

ALL TOTAL TECHNICAL PROJECTS

Undergraduate

- Project 1. Building the Python Equivalent of the NCL Visualization Gallery
- ~~Project 2. Code Profiling and Optimization for MURaM Project~~ Closed 11/14
- Project 3. Cross Reference Monitoring of Supercomputers and Support Infrastructure **intern will be in Cheyenne, WY
- Project 4. Develop Fortran-aware Backend for Rose
- Project 5. Developing a Marker Based Augmented Reality (AR) System for an Earth Science Education and Outreach Application
- Project 6. Real-time Weather Museum Touchscreen
- Project 7. Using Machine Learning to Improve Preprocessing of Satellite Observations for Data Assimilation

Graduate

- Project 8. Applying Advanced Search Techniques to Enable Scientific Data Discovery and Exploration
- Project 9. Analysis of Weather and Climate Model Data on GPUs
- Project 10. Automatic Load Balancing for an Earth System Model
- Project 11. Estimating the Ocean Boundary Layer Depth with Machine Learning
- Project 12. Evaluating Two Approaches to Automated Code Refactoring
- Project 13. Open Source Technologies Supporting Data Science and Decision Analysis with Internet-of-Things 3D Printed Weather (IoTwx/3d) Stations
- Project 14. Using Neural Networks for Scientific Data Compression

NON-TECHNICAL PROJECT

- CISL Outreach, Diversity, and Education *G

1. Building the Python Equivalent of the NCL Visualization Gallery

Areas of Interest: Visualization, Software Engineering

For nearly two decades researchers in the atmospheric, oceanic, and related sciences have employed the NCAR Command Language (NCL) to analyze and plot their data. With the emergence of Python as the scripting language of choice for scientific workflows, NCAR has set a course to migrate many of NCL's highly specialized capabilities into the Python ecosystem. NCL's plotting functions are of particular interest. The NCL site (ncl.ucar.edu) hosts a gallery of visualization examples for different kinds of plots. The lack of an equivalent resource in python is one of the most common concerns expressed by NCL users transitioning to python. Currently, scientists have to combine information from the matplotlib and cartopy module documentation to build the publication-quality figures they desire.

Over this summer internship, the student will explore and learn about data visualization in the atmospheric and oceanic sciences using matplotlib, cartopy, and holoviews. The student will generate plotting templates inspired by the NCL gallery including: box plots, contour plots, (r,theta) radar plots, vector plots, and 3-dimensional plots. The student will have to plot using different map projections and with different overlays of satellite, measured, and modeled data. Over the course of the internship, we will see the library of python plotting examples grow.

Students: The project is open to undergraduate students. There are 2 open positions.

Skills and Qualifications: Experience with Python programming. Familiarity with Jupyter Notebooks. User level familiarity with Linux and Unix-based tools for scripting and file manipulation. Ability and willingness to work with a team. Good communication and writing skills

Optional: Familiarity with NCL (NCAR Command Language). Experience with NumPy, Matplotlib, Cartopy, holoviews, and Xarray

~~2. Code Profiling and Optimization for MURaM Project~~

Project closed as of 11/14/19

3. Cross Reference Monitoring of Supercomputers and Support Infrastructure

Areas of Interest: Supercomputer Systems Operations, Data Science, Building automation systems and controls optimization

As supercomputers have evolved, the monitoring capabilities and support systems have increased in complexity as well as the volume of metric data that is produced and collected. Often systems have distinct frameworks for monitoring various individual components and subcomponents of the systems. Developing an aggregation framework and connecting it to a monitoring framework that cross references information collected from building automation systems, building and supercomputer cooling systems (thermal sensors, fan sensors, etc.), electrical systems, batch system software and file systems will aid in more robust operations and management of complex computing facilities such as the ones at NCAR.

NCAR currently operates and maintains a data center in Cheyenne, WY known as the NWSC that houses the current flagship supercomputer, Cheyenne, a data analysis and visualization cluster, Casper, and several large scale file systems, GLADE, Campaign Store. A key to continued successful operation of such systems is to correlate events that impact availability of access or performance of the systems. The aggregation of events can lead to quicker diagnostics and responses from support and administration teams as well as understand transient behaviors of systems.

The student project includes leveraging tools and site expertise to fabricate a portal to view and analyze data of the various monitoring systems and making it actionable to various parties at NCAR and allowing a whole systems reporting framework. The student will have opportunities to gain comprehensive knowledge of building and data center operational metric monitoring and supercomputing metrics while working through various analysis in data science using prominent open-source toolkits.

Students: This project is open to undergraduate students.

Skills and Qualifications: A student who is proficient in the fields of mechatronics with familiarity with a data science programming capability in the realm of time series data. The student should have a basic understanding in areas of thermodynamics fluid dynamics. Additionally, the student should have applicable knowledge of statistical analysis.

**intern will be in Cheyenne, WY

4. Develop Fortran-aware Backend for Rose

Areas of Interest: Software engineering, Supercomputer Systems Operations, Application optimization / parallelization

The Rose Compiler (rosecompiler.org) is a useful tool for transforming program sources.

However, the backend of the Rose Compiler does not support Fortran very well, and major applications of this technology for NCAR involves Fortran.

The SIParCS Intern will modify the Rose-supplied Fortran backend so it preserves a user's source formatting options. This involves learning enough about Rose to make the modifications needed.

There are several tools created at NCAR that can use a working Fortran backend, so there are tests immediately available for this work.

Students: This project is open to undergraduate students.

Skills and Qualifications: Basic computer science skills, compiler skills are helpful

5. Developing a Marker Based Augmented Reality (AR) System for an Earth Science Education and Outreach Application

Areas of Interest: Visualization, Software Engineering, Augmented Reality

NCAR has been developing Augmented and Virtual Reality (AR/VR) applications to help make NCAR science more engaging and accessible to a wider audience. Inspiring, educating, and informing the public about NCAR research and about the wonder and relevance of science is a primary mission of our organization. Implementing new technologies that enhance our storytelling capabilities and engage our audiences plays a key role in making that possible.

The student will work on developing an AR Marker Detection and Tracking routine to estimate the position and orientation, in 3D space, of an object in a video stream for the Meteo AR application, an app for exploring geoscience data from a mobile device. This project will offer the intern valuable experience developing custom cross-platform OpenCV plugins for game engines such as Unity, and extending computer vision and machine learning capabilities to mobile applications/games.

Students: This project is open to undergraduate students.

Skills and Qualifications: Excellent C++ and C# programming skills. Knowledge of Computer Vision and Image processing techniques. Experience with OpenCV library. Familiarity with Camera Calibration. Experience with mobile app development in Unity. Intermediate to advanced computer skills. Excellent written and verbal communication skills.

6. Real-time Weather Museum Touchscreen

Areas of Interest: Data Science, Visualization, Meteorology

Staff from the UCAR Center for Science Education, Unidata, and CISL are interested in developing a new touchscreen interface that displays real time weather data. The student working on this project will work with existing data sources and programming tools to develop a new user interface to visualize maps and graphs of real-time weather. This interface will be included in the exhibits at the NCAR Mesa Lab in Boulder, CO, and at the NWSC Visitor Center in Cheyenne, WY. This display will allow visitors to these facilities to access current weather conditions at locations all over Earth, and it will include background content about weather data and features seen on weather maps.

Students: This project is open to undergraduate students.

Skills and Qualifications: User interface design for desktop or web applications;
Programming experience: Experience working with data APIs, programming experience with Java, Python or similar languages

7. Using Machine Learning to Improve Preprocessing of Satellite Observations for Data Assimilation

Areas of Interest: Data Science, Geostatistics, Numerical Methods

The Data Assimilation Research Testbed (DART) is a mature and widely used community software facility for data assimilation. One application of data assimilation is improving numerical weather prediction (NWP) forecasts. An atmospheric model prediction is run on a supercomputer with hundreds or thousands of nodes, and the output from this model is then statistically combined with atmospheric measurements such as temperature or winds. The measurements may also be from much more sophisticated instruments like radars or satellite radiometers. The process of combining the model forecast and observations is known as data assimilation.

95% of the measurements of the atmosphere potentially available for data assimilation come from meteorological satellites orbiting the earth looking down. However, these observations remain underutilized due to a variety of issues such as uncertainty regarding water and ice clouds. According to some estimates, more than 70% of these satellite observations are currently discarded by NWP forecasting centers around the world due to “cloud contamination.” Yet clouds are precisely where the severe weather is occurring. This project aims to help address these issues by combining sophisticated machine learning techniques with data assimilation.

This project will use machine learning to find and exploit relationships between atmospheric fields like water vapor, temperature, clouds, etc. and satellite observations. These extracted relationships will then be input into data assimilation systems in order to improve NWP forecasts, where improved forecasts could save life, limb, and property. These relationships also have the potential to help benefit climate studies and improve model parameterizations.

As there are many more satellite observations available than are currently used, and because the models involved work with billions of variables, this is a “big data” problem that is highly relevant to today’s world. This internship is a chance to apply proven machine learning packages and techniques on an important unsolved problem with substantial real-world consequences.

Students: This project is open to undergraduate students.

Skills and Qualifications: Programming (Python and/or Matlab), machine learning, data analysis

8. Applying Advanced Search Techniques to Enable Scientific Data Discovery and Exploration

Areas of Interest: Software Engineering, Digital Asset Management, Scientific Data Discovery

Scientific data search and discovery continues to be a challenge for people interested in using climate and weather data, from expert domain scientists to citizen scientists and the science-interested public. A common roadblock is the difficulty in understanding the language and vocabulary of specific earth science domains. Despite the advent of new services for data discovery such as the Google Dataset Search and other scientific data search platforms, scientific data discovery remains difficult for a significant portion of our users.

This project will explore the use of current search platforms, such as Apache Solr and Elasticsearch, to make search easier to use by our users. We will enhance our current search capabilities and User Interface and explore areas such as query suggestions, spelling, search term autocomplete, suggested hints, synonyms, and other semantic enhancements. The project will include measuring our work with Usability Testing and the project will be highly collaborative and team based, using the Agile Scrum methodology. This work will be an opportunity to add new end user features to an existing production scientific data repository, work with a software engineering team in an Agile model and learn software engineering practices.

Students: The project is open to graduate students.

Skills and Qualifications: Basic understanding of programming. With basic experience in languages such as Java and Javascript. Basic understanding of controlled vocabularies and metadata schemas. Basic understanding of XML and HTML markup languages. Basic understanding of web services, dynamic and web UI. Basic understanding of query languages like SQL. Ability to interact with mentors and peers in a manner that supports collaboration and inquiry. Ability to work with diverse staff. Good problem solving skills. Good oral and written communication skills. Willingness to learn and use computing tools and programs. Curiosity to explore new things.

9. Analysis of Weather and Climate Model Data on GPUs

Areas of Interest: Data Science, Visualization, Application Optimization/Parallelization

With recent software developments, it's possible to perform end-to-end data simulation and analysis entirely on GPUs. This technique reduces the need for data storage and unnecessary input/output (I/O), which remain challenging bottlenecks for high performance computing. The project will explore techniques for data processing on GPUs, including topics such as feature detection for extreme weather events and In situ analysis techniques. Application Programming Interfaces (APIs) such as RAPIDS along with other other In situ processing libraries will be used.

The goal of this 2020 summer internship will focus on developing the post processing code with the help of analysis tools for a certain weather and climate model. The student's primary focus will be on using RAPIDS, (or a similar library), that supports classic analysis techniques as well as machine learning approaches to analyze large datasets on GPU memory. Additionally, the student will explore methods for doing In situ processing on GPUs.

Students: The project is open to graduate students.

Skills and Qualifications: Strong programming skills with at least one of the following languages - C, C++, and/or Fortran - is required. Familiarity with Python scripting. An understanding of parallel programming, machine learning concepts, and knowledge of GPGPU is preferred. Familiarity with In situ processing libraries is considered a bonus.

**If applying to Project 9 "Analysis of Weather and Climate Model Data on GPUs", the online application will list it as "Auto Mapping of OpenACC to OpenMP-GPU Directives and Vice versa". We are aware of this name discrepancy but please choose this option for Project 9.

10. Automatic Load Balancing for an Earth System Model

Areas of Interest: Application Optimization/Parallelization, Software Engineering

Modern earth system models are incredibly complex applications containing individual 'components' like the atmosphere, ocean, ice or land models that work together to accurately simulate the entire earth system. Each of these components model different physical processes, and some calculations are more expensive than others. This makes balancing the work to be done across many processors difficult, and thus has typically been the domain of a few experts with first-hand knowledge about the systems and the models. A new approach is needed that does not rely on a limited number of individuals who are experts in the load balancing of these models. Instead we intend to develop a dynamic load balancing capability that will support both traditional high-performance computing environments as well as emerging 'cloud' environments.

This project will seek to prototype a dynamic load balancing capability by performing short runs of component models, and from their performance, calculate an improved configuration. The end goal of this project will be to develop a robust approach to automatic load balancing that will result in cost savings for the broader user community. Students will develop experience with a leading earth system model, cloud & high-performance computing (HPC) systems, load-balancing, and scientific programming.

Students: The project is open to graduate students.

Skills and Qualifications: Programming experience (Fortran, C/C++ or Java are ideal). Linux shell experience. Good problem-solving skills.

*****NOTE*****

This project has the possibility of an extension for a second summer. Details TBA.

11. Estimating the Ocean Boundary Layer Depth with Machine Learning

Areas of Interest: Data Science, Machine Learning, Software Engineering

The ocean boundary layer is a crucial mediator of Earth system dynamics. However, the ocean boundary layer depth is poorly simulated in global Earth system models. These errors are an important source of uncertainty in Earth system projections from subseasonal to centennial timescales. Although we know the models are biased, we cannot currently observe the ocean boundary layer on the short time and small space scales needed to understand all of the processes controlling its depth and dynamics. In this project, we would like to develop and evaluate a machine learning model of the ocean boundary layer depth with inputs from surface satellite observations and sparse in-situ vertical Argo profiles. The intern will be responsible for pre-processing available satellite and Argo boundary layer data and building several machine learning models and evaluating them against existing interpolated products.

Students: The project is open to graduate students.

Skills and Qualifications: Experience with Python scientific computing libraries is required. A background in geosciences or oceanography and experience with a machine learning library, e.g. TensorFlow, PyTorch, scikit-learn, are preferred but not required.

12. Evaluating Two Approaches to Automated Code Refactoring

Areas of Interest: Compiler construction, Application Optimization/Parallelization, Software Engineering

Currently there exists a broad diversity in computer architecture that significantly complicates the creation of robust high performance scientific software. Due to the diversity, scientific applications that work well on one architecture do not necessarily work well on other architectures. Frequently, refactoring of source code to match the target architecture is necessary and resource intensive activity. There is a strong need for automating the code refactoring task. In this project, we will evaluate two complimentary approaches for automated code refactoring: 1) to use source to source transformation technology to assist a conventional compiler, and 2) to augment conventional compiler technology. First approach is a short- to mid-term solution that automatically generates architecture-optimized source code, which can be compiled by a conventional compiler. Second approach is a longer-term solution that a compiler automatically performs architectural specific transformations before making conventional optimization during compilation. Both approaches require similar static analysis techniques to get enough information for direction of code transformations.

Students: The project is open to graduate students.

Skills and Qualifications: This project will involve the working with compiler technology to analyze source code using both a custom Fortran parser as well as the LLVM compiler infrastructure. Students will work with important pieces of code extracted from real scientific applications. Students should have experience with internals of compiler technology.

*****NOTE*****

This project has the possibility of an extension for a second summer. Details TBA.

13. Open Source Technologies Supporting Data Science and Decision Analysis with Internet-of-Things 3D Printed Weather (IoTwx/3d) Stations

Areas of Interest: Data Science, Visualization, Internet of Things (IoT) and sensors/instrumentation

Fueled by the growth of Internet of Things (IoT), automatic weather stations are everywhere, and are integrated into our daily activities, whether we know it or not. These stations have fueled citizen science observation networks and DIY weather stations that have come into prominence around the world. With 3D-printed technology, low-cost microcontrollers and high precision digital sensors, supported by open software and hardware, a new era for continuous, advanced, automated weather observation is underway. When IoT weather stations are properly coupled with Free and Open Source Software (FOSS), they have great potential to accelerate the pace of research by lowering the cost of entry for researchers seeking to integrate field-deployable research instrumentation with data access and analysis software.

Most of the data from IoT and automated weather stations end up being archived and left for others to download raw files instead of being put to active use in supporting end user decision processes. And while there are commercial services to make the data more accessible via programmatic APIs, they often require programming expertise and many have limits on free access and usage of the data. There is a maturing ecosystem of open source data storage and IoT device management platforms, and many of them support real-time data collection, access and display, and some are beginning to support integrated decision analysis.

This research project has three goals: (1) perform a comprehensive investigation of open source IoT management platforms to support an IoT-connected weather station, (2) develop software strategy build on open source tools for managing, accessing, disseminating and realtime and historical data collected by the weather station, and (3) develop re-usable data analysis platform to support decision analysis atop tools such as Jupyter Notebooks or custom end-user dashboards, with the ultimate goal of making IoTwx data more immediately useful and actionable for end users.

This project is open to graduate students who have a passion for making data accessible and useful to everyone, and an interest in real-time devices. The successful candidate will be well-organized, highly motivated, self-driven, wickedly curious and have a keen interest in the future of ubiquitous environmental sensing in any context.

Students: The project is open to graduate students.

Skills and Qualifications: Required: proficiency with open source software, comfortable working in *nix (Linux) environments; proficiency with one modern programming language; familiarity with a data science stack such as Python/Anaconda; familiarity with Github and git; familiarity with data munging and data analysis. Desired: background working with IoT data protocols (e.g. MQTT); proficiency with Jupyter Notebooks, Python and MicroPython; command line scripting, compiling, and software packaging; working knowledge of statistical concepts for data science and machine learning, writing and developing object-oriented programs for open source

14. Using Neural Networks for Scientific Data Compression

Areas of Interest: Data Science, Machine Learning

The amount of numerical scientific data has grown so rapidly that lossy data compression is now being explored in many scientific domains. While lossy compression is traditionally performed by transformation (e.g. wavelet transform), predictor (linear, non-linear, etc.) based methods, neural networks have started to rise as a contender in this field. For example, in the image space, recent research from Google has shown compressors based on recurrent neural networks to be as effective as JPEG compression for certain use cases.

This SIParCS project aims to better understand the potential of neural networks in scientific data compression tasks. There will be several steps involved in this project. First, the compression experiment environment needs to be set up on NCAR's HPC systems, namely the Casper cluster with Nvidia Tesla V100 GPUs. Second, the existing research on neural-network-based compression, most likely techniques developed by Google, needs to be better understood and possibly reproduced on-site. Third, the image compression networks need to be evaluated (with possible adjustments) on numerical scientific data with respect to various metrics and use cases. Fourth, given adequate understanding on this topic, we aim to improve these neural networks with a focus on scientific data compression.

Note that the proposed work of the first two steps are considered to be realistic within the time frame and resources of SIParCS 2020, and the second two steps are not required for a single summer project. The execution of the third and fourth steps will be dependent on the progress of the first two steps.

Students: The project is open to graduate students.

Skills and Qualifications: Knowledge of deep learning theories and experience with neural network projects using one of the deep learning frameworks, such as TensorFlow, PyTorch, Keras, etc. Experience with data compression is desirable but not required.

Non-technical: CISL Outreach, Diversity, Education (CODE) Intern

Areas of Interest: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work or related field.

As the CODE Intern, you will provide administrative support to the SIParCS Program office and affiliated programs. You will also assist with planning and preparation for education and outreach programs to occur during the 2020-2021 school year. This is a full time student intern position that runs from mid-May through early August. End date may be flexible.

Student Intern Support

- Be an active participant on the SIParCS team to provide support and mentoring for students.
- Live-in at the apartments with the interns, and plan and participate in after-hours teambuilding activities. Apartments are suite-style/gender inclusive.
- Keep program leadership apprised of any issues that arise.
- May assist students/participants with special needs.
- May travel to assist with intern recruitment during fall months.
- Attend the Rocky Mountain Advanced Computing Consortium with the SIParCS program

Summer program logistics:

- Assist with summer program support including planning and running events. Events include orientation, professional development workshops, field trips, and other learning opportunities for interns.
- Assist with apartment move-in and move-out logistics, distributing and collecting student supplies, and coordinating with apartment administration.
- Help write and edit SIParCS Annual Report.
- Update program alumni tracking documents for program assessment and evaluation process

General administrative support:

- Maintain program databases and updating web pages.
- Edit and compile information for documents, posters, brochures and newsletters.
- Provide multimedia support at program events (photos, video) and keep inventory of program pictures.
- Provide meeting support and prepare meeting materials.

School-Year program support

- Update presentation slides and other educational materials for school-year programs.
- May write and or edit educational materials such as text updates for HPC-TV or Meteo-AR. <https://www2.cisl.ucar.edu/outreach/resources>
- May test and evaluate new educational materials.

Additional tasks and professional development opportunities may be assigned according to the intern's skills and interests.

Students: The project is open to graduate students.

Skills and Qualifications:

- Ability to work 40 hours/week from May 11 - July 31.
- Experience with word processing, database, and spreadsheet applications in a Windows environment (Microsoft Word, PowerPoint, Excel, Access) as well as a Google Apps environment (Sheets, Docs, etc).
- Proven ability to organize, prioritize, and follow through on multiple tasks, with close attention to detail.
- Good written and verbal communication skills with the ability to convey information to interns, visitors, and partners in a welcoming and professional manner. Skill in proofreading.
- Interest and sensitivity in working in diverse settings that include interns from a broad spectrum of underrepresented groups.
- Proven ability to work in a team and individually. A strong work ethic.
- Patience and adaptability.
- Discretion in handling confidential information.
- Occasional travel will be required for this position

REQUIRED: Must be enrolled in a graduate level college, university or accredited professional program in: Higher Education Administration / Student Affairs, Science Education, Education Policy, Social Work or related field.

DESIRED: Experience working with undergraduate and graduate students.