Reproducibly Analyzing Wildfire, Drought, and Flood Risk with NASA Earthdata Cloud

Table of Contents

1. Project Summary

As the climate changes, prediction and management of the risk of wildfire, drought, and floods has become increasingly challenging. It is no longer sufficient to assume that what has been normal and historic for the last century will occur with the same frequency into the future. These natural risks are intrinsically linked to the changing distributions of surface water, precipitation, vegetation, and land use in both time and space. At the same time, there are now hundreds of petabytes of relevant Earth science data available through the NASA Earthdata Cloud that can be used to understand and forecast these water-dependent environmental risks. With the volume of Earth science data growing dramatically year over year, it is important for scientists to understand how open science and cloud-based data intensive computing can be used to reproducibly analyze and assess the changing risk profile of wildfire, drought, and floods.

In this proposed TOPS ScienceCore module, learners will learn to identify, extract, analyze, visualize, and report on data available through NASA Earthdata Cloud for three different scenarios: wildfire, drought, and flood risk. The module will build upon TOPS OpenCore and reinforce principles of reproducibility and open science-based workflows. Computationally, the scenarios will estimate changes in the hydrological water mass balance for defined regions primarily using remote sensing data. We will demonstrate best practices in “data-proximate computing” by considering examples that involve computing climatologies and other statistics from long-time series using numerical methods that scale well with the data being available on the cloud. This module will leverage scientific Python libraries such as Xarray and Dask to perform the computations. The focus of this module will be on data processing and visualization and doing so in a reproducible and transparent way.

After completing this module, learners will be able to adapt and remix the scenarios for their own open science objectives regarding environmental risks such as wildfire, drought, and flood. These risks are common worldwide yet need to be each analyzed in a regional context. The module will provide concrete examples that showcase how open science can be done.

The module will be written as an extension to the OpenCore framework and all course materials will be open, available in English and Spanish, and accessible in the vision, hearing, mobility, and attention dimensions. The ScienceCore module will be released as one or more Jupyter notebooks on GitHub with supporting material for delivering the course using the cloud either for in-person or for virtual cohorts.

2. Scientific/Technical Management

2.1. Introduction

With a changing climate, wildfires, droughts, and floods continue to be significant risks across the United States and the rest of the world [1,2]. Events that used to occur only once a century are now occurring every few years. Historical norms for frequency of extreme climate leading to episodic disaster are not sufficient to infer the frequency into the future.

Flooding is the most significant environmental disaster affecting more than two billion people a year. Floods cause significant damage to infrastructure, displace people, and lead to disease. The frequency of flooding has increased in recent years due to changes in rainfall and land use.

The data from the National Interagency Fire Center shows significant growth in the size of wildfires in the US over the last 25 years. In some years, over 10 million acres in the US have been burned by wildfires. Canada has also experienced significant recent wildfires resulting in loss of property, livestock, and industry.

Floods and wildfires are intrinsically linked to underlying water conditions: too much or too little water. Droughts, although on a longer timescale, are also caused by too little water being retained in the environment. Droughts, wildfires, and floods can all occur over the course of a few short months in the same region. The key factor is the abundance, or absence, of precipitation (rain and snowfall) either in short-term events or long-term shifts of how that water is retained by the land.

Floods, wildfires, and droughts are at risk to increase in frequency and intensity due to fundamental shifts in extreme precipitation caused by climate change. Fundamentally, these three environmental risks are about water no longer having the distribution that they have had in the past. We can recognize that our world is changing and causing risk. How can we mitigate that risk?

NASA Earthdata Cloud has moved petabytes of data into the cloud and that data is ideally suited to answering questions about climate change risk. However, practitioners don't yet have the proper skills and training to take advantage of these amazing resources. It is no longer sufficient to search for a dataset and then download it locally. The sheer size of the available data makes this not just impractical but, in many cases, impossible. But how do scientists attempting risk assessment complete their work when it is so difficult to download the needed data to a computer to allow running their analysis code locally? The answer is to instead perform *data-proximate computing* by pushing the analysis code to a computer that is very close (in terms of both costs and latency using the Internet) to where the data is hosted. For data that is hosted in the commercial cloud (such as NASA Earthdata), this means running data analysis on computers in that same cloud environment.

This ScienceCore module, building on top of the OpenCore modules, will teach learners how to access the NASA Earthdata Cloud and produce dashboard-based visualizations and analyses of water sensing data. This data can then be compared with model outputs to forecasts, the new 'normals' for wildfire, drought, and flood risk across the world.

This risk assessment is highly localized and needs to be repeated for every country, state, city, and village. Demonstrating how to leverage NASA Earthdata cloud data in an open, reproducible way will aid thousands of scientists and analysts to produce the reports they need for their own communities.

There is currently a barrier for scientists to use NASA Earthdata Cloud. They do not have the skills and expertise to analyze data at scale in the cloud. This ScienceCore module will teach that skill.

2.2. Objectives and Expected Significance

Teaching a large number of users about reproducibly analyzing earth data will accelerate our ability to mitigate and adapt to climate change. So a 'science objective' is determining if this ScienceCore module will actually help with those adaptation efforts.

This new ScienceCore module will solve the need to apply earth sensing data to climate risk assessment. More generally, it will serve as a template for future ScienceCore modules in this domain. As part of this work, we will not only develop the ScienceCore module but measure its effectiveness in meeting its learning objectives.

Open science is important for climate change because it helps to ensure that the research and data related to climate change is accessible to everyone. This means that anyone, regardless of their expertise or background, can review and verify the research, which helps to build trust in the findings. In addition, open science promotes collaboration and sharing of ideas among researchers, which can lead to more rapid progress in understanding and addressing climate change. By making research open and transparent, we can ensure that the best science is being used to inform decisions about how to address climate change.

2.3. Impact of proposed work to state of the art

Climate risk needs to be reevaluated at the national, state, and municipality levels. Companies and non-governmental organizations also need to assess climate risk. NASA data contained in the Earthdata cloud is highly relevant to answering the question of assigning climate risk. This ScienceCore module will provide a template for accessing the data and then analyzing it reproducibly, in a consistent and open manner that exemplifies the best practices identified in the TOPS OpenCore content.

The intended audience for this ScienceCore module is people tasked with producing climate risk assessments. There is potential for taking the tooling and data that we collectively now have available from these global models and downscaling that information to every country, region, state, city, town, and village on anticipated climate impacts. Policy and planners at all levels are beginning to tackle the questions of taking these large, global scale data products and figuring out what they mean for their locality. The community is thinking about how the physical climate variables (e.g. temperature and precipitation) affect risks such as flooding, wildfires, droughts, sea level rise, food sustainability, among many others.

JupyterHub infrastructure will provide a platform to deliver this module. This infrastructure can be deployed by any team with intermediate DevOps and Kubernetes knowledge. Using the open infrastructure allows the content to be run by any organization.

2.4. Relevance of proposed work to announcement

This ScienceCore module is focused on helping people access and analyze data from NASA, including data that is stored in the cloud. This includes using open-source tools and libraries to analyze and visualize the data, and to create and share reproducible research workflows. The project also includes modules that build on existing OpenCore concepts, and that cover important topics in different scientific disciplines.

These modules can be accessed through the TOPS Open edX platform or as Jupyter Books on the TOPS GitHub. The goal of ScienceCore is to make it easy for people to work with NASA data, and to collaborate and share their research with others. The project is designed to be accessible, open, collaborative, multilingual, and interactive. All of the final products created through ScienceCore will be openly licensed and shared on the TOPS GitHub, and proposals for funded projects must include plans for collaboration and participation in annual coordination meetings.

2.5. Technical approach and methodology

To create this OpenCore module we will bring together **pedagogical experts,** with training in open and inclusive content design, alongside **scientific content specialists** to design course material that follows best-practices in open science, data analysis, and domain methodologies.

This module will contain three fully worked examples of reproducible analysis of an environmental risk. During course delivery, an instructor may choose to only go through one of these worked examples and leave the others for reference. Each example will be independent.

The largest risk is that the technology for accessing cloud enabled data is rapidly changing. It is possible that the specific code examples developed in this ScienceCore module will be obsolete or otherwise out of date in the near term. However, the science content and especially the open science content has a much longer persistence.

Any of the ‘code exercises’ will be written so that they can be replaced in the future without having to fundamentally change the narrative of the module. The science objectives (data discovery, data reduction, visualization, comparison to model output, use in machine learning) will remain even if the libraries we use continue to undergo rapid development.

Another problem is how to scope this module so that participants have the right background for it to be useful. In a relatively short half-day course, learners will already have to have a sufficient background knowledge in data driving computing, programming, visualization as well as risk assessment, remote sensing data, and statistical analysis to be able to directly apply this module. This ScienceCore module will contain links and reference to ‘additional material’ that can help learners to prepare in advance with background material as needed. So the module is taught to those for whom it was designed, the module will provide six learner personas [3] including the person’s general background, what they already know, what they want to achieve, and any special needs they have.

As well, the module will include pre-assessment tools for assessing learners' knowledge of climate science and risk assessment, proficiency with Jupyter notebooks and Python programming, as well as years of experience related with these areas to control for those piloting the ScienceCore module who may have not yet developed the adequate skills.

The module will be designed using a backward design [4,5,6] incorporating guidelines like those in [7], [8],and [9]. Module backward design entails:

1) Creating learner personas to determine and clearly communicate what audience the module is designed for.
2) Writing an initial draft of topics that will be included and not included.
3) Creating one summative assessment including the contents learners will have to master to obtain this module badge.
4) Creating all the formative assessments so learners practice contents throughout the module. About one formative assessment per teaching unit, comprising 5 to 9 new concepts and at most after every 15 minutes of content, ensures the module uses an active teaching style and allows both trainers and learners to assess their progress and adapt to the requirements of the occasion.
5) Ordering formative assessments considering complexity, dependencies, and how a topic motivates learners.
6) Writing the content that will guide learners between formative assessments.
7) Generating a succinct module description for promoting it and reaching interested learners.

2.5.1. Task 1: Creation of module content in English and Spanish

1. Review ‘reproducible’ workflows from OpenCore
2. Introduce working with cloud data in general and NASA Earthdata Cloud
3. Problem Statement: Environmental Risk Assessment
4. Fully worked examples:
    a. Example 1: Flood
    b. Example 2: Wildfire
    c. Example 3: Drought

2.5.2. Task 2: Assessment of module content

1. Development of a pre- and post-survey
2. Ensure alignment with OpenCore modules
3. Piloting of new ScienceCore over at least eight offerings (need to cover English/Spanish, Virtual/In-person, Wildfire/Flood/Drought)

2.6. Plan of Work

July - December 2023

- First pass at content creation of ScienceCore materials with one fully worked example (Flooding)
- Preparation of a survey tool to assess the learning outcomes of the module
- First Pilot of ScienceCore Material at or related to AGU2023
    - Virtual and In-Person, Spanish and English (4 offerings)
    - Use survey to assess utility of the ScienceCore module

January - June 2024

- Development of second worked examples (Wildfire)
- Revise curriculum based on results of first ScienceCore pilot
- Refinement of visual assets and graphics
- Filming of recorded videos for online learning (English and Spanish versions)

July – December 2024

- Development of a third fully worked example (Drought)
- Second Pilot of ScienceCore Material
    - Look to partner with organizations who are leading both Virtual Cohorts or In-person cohorts of OpenCore

- Possible delivery schedule for second pilot. Minimum goal of 6 offerings of the module.

Format	Language	Worked Example
In person	English	Wildfire
Virtual	English	Flood
In person	English	Drought
Virtual	Spanish	Wildfire
In person	Spanish	Flood
Virtual	Spanish	Drought

- Use survey again to assess utility of the ScienceCore module

January – June 2025

- Deployment of a ‘learner-directed’ version of the ScienceCore module via OpenEdX or equivalent program
- Refinement and final release of course materials based on the second pilot.
- Creation of sufficient documentation that the required JupyterHub infrastructure could be deployed on the commercial cloud to allow the running of this ScienceCore module.

2.7. Management Structure

The project team (Principal Investigator, Content Specialists, Instructional Design team) will meet virtually every two weeks for the duration of the project. The specific tasks will be organized using GitHub issues on a publicly accessible repository. The PI will interface directly with Engineering support as needed.

2.8. Description of Contributions

PI will act as project manager for the duration of the project and be responsible for ideation and oversight of the project.

Content Specialists (having advanced training in climate risk assessment and remote sensing experience) will be recruited to author the course materials. Additional work for specialized skills such as illustrators or voice talent will be subcontracted as required. Engineering support will set up and maintain the JupyterHub environment needed to deliver this ScienceCore module.

The Instructional Designer team, including an accessibility specialist with previous experience in both designing and delivering OpenCore content, will ensure that the ScienceCore module aligns with OpenCore's learning style and format. We will ensure materials are also prepared in Spanish and contextualized at all stages of the development process.

Administrative staff will provide financial, administration, and process support to the project.

3. References

[1] Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., & Zhou, B. (2021). Weather and Climate Extreme Events in a Changing Climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1513–1766). Cambridge University Press.

[2] Hicke, J. A., Lucatello, S., L.D., M., Dawson, J., Aguilar, M. D., Enquist, C. A. F., Gilmore, E. A., Gutzler, D. S., Harper, S., Holsman, K., Jewett, E. B., Kohler, T. A., & Miller, KA. (2022). North America. In H. O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1929–2042). Cambridge University Press.

[3] Wilson, G. (2020). *Teaching tech together: How to make lessons that work and build a teaching community around them*. CRC Press.

[4] Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed). Association for Supervision and Curriculum Development.

[5] Biggs, J. B., & Tang, C. S. (2011). *Teaching for quality learning at university: What the student does* (4th edition). McGraw-Hill, Society for Research into Higher

Education & Open University Press.

[6] Fink, L. D. (2013). *Creating significant learning experiences: An integrated approach to designing college courses* (Revised and updated edition). Jossey-Bass.

[7] MetaDocencia Team. (2022). *Hoja de rota modelo para desarrollo de Cursos*. https://doi.org/10.5281/ZENODO.7390559

[8] Via, A., Palagi, P. M., Lindvall, J. M., Tractenberg, R. E., Attwood, T. K., & Foundation, T. G. (2020). *Course design: Considerations for trainers – a Professional Guide*. https://doi.org/10.7490/F1000RESEARCH.1118395.1

[9] *Collaborative Lesson Development Training: Summary and Setup*. (n.d.). Retrieved December 7, 2022, from https://carpentries.github.io/lesson-development-training/

4. Open-Source Science Development Plan

All materials created under this proposal will be released under a CC-BY permissive license and made available on a publicly accessible GitHub repository.

All training outputs will be made publicly available as they are created. Individual calls have both blank “shared document templates” and completed post-meeting shared documents, where learners can take training notes and solve some of the formative assessments, which will be deposited online on GitHub and Zenodo, under a CC-BY 4.0 license. Slides from presentations will be deposited on Zenodo. All online training will be recorded, except for private group discussions in breakout rooms. All recordings will be deposited on Youtube, again with a CC-BY license and full Spanish or English transcriptions to facilitate hearing, screen-reading and language accessibility. Where code and scripts are used, all code content will be stored on GitHub under a permissive OSI Licence. Code content is accompanied by detailed documentation such that others can contribute to it.

4.1. ScienceCore Module Impact Measurement

We will measure the impact and performance of this ScienceCore module. The impact study will produce data which will be in the form of survey responses. We will share the survey protocols and methods openly on protocols.io, including any recruitment advertisements and survey questions.

Where ethically appropriate and safe we will share impact study response data anonymously or aggregated, as permitted by participants consent. Raw data will be stored securely on cloud-based servers (such as dedicated Dropbox for Business or GSuite) until processed (i.e. anonymised where necessary). Raw data will be retained for a minimum of five years unless ethics review or institutional repository requirements stipulate otherwise.

Once processed, data will be deposited on an open repository such as Zenodo, alongside metadata "README" files that describe the shape of data and the meaning of specific columns.

Where possible, data will be stored in re-usable formats, i.e. CSV and JSON rather than proprietary formats, to facilitate re-use without the need for expensive or proprietary software.

We may use codes and scripts to analyze data, tidy the data, and produce visualizations. Where this occurs, the scripts and clear running documentation and dependency data will be stored on GitHub under a permissive OSI license, and versioned snapshots of the code will be deposited on Zenodo.

4.2. Results dissemination

In addition to sharing the data, protocol, and scripts, we will ensure that all publications associated with our impact study and training program are posted as preprints to facilitate early comment and distribution, and finally in open-access journals.

4.3. This proposal

This proposal will be deposited on Zenodo to get a DOI as well as shared via the Open Grants portal (https://www.ogrants.org/).

5. Table of Personnel and Work Effort

Team	**Person / Role**	**FTE**	**Description and specialized skills**
Principal Investigator	PI	5%	Scientific oversight, ideation, and project management
Instructional Design	Co-I-1	10%	Training design and oversight
Instructional Design	Co-I-2	5%	Accessibility of training
Instructional Design	Co-I-3	5%	Training coordination
Instructional Design	Co-I-4	5%	Training infrastructure
Support	Cloud engineer	5%	Initial setup and maintenance
Support	Administrator	5%	Grant administration
Content Specialist	Research associate	10%	Data access with NASA Earthdata Cloud
Content Specialist	Research associate	5%	Floods
Content Specialist	Research associate	5%	Wildfire
Content Specialist	Research associate	5%	Drought

6. Equal Access Plan

All team interactions under this project will follow the existing code of conduct in place at the applicant's organization.

For any virtual or in-person pilot delivery of this ScienceCore module will follow both this code of conduct and the Equal Access Plan of the relevant partner organization.

The Instructional Design Team will anticipate general accessibility needs (e.g. alt-text for visually impaired people, or automatic captioning for those hard of hearing) and will also be available to facilitate participation taking into account individual requirements. We will provide live interpretation between Spanish and English as needed to increase engagement and the exchange of ideas between participants.

7. Redacted Budget Justification: Narrative and Details

This project is scoped as a two year project that will

1) Develop a new ScienceCore module
2) Evaluate the effectiveness of the module through a series of pilots
3) Use this experience to refine the module for use by others

Developing this ScienceCore module requires bringing together a diverse team with specialized skills in several areas, and the majority of our costs will cover the time of several teams to complete this work. Below is a brief explanation of the roles and teams we've defined for this work.

- Principal Investigator/Project Manager ~3hr/week
- Instructional Design Team ~10hr/week
- Content Specialists Team ~10hr/week
- Engineering Support ~2hr/week
- Project Administration ~2hr/week

Our budget covers two years along with an 8% increase from year 1 to year 2 due to anticipated inflation. We provide a description of major budget items below.

The **Instructional Design Team** brings extensive experience in curriculum design and implementation. This team has been involved in delivery of the Carpentries program and the creation of OpenCore content, and has a specialty in designing learning material that is inclusive and accessible. This team is completely bilingual in English and Spanish.

The **Content Specialist Team** will have expertise in the scientific workflows and questions that are covered by this module (for example, remote sensing data, machine learning expertise applied to climate modeling, and hydrological modeling for content such as droughts and flooding). They will be recruited from pre-existing community partners of the applicant organization. Each of these partners are research oriented communities currently using managed JupyterHub service to do their science, leveraging notebooks and cloud based data, including data available on NASA Earthdata Cloud. We will leverage our existing relationships with these groups to identify and recruit postdocs, senior graduate students, or other researcher associates to create content in collaboration with our Instructional Design Team. We will identify partners with both a general expertise in accessing data with NASA Earthdata Cloud using Jupyter notebooks, as well as specialized knowledge about the environmental hazards of wildfires, floods, and drought. This ScienceCore module will draw examples and content from the scientific community currently using these tools. We will identify two or three members of partner communities with the necessary background, , and will fill in the gaps in domain content with short term contracts as needed. At least one of these scientific Content Specialists will be bilingual in English and Spanish as well.

The use of **cloud data** presumes there is computational infrastructure available in the public cloud to develop and pilot this new ScienceCore module. As such, we include costs for **cloud infrastructure, deployment, and operational support**, which will be provided by the organization's engineering team to support the development of content as well as our pilot training sessions.

Project Coordination and **scientific oversight** will be provided by the Principal Investigator. **Administrative support** for the project will be provided by the organization's administrative team.

We include **travel budget** for representatives of the project team to conferences and workshops such as AGU to pilot in-person this new ScienceCore module training. We will also send members of the team (PI, Content Specialists, Instructional Design Team) to the annual 4-day TOPS coordination meetings in Washington DC.

Finally, we include a budget for **live interpretation, graphical design and other creative services** to cover production costs related to course materials.

8. Facilities and Equipment

The applicant organization has extensive expertise in deploying and operating cloud-based infrastructure for interactive computing in the sciences, and has several partnerships with scientific communities of practice for cloud-based workflows in domains that are relevant to this proposal. This team will leverage the same scalable and open deployment infrastructure used for other communities in order to provide cloud infrastructure for this proposal. They will provide a managed JupyterHub for the duration of the project (July 2023 – June 2025), which will support developing course materials and repeated offerings of the ScienceCore module.