WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Process-based lake modeling in R using GLM (General Lake Model)

Experiment on your lakes safely from home on your laptop

Robert Ladwig, A. Adhlakha, H.A. Dugan & P.C. Hanson

University of Wisconsin-Madison

rladwig2@wisc.edu

@hydrobert

supported by NSF ABI development grant, #DBI 1759865

Who’s who?

Robert, postdoc at
Center for Limnology

Aryan, undergrad in
Computer Science

Hilary, Assistant Prof.
UW-Madison

Paul, Research Prof.
UW-Madison

With very valuable input from Matt Hipsey, Cayelan Carey, Jordan Read, Dale Robertson, Joe Stachelek, Tadhg Moore, Jorrit Mesman and Johannes Feldbauer

Welcome!

- If you want to run the simulations during the workshop, you will need to install the following software on your computer. If you just want to watch, ask questions, and drive from the back seat, that's fine, too!
- **Future questions?** Email Robert at rladwig2@wisc.edu

Two paths to do the workshop examples:

(1) Use the docker: https://hub.docker.com/r/hydrobert/glm-workshop (requires docker)
  (a) this includes Rocker, all packages, all scripts and all data; then open a terminal

```
docker pull hydrobert/glm-workshop

docker run --rm -d  -p 8000:8000 -e ROOT=TRUE -e PASSWORD=password hydrobert/glm-workshop:latest
```

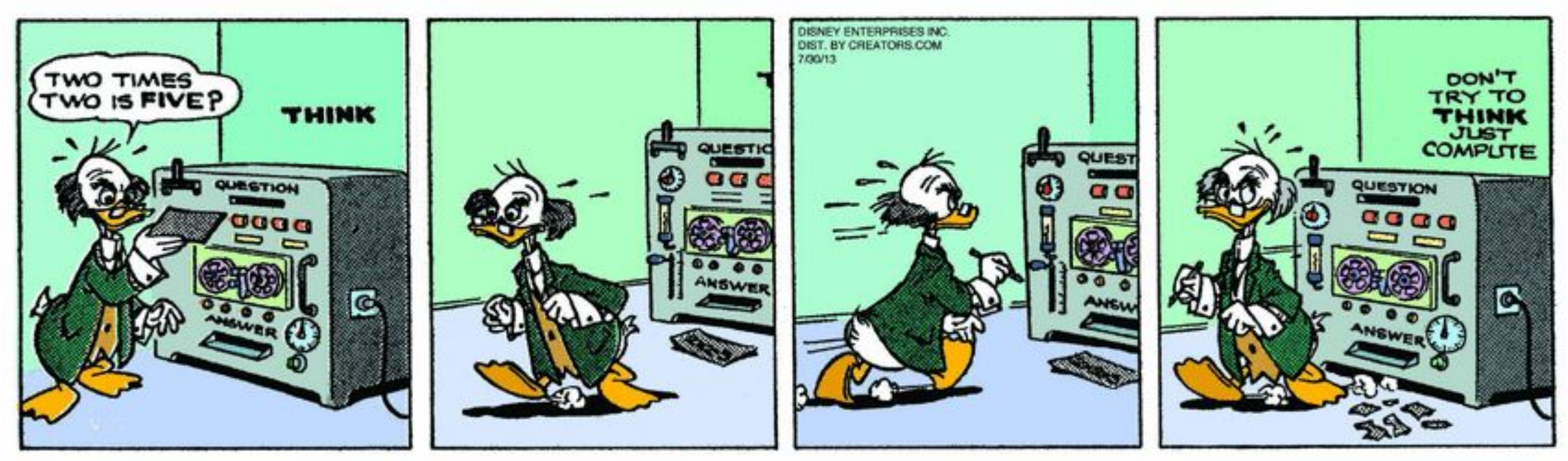
open any web browser and type 'localhost:8000' (user: rstudio, password: password)

(2) Clone or download files from: https://github.com/robertladwig/GLM_workshop
  (a) you'll need R (>= 3.5) and these packages: GLM3r, glmtools, rLakeAnalyzer, tidyverse

Expectations for today

- **Introduction** to process-based modeling intended for all skill levels and background
- **Overview** of concepts and lake modeling theory → **G**eneral **L**ake **M**odel (**GLM**)
- **Coding examples** in R using data from Lake Mendota (thanks to NTL-LTER!)

Great textbook: *Steven C. Chapra (2008) “Surface Water-Quality Modeling” Waveland Pr Inc*



Disney’s Donald Duck comics

Time schedule today

9:30-10:15	**Introduction to process-based lake modeling**	• What is process-based modeling? • GLM theory and applications • AED2 theory and examples for O_2 and C • Overview of R-packages
10:20-11:15	**Using the model in R**	• Running GLM in R • Visualising results • Calibrating water temp. and oxygen parameters • Checking your phytoplankton
11:15-11:30	**Questions and problems**	• Stick around to talk about questions and raise issues

Time schedule today

9:30-10:15	Introduction to process-based lake modeling	• What is pro • GLM theor • AED2 the • Overvie
10:25-11:15	Using the model in R	• Runnin • Visua • Calib • Che
11:15-11:30	Questions and problems	• Stick around t ues

Please ask urgent questions by "raising hand" in Zoom;

Otherwise you can write your question/issue here and we will talk about it at the end

Time schedule today

Time for technical issues

9:30-10:15	Introduction to process-based lake modeling	• What is pro • GLM theor • AED2 the • Overview
10:25-11:15	Using the model in R	• Runnin • Visua • Calib • Che
11:15-11:30	Questions and problems	• Stick around to ... ues

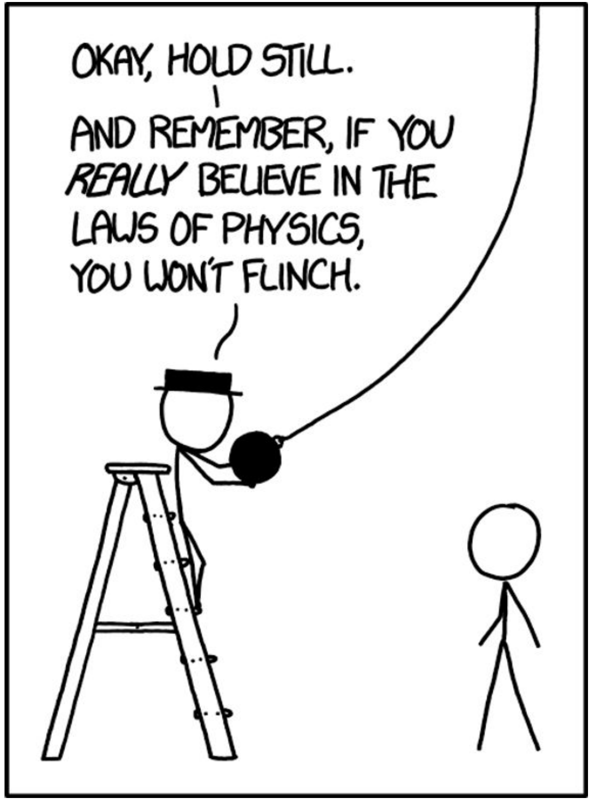
Please ask urgent questions by "raising hand" in Zoom;

Otherwise you can write your question/issue here and we will talk about it at the end

Modeling introduction

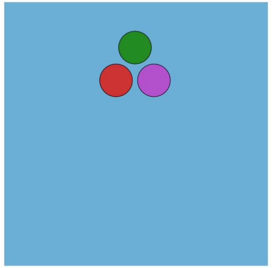
Process-based modeling in a nutshell

- Models are simplified reflections of reality
- Deterministic (events at next time step depend on events from previous time step)
  - and describe processes with mathematical equations
  - that are either based on empirical knowledge or physical principles
- Numerical models need time and space discretization
- As well as initial data (to start from) and boundary data (as driving data)



Process-based modeling in a nutshell

- **Conservation of mass (continuity)**
  - Inflow(s)= outflow(s)
  - Mass cannot be created or destroyed
- **Conservation of momentum**
  - Velocity based on balances of forces (gravity, pressure, friction, earth-rotation)



State of the environment

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = q_0$$

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\nabla\mathbf{v} = -\frac{1}{p}\nabla p + g + \nu\Delta\mathbf{v}$$

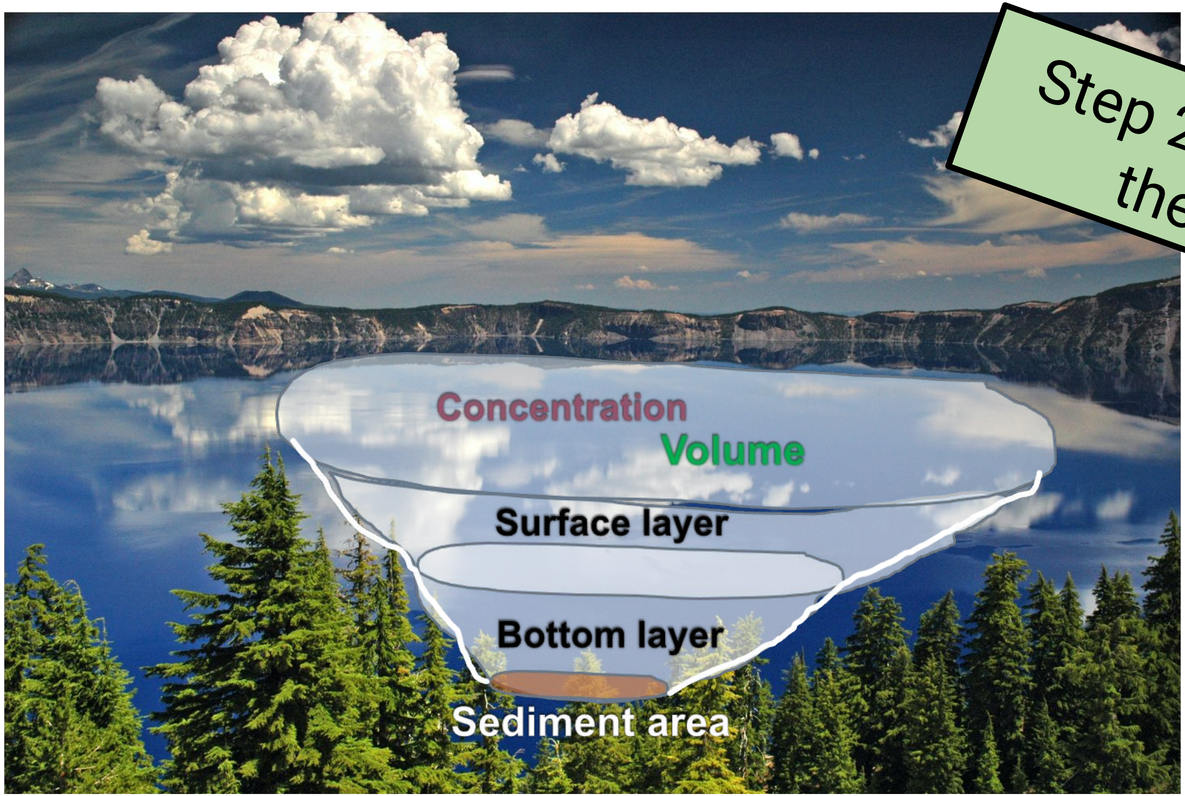
State of the system

CFD animation by WaterLily (https://github.com/weymouth/WaterLily); awesome lake experiment by Colin and Austin

Process-based modeling in a nutshell

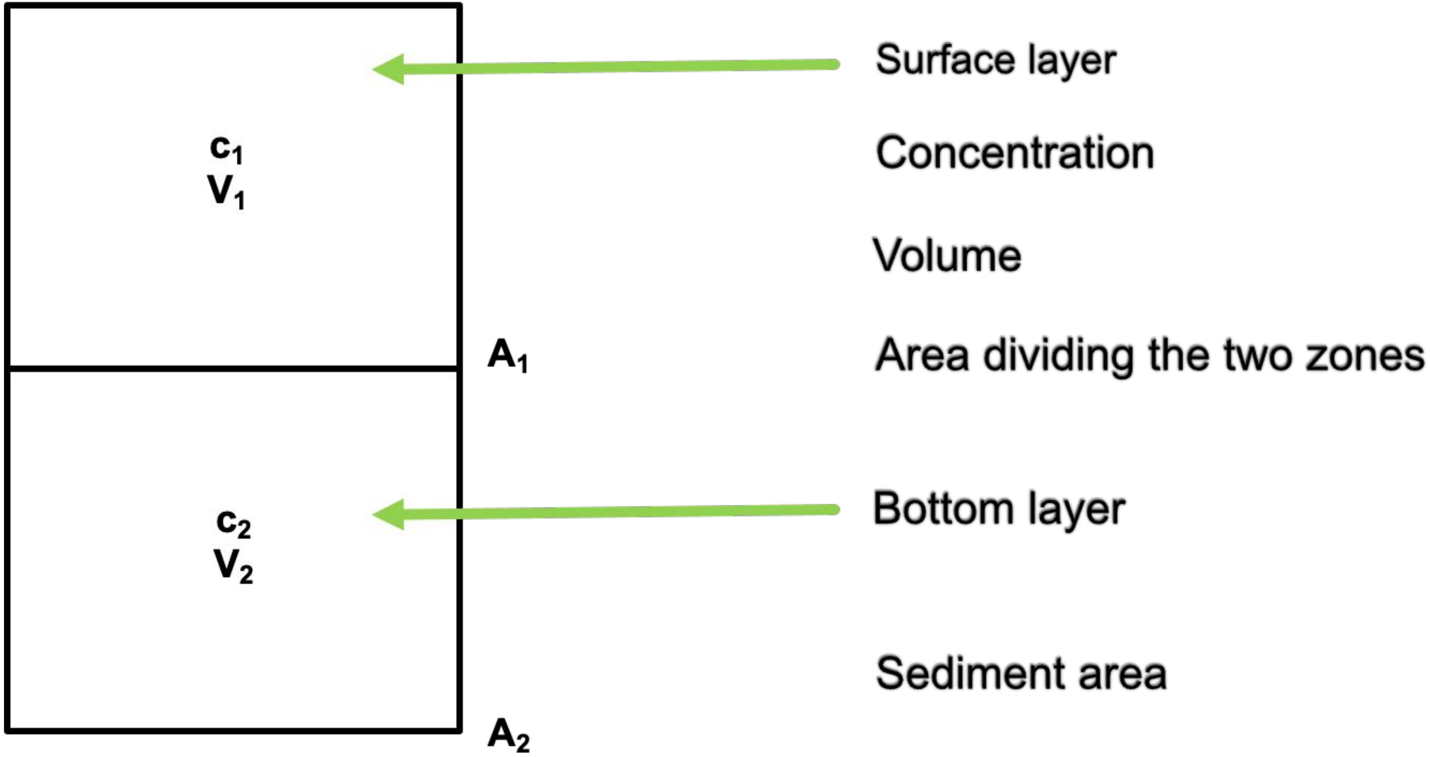
Step 1: imagine a (nice) lake

Process-based modeling in a nutshell



Step 2: abstract the lake

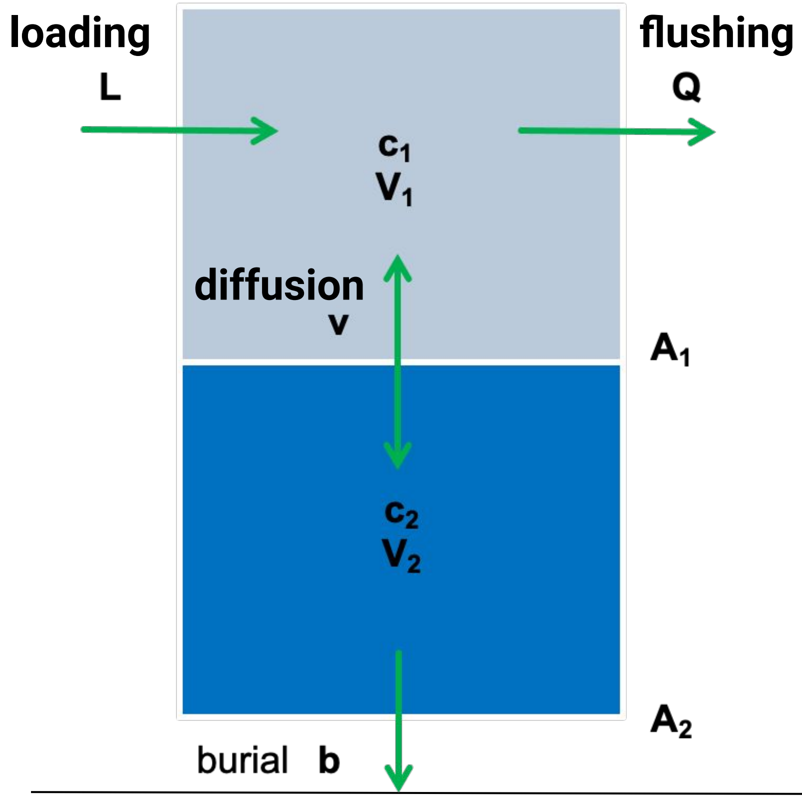
Process-based modeling in a nutshell



most are bookkeeping exercises to keep inflow = outflow (mass-balance)

Process-based modeling in a nutshell

Step 3: add math



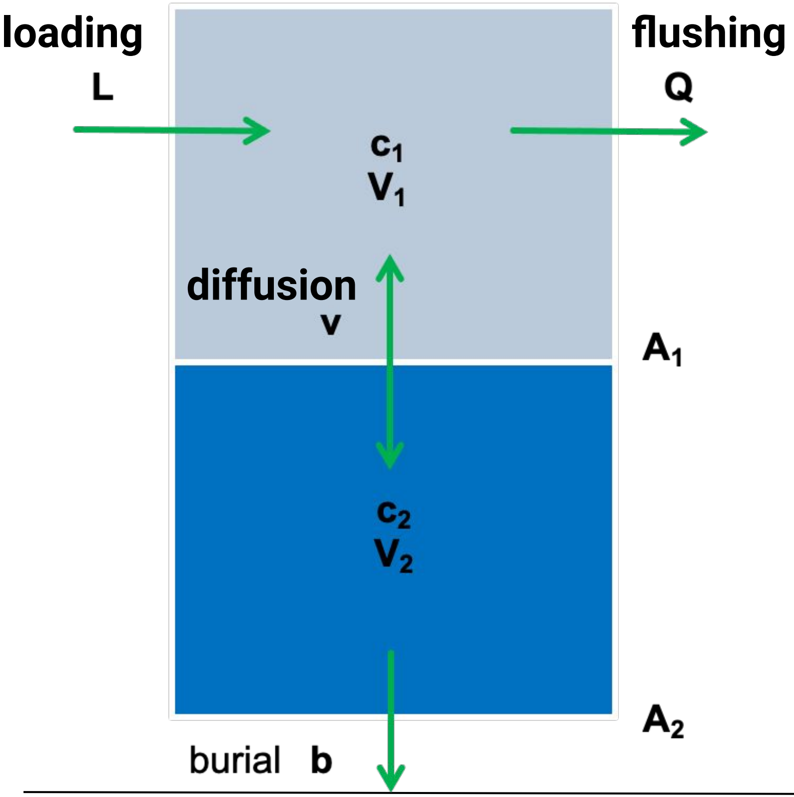
$$\frac{dc_1}{dt} = \frac{L - Qc_1 + vA_1(c_2 - c_1)}{V_1}$$

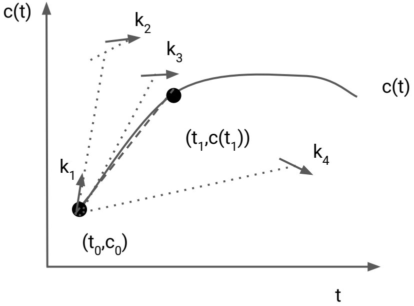
$$\frac{dc_2}{dt} = \frac{vA_1(c_1 - c_2) - bA_2c_2}{V_2}$$

most are bookkeeping exercises to keep inflow = outflow (mass-balance)

Process-based modeling in a nutshell

Step 4: solve with numerics, e.g., Runge-Kutta





$$c_{i+1} = c_i + f(c_i, t_i) * \Delta t \; h = \Delta t$$

$$c_{i+1} = c_i + [\tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)]\Delta t$$

$$k_1 = f(t_i, c_i)$$

$$k_2 = f(t_i + \tfrac{1}{2}h, c_i + \tfrac{1}{2}hk_1)$$

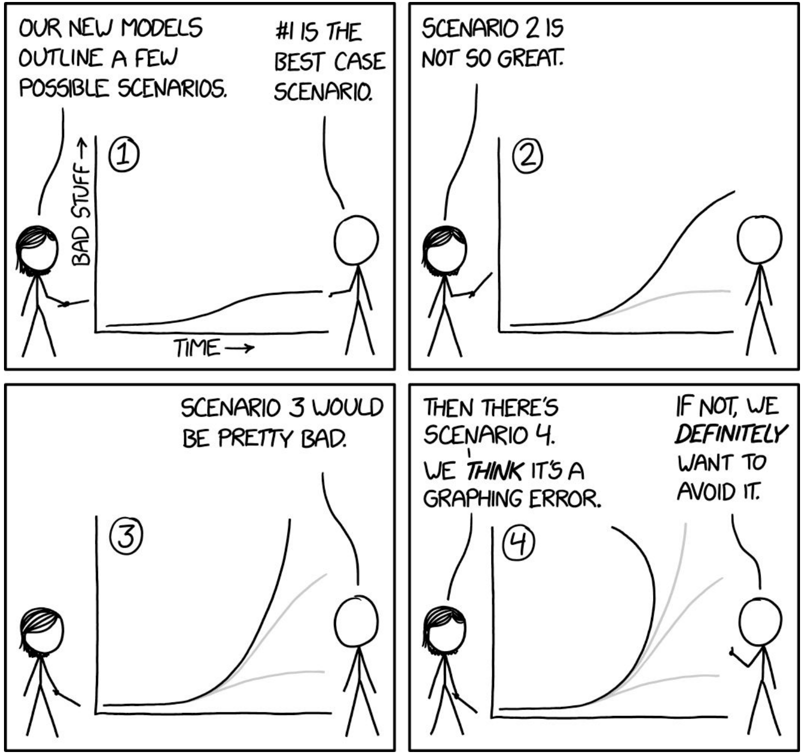
$$k_3 = f(t_i + \tfrac{1}{2}h, c_i + \tfrac{1}{2}hk_2)$$

$$k_4 = f(t_i + h, c_i + hk_3)$$

most are bookkeeping exercises to keep inflow = outflow (mass-balance)

Process-based modeling in a nutshell

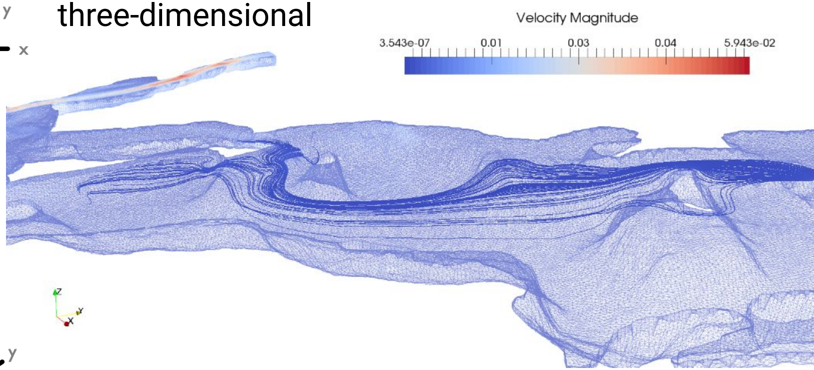
- Why should you use a model?
- Explore ideas regarding ecological systems that may not be possible/feasible to field-test for
    - logistical reasons
    - political reasons
    - financial reasons
    - physical reasons
- Formulating a model helps to better understand the system and can identify data needs
- “Your model is your hypothesis.” (Kate)



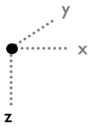
GLM introduction

Multi-dimensional models

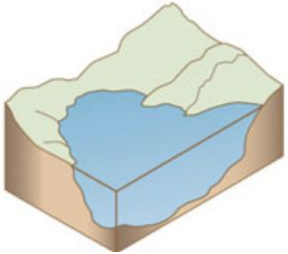
three-dimensional



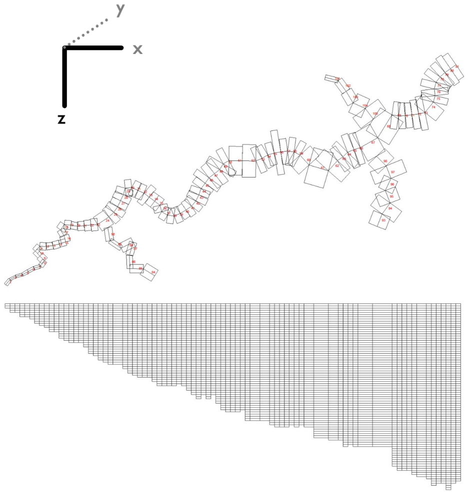
zero-dimensional



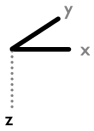
0D: Box model

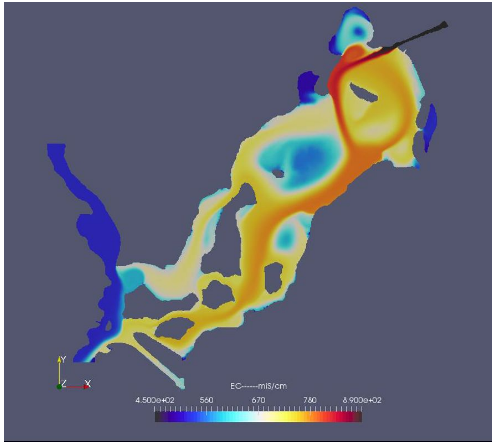


longitudinal-vertical 2D



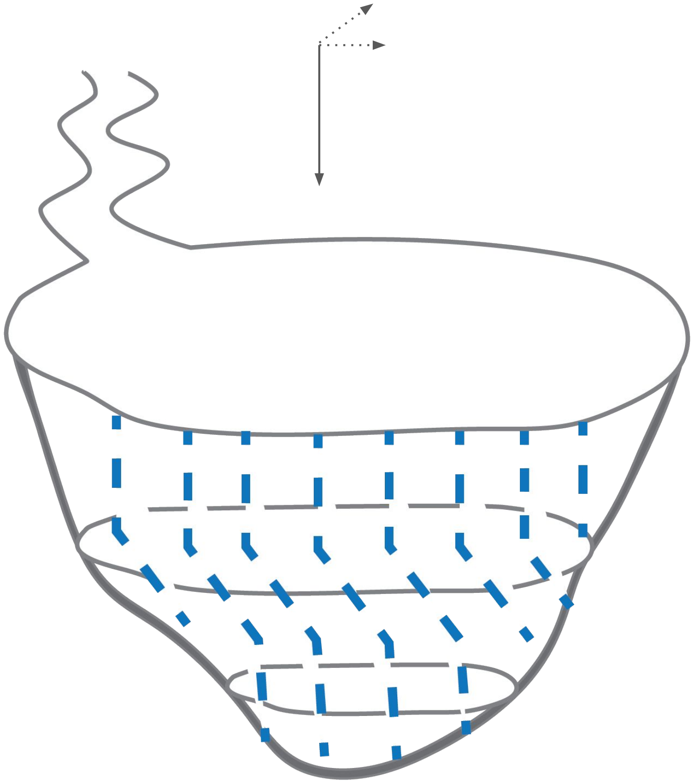
depth-averaged 2D

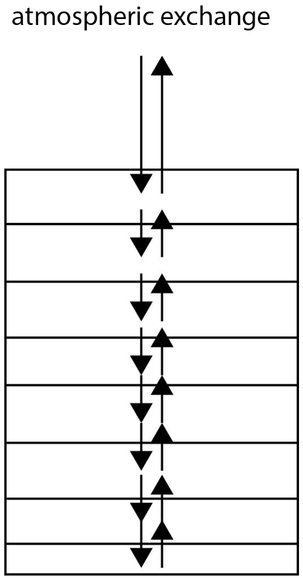


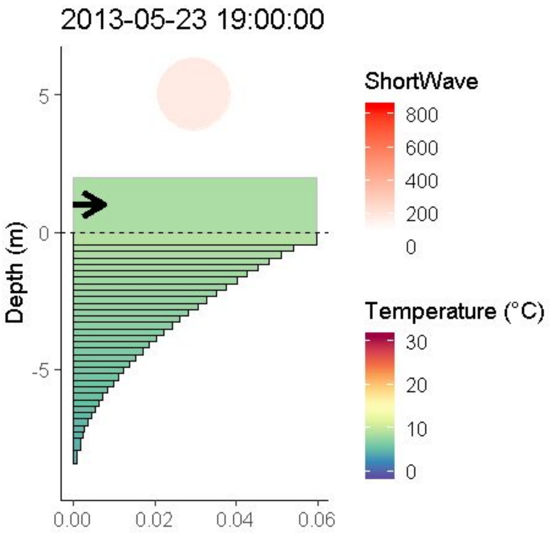


Process-based lake modeling

- **Vertical 1D model**:
  - Horizontal homogeneity
  - No Coriolis force
- Layering over vertical axis
- Every layer = volume of lake
- **Focus on heat transport**







layer animation by Tadhg Moore

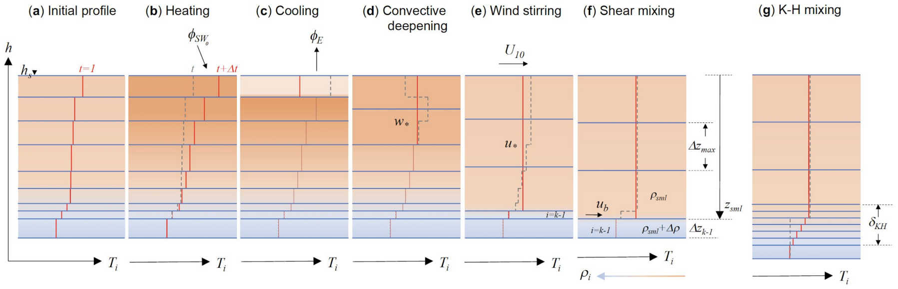
Process-based lake modeling

- Reactive-diffusive transport equation for water temperature
- Turbulence-closure scheme to solve for momentum terms and diffusivities (e.g., k-ε in Simstrat)
- Molecular diffusivity is mostly neglected → approximation of K_z (e.g., MyLake)
- Alternatively, energy-balance approaches (e.g., GLM): potential ag. available kinetic energy

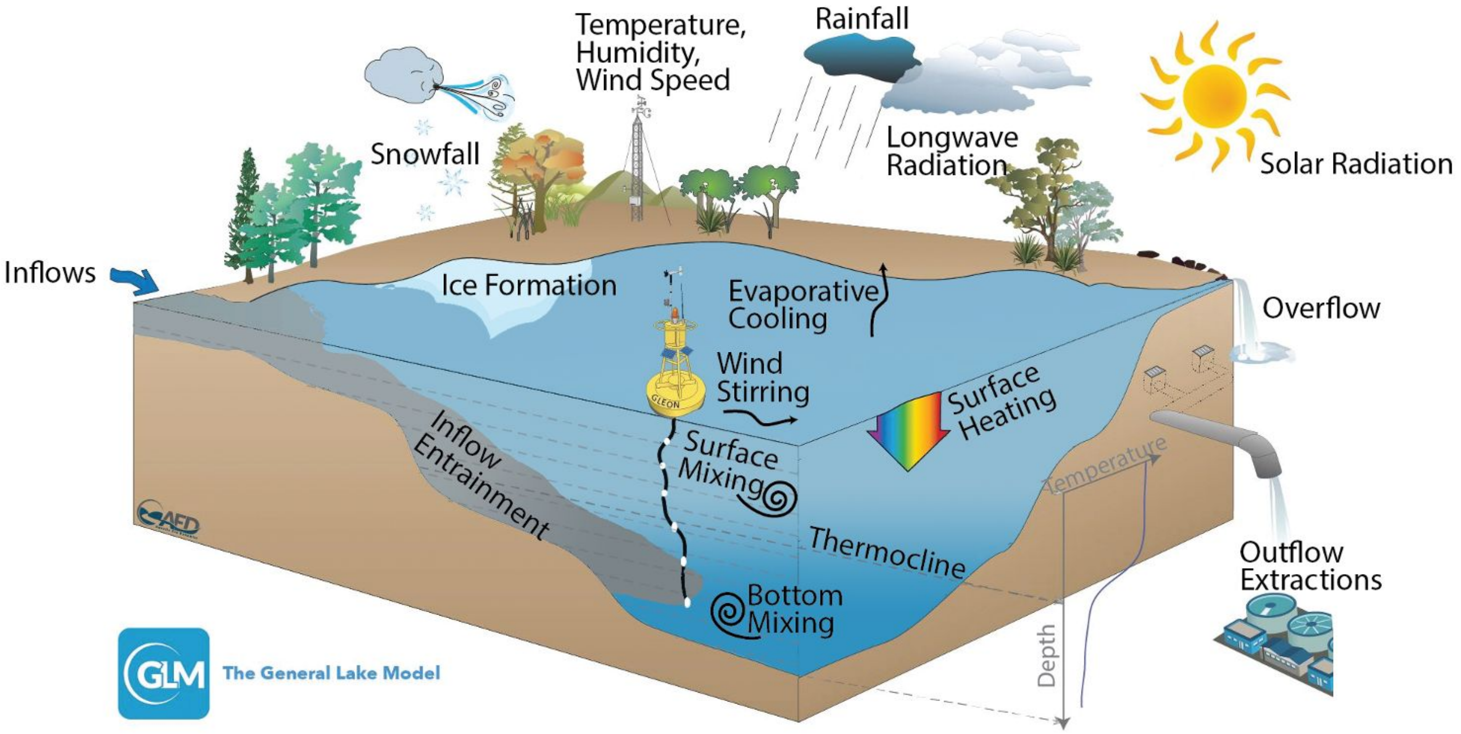
$$\frac{\partial T}{\partial t} = \frac{1}{A}\frac{\partial}{\partial z}\left(A(\nu_t' + \nu')\frac{\partial T}{\partial z}\right) + \frac{1}{\rho_0 c_p}\frac{\partial H_{sol}}{\partial z} + \frac{dA}{dz}\frac{H_{geo}}{A\rho_0 c_p}$$

diffusive transport — atmospheric heating — sediment heating

$$\frac{\partial T}{\partial t} = \frac{1}{A}\frac{\partial}{\partial z}\left(AK_z\frac{\partial T}{\partial z}\right) + \frac{1}{\rho_0 c_p}\frac{\partial H_{sol}}{\partial z}$$



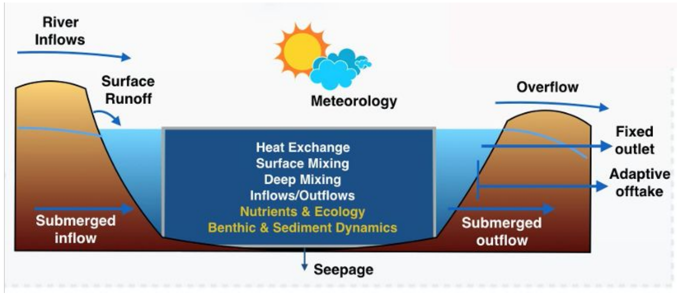
General Lake Model (GLM)



Hipsey et al. 2019

General Lake Model (GLM)

- GLM started as a GLEON project
- Developed by Matt Hipsey, Louise Bruce and Casper Boon at UWA
- Designed to operate with the Aquatic EcoDynamics Model Libraries (AED2) → water quality model
- Written in C and Fortran
- Freely available as open-source program: https://github.com/AquaticEcoDynamics/GLM



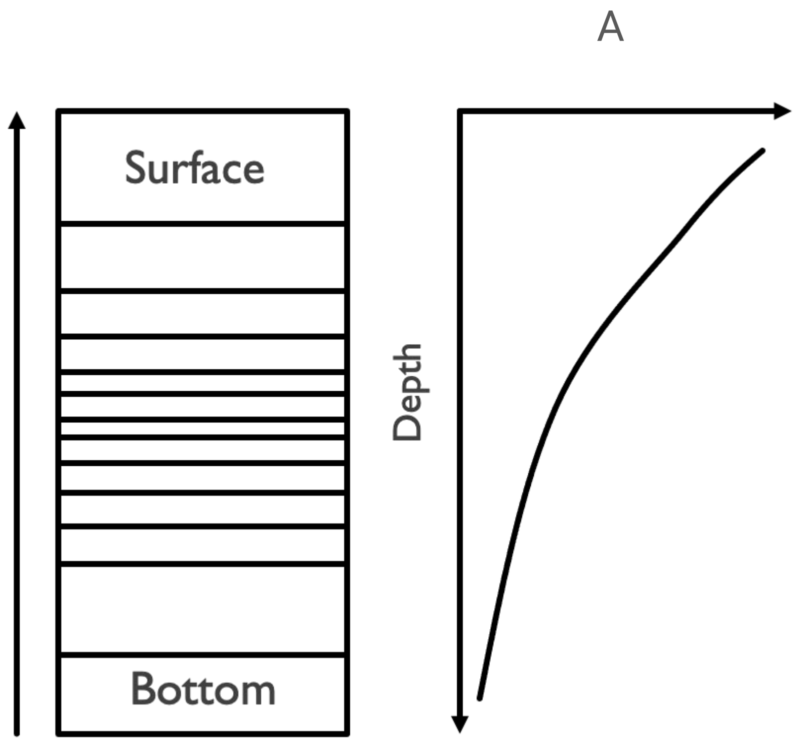
Hipsey et al. 2019

Grid

- Each vertical layer (= control volume) adapts its thickness due to flows, mixing and mass fluxes

  → flexible Lagrangian structure

- Each layer has ‘unique’ density
- When there’s enough energy to overcome density gradient → layers merge
- Layer’s volume changes depending on hypsography

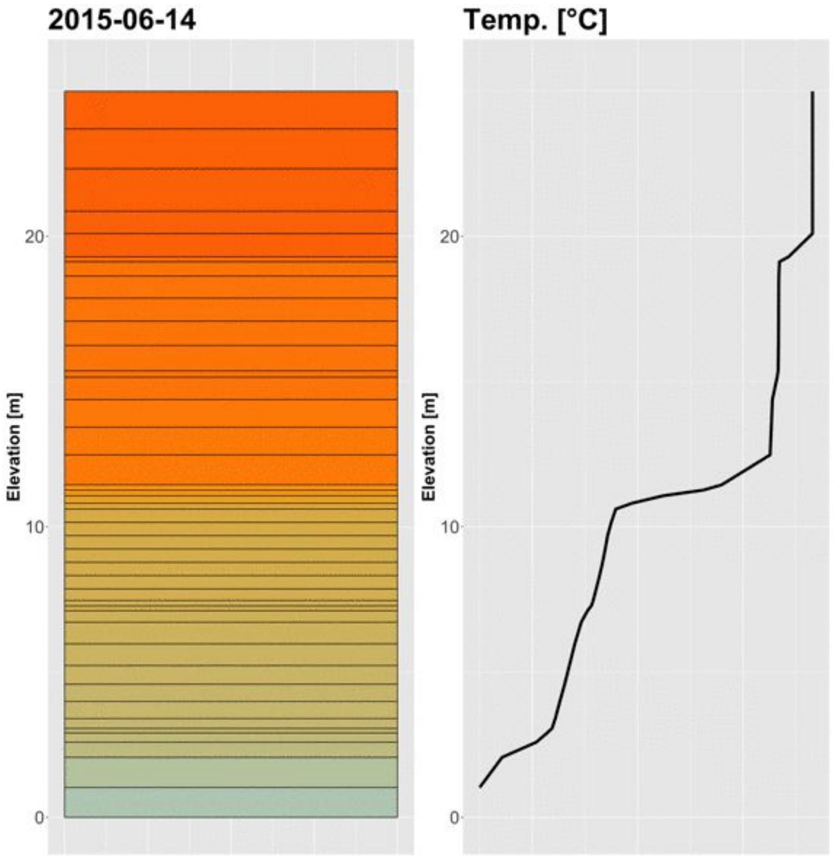


Grid

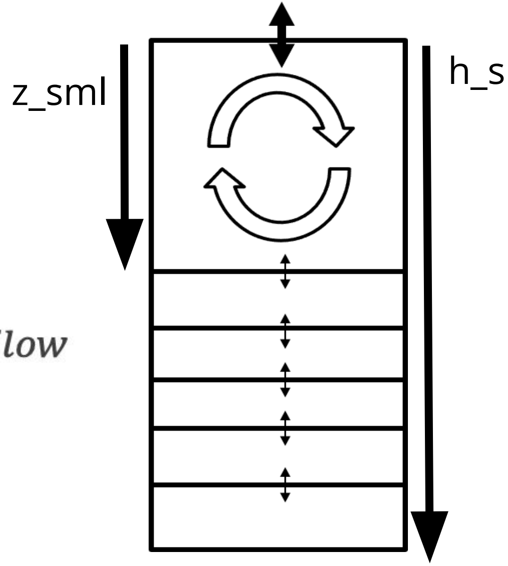
- Each vertical layer (= control volume) adapts its thickness due to flows, mixing and mass fluxes

  → flexible Lagrangian structure

- Each layer has ‘unique’ density
- When there’s enough energy to overcome density gradient → layers merge
- Layer’s volume changes depending on hypsography



Water balance



- Net water flux over total lake (solved by integration over time):

$$\frac{dV_s}{dt} = A_s \frac{dh_s}{dt} + \sum_I^N Q_{inf,I} - \sum_I^N Q_{out,I} - Q_{seepage} - Q_{overflow}$$

Surface layer height

- Code solves first for surface flux changes, then for all other fluxes

Mass balance of surface layer = precipitation + snowfall + run-off – evaporation - ice

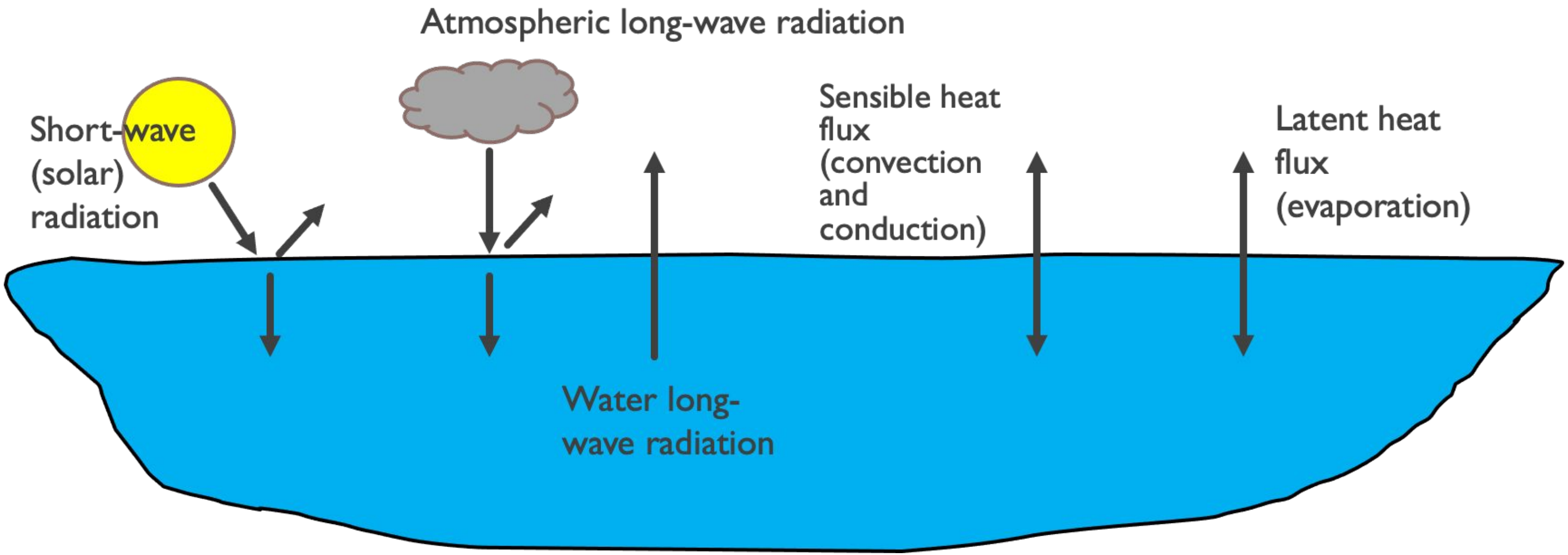
$$\frac{dh_s}{dt} = P + S + \frac{Q_R}{A_S} - E - \frac{d\Delta z_{ice}}{dt}$$

- Mixing only affects thickness of surface layer (z_sml) not its height

Hipsey et al. 2019

Surface energy balance

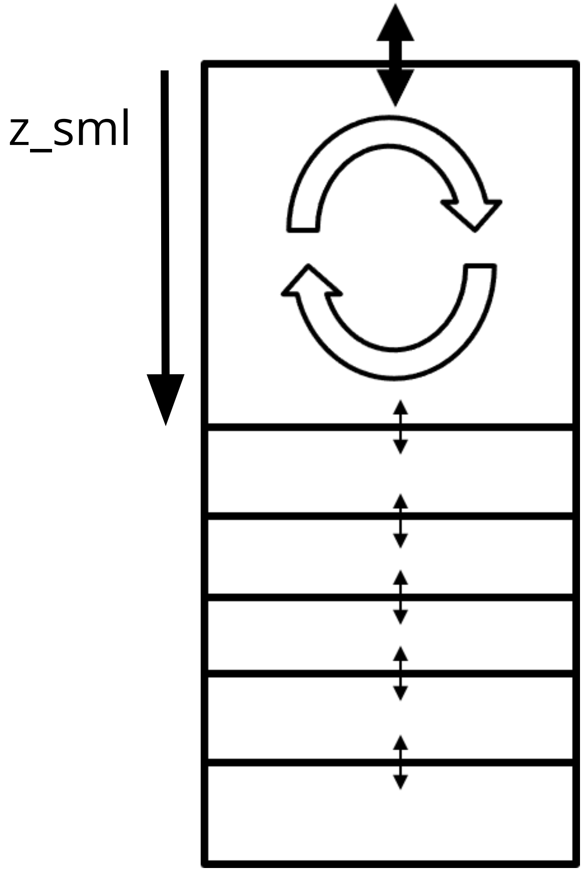
$$\rho C_p z_s \frac{dT_s}{dt} = J_{SW} + J_{LW,in} - J_{LW,out} - J_E + J_H$$



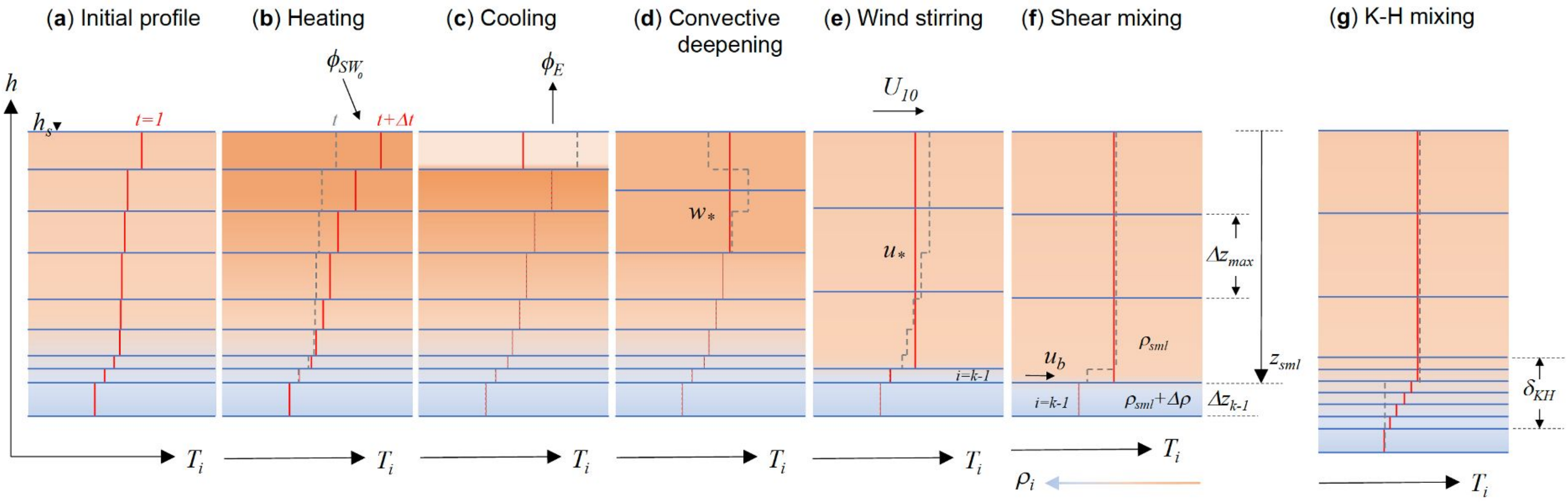
Hipsey et al. 2019

Mixing I

- Surface layer: energy balance approach
  - Energy balance determines thickness of surface mixed layer
  - Below surface mixed layer, turbulent diffusion approach is used
- Here, mixing depends on <u>difference between potential energy and available kinetic energy</u>
- Kinetic energy = convective overturn, wind stirring and shear production/Kelvin-Helmholtz billowings

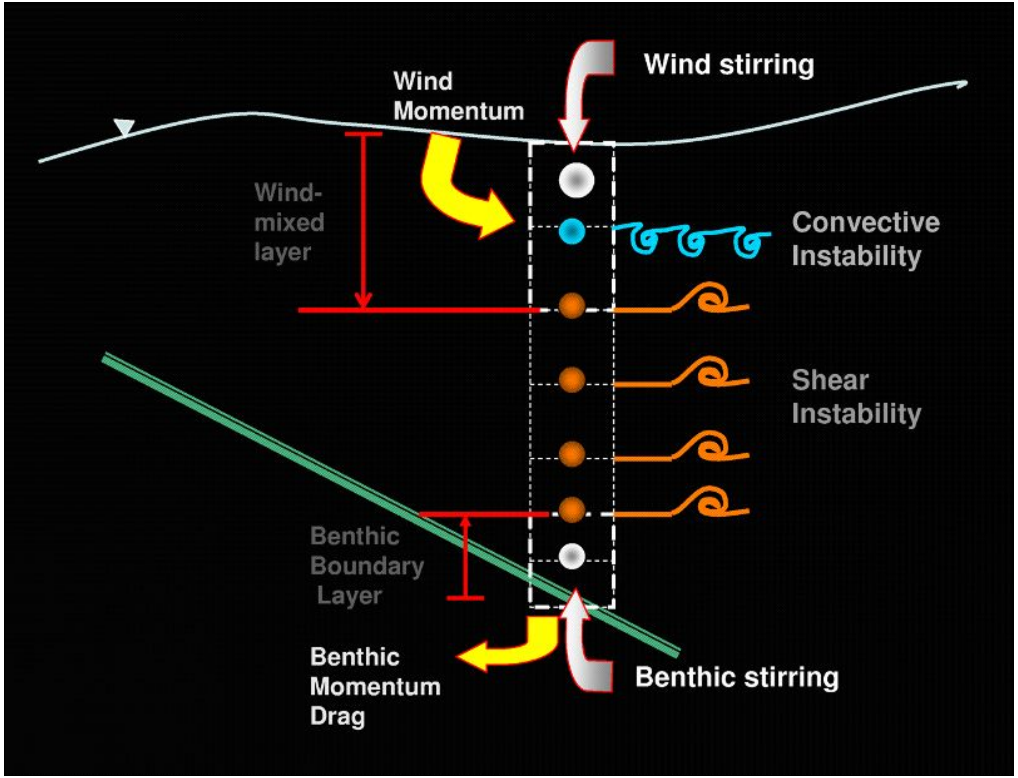


Mixing II



Hipsey et al. 2019

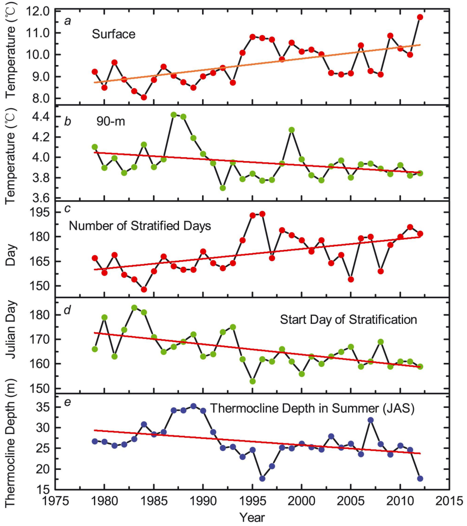
Mixing III



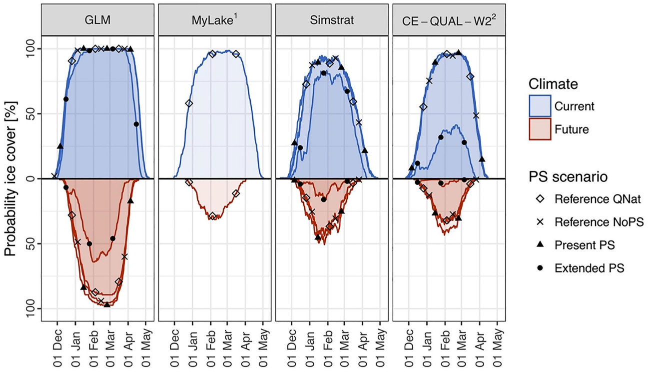
Kelvin-Helmholtz instabilities

Example applications

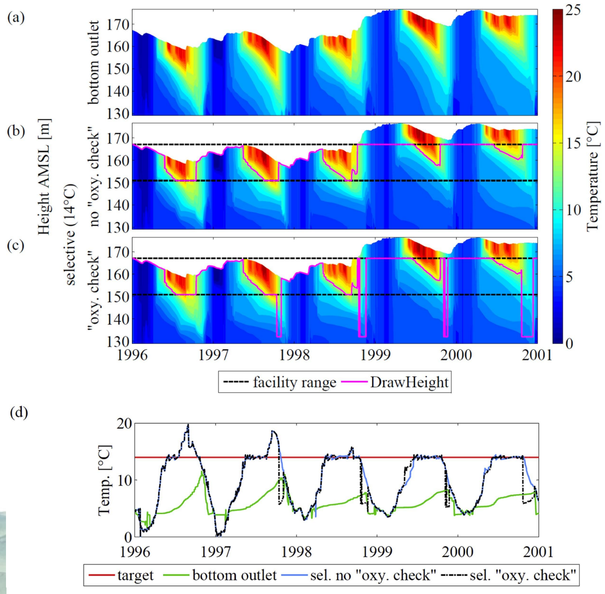
Climate change effects on lakes[1]

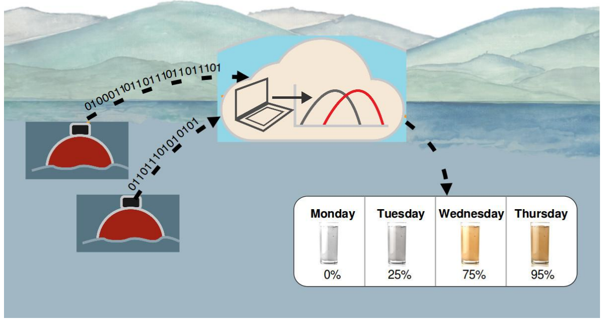


Future ice cover shifts[2]



Optimal withdrawal depth in reservoirs[4]





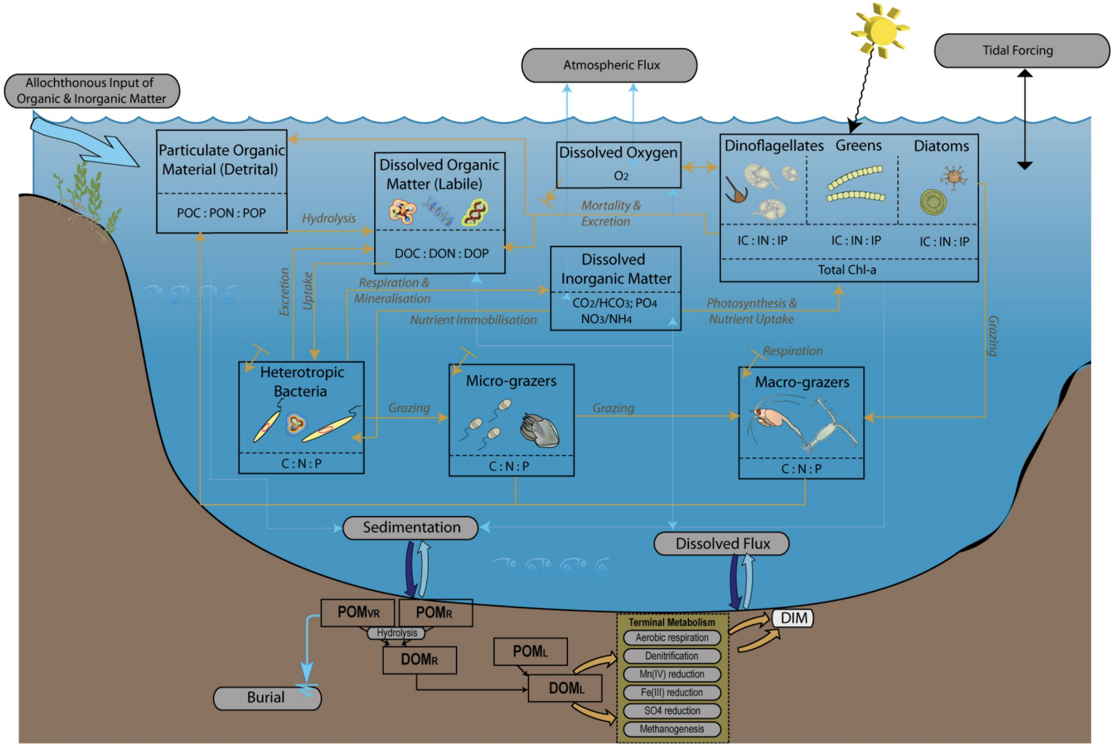
Smart ecological forecasting (FLARE)[3]

[1]Huang et al. 2017; [2]Kobler et al. 2019; [3]Figueiredo, Thomas, Carey & Weathers; [4]Weber et al. 2017

AED2 theory

AED2: water quality model

- Conceptualize your individual ecosystem:
  - Oxygen, nutrients, phytoplankton
  - Sediment biogeochemistry
  - Benthic communities

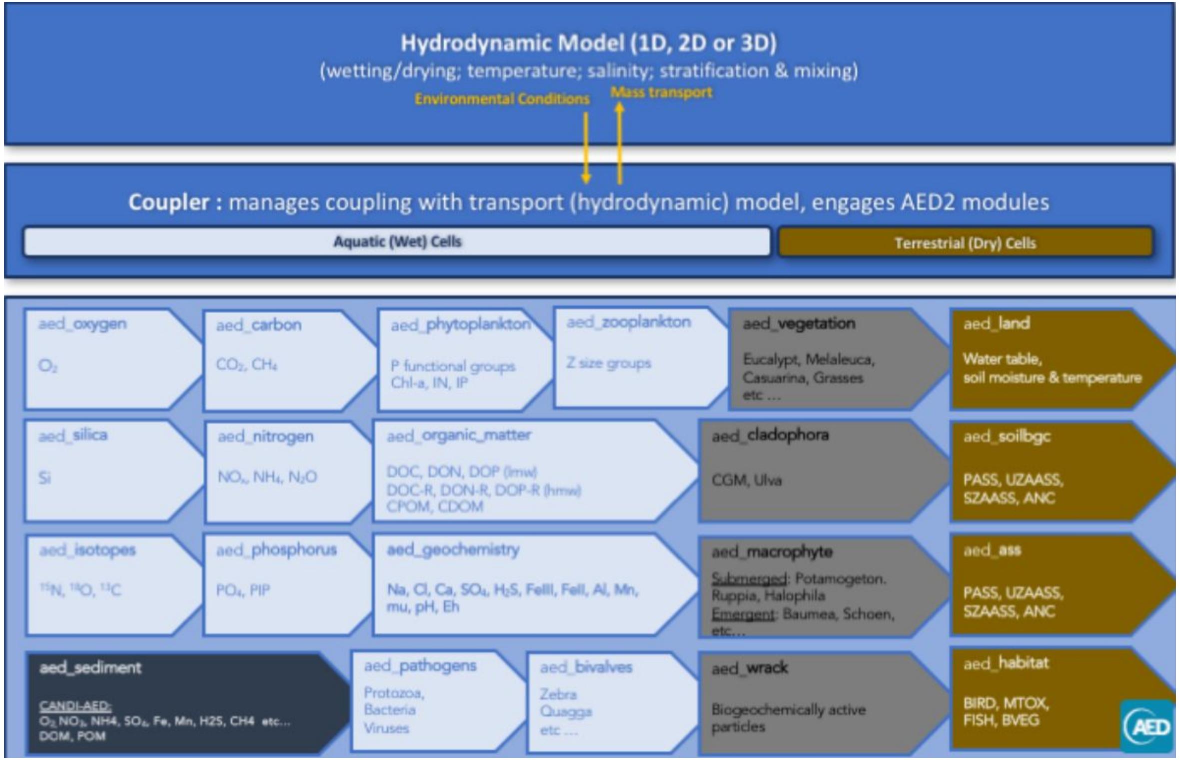


Carbon & Nutrient Flux Pathways

AED2 manual: Hipsey et al. 2013

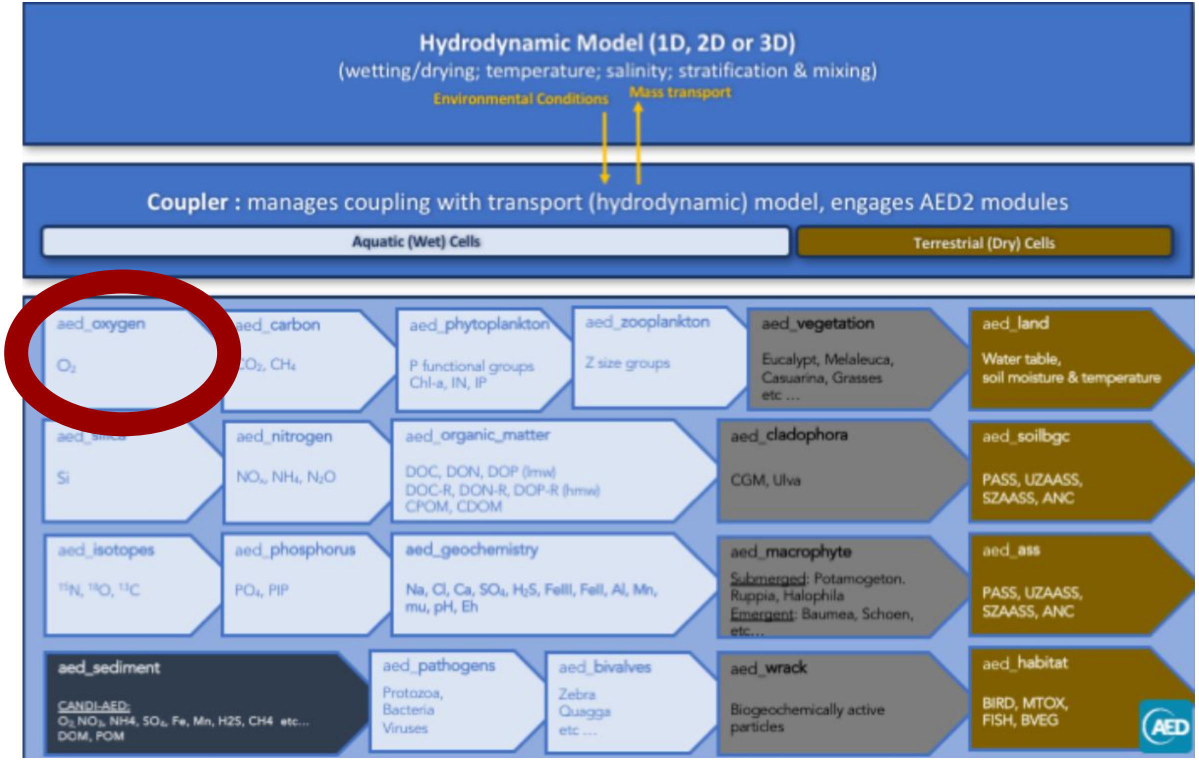
AED2: modularised structure

- Interacts with GLM (two-way coupling)
- Array of modules are available
- Hierarchy of dependencies between modules, e.g., nitrogen reactions can depend on availability of oxygen



AED2: modularised structure

- Interacts with GLM (two-way coupling)
- Array of modules are available
- Hierarchy of dependencies between modules, e.g., nitrogen reactions can depend on availability of oxygen

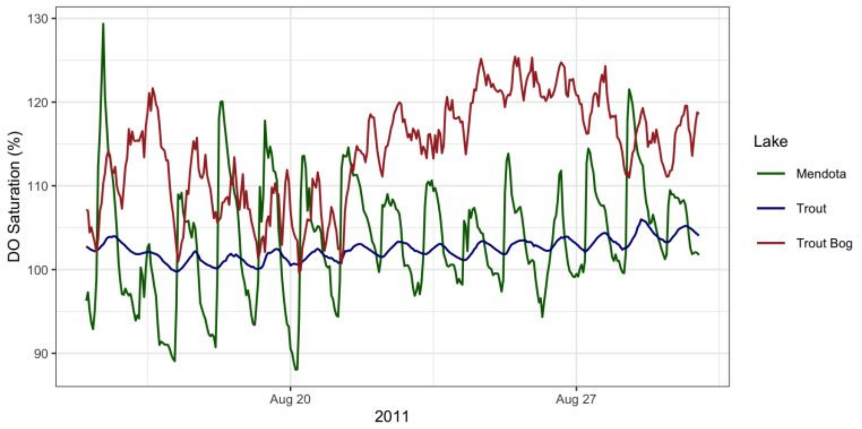


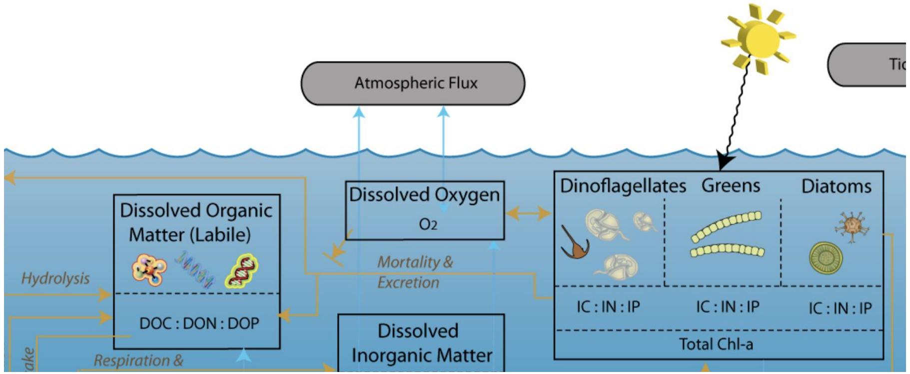
AED2: dissolved oxygen

atmospheric exchange   sediment oxygen demand

$$F_{[DO]}^{total} = k_{O_2}([DO]_{air} - [DO]_{water}) + S_{SOD} f_{SOD}^{T}(T) f_{SOD}^{DO}(DO) +$$
$$F_{mineralisation} +$$
$$F_{nitrification} +$$
$$F_{photosynthesis} +$$
$$F_{respiration}$$

- Fast moving pool (minutes to days)
- Solved over vertical axis
- Plus advection and diffusion
- Photosynthesis/respiration = phytoplankton, zooplankton, seagrass, bivalves

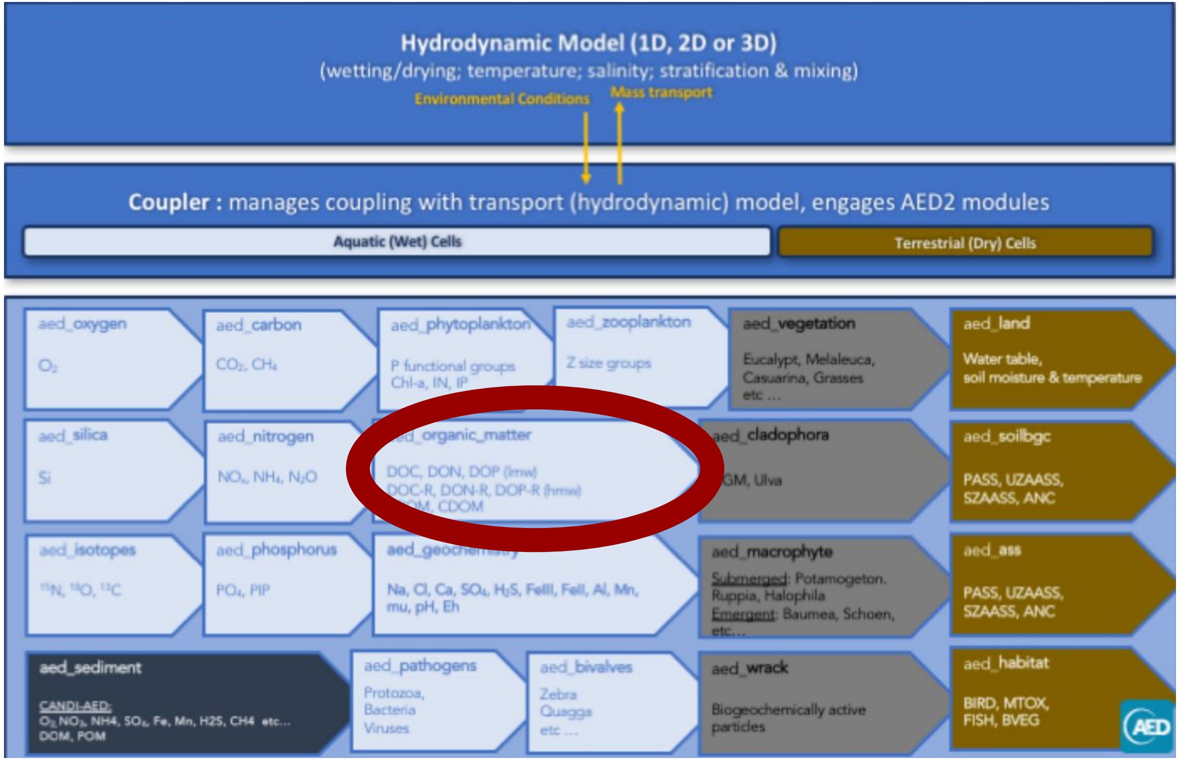




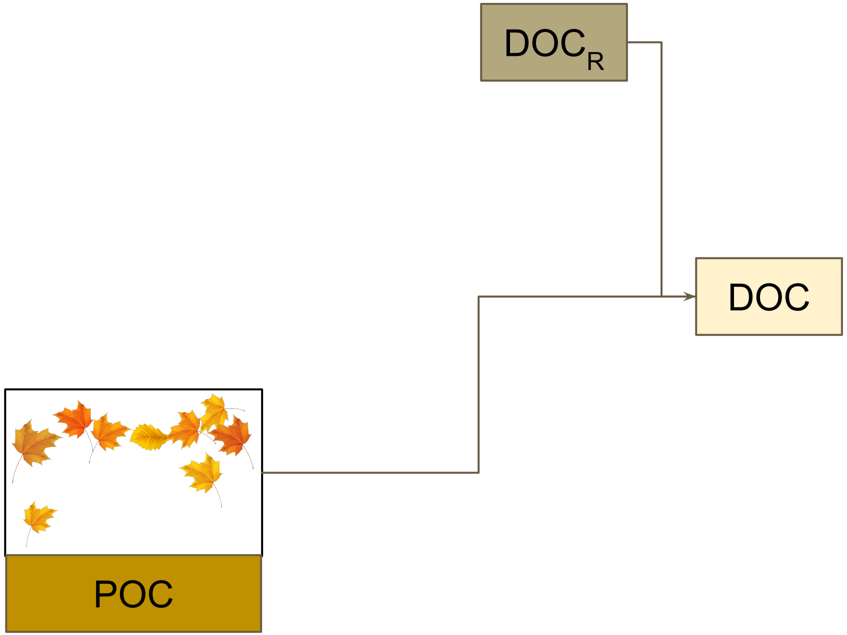
AED2 manual: Hipsey et al. 2013; oxygen plot by Hilary Dugan @hildug

AED2: modularised structure

- Interacts with GLM (two-way coupling)
- Array of modules are available
- Hierarchy of dependencies between modules, e.g., nitrogen reactions can depend on availability of oxygen



AED2: dissolved organic carbon - concept



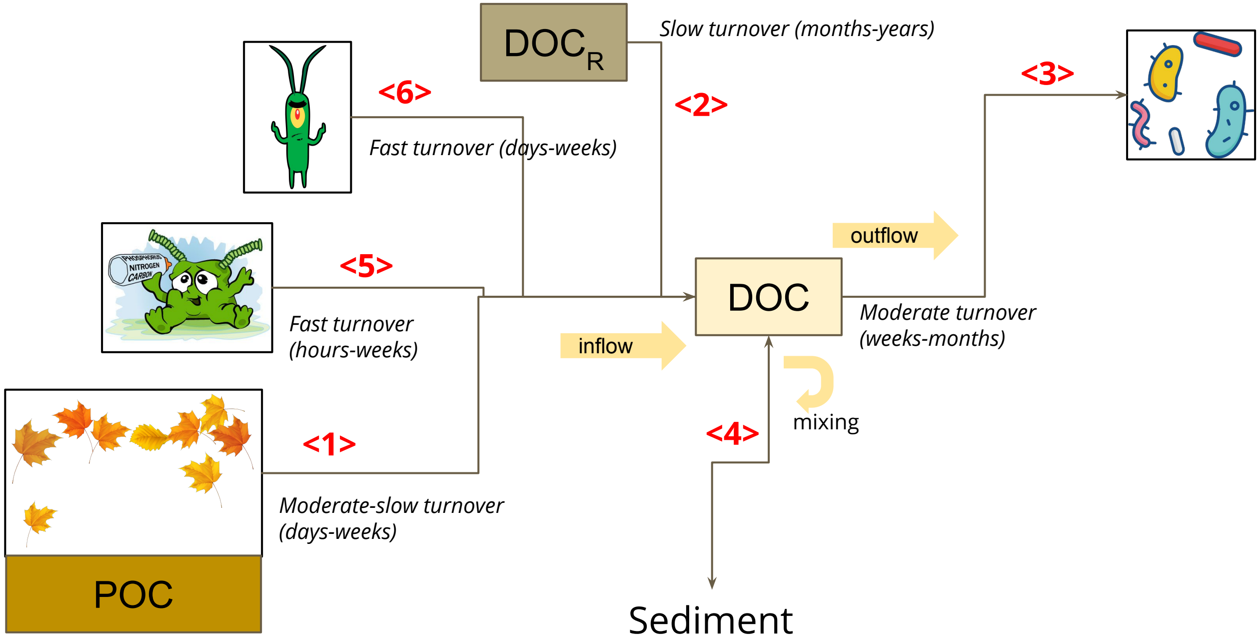
Three pools:

- Labile dissolved OC (**DOC**): fast turnover
- Refractory dissolved OC (**DOCr**): resistant to decomposition
- particulate OC (**POC**)

AED2: dissolved organic carbon - concept

$$\frac{dDOC}{dt} = (1 - f_{ref}) f^{POC}_{decom} + (1 - f_{photo}) f^{DOCR}_{photo} - f^{DOC}_{miner} \pm f^{DOC}_{sed} + \sum_{a}^{N_{PHY}} f^{PHY\_C_a}_{excr} + \sum_{z}^{N_{ZOO}} f^{z}_{excr}$$

<1> <2> <3> <4> <5> <6>



- Boxes are variables (states) you can observe
- In the model, C exists only in the boxes
- Arrows are fluxes, transformations, processes you generally cannot observe directly
- Fluxes determine how the boxes change through time
- Fluxes have equations with parameters
- Recalcitrant means the arrow is slow, labile fast
- Fitting a model requires changing parameter values so that boxes match observations, given boundary conditions
- All of these processes happen in all layers

AED2: dissolved organic carbon - equations

$$\frac{dDOC}{dt} = (1 - f_{ref}) f^{POC}_{decom} + (1 - f_{photo}) f^{DOCR}_{photo} - f^{DOC}_{miner} \pm f^{DOC}_{sed} + \sum_{a}^{N_{PHY}} f^{PHY\_C_a}_{excr} + \sum_{z}^{N_{ZOO}} f^{z}_{excr}$$

= + decomposition from particulate detritus (POC)
+ phototransformation of chromophoric DOM (DOC-R)
– mineralisation by bacteria
± sediment flux
– excretion by phytoplankton groups
– excretion by zooplankton groups

$$\frac{dDOC\_R}{dt} = f_{ref} f^{POC}_{decom} - f^{DOCR}_{miner} - f^{DOCR}_{photo} \pm f^{DOCR}_{sed}$$

= + accumulation during particulate detritus (POC) mineralisation
– slow mineralisation by bacteria
– photolysis of chromophoric DOM (DOC-R)
± sediment flux

$$\frac{dPOC}{dt} = f^{CPOM}_{bdown} - f^{POC}_{decom} - f^{POC}_{sett} + \sum_{a}^{N_{PHY}} f^{PHY\_C_i}_{mort} + \sum_{z}^{N_{ZOO}} \left[(1 - k^{z}_{assim}) f^{z}_{assim} + (1 - k^{z}_{fsed}) f^{z}_{fecal} + f^{z}_{mort}\right]$$

= + breakdown of CPOM
– decomposition to DOC
± sedimentation
+ mortality from phytoplankton groups
+ messy feeding, faecal pellet release and mortality from zooplankton groups

$$\frac{dCPOM}{dt} = -f^{CPOM}_{bdown} - f^{CPOM}_{sett}$$

= – breakdown of CPOM
± sedimentation

Aerobic mineralization of DOM

f(O2) = scales from 0-1 based on [O2]

f(T) = scales from ~0.2-2.0 based on T

$$f^{VAR}_{miner} = R^{VAR}_{miner} \frac{[O_2]}{K_{miner}+[O_2]} (\theta_{miner})^{T-20} [VAR]$$

(15)

$$+ f^{NO_3}_{denit} \chi^{VAR}_{denit}$$

(16)

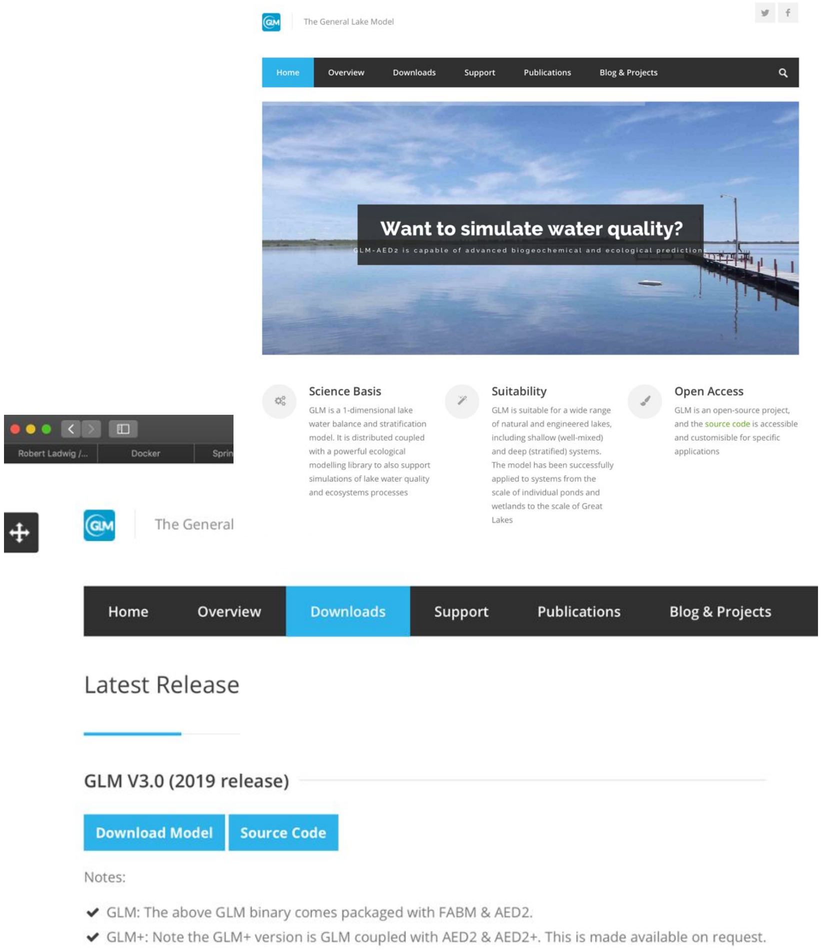
Anaerobic mineralization of DOM

RDOC, minimum-maximum rate of aerobic mineralization of labile DOC @20C (0.01-0.10 d^{-1})

Running GLM (in R)

Setting up GLM

- Installing GLM:
  - Download software package via AED GLM website: http://aed.see.uwa.edu.au/research/models/GLM/latest_release.html (precompiled binaries)
  - Source code and compilation instructions are available, accessible via the GitHub repository - https://github.com/AquaticEcoDynamics/GLM
  - GLM3r in R devtools::install_github("GLEON/GLM3r")



Setting up GLM

- Main file: glm3.nml
- Ordered into blocks
  - Begin with: &
  - End with: /
- Includes model setup and relative paths to water quality and driver data
- GLM executable needs the setup file

```
!-------------------------------------------------------------------------------
! general model setup
!-------------------------------------------------------------------------------
&glm_setup
   sim_name = 'Mendota'
   max_layers = 75
   min_layer_vol = 0.1
   min_layer_thick = 0.15
   max_layer_thick = 1.5
   density_model = 1
/
&mixing
   surface_mixing = 1
   coef_mix_conv = 0.2
   coef_wind_stir = 0.23
   coef_mix_shear = 0.3
   coef_mix_turb = 0.51
   coef_mix_KH = 0.3
   coef_mix_hyp = 0.4689587
   deep_mixing = 2
   diff = 0
/
&light
   light_mode = 0
   n_bands = 4
   light_extc = 1, 0.5, 2, 4
   energy_frac = 0.51, 0.45, 0.035, 0.005
   Benthic_Imin = 10
   Kw = 0.4315141
/
!-------------------------------------------------------------------------------
! water quality setup
! if this block is read, water quality functionality will be enabled
!-------------------------------------------------------------------------------
```

GLM homepage: https://aed.see.uwa.edu.au/research/models/GLM/index.html

Setting up GLM

- Define the &morphometry
  - Area-depth relationship (&H and &A)
- Define what you want to simulate
  - State variables (temperature, salinity, etc.)
  - Grid resolution and time step (&glm_setup, &time)
- Define external environment
  - Boundary conditions (inflows, meteorology, outflows) (&meteorology, &inflows, &outflows)
- Provide an initial condition (&init_profiles)
- Start your simulation

glm3.nml file

&glm_setup:	General simulation info
&wq_setup:	Water quality
&time:	Time control
&morphometry:	Lake morphometric information
&output:	Output file details
&init_profiles:	Initial vertical profiles
&meteorology:	Surface forcing and meeorology
&inflows:	Information about inflowing rivers
&outflows:	Information about outflows
&light:	Information about light climate

GLM homepage: https://aed.see.uwa.edu.au/research/models/GLM/index.html

Setting up GLM

- Boundary conditions: data that drive the model
- At minimum, GLM needs meteorological data (hourly or daily) in CSV-format

```
Date,ShortWave,LongWave,AirTemp,RelHum,WindSpeed,Rain,Snow
2009-01-01 00:00:00,0,191.50999,-15.609991,88.86569,2.9134343,0,0
2009-01-01 01:00:00,0,217.56,-14.750006,88.178638,3.5679266,0,0
2009-01-01 02:00:00,0,217.56,-14.15,88.305945,3.9796608,0,0
2009-01-01 03:00:00,0,217.57001,-13.560004,88.286784,4.3862171,0,0
2009-01-01 04:00:00,0,230.42999,-12.959998,88.028059,4.8016873,0,0
2009-01-01 05:00:00,0,230.42999,-12.449988,87.745015,5.3791169,0,0
2009-01-01 06:00:00,0,230.42999,-11.929999,87.288882,5.9664732,0,0
2009-01-01 07:00:00,0,246.75,-11.41001,86.734669,6.5488395,0,0
2009-01-01 08:00:00,0,246.75999,-10.620001,86.510658,6.995177,0,0
2009-01-01 09:00:00,50.312,246.75999,-9.8200134,85.935506,7.4708298,0,0
2009-01-01 10:00:00,168.256,253.35001,-9.0199951,85.14882,7.9805386,0,0
2009-01-01 11:00:00,255.60001,253.35001,-7.8599915,85.236462,8.1806115,0,0
2009-01-01 12:00:00,325.892,253.35001,-6.6900085,84.658597,8.4221193,0,0
2009-01-01 13:00:00,349.651,265.15002,-5.5300049,83.680197,8.6924564,0,0
2009-01-01 14:00:00,318.17099,265.14999,-4.9799866,84.281007,8.1392691,0,0
2009-01-01 15:00:00,264.33701,265.14999,-4.4299988,84.709775,7.6592037,0,0
2009-01-01 16:00:00,153.472,271.73001,-3.880011,84.970459,7.2817926,0,0
2009-01-01 17:00:00,51.591999,271.73001,-4.1399902,86.532806,6.6806363,0,0
```

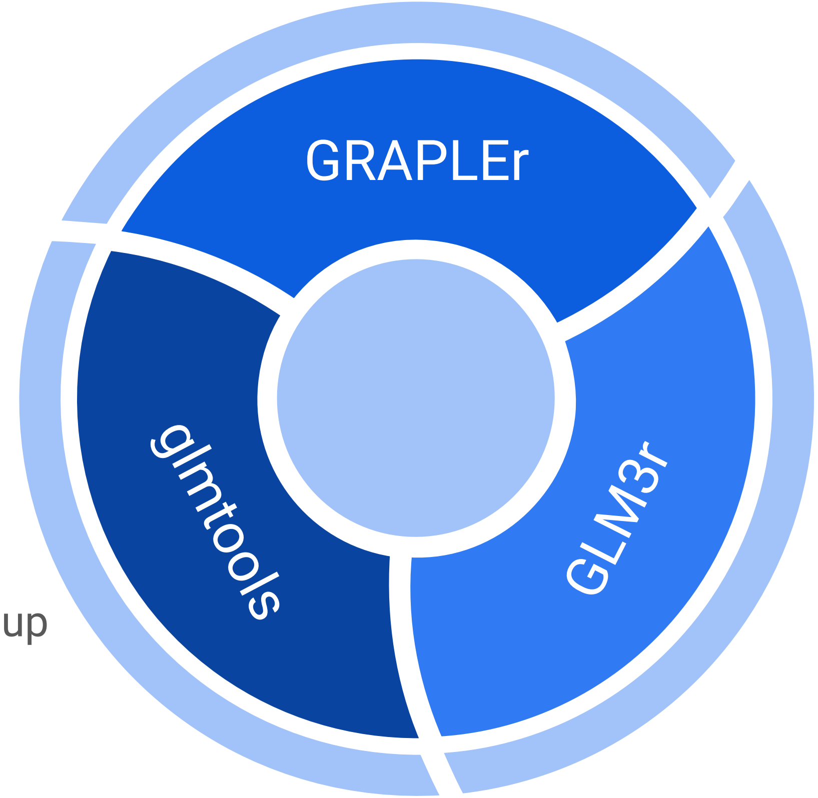
```
Time,FLOW,TEMP,SALT
2009-01-01 00:00:00,3.560861,-0.526716,0
2009-01-02 00:00:00,6.162997,-0.585426,0
2009-01-03 00:00:00,3.826635,-0.639793,0
2009-01-04 00:00:00,5.397271,-0.696168,0
2009-01-05 00:00:00,5.417987,-0.742644,0
2009-01-06 00:00:00,3.886593,-0.784772,0
2009-01-07 00:00:00,5.806259,-0.821455,0
2009-01-08 00:00:00,5.65432,-0.85336,0
2009-01-09 00:00:00,4.725836,-0.873803,0
2009-01-10 00:00:00,5.821722,-0.906718,0
2009-01-11 00:00:00,5.724123,-0.926215,0
2009-01-12 00:00:00,5.495776,-0.936452,0
2009-01-13 00:00:00,5.473368,-0.947798,0
2009-01-14 00:00:00,5.513825,-0.955154,0
```

- And flow data (inflow and/or outflow, hourly or daily) in CSV-format

GLM homepage: https://aed.see.uwa.edu.au/research/models/GLM/index.html

GLM in R

- Distributed computing system
- Run multiple GLM simulations



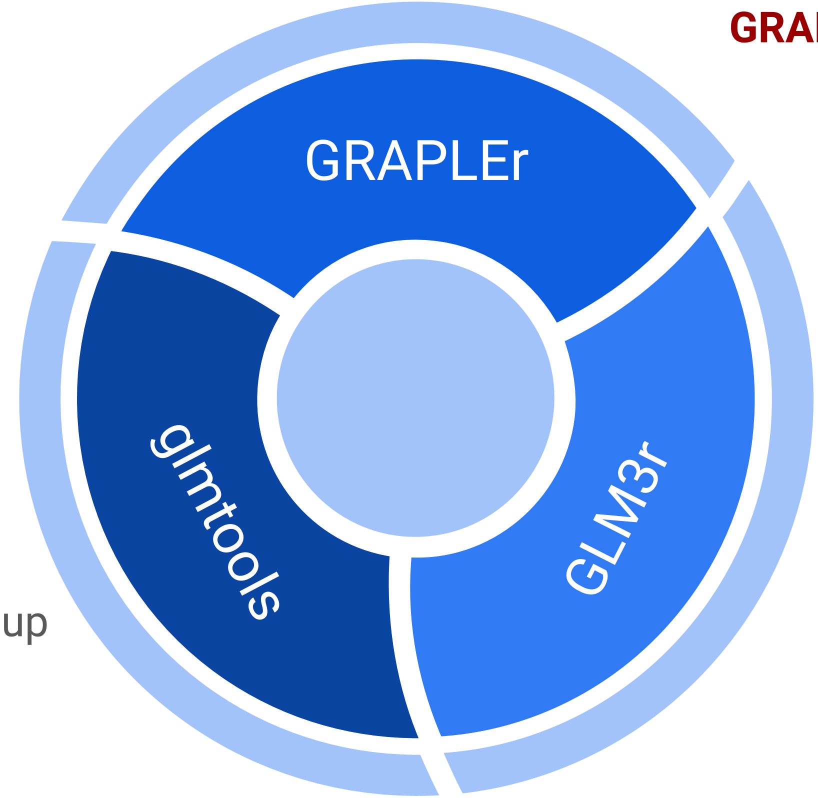
- Run GLM in R

- Changing model setup
- Visualizing results
- Physical derivatives
- Calibration

GLM in R

- Distributed computing system
- Run multiple GLM simulations

GRAPLE/GRAPLEr



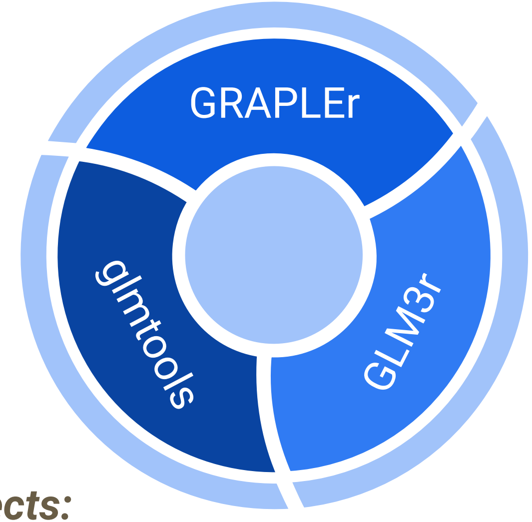
USGS-R/glmtools

- Changing model setup
- Visualizing results
- Physical derivatives
- Calibration

GLEON/GLM3r

- Run GLM in R

GLM in R



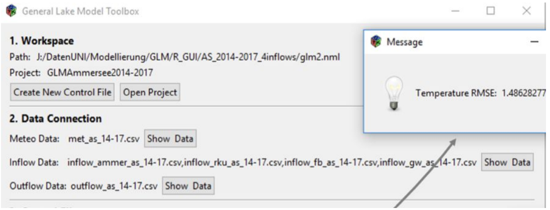
More awesome R-related GLM projects:

Macrosystems EDDIE: classroom modules to teach macrosystems ecology

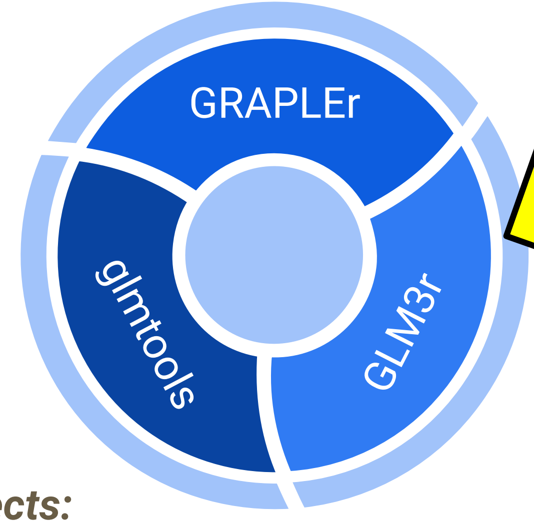
LakeEnsemblR: R-package to run lake ensembles (release soon)

glmGUI: GUI and toolbox to run GLM

Columbus4Limnology: Run GLM online in the cloud (release soon)



GLM in R



#LakeEnsemblR workshop at this GLEON!

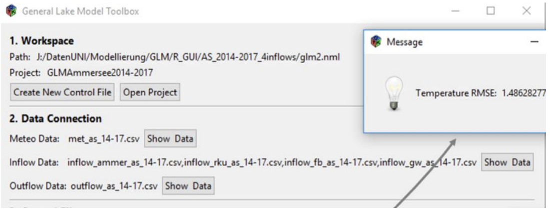
More awesome R-related GLM projects:

[Macrosystems EDDIE](): classroom modules to teach macrosystems ecology

[LakeEnsemblR](): R-package to run lake ensembles (release soon)

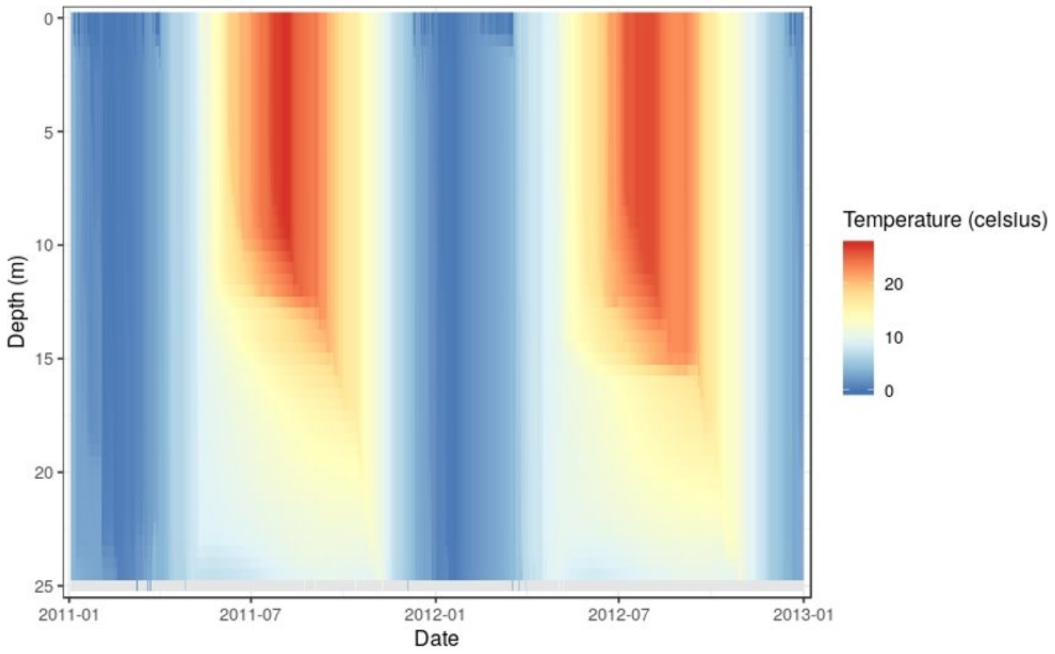
[glmGUI](): GUI and toolbox to run GLM

[Columbus4Limnology](): Run GLM online in the cloud (release soon)



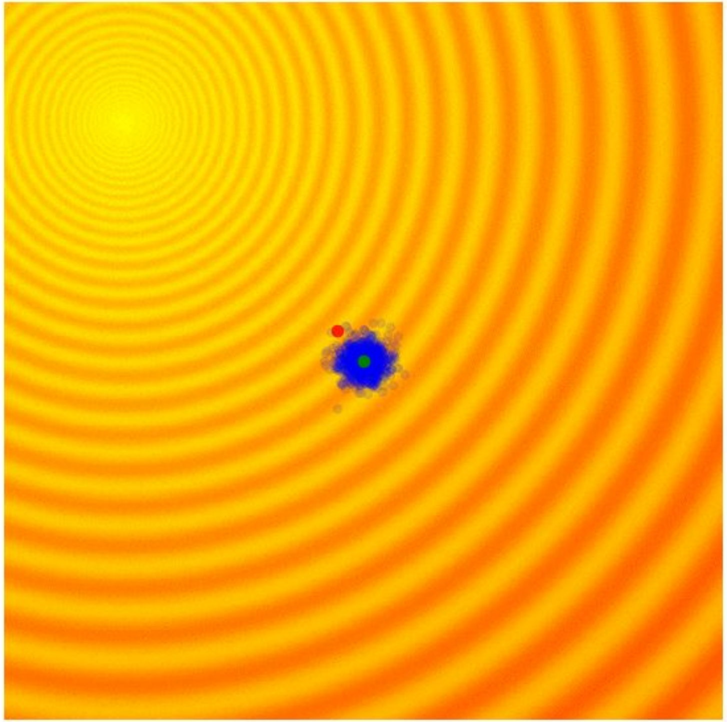
glmtools

- Either from github or GRAN
- Offers lots of functionalities:
    - Manipulate model setup
    - Retrieve output data
    - Compare observed to simulated data
    - Visualize results
    - Calibration functions
- ‘ggplot_overhaul’ branch includes up-to-date version with experimental features (e.g., water quality calibration)



glmtools: calibration

a) CMA-ES: derivative-free optimization, which adaptively changes the search path for every generation (modifying normal distribution and covariance matrix), covers wide search space, converges fast but needs more computational time (~O(N2))
b) Nelder-Mead method: derivative-free optimization, direct search method



Github: https://github.com/USGS-R/glmtools, CMA-ES algorithm: Hansen 2006, animation by https://blog.otoro.net

Finally some R

Wrapping up

Questions, issues, problems & feedback?

Join the official GLM slack

Thanks for joining!

Have fun lake modeling!