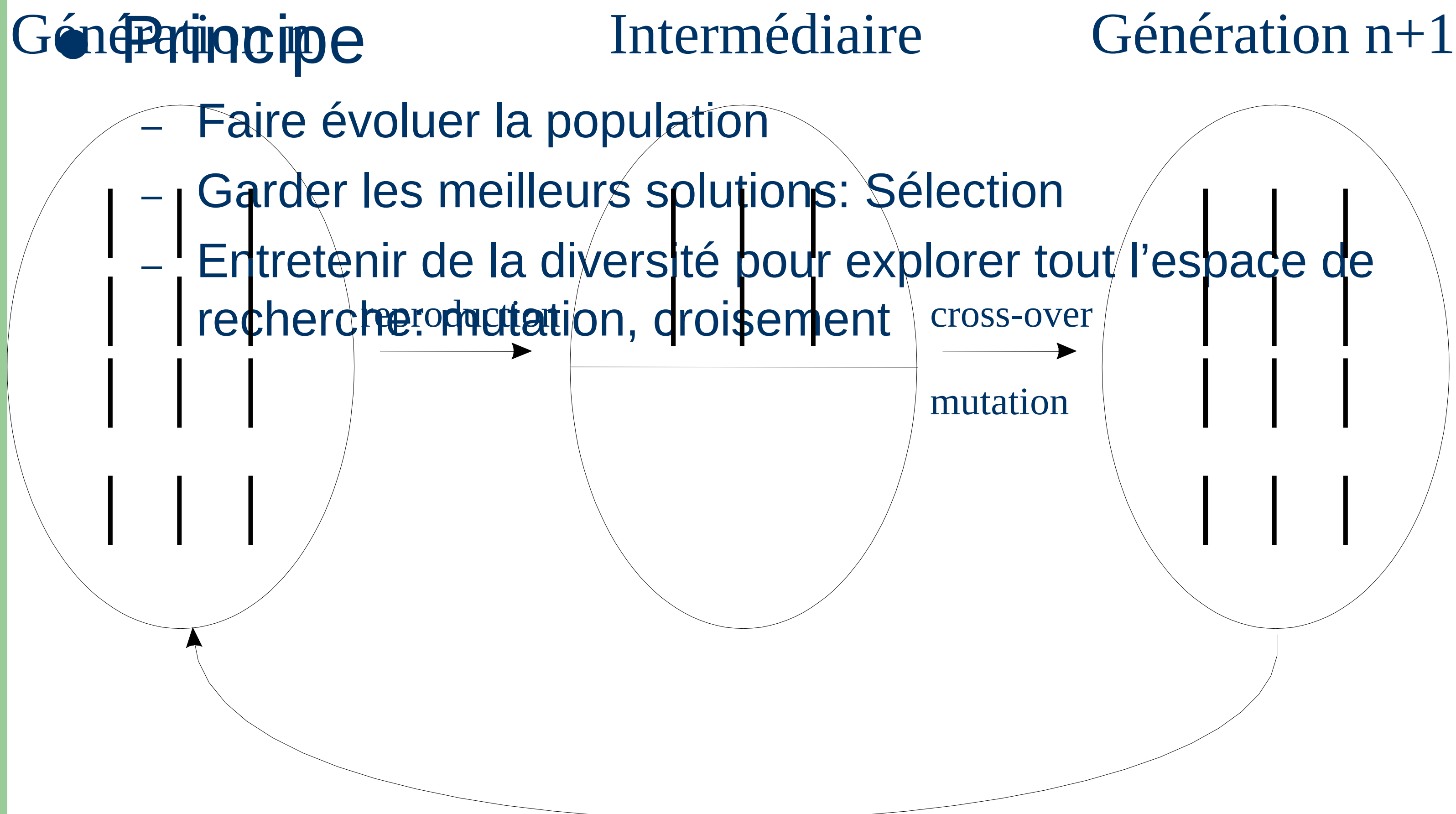
Principe de l'algorithme génétique

- Définitions:
  - Gène: la valeur d'un paramètre
  - Chromosome: une solution potentielle: ensemble de gènes
  - Fitness: « qualité » de la solution, ici sa vraisemblance
  - Population: un ensemble de chromosomes, pool de diversité



Principe de l’algorithme génétique

- Principe
  - Faire évoluer la population
  - Garder les meilleurs solutions: Sélection
  - Entretenir de la diversité pour explorer tout l’espace de recherche: mutation, croisement



Principe de l’algorithme génétique

- Sélection (de 2n à n):
  - On tire deux chromosomes de la pop de départ, le meilleur est mis dans la pop intermédiaire
- Cross-over (de n à 2n):
  - On appareille les n chromosomes 2 à 2
  - Le meilleur est copié une deuxième fois
  - On fait une combinaison linéaire des deux chromosomes
- Mutation 1 et 2 (2n):
  - Pour chaque gène, on tire un nombre aléatoire entre 0 et 1
  - Si le nombre est inférieur à une proba de mutation =>mute
  - Mutation 1: On tire une valeur entre les bornes de recherche du paramètre
  - Mutation 2: On tire une valeur dans un intervalle resserré autour de la valeur courante

Intérêts

- Bonne exploration de l’univers de recherche

$\Longrightarrow$ évite bien les optima locaux

- Pas de calcul de dérivées

$\Longrightarrow$ Facilite le codage + vitesse de calcul

- Semble donner des résultats très satisfaisants!

Conclusion

Conclusion

- Analyse de l’impact de la discrétisation a permis de choisir un modèle robuste

- Modèle complet comporte de nombreux paramètres inconnus:
  - Étudier l’identifiabilité des paramètres
  - Réduire le nombre de paramètres en ajoutant des hypothèses supplémentaires
  - Choisir un bon algorithme d’optimisation vu la complexité de la fonction de vraisemblance