# Preview of Award 2209711 - Annual Project Report

## Cover

| | |
|---|---|
| Federal Agency and Organization Element to Which Report is Submitted: | 4900 |
| Federal Award or Other Identifying Number Assigned by Agency: | 2209711 |
| Project Title: | Frameworks: Large Scale Atmospheric Research Using an Integrated WRF Modeling, Visualization, and Verification Container Framework (I-WRF) |
| PD/PI Name: | Richard Knepper, Principal Investigator<br>Sue E Haupt, Co-Principal Investigator<br>Jared A Lee, Co-Principal Investigator<br>Sara C Pryor, Co-Principal Investigator |
| Recipient Organization: | Cornell University |
| Project/Grant Period: | 08/01/2022 - 07/31/2026 |
| Reporting Period: | 08/01/2022 - 07/31/2023 |
| Submitting Official (if other than PD\PI): | Richard Knepper<br>Principal Investigator |
| Submission Date: | 06/09/2023 |
| Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions) | Richard Knepper |

## Accomplishments

### * What are the major goals of the project?

The I-WRF project for the NSF Cyberinfrastructure for Sustained Scientific Innovation program will provide a framework for application containers that allow the WRF Weather Research and Forecasting software and accompanying MET and METplus validation software to be run on a wide range of resources with minimal installation requirements. I-WRF will support three science use cases that quantify impacts of environmental and human developments together with climate change on critical outcomes. I-WRF is also intended to facilitate outreach by making it easier to provide demonstrations and training and by building understanding and interest for potential atmospheric scientists.

The major goals of the I-WRF project are:

- Implement a coordinated containerized framework for the Weather Research and Forecasting Model (WRF) that seamlessly integrates a new multi-node WRF container, an optimized Model Evaluation (METplus) container, and a visualization container for more productive research.
- Support multi-node simulations enabling research-grade applications, i.e., covering large domains at high spatial discretization.
- Facilitate ease of use that enables a wider range of researchers in fields such as environmental engineering, transportation, civil engineering, air quality policy, agriculture, urban planning, etc. to run their own modeling activities, followed by convenient interaction with their results.
- Create a container framework in which users will not have to configure and deploy individual elements separately. Containers will include the entire environment and recipes required for conducting collaborative and extremely complicated workflows.
- Test and validate the integrated framework and container features on the latest parallel HPC and cloud platforms with use cases that scale studies such as the evolution of renewable energy generation in a changing climate, the effect of land use and climate change on severe weather events, and the relation between air quality and human morbidity and mortality.
- Use these containers as vehicles for introducing students to numerical atmospheric simulations and output evaluation at WRF and METplus tutorials and classroom curricula.
- Increase awareness of the I-WRF project and the framework's capabilities in order to put more usable tools for atmospheric science research in the hands of a larger number of next-generation researchers.

### * What was accomplished under these goals and objectives (you must provide information for at least one

**of the 4 categories below)?**

Major Activities:

Specific Objectives:

Beneath each objective listed below is a summary of what was accomplished during PY1 of the 4-year project.

**Build an integrated framework for a new multi-node WRF container and METplus container**

- Gathered user requirements from use case scientists and HPC/cloud systems managers.
- Ran the current WRF container on NCAR's Cheyenne HPC using a Singularity container runtime and on Cornell's Red Cloud using a Docker CE container runtime and identified/detailed all shortcomings.
- Mapped out requirements for expanding the capabilities of the container to further execution environments with intervention-free interoperability between WRF and MET/METplus.
- "Big lift" milestone of making MET/METplus able to read in native grid output from WRF is targeted for completion by the end of PY1.
- Worked on defining the base multi-node WRF and MET/METplus configuration, developing the compiler strategy (e.g., Intel OneAPI), examining security requirements, and defining attributes of the visualization container. Several visualization candidates are under investigation, such as VAPOR which reads NetCDF files to create visualization animations. We will also provide instructions for users on how to make their own visualizations and scripts for creating spatial plots. Examples will be provided.
- Once the multi-node framework is completed, we plan to add features to handle large-scale runs that require different types of inputs, e.g., high-resolution terrain data, aerosols, clouds, radiation, etc., and data such as ERA5 that provide hourly estimates of a large number of atmospheric, land and oceanic climate variables covering 1940 to present and is somewhat representative of what a lot of weather and climate model data sets would look like.
- Execution scripts will be prepared for and tested on supercomputers at the Texas Advanced Computing Center and NCAR-Wyoming Supercomputing Center, and on Red Cloud and the Jetstream2 Regional Cloud at Cornell. We requested hundreds of terabytes of storage for our Cheyenne and future Derecho allocations. I-WRF containers are also being architected to run on laptops to facilitate tutorial training and to help overcome training obstacles such as students failing to compile software during training sessions.
- The I-WRF User's Guide and I-WRF Discussion Forum are being established on NCAR's GitHub to facilitate instruction and provide a collaborative communication forum for the community.
- All model code will be open source and project outcomes will be distributed without limits.

**Support multi-node simulations enabling research-grade applications, i.e., covering large domains at high spatial discretization**

- Sara C. Pryor in collaboration with Melissa Bukovsky prepared and ran a number of test cases (storylines) of historically important events and re-simulated them under changing Land Use/Land Cover (LULC). These storylines of deep convection over the Southern Great Plains and their response to changing LULC have been performed on a range of different HPC (e.g., Cheyenne and Perlmutter) to develop timing statistics that can be benchmarked relative to the container and also to evolve understanding of the atmospheric response to enhanced levels of urbanization. This action is linked to a flagship project under the auspices of the World Climate Research Programme that is assessing the use of regional models in a storyline framework for understanding climate hazards. Other science use cases will further inform the I-WRF architecture design by identifying complexities within workflows in order to reduce them.
- Dr. Sue Ellen Haupt, Senior Scientist at the NCAR Research Applications Laboratory is now co-PI on the project overseeing all science use cases. Dr. Jared Lee, a project use case scientist focused on modeling for renewable energy (WRF-Solar-Wind) and atmospheric chemistry (WRF-Chem), has also assumed a co-PI leadership role and will be coordinating the operations of the NCAR technical team.
- In addition, a postdoctoral researcher joined the I-WRF team at Cornell to support use case development, compiling WRF on various platforms, and conducting and evaluating numerical downscaling simulations using WRF. Containerization engineer Ben Trumbore and cloud systems engineer Bennett Wineholt also joined the project. Wineholt worked previously on a multi-node WRF container test project using Google Cloud.
- Creating a container framework that makes using WRF easier for less advanced users while at the same time adding the flexibility and features desired by more advanced users is a balancing act.

**Facilitate ease of use to enable a wider range of researchers to run their own modeling activities**

- We ran the current-generation Numerical Weather Predictions container by closely following the NWP instructional guide. The goal was to identify usability issues that would be especially frustrating to new users and apply them to the development of the new I-WRF container. The current container was judged to be not much easier than compiling and installing WRF normally. Issues identified included incomplete documentation for Singularity use cases, difficulty duplicating files across different folders, hard-to-find configurations, inconsistent bash and tcsh instructions, difficulty finding namelists, no directions on where to modify I/O in namelists, and no job submission instructions or scripts. These and other usability issues will be addressed and help meet our goal of making I-WRF software installation easier to initiate, use, and maintain. Extensibility will be achieved by varying execution scripts that provide alternative execution scales, data sets, and namespaces.
- Instructional data sets and demo scripts will be included in the I-WRF containers to further facilitate ease of use.
- After writing scripts to run on Cheyenne (and Derecho, when it becomes available), we will decide whether to convert them into a GUI.
- The containers will also be integrated into NCAR's WRF and MET tutorials to introduce students to numerical atmospheric simulations and output analysis.

**Test and validate the framework on the latest parallel HPC/cloud platforms with scalable use cases**

- A WRF simulation test case was run on NCAR's Cheyenne supercomputer and Cornell's Red Cloud OpenStack private cloud platform to establish basic simulation function with containerized code and storage mounting locations.
- Scalable compute, networking, and file storage systems for private and public cloud platform deployment have been identified, with initial development work focused on deploying Kubernetes in Red Cloud backed by the Ceph File System (CephFS) distributed SSD file storage.

**Increase I-WRF project awareness and put more usable tools in the hands of more researchers**

- Launched the I-WRF portal at https://www.i-wrf.org.
- Publicized the I-WRF project in media outlets including *HPCwire* (https://www.hpcwire.com/off-the-wire/nsf-funds-cornell-and-ncar-to-implement-an-integrated-framework-for-atmospheric-and-climate-modeling/), *Energywire* (https://i-wrf.org/about/zero-carbon-grid.pdf), and the *Cornell Chronicle* (https://news.cornell.edu/stories/2022/09/integrated-framework-atmospheric-and-climate-modeling) and increased project awareness via social media (e.g., 256 likes/reposts on LinkedIn).
- Provided an I-WRF project summary to the project director for NSF briefings. A Dear Colleague Letter updating the community on meteorology modeling tools referenced I-WRF.
- Published "The I-WRF Framework: Containerized Weather Modeling, Validation, and Verification" in the *Proceedings of the Practice and Experience in Advanced Research Computing*. This publication will be presented at the PEARC23 Conference in July 2023.

Significant Results:

Key outcomes or Other achievements:

*** What opportunities for training and professional development has the project provided?**

**Cross-Training & Knowledge Sharing**

Expertise was shared between NCAR and Cornell and use case scientists during our I-WRF PI meetings and between SMEs and sub team engineers during our I-WRF Tech Team meetings. PI and Tech Team meetings occur monthly to ensure timely cross-training and knowledge sharing. In-depth follow-up discussions occurred between meetings to solve specific technology implementation issues and to share lessons learned.
Discussions occurred with Wim Munters who collaborates with Sara Porchetta and Jeroen van Beeck at the von Karman Institute for Fluid Dynamics in Belgium on the impact of ocean waves on offshore wind farm power production. They are planning to perform multi-year WRF simulations on in-house and national HPC systems, deploy their simulations on cloud infrastructure, and utilized containerize WRF tools in their education programs. Since their objectives align closely with those of the I-WRF project, they have expressed interest in becoming an I-WRF pilot user. Their renewable energy research (https://www.sciencedirect.com/science/article/abs/pii/S0960148121012799?via%3Dihub) would be a welcomed addition to our use cases.

**How-To Documentation & Training**

I-WRF containers will be made publicly available through the project website, the NCAR WRF GitHub site, and the Docker Hub Registry. The project team will also work to make the containers and associated execution scripts discoverable and

available through resources such as the ACCESS Research Software Portal (https://software.xsede.org).
The NCAR team will provide support on WRF compilation and configuration for the WRF, MET/METplus, and visualization containers, as well as post-processing scripts that allow for data to pass between the containers without modification. The Cornell CAC team will create execution scripts and orchestration that allow the containers to run on the targeted execution environments, as well as testing and integration suites that will prepare the application containers themselves for how-to training tutorials and research-grade applications.

**I-WRF Portal**

The I-WRF portal was launched and will provide scientists and the cyberinfrastructure community extensive information on the project, including use case accomplishments/plans/products, publications, news and events, technical documentation, and project reports (https://i-wrf.org/).

### * Have the results been disseminated to communities of interest? If so, please provide details.

**Scientific Meetings, Publications, and Conferences**

I-WRF PIs, use case scientists, and sub team engineers are actively engaged in professional societies and conferences such as the International Conference on High Performance Computing, Networking, Storage, and Analysis (SC), Practice & Experience in Advanced Research Computing (PEARC), Coalition for Academic Scientific Computing (CASC), Campus Research Computing Consortium (CaRCC), American Meteorological Society (AMS), and American Geophysical Union (AGU), all of which will afford opportunities to disseminate results to communities of interest.
Outreach during our launch year included:

- SC22 Conference exhibit featuring the I-WRF project, Dallas TX.
- Outreach to members of the Campus Research Computing Consortium (CaRCC) Emerging Centers track which I-WRF PI Knepper co-founded.
- Presented science use case #1 “Urban Impacts on Deep Convection in the Southern Great Plains” by Xin Zhou, Fred Letson, Paola Crippa & Sara C. Pryor at the *American Geophysical Union Fall Meeting*, Chicago IL, 13 December 2022.
- Publishing/presenting “The I-WRF Framework: Containerized Weather Modeling, Validation, and Verification” by Richard Knepper, Sara C. Pryor, Bennett Wineholt, Melissa Bukovsky & Jared Lee at the *Practice & Experience in Advanced Research Computing Conference (PEARC23)*, Portland, OR.

### * What do you plan to do during the next reporting period to accomplish the goals?

**Build an integrated multi-node WRF and METplus framework**

- Complete container generation and WRF and METplus compilation.
- Build the visualization container.
- Modify the containerized systems to improve alignment with HPC center best practices suggestions for container security.
- Reduce container footprints by removing extraneous software elements used in development.
- Determine where to store Docker images.
- Test passing native WRF output to MET tools using containers.
- Resolve issues with permissions for files outside of containers.
- Enhance the WRF Dockerfile to allow users to supply their own files to build.

**Support multi-node simulations**

- Add features to handle large-scale runs that require different types of inputs.
- Complete preparation/testing of execution scripts for TACC supercomputer(s), NCAR’s Cheyenne and Derecho supercomputers, and Cornell’s Red Cloud and Jetstream2 Regional Cloud.
- Run/rerun case studies, starting with the Land Use/Land Cover (LULC) use case whose team plans to: (1) identify historically important weather event(s) in the Northeast U.S. (e.g., winter storm that caused major economic losses, transportation disruption, loss of life); (2) use WRF to simulate this event at 1–2 km resolution using ERA5 as initial and lateral boundary conditions (ICs/LBCs); (3) assess the fidelity of the simulation and adjust the WRF configuration if necessary; (4) once fidelity is demonstrated, perturb the simulation/event so that the ICs/LBCs represent future climate change (pseudo-global warming approach) and/or future projections of LULC and society; and (5) compare simulations.
- Spatial plots of simulation differences and biases for multiple variables and dynamic visualizations (VAPOR or VAPOR-like) are desired. METviewer already performs a number of statistics; we will determine what to add, if anything.
- Compare simulations run on the native HPC systems and also from within the container in the same system to identify opportunities for container optimization.

**Facilitate ease of use to enable a wider range of researchers**

- Populate the I-WRF User Guide with documentation and lessons learned.
- Correct weaknesses in current documentation (i.e., incomplete Singularity documentation, inconsistent bash and tcsh instructions, etc.).
- Make scripts easy to use and/or convert them into a GUI by leveraging Jupyter Notebook Python execution, data processing, and visualization tools.
- Determine how to integrate I-WRF containers into NCAR’s WRF and MET tutorials.

**Test and validate the framework**

- Integrate GitHub code version control with GitHub Actions automatic test system pipeline that will be used to regularly assess whether newly created containers are running correctly.

**Increase project awareness and put more usable tools in the hands of more researchers**

- Convene an Advisory Board representing multiple disciplines in the first quarter of PY2 to discuss accomplishments to date and solicit opinions on our plans going forward and how to increase project awareness and impact.
- Respond to calls for participation from American Meteorological Society, American Geophysical Union, and other relevant community conferences.
- Make ongoing updates to the I-WRF User's Guide and I-WRF portal.

---

## Products

### Books

### Book Chapters

### Inventions

### Journals or Juried Conference Papers

**View all journal publications currently available in the NSF Public Access Repository for this award.**

The results in the NSF Public Access Repository will include a comprehensive listing of all journal publications recorded to date that are associated with this award.

Richard Knepper, Sara C. Pryor, Bennett Wineholt, Melissa Bukovsky, and Jared Lee. 2023. The I-WRF Framework: Containerized Weather Modeling, Validation, and Verification. In Practice and Experience in Advanced Research Computing 2023. ACM, New York, NY, USA,. Status = AWAITING_PUBLICATION.

### Licenses

### Other Conference Presentations / Papers

Xin Zhou, Fred Letson, Paola Crippa & Sara C. Pryor (2022). *Urban Impacts on Deep Convection in the Southern Great Plains*. American Geophysical Union Fall Meeting 2022. Chicago, IL, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

### Other Products

### Other Publications

### Patent Applications

### Technologies or Techniques

The I-WRF project team developed the following technique:

- MET/METplus integration with WRF – created the ability for MET/METplus to read native output from WRF.

### Thesis/Dissertations

### Websites or Other Internet Sites

*I-WRF*
https://i-wrf.org

The I-WRF portal features the I-WRF team, project summaries, publications, news/events, and will provide use case accomplishments/plans/products and links to the I-WRF User Guide and I-WRF Discussion Forum which will be new branches on the NCAR website. NSF reports will also be available at the portal.

---

## Participants/Organizations

### What individuals have worked on the project?

| Name | Most Senior Project Role | Nearest Person Month Worked |
|---|---|---|

| Name | Most Senior Project Role | Nearest Person Month Worked |
|---|---|---|
| Knepper, Richard | PD/PI | 1 |
| Haupt, Sue | Co PD/PI | 1 |
| Lee, Jared | Co PD/PI | 2 |
| Pryor, Sara | Co PD/PI | 1 |

**Full details of individuals who have worked on the project:**

**Richard Knepper**
**Email:** rich.knepper@cornell.edu
**Most Senior Project Role:** PD/PI
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Programmatic oversight of the I-WRF project ensuring deliverables outlined in the proposal are met on schedule.

**Funding Support:** No funding support from other projects used for this award.

**Change in active other support:** No

**International Collaboration:** No
**International Travel:** No

**Sue E Haupt**
**Email:** haupt@ucar.edu
**Most Senior Project Role:** Co PD/PI
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Oversees all science use cases and is co-developer of the Renewable Energy use case.

**Funding Support:** No funding support from other projects used for this award.

**Change in active other support:** No

**International Collaboration:** No
**International Travel:** No

**Jared A Lee**
**Email:** jaredlee@ucar.edu
**Most Senior Project Role:** Co PD/PI
**Nearest Person Month Worked:** 2

**Contribution to the Project:** Coordinates the operations of the NCAR technical team and is the general project manager for WRF development.

**Funding Support:** No funding support from other projects used for this award.

**Change in active other support:** No

**International Collaboration:** No
**International Travel:** No

---

**Sara C Pryor**
**Email:** sp2279@cornell.edu
**Most Senior Project Role:** Co PD/PI
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Scientific guidance, management of a post-doctoral researchers, and interface with the WRF user community among researchers to provide information about the I-WRF Framework for adopters.

**Funding Support:** No funding support from other projects used for this award.

**Change in active other support:** No

**International Collaboration:** No
**International Travel:** No

---

**What other organizations have been involved as partners?**
Nothing to report.

**Were other collaborators or contacts involved? If so, please provide details.**
Nothing to report

---

## Impacts

**What is the impact on the development of the principal discipline(s) of the project?**

The project team is developing, testing, and deploying an integrated multi-container framework for the WRF model to enable for the first-time *multi-node* containerized simulations complete with verification and visualization capabilities, thus “lowering the bar” for atmospheric researchers who wish to more easily and effectively use WRF in parallel on multiple platforms, ranging from desktops to clouds and supercomputers.

The UN Climate Change Conference and the NSF Engineering Visioning Alliance on *The Role of Engineering in Addressing Climate Change* clearly signal the urgent need for more and better simulations of possible future climate states. This in turn requires greater participation in research-grade simulations of possible future climate conditions to enhance climate resilience through the U.S. economy. Our work directly addresses some of these research needs through our three science case studies but more critically will enable more researchers to engage with undertaking the wide range of simulations needed to provide actional information to stakeholders, to explore uncertainty, and to fundamentally advance climate science.

**What is the impact on other disciplines?**

I-WRF’s coordinated capabilities and ease of use will enable a wider range of researchers in fields such as environmental engineering, transportation, civil engineering, air quality policy, agriculture, urban planning, etc., to run their own modeling activities, followed by convenient interaction with their results. The containerized system will broaden participation and support early career scientists in multiple disciplines by reducing the activation energy required to undertake numerical simulation and evaluation. These same systems will enable new users to independently investigate the WRF system due to the removal of Linux mastery as a necessary step. Importantly, participation by distributed domain experts will be encouraged as reproducible results from identical software environments will reduce “architecture climate.” The containerized systems will promote science and engineering excellence across disciplines and enable a wider diversity of people to engage in simulation with the WRF model and the METplus verification system.

### What is the impact on the development of human resources?

The team will engage in a program to develop students and early career researchers through direct interaction with the technologies involved as well as the research problems, to support the development of new scholars and technologists through programs provided at NCAR and partnership institutions, as well as through outreach and activities with professional networks.

We will also impact the development of human resources by pursuing hosting short courses at the American Meteorological Society (AMS) Annual Meetings as we have in the past. The resulting activities will put usable tools for atmospheric science in the hands of a larger number of next-generation researchers and incorporate the best practices for compilation, configuration, and deployment for cyberinfrastructure professionals.

Three Post-Doctoral Fellows will be supported by the project. Cornell Masters of Engineering students and undergraduate students will be encouraged to participate as well.

### What was the impact on teaching and educational experiences?

NCAR tutorials will leverage I-WRF containers to more effectively train a broader array of graduate students and early career researchers in the atmospheric sciences, thus expanding the S&E workforce. The team also plans to leverage its past experiences and connect directly with institutions such as the Metropolitan State University of Denver (MSU Denver), Central Michigan University (CMU) and other institutions to encourage the integration of the I-WRF containerized modeling system into course curricula. To further expand educational opportunities, we will create and maintain an on-demand I-WRF Virtual Workshop to democratize access to learning for less resource rich institutions.

### What is the impact on physical resources that form infrastructure?

The I-WRF project is focused on developing a greater degree of platform portability for running WRF on all types of cyberinfrastructures: desktops, clouds, and HPC resources. Student use of pre-compiled and tested Docker images can save time on desktops and small cloud installations. Researcher productivity using WRF on HPC or cloud physical resources will be improved through per cluster and per cloud provider example implementations to launch WRF jobs. We plan to accomplish this by using Python as the common top-level invocation layer and programming scheduling script generation using the Signac framework to generate submission scripts for Slurm or PBS scheduler on HPC resources, using Terraform Python CDKTF resource provisioning scripts for cloud resources, and the Docker Engine Python interface docker-py for local desktop execution.

Our primary platform targets are HPC resources which feature CPU hardware and support running Singularity containers, as well as cloud systems and desktops running CPU hardware which can run the Docker container runtime engine. This is due to current supported platforms for WRF MPI Fortran code compilation being limited to CPU architectures, and container runtime overheads being either more performant due to optimized library availability during real implementation comparison or within a few percent of total runtime accounting for system interactions including I/O. Our ability to support WRF on target physical infrastructure with specialized hardware and/or limited container capabilities will be more limited.

### What is the impact on institutional resources that form infrastructure?

By improving the portability of WRF containers, institutions will have more choices on where they run WRF simulations. Improved training should mean less institutional resources are needed to support researchers using WRF. More researchers will be able to get started with multi-node WRF without institutional staff resources. The flexibility afforded by the new containerized workflow means that researchers will be able to access the broadest set of CI resources, including HPC and cloud elements, with minimal demands on cyberinfrastructure providers.

### What is the impact on information resources that form infrastructure?

I-WRF use case scripts, build files, etc. will be shared with sample data and storage configurations.

### What is the impact on technology transfer?

Technologies generated by this project are open source, therefore technology transfer licensing, patent applications, etc. are not applicable.

**What is the impact on society beyond science and technology?**

The I-WRF project will support critically important inquiry that impacts the day-to-day life of U.S. citizens, agriculture, power generation, and public health, by making it easy to run large-scale simulations, and by making it simpler for a larger range of students to use and understand computational approaches to atmospheric research. Use cases will focus on how climate variability and weather extremes impact society, how climate change impacts wind and solar resources, and how changes in atmospheric dynamics and emissions linked to global climate change may reverse recent trends toward improved air quality and declining human mortality and morbidity.

**What percentage of the award's budget was spent in a foreign country?**
Nothing to report.

---

## Changes/Problems

**Changes in approach and reason for change**
Nothing to report.

**Actual or Anticipated problems or delays and actions or plans to resolve them**
Nothing to report.

**Changes that have a significant impact on expenditures**
Nothing to report.

**Significant changes in use or care of human subjects**
Nothing to report.

**Significant changes in use or care of vertebrate animals**
Nothing to report.

**Significant changes in use or care of biohazards**
Nothing to report.

**Change in primary performance site location**
Nothing to report.