Finite mixture of size projection matrix models for highly diverse rainforests in a variable environment

F. MORTIER, D.-Y. OUÉDRAOGO & N. PICARD

with S. Gourlet-Fleury and V. Freycon

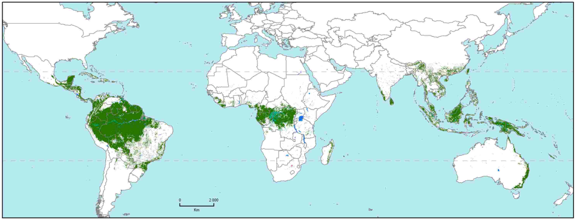
May 05-09, 2012

Outline

1. Ecological Context
2. Forest dynamics models
   - The Usher model
   - Environmental variability
   - High biodiversity
3. The M'Baïki experimental site: unique data in Central Africa
   - Mixture models outputs
   - Dynamics species groups
4. Conclusions

outline

Tropical forests I



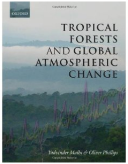
Generality and Specificity

1. 6% of the worlds surface area
2. 50% of all living organisms on Earth
3. more than 1.5b people depending on the forest

Tropical forests II

Multiple uses and actors

- local: resource (wood, NWFP)
- national: foreign exchange (timber)
- global scale: CO_2 concentration regulation through carbon sequestration
  (Millennium Ecosystem Assessment, 2005)



Tropical forests III

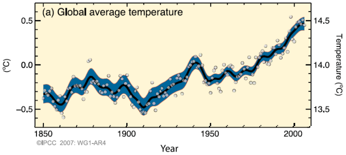
Conservation and management

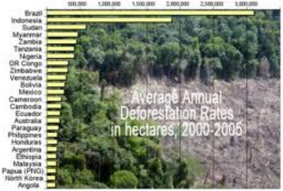
increasing atmospheric CO_2 concentration:

- increasing global temperatures
- climate change Solomon et al. (2007)

increasing land uses

- Agricultural, fuel
- ... FAO (2010)



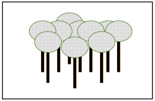


outline

1. Ecological Context
2. Forest dynamics models
   - The Usher model
   - Environmental variability
   - High biodiversity
3. The M'Baïki experimental site: unique data in Central Africa
   - Mixture models outputs
   - Dynamics species groups
4. Conclusions

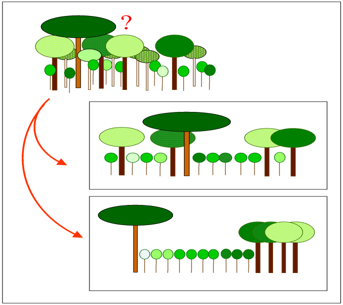
Different type of models I

Stand models: all trees are equivalents

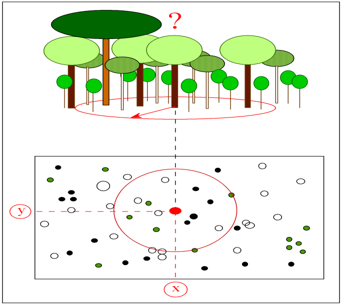


(V, G, N) = f(age, species, site, density)

Individual tree models



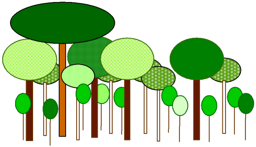
distance independant

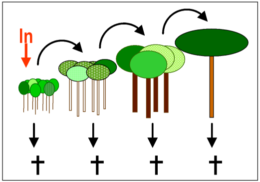


distance dependent

Different type of models II

Size projection matrix





Discrete distributions: tree equivalents inside class

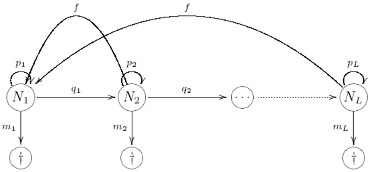
The Usher matrix I

The Usher model (Usher, 1966, 1969)

- matrix population model for size-structured populations
- describes the evolution of the population by a vector $N(t)$
- $N(t)$ vector of the number of individuals in L ordered state class at discret time t.
- This matrix population model relies on the four following hypotheses:
  - Hypothesis of independence: evolution of individuals is independent.
  - Markov hypothesis: evolution of an individual between two time steps t and $t+1$ only depends on its state at t.
  - Usher hypothesis: during each time step, an individual can stay in the same class, move up a class, or die; each individual may give birth to a number of offspring.
  - Hypothesis of stationarity: evolution of individuals between two time steps is independent of time.

The Usher matrix II

graphical representation of the Usher model



$$\mathbb{E}(\mathbf{N}_{t+1}|\mathbf{N}_t) = \mathbf{PS}\mathbb{E}\mathbf{N}_t + \mathbf{R}$$

$$\mathbf{P} = \begin{pmatrix} 1-q_2^\bullet & & 0 & 0 \\ q_2^\bullet & \ddots & & \vdots \\ & \ddots & 1-q_l^\bullet & 0 \\ 0 & & q_l^\bullet & 1 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 1-m_1 & & 0 \\ & \ddots & \\ 0 & & 1-m_l \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- $q_{i+1}^\bullet$ conditional upgrowth rate
- m_i probability to die
- r number of recruited trees

Environment and Usher model I

Temporal changes of the diameter distribution of a tree population

$$\mathbf{N}(t+1) = \mathbf{P}(X_t)\mathbf{S}(X_t)\mathbf{N}(t) + \mathbf{R}(X_t)$$

$$\mathbf{P}(X_t) = \begin{pmatrix} 1 - q_2^{\bullet}(X_t) & & 0 & 0 \\ q_2^{\bullet}(X_t) & \ddots & & \vdots \\ & \ddots & 1 - q_l^{\bullet}(X_t) & 0 \\ 0 & & q_l^{\bullet}(X_t) & 1 \end{pmatrix}$$

$$\mathbf{S}(X_t) = \begin{pmatrix} 1 - m_1(X_t) & & 0 \\ & \ddots & \\ 0 & & 1 - m_l(X_t) \end{pmatrix}$$

$$\mathbf{R}(X_t) = \begin{pmatrix} r(X_t) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- $q_{i+1}^{\bullet}(X_t)$ conditional upgrowth rate
- $m_i(X_t)$ probability to die
- $r(X_t)$ number of recruited trees

Regression estimator

(Rogers-Bennett and Rogers, 2006; Picard et al., 2008; Zetlaoui, 2006)

$$q_{i+1}^{\bullet}(X_t) = \frac{\Delta D_i(X_t)}{d_i}$$

$\Delta D_i(X_t)$ "typical" dbh growth rate
d_i width of diameter class i

Environment and Usher model II

Modeling growth, mortality and recruitment

- Let ΔD be the Diameter increment,

$$\Delta D_{sti} = \mathbf{X}_{ti}^{D}\beta_s + \varepsilon_s$$

  where β_s is the vector of unknown parameters, $\mathbf{X}^D = (\mathbf{X}_{ti}^{D})_{t,i}$ the kown incidence

- Let N_{st} be the number of recruits:

$$\begin{aligned} N_{st} &\sim \mathcal{P}\left(r_{st}\right) \\ \log(r_{st}) &= \mathbf{X}_{t}^{N}\gamma_s \end{aligned}$$

- Let M_{sti} be the mortality

$$\begin{aligned} M_{sti} &\sim \mathcal{B}er\left(m_{sti}\right) \\ \mathrm{logit}(m_{sti}) &= \mathbf{X}_{ti}^{M}\alpha_s \end{aligned}$$

Main concern with tropical forests: high specific richness

+ 300 species/ha

- many species with very few individuals
- high intra specific variability of dynamic variables
- ➥ one species one model: poor fit of models

☞ **building groups of species according to their dynamics**

Mixture of growth, mortality and recruitment processes

Mixture models

- For growth and mortality processes, let assumed:

$$\ell_n(\psi|\mathbf{Y}) = \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{i=1}^{n_{st}} \log\left[\sum_{k=1}^{K} \pi_k\, f(Y_{sti}|\mathbf{X}, \psi_k)\right]$$

  with f Gaussian density and $Y_{sti} = \Delta D_{sti}$ or the masse function associated to the Bernoulli distribution and $Y_{sti} = M_{sti}$.

- For the recruitment process, let assumed

$$\ell_n(\psi|\mathbf{Y}) = \sum_{s=1}^{S}\sum_{t=1}^{T} \log\left[\sum_{k=1}^{K} \pi_k\, f(N_{st}|\mathbf{X}, \psi_k)\right]$$

  where f is the masse function associated to the Poisson distribution.

Mixture Models and variable selection

Co-variates can differ from a group to another
stepwise selection can be computational intensive
- LASSO (**L**east **A**bsolute **S**hrinkage & **S**election **O**perator)

Selection (Khalili and Chen, 2007)

The estimator $\hat{\psi}$ of the model's parameters ψ corresponds to the maximum of a penalized version of the log-likelihood:

$$\hat{\psi} = \arg\max_{\psi} \left\{ \ell_n(\psi|\mathbf{Y}) - p_n(\psi) \right\}$$

where p_n is a penalty. We used the lasso penalization to perform variable selection in each component:

$$p_n(\psi) = \sum_{k=1}^{K} \pi_k \left(\sum_{j=1}^{p+1} \gamma_{nk} \sqrt{n} |\beta_{kj}| \right)$$

where β_{kj} is the jth element of $\boldsymbol{\beta}_k$ and γ_{nk} is a tuning parameter.

EM algorithm with LASSO

- **Estimate** $\hat{c}_{sik} = P\{$ tree of species s $\in k|$ data $(\mathcal{X})$, parameters$(\psi)\}$
  to start we choose randomly $\hat{c}_{sik}$

- **Maximise** the penalized log likelihood (Khalili and Chen, 2007)

$$\log L^*(\psi|\mathcal{X}) = \log L(\psi|\mathcal{X}) - p(\psi|\mathcal{X})$$

with $p(\psi|\mathcal{X})$ penalty term

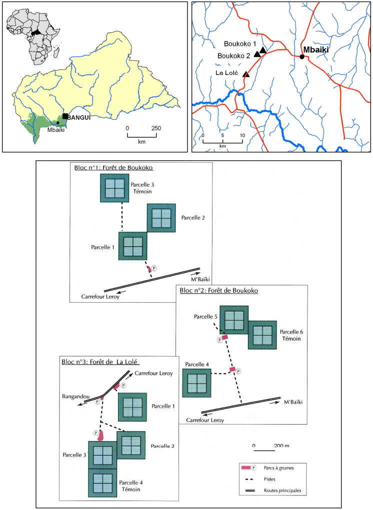
covariates selection by shrinkage : coefficient $\beta_{kp} <$ threshold $= 0$

$\rightarrow \hat{\psi}_k \rightarrow \hat{\pi}_k$

- **Classification** of **species** (not individuals) into groups
  $\rightarrow$ tree i of species s belongs to groups k for $\max(\hat{c}_{sik})$

outline

The M’Baïki experimental site (Central African Republic)



N. Fauvet CIRAD - UR 105
Septembre 2011

- Permanent sample plots (annual)
- 1982
- 40 ha
- semi-deciduous forest
- 239 species / morphospecies
- silvicultural treatments (undisturbed, logging, logging + thinning)
  $\rightarrow$ disturbance gradient

The M'Baïki experimental site (2)

Every year since 1982 (except in 1997, 1999, 2001), all trees $\geq$ 10 cm diameter at breast height (dbh)

- individually marked
- measured for dbh
- mapped
- identified

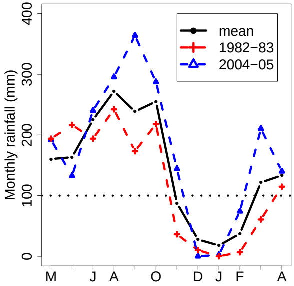
+ inventory of dead trees and newly recruited trees with dbh $\geq$ 10 cm
annual diameter growth, mortality, and recruitment

Data

- growth, mortality more than 200,000 observations
- recruitment more than 100,000 observations
- more than 200 species

Covariates (1) : Drought indices

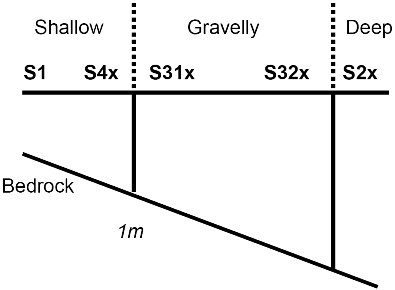
- Length of *the dry season* (**LDS**, nb months)
- Average rainfall during *the dry season* (**RDS**, mm)



Positive correlation

- Average annual soil water content (**MSW**, mm)
  simple water balanced model
  5 soil depth categories

$$SW_{t+1} = SW_t + P_t - E_t$$



Covariates (2): Light availability indices

Indirectly measured

- stand basal area (**BAst**, $m^2\ ha^{-1}$)
- stand density (**Dst**, number of stems ha^{-1})

Positive correlation

Covariates (3): Tree development stage

- tree diameter (**Di**, cm)

Mixture models outputs I

Separate species groups for growth, mortality, recruitment

- Species groups **characteristics**

- Species groups **response to drought**

Coefficients

$$\begin{aligned} \Delta D_{kti} &= \mathbf{X}_{ti}^{D} \beta_k + \varepsilon_k \\ \mathrm{logit}(m_{kti}) &= \mathbf{X}_{ti}^{M} \alpha_k \\ \log(r_{kt}) &= \mathbf{X}_{t}^{N} \gamma_k \end{aligned}$$

Covariates

Length of the dry season (**LDS**)

Average rainfall during the dry season (**RDS**)

Average annual soil water content (**MSW**)

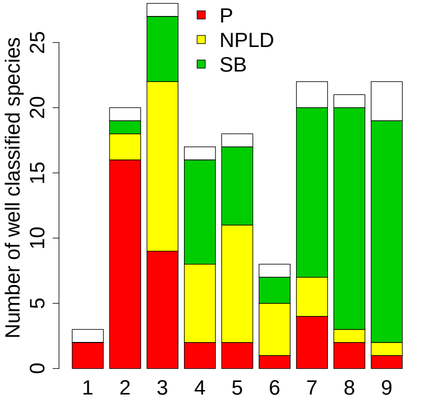
Mixture models outputs II

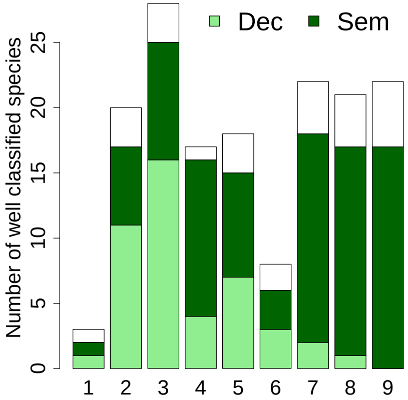
Classification results

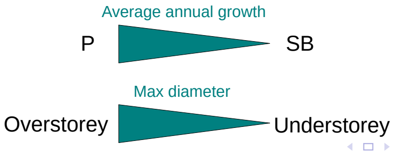
- 9 **growth species groups**
- 3 **mortality species groups**
- 5 **recruitment species groups**

54 non-empty matrix population groups

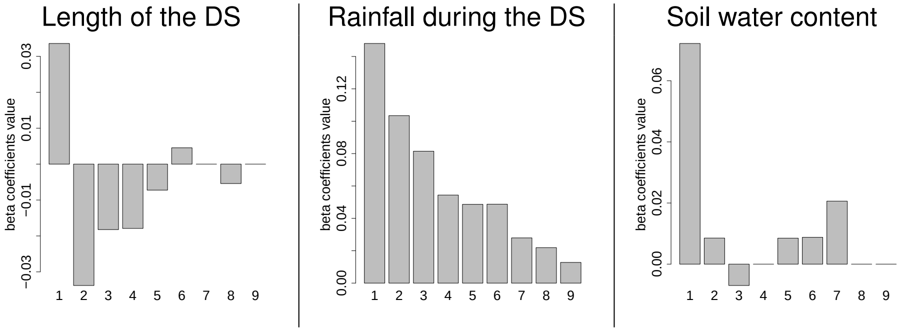
9 growth species groups

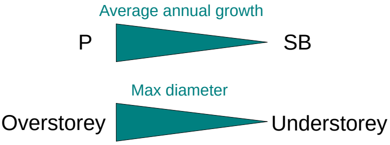






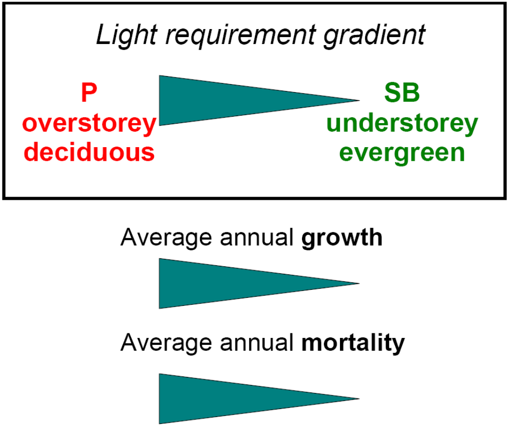
Growth species groups response to drought





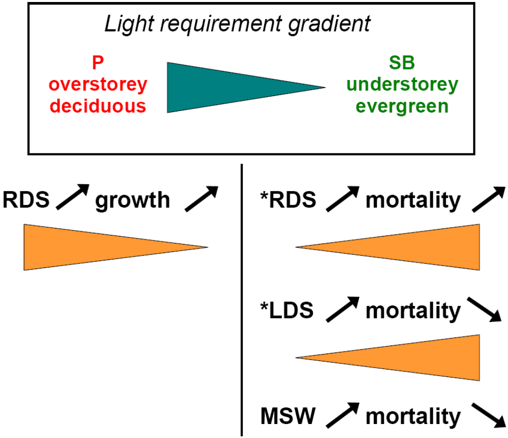
Comments I

Drought indices: capture several effects



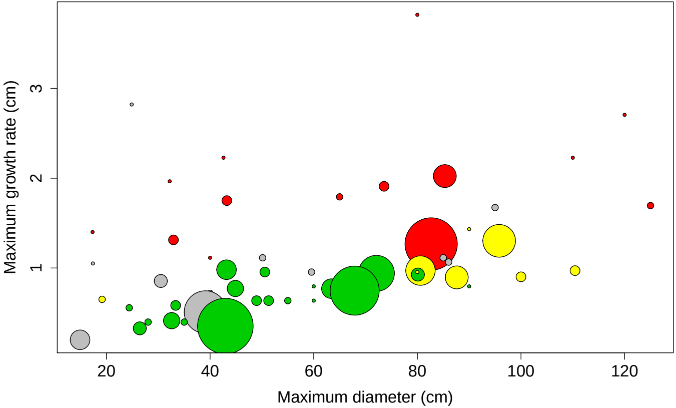
Comments II

Drought indices: capture several effects

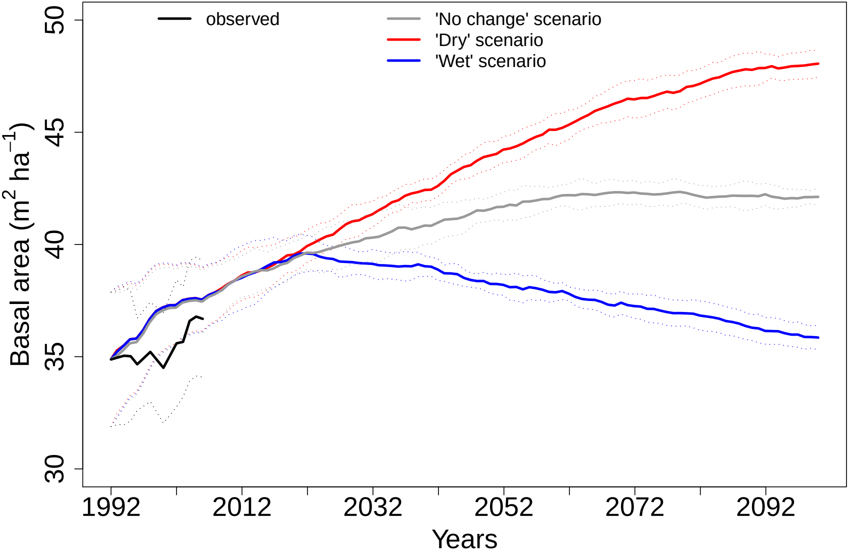


- **LDS/RDS**
  indirect measure of light availability for the understorey
  (Lingenfelder and Newbery, 2009; Newbery et al., 2011)
- **MSW**
  measure of water stress

Dynamics species groups



Dynamics response to drought: basal area



outline

Conclusions I

- **A species classification without *a priori***

  based on (growth + mortality + recruitment) response to

  light, drought (+ size)

$\Rightarrow$ leads to ecological groups of tree species

- M'Baïki **undisturbed** forest seems to be resilient to drought

+ pioneer trees more sensitive to drought
Disturbance increased the proportion of pioneer trees

Ouédraogo et al. (2011)

$\hookrightarrow$ **combination of logging + drought disturbance?**

Conclusions II

- **A powerful method for species classification**
  - species classification
  - estimation of the response to light, drought, and size
  - covariates selection

simultaneously!

- but Estimation and selection realized process by process

➳How to combine estimation, classification ans selection for the three processes simultaneously?

Bibliography

FAO. Global forest resources assessment 2010. Number 163 in Forestry Papers. United Nations Food and Agriculture Organization, Rome, 2010.

A. Khalili and J. Chen. Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, 102(479):1025–1038, 2007.

M. Lingenfelder and D. Newbery. On the detection of dynamic responses in a drought-perturbed tropical rainforest in Borneo. *Plant Ecology*, 201(1): 267–290, 2009.

Millennium Ecosystem Assessment. *Ecosystems and Human Well-being : Biodiversity Synthesis*. World Resources Institute, Washington, DC, 2005.

D. Newbery, M. Lingenfelder, K. Poltz, R. Ong, and C. Ridsdale. Growth responses of understorey trees to drought perturbation in tropical rainforest in Borneo. *Forest Ecology and Management*, 262(12):2095–2107, 2011.

D. Ouédraogo, D. Beina, N. Picard, F. Mortier, F. Baya, and S. Gourlet-Fleury. Thinning after selective logging facilitates floristic composition recovery in a tropical rain forest of central africa. *Forest Ecology and Management*, 262(12):2176–2186, December 2011.

N. Picard, F. Mortier, and P. Chagneau. Influence of estimators of the vital rates in the stock recovery rate when using matrix models for tropical rainforests. *Ecological Modelling*, 214:349–360, 2008.

L. Rogers-Bennett and D. W. Rogers. A semi-empirical growth estimation method for matrix models of endangered species. *Ecological Modelling*, 195: 237–246, 2006.

S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller. IPCC, 2007: Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change, 2007.

M. Usher. A matrix approach to the management of renewable resources, with special reference to selection forests. *Journal of Applied Ecology*, 3: 355–367, 1966.

M. Usher. A matrix model for forest management. *Journal of Biometric Society*, 25:309–315, june 1969.

M. Zetlaoui. *Aspects statistiques de la stabilité en dynamique des populations : application au modèle de Usher en foresterie*. PhD thesis, Université Paris XI, décembre 2006.