Site Index, the SDE approach

Oscar García
University of Northern British Columbia

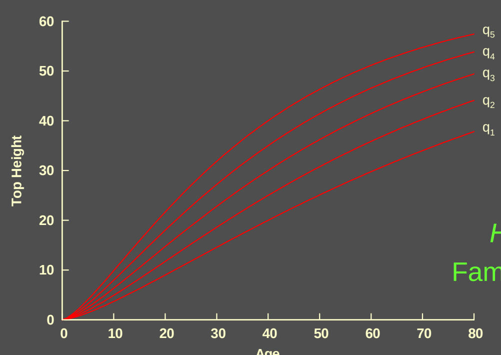
2nd International Conference on Forest Measurements and Quantitative Methods and Management
June 15-18, 2004
Hot Springs, Arkansas, USA

Outline

- Site index
  - Deterministic
  - Stochastics
  - SDE
  - Estimation approaches
- Implementation
  - SDE
  - Estimation
- Example
- Conclusions

SDE = stochastic differential equation. Fortunately, the mathematical theory of SDEs is not needed, just some of the jargon.
Focus on the fundamental site index concepts. Mostly simple, but some subtle pitfalls.

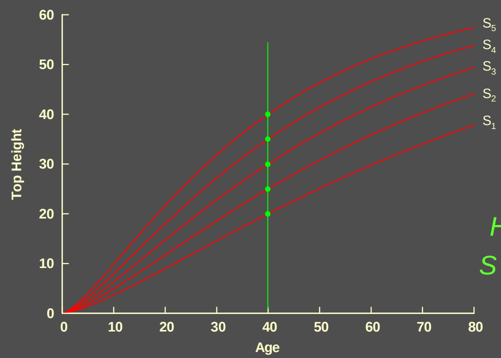
Site Index - Deterministic



H = f (t, q)
Family of curves

Basic idea: different curves for different site qualities, better quality at the top.
Infinite number of curves, labeled by parameter *q*.
Any one-to-one transformation of *q* would serve as well.

Site Index - Deterministic



$H = f\,(t,\, q)$

$S = f\,(40,\, q)$

Most common labeling scheme: *Site index* = top height at an *index* or *base age*.
Obtainable from any other q, and vice-versa.
Nothing magic about the base age.

Site Index - Deterministic

- In specific models $H = f\,(t,\, q)$ there are adjustable parameters $\boldsymbol{p} = (p_1,\, p_2,\, ...)$:

  $H = f\,(t,\, \boldsymbol{p},\, q)$

- The p_i are common to all stands or plots

  "Global"

- q is site-dependent, specific to each stand or plot

  "Local"

Global-local terminology not standard, used just as shorthand.
Sometimes will omit writing $\boldsymbol{p}$.

Site Index - Deterministic

- Examples:
  - Anamorphic Schumacher

    $H = a \exp(-b\,/\,t)$, a (=q) local, b global

    $S = a \exp(-b\,/\,40) \;\rightarrow\; a = S\,/\exp(-b\,/\,40)$

    $\rightarrow\; H = S \exp[-b\,(1\,/\,t - 1/40)]$
  - "Polymorphic" Richards

    $H = a\,[1 - \exp(-b\,t\,)\,]^c$, b local, a and c globals

    $S = a\,[1 - \exp(-40\,b)\,]^c \;\rightarrow\; b = -\ln\,[1 - (S/a)^{1/c}]\,/\,40$

    $\rightarrow\; H = a\,\{1 - [1 - (S/a)^{1/c}]^{\,t/40}\,\}^c$

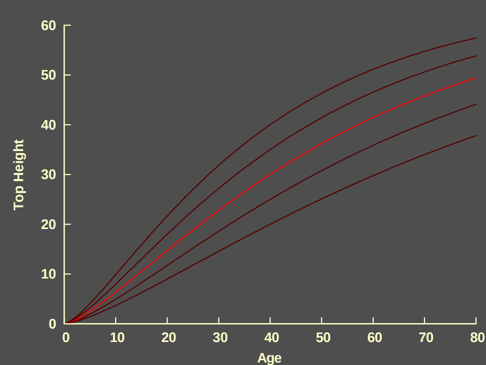
Equivalent forms with q or S.
In a site index model only the globals need to be given specific numeric values. S or other local parameter chooses a particular curve from the family.
In general, it might not be possible to solve analytically for the local.

Site Index - Deterministic

- In general, we may have $a = a(q)$, $b = b(q)$, $c = c(q)$, with these functions involving other global parameters
  (reparameterization)
- Only the global values are needed for the final model
- In these examples the equation form is base-age invariant. Not always the case
- The estimation procedure may or may not be base-age invariant

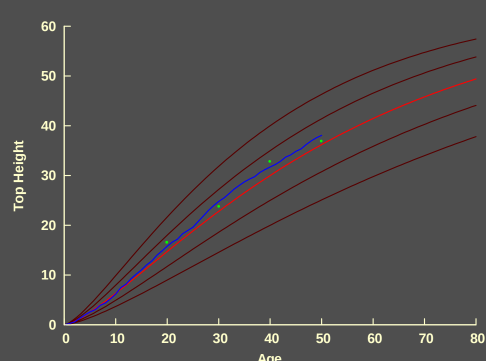
"Base-age invariant": Bailey and Clutter, *Forest Science 20:* 155-159, 1974.
Two sides/aspects: changing base age does not change the equation form, and/or does results in the same (global) parameter estimates.

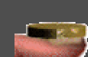
Site Index - Deterministic



Deterministic: stand supposed to follow the curve. Straightforward.

Site Index - Stochastics





Site index?

In reality, growth rate will vary because of weather, etc. (blue).
In addition, there may be measurement and/or sampling errors (green).
Confusion and controversy.
Question 1: which/what is the site index now?

Site Index - Stochastics

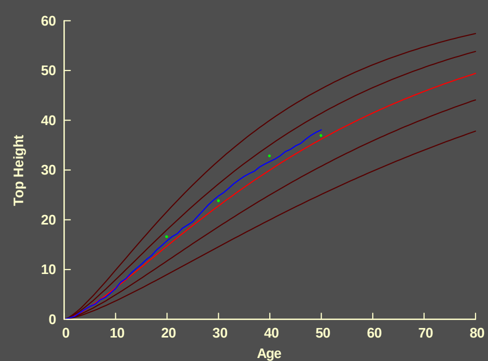
1. Stand height at age 40
   "Stand site-index"
2. "Expected" height at age 40
   "Site site-index"

- Site index: Most likely top height at a base age among all the hypothetical stands that could grow on the site.
- Analogously for the site curve and q

Two camps:
1. Literal. Point on blue curve. Index is a property of the stand!
2. More convoluted, trying to stick to the original concept (property of the site): point on the red curve.

Definitions cannot be wrong; they may be more or less natural, more or less useful.
I choose definition 2.

Site Index - Stochastics



ΔH, dH/dt

Question 2: estimation.
Rational estimation requires a stochastic model.
To understand the blue curve, think in terms of increments.
Discrete *vs.* continuous largely a matter of taste?
Sometimes in math, embedding a particular case (finite differences) into a more general class of problems (continuous) makes it easier to handle.
Again fortunately, we need little or no differential equation (DE) theory. Most are separable: move all H 's to the left-hand-side, all t 's to the right, and integrate both sides.

Site Index - SDE

- dH/dt $= g(H, t, \boldsymbol{p}, q, u(t))$
  $u(t)$ is "environmental noise" (a stochastic process)
  $H(t_0) = H_0$
- $h_i = H(t_i) + \varepsilon_i$
  (measurement / sampling error)
- Estimate $\boldsymbol{p}$
  (and perhaps t_0 or H_0)

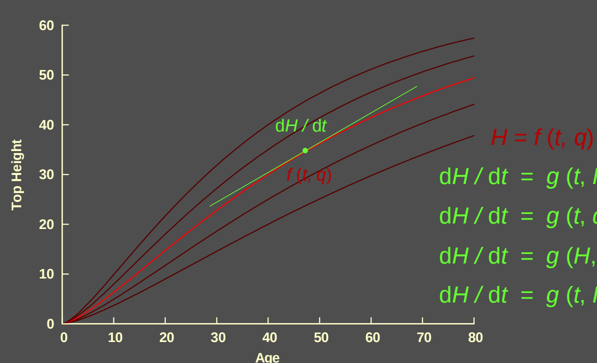
General formulation.

S.I. - Estimation approaches

- Parameter prediction
  1) All locals
  2) Keep the q_i, re-estimate globals
- Mixed effects
  - Assume q "random", with given distribution
  - Assume data is a random sample
- Difference equation (Bailey-Clutter)

Most popular methods.
PP: Two stages; fit each plot separately first.
ME: Growth data is rarely, if ever, a simple random sample from the population.
To explain the last one we digress...

S.I. - Differential equations



$dH / dt = g(t, H, q)$

$dH / dt = g(t, q)$

$dH / dt = g(H, q)$

$dH / dt = g(t, H)$

It is often incorrectly assumed that there is a unique differential equation associated with any growth function.
Actually, because we are on the curve, we can substitute for H in the DE, and get a DE that contains t but not H.
Or eliminate t between the DE and the growth function, and obtain a DE in terms of H only. There are good physical and biological arguments to prefer this form.
Or anything in between (e.g., write $H = H^k H^{1-k}$, and substitute only for the first H).
A particularly neat trick is to eliminate q between the DE and the function. This is behind the Bailey-Clutter technique.
Integrating any of these DEs generates the same growth function. I.e., they are all equivalent in the deterministic case. However, once we are off the curve, they differ.
In particular, the very convenient q-free Bailey-Clutter form would imply that growth rate is independent of site quality!

S.I. - Estimation approaches

- Parameter prediction
  1) All locals
  2) Keep the q_i, re-estimate globals
- Mixed effects
  - Assume q "random", with given distribution
  - Assume data is a random sample
- Difference equation (Bailey-Clutter)
  - Differential equation of family (q-free)
  - Integrate between successive measurements

Eliminating the local q from the problem makes estimating the globals easy.
These methods are relatively simple and convenient, and may (or may not) produce satisfactory results in practice.
They do ignore the error structure, or use a rough approximation.

Site Index - SDE

- dH/dt = $g(H, t, \mathbf{p}, q, u(t))$

  $u(t)$ is "environmental noise" (a stochastic process)

  $H(t_0) = H_0$

- $h_i = H(t_i) + \varepsilon_i$

  (measurement / sampling error)

- Estimate $\mathbf{p}$

  (and perhaps t_0 or H_0)

I tried to use a more realistic approximation to the error structure implied by the SDE above.
Biometrics 39: 1059-1072, 1983.
K.Rennolls, *For. Ecol. and Management 71:* 217-225, 1995.

Implementation - SDE

dH/dt = $b\,(a - H)$ + noise ?

- dH^c/dt = $b\,(a^c - H^c) + \sigma\, w'(t)$
  - White noise, Wiener, Brownian motion process
  - Von Bertalanffy – Richards
  - $H = a\,[1 - \exp(-bt)]^{1/c}$ + error (if $H_0 = t_0 = 0$)
- $h_i^c = H^c(t_i) + \sigma_m \varepsilon_i$

  (the ε_i are independent normal)

Simplest would be a linear (S)DE. Not too bad, but not very flexible.
Much improved by allowing a power transformation of H (c is another parameter to be estimated). Thus, the DE happens to be equivalent to the Richards DE. Check: calculate the derivative on the left-hand-side and simplify.
The essential assumption in the "noise" term (with various names) is that non-overlapping time intervals are independent.
Integration of the SDE results in a Richards, plus an error with the serial correlation and increasing variance that one might expect.
Measurement/sampling error component chosen mostly for mathematical convenience, but reasonable: for typical values $0.5 < c < 1$, this error increases less-than-proportionally with H.
Aside: Growth functions other than the Richards could be used. There is a two-parameter transformation that linearizes practically all the models used in forestry (http://web.unbc.ca/~garcia/unpub/unigrow.pdf).

Implementation - SDE

- In general, parameters a, b, c, H_0, t_0, σ, σ_m, can be either global, local, or fixed, possibly after reparameterization
- Simplest, most useful cases:
  - a local, and b, c global (anamorphic)
  - b local, and a, c global
  - Non-zero origin (include t_0 or H_0 parameter)
  - Breast-height age: H_0 = 1.3 m

Implementation - Estimation

- To be estimated: several global parameters, plus (usually) one local for each sample plot
- Actually, the local(s) and the variances are not used in the final model (“nuisance parameters”), but still need to be estimated
- Choose the parameter values that maximize the probability of observing our data (likelihood function)

Implementation - Estimation

- Expression for the likelihood can be obtained from the SDE and measurement error model
- Maximization with a customized algorithm that takes advantage of special structure
- Invariance of ML estimators: maximum does not change when re-defining what needs to be estimated (e.g., height, site). Just plug-in the parameter estimates.

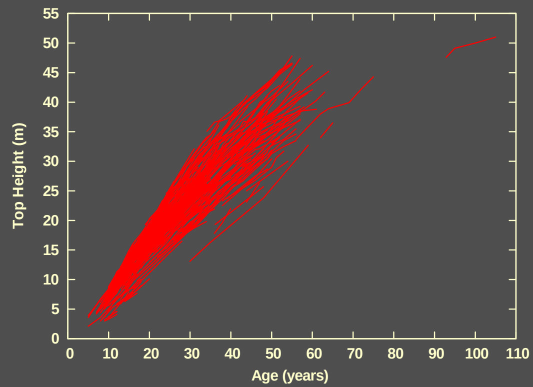
Invariance does away with the question of using the same or different models for estimating either height or site index.

Example

- Software: www.unbc.ca/forestry/forestgrowth/sde
- GUI front-end added recently
- Data: Douglas fir plantations. New Zealand, South Island (1990)
  - 288 plots, 1421 measurements

Original programs not very user-friendly, developed for in-house use. Only a few brave souls tried them elsewhere. GUI and other changes now makes it easy.

Example



Data from 1990 Douglas fir growth model (unpublished), used by permission from Forest Research.

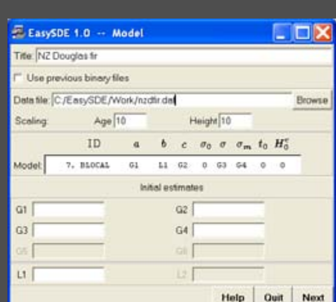
Note 105 years-old stand planted in 1865, last measured in 1970!

Example

Plot ID	Age	Height
37000100	26.87	17.1
37000100	31	19.8
37000100	36.94	26.1
37000100	40	27.9
37000100	43	30.4
37000100	46	32
37000100	47	32.1
37000100	51.94	34.1
37000200	31	18.7
37000200	36.94	24
37000200	40	26.1
37000200	47	30.7
37000300	31	19.1
37000300	36.94	24.9
37000300	40	[illegible]

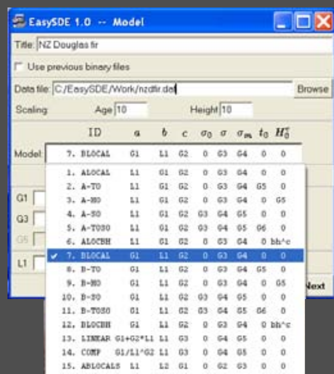
Input data file. Plots separated by blank lines.

Example



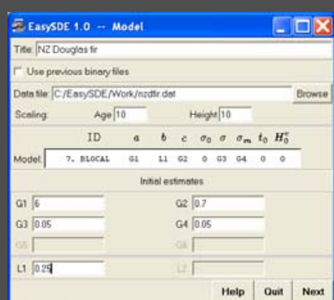
Scaling: Age and height divided by 10 to improve numerical conditioning and convergence (optional).

Example



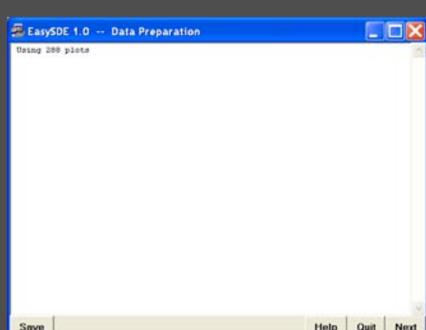
Most useful model variants are built-in. Unusual formulations would require re-compilation.

Example



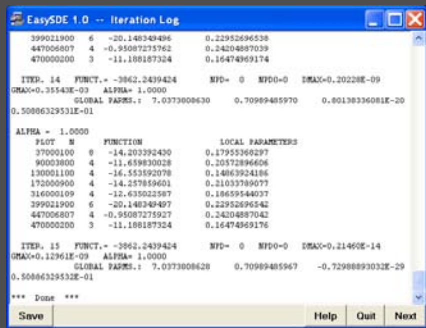
Initial parameter estimates.

Example



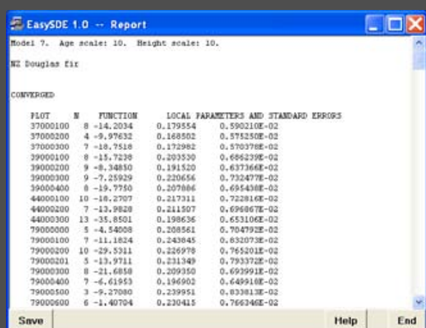
Data check and pre-processing: OK.

Example



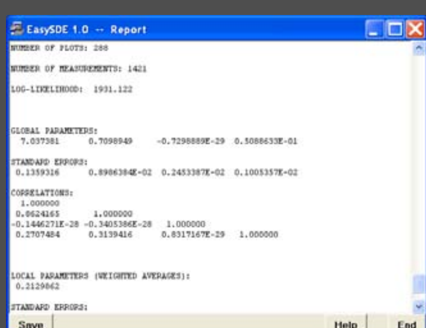
Iteration log. Converged in 15 iterations.
Mostly esoteric/useless info. Something to look at in the old days when this was an over-night run. Now it flashes through the screen in a couple of seconds!

Example



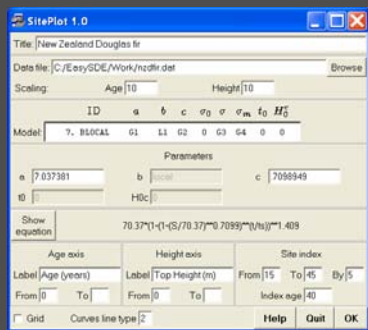
Output.
Locals may be used to estimate plot site indices for some modelling applications.
Contribution to the likelihood (“FUNCTION”) can flag outliers.

Example



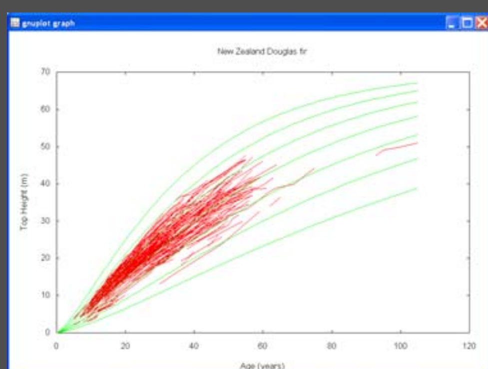
The globals (except sigmas) are what goes into the model.
Log-likelihood is used for model comparisons and hypothesis testing.

Example

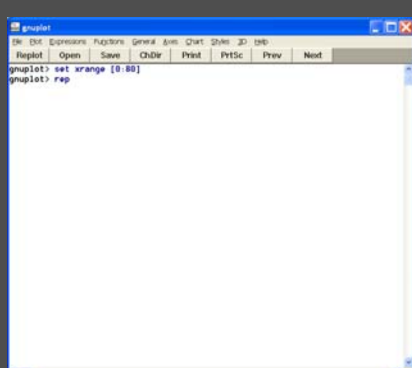


Graphing.

Example

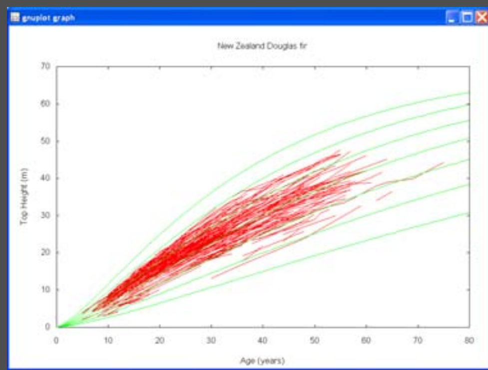


Example

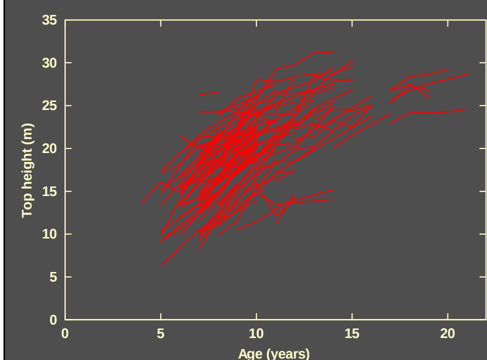


Gnuplot (www.gnuplot.info) used for graphing.
Flexible, many options.
E.g., change age range to 0-80 years.

Example



Example

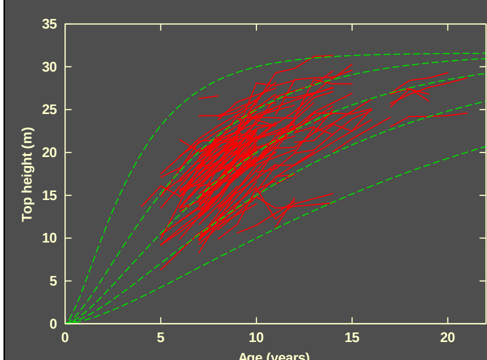


Eucalypt in Spain

Efficient estimation may be more important with lower quality data.

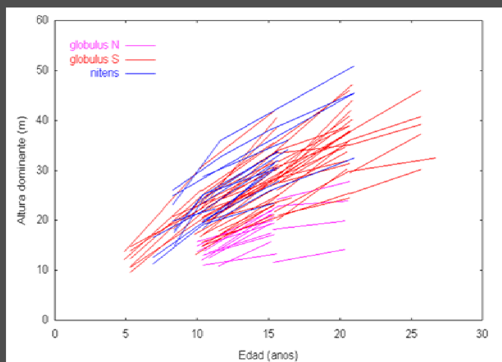
Small plots from continuous forest inventory, highly variable.

Example



For.Ecol.Man. 173: 49-62, 2003

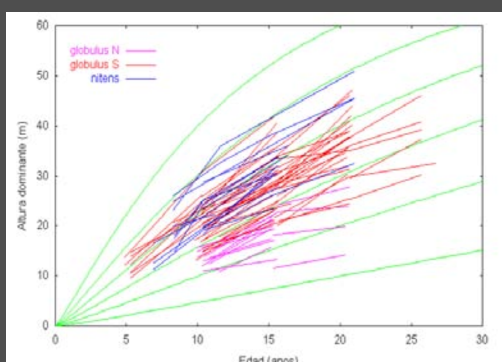
Example



Eucalypts in Chile

100 m^2 and some 225 m^2 plots from species introduction trials, mostly with just one remeasurement.

Example



Ciencia e Investigación Forestal 9: 5-21, 1995.

Conclusions

- Reasonable methodology and results
- Foundations
  - Deterministic aspects
    - Family of curves $H = f(t, q)$ (or 3-D surface)
    - Global and local parameters: $H = f(t, \mathbf{p}, q)$
  - Stochastics
    - Stand SI *vs* site SI
    - SDE + error: $dH/dt = g(H, \mathbf{p}, q, u(t))$, $h_i = H(t_i) + \varepsilon_i$

Most methods may give acceptable results, at least with good data. More important are firm foundations.

Note: H = f(t, q) is sometimes thought of as a 3-D surface instead of as a family of curves. OK, but it requires more mental gymnastics, and we do not really gain anything.