Some guidance on using mathematical notation in biology

Andrew Edwards[1] & Marie Auger-Méthé[2]

[1]Pacific Biological Station & University of Victoria; [2]University of British Columbia

UVic Biology Department Seminar

Friday 1st November, 2019

Fisheries and Oceans Canada Pêches et Océans Canada

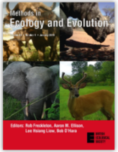
Paper

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Some guidance on using mathematical notation in ecology

Andrew M. Edwards[1,2] | Marie Auger-Méthé[3,4]



Volume 10, Issue 1
January 2019
Pages 92-99

Code for building this talk: https://github.com/andrew-edwards/notation-talk

Feedback

Brian J. Enquist
@bjenquist

Thank goodness someone published this. Has been on my to wish list to do something similar. 'Some guidance on using mathematical notation in ecology'- "the clarity of mathematical notation varies between papers and can often be inconsistent" https://t.co/f

15 Oct 2018

Ryan S. Miller
@RyanSMiller_

As someone that became quantitative later in their career I wish someone had published this long ago! https://t.co/wsrHtBE9v8

30 Oct 2018

Kiva Oken
@KivaOken

Bad notation impedes communication of your ideas. Sloppy notation is at *least* as annoying as poor writing. Every ecologist who uses equations should read this! https://t.co/rZZ5BFKc3P

03 Apr 2019

Feedback

Dr Gavin Simpson
@ucfagls

For someone like me, without a strong math or stats background, reading, interpreting, and writing equations in my own work is often a dark art. I could have done with a guide like this years ago!

15 Oct 2018

Dr. Vanessa Tobias
@marshprincess

I thought this was helpful for reading papers with lots of equations as well as writing them. Here's the link to the pdf: https://t.co/6Bn2lMnPUh

30 Apr 2019

Ian Jonsen
@IanJonsen

Well done @AndrewM_Edwards and @AugerMethe this has been needed for quite some time! 👏👏👏

25 Mar 2019

Motivation

We have both reviewed manuscripts in which unclear notation:

- made it impossible to understand modelling details
- meaning we could not evaluate the results or conclusions
- so could not properly review the work

Motivation

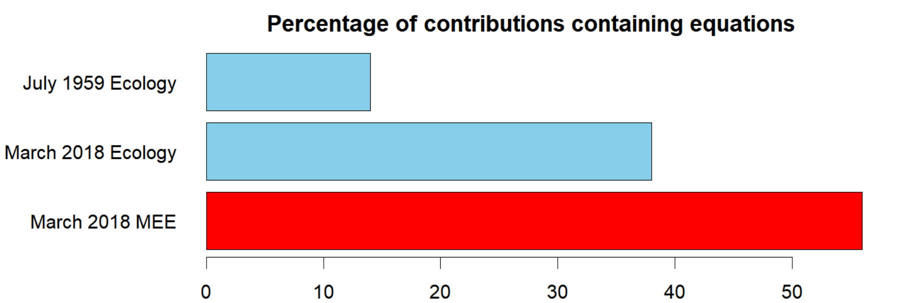
- mathematical modelling playing an increasing role in biology
- requires communication of details of models
- this includes notation
- clarity of notation varies between papers

Motivation

- mathematical modelling playing an increasing role in biology
- requires communication of details of models
- this includes notation
- clarity of notation varies between papers
- poor notation can:
  - make models appear more complicated than necessary
  - this leads to cluttered equations that impede understanding
  - impedes communication of ideas
  - can prevent work being properly reviewed – analogous to an incomplete description of the methods of a laboratory experiment

Motivation

Lowry (1959): increase "use of mathematical notation where appropriate in ecological literature"



Lowry said advantages include:

- precise communication of logical thought (beyond that afforded by the written word)
- reproducibility of methods and results

and they "accrue in proportion to the care used by the author in the use of notation".

Motivation

Many skills involved in producing ecological modelling paper, with books covering:

- introductory ecology
- general mathematics
- ecological modelling
- implementing models in R
- writing and publishing a paper

Recommendations exist for:

- streamlining workflows (tools such as Git and GitHub)
- making computer code available and reproducible

But no tips on developing the mathematical notation to use in a model.

Motivation

Improvements have impacts – for Individual Based Models:

- previously criticised as generally being so poorly documented that they could not be evaluated or reproduced
- motivated development of standardised protocols
- led to a more rigorous formulation of models
- enhanced understanding

Motivation

Why us?

Similar backgrounds:

- all research highly quantitative
- lead authors of >20 ecological modelling papers
- taught modelling courses to biology students

Differences:

- one undergrad in math
- one undergrad in biology

We have papers with notation that could, in hindsight, be improved.

What is notation?

Notation – we are specifically referring to:

- letters used to represent quantities in equations
- use of subscripts and superscripts
- parentheses etc.
- related concepts

Letters can be

- Roman – $a, b, c, \dots$
- Greek – $\alpha, \beta, \gamma, \dots$
- lower case – $a, b, c, \alpha, \beta, \gamma, \dots$
- upper case – $A, B, C, \Gamma, \Delta, \Omega, \dots$

What do journals say?

We reviewed the Author Guidelines for a sample of 14 journals:

- *Bulletin of Mathematical Biology*
- *Canadian Journal of Fisheries and Aquatic Sciences*
- *Ecology*
- *Ecology Letters*
- *Evolution**
- *Functional Ecology*
- *Journal of Animal Ecology*
- *Journal of the Royal Society Interface**
- *Marine Ecology Progress Series*
- *Methods in Ecology and Evolution*
- *Molecular Biology and Evolution*
- *Nature*
- *PLOS ONE*
- *Science*

*Only two that have no mention of equations.
Minimal guidance in others – almost exclusively restricted to typesetting aspects (such as bold for vectors).

Authors' responsibility

Onus for good notation should be on authors (rather than journals) because:

- notation should be decided early on in a study (possibly before deciding on a particular journal)
- ideally computer code will match the notation
- work may first appear in a technical document such as a thesis before being submitted to a journal
- Supporting Information of a paper often contains the full mathematical details of models and is not typeset or edited by publishers – authors are responsible for clarity

Guidelines

- based on those traditionally used in mathematics
- are in a somewhat logical order
- our aim is for them to be useful and adopted
- though we anticipate exceptions for which they are purposefully overlooked
- use examples from common ecological models, our own fields of research and from evolutionary biology

1. Define all terms

Who recognises this equation?

$$E = mc^2$$

1. Define all terms

Who recognises this equation?

$$E = mc^2$$

Who knows what it means?

1. Define all terms

Who recognises this equation?

$$E = mc^2$$

Who knows what it means?

It does not convey any information, since the letters are not defined.

1. Define all terms

Common ecological example:

$$S = cA^z$$

1. Define all terms

Common ecological example:

$$S = cA^z$$

where:

- S represents the number of species (of a particular taxonomic category) in area A
- constant c is the number of species that would be in one square unit
- dimensionless exponent z quantifies the change in species number with area

1. Define all terms

Common ecological example:

$$S = cA^z$$

where:

- S represents the number of species (of a particular taxonomic category) in area A
- constant c is the number of species that would be in one square unit
- dimensionless exponent z quantifies the change in species number with area

So equation represents an increase in species richness with area.

1. Define all terms

Common ecological example:

$$S = cA^z$$

where:

- S represents the number of species (of a particular taxonomic category) in area A
- constant c is the number of species that would be in one square unit
- dimensionless exponent z quantifies the change in species number with area

So equation represents an increase in species richness with area.

Immediately after an equation (as part of the same sentence) any previously undefined symbols should be defined using the phrase "where ...".

1. Define all terms

The number one guideline is to define every term that is used in an equation.

1. Define all terms

The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.

1. Define all terms

The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.

1. Define all terms

The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.

1. Define all terms

The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.

Maybe introduce a second time if appears again much later in a lengthy work.

Table of notation helpful for complicated models.

1. Define all terms

The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.
The number one guideline is to define every term that is used in an equation.

Maybe introduce a second time if appears again much later in a lengthy work.

Table of notation helpful for complicated models.

Desirable for tables and figures to be understandable on their own, so any notation in them should be additionally defined in their captions.

2. Use italics, boldface and capitalisation appropriately

By convention, mathematical symbols (except Greek letters) should be *italicised*.
Distinguishes text from mathematical notation:

- a large value of a
- a large value of a

2. Use italics, boldface and capitalisation appropriately

By convention, mathematical symbols (except Greek letters) should be *italicised*.
Distinguishes text from mathematical notation:

- a large value of a
- a large value of a

Vectors and matrices should be Roman script and bold:

- $\mathbf{a}$ is a vector
- $\mathbf{A}$ is a matrix
- element i of $\mathbf{a}$ is often a_i
- element in row i and column j of $\mathbf{A}$ is A_{ij} (or sometimes a_{ij})

Generic random variables are usually upper case: X.
Possible numeric values represented by the corresponding lower-case letter: x.

2. Use italics, boldface and capitalisation appropriately

Roman letters for standard mathematical functions:

- $\sin x$, $\cos x$, $\log x$, $\ln x$, e^x

and for other words such as:

$$X \sim \mathrm{Normal}(\mu, \sigma^2)$$

2. Use italics, boldface and capitalisation appropriately

Roman letters for standard mathematical functions:

- $\sin x$, $\cos x$, $\log x$, $\ln x$, e^x

and for other words such as:

$$X \sim \mathrm{Normal}(\mu, \sigma^2)$$

for variable X coming from a normal distribution with mean μ and standard deviation σ.

2. Use italics, boldface and capitalisation appropriately

Roman letters for standard mathematical functions:

- $\sin x$, $\cos x$, $\log x$, $\ln x$, e^x

and for other words such as:

$$X \sim \mathrm{Normal}(\mu, \sigma^2)$$

for variable X coming from a normal distribution with mean μ and standard deviation σ.

Also use a Roman d for derivative and integral of variable X with respect to time t:

$$\frac{\mathrm{d}X}{\mathrm{d}t} \quad \text{and} \quad \int f(x)\mathrm{d}x$$

2. Use italics, boldface and capitalisation appropriately

Roman letters for standard mathematical functions:

- $\sin x$, $\cos x$, $\log x$, $\ln x$, e^x

and for other words such as:

$$X \sim \mathrm{Normal}(\mu, \sigma^2)$$

for variable X coming from a normal distribution with mean μ and standard deviation σ.

Also use a Roman d for derivative and integral of variable X with respect to time t:

$$\frac{\mathrm{d}X}{\mathrm{d}t} \quad \text{and} \quad \int f(x)\mathrm{d}x$$

Units should be given in Roman type to distinguish them from mathematical variables:

- the speed of the polar bear was 1 $\mathrm{km\ h^{-1}}$
- let the speed be x $\mathrm{km\ h^{-1}}$

3. Use subscripts appropriately

Subscripts represent different values of a quantity.

- B_t – biomass of a population in year t, where $t = 1, 2, 3, ..., T$, and T is the maximum year.

3. Use subscripts appropriately

Subscripts represent different values of a quantity.

- B_t – biomass of a population in year t, where $t = 1, 2, 3, \ldots, T$, and T is the maximum year.

Cannot then use:

- B_s – biomass in spatial area s, where $s = 1, 2, 3, \ldots, S$, and S is the number of areas.

3. Use subscripts appropriately

Subscripts represent different values of a quantity.

- B_t – biomass of a population in year t, where $t = 1, 2, 3, \ldots, T$, and T is the maximum year.

Cannot then use:

- B_s – biomass in spatial area s, where $s = 1, 2, 3, \ldots, S$, and S is the number of areas.

Because:

- $t = 1$ gives B_1 as the biomass in year 1
- $s = 1$ gives B_1 as the biomass in spatial area 1

Obvious confusion.

3. Use subscripts appropriately

Should use:

- B_{st} – biomass in area s in year t (for various combinations of s and t)

3. Use subscripts appropriately

Should use:

- B_{st} – biomass in area s in year t (for various combinations of s and t)

Extending this, for data from fish trawl surveys:

- B_{ijkmn} – biomass caught per hour (g h^{-1}) of:
  - taxonomic group i
  - length class j
  - haul k
  - by vessel m
  - in year n

Succinctly describes detailed structure of the data, and made subsequent calculations unambiguous.

3. Use subscripts appropriately

Should use:

- B_{st} – biomass in area s in year t (for various combinations of s and t)

Extending this, for data from fish trawl surveys:

- B_{ijkmn} – biomass caught per hour (g h^{-1}) of:
  - taxonomic group i
  - length class j
  - haul k
  - by vessel m
  - in year n

Succinctly describes detailed structure of the data, and made subsequent calculations unambiguous.

Commas: can use B_{st} or $B_{s,t}$, but always use them to avoid ambiguity: $B_{3,17}$ rather than B_{317}.

3. Use subscripts appropriately

Abbreviations as (non-italicised) subscripts to indicate related definitions.

For example, when modelling DNA records from sediments:

- P_{fp} – probability of **f**alse **p**ositives
- P_{fn} – probability of **f**alse **n**egatives

3. Use subscripts appropriately

Abbreviations as (non-italicised) subscripts to indicate related definitions.

For example, when modelling DNA records from sediments:

- P_{fp} – probability of **f**alse **p**ositives
- P_{fn} – probability of **f**alse **n**egatives

Fisheries:

- B_{MSY} – biomass of a stock at the maximum sustainable yield (MSY)

So MSY is not an index (like B_t), but an acronym that is non-italicised to distinguish from the related B_t.

3. Use subscripts appropriately

Abbreviations as (non-italicised) subscripts to indicate related definitions.

For example, when modelling DNA records from sediments:

- P_{fp} – probability of **f**alse **p**ositives
- P_{fn} – probability of **f**alse **n**egatives

Fisheries:

- B_{MSY} – biomass of a stock at the maximum sustainable yield (MSY)

So MSY is not an index (like B_t), but an acronym that is non-italicised to distinguish from the related B_t.

Please please please avoid using acronyms such as MSY to represent a variable.

3. Use subscripts appropriately

Usually, a subscript as an index should appear on both sides of an equation:

$$y_t = 2x_t + 3,$$

for the relationship between two variables x_t and y_t at each time t.

3. Use subscripts appropriately

Usually, a subscript as an index should appear on both sides of an equation:

$$y_t = 2x_t + 3,$$

for the relationship between two variables x_t and y_t at each time t.
Should generally not be a mixture of subscripts:

$$y_t = 2x_i + 3.$$

Can lead to confusion:

- $i = 1, x_1 = 3$ gives $y_t = 9$
- $i = 2, x_2 = 5$ gives $y_t = 13$

Should use y_{it}.

3. Use subscripts appropriately

Usually, a subscript as an index should appear on both sides of an equation:

$$y_t = 2x_t + 3,$$

for the relationship between two variables x_t and y_t at each time t.
Should generally not be a mixture of subscripts:

$$y_t = 2x_i + 3.$$

Can lead to confusion:

- $i = 1, x_1 = 3$ gives $y_t = 9$
- $i = 2, x_2 = 5$ gives $y_t = 13$

Should use y_{it}.
An expression such as

$$y_{ij} = 2x_i + 3$$

is valid.

3. Use subscripts appropriately

Source of confusion: i and j as indices, but using one again in a summation:

$$y_{ij} = 2x_{ij} + \sum_{i=1}^{N} x_{ij}$$

3. Use subscripts appropriately

Source of confusion: i and j as indices, but using one again in a summation:

$$y_{ij} = 2x_{ij} + \sum_{i=1}^{N} x_{ij}$$

- i used on both sides here as an index (equation valid for all values of i)

3. Use subscripts appropriately

Source of confusion: i and j as indices, but using one again in a summation:

$$y_{ij} = 2x_{ij} + \sum_{i=1}^{N} x_{ij}$$

- i used on both sides here as an index (equation valid for all values of i)
- i used in summation as a dummy index

3. Use subscripts appropriately

Source of confusion: i and j as indices, but using one again in a summation:

$$y_{ij} = 2x_{ij} + \sum_{i=1}^{N} x_{ij}$$

- i used on both sides here as an index (equation valid for all values of i)
- i used in summation as a dummy index

Solution: use k as a dummy index:

$$y_{ij} = 2x_{ij} + \sum_{k=1}^{N} x_{kj}$$

3. Use subscripts appropriately

Source of confusion: i and j as indices, but using one again in a summation:

$$y_{ij} = 2x_{ij} + \sum_{i=1}^{N} x_{ij}$$

- i used on both sides here as an index (equation valid for all values of i)
- i used in summation as a dummy index

Solution: use k as a dummy index:

$$y_{ij} = 2x_{ij} + \sum_{k=1}^{N} x_{kj}$$

Another way to understand:

$$\sum_{k=1}^{N} x_{kj} = \sum_{l=1}^{N} x_{lj} = \sum_{m=1}^{N} x_{mj}$$

since k, l and m are just dummy indices.

4. Be careful with superscripts

Superscripts can distinguish related variables:

- X
- X'

Though be careful – single quotes sometimes used for derivatives of functions:

- $f'(x)$

or transpose of vectors and matrices:

- $\mathbf{X}'$

but transpose can also be

- $\mathbf{X}^{\mathrm{t}}$
- $\mathbf{X}^{\mathrm{T}}$

Asterisks traditionally represent the steady state of a dynamic variable:

- X^*

4. Be careful with superscripts

Please please please do not use a number or a letter (as an index) as a superscript:

- B^t – biomass in year t

but this looks like B raised to the power t.

Setting $t = 2$ gives

- B^2

which is interpreted as $B \times B$, not the desired biomass in year 2.

4. Be careful with superscripts

Please please please do not use a number or a letter (as an index) as a superscript:

- B^t – biomass in year t

but this looks like B raised to the power t.

Setting $t = 2$ gives

- B^2

which is interpreted as $B \times B$, not the desired biomass in year 2.

Even worse example:

- u_{at}^{sg} – exploitation rate of fish of:
  - sex s
  - age a
  - caught by fishing gear g
  - in year t

Just use subscripts and go with

- u_{atsg}

like the earlier example of B_{ijkmn}.

4. Be careful with superscripts

Subscripts and superscripts should always come after the variable. Not:

- ${}_j v_t$

Otherwise, if θ multiplied by ${}_j v_t$ gives

- $\theta_j v_t$

but ambiguous: is first component θ_j or θ?

5. Helpful to distinguish variables from parameters

Example – Ricker model:

$$R = \alpha S e^{-\beta S}$$

where

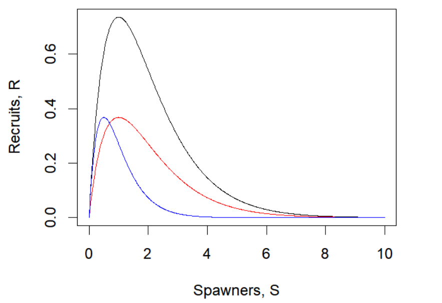
- R – the recruitment of new fish (dependent variable on the left-hand side)
- S – the spawning stock biomass (independent variable on the right-hand side)
- α – maximum number of recruits produced by each unit of spawning stock biomass (parameter)
- β – scales the intensity of density dependence (parameter).

5. Helpful to distinguish variables from parameters

$$R = \alpha S e^{-\beta S}$$

The model says how the variables R and S relate to each other.

Using particular parameter values of α and β distinguishes one realisation of the model from another:



5. Helpful to distinguish variables from parameters

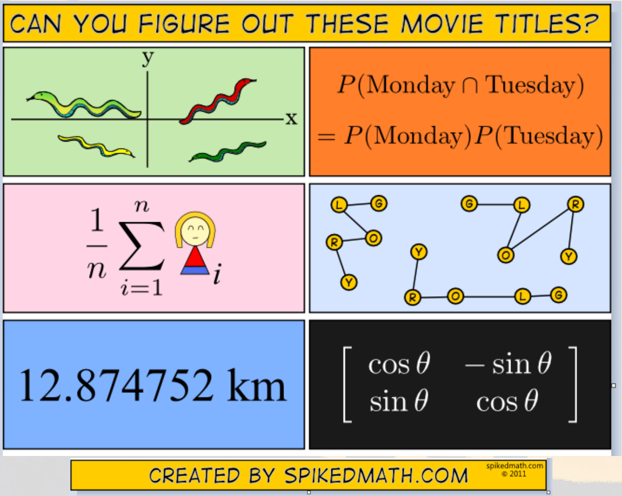
Useful, if possible:

- variables – upper case
- parameter and constants (like gravitational constant g) – lower case and Greek letters

However, stick with established convention when this is not followed, or for simple equations such as

$$y = 3x + 7$$

with variables x and y (often the first choice to represent variables).


CAN YOU FIGURE OUT THESE MOVIE TITLES?
y
x
$P(\text{Monday} \cap \text{Tuesday})$
$= P(\text{Monday})P(\text{Tuesday})$
$\frac{1}{n}\sum_{i=1}^{n}$ i
L G R O Y
G L R O Y
Y R O L G
12.874752 km
$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
CREATED BY SPIKEDMATH.COM
spikedmath.com © 2011

6. Avoid multi-letter variable names

Use only one letter (rather than two or more) to represent a quantity.
Common abbreviation: SSB for spawning stock biomass.
Fine as an acronym in a sentence, but problematic when SSB used as a mathematical quantity in an equation:

- SSB_t – spawning stock biomass in year t
- S – selectivity (proportion of biomass caught by fishing)
- B_t – total (spawning plus non-spawning) biomass in year t

But this breeds ambiguity:

$$\frac{SB_t}{SSB_t}$$

6. Avoid multi-letter variable names

Use only one letter (rather than two or more) to represent a quantity.
Common abbreviation: SSB for spawning stock biomass.
Fine as an acronym in a sentence, but problematic when SSB used as a mathematical quantity in an equation:

- SSB_t – spawning stock biomass in year t
- S – selectivity (proportion of biomass caught by fishing)
- B_t – total (spawning plus non-spawning) biomass in year t

But this breeds ambiguity:

$$\frac{SB_t}{SSB_t} = \frac{\cancel{S}B_t}{\cancel{S}SB_t} \qquad ?$$

6. Avoid multi-letter variable names

Use only one letter (rather than two or more) to represent a quantity.
Common abbreviation: SSB for spawning stock biomass.
Fine as an acronym in a sentence, but problematic when SSB used as a mathematical quantity in an equation:

- SSB_t – spawning stock biomass in year t
- S – selectivity (proportion of biomass caught by fishing)
- B_t – total (spawning plus non-spawning) biomass in year t

But this breeds ambiguity:

$$\frac{SB_t}{SSB_t} = \frac{\cancel{S}B_t}{\cancel{S}SB_t} = \frac{S\cancel{B_t}}{SS\cancel{B_t}} \qquad ?$$

6. Avoid multi-letter variable names

Use only one letter (rather than two or more) to represent a quantity.
Common abbreviation: SSB for spawning stock biomass.
Fine as an acronym in a sentence, but problematic when SSB used as a mathematical quantity in an equation:

- SSB_t – spawning stock biomass in year t
- S – selectivity (proportion of biomass caught by fishing)
- B_t – total (spawning plus non-spawning) biomass in year t

But this breeds ambiguity:

$$\frac{SB_t}{SSB_t} = \frac{\cancel{S}B_t}{\cancel{S}SB_t} = \frac{S\cancel{B}_t}{SS\cancel{B}_t} \qquad ?$$

Solution: pick, say, B_t and T_t for, respectively, the spawning and total biomasses at time t.

6. Avoid multi-letter variable names

Occasionally may be okay to use an acronym or word as a variable name.
Often used in simple statistical models:

$$e^{\beta_1 + \beta_2 \times \mathrm{MeanDepth}_i}$$

which intuitively represents an exponential effect of a linear function of mean depth (β_1 and β_2 are parameters).

6. Avoid multi-letter variable names

Occasionally may be okay to use an acronym or word as a variable name.
Often used in simple statistical models:

$$e^{\beta_1 + \beta_2 \times \mathrm{MeanDepth}_i}$$

which intuitively represents an exponential effect of a linear function of mean depth (β_1 and β_2 are parameters).

- such statistical models generally require no further subsequent mathematical manipulation
- may avoid the problems outlined in previous slide

6. Avoid multi-letter variable names

Occasionally may be okay to use an acronym or word as a variable name.
Often used in simple statistical models:

$$e^{\beta_1+\beta_2\times\text{MeanDepth}_i}$$

which intuitively represents an exponential effect of a linear function of mean depth (β_1 and β_2 are parameters).

- such statistical models generally require no further subsequent mathematical manipulation
- may avoid the problems outlined in previous slide

Words or acronyms can make models more understandable:

- e.g. for stakeholders who are not quantitatively trained
- for a presentation or a poster, especially if not all details of a model are going to be given

6. Avoid multi-letter variable names

Occasionally may be okay to use an acronym or word as a variable name.
Often used in simple statistical models:

$$e^{\beta_1 + \beta_2 \times \text{MeanDepth}_i}$$

which intuitively represents an exponential effect of a linear function of mean depth (β_1 and β_2 are parameters).

- such statistical models generally require no further subsequent mathematical manipulation
- may avoid the problems outlined in previous slide

Words or acronyms can make models more understandable:

- e.g. for stakeholders who are not quantitatively trained
- for a presentation or a poster, especially if not all details of a model are going to be given

Just ensure:

- no potential for confusion (can be hard to guarantee when first defining notation)
- equations don't become cumbersome and hard to understand

6. Avoid multi-letter variable names

One solution is to simultaneously give equations in word and notation form:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\text{uptake} + \text{respiration} + Z\,\text{excretion} + Z\,\text{predation excretion} + \text{mixing}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \text{uptake} - \text{respiration} - \text{grazing by } Z - \text{sinking} - \text{mixing}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \text{growth} - \text{higher predation}$$

The specific functional forms used (discussed in Section 3) are:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\frac{N}{e+N}\frac{a}{b+cP}P + rP + \frac{\beta\lambda P^2}{\mu^2+P^2}Z + \gamma d Z^2 + k(N_0 - N) \tag{1}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{N}{e+N}\frac{a}{b+cP}P - rP - \frac{\lambda P^2}{\mu^2+P^2}Z - (s+k)P \tag{2}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \frac{\alpha\lambda P^2}{\mu^2+P^2}Z - dZ^2 \tag{3}$$

7. Fully define probability distributions

- Discrete random variable – takes discrete values (e.g. 1, 2, 3, ...)

$f(x)$ – *probability mass function* of a discrete variable X:

$$f(x) = \mathrm{P}(X = x)$$

where

- $\mathrm{P}(\cdot)$ – probability of occurrence of the event in parentheses (X takes the value x)

For example, Poisson distribution for count data:

$$f(x) = \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

where x are the possible values and there is just one parameter $\lambda > 0$.

7. Fully define probability distributions

- Continuous random variable – can take any value within a specified range (e.g. between 0 and 10).

$f(x)$ – continuous probability density function, for example:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}, \quad x > 0$$

for parameters μ and σ, and log is natural logarithm.

Worth defining 'log' and using $\log_{10}$ for base-10 logarithm (realising ln is also used for the natural logarithm).

7. Fully define probability distributions

Note in both examples the domain is specified:

$$
\begin{aligned}
f(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \\
f(x) &= \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}, \quad x > 0
\end{aligned}
$$

This also helps confirm the distribution is appropriate for the question at hand.

7. Fully define probability distributions

Note in both examples the domain is specified:

$$
\begin{aligned}
f(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots \\
f(x) &= \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}, \quad x > 0
\end{aligned}
$$

This also helps confirm the distribution is appropriate for the question at hand.

Normal distribution – generally no need to explicitly specify that $-\infty < x < \infty$.

7. Fully define probability distributions

For discrete and continuous distributions, can define

- $F(x)$ – cumulative probability distribution:

$$\text{discrete} \quad F(x) = \mathrm{P}(X \le x) = \sum_{i=0}^{x} f(i)$$

$$\text{continuous} \quad F(x) = \mathrm{P}(X \le x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$$

7. Fully define probability distributions

For a second random variable, Y, possible options are:

- $g(y)$
- $f_Y(y)$, and use $f_X(x)$ for X
- $f(y)$ – does not work as not distinguishable from $f(x)$

7. Fully define probability distributions

For a second random variable, Y, possible options are:

- $g(y)$
- $f_Y(y)$, and use $f_X(x)$ for X
- $f(y)$ – does not work as not distinguishable from $f(x)$

Some distributions have a conventional shorthand, e.g.:

- $X \sim \mathrm{N}(\mu, \sigma^2)$ – variable X is normally distributed with mean μ and standard deviation σ
- $X \sim \mathrm{N}(0, 2)$ – here specify whether the 2 is the standard deviation or the variance to avoid ambiguity
- $X \sim \mathrm{Gamma}(a, b)$ – Gamma distribution, but define parameters as can be parameterised using shape and either rate, scale or mean

8. Give equations of a model rather than just computer code

One reason that acronyms or words get used to identify variables may be that this is how they are written in computer code.

Words in code can make code easier to read and avoid typographical errors.

But words in equations are not recommended. Solutions:

- write equations first and then have a comment in the corresponding code that links the words used in the code to the corresponding mathematical notation
- modern simpler alternative is the R package `knitr`
- `knitr` interweaves text and computer code in a single file, easily enabling the same succinct notation to be used throughout
- `Rmarkdown` now allows easy interweaving math, code, write up, references, figures etc.

8. Give equations of a model rather than just computer code

- Writing equations may seem a necessary prerequisite to writing code
- Some people are proficient programmers but are sometimes unable to translate the code into mathematical notation

For example, for a numeric vector `x` in R, the command

```
y <- cumsum(x)
```

is defined as creating a vector `y` where each element is the cumulative sum of the elements of `x`. Equivalent math: for a vector $\mathbf{x}$ of length n,

$$y_i = \sum_{j=1}^{i} x_j, \quad i = 1, 2, 3, \ldots, n.$$

8. Give equations of a model rather than just computer code

- Writing equations may seem a necessary prerequisite to writing code
- Some people are proficient programmers but are sometimes unable to translate the code into mathematical notation

For example, for a numeric vector `x` in R, the command

```
y <- cumsum(x)
```

is defined as creating a vector `y` where each element is the cumulative sum of the elements of `x`. Equivalent math: for a vector $\mathbf{x}$ of length n,

$$y_i = \sum_{j=1}^{i} x_j, \quad i = 1, 2, 3, \ldots, n.$$

However, the verbal description is ambiguous, unlike the equation.

8. Give equations of a model rather than just computer code

The `cumsum(x)` example is usually just for book-keeping.

But for more complex code (e.g. log-likelihood functions), give the equation not the code, since:

- code relies on reader having knowledge of that particular language
- languages evolve and fall out of favour
- properly documented equations will stand the test of time
- useful to supply code as Supporting Information or as a package

9. Abide by conventions (but still define everything)

Common mathematical usage of particular letters and symbols

Letter/symbol	Common usage
e	usually avoided to prevent confusion with non-italicised e (=2.718...)
f, g	function, e.g. $f(x) = x^3 + 7$
i, j, k	index, e.g. the ith element of vector $\mathbf{x}$ is x_i
n, N	sample size
o, O	usually avoided to prevent confusion with number 0
t	time
u, v, w	speeds
x, y, z	variables, or co-ordinates in space
$\mathrm{P}(\cdot)$	probability of occurrence of the event in parentheses
X, Y, Z	variables

9. Abide by conventions (but still define everything)

Common mathematical usage of particular letters and symbols

Letter/symbol	Common usage
$\alpha, \beta, \gamma, \theta$	parameters
δ, Δ	difference or change in a variable, ΔX, or a parameter
ε	a small value, or random noise term
μ	mean
π	the value 3.141...
$\prod$	product of the proceeding values
σ	standard deviation
$\sum$	summation of the proceeding values
X^*	steady-state value of X
$\dot{X}$	derivative of X
$f'(x)$	derivative of $f(x)$ with respect to x
$\partial f / \partial x$	partial derivative of $f(x,y)$ with respect to x
$\hat{\theta}$	an estimate of θ

9. Abide by conventions (but still define everything)

Compare:
1. The population size is given by

$$t_\varepsilon = g t_{\varepsilon-1}(1 - t_{\varepsilon-1}) + f_\varepsilon$$

where

- t_ε – population in year ε ($\varepsilon = 1, 2, 3, \ldots, \Gamma$)
- g – intrinsic growth rate at low population size
- f_ε – normally distributed random noise with mean σ and variance μ^2

9. Abide by conventions (but still define everything)

With:

2. The population size is given by

$$X_t = rX_{t-1}(1 - X_{t-1}) + \varepsilon_t$$

where

- X_t – population in year t ($t = 1, 2, 3, ..., T$)
- r – intrinsic growth rate at low population size
- ε_t – normally distributed random noise with mean μ and variance σ^2.

9. Abide by conventions (but still define everything)

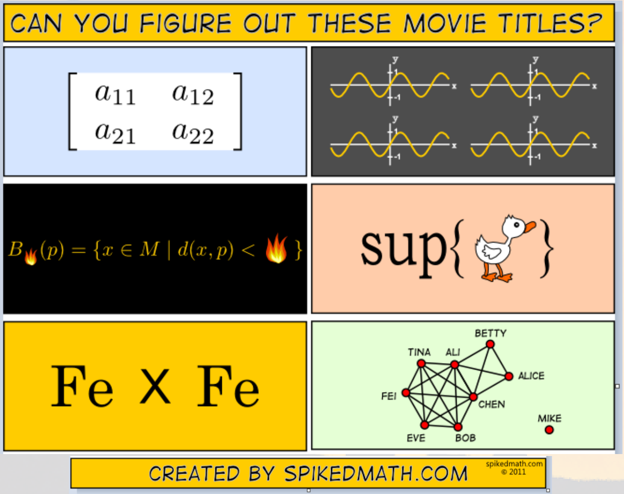
Within some fields, certain notation may be standard.

But still clearly define notation, in particular because of the multidisciplinary nature of ecology.

Established conventions:

- may be best to try to retain
- but maybe not if conventional notation was not well thought-out
- can help when learning a new area
- multidisciplinary work – can be hard to retain all conventions (and appease everyone)

Quiz time



CAN YOU FIGURE OUT THESE MOVIE TITLES?
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
$B(p) = \{x \in M \mid d(x,p) < \}$
sup{ }
Fe X Fe
BETTY
TINA
ALI
ALICE
FEI
CHEN
MIKE
EVE
BOB
CREATED BY SPIKEDMATH.COM
spikedmath.com
© 2011

10. Use parentheses and brackets only as necessary

Parentheses ()

- around the arguments of a function, e.g. $f(x)$
- denote which calculations in an equation need to be done first, e.g. $3(x+5)$

To avoid too many slightly-different sized parentheses also use:

- square brackets [ ]
- braces {}

10. Use parentheses and brackets only as necessary

Parentheses ()

- around the arguments of a function, e.g. $f(x)$
- denote which calculations in an equation need to be done first, e.g. $3(x+5)$

To avoid too many slightly-different sized parentheses also use:

- square brackets [ ]
- braces { }

But don't overuse parentheses:

$$f(x) = \frac{\left(\lambda^{x}\right)\left(\mathrm{e}^{-\lambda}\right)}{x!}, \quad x = 0, 1, 2, \ldots$$

makes the equation appear more complicated than it is.
This often happens in practice – but extra clutter can impede comprehension.

10. Use parentheses and brackets only as necessary

Almost always no need to use a symbol such as $\times$ or $\cdot$ for multiplication, unless

- using words as variable names
- breaking up long equations for readability

10. Use parentheses and brackets only as necessary

Almost always no need to use a symbol such as $\times$ or $\cdot$ for multiplication, unless

- using words as variable names
- breaking up long equations for readability

Parentheses also used:

- $x \in (0,1)$ means $0 < x < 1$
- $x \in [0,1]$ means $0 \leq x \leq 1$
- $x \in [0,1)$ means $0 \leq x < 1$

11. Equations should be part of sentences

Sometimes require punctuation:

- maybe a comma when in middle of sentence
- a period when completing a sentence

11. Equations should be part of sentences

Sometimes require punctuation:

- maybe a comma when in middle of sentence
- a period when completing a sentence

Worth numbering all equations (even if not referred to later):

- easier for others to refer to
- easier for your future self to refer to

11. Equations should be part of sentences

Sometimes require punctuation:

- maybe a comma when in middle of sentence
- a period when completing a sentence

Worth numbering all equations (even if not referred to later):

- easier for others to refer to
- easier for your future self to refer to

Single terms in equations (or very simple equations) appear within text and not on their own line.

Fractions within such lines should be a/b not $\frac{a}{b}$.

12. Revise notation early on if necessary

- Think about notation before defining a model
- Can be useful to revisit notation *early on* in a project once some details are fleshed out
- But hard to change notation once proceeded far enough, such as:
  - two papers already published
  - thousands of lines of computer code shared with others
    - potentially confusing to switch notation in related works

Thus, time spent revisiting notation early on may be time well spent.

12. Revise notation early on if necessary

- Think about notation before defining a model
- Can be useful to revisit notation *early on* in a project once some details are fleshed out
- But hard to change notation once proceeded far enough, such as:
  - two papers already published
  - thousands of lines of computer code shared with others
    - potentially confusing to switch notation in related works

Thus, time spent revisiting notation early on may be time well spent.

Similar to functionalising computer code – taking a pause and rewriting code in a more user-friendly way tends to pay off in future.

12. Revise notation early on if necessary

Related point – PROOFREAD.

Different publishers have different minor typsetting rules (that may or may not impact comprehension).

Letters should have unique definition in single piece of work.

Example of confusing notation

Highlights several problems:

$$\sigma_a^{s^2} = \left(\frac{sd_a^S}{L_a^S}\right)^2$$

where:

- sd_a^S – standard deviation of the length of a fish of age a and sex S
- σ_a^s – standard deviation of the distribution of log(length) at age a for sex s
- L_a^S – length at age a for sex s
- s – sex, $s = 1$ for females and $s = 2$ for males

Example of confusing notation

Highlights several problems:

$$\sigma_a^{s^2} = \left(\frac{sd_a^S}{L_a^S}\right)^2$$

where:

- sd_a^S – standard deviation of the length of a fish of age a and sex S
- σ_a^s – standard deviation of the distribution of log(length) at age a for sex s
- L_a^S – length at age a for sex s
- s – sex, $s = 1$ for females and $s = 2$ for males

Example of confusing notation

Highlights several problems:

$$\sigma_a^{s^2} = \left(\frac{sd_a^S}{L_a^S}\right)^2$$

where:

- sd_a^S – standard deviation of the length of a fish of age a and sex S
- σ_a^s – standard deviation of the distribution of log(length) at age a for sex s
- L_a^S – length at age a for sex s
- s – sex, $s = 1$ for females and $s = 2$ for males

Example of confusing notation

Highlights several problems:

$$\sigma_a^{s^2} = \left(\frac{sd_a^S}{L_a^S}\right)^2$$

where:

- sd_a^S – standard deviation of the length of a fish of age a and sex S
- σ_a^s – standard deviation of the distribution of log(length) at age a for sex s
- L_a^S – length at age a for sex s
- s – sex, $s = 1$ for females and $s = 2$ for males

Closing remarks

Hope these guidelines:

- are useful
- help improve future comprehension of biological models
- help improve reproducibility

Re-iterate: these are guidelines but not rules, and should be overlooked when appropriate.

Acknowledgements
Jaclyn Cleary, Brooke Davis, Wayne Hajas, James Robinson, Catarina Wor, Brianna Wright.
Three anonymous reviewers, Sean McMahon and Robert O'Hara.