Citizen Climate Science on CloudLab

Semil Jain, Dr. Robert Ricci, and Hamza Sheikh

[1] University of Utah
semil.jain@utah.edu
[2] University of Utah
ricci@cs.utah.edu
[3] University of Utah
hamza@cs.utah.edu

Abstract. Citizen science offers a way for everyone to get involved in climate research, yet challenges persist in providing access to computational resources essential for modeling complex climate phenomena. In response, the National Discovery Cloud for Climate (NDC-C) initiative is proposed, leveraging CloudLab's infrastructure to facilitate citizen access to climate models. Through the creation of user-friendly model profiles and streamlined portal development, the NDC-C aims to democratize access to climate simulations. Additionally, efforts to federate resources and launch an inclusive access program underscore the initiative's commitment to fostering a diverse community of citizen scientists and educators. This project outlines our work towards creating NDC-C and how it could change the game for citizen climate science.

Keywords: Citizen Science · Climate Models · Cloudlab.

1 Introduction

1.1 Citizen Science

Citizen science engages volunteers in scientific research, enabling them to collect, analyze, and interpret data alongside professionals across disciplines like environmental monitoring, biodiversity, and climate science. It empowers non-professionals to contribute, fostering collaboration between scientists and the public through technology-driven tools like mobile apps and online platforms. Projects span diverse fields, reflecting participants' interests, from environmental monitoring to climate research. They can be scientific authority-based, following traditional methods, or social movement-based, prioritizing community engagement and advocacy [15]. In essence, citizen science promotes inclusivity, collaboration, and innovation, advancing scientific knowledge and engaging society in the scientific process.

1.2 CloudLab: Advancing Cloud Computing Research

CloudLab[4] is a dynamic platform empowering researchers to explore the future of cloud computing. With its flexible infrastructure[1], researchers can experiment with custom cloud architectures, from provisioning to execution, with granular control and visibility. Leveraging resources across multiple sites in the US and globally, CloudLab offers nearly 1,000 machines for comprehensive research endeavors. Researchers can deploy standard cloud software stacks or craft custom architectures, pushing the boundaries of cloud capabilities. Interconnecting with existing testbeds like GENI and Emulab facilitates collaboration across a global network. CloudLab's impact spans various fields, with hundreds of published papers showcasing its role in transformative research. Through its profile-based approach, CloudLab streamlines the setup of cloud environments, enabling precise control over system components for experimentation. Led by a consortium of academic institutions and supported by the NSFCloud program, CloudLab drives innovation in cloud computing research.

2 Proposal

Climate science research necessitates more than just access to models; it requires appropriate facilities for storage, computation, and automated code packaging. To address this need, we propose leveraging CloudLab's computing resources and profile[2] mechanism to create a platform where scientists can initiate small-scale experiments and prototype solutions.

- we will develop profiles based on NCAR's containerized atmosphere models[13], offering one-click access to execute these models.
- we will explore the requirements for federating code and data within these profiles with other environments, laying the groundwork for future integration with the National Discovery Cloud for Climate (NDC-C)
- we will launch a pilot program aimed at enabling citizens and educators to conduct experiments without extensive CloudLab expertise, fostering broader participation in climate science research

3 Implementation

We have structured our implementation into three distinct projects to streamline our efforts:

3.1 NSF NCAR NWP Models

Step 1: Understanding the NWP Model

- The WRF Preprocessing System (WPS) takes existing 4-D atmospheric data from GRIB-format files and interpolates it onto the user's specified WRF domain grid
- Gridpoint Statistical Interpolation (GSI) is the operational data assimilation system, which takes a forecast (aka, first guess or background field) and modifies the model state based on observations.

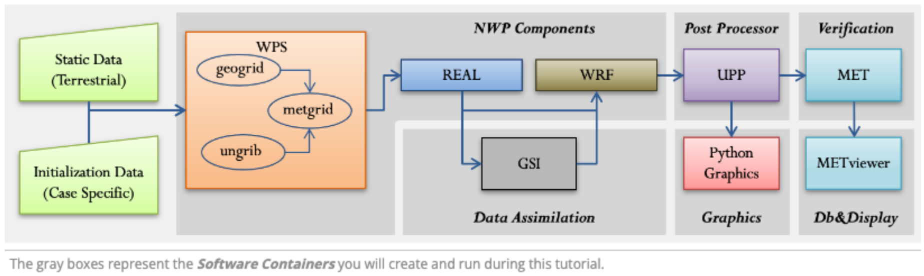

Fig. 1. Containerized Numerical Weather Prediction Model[12]

- The Weather Research and Forecasting (WRF) model is a numerical weather prediction model for weather research and forecasting.
- The Unified Post-Processor (UPP) is a post-processor for WRF that processes the raw model output into more useful forms.
- Python graphics converts the UPP output to visualizations that can be interpreted by humans.
- MET is a freely available community verification package supported by the DTC. A suite of verification tools with over 70 statistics using both point and gridded data.
- MetViewer is a Database and display analysis tool used to view the verifications produced by MET. datasets

Step 2: Understanding the data GRIB, a file format commonly used in meteorology, serves as a standardized method for storing and transporting gridded meteorological data, particularly output from Numerical Weather Prediction (NWP) models. With its compact structure and efficient encoding, GRIB files are ideal for representing vast arrays of meteorological variables, including temperature, humidity, wind speed, and atmospheric pressure, across spatial and temporal domains. This format facilitates seamless exchange and analysis of meteorological data, supporting a wide range of applications in weather forecasting, climate research, and environmental modeling. We scraped the data from the NOAA website which hosts it in the form of a directory-based structure[10].

Step 3: Getting the default settings up and running The next step is to get the architecture up and running on CloudLab. The tutorial[13] can be referred to prepare the automation scripts. We divided the work into two phases:

1. Preparation phase[9]
   - Download Prerequisite Tools and Modules: Begin by downloading the necessary tools and modules required for setting up the environment. eg. Docker, node, python, etc.
   - Setup Repositories and Variables: Configure NCAR NWP repositories and define variables to streamline the setup process.

  - Build Individual Containers: Construct individual containers discussed in the previous step tailored to specific requirements and functionalities.
2. Running a specific case[6] (eg. Sandy Hurricane[11])
  - Setup Case-Specific Repos and Variables: Customize repositories and define variables specific to the case being executed.
  - Run Docker Containers: Initiate the Docker containers configured in the previous steps to execute the case.
  - Verify Output: After running the case, verify the output to ensure that the desired results have been obtained accurately and efficiently.

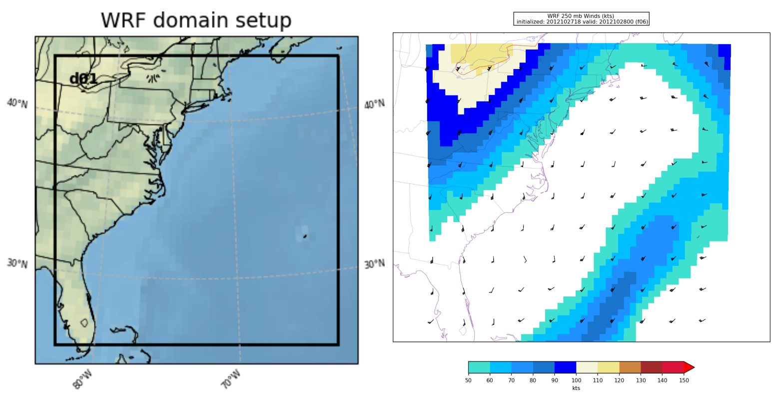

Fig. 2. Sample Output (Left: Domain on which the model ran, Right: Wind speed and direction during the 6th hour of the hurricane)

Step 4: Understanding Parameter Modifications After setting up the default tutorial environment successfully, the next step involves understanding and implementing parameter modifications to tailor the system to specific use cases. These modifications can significantly impact the outcome of the simulations and ensure that desired results are achieved effectively. Some typical modifications can include changing the domain of the forecast or the date and time of the event. The changes typically occur in several key files:

1. set_env.ksh
  - Change Versioning: Update version numbers of software components to ensure compatibility and access to the latest features.
  - Modify Container-Wise Settings: Adjust container-specific configurations to optimize performance and resource utilization.
2. namelist.input
  - Adjust Time Settings: Modify time-related parameters to control the time axis that the model should run against.
  - Modify Domain Settings: Alter domain parameters to define the spatial extent and resolution of the simulation domain.

  - Adapt Model Settings: Adjust model-specific parameters to fine-tune the behavior and accuracy of the numerical model or change the physics model itself.
3. namelist.wps
  - Modify WPS Settings: Customize settings related to the Weather Research and Forecasting (WRF) Preprocessing System (WPS) to refine data preprocessing and input generation.

An exhaustive list of modifications can be found here [7]. By understanding and implementing these parameter modifications, users can optimize the simulation setup to better align with their specific research objectives and requirements, ultimately enhancing the accuracy and relevance of the simulation results.

Step 5: Modify the parameters to obtain new results The following changes were made to the parameters to obtain new results:

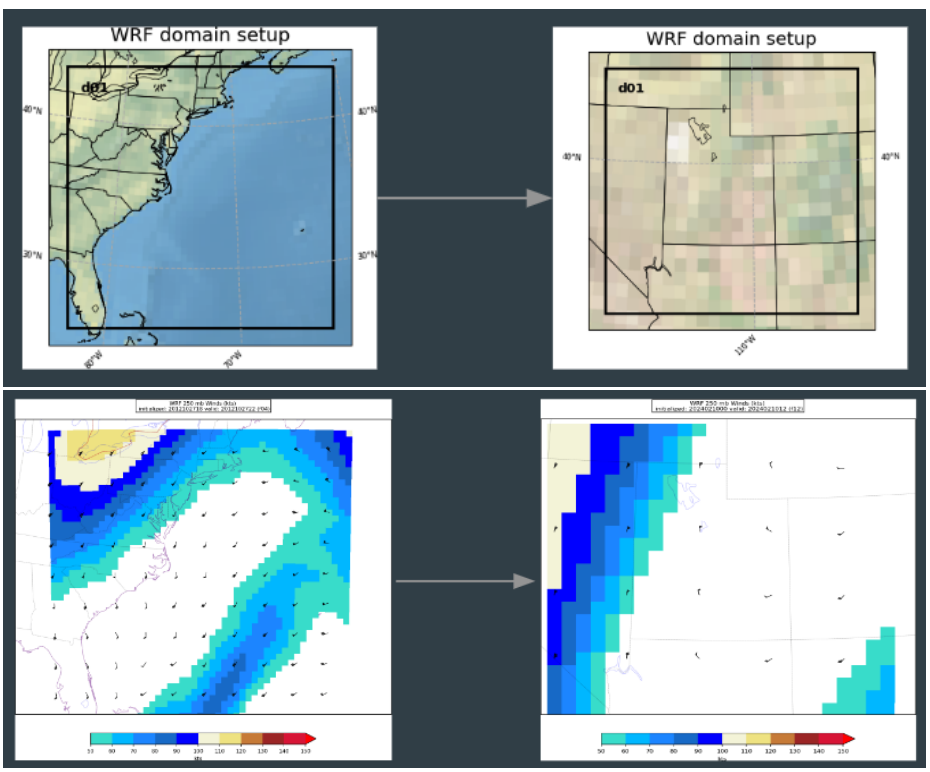

Fig. 3. Changes after Modifications. The above image shows domain modification and the image below shows time modification

Step 6 - Create an automated GitHub-based profile On CloudLab, accessing the Cloud has been streamlined through one-click access profiles, simplifying deployment for users. We've taken this a step further by automating the entire process of model initialization and modularizing it into a GitHub-based profile[3]. By integrating GitHub repositories with CloudLab profiles, users can effortlessly access the latest forecast models and deploy them with ease. Our automation code encapsulates all the preceding steps, ensuring modularity and flexibility in managing forecast models. This automated approach not only enhances user experience but also promotes collaboration and innovation in climate research by facilitating seamless access to cutting-edge tools and resources. You can find the entire github-based profile here[8].

Step 7: Create a User-Friendly Web Server to Expose Forecast Models To enhance accessibility and usability, we'll establish a user-friendly web server that exposes the forecast models. This server will empower users to interact with the models directly, enabling them to input their own latitude, longitude, date, and time to generate personalized forecasts. By providing an intuitive interface, users can easily navigate and test the forecast models according to their specific needs and interests. This initiative aims to democratize access to climate forecasting tools and foster greater engagement and collaboration within the community.

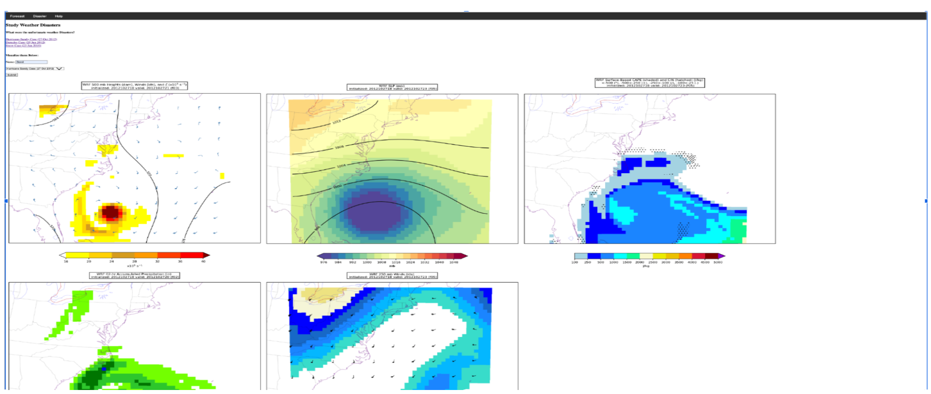

Fig. 4. WebServer written in Node.JS and hosted on CloudLab

Step 8: Apply to a Use-Case (Citizen Science Course) In my Citizen Science project, I've embarked on a journey to enhance public engagement with climate science by exploring effective strategies. At the heart of my research lies an investigation into whether individuals are more motivated to learn about climate science when they understand its tangible, day-to-day benefits, such as mitigating disruptions like flight delays. This endeavor is crucial in addressing the urgent challenges of climate change by fostering widespread understanding and action.

Through my presentation, I've shared key insights gleaned from my research journey. I've underscored the innate motivation individuals exhibit toward contributing to science, even in the absence of monetary incentives and their desire for feedback on their contributions. Additionally, I've emphasized the importance of selecting testbed environments that highlight tangible personal gains, such as the reduction of flight disruptions, to effectively engage participants.

I used the webserver created in the above steps to educate people about Climate Science Models.

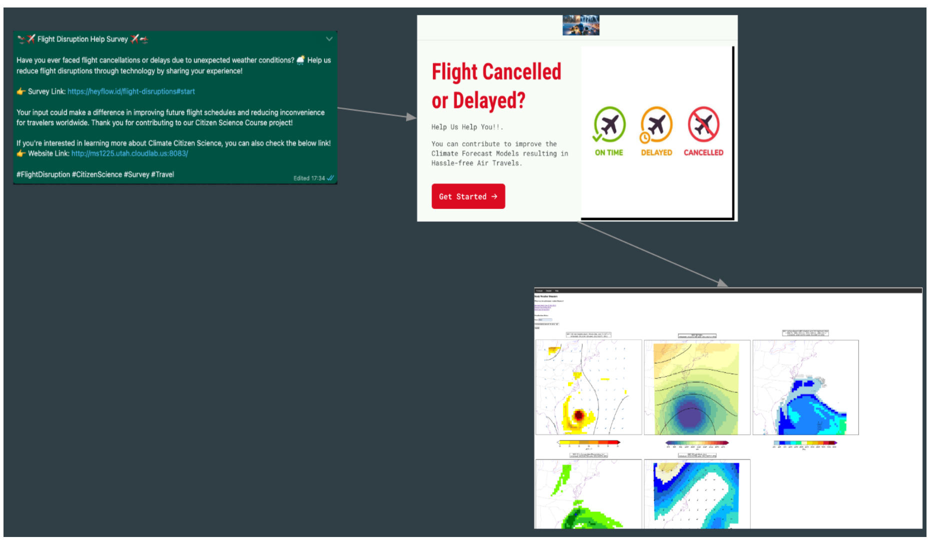

Fig. 5. People filled up the Flight Disruption Survey, Were given some cool visual feedback and were motivated to learn more about the models. We tracked website engagement metrics for the research

3.2 Data Hosting

The National Centers for Environmental Prediction (NCEP) within the National Oceanic and Atmospheric Administration (NOAA) in the United States develop and operate multiple numerical weather prediction (NWP) models. These models, including the Global Forecast System (GFS), North American Mesoscale Model (NAM), Rapid Refresh (RAP), and High-Resolution Rapid Refresh (HRRR), play vital roles in providing weather forecasts across various time scales, from short-term to long-term. They are extensively utilized by meteorologists, researchers, industries, and the general public for weather prediction and monitoring purposes. The link to these datasets[14].

CloudLab presents an ideal medium for hosting and providing easy access to these NWP model datasets. By leveraging CloudLab's storage capabilities, the datasets can be securely stored and efficiently managed, ensuring accessibility for end-users. Through CloudLab's infrastructure, users can access the datasets seamlessly, facilitating research, analysis, and the development of innovative applications in weather prediction and related fields. By utilizing CloudLab as a data hosting solution, we can streamline access to essential NWP model datasets, advancing research and applications in meteorology and related disciplines.

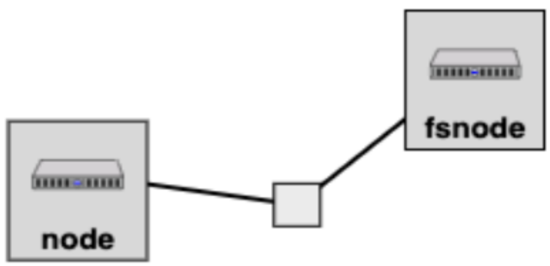

Fig. 6. Graphic of the two nodes from Cloudlab Website indicating the code and data separation.

3.3 NSF: DART-WRF (In Progress)

The Data Assimilation Research Testbed (DART) serves as an open-source, freely available community facility for ensemble data assimilation (DA). Ensemble DA combines observations with numerical models to estimate the state of a physical system, offering various applications such as generating initial conditions for forecasts, producing reanalysis of system states, and assessing the impact of specific observations on forecast skill.

CloudLab provides an ideal platform for hosting output from real data examples, such as identifying potential hurricanes from forecast visualizations, as part of the ongoing NSF DART-WRF project. By leveraging CloudLab's capabilities, we aim to educate and engage individuals in identifying potential causes of hurricanes and ways to detect them based on forecast images. Through interactive demonstrations and educational resources, CloudLab serves as a medium for promoting scientific literacy and fostering a deeper appreciation for the role of data assimilation in improving forecast accuracy and understanding weather phenomena.

As this project progresses, CloudLab will continue to play a pivotal role in facilitating access to data, enabling experimentation, and empowering individuals to explore and learn about the intricate processes involved in weather forecasting and data assimilation.

4 Conclusion

In the realm of climate science, the synergy between climate models and CloudLab will pave the way for transformative advancements. By harnessing CloudLab's computing power and storage capabilities, researchers can unlock new avenues for experimentation and innovation in climate modeling.

The seamless integration of climate models' code, data, and memory requirements with CloudLab's infrastructure can streamline the research process, enabling researchers to focus on their scientific inquiries without the burden of computational constraints. Additionally, CloudLab's profile-based mechanism can provide an intuitive and efficient means for researchers to access and deploy complex climate models, enhancing the ease of experimentation and collaboration.

Furthermore, CloudLab can play a pivotal role in facilitating data collection and access for climate research. Through its file system and dataset nodes, CloudLab offers researchers robust mechanisms for storing, managing, and sharing large volumes of climate data, ensuring accessibility and reproducibility across scientific endeavors.

In conclusion, the collaboration between climate models and CloudLab represents a powerful union, driving progress and innovation in climate science. Together, they empower researchers to push the boundaries of knowledge and contribute meaningfully to addressing the complex challenges posed by climate change. We also created a website[5] that can act as a base for NDCC.

5 Potential Work

- Creation of Base Profile with Startup Code: Developing a base profile with essential startup code pre-configured will streamline the onboarding process for users on CloudLab. By providing a ready-to-use environment, researchers and citizens can quickly dive into their projects without the hassle of setting up infrastructure, accelerating the pace of innovation and discovery.
- User Access for Citizens and Researchers: Extending access to CloudLab for potential citizens and researchers will democratize participation in scientific endeavors. By removing barriers to entry, such as technical expertise or institutional affiliations, CloudLab can empower a broader community to engage in meaningful research and contribute to scientific advancements in climate science and beyond.
- Hosting Events/Hackathons to Increase Familiarity: Organizing events and hackathons focused on CloudLab's capabilities and use cases will increase familiarity and adoption among potential users. By showcasing real-world

applications and providing hands-on experience, these events will foster a vibrant ecosystem of innovation, collaboration, and knowledge-sharing within the scientific community.

By pursuing these future initiatives, CloudLab can further solidify its role as a catalyst for scientific discovery and societal impact, empowering individuals from diverse backgrounds to contribute meaningfully to the advancement of knowledge and technology.

References

1. CloudLab, T.: The cloudlab manual: Hardware, https://docs.cloudlab.us/hardware.html
2. CloudLab, T.: Cloudlab profiles, https://www.cloudlab.us/example-profiles.php
3. Cloudlab, T.: Git-based profiles, https://github.com/emulab/my-profile
4. Duplyakin, D., Ricci, R., Maricq, A., Wong, G., Duerig, J., Eide, E., Stoller, L., Hibler, M., Johnson, D., Webb, K., Akella, A., Wang, K., Ricart, G., Landweber, L., Elliott, C., Zink, M., Cecchet, E., Kar, S., Mishra, P.: The design and operation of CloudLab. In: Proceedings of the USENIX Annual Technical Conference (ATC). pp. 1–14 (Jul 2019), https://www.flux.utah.edu/paper/duplyakin-atc19
5. Jain, S.: Citizen-cloudlab website, https://github.com/SemilJain/Citizen-CloudLab
6. Jain, S.: Hurricane sandy script, https://github.com/SemilJain/ndcc-climate-research/blob/main/sandy.sh
7. Jain, S.: Modification scripts templates, https://github.com/SemilJain/ndcc-climate-research/tree/main/scripts$_{t}emplate$
8. Jain, S.: Ndcc research profile, https://github.com/SemilJain/ndcc-climate-research/tree/main
9. Jain, S.: Nwp model preparation script, https://github.com/SemilJain/ndcc-climate-research/blob/main/preparation.sh
10. NOAA: Noaa forecast data, https://nomads.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/
11. NSF, T.: Hurricane sandy case(27-oct-2012), https://dtcenter.org/nwp-containers-online-tutorial/hurricane-sandy-case-27-oct-2012
12. NSF, T.: Nwp architecture, https://dtcenter.org/sites/default/files/community-code/nwp$_{c}ontainers/Documentation/AMS_{2}020/03_{s}ystem_{c}ase_{s}tudy_{o}verview.pdf$
13. NSF, T.: Nwp containers, https://dtcenter.org/nwp-containers-online-tutorial/software-containers
14. NSF, T.: Publicly available datasets, https://dtcenter.org/nwp-containers-online-tutorial/publicly-available-data-sets
15. Ottinger, G.: Reconstructing or Reproducing? Scientific Authority and Models of Change in Two Traditions of Citizen Science. Routledge (2017)