Advanced Environmental Modeling
by
Razeeb Sarker

C | A | U
Christian-Albrechts-Universität zu Kiel

Taught by:

Prof. Dr. Kai Wirtz

Helmholtz-Zentrum Hereon
Abteilungsleiter: Ökosystemmodellierung
Analyse und Modellierung Institut für Küstensysteme

Contents

Preparatory course for Ecosystem Modeling 3

Exercise 2: R-plot functions 3

Exercise 3: Differential calculus and simple numerics 5

Exercise 4: Exponential spread of a viral disease 6

Primary production in the water column 6

Exercise 5: Photosynthesis curves ("Light response") 7

Exercise Exercise 6a: Light attenuation in the water column 8

Exercise 6b: Integral primary production 9

Exercise 7: Working with NetCDF data 10

Exercise 7b: Map the North Sea bathymetry 11

Primary production of the German Bight 12

Exercise 1: Light gradients in the coastal ocean 12

Exercise 2: Simulating phytoplankton growth in space and time . . . 14

Exercise 3: Nutrient limitation of primary production 15

Exercise 4a: Total primary production based on data and models . . 16

Exercise 4b: Global implications of coastal/shelf production 17

Exercise 4c: Reducing eutrophication 17

List of Figures

Preparatory course for Ecosystem Modeling

Exercise 2: R-plot functions

Mathematical models are in general built on functions. Functions express dependencies among variables (e.g., how the WLAN signal weakens at increasing distance to the router). Here a list of functions often used in ecosystem or biological modeling:

		equation	parameters
1	linear	$y(t) = y_0 + \alpha\, t$	y_0: off-set α: slope/velocity/(regression coefficient)
2	exponential	$y(t) = y_0 \cdot \mathrm{e}^{\alpha\, t}$	y_0: initial off-set α: growth/attenuation/decay coeff.
3	logarithm	$y(t) = \log(t),\ \log\left(\mathrm{e}^{y}\right) = y$	(non-linear regression)
4	Michaelis–Menten	$y(t) = y_m \dfrac{t}{t^* + t}$	y_m: saturation value t^*: half-saturation

1. What are reasonable ranges for the argument t? Consider the functional behavior at very low and large t and choose reasonable x-and y-limits accordingly. Start with the parameters to $y_m = y_0 = c = 1$; $\alpha = -1, 1, 3$; $t^* = 0, 1, 5$.
2. Plot the functions in the table using one or several script(s)
3. Add 1–3 curves to each graph for different parameter values
4. Repeat the last task by implementing a loop, i.e. using `for`.
5. Add legends and proper labels; also add/invent units.
6. Re-adjust plot settings such as line thickness, outer frame or font size.
7. Use a logarithmic y-axis. When do you gain new insights?
8. Produce a graphical output file in a standard device format (e.g., eps, pdf, png). Check its quality. Embed the graph into your favorite text processing program (WORD, latex, OpenOffice...)
9. Select one function from 1-5 and write a short article (=one page incl. 1-2 figures), about possible interpretations and provide one or two scientific example applications.

Exercise 3: Differential calculus and simple numerics

The derivative

$$\frac{\mathrm{d}y}{\mathrm{d}t}: \quad \text{"derivative of } y \text{ with respect to } t\text{ "} \tag{1}$$

tells you how much a function $y(t)$ changes when varying the argument t. In contrast, the integral

$$\int_{t_0}^{t_1} y\,\mathrm{d}t: \quad \text{" integral of } y(t) \text{ over } t\text{ "} \tag{2}$$

gives the accumulated sum of $y(t)$ over an interval $t_0 \ldots t_1$. Each calculus application is the reciprocal of the other: the differential change in the accumulated sum again gives the original function:

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_t y(t')\,\mathrm{d}t' = y(t) \tag{3}$$

This is also apparent from the numerical scheme of differentiation and integration, where we employ a small but finite step size Δt:

$$\frac{\mathrm{d}y}{\mathrm{d}t} \simeq \frac{y(t+\Delta t) - y(t)}{\Delta t} \tag{4}$$

$$\int_{t_0}^{t_1} y\,\mathrm{d}t \simeq \sum_{t=t_n} y(t)\cdot\Delta t \quad \text{with} \quad t_n = t_0, t_0+\Delta t, t_0+2\Delta t, \ldots t_1 \tag{5}$$

Now let us assume that the derivative of the variable y with respect to time t is known. The analytical form of this differential – denoted as $f(y)$) – contains our mechanistic understanding of what drives the dynamics of y. In biology, the function $f(y)$ mostly depends on the variable itself (in general we have many of them) and on environmental conditions, here denoted as E.

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(y, E) \tag{6}$$

The resulting ODE Eq. (6) can be regarded as the fundamental form of almost all process oriented, dynamical models in environmental science. How can we solve it numerically? Please rewrite Eq. (4) to get an expression form

$$y(t+\Delta t) = \ldots \tag{7}$$

Exercise 4: Exponential spread of a viral disease

The number of infected inhabitants N within a country (e.g., Germany) changes over time t dependent on the mean transmission rate β and "recovery" (including deceasing) rate α. For the Corona virus disease (COVID-19) during its initial wave in early 2020, the recovery period appeared to be about 14 days, thus α=0.07d^{-1}.

At time t=0 we assume immigration of 10 infected individuals (e.g. from returning from holidays, family or business meetings). The following spread of the virus and the concomitant change in N can be described by an ordinary differential equation (ODE):

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \beta N - \alpha N = (\beta - \alpha)N \tag{8}$$

1. Explain the origin of this ODE. Why is the right hand side linear in N and not, for example, given by another function such as N^2 or $1/N$?
2. What are the units of variables and parameters?
3. Please write down an analytical solution $N(t)$ of the ODE using one of the functions discussed before.
4. Plot the solution for three different values of the transmission rate β (β=0.03d^{-1}, β= α, β=0.17d^{-1}) for 120d. Please also switch to a logarithmic scale of the y-axis.
5. Based on our simple mathematical infection model we hindcast the number of infected inhabitants in Germany. For simplicity, we assume a uniform spread through the country.

   Plot the trajectory of German case numbers $N(t)$. Estimate the time period until the entire population of 82 Mill. is infected. Calculate the time period in which one of our course is –statistically– infected. Use both graphical and analytical methods.

Primary production in the water column

Primary production by photosynthesizing unicellular organisms (phytoplankton) fuels the global carbon cycle, and enables most of aquatic life. At times, excessive production can harm humans or lead to environmental degradation. Understanding the dependency of photosynthesis on driving factors is thus highly relevant in environmental research.

Exercise 5: Photosynthesis curves (”Light response”)

The (biomass-specific) rate of gross primary production P by phytoplankton responds in a non-linear way to variations in photosynthetically active radiation (PAR), here denoted as I. $P(I)$ is zero in the dark ($I = 0$), and approaches the maximal rate P_m at high PAR – if we neglect the process of photo-inhibition.

1. From the above function table, choose an appropiate function to describe the light response $P(I)$. Plot it for maximal photosynthesis rate P_m=2d^{-1} and PAR interval $0 \leq I \leq 300$ µEm^{-2}s^{-1}.

2. For comparison, we also construct a piecewise linear (”ramp”) response curve

$$P = P_m \cdot \begin{cases} \dfrac{I}{2K_I} & : \quad \text{if} \quad I < 2K_I \\ 1 & : \quad \text{else} \end{cases} \tag{9}$$

   with half-saturation ”constant” K_I (here 25 µEm^{-2}s^{-1}).

3. Visually and numerically compare the ramp function to the smooth ”original”. Plot both functions with identical parameters into one graph. Calculate the maximal relative deviation between the two response functions.

Exercise Exercise 6a: Light attenuation in the water column

The flux density of PAR within the water column decreases with depth z, from the surface value $I(z = 0) = I_0$=200 µEm^{-2}s^{-1} to smaller values $I(z)$. Very similar to Eq. (8) in the preceding exercise, the change of PAR I with depth z can be described by the ODE:

$$\frac{\mathrm{d}I}{\mathrm{d}z} = -\epsilon I \tag{10}$$

where ϵ is the attenuation coefficient.

1. List the units of variables and parameters.
2. Plot the solution for two different values of the attenuation coefficient ϵ along the water column from the surface ($z = 0$) to the bottom ($z = H$). Consider an apprehensive choice of the x- and y-axis. What is the light intensity $I(H)$ at the bottom?

Exercise 6b: Integral primary production

Now we calculate the photosynthesis rate P Eq. (9) as a function of PAR. Again, set the PAR half-saturation "constant" K_I to 25 µEm^{-2}s^{-1}) and maximal photosynthesis rate P_m=2d^{-1}.

1. Using your solution from Eq. (10), plot the depth dependent primary production $P(z)$ for a water column of H=15m depth, with ϵ=1 m^{-1}. How large is the gross primary production rate at the bottom?

2. Calculate the integral primary production $\overline{P} = \int_0^H P(I(z))\mathrm{d}z$ using an analytical solution, which can be retrieved by combining Eq. (10) and Eq. (9)

$$\begin{aligned}
\overline{P} = & \int_0^{Z_1} P(I(z))\,\mathrm{d}z & + \int_{Z_1}^{H} P(I(z))\,\mathrm{d}z & \\
= & \; P_m \int_0^{Z_1} \mathrm{d}z & + \frac{P_m}{2K_I}\int_{Z_1}^{H} I(z)\,\mathrm{d}z & \\
= & \; P_m \cdot Z_1 & - \frac{P_m}{2K_I}\int_{Z_1}^{H} \epsilon^{-1}\frac{\mathrm{d}I}{\mathrm{d}z}\,\mathrm{d}z & \\
= & \; P_m \cdot Z_1 & - \frac{P_m}{2\epsilon K_I}\cdot \Big(I(H) - I(Z_1)\Big) & = P_m \cdot \Big(Z_1 + \epsilon^{-1} - \frac{I(H)}{\epsilon I(Z_1)}\Big) \qquad (11)
\end{aligned}$$

$$\overline{P} = P_m\,\epsilon^{-1}\cdot\Big(\epsilon Z_1 + 1 - \frac{I(H)}{I(Z_1)}\Big) \qquad (12)$$

3. What is the difference between the integral production of the 15m-column and the one of a deep site with H=150m. So, how can we further simplify the analytical result in Eq. (12) ?

4. Plot the integral primary production $\overline{P}$, again as a function of depth H and of attenuation ϵ using different values of K_I (e.g., K_I=5, or 40 µEm^{-2}s^{-1}).

Exercise 7: Working with NetCDF data

1. In addition to your own *R*-scripts, we recommend to employ existing software for quickly assessing spatial data provided as netCDF. If not existing on your machine, try to install one of the following netCDF-viewers:

   - Panoply (`https://www.giss.nasa.gov/tools/panoply`)
   - Ncview (if you have a linux terminal) (`http://meteora.ucsd.edu/p̃ierce`)
   - SNAP toolbox (`http://step.esa.int/main/toolboxes/snap`)
   - BEAM (`http://earth.esa.int/beam`)

   Please don't persist in installation attempts that need more than 10 minutes. As said, these tools are nice to have, but not necessary.

2. Implement access to a storage folder for (netCDF) files larger than 10MB. Also create a CLOUD backup of this folder for facilitating your shared team work.

3. You may need to install few *R* packages "ncdf4' and "plot3D" (see `install.packages('..')` in Ex.3).

Exercise 7b: Map the North Sea bathymetry

Learn more about the functions `nc_open`, `ncvar_get`, and `image2D`.

- Start to write a generic function that opens a netCDF file and reads all variables known from the structural information in the netCDF "header" (or a subset of those).
- Use this script/function to load the bathymetry of the southern North Sea (SNS) from the netCDF file `esacci_depth.nc`.
- Develop another generic function that visualizes 2D data matrices. This function should include the value range as input and be able to save the result in a graphics format (e.g., PNG, PDF)
- How does the content of `esacci_depth.nc` look like? What is an appropriate choice of the value range? Describe the major depth gradients. Compare the depth of the SNS range with that of other coastal areas. How did Northern Europe look like in the Pleistocene? Can you see two old mega-estuaries?
- Shrink the map size for minimizing the amount of white landmask pixels. You may skip the English Channel and northern parts of the domain.

Primary production of the German Bight

Exercise 1: Light gradients in the coastal ocean

1. Download remote sensing products provided by ESA-CCI from http://www.esa-oceancolour-cci.org Navigate from https://www.oceancolour.org to OPenDAP, CCI ALL-v3.1-DAILY, NetcdfSubset to the site https://www.oceancolour.org/thredds/ncss/grid/CCI_ALL-v3.1-DAILY/dataset.html. Looking aside during the intermediate steps provides an insight into the richness of available ESA products and data formats.

   Select chlorophyll "chlor_a" and water attenuation "kd" (ϵ in our notation) from the variable list. Focus on the southern North Sea (SNS) and cut the data within the geo-coordinates of our area of interest

   - min. and max. longitude: 0.52 °E – 9.78 °E
   - min. and max. latitude: 50 °N – 58.03 °N

   The time period should reflect the available bandwidth of today. In case of fast internet connection, load all time slices; alternatively you may start with the spring of 2003.

2. Re-build a script/function that opens and reads the remote sensing data from netCDF files. The function returns a list/structure that contains a number of fields such as 'latitude',

'longitude', or 'chl'. You should understand how your script operates on netCDF-type input data and how to modify variable names or dimension of the data.

By adjusting the plotting script, visualize light attenuation in the SNS. Describe major gradients in water transparency verbally.

3. "Ligth penetration depth" (D, sometimes called "Secchi depth") is defined as the inverse of total light attenuation ϵ:

$$D = \epsilon^{-1} \tag{1}$$

What is the unit of D? Please describe its characteristic lateral gradients.

4. Perform a point-wise comparison between water depth H and water clarity (D) in a scatter plot. Compute the linear statistics (e.g., by using the Linear Model `lm()` function) and thereby retrieve a simple linear relationship

$$D = D_0 + D' \cdot H \tag{2}$$

Depending on the size (time span) of the netCDF file, the scatter plot may become non conclusive, which motivates the usage of a 2D histogram such as `hexbin`.

5. (For specialists): a close look reveals a non-linear relation between D and H. Which would be a better fitting function, then?

6. Collect all coefficients D_0 and D' in class. Do we see interannual or seasonal differences? Are these significant? What is the resulting average representation of $\epsilon(H)$ ("climatology") ?

Exercise 2: Simulating phytoplankton growth in space and time

1. Plot and compare a series of consecutive satellite CHL scenes during spring (days 80–130) for an arbitrary year. Repeat this for few other years. A list of promising candidates will be disclosed in class.

2. Focus on one period/year and select a starting date t_0. Plot both CHL and ϵ for days relatively free of clouds.

3. Calculate surface PAR I_0 using the script `surface_PAR.R` depending on the day(s) of year. What about shading by clouds?

4. Retrieve your function for vertically averaged relative photosynthesis rate $\overline{P}$ depending on depth H, surface PAR I_0 and attenuation ϵ (Ex-Eq. (12)). Calculate $\overline{P}$ for the starting date t_0 over the entire domain.

5. Choose one interval with two (or more) consecutive good scene days and extrapolate from the observed CHL field at time t_0 to the the CHL distribution at time $t_0 + \Delta t$ in the spirit of Ex-Eq. (**??**)

$$\mathrm{CHL}(t_0 + \delta t) = (1 + \overline{P} \cdot \delta t) \cdot \mathrm{CHL}(t_0) \tag{3}$$

   Note the difference between Δt and δt. Compare and discuss the two settings δt=Δt and δt=$\Delta t/50$. Plot the simulated phytoplankton distribution CHL$(t_0 + \Delta t)$.

6. Compare CHL$(t_0+\Delta t)$ with the observed distribution at $t_0+\Delta t$, first visually then numerically by creating the difference plot. Can the differences be lowered by changing a parameter such as P_m?

7. Discuss the differences between predicted and observed surface CHL. For example, is primary production identical to the relative growth rate of phytoplankton? List the major simplifying assumptions.

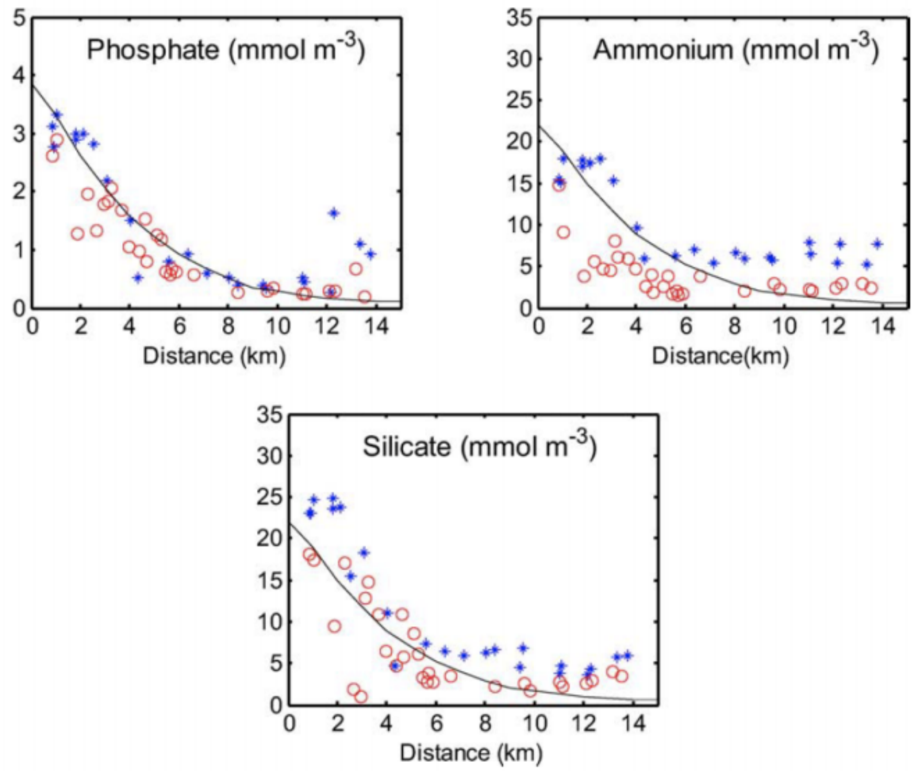
8. If time allows, repeat the simulation for another period.

9. Propagate the phytoplankton distribution further in time, by looping over Eq. (3) from $t_0+\Delta t$ to $t_0+2\Delta t$, $t_0+3\Delta t$ and so forth. What happens? Is this realistic? Which important processes or factors are missing?

10. If still time allows, you may also use the averaged reconstructed light field based on bathymetry, Eq. (2) , for calculating photosynthesis. Where in the coastal ocean does this affect the accuracy of the prediction?

Exercise 3: Nutrient limitation of primary production

Here we add an important growth limiting factor besides light, which is nutrient limitation. Nutrient concentration in the GB decreases with increasing distance to the coast, as seen in the data presented by Floeser et al 2002 and Ebenhoeh et al 2004:



Assuming a linear increase of depth with distance, we have

$$N(H) = N_0 \cdot H_0/(H_0 + H) \tag{4}$$

with H_0=15m and as a proxy for the depth dependence in nutrient limitation

$$f_N(H) = K_N/K_{N0} \cdot N(H)/(N(H) + K_N) \tag{5}$$

with unknown nutrient saturation constant K_N and a scaling parameter K_{N0}.

1. Plot nutrient limitation f_N over depth H for different choices of the saturation constant (e.g., $K_N = N_0$, K_N very large).

2. Combine our previous estimates of 'relative' primary production ($P(H)$, relative to standing stock) in the southern North Sea with the offshore increase in nutrient limitation. Visualize the resulting effect on $P(H)$,

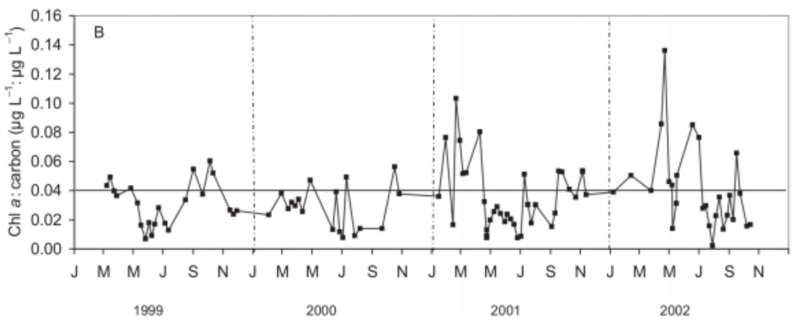
$$P_N(H) = f_N(H) \cdot \overline{P} \tag{6}$$

If we use the bathymetry based averaged primary production (Eq. (2)) for $\overline{P}$, what is the functional form of the result? Where is the phytoplankton growth rate maximal?

Exercise 4a: Total primary production based on data and models

Here we finally combine basic ingredients to estimate the total primary production P_{tot} of the southern North Sea. Calculate and then combine ...

1. the average surface irradiance I_0 for the SNS region (from internet sources and/or our).
2. an annual map for the standing stock of primary producers in units gC m^{-3} using the average CHL:C ratio in the turbid North Sea around 0.04 gCHL/gC estimated from the study of Llewellyn et al 2004:



3. the integral production $\overline{P}_N$ relative to C stock while assuming P_m=1.5d^{-1} and inserting either the "attenuation climatology" $\epsilon(H)$ or a direct long-term average of the downloaded ESA product.
4. optional: base your calculation on seasonal estimates of CHL, CHL:C, light, nutrients, and ϵ

How sensitive is P_{tot} to input uncertainty? Compute effects of changing the conversion ratio (gC/gCHL) or the specific attenuation coefficient D' by 20% each.

Exercise 4b: Global implications of coastal/shelf production

Estimate from internet sources

1. the global total primary production in units of tC/y, and
2. the annual total CO_2 release by all European countries.

Relative to both numbers, what is the ecosystem service of primary producers in the southern North Sea? Assume that 1% of P_{tot} is sequestrated or preserved (thus removed from active cycling) at longer time scales.

Exercise 4c: Reducing eutrophication

Let us assume that the "estuarine" nutrient concentration N_0 reflects anthropogenic loads ('eutrophication')

1. Increase and decrease the "estuarine" nutrient concentration N_0 in Eq. (4) . Visualize the effects on total primary production of the SNS. Why is the response non-linear?
2. Where do we see the largest effect in the coastal ocean? (can be addressed in 2D or depending on H)
3. Repeat the last steps after increasing K_N to a very large value. What happens?
4. Researchers currently propose to increase eutrophication to foster carbon sequestration. Would this kind of geo-engineering be wise? By contrast, some authorities intend to further reduce riverine nutrient input by 50%. Discuss both the economic and the ecological dimension of the two scenarios.