

FREE SOFTWARE FOR ENVIRONMENTAL SCIENCES

Weiming Hu

Center for Western Weather and Water Extremes
Scripps Institution of Oceanography
UC San Diego

March 19, 2023

UC San Diego

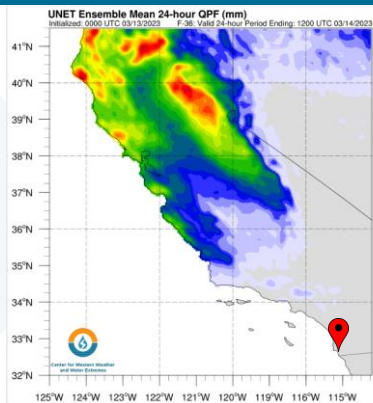


SCRIPPS INSTITUTION OF
OCEANOGRAPHY

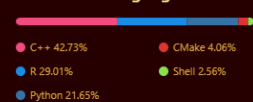
R/V Sally Ride, Anacortes, Washington. Credit: Jeff Dillon.

BACKGROUND

- Educator
- Geographer
- Machine Learning Scientist
- Open-Source Developer
- Free Software Enthusiast
- Electric Unicycle (EUC) Rider
- Guitar

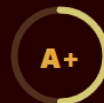


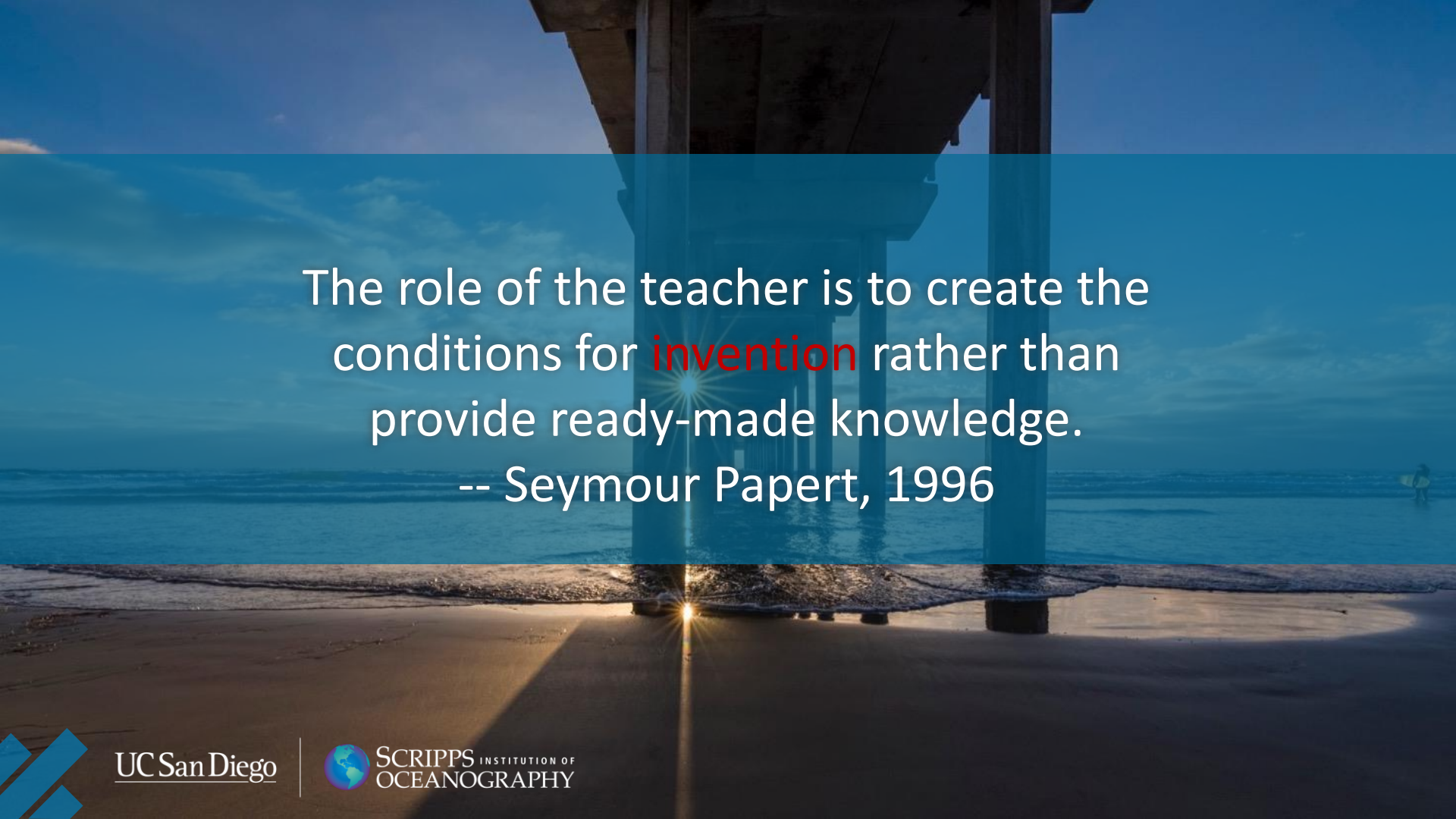
Most Used Languages



Weiming's GitHub Stats

☆ Total Stars Earned:	27
🕒 Total Commits (2023):	296
🔗 Total PRs:	34
📄 Total Issues:	152
👤 Contributed to (last year):	2





The role of the teacher is to create the conditions for **invention** rather than provide ready-made knowledge.

-- Seymour Papert, 1996

OUTLINE

- Motivation
- Open Science
- Open Access
- Free Software for Open Science
- Teaching Free Software for Environmental Science
- What Can the Dev Community do

MOTIVATION

1. Teaching commercial software in public institutions
2. Using commercial software in public institutions

MOTIVATION

This talk is not about:

- Specific software
- Science topic
- Pointing fingers

This talk is about:

- Similarities between science and free software
- Understanding people's decisions
- Change and better practice

OPEN SCIENCE

Definition of *Open* in Open Data, Open Content, and Open Knowledge:

“Open means anyone can freely access, use, modify, and share for any purpose.” [1]

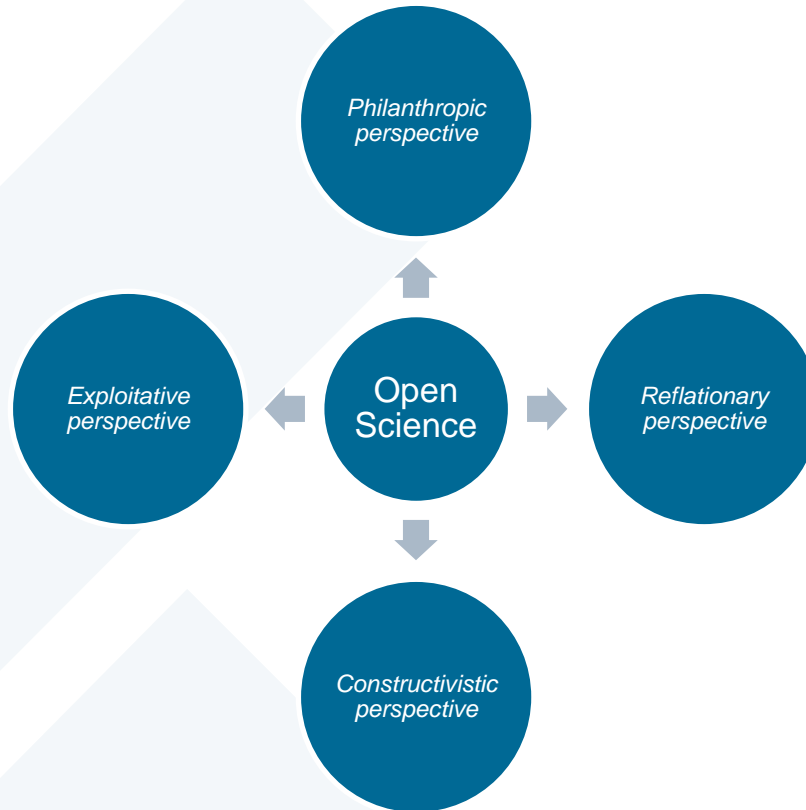
Definition of *Free* in Free Software:

“Free software means software that respects users' freedom and community. Roughly, it means that the users have the freedom to run, copy, distribute, study, change and improve the software.” [2]

[1] Open Knowledge Foundation, <https://opendefinition.org/>

[2] GNU Operating System, <https://www.gnu.org/philosophy/free-sw.en.html>

PERSPECTIVES OF OPEN SCIENCE



PERSPECTIVES OF OPEN SCIENCE

Philanthropic Perspective

- Public lectures or courses



- Open access journals

A screenshot of the Penn State GEOG 160 website. The header includes the Penn State logo and 'DEPARTMENT OF GEOGRAPHY'. The main title is 'GEOG 160 MAPPING OUR CHANGING WORLD'. A navigation bar contains links for HOME, SYLLABUS, LESSONS, CANVAS, RESOURCES, and LOGIN. The main content area features the title 'Geography 160: Mapping our Changing World' and a large map of the world showing wind patterns with a color scale from blue to red. A search bar is visible on the right side. Below the map, there is a 'Quick Facts about GEOG 160' section and a list of chapters under the heading 'Lessons'.

PennState
College of Earth and Mineral Sciences

DEPARTMENT OF
GEOGRAPHY

GEOG 160
MAPPING OUR CHANGING WORLD

HOME SYLLABUS LESSONS CANVAS RESOURCES LOGIN

Geography 160: Mapping our Changing World [Print](#)

Logged in as Anonymous.

Search

Lessons

- Chapter 1: Location is Where It's At: Introduction to GIScience and Technology
- Chapter 2: Shrinking and Flattening the Globe: Scale, Projections, and Datums
- Chapter 3: Can I Map That? Maps to Depict Anything in Our World
- Chapter 4: Encoding Our World: Geographic Data Representation
- Chapter 5: How We Know Where We Are: Land Surveying, GPS, and Technology

Check out nullschool.net to view current wind patterns across the globe

Quick Facts about GEOG 160
The class that currently uses this website as a required text is:

PERSPECTIVES OF OPEN SCIENCE

Reflationary perspective

- making scientific results freely available during pre-publication.

The screenshot shows the EarthArXiv homepage. At the top, there is a navigation menu with a hamburger icon and the EarthArXiv logo. The main heading is "EarthArXiv". Below it is a search bar with the placeholder text "Search Preprints" and a magnifying glass icon. A prompt below the search bar says "Type your query and press enter to search." At the bottom, there is a green button labeled "START NEW SUBMISSION" and a link that says "Read about EarthArXiv or view list of Preprints."

The screenshot shows the arXiv website homepage. At the top, there is a navigation bar with the Cornell University logo and the text "We gratefully acknowledge support from the Simons Foundation and member institutions." The arXiv logo is prominently displayed. Below the logo is a search bar with a "Search" button and a "Login" link. A "Help | Advanced Search" link is also present. The main content area features a "Subject search and browse:" section with a dropdown menu set to "Physics" and a "Search" button. Below this is a "News" section with a link to "arXiv's blog". A red-bordered box on the right contains "COVID-19 Quick Links" and a list of links for "arXiv" and "medRxiv and bioRxiv".

The screenshot shows the medRxiv website homepage. At the top, there are logos for CSH, Cold Spring Harbor Laboratory, and BMJ Yale. The main heading is "medRxiv" with the tagline "THE PREPRINT SERVER FOR HEALTH SCIENCES". Below the heading is a search bar with a magnifying glass icon and a link to "Advanced Search". At the bottom, there is a red-bordered box with a cautionary note: "Caution: Preprints are preliminary reports of work that have not been certified by peer review. They should not be relied on to guide clinical practice or health-related behavior and should not be reported in news media as established information."

The screenshot shows the bioRxiv website homepage. At the top, there are logos for CSH, Cold Spring Harbor Laboratory, and BMJ Yale. The main heading is "bioRxiv" with the tagline "THE PREPRINT SERVER FOR BIOLOGY". Below the heading is a search bar with a magnifying glass icon and a link to "Advanced Search". At the bottom, there is a red-bordered box with a link: "COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv".

PERSPECTIVES OF OPEN SCIENCE

Constructivistic perspective

- Crowdsourcing & interdisciplinary



Missing Maps

LEARN EVENTS ABOUT BLOG EXPLORE en



Putting the World's Vulnerable Communities on the Map

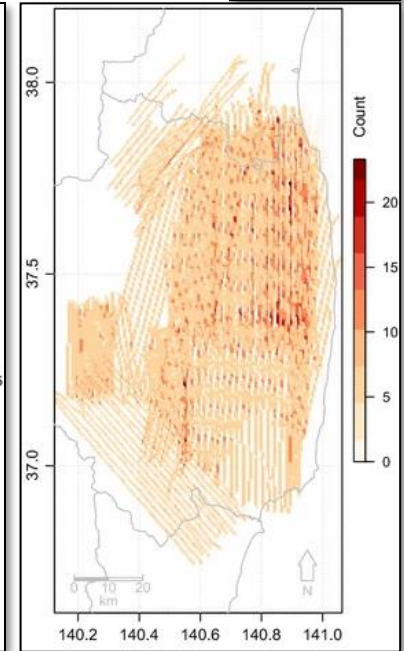
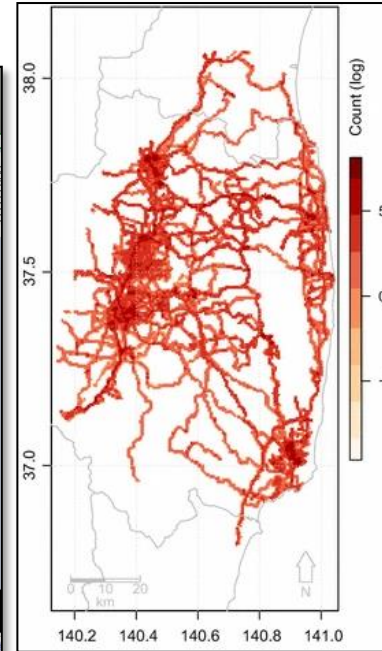
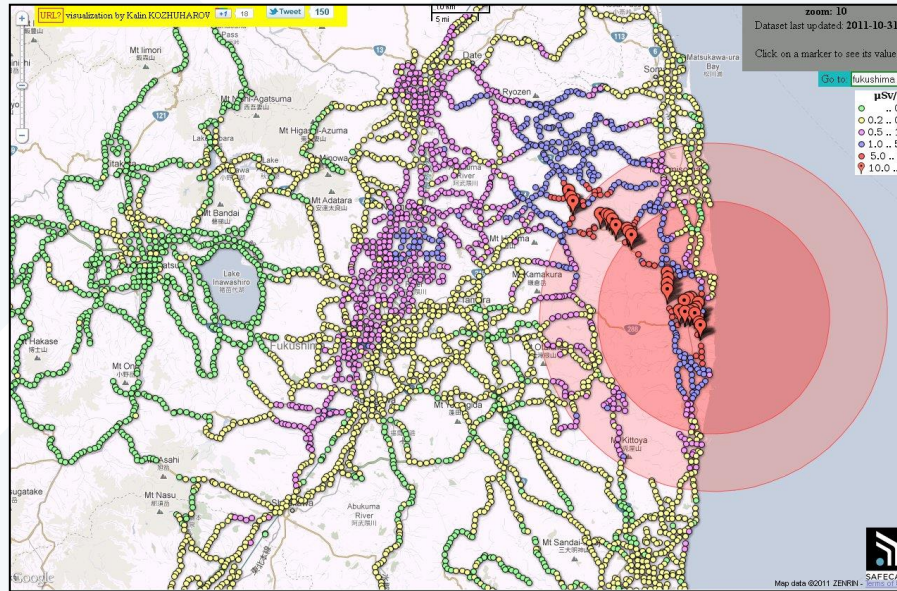
Each year, disasters around the world kill nearly 100,000 and affect or displace 200 million people. Many of the places where these disasters occur are literally 'missing' from open and accessible maps and first responders lack the information to make valuable decisions regarding relief efforts. Missing Maps is an open, collaborative project in which you can help to map areas where humanitarian organisations are trying to meet the needs of people who live at risk of disasters and crises.



PERSPECTIVES OF OPEN SCIENCE

Constructivistic perspective

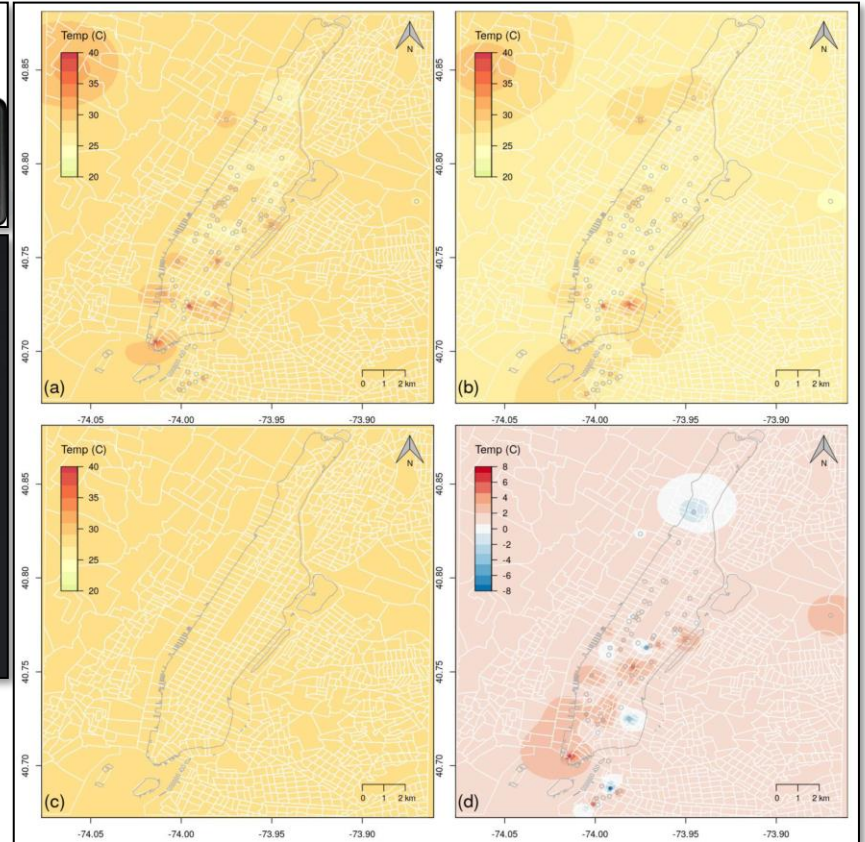
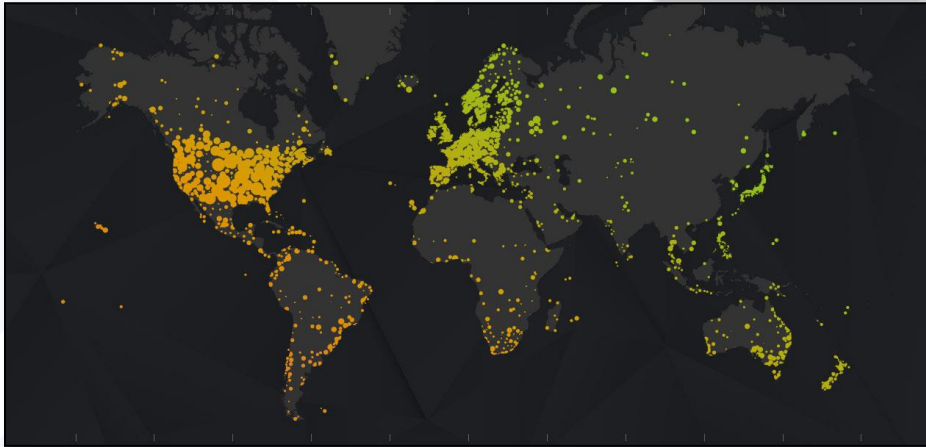
- Crowdsourcing & interdisciplinary



PERSPECTIVES OF OPEN SCIENCE

Constructivistic perspective

- Crowdsourcing & interdisciplinary




PERSPECTIVES OF OPEN SCIENCE

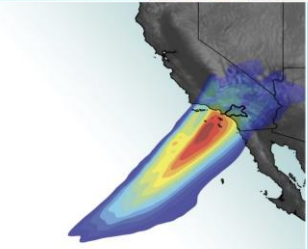
Exploitative perspective

- application-oriented knowledge

Prado Dam FORECAST INFORMED RESERVOIR OPERATIONS


Preliminary
Viability
Assessment
August 2021

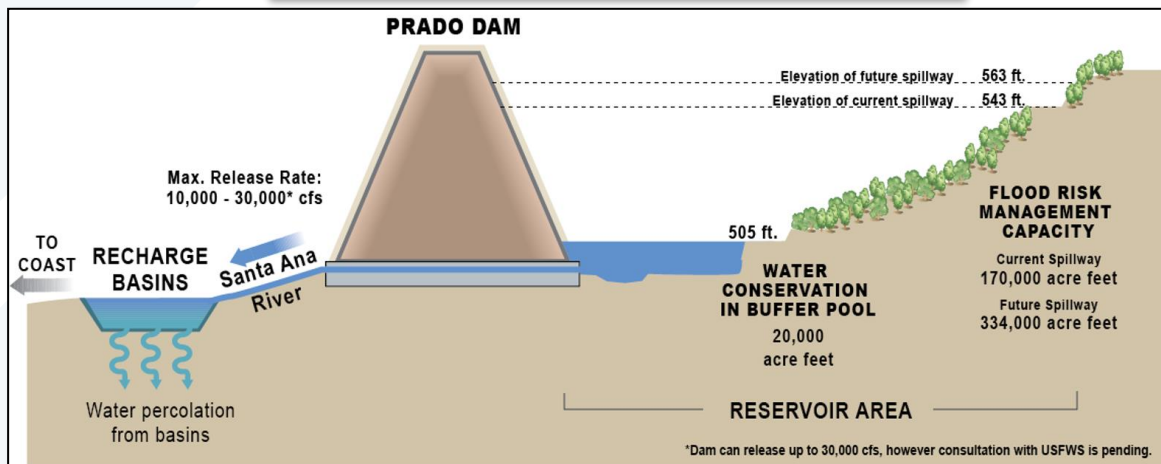
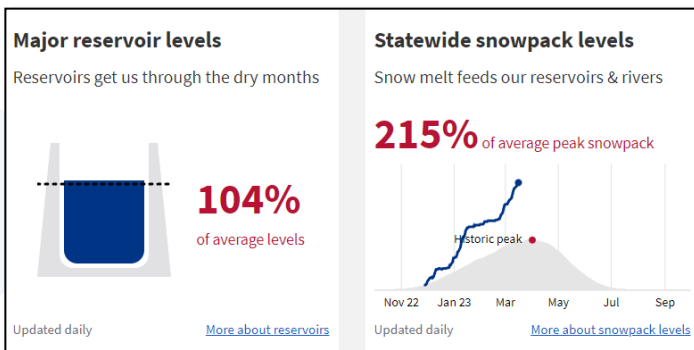




Prado Dam FIRO Steering Committee

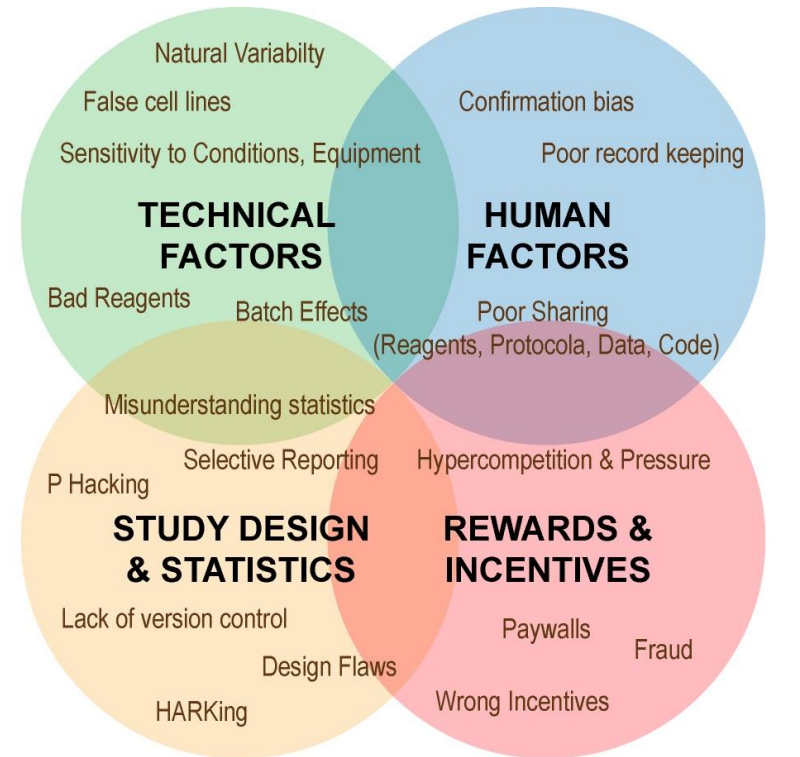
- **F. Martin Ralph**: CWSE Co-chair
- **Greg Woodside**: Orange County Water District Co-chair
- **Jay Jasperse**: Sonoma Water
- **Michael Anderson**: California Department of Water Resources (CDWR)
- **Cary Talbot**: USACE, Engineer Research and Development Center
- **Alan Haynes**: California Nevada River Forecast Center
- **Rene Vermeore**: USACE Los Angeles District
- **Jan Sweetman**: USACE Los Angeles District
- **James Tyler**: Orange County Public Works
- **Karin Cleary Bass**: U.S. Fish and Wildlife Service, Palm Springs





WHY DO WE NEED OPENNESS IN SCIENCE

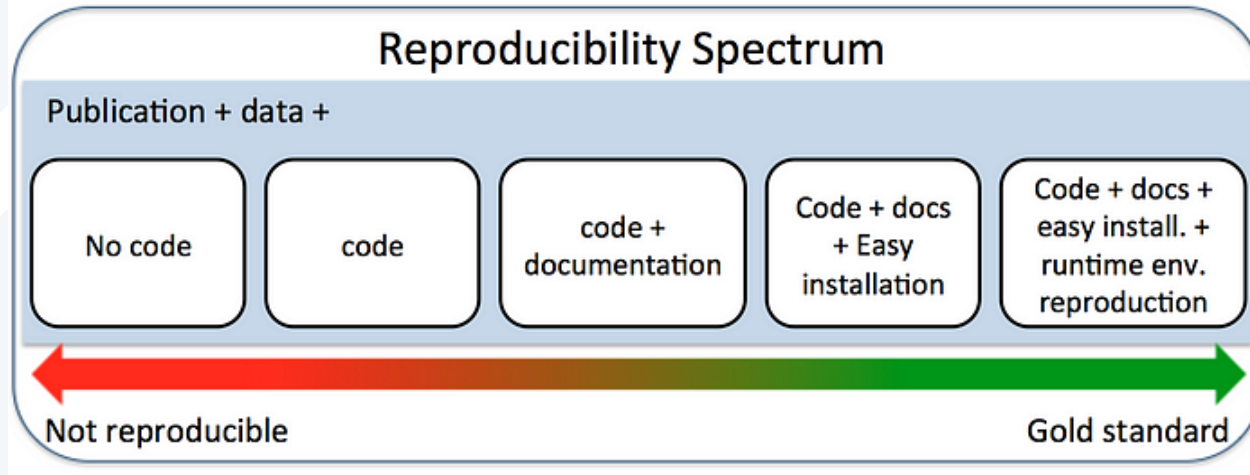
1. Validity (Reproducibility) of scientific findings



FACTORS DECREASING REPRODUCIBILITY

WHY DO WE NEED OPENNESS IN SCIENCE

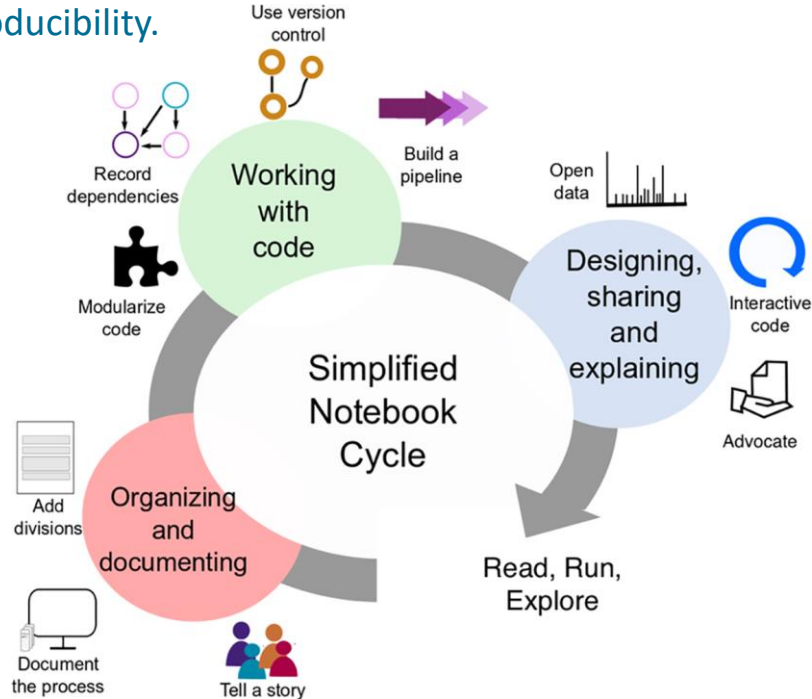
1. Validity (Reproducibility) of scientific findings



WHY DO WE NEED OPENNESS IN SCIENCE

1. Validity (Reproducibility) of scientific findings

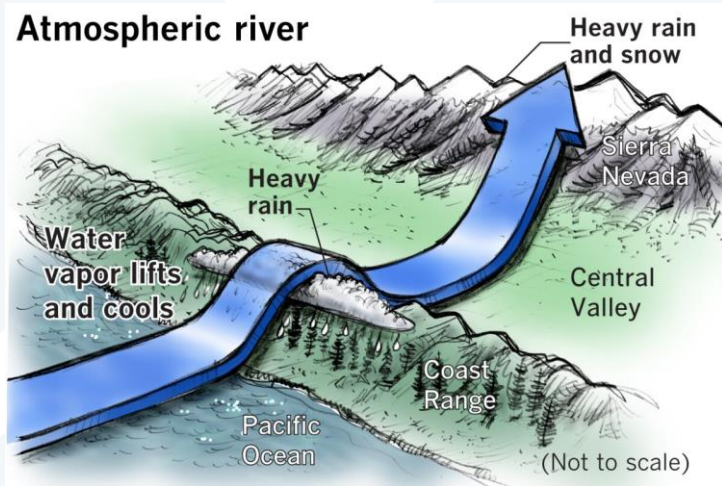
Relevance with Free Software: Free software promotes reproducibility.



WHY DO WE NEED OPENNESS IN SCIENCE

2. Increasing complexity and specialization of scientific problems

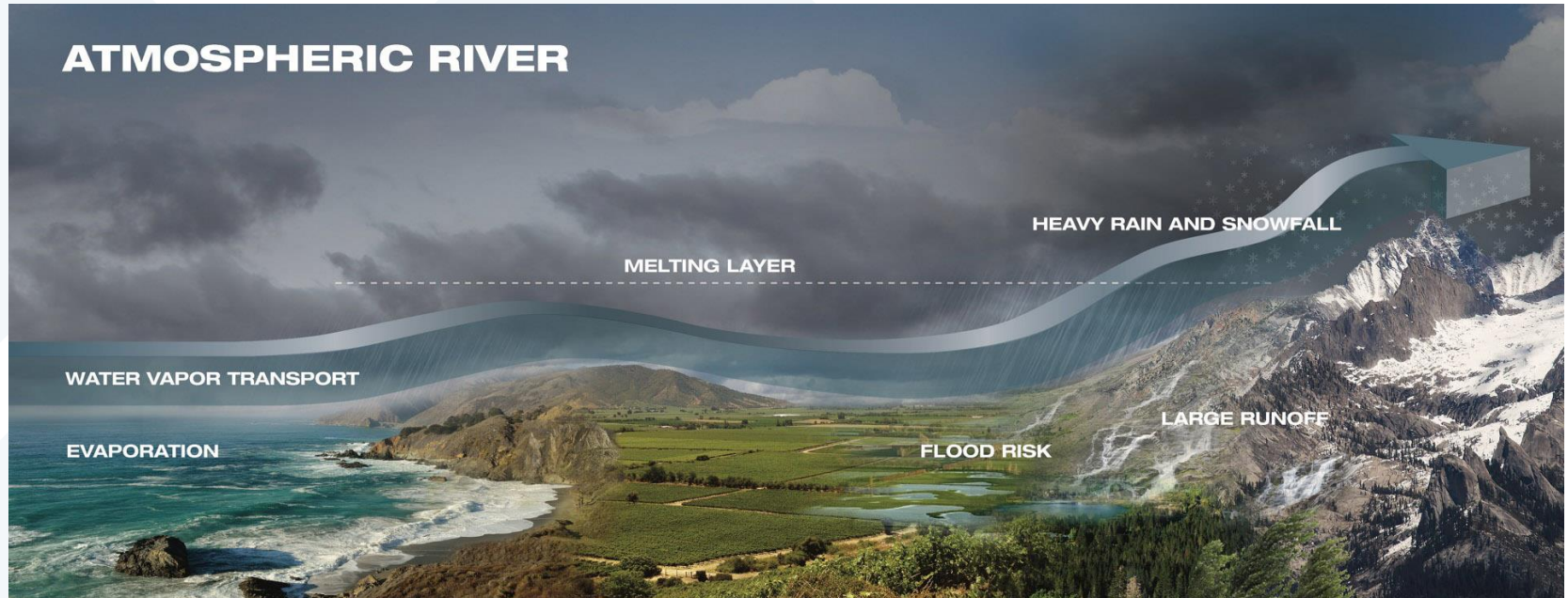
Example: How can we better understand atmospheric river?



WHY DO WE NEED OPENNESS IN SCIENCE

2. Increasing complexity and specialization of scientific problems

Example: How can we better understand atmospheric river?



WHY DO WE NEED OPENNESS IN SCIENCE

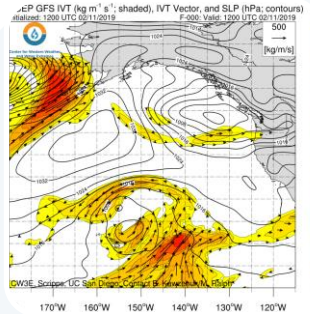
2. Increasing complexity and specialization of scientific problems

Example: How can we better understand atmospheric river?



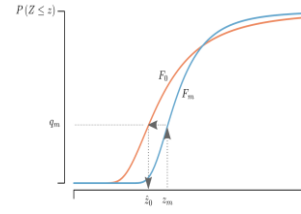
Observation

- Field researcher
- Engineer



Modelling

- Meteorologist
- Computer Scientist



Post-Processing

- Statistician
- Data scientist



Downstream Applications

- Policy maker
- Hydrologist
- Operator
- ...

WHY DO WE NEED OPENNESS IN SCIENCE

2. Increasing complexity and specialization of scientific problems

Example: How can we better understand atmospheric river?

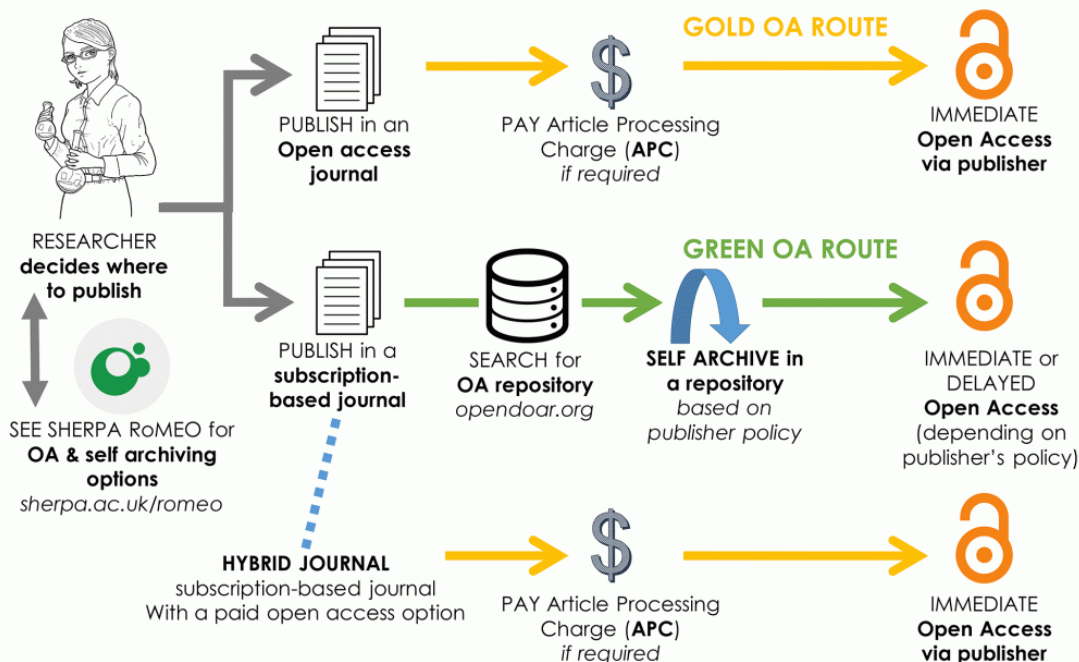
Relevance with Free Software: Free software facilitate communication and sharing.

OPEN SCIENCE IN A COMPETITIVE ENVIRONMENT

1. Sharing is risky, but the many benefits ultimately outweigh these risks.
 - Visibility
 - Collaboration
2. It drives innovation and prevents over-complexity.
3. Knowledge sharing doesn't necessarily mean giving away all your trade secrets.

OPEN ACCESS IS NOT FREE

Open Access Publishing



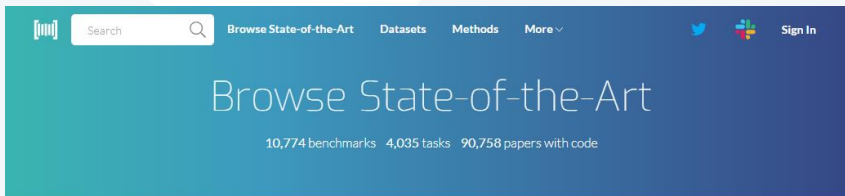
OPEN ACCESS IS NOT FREE

- Predatory journals
- On-going battles with peer review
 - Pre-publication and post-publication review
 - Do you know these phrases? “counterfeit consciousness”, “profound neural organization” and “colossal information”
- ...


FREE SOFTWARE FOR OPEN SCIENCE

1. Reproducibility

- Example: Paper with Code



Computer Vision



Semantic Segmentation
208 benchmarks
3642 papers with code






Image Classification
411 benchmarks
2927 papers with code



Object Detection
278 benchmarks
2734 papers with code



Contrastive Learning
2 benchmarks
1281 papers with code




Image Generation
211 benchmarks
1201 papers with code

▶ See all 1472 tasks

Natural Language Processing

U-Net: Convolutional Networks for Biomedical Image Segmentation

18 May 2015 · Olaf Ronneberger, Philipp Fischer, Thomas Brox · [18th ICV 2015](#)

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmural light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at <http://mb.informatik.uni-freiburg.de/people/ronneber/u-net>.

PDF Abstract

Code

- [labml/annotated_deep_learning_paper_implementations](#) 17,673 PyTorch
- [PaddlePaddle/PaddleSeg](#) 6,584 Python
- [mitelisl/Pytorch-U-Net](#) 6,513 PyTorch
- [open-mmlab/mmssegmentation](#) 5,396 PyTorch
- [quvvel/segmentation_models](#) 4,249 Python

Datasets



Results from the Paper

Ranked #1 on Semantic Segmentation on Kvasir-Instrument

Task	Dataset	Model	Metric Name	Metric Value	Global Rank	Result	Benchmark
Lesion Segmentation	Anatomical Tracings of Lesions After Stroke (ATLAS)	U-Net	Dice	0.6606	#3	🏆	Compare
			IoU	0.3447	#3	🏆	Compare
			Precision	0.5994	#1	🏆	Compare
			Recall	0.6449	#2	🏆	Compare
Retinal Vessel Segmentation	CHASE_DB1	U-Net	F1 score	0.7783	#11	🏆	Compare
			AUC	0.9772	#11	🏆	Compare

FREE SOFTWARE FOR OPEN SCIENCE

1. Reproducibility

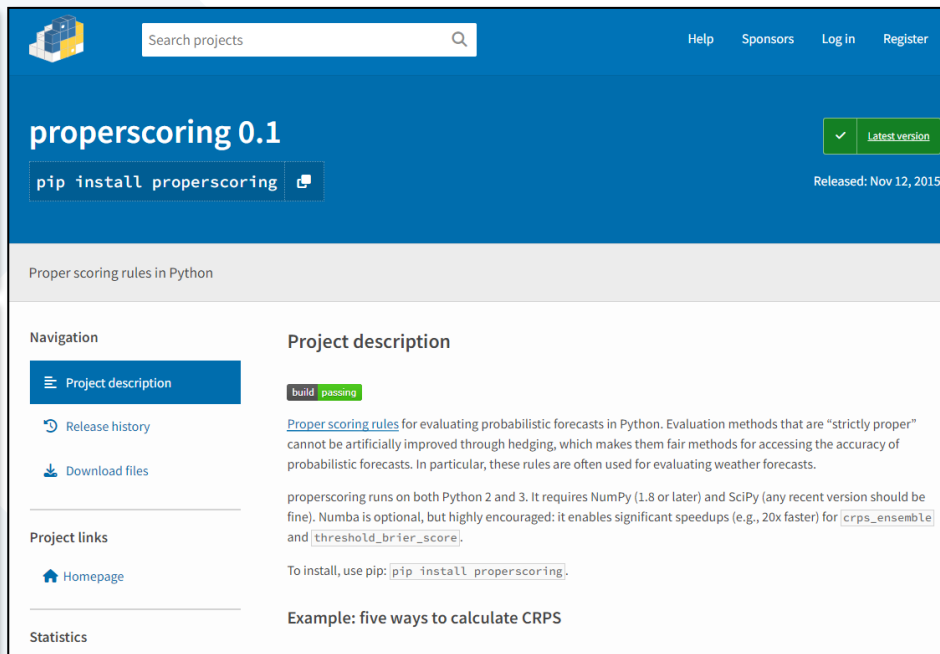
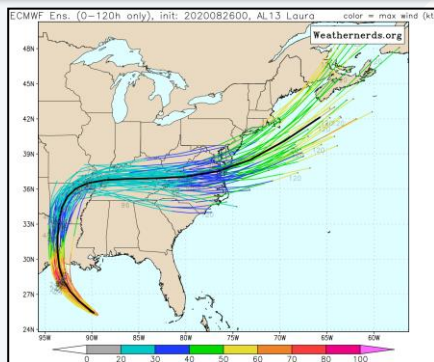
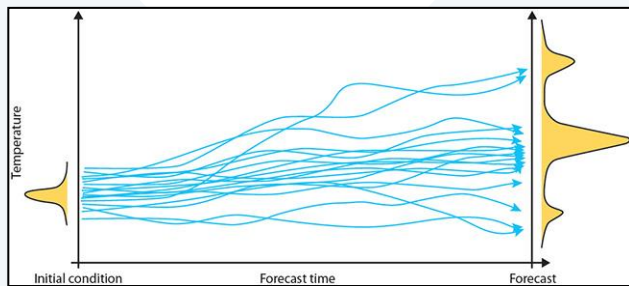
- Hosting platforms?



FREE SOFTWARE FOR OPEN SCIENCE

2. Collaboration

- Example: Weather verification



Search projects

Help Sponsors Log in Register

proberscoring 0.1

✓ Latest version

Released: Nov 12, 2015

```
pip install proberscoring
```

Proper scoring rules in Python

Navigation

- Project description
- Release history
- Download files

Project links

- Homepage

Statistics

Project description

build: passing

[Proper scoring rules](#) for evaluating probabilistic forecasts in Python. Evaluation methods that are “strictly proper” cannot be artificially improved through hedging, which makes them fair methods for accessing the accuracy of probabilistic forecasts. In particular, these rules are often used for evaluating weather forecasts.

proberscoring runs on both Python 2 and 3. It requires NumPy (1.8 or later) and SciPy (any recent version should be fine). Numba is optional, but highly encouraged; it enables significant speedups (e.g., 20x faster) for `crps_ensemble` and `threshold_brier_score`.

To install, use `pip install proberscoring`.

Example: five ways to calculate CRPS

The rapid intensification of hurricane Laura is underway, devastating landfall expected: <https://www.severe-weather.eu/tropical-weather/hurricane-laura-landfall-mk/>
Twenty-five years of ensemble forecasting: <https://www.ecmwf.int/en/about/media-centre/news/2017/twenty-five-years-ensemble-forecasting>

FREE SOFTWARE FOR OPEN SCIENCE

3. Accessibility and Pricing

- Example: Licensing

TEACHING FREE SOFTWARE FOR ENVIRONMENTAL SCIENCE

Is it possible to only use free / open-source software for environmental science?

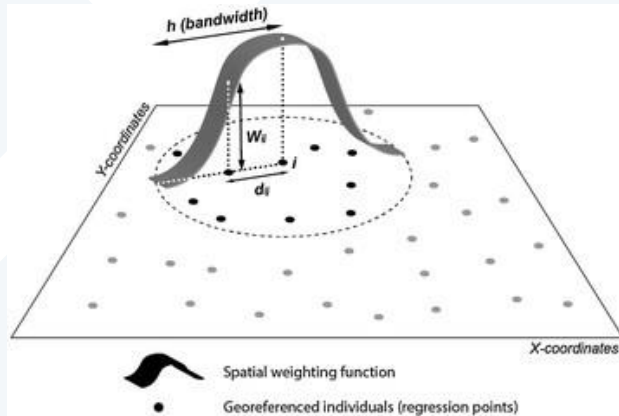


L^AT_EX



TEACHING FREE SOFTWARE FOR ENVIRONMENTAL SCIENCE

An example: Geographically Weighted Regression (GWR)



Add temporal dimension?

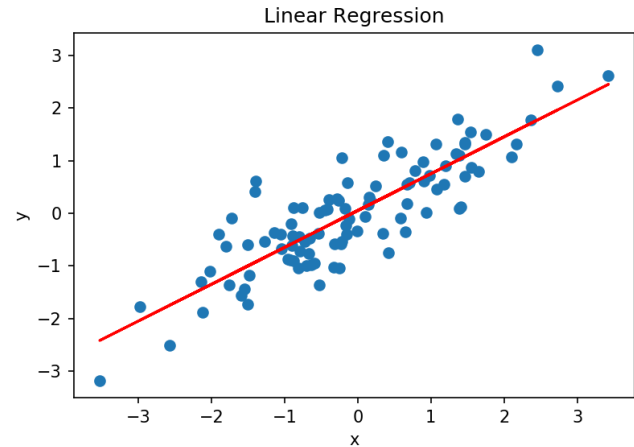
Process big data?

Source code?

TEACHING FREE SOFTWARE FOR ENVIRONMENTAL SCIENCE

- We teach the math first and then teach then how to use the toolbox.

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_tX_t + u$$



TEACHING FREE SOFTWARE FOR ENVIRONMENTAL SCIENCE

- We don't need to read the source code for learning purposes.
- Reading and teaching the source code is too time-consuming.
- ...

```
// C++ program for implementation
// of Bubble sort
#include <bits/stdc++.h>
using namespace std;

// A function to implement bubble sort
void bubbleSort(int arr[], int n)
{
    int i, j;
    for (i = 0; i < n - 1; i++)

        // Last i elements are already
        // in place
        for (j = 0; j < n - i - 1; j++)
            if (arr[j] > arr[j + 1])
                swap(arr[j], arr[j + 1]);
}

// Function to print an array
void printArray(int arr[], int size)
{
    int i;
    for (i = 0; i < size; i++)
        cout << arr[i] << " ";
    cout << endl;
}

// Driver code
int main()
{
    int arr[] = { 5, 1, 4, 2, 8};
    int N = sizeof(arr) / sizeof(arr[0]);
    bubbleSort(arr, N);
    cout << "Sorted array: \n";
    printArray(arr, N);
    return 0;
}
// This code is contributed by rathbhupendra
```

```
// C++ program for implementation
// of Bubble sort
#include <bits/stdc++.h>
using namespace std;

// A function to implement bubble sort
void bubbleSort(int arr[], int n)
{
    int i, j;
    for (i = 0; i < n - 1; i++)

        // Last i elements are already
        // in place
        for (j = 0; j < n - i - 1; j++)
            if (arr[j] > arr[j + 1])
                swap(arr[j], arr[j + 1]);
}

// Function to print an array
void printArray(int arr[], int size)
{
    int i;
    for (i = 0; i < size; i++)
        cout << arr[i] << " ";
    cout << endl;
}

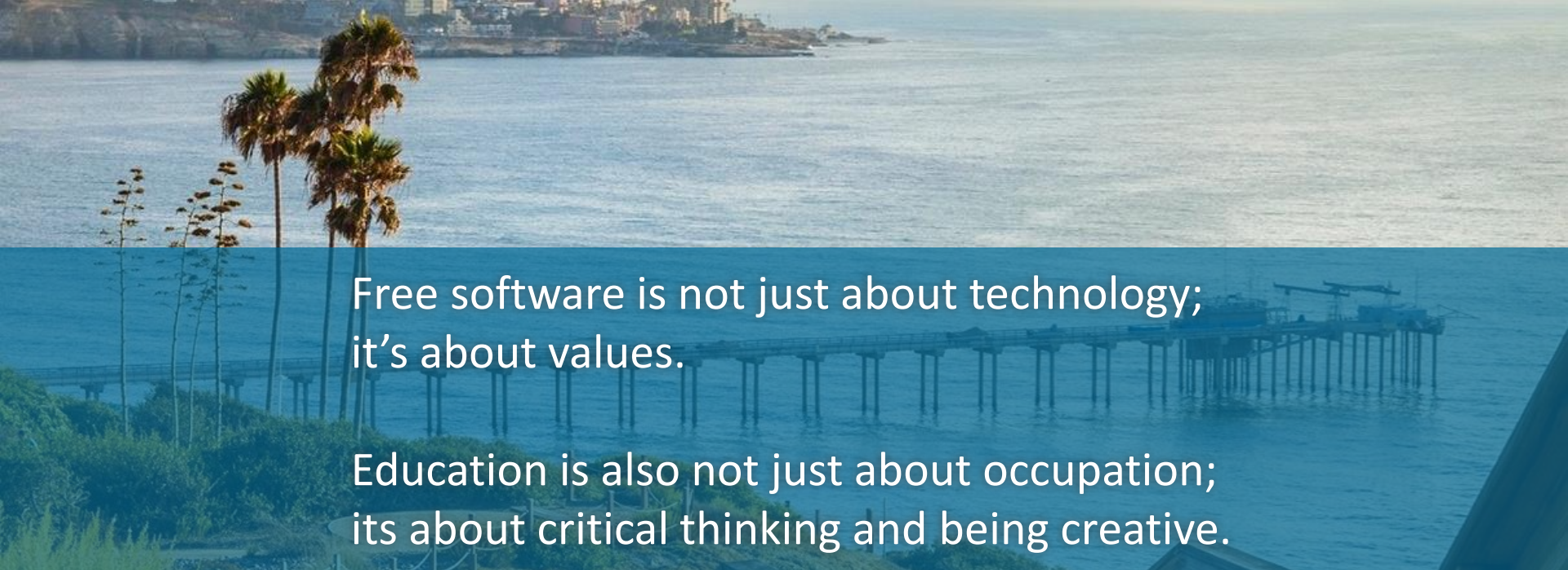
// Driver code
int main()
{
    int arr[] = { 5, 1, 4, 2, 8};
    int N = sizeof(arr) / sizeof(arr[0]);
    bubbleSort(arr, N);
    cout << "Sorted array: \n";
    printArray(arr, N);
    return 0;
}
// This code is contributed by rathbhupendra
```


TEACHING FREE SOFTWARE FOR ENVIRONMENTAL SCIENCE

1. Choose the right software
2. Provide context
3. Start with examples and hands-on tutorial
4. Guide them to investigate the implementation
5. Collaborative learning
6. Make it fun

WHAT CAN THE DEV COMMUNITY DO

1. Be patient
2. Create educational resources
3. Collaborating with educators



Free software is not just about technology;
it's about values.

Education is also not just about occupation;
its about critical thinking and being creative.

