

# Build, customize and run models with Xarray-Simlab

**Benoît Bovy** (Freelance, GFZ-Potsdam)

Pangeo Showcase Webinar Series, April 21st 2021



*Can we have a single tool that covers all the lifecycle of a scientific idea, from data collection to publication?*



Matthias Bussonnier, Jupyter talk @ UC Berkeley, September 2015

*Can we have a single tool that covers all the lifecycle of a scientific idea, from data collection to publication?*

*model run*  
*numerical experiment*



Matthias Bussonnier, Jupyter talk @ UC Berkeley, September 2015

# Xarray-Simlab

- Python framework for *interactive, exploratory and scalable* modelling
- Tightly integrated with  xarray ,  DASK ,  Zarr & friends
- Domain-agnostic
  - Geomorphology, hydrology, ecosystem modelling...
- Project started in 2017, (im)mature



<https://github.com/xarray-contrib/xarray-simlab>

# Inspired by

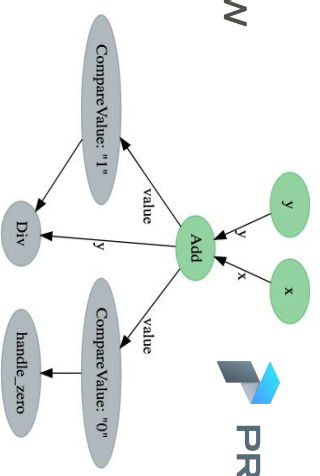
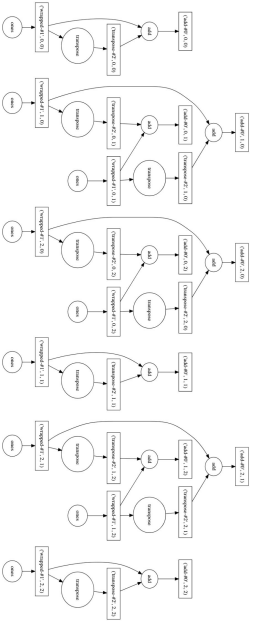
Domain-specific modelling frameworks (Python)



CLIMLAB

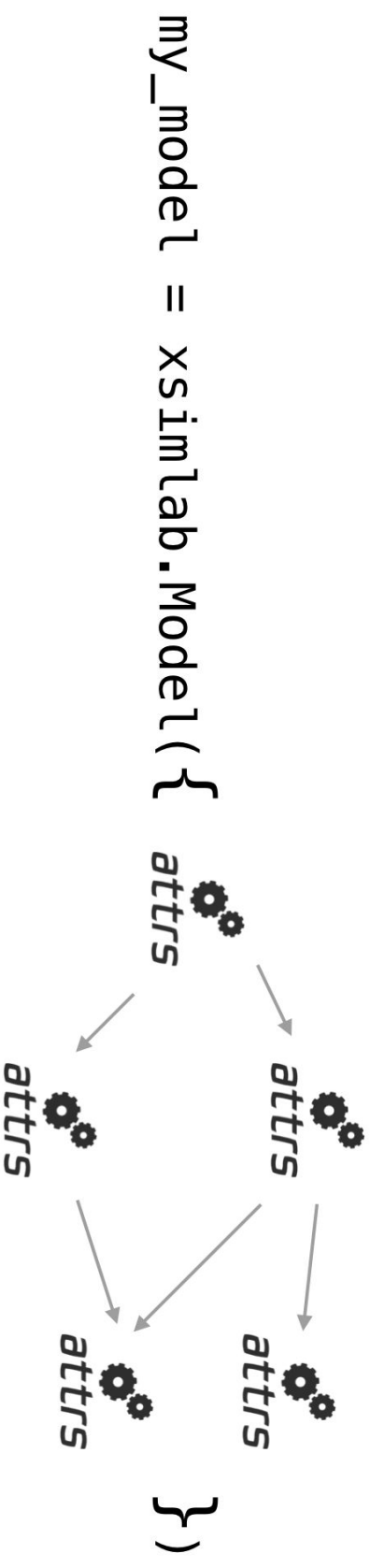


Workflow / task management libraries and frameworks (DAGs everywhere)

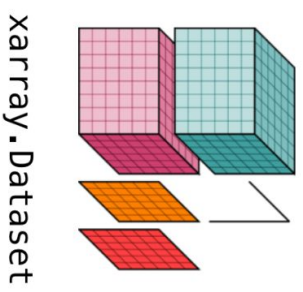


# A Model = a collection of “dataclasses” called processes

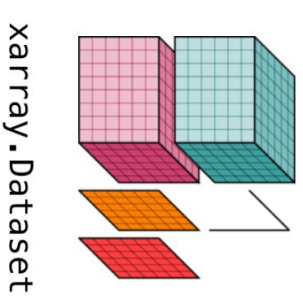
With implicit DAG (re)construction



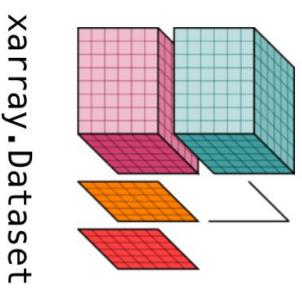
# Simulation(s): Dataset IN, Dataset OUT



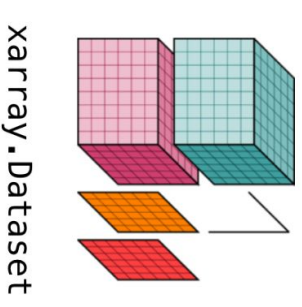
`Dataset.xsimlab.run(model=my_model)`



# Simulation(s): Dataset IN, Dataset OUT

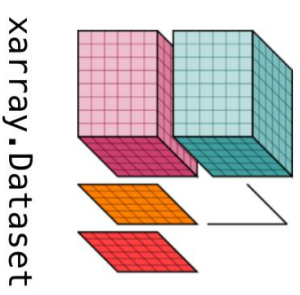


```
(Dataset
  .<some_preprocessing>
  .xsimlab.run(model=my_model)
  .<some_postprocessing>)
```

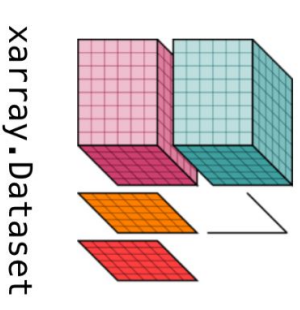




# Simulation(s): Dataset IN, Dataset OUT



```
Dataset.xsimlab.run(  
    model=my_model,  
    batch_dim="batch",  
    parallel=True,  
    store="run.zarr"  
)
```



**DEMO**

## **Many more features**

- Model customization by process class inheritance
- Execute the DAG of process classes in parallel with Dask
- Flexible variable dimensions
- Other kinds of variables (group, on\_demand, etc.)
- Time-varying input values
- Simulation monitoring and control via hooks
- Zarr encoding options

## What's next?

- Explicit process dependencies (ongoing, @Joeperdefloep)
  - Will enable workflows more familiar to modellers...
- More flexible simulation stages
  - Xarray-centric data workflow system? (one “run” stage)
- Command-line and/or graphical (Jupyter widgets) interface
- More users and contributors from various domains! (hopefully)

# Thank you for your attention!

Xarray-simlab: <https://github.com/xarray-contrib/xarray-simlab>



  @benbovy