Fundamentals of Modeling

EESS211 / EARTHSYS 211

Definitions

A model is…

- “an abstract representation of a system or process”
- “a simplified representation of the real world”

Definitions

A model is …

- a way to explore and add to our understanding of reality
- a way to communicate expert knowledge without the expert present

A model is not…

- reality
- a substitute for measurements or experiments

Goals of Modeling

- To aid measurement and experiment designs
- To summarize understanding by forcing the abstraction, integration, and formalization of scientific ideas
- To simulate “virtual” experiments
- To communicate scientific understanding to non-experts
- To make predictions
- Modeling is NOT a goal in itself

Why are Models so Important for Environmental Science?

- We are often interested in temporal and spatial scales where measurements and experiments are difficult or impossible.
- Many factors are not easily controlled, so experiments are often impossible
- There are often many interacting system components, so that the implications of our understanding of each component and relationship are not immediately clear.

But be careful not to overdo it. A couple of useful warnings…

1) All models are wrong, some are useful (George Box)

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

In applying mathematics to subjects such as physics or statistics we make tentative assumptions about the real world which we know are false but which we believe may be useful nonetheless. The physicist knows that particles have mass and yet certain results, approximating what really happens, may be derived from the assumption that they do not. Equally, the statistician knows, for example, that in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, he can often derive results which match, to a useful approximation, those found in the real world.

2) Don't fall in love with any one model, or:

"The moment one has offered an original explanation for a phenomenon which seems satisfactory, that moment affection for his intellectual child springs into existence; and as the explanation grows into a definite theory, his parental affections cluster about his intellectual offspring, and it grows more and more dear to him, so that, while he holds it seemingly tentative, it is still lovingly tentative, and not impartially tentative.

So soon as this parental affection takes possession of the mind, there is a rapid passage to the adoption of the theory. There is an unconscious selection and magnifying of the phenomena that fall into harmony with the theory and support it, and an unconscious neglect of those that fail of coincidence. The mind lingers with pleasure upon the facts that fall happily into the embrace of the theory, and feels a natural coldness toward those that seem refractory. Instinctively there is a special searching-out of phenomena that support it, for the mind is led by its desires."   Chamberlin 1890

2 General Uses of Modeling

- Exploratory Modeling
  - explore implications of (often untested) hypotheses
- Predictive Modeling
  - estimate future or unobserved values
  - requires estimates of uncertainties to be useful
  - harder but also more powerful

2 General Uses of Modeling

- Exploratory Modeling
  - explore implications of (often untested) hypotheses

Your projects will focus mainly on the first, because using data to calibrate and test predictions takes more time. But the assignments and lectures will focus on both, with an emphasis on prediction.

- Predictive Modeling
  - estimate future or unobserved values
  - requires estimates of uncertainties to be useful
  - harder but also more powerful

Start thinking about your model project:

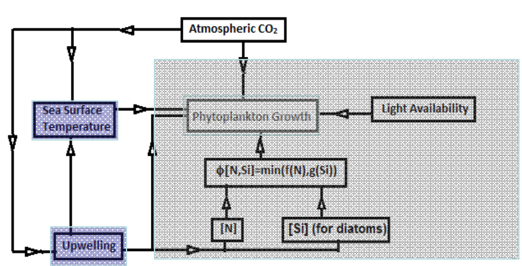
-What is the question(s) you want to ask with your model?

Some things to remember for projects

- Can use literature to help define equations and parameters
- Does not have to be a completely original model
- Use of data you don't already have in hand is not generally advised
- Goal is to reinforce concepts and at the same time help you progress on research
- Getting early feedback will make it easier

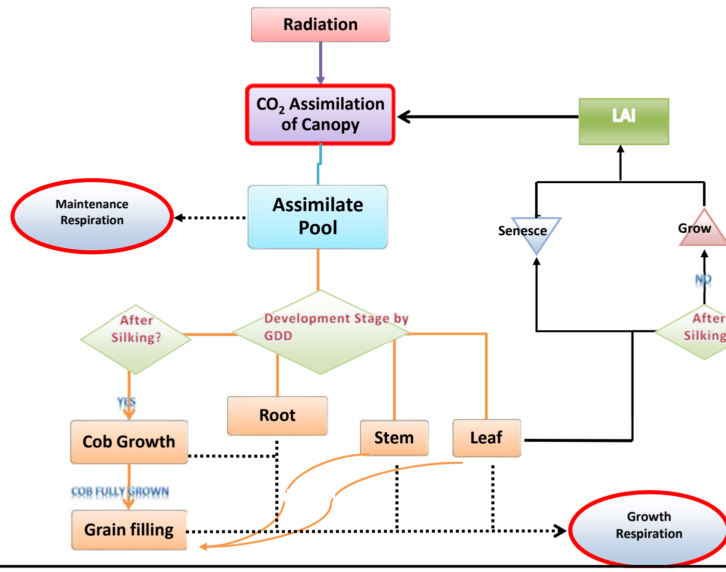
A few examples of past projects

- How does rising SST affect phytoplankton growth and their ability to convert the atmospheric CO_2 to biomass?
- How does upwelling influence the biological activity?



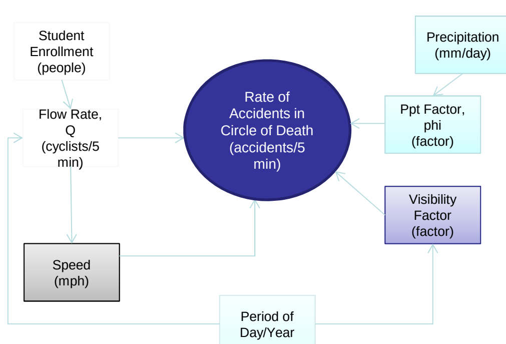
[N], [Si] - Nutrient Concentrations

• What is the relative importance of temperature, radiation, and plant genetics in determining corn yield?



• During which periods of the day/year are accident rates in the Circle of Death the highest?

•What will happen to these accident rates under future scenarios of population growth and climate change?



Outline

- Some more definitions
- Brief review of statistics
- Overview of the modeling process
- An example

Modeling Definitions

A system is a set of inter-related components and the relationships between them

A variable is a value that changes freely in time and space

A constant is a value that does not vary between systems (e.g. gravity)

A parameter is a value which is constant for a single case (model run) but may vary between cases

Modeling Definitions

Scale refers to the grain and extent of a model.

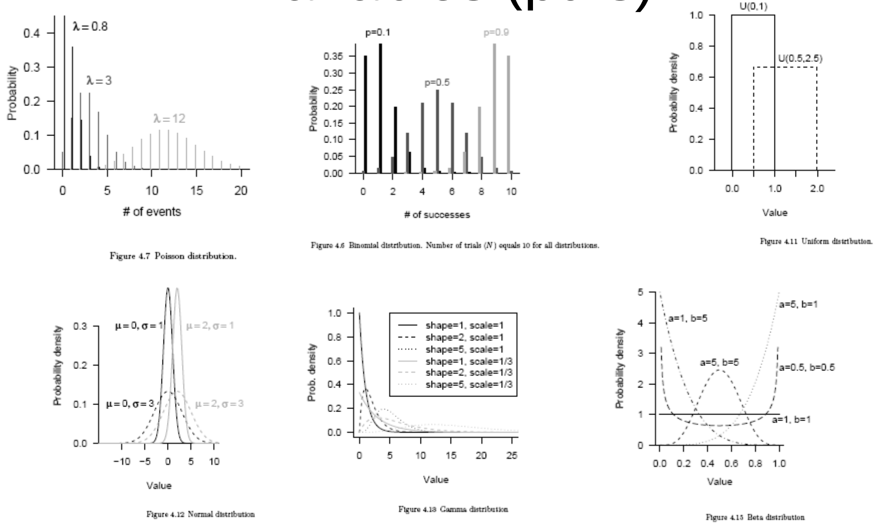
Grain is the spatial or temporal resolution, below which model inputs and outputs are assumed to be homogeneous

Extent is the total area or time period covered by a model

Brief Review of Statistics Concepts Relevant to this Class

- Mean
- Variance
- Standard deviation, standard error
- Random variables
- Probability distribution functions (pdfs)

Lots of flavors of random variables (pdfs)



(bolker)

A brief aside:

- How do computers generate random numbers??

A brief aside:

•A common way for uniform distributions is linear congruential generator (LCG) which uses the recursive formula:

$$x_{n+1} = ax_n + b(mod\ m)$$

•A common way for Gaussian (normal) distributions is Box-Muller, which uses 2 uniforms as input

•See help for runif, rnorm, or RNG for more info

A brief aside:

•Do they work well?

•Lot’s of ways to test this (used to be an old homework)

E.g…

A brief aside:

```
#let's see if there are the right fraction of very high values

n=1e4                   #length of random sequence
nrun=1e2                #number of times to generate sequence and compute
statistic
temp = numeric (nrun)   #vector to save results
for (i in 1: nrun) {
        x = runif (n)
        temp[i] = sum ( x > 0.99) / n       #save statistic
}
hist(temp)

paste('confidence interval:',mean(temp)-
2*sd(temp)/sqrt(nrun),mean(temp)+2*sd(temp)/sqrt(nrun))
```

A brief aside:

```
#let's see if there tend to be differences depending on if previous number was positive

nsamp=1e5
x = rnorm(nsamp)
index = which(x[1:(nsamp-1)] <= 0)#find which values are negative
x1 = x[index+1]        #take a subset of x, equal to all those values following negative values
x2 = x[-(index+1)]     #take the other values, note in r a negative index means to omit these terms
#now let's plot them one on top of the other
x11()                          #open new window
hist(x1, col="gray", xlim=c(-4,4))
hist(x2, col=NA,border=4,add=T)

#We can test if x1 and x2 have statistically different distributions using the kolmogorov-smirnov test

ks.test(x1,x2)
```

Outline

- Some more definitions
- Brief review of statistics
- Overview of the modeling process
- An example

Steps to Modeling

1) Define purpose of model
2) Develop conceptual model / abstraction
3) Select the model type and scale
4) Write computer code to implement conceptual model

Steps to Modeling

5) Estimate “free” parameters in model (calibration)

6) Evaluate model with sensitivity and uncertainty analysis

7) Evaluate model predictions with measurements (not those used in #5)

8) Model Application (prediction, inversion, etc.)

Steps to Modeling

Untested Hypothesis

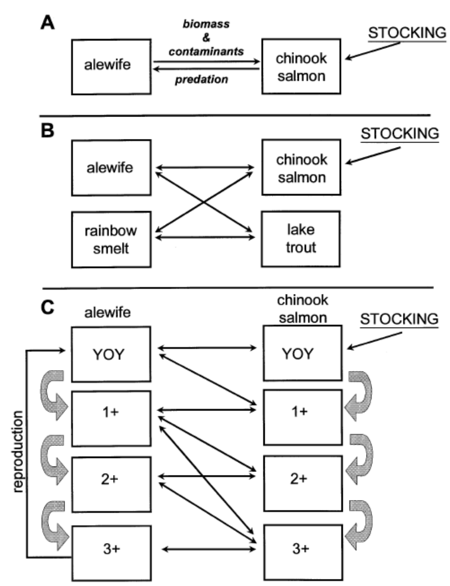
1) Goal Definition
2) Conceptualization
3) Model Type/Scale Decisions
4) Coding
5) Calibration ← Data
6) Sensitivity Analysis
7) “Validation” ← Data
8) Application ← Data

Partially Tested Hypothesis

Conceptual Models

Usually some sort box-and-arrow plot showing variables and relationships

Conceptual Models



Jackson et al. 2000

Types of Models

There is not a clear taxonomy, because models vary on a continuum. But some useful (?) distinctions:

Mathematical:	deterministic vs. stochastic
	analytical vs. simulation
Spatial:	distributed vs. lumped
Temporal:	dynamic vs. static

Types of Models

Conceptual:	mechanistic vs. statistical
	proccess-based vs. statistical
	bottom-up vs. top-down
	white-box vs. black-box

Note that nearly all models involve some level of empiricism (i.e. parameter estimation).

$$Y = F\ (X, \beta)$$

desired output(s)

functional form of relationship

model input(s)

model parameter(s)

A Modeling Example

1) Goal Definition
2) Conceptualization
3) Model Type/Scale Decisions
4) Coding
5) Calibration ← Data
6) Sensitivity Analysis
7) "Validation" ← Data
8) Application ← Data

Untested Hypothesis

Partially Tested Hypothesis

Some Practical Modeling Tips

- Begin each code with name, description, calling sequence, inputs, outputs, dates written/modified, contact info, and anything else that might be useful to someone else (incl. you in the future).
- Add comments throughout code
- Treat all parameters as input or, if "hard-coded" at top of code (to make changing easier)

A Modeling Example

1) Goal Definition — Untested Hypothesis
2) Conceptualization
3) Model Type/Scale Decisions
4) Coding
5) Calibration ← Data
6) Sensitivity Analysis
7) "Validation" ← Data

---

8) Application ← Data

Partially Tested Hypothesis

A Modeling Example

Goal Definition

-To predict the amount of water that can be transferred from land to atmosphere (potential evapotranspiration) at a given location and day...

Modeling potential evapotranspiration (PET)

Building a conceptual model:

-what variables should we consider?

-what type of equation(s) should we use?

Modeling PET

Some variables to consider:

Solar radiation
Air temperature
Air humidity
Wind Speed
Vegetation hydraulic capacity
Height of vegetation (?)
Height of nearby vegetation (?)
Direct vs. diffuse radiation (?)
Rotation of the earth (?)

Modeling PET

Penman-Monteith:

It takes energy to evaporate water.
Start with energy balance

$R_n = H + \lambda E + G$

R_n = Net irradiance (W m^{-2})
H = sensible heat flux
λE = latent heat flux
G = ground heat flux

Modeling PET

Penman-Monteith:

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma\left(1 + \frac{r_s}{r_a}\right)} \quad (3)$$

where R_n is the net radiation, G is the soil heat flux, $(e_s - e_a)$ represents the vapour pressure deficit of the air, ρ_a is the mean air density at constant pressure, c_p is the specific heat of the air, Δ represents the slope of the saturation vapour pressure temperature relationship, γ is the psychrometric constant, and r_s and r_a are the (bulk) surface and aerodynamic resistances. The parameters of the equation are defined in Chapter 3.

Modeling PET

The FAO Penman-Monteith:

The FAO Expert Consultation on Revision of FAO Methodologies for Crop Water Requirements accepted the following unambiguous definition for the reference surface:

> **"A hypothetical reference crop with an assumed crop height of 0.12 m, a fixed surface resistance of 70 s m^{-1} and an albedo of 0.23."**

Modeling PET

The FAO Penman-Monteith:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34 u_2)} \quad (6)$$

'e

ET_o reference evapotranspiration [mm day^{-1}],
R_n net radiation at the crop surface [MJ m^{-2} day^{-1}],
G soil heat flux density [MJ m^{-2} day^{-1}],
T mean daily air temperature at 2 m height [°C],
u_2 wind speed at 2 m height [m s^{-1}],
e_s saturation vapour pressure [kPa],
e_a actual vapour pressure [kPa],
e_s - e_a saturation vapour pressure deficit [kPa],
Δ slope vapour pressure curve [kPa °C^{-1}],
γ psychrometric constant [kPa °C^{-1}].

Modeling PET

So to use FAO Penman-Monteith you need to know daily mean temperature, wind speed, relative humidity, and solar radiation.

As you can image, many others have sought simpler models...

Modeling PET

Priestly-Taylor (1972):

a simplification of penman-monteith that uses a constant to replace wind & humidity inputs

$$\lambda E = a\frac{s(R_n - G)}{s+\gamma}$$

a = calibrated constant = often 1.26 is used as default

Modeling PET

Thornwaite equation for monthly PET (1948):

$$PET = 16\left(\frac{L}{12}\right)\left(\frac{N}{30}\right)\left(\frac{10T_a}{I}\right)^{\alpha}$$

$$\alpha = (6.75 \times 10^{-7})I^3 - (7.71 \times 10^{-5})I^2 + (1.79 \times 10^{-2})I + 0.49$$

$$I = \sum_{i=1}^{n}\left(\frac{T_{ai}}{5}\right)^{1.514}$$

Sum for all months when $T_a > 0$

L = daylength (hours)
N = # days in month
Ta = mean monthly temperature

Modeling PET

Hamon (1963)

PET = 715.5 * DL * esat(T_m) /(T_m + 273.2)

DL = daylength (days) (from latitude and day of year)
esat = saturated vapor pressure
Tm = mean temperature

Modeling PET

Hargreaves (1985):

$$ET_o = 0.0023(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} R_a \quad (52)$$

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r [\omega_s \sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\sin(\omega_s)] \quad (21)$$

where

R_a extraterrestrial radiation [MJ m^{-2} day^{-1}],
G_{sc} solar constant = 0.0820 MJ m^{-2} min^{-1},
d_r inverse relative distance Earth-Sun (Equation 23),
ω_s sunset hour angle (Equation 25 or 26) [rad],
φ latitude [rad] (Equation 22),
δ solar decimation (Equation 24) [rad].

Modeling PET

Linacre (1977):

The Penman formula for the evaporation rate from a lake is simplified to the following:

$$E_0 = \frac{700\ T_m/(100 - A) + 15\ (T - T_d)}{(80 - T)} \quad (\text{mm day}^{-1})$$

where $T_m = T + 0.006h$, h is the elevation (metres), T is the mean temperature, A is the latitude (degrees) and T_d is the mean dew-point. Values given by this formula typically differ from measured values by about 0.3 mm day^{-1} for annual means, 0.5 mm day^{-1} for monthly means, 0.9 mm day^{-1} for a week and 1.7 mm day^{-1} for a day. The formula applies over a wide range of climates. Monthly mean values of the term $(T - T_d)$ can be obtained either from an empirical table or from the following empirical relationship, provided precipitation is at least 5 mm month^{-1} and $(T - T_d)$ is at least 4°C:

$$(T - T_d) = 0.0023h + 0.37\ T + 0.53\ R + 0.35\ R_{ann} - 10.9°\text{C}$$

where R is the mean daily range of temperature and R_{ann} is the difference between the mean temperatures of the hottest and coldest months. Thus the evaporation rate can be estimated simply from values for the elevation, latitude and daily maximum and minimum temperatures.