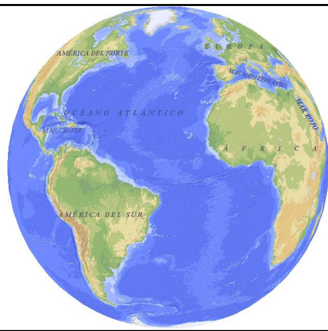
Neat Models for Scruffy Data

Oscar García
University of Northern British Columbia

Presented at
The 2000 Southern Mensurationists Conference
Jekyll Island, GA, 27-29 November 2000.



What to do with scarce and messy data?
Try harder: parsimonious models, prior knowledge.
Models for eucalypts in Spain and Chile.
Aided by growth pattern observations from New Zealand.

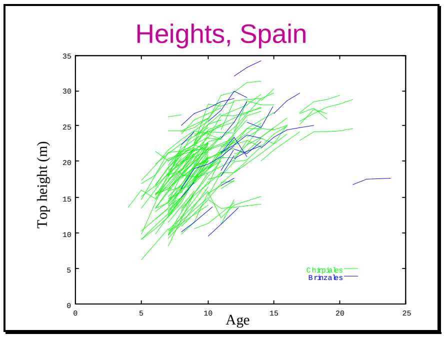
Outline

- Context
- Site index
- Growth and mortality
- Output and auxiliary relationships
- Synthesis and usage
- Concluding remarks

The context

- Eucalypt in Galicia, N.W. Spain
  - *E. globulus*, coppice, minimal tending
  - 200 m^2 (0.05 acre) plots
  - National Cellulose Company (ENCE)
- Provisional eucalypts model in Chile
  - *E. globulus*, *E. nitens*, others. Planted
  - 100 and 250 m^2 plots, from species introduction trials
- *Pinus radiata* in New Zealand
  - García, O. (1990) "*Growth of Thinned and Pruned Stands*". IUFRO symposium, FRI Bulletin No. 151

Spain: small CFI plots, not intended for growth modelling.



First step: height growth / site index sub-model.
Green: coppice, blue: seed origin.
No significant differences found by likelihood ratio test (surprise!), but seed-origin excluded from final model anyway.

Height growth / Site

Richards:

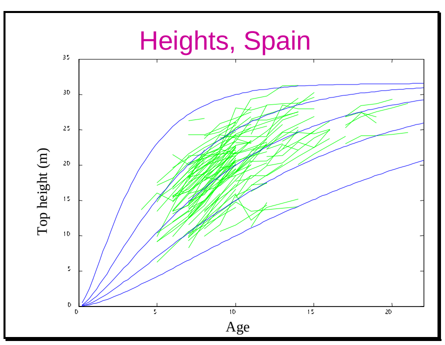
$$\frac{dH^c}{dt} = b(a^c - H^c) \text{ + random} \qquad h_i^c = H_i^c + \varepsilon_i$$

Simultaneous ML estimation

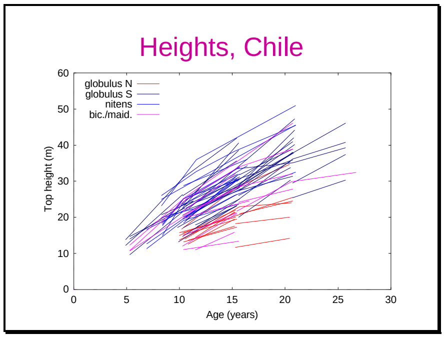
$$\frac{dH^{0.6339}}{dt} = b(31.60^{0.6339} - H^{0.6339})$$

$$H = 31.60(1 - e^{-bt})^{1/0.6339}$$

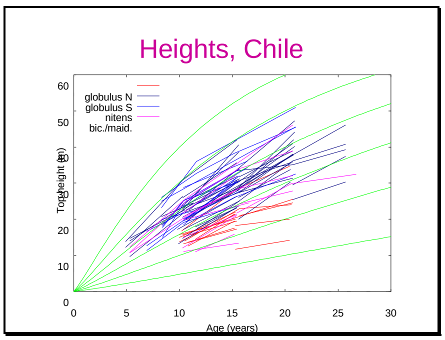
Bertalanffy-Richards happens to be a linear differential equation in a power of H.
The concept of site index implies that *one* parameter (possibly after a reparameterization) is specific to each plot ("local"), varying with site quality. The rest are common to all plots ("global").
For estimation purposes, environmental variation modelled as white noise on the right-hand-side. Measurement/sampling error also included (H=true, h=observed).
Resulting stochastic differential equation (SDE) integrated to evaluate the likelihood.
Customized optimization procedure (exploiting sparsity) maximizes over all the (hundreds of) parameters.
Easier done than said!
Best results with *b* as local.
Relationship between *b* and site index left as exercise.



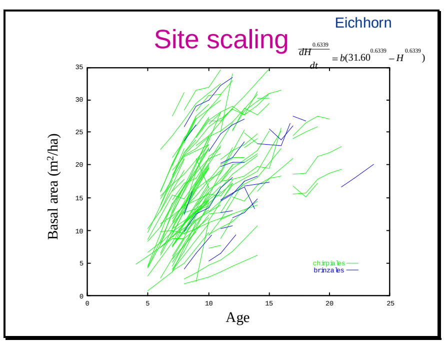
Reasonable?



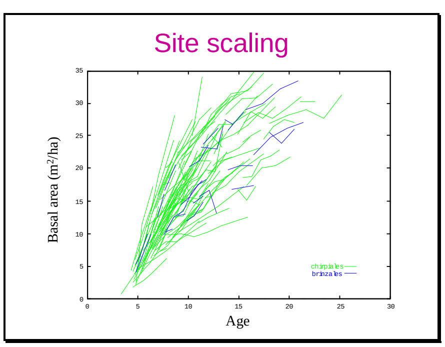
Similar (or worse), from Chile.



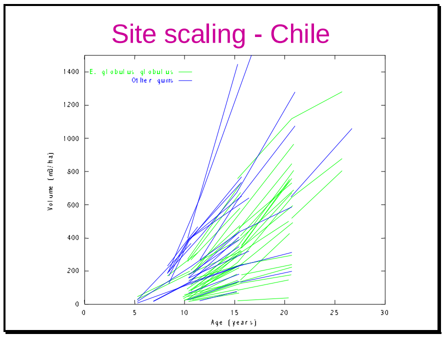
Reasonable?
Overkill, too complicated? Maybe, but try other methods with these data!



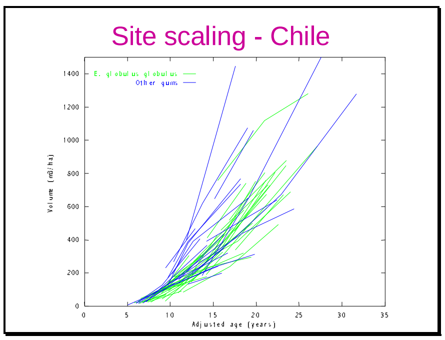
Now to the rest of the model (basal area/volume, mortality).
Part of the scatter due to site. How to take care of it?
Eichhorn’s (1904) “law” often works: trends in terms of height (instead of age) are about the same for all sites.
Height measurements are imprecise, however, and graphs against height tend to be a hopeless mess.
But note that *b* in the height model is a site-dependent time-scale factor. Adjust the ages for each plot to a common reference using the estimated *b*.



Done. “Age” from here on is “adjusted age”.
Normalized to the average site of 20.5 (base age 10).
Validation showed that this trick got rid of most or all the site-induced variation.



Again, for Chile. Before scaling.

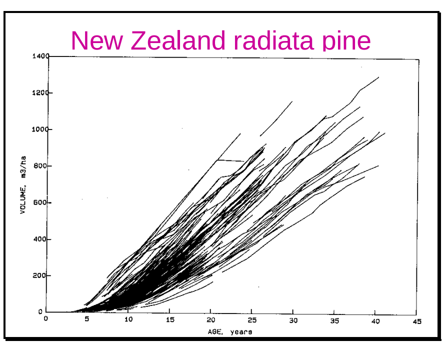


After scaling.

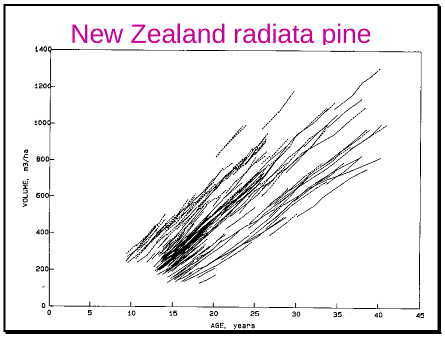
Growth

dV / dt = gross increment - mortality

We ignore site now, *t* is site-scaled.
Volume increment per hectare in two parts.
Look at gross increment first.
Not much info in our data. Let's get some guidance from another source.

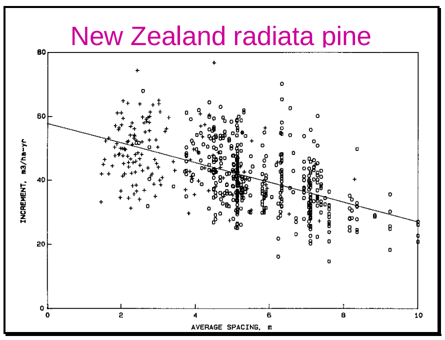


Volumes per hectare for radiata pine in Kaingaroa Forest, under a very wide range of thinning regimes (densities from 100 up to near 5000 stems per hectare). Site-adjusted ages.
After canopy closure, growth rate seems fairly constant (forget about the textbook sigmoids!)
Slopes are a little lower toward the South-East (older and/or lower density stands).
Mortality would affect the upper part.
Normal rotation ages are between 20 and 30 years.



Same, but only for “dosed” stands.

And adding estimated mortality, so this reflects gross increments (dashed trends include mortality).



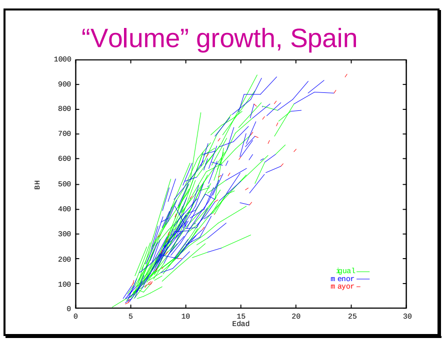
Increments calculated from the one- to three-year intervals in the previous graph.

Crosses include mortality.

Best explanatory variable was number of trees (or spacing).

After adjusting for N, no age decline!

In fact, non-declining gross increments are widely supported by long-term European thinning experiments, despite rumors and recent science folklore to the contrary .



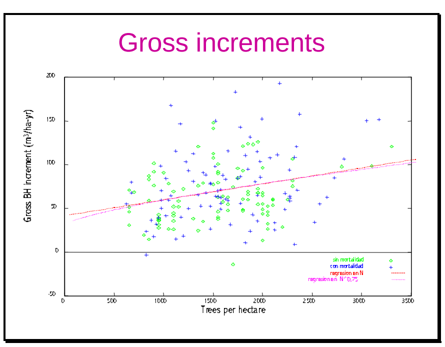
The Spanish data, basal area times top height (BH) over adjusted age.

The proxy BH used in preference to V, for convenience and to avoid dependence on particular volume tables, utilization standards, etc.

Nett values. Green: no mortality, blue: with mortality, red: increasing N (from ingrowth and/or errors).

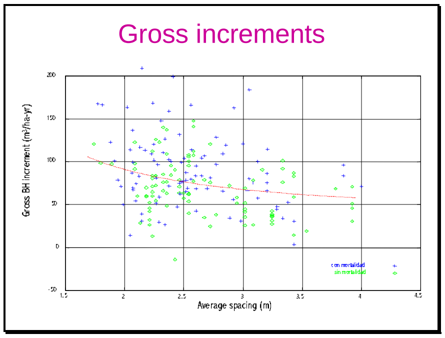
No thinning.

Can’t say much, but does not seem incompatible with the age-independent gross increment hypothesis.



Calculated gross annual increments (mortality estimated as described later).

Not much to choose between a regression linear in N, or in $N^{0.75}$



Over spacing.

Growth

dV/dt = gross increment - mortality

$dV/dt = g(N) + k\,(V/N)\;dN/dt$ $\quad k \approx 0.75$

$d\,BH/dt = a + b\,N^{\alpha} + k\,BH\;d\ln N/dt$

(after canopy closure)

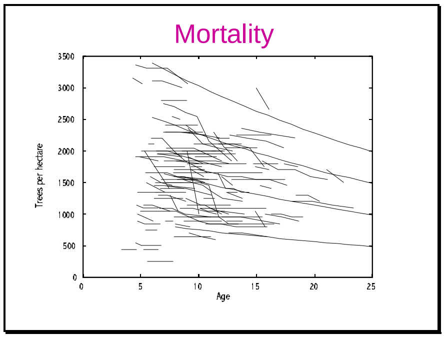
Following canopy closure, gross volume (or BH) increment assumed a function of stems per hectare (N).

Volume mortality equals mortality in number of trees times mean tree volume, times a reduction factor because dead trees are smaller than average.

Factor *k* guessed around 0.75, from literature info. Not critical.

As shown before, the exact α does not matter much. A value of 0.75 was initially chosen, thinking (wrongly) that it would simplify things, and is shown in the equations that follow. In fact, any value could be handled just as easily. The final model used $\alpha = 1$.

What about stems mortality (dN/dt)?
Typical mess.



Ignore ingrowth (or phantom trees). If real, the trees would likely to be too small to make a difference.
Assumption of a constant relative mortality of about 2.8% per year seems as good as any.
Calculated mean: d ln N / dt = -0.0281

The model

$$\frac{dH^{0.6339}}{dt} = 0.1429(31.60^{0.6339} - H^{0.6339})$$

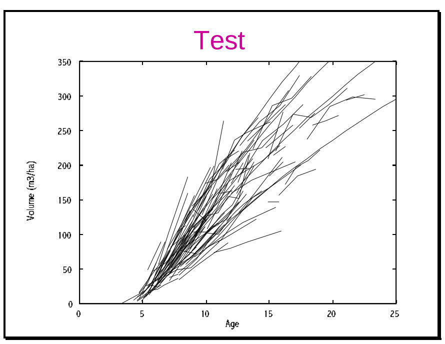
$$\frac{dBH}{dt} = 32.10 + 0.1529\,N^{0.75} - 0.02108\,BH$$

$$\frac{dN^{0.75}}{dt} = -0.02108\,N^{0.75}$$

Linear in ($H^{0.6339}$, BH, $N^{0.75}$)

Years to "canopy closure" (BH=150): $17.49 - 1.390 \ln N_0$
Thinnings: $B/B_0 = (N/N_0)^{0.75}$
Volume: $V = f(B, H, N)$

The model so far: a system of three differential equations.
Note that the third one can be written with any power of N. Using the same exponent as in the second equation results in a linear system with the indicated variable transformations.
It remains to predict growth before canopy closure. We assumed the canopy was closed when BH=150 (V approx. 50 sq.m/ha). Then obtained a regression for the time to reach that BH, depending on the initial stems per hectare.
It is convenient to have a relationship between basal area before and after thinning for when thinning is specified in terms of numbers of trees, and vice-versa.
Leap of faith: all unthinned data.
Volume is estimated from the state variables through a conventional stand volume table (regression).



Example: three predicted unthinned trajectories, with different initial densities.

System Dynamics

Closed canopy. Site index 20.5 (base age 10)

(H, B, N) (state)

$$\dot{H} = 2.012\, H^{0.3661} - 0.2254\, H$$
$$\dot{B} = 0.2043\, B + 32.10 \frac{1}{H} - 2.012 \frac{B}{H^{0.6339}} + 0.1529 \frac{N^{0.75}}{H}$$
$$\dot{N} = -0.0281\, N$$

(transition)

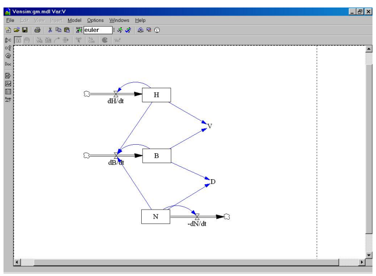
$$V = 0.7723\, B + 0.3334\, B H - 0.0004361\, H N$$

(output)

+ auxiliary functions

For clarity (?), the equations can also be written in terms of the untransformed state variables.

System Dynamics



The linear differential equations can be easily integrated analytically, giving formulas for calculating the projected state variables given any initial values and elapsed time.

For training and communication purposes, however, it may be useful to enter the equations from the previous slide into a graphical "System Dynamics" modelling system such as Stella, Dynamo, or (shown here) Vensim. The software performs numerical integration and displays results in graphical and tabular forms.

Concluding remarks

- Multiplier and closure variable, to model open stands
- Logical structuring for data-poor models
- Full ML estimation for SDE is possible
- Support for multivariate Richards
- Other mortality models?
- Need to squeeze maximum of info out of limited data

We produced a more elegant generalization that models also the dynamics of the open-canopy phase, using an additional "occupancy" state variable. Usage is a little more complex and inconvenient, though.

More elaborate parameter estimation procedures are feasible, but perhaps not worthwhile.

The equations happen to be a particular instance of the "multivariate Richards" model, previously used on purely empirical grounds (García 1979, 1994).

Challenge: more general mortality, but still analytically integrable.

Somewhat paradoxically, efficient and sophisticated methods may be more necessary in data-poor situations than with extensive high quality data.

Not everybody hates eucalypts!