HYDROLOGIC MODELING OVERVIEW

ERT 474/574

Open-Source Hydrologic Data Analytics

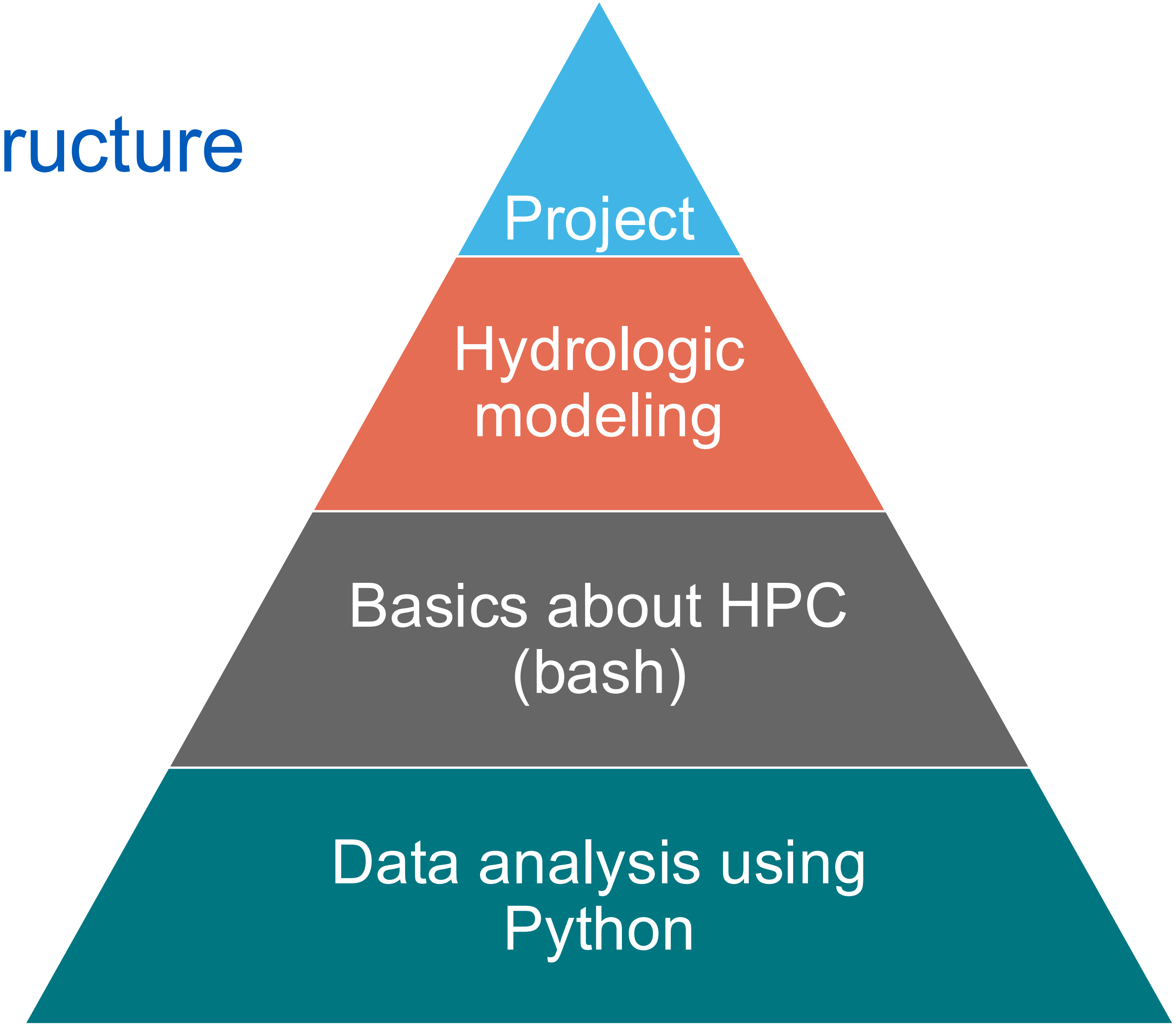
Sep 3rd 2025

University at Buffalo The State University of New York

Recap

- Coding languages
  - What language is most used for GCMs and hydrological models?
- Git & GitHub
  - What is the difference between Git and GitHub?
  - When we make edits to a file locally, what steps do we need to take to push the changes to GitHub?

Class structure



Hydrologic models

Hydrology 101

Water balance equation

$P =$

Precipitation

4

Hydrologic models

Hydrology 101

Water balance equation

$$P = Q + E + \Delta S$$

Runoff

Storage change

Precipitation Evapotranspiration

5

Hydrologic models

Hydrology 101

Water balance equation

$$P = \boxed{Q} + E + \Delta S$$

Runoff

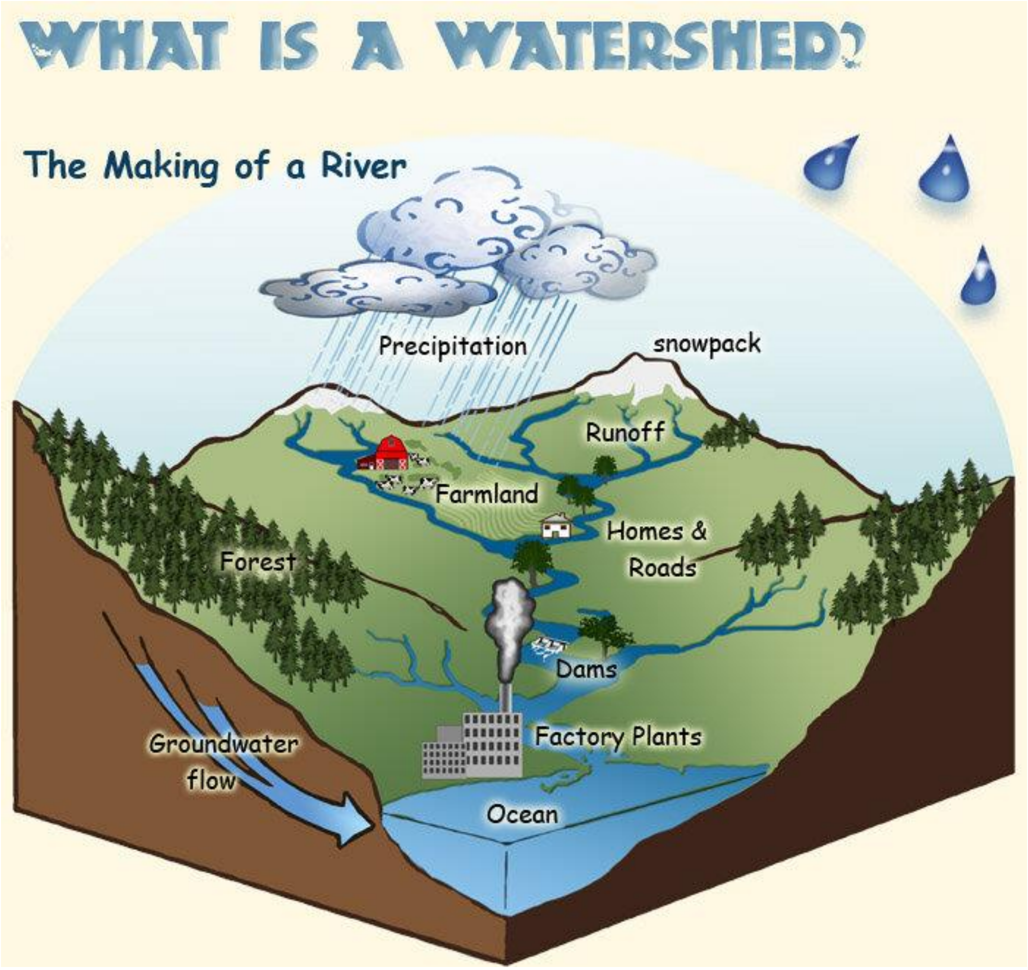
Storage change

Precipitation

Evapotranspiration

Runoff is the main variable of interest to hydrologist!

Watershed



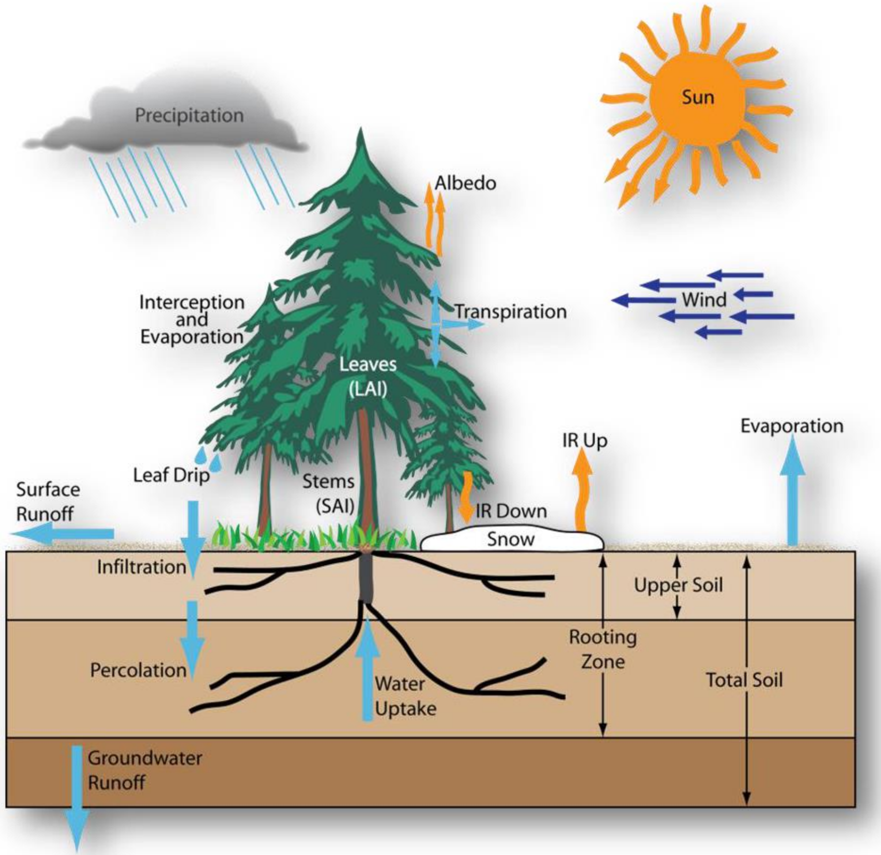
- **Watershed** describes an area of land that drains downslope to the lowest point.
  - Imagine a water drop falls on a mountain: where will it flow?
- Watershed boundaries follow major ridgelines around channels and meet at the bottom.
- Watersheds can be large or small.

7

What watershed are we located at?

- United States Geological Survey (USGS) National Watershed Boundary Dataset
  - Buffalo (https://hub.arcgis.com/maps/esri::usgs-watershed-boundaries/explore?location=42.752919%2C-78.410536%2C8.05)

Process-based hydrologic models

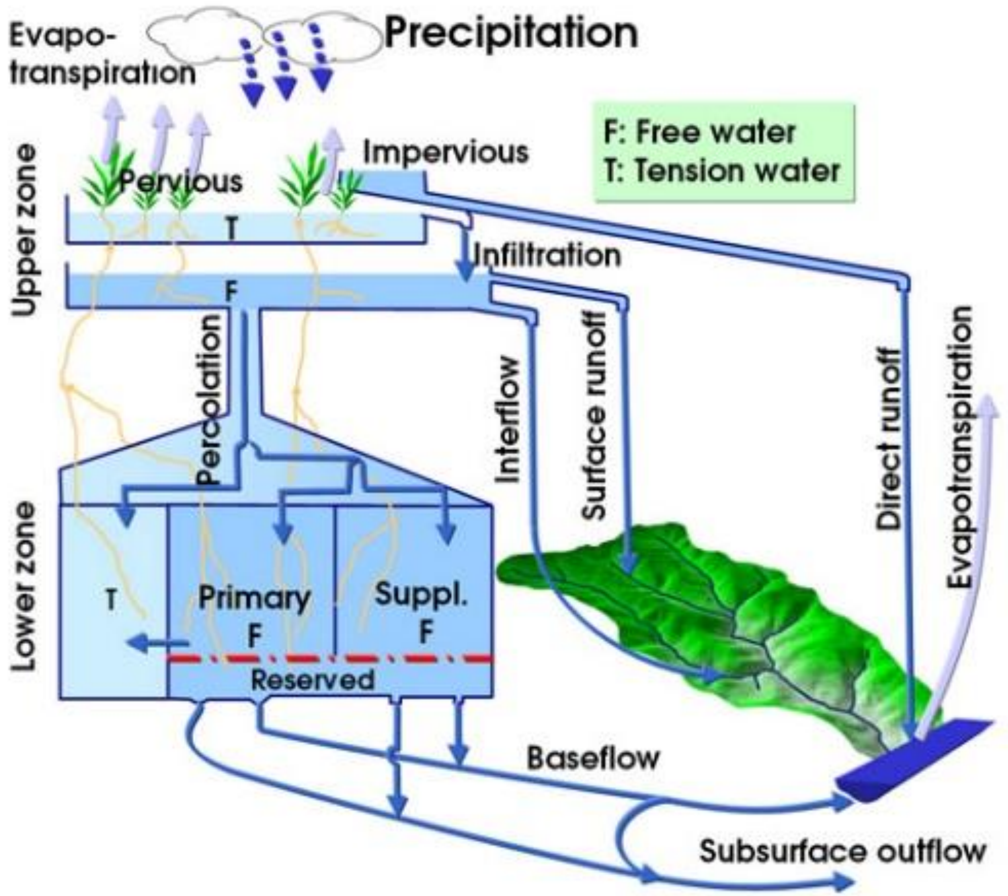


https://regclim.coas.oregonstate.edu/teaching-examples/climatology-201/energy-and-water-balance-background/

- Process-based hydrologic models represents a collection of connected processes, such as soil infiltration, soil evaporation, transpiration from vegetation, etc.
- Closure of water balance and energy balance
- Complexities of models (different perspectives)
  - Lump model or spatially distributed models
  - Physical process representation

9

Sacramento Soil Moisture Accounting Model (SAC-SMA)



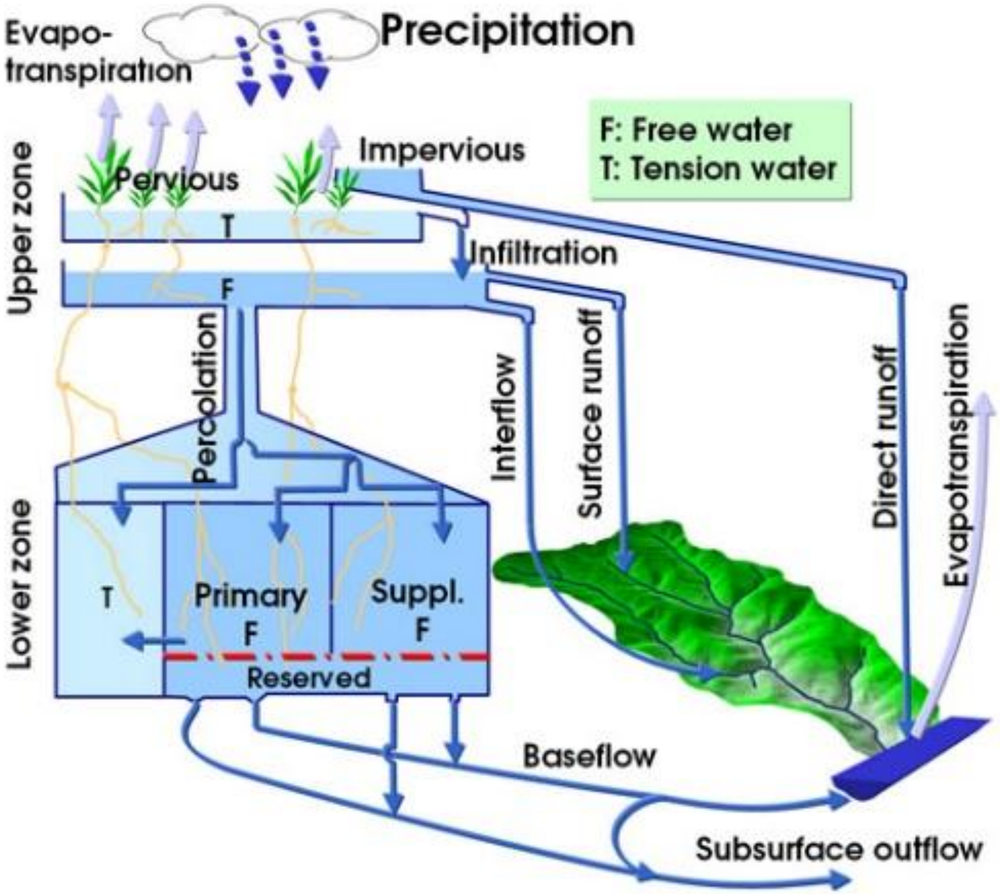
- SAC-SMA is a lumped hydrologic model
  - Newer development might enable it to be semi-distributed.
- The history of model development goes back to 1970s.
- Highly abstraction of physical processes related to real-world water cycles
- It is probably one of the most famous and widely used hydrological models

10

Sacramento Soil Moisture Accounting Model (SAC-SMA)

Languages https://github.com/NOAA-OWP/sac-sma

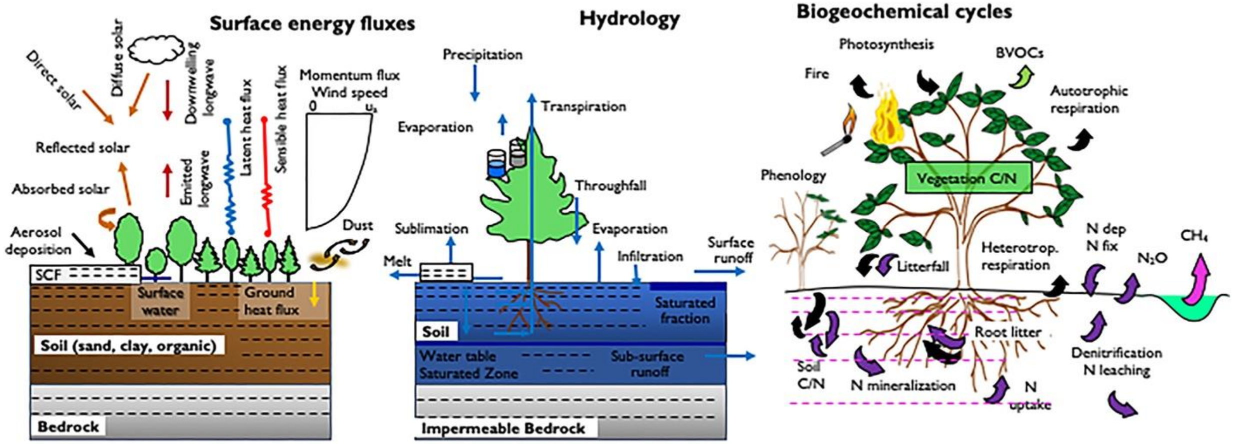
- Fortran 95.5%
- CMake 2.3%
- Makefile 2.2%

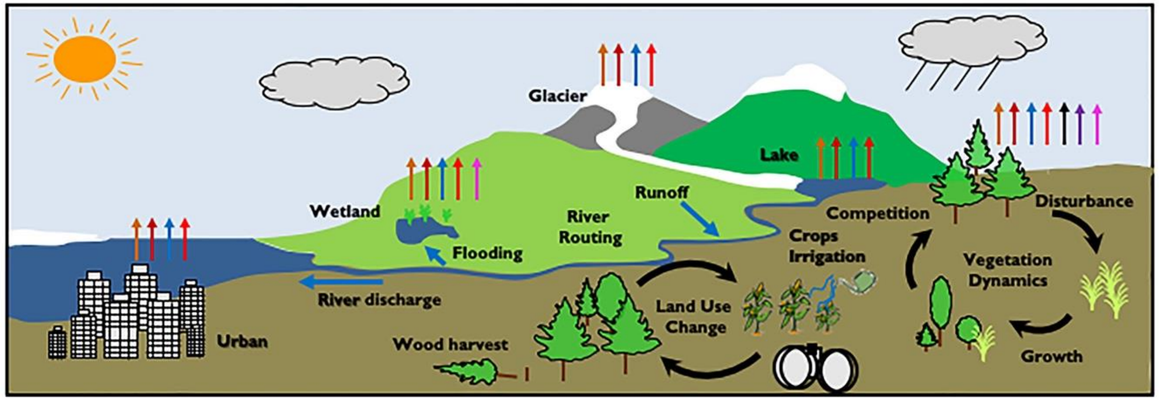


- SAC-SMA is a lumped hydrologic model
  - Newer development might enable it to be semi-distributed.
- The history of model development goes back to 1970s.
- Highly abstraction of physical processes related to real-world water cycles
- It is probably one of the most famous and widely used hydrological models

11

Community Terrestrial Systems Model (CTSM)





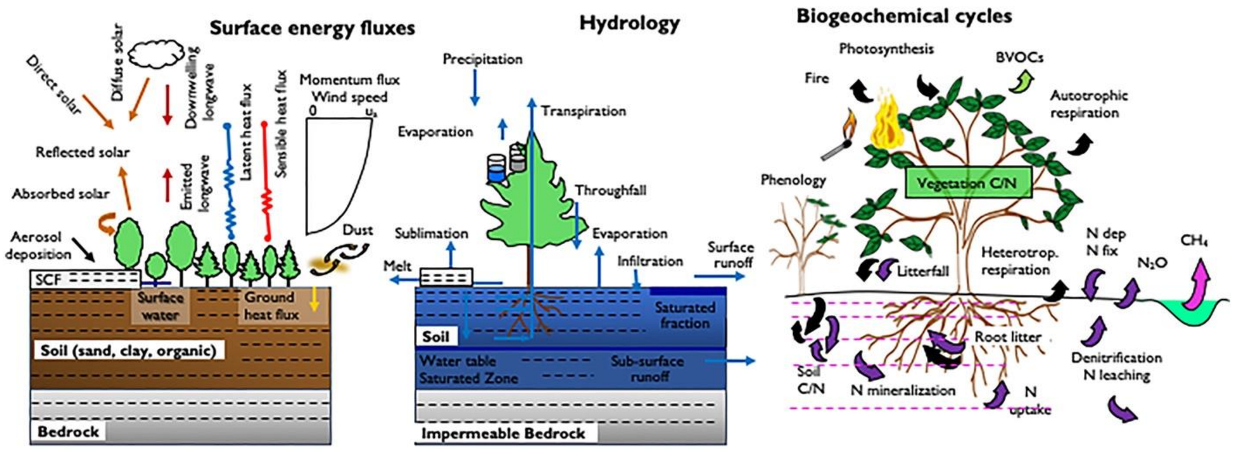
- CTSM is a distributed hydrologic model
- The history of model development goes back to 1996.
- State-of-the-science land models that more closely mimic the real-world physical processes, not only for water but energy and biogeochemical cycles.
- It is widely used in earth system modeling community.

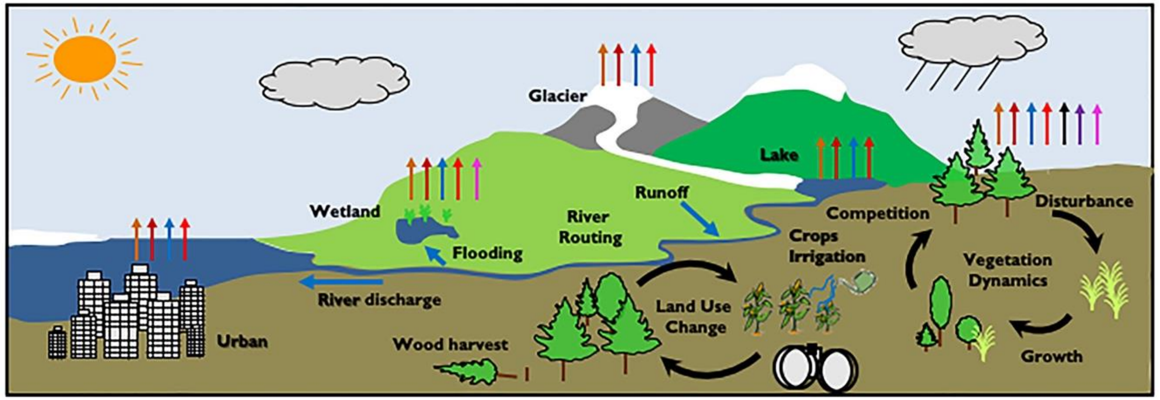
 12

Community Terrestrial Systems Model (CTSM)

Languages

- Fortran 86.0%
- Python 9.0%
- Perl 3.1%
- Shell 0.7%
- NCL 0.6%
- CMake 0.4%
- Other 0.2%



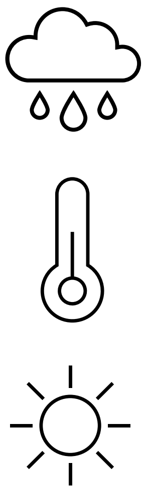


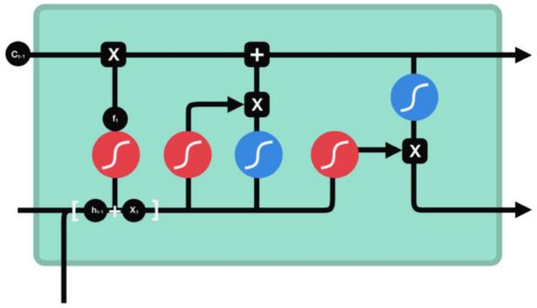
- CTSM is a distributed hydrologic model
- The history of model development goes back to 1996.
- State-of-the-science land models that more closely mimic the real-world physical processes, not only for water but energy and biogeochemical cycles.
- It is widely used in earth system modeling community.

13

ML hydrologic models

Input **ML-model** **Output**





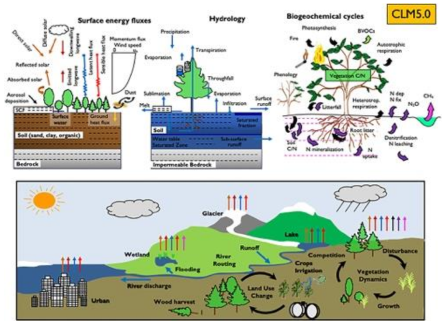
River flow

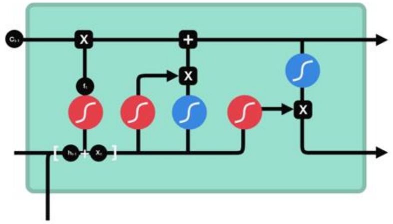
Meteorological forcing data

- Data-driven model
- Directly used meteorological forcing data to predict runoff
  - Detailed physical processes are usually not explicitly represented
- Black box nature
- Water balance and energy balance are not explicitly represented.

14

It is not just black or white!

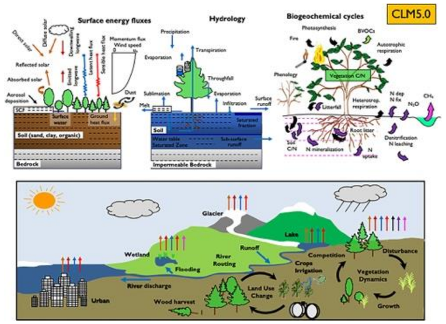




Pure process-based models | Pure ML-AI models

15

It is not just black or white!

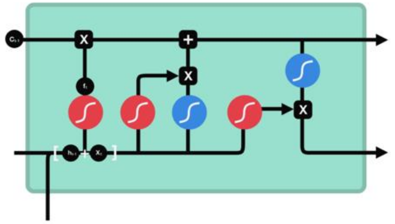


Deep Learned Process Parameterizations Provide Better Representations of Turbulent Heat Fluxes in Hydrologic Models

Andrew Bennett, Bart Nijssen

First published: 12 May 2021
https://doi.org/10.1029/2020WR029328

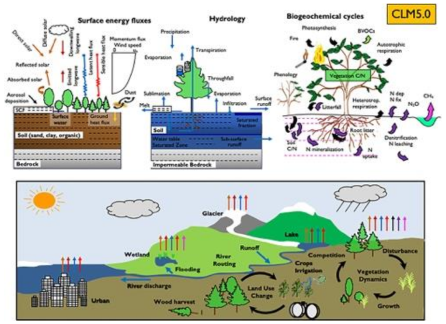
Use ML to represent one process in the process-based hydrologic model



Pure process-based models | Pure ML-AI models

16

It is not just black or white!

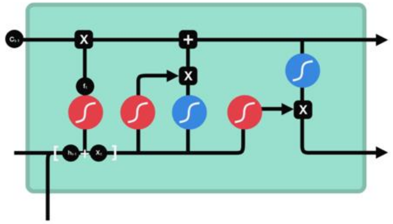


Deep Learned Process Parameterizations Provide Better Representations of Turbulent Heat Fluxes in Hydrologic Models

Andrew Bennett, Bart Nijssen

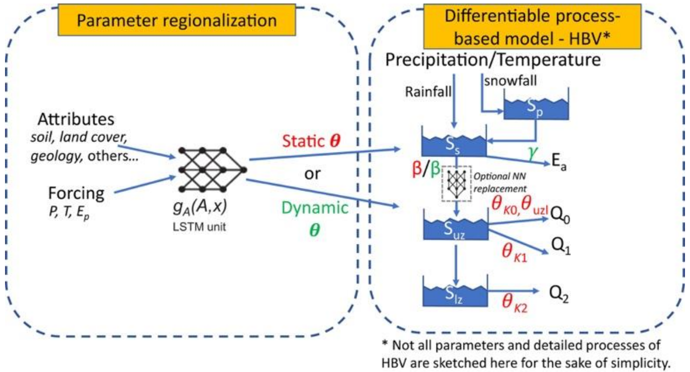
First published: 12 May 2021
https://doi.org/10.1029/2020WR029328

Use ML to represent one process in the process-based hydrologic model



Pure process-based models — Pure ML-AI models

Instead of having one big black box, multiple smaller black boxes are used to mimic the process of physically based hydrologic models



17

Which models do you prefer?

18

What platforms can we use to run hydrologic models?

PC / Laptops

Challenges to run hydrologic models on PCs

- Software / module dependencies
  - Software installation and system configuration can be tricky and time-consuming
- Data accessibility
  - Voluminous input (such as meteorological forcing) for distributed hydrologic models
- Limited computing resources
  - Basic concept: core-hour (A unit of measurement for the amount of computational activity that occurs on a single core for a period of one hour)

What platforms can we use to run hydrologic models?

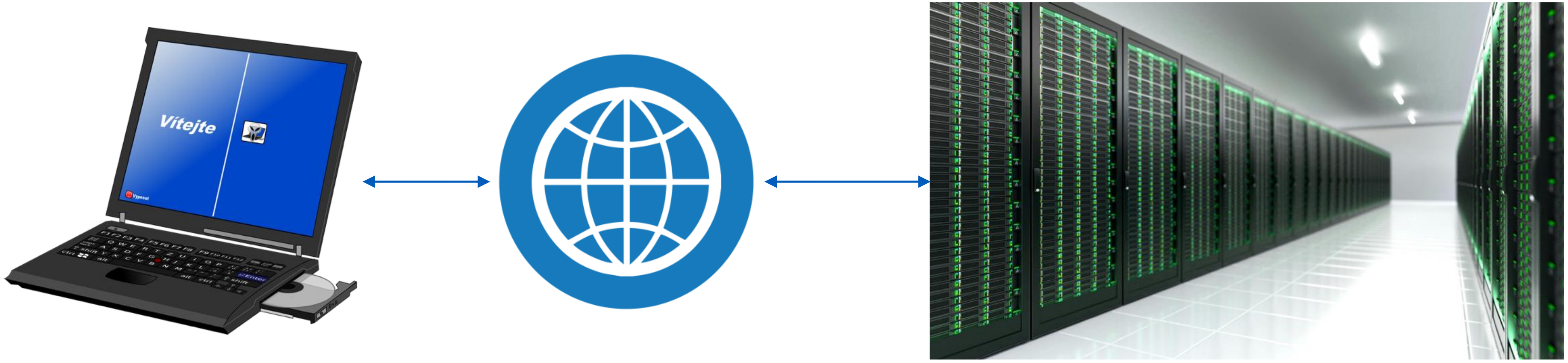
- Major research institutions have their own HPC centers or servers

University at Buffalo Center for Computational Research

- Great technical support
- Large data storage
- Limited to researchers (Usually not accessible to the general public)

High Performance Computing / Servers

What platforms can we use to run hydrologic models?



- Cloud computing has more flexibilities than HPC centers
- However, the cloud computing can be more expensive per core-hour than HPC.

Cloud computing

Many available resources

Google colab

GETTING STARTED WITH GOOGLE COLAB

Azure Notebooks

- Google Colab and Azure Notebooks are great resources to start learning Python.
- They both have free versions.
- However, in the free versions, the users cannot customize the virtual environment for Python (i.e., you cannot install the desired packages) and have no access to terminals

Many available resources

CUAHSI
allied for water science
www.cuahsi.org

Google colab
GETTING STARTED WITH GOOGLE COLAB

PANGEO

Azure Notebooks

Google Earth Engine

Google Earth Engine

- Widely used in remote sensing community
  - Hosts satellite imagery and stores it in a public data archive that includes historical earth images going back more than forty years
  - Landsat and Sentinel-2
- The available API for Google Earth Engine is Python and JavaScript.
  - Customization of Python environment is supported
- GEE is usually not used to run hydrologic models

PANGEO

- Pangeo is first and foremost a community promoting open, reproducible, and scalable science.
  - This community provides documentation, develops and maintains software, and deploys computing infrastructure to make scientific research and programming easier.
- The Pangeo software ecosystem involves open source tools such as **xarray**, iris, dask, **jupyter**, and many other packages.

Parallel computing

Interactive computing

Geospatial datasets

Analyzing and visualizing meteorological and oceanographic data sets

CUAHSI

Consortium of Universities for the Advancement of Hydrologic Science, Inc. (Sponsored by National Science Foundation)

- Support water science through education and data services
- Provide free and open source software for managing, archiving, sharing, discovering, publishing, and analyzing all types of water data
  - Hydroshare
  - Jupyterhub
  - MATLAB Online
  - Hydrologic Information System (HIS)
  - Model Domain Subsetter

CUAHSI allied for water science
www.cuahsi.org

CUAHSI

allied for water science

www.cuahsi.org

Consortium of Universities for the Advancement of Hydrologic Science, Inc. (Sponsored by National Science Foundation)

- Support water science through education and data services
- Provide free and open source software for managing, archiving, sharing, discovering, publishing, and analyzing all types of water data
  - **Hydroshare**
  - **Jupyterhub**
  - MATLAB Online
  - Hydrologic Information System (HIS)
  - Model Domain Subsetter

HYDROSHARE

- **Discover** multiple types of water data published by others.
- **Share** data with colleagues and groups.
- **Formally publish** data with a Digital Object Identifier (DOI) so your work can be easily cited.

GitHub Codespaces

- GitHub Codespaces is a **cloud-based** development environment provided by GitHub. It allows users to **create and manage development environments directly within a web browser** or through Visual Studio Code desktop, eliminating the need for extensive local setup.
- Available cloud computing resources
  - With the student developer pack, you have 180 free core-hours & 20 GB of storage
  - Without the student developer pack, 120 free core-hours & 15 GB of storage

Homework #2

In this homework, we will read two manuscripts about a widely used hydrologic model, the Variable Infiltration Capacity (VIC) model, i.e., Xu et al. (1994), and Hamman et al. (2018).

- You will need to answer questions. Please describe it in your own words.
- To answer the questions, please create a **Jupyter notebook**, copy the questions to the notebook, and type down the answer in **Markdown.**