Environmental Modelling and Software 123 (2020) 104559

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Advanced cyberinfrastructure for intercomparison and validation of climate models

Ziheng Sun [a], Liping Di [a,*], Benjamin Cash [b], Juozas Gaigalas [a]

[a] Center for Spatial Information Science and Systems, George Mason University, USA
[b] Center for Ocean-Land-Atmosphere Studies, George Mason University, USA

ARTICLE INFO

Keywords:
Model comparison
Model validation
Climate science
Web service
Cyberinfrastructure
Big data

ABSTRACT

The current routine of comparison and validation in climate science is frequently static and of low efficiency, which hinders evidence-based decision making and scientific confidence. Due to the aggressively increasing resolution, complexity, and associated data volumes of climate models, objectively comparing multiple models and assessing their accuracy against observations is an ever-increasing challenge. We propose an integrated framework for harmonizing state-of-the-art cyberinfrastructure techniques with the user habits formed by long-term familiarity with existing community-oriented software. An open source prototype named COVALI is implemented and used to compare and validate the results of several widely-used climate models and datasets. Our results show that the proposed cyberinfrastructure-based strategy can significantly automate the comparison and validation processes in climate modeling. More importantly, the new strategy retains the existing user habits in the climate community while making it easier for scientists to adopt new technology in their research routine.

Software availability

Name of software: COVALI
Developer: Center for Spatial Information Science and Systems, George Mason University
Source language: Java
Availability: The source code and application jar can be accessed via Github: http://github.com/CSISS/cc

1. Introduction

Big data features heavily in two essential and constantly applied steps in climate modeling workflows: ***1) model comparison***, which evaluates the level of agreement between models, and ***2) model validation***, which evaluates the level of agreement between models and observations (Anderson and Bates, 2001; Flato et al., 2014). Climate models are comprised of complex systems of equations that include a number of assumptions and approximations that differ from model to model, and intermodel comparison is a means of assessing their collective impact (e.g., Randall et al., 2007). In addition to the obvious need to assess how accurate models are at reproducing the real climate system, comparison between models is also essential.

Currently, comparison and validation of climate models is generally performed by individual scientists using their own unique codes and workflows. The process frequently requires locating, accessing, transferring, regridding, and analyzing tens to hundreds of terrabytes of data. While some standard community analysis packages do exist (e.g. PMP, CESM diagnostics), the large number of steps and datasets lead to a situation where results obtained by one researcher cannot be readily reproduced and/or evaluated by third parties (Gleckler et al., 2008; Kennedy et al., 2011). A standard, widely-available strategy (tools, libraries) for comparison and validation of model results would greatly improve our understanding of model accuracy, applicability, scope, and error sources (Rood, 2011).

To standardize the comparison and validation processes, thus improving reliability and reproducibility as well as saving considerable time and effort, we present a cyberinfrastructure-based framework to enable streamlined one-stop comparison and validation in climate modeling. Inspired by state-of-the-art big data manipulation strategies in computer science, solutions involving cloud-based data storage and high-performance on-demand computing can be transplanted to scientific modeling (Zhang, 2019). The scripts, commands, and libraries used for comparison and validation will be managed remotely by cyberinfrastructure. Once any of the scripts, commands, or libraries is changed,

* Corresponding author. 4087 University Dr STE 3100, Fairfax, VA, 22030, USA.
E-mail address: ldi@gmu.edu (L. Di).

https://doi.org/10.1016/j.envsoft.2019.104559
Received 4 August 2019; Accepted 12 October 2019
Available online 17 October 2019
1364-8152/

the cyberinfrastructure will automatically re-compile and re-run the entire workflow (Sun, 2012). The results will be delivered to users once the execution is completed. Scientists can browse the results and side-by-side compare them with other model results or on-field observations discovered by the cyberinfrastructure.

The proposed framework will bring a change of routine to climate scientists and will make the numeric models more intercomparable and their results easier to validate against observations. A prototype system has been implemented to realize the framework and has been used to compare the results of several popular climate models and data products, e.g., WRF (Done et al., 2004), GFS (Saha et al., 2010), HRRR (Smith et al., 2008), NARR (Mesinger et al., 2006), ASR (Bromwich et al., 2018), ERA5 (Hersbach and Dee, 2016), via searching and visualizing in the web browser. Observational data from weather stations, radar, and sensor networks, are indexed, discovered, and displayed in the prototype to validate the model results and calculate the model errors. The results show that the proposed framework can greatly improve the efficiency of comparing and validating those models. The expected advantages including reduced time cost, increased scientific productivity, standardized processes, and greater confidence in results have been verified in the experiments. For the community at large, the proposed cyberinfrastructure will make it easier and faster to deliver model simulations from research to operation for climate-related activities, such as air quality monitoring, water availability, and climate change projections.

2. Urgent challenges in climate modeling

2.1. Model intercomparison

Many models have been created by different groups to simulate the climate system. At the time of writing, thirty-two separate models are listed as participating in the Coupled Model Intercomparison Project, phase 6 (CMIP6[1]) (Eyring et al., 2016). CMIP6 is a community-established framework for studying the output of coupled atmosphere-ocean general circulation models. It facilitates assessment of the strengths and weaknesses of climate models which can enhance and focus the development of future models (Zanchettin et al., 2016). Within CMIP6 are multiple subprojects focusing on exploring particular components of the climate system in depth. For example, the Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP) (Jones et al., 2016) consists of a subset of eleven coupled climate-carbon cycle models using a common protocol to study the coupling between climate change and the carbon cycle. For each model, two simulations were performed in order to isolate the impact of climate change on the land and ocean carbon cycle, and therefore the climate feedback on the atmospheric CO_2 concentration growth rate. The results of different models are compared and analyzed to identify the agreement and disagreement among them and to help understand their causes. Another project, the Cloud Forcing Model Intercomparison Project (CFMIP) focuses on quantifying and comparing the critical cloud radiative feedbacks across the suite of participating models (Webb et al., 2017). Taken together, hundreds of derived versions of various models exist across the climate community. As climate models have become increasingly complex, and have moved to higher resolution, the need for innovative methods of data discovery as well as intercomparison and analysis of the resulting simulations has grown more urgent.

2.2. Model validation

Governments, organizations, and individuals frequently ask if climate models are reliable (Hutton et al., 2016). Rigorous model validation is crucial to demonstrate that simulations are sufficiently accurate to inform decisions on grand challenges on behalf of both human society and our living environment (Randall et al., 2007). In addition to making forecasts, models are regularly tested on their ability to reproduce past weather and climate events (Karmalkar et al., 2011; Rood, 2011), a process referred to as "hindcasting". Hindcasting is particularly important for assessing climate change projections, as an alternative to waiting for 30 years to verify the accuracy of a given projection. If a model can correctly predict trends from a starting point somewhere in the past, there is less reason to think its predictions of future climate states will be wrong (Pedersen and Winther, 2005). Reanalysis products such as ERA5 (Hersbach, H., 2018), which are blended datasets comprised of both observed and simulated data, are frequently used for validating models as well as providing initial conditions for forecasts and boundary conditions for regional models.

Climate models are also assessed on their ability to reproduce the average state of the climate system, known as the climatology. For example, researchers check to see if the average temperature of the Earth in winter and summer is similar in the models and reality. They also compare quantities such as sea ice extent between models and observations and may choose to use models that do a better job of representing the current amount of sea ice when trying to project future changes. Similarly, specific events that have a large impact on the climate, such as volcanic eruptions, can also be used to test model performance. The climate responds relatively quickly to volcanic eruptions, so modelers can see if models accurately capture what happens after big eruptions, after waiting only a few years. Studies show that significant uncertainties remain in model simulations of changes in temperature and in atmospheric water vapor after major volcanic eruptions (Zanchettin et al., 2016).

2.3. New technology vs old technology

In recent years, climate science has struggled with the growing disparity between conventional tools and new emerging techniques (Sun, 2019), neither of which is sufficient alone (Gimeno, 2013; Ng et al., 2017). The conventional tools are solid, robust, and have a large number of users (Craig et al., 2005; Kouzes et al., 2009; Lupo and Kininmonth, 2013; Tsipis, 2019). However, as technology evolves and data accumulates, the conventional tools are falling short of addressing new challenges such as higher-resolution big data (Lopez and Manogaran, 2016; Schnase et al., 2016; Sellars et al., 2013) and the inherent limits of numeric models (Abramson et al., 2005; Mizielinski et al., 2014). New technologies are growing in popularity because of their potential to meet those challenges (Hashem et al., 2015; Reed and Dongarra, 2015), and therefore it is to be hoped that they will gradually replace the dominant role of conventional tools in the climate research (Evangelinos and Hill, 2008; Haupt et al., 2008; Huntington et al., 2017; Kouzes et al., 2009; Schnase et al., 2017; Stewart et al., 2015).

However, it will never be simple to convince scientists to drop tools that they have been using for decades and use something new and unfamiliar, even when the new technologies are more powerful and efficient (Fernández-Quiruelas et al., 2011). This conflict follows the birth of all new technologies. The chaotic multitude of languages, tools, and libraries makes the situation worse and leads to a situation where the abundancy of big data (petabytes of observational, reanalysis, simulated datasets online) cannot be fully exploited to benefit climate researchers.

3. Advanced cyberinfrastructure solution

Web systems have taken on a big portion of the market in scientific data processing today because of their high availability, scalability, and efficiency. Cyberinfrastructure, which is a term summarizing these web systems and their served capabilities, is not a new technology but a merger of tools, data, and human resources into a seamless interoperable platform. While processors, storage devices, sensors, and other physical assets are part of CI, it is more than the practice of connecting people

[1] https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/.

with advanced networks and sophisticated applications running on powerful computer systems-it aims to involve those people as participants in the generation of knowledge, giving them the opportunity to share expertise, tools, and facilities. CI is also known as e-research, e-science, and e-infrastructure in Europe, Australia, and Asia, bringing together high-performance computing, remote sensors, large data sets, middleware, and sophisticated applications (modeling, simulation, visualization)(Cyberinfrastructure vision for 21st century discovery). Beyond the technology, an essential part of CI is the way it allows distributed teams to turn "flops, bytes, and bits into scientific breakthroughs"(Cyberinfrastructure vision for 21st century discovery).

CI provides a technical infrastructure which knits together high-speed networks with high-performance, high-availability, and high-reliability computational resources (Hey, 2005). Data can be large aggregations of previously collected data (e.g., reanalysis) or live feeds from remote sensors (e.g., satellite observations, NEXRAD network). Many of the systems are housed in different locations, and experiments typically run on virtual machines in which spare cycles from dozens or hundreds of computers are used for a single task. Data sets can be distributed as well, with data coming from multiple institutions (Schnase, 2016).

CI provides an opportunity for a new kind of scholarly inquiry and education, empowering communities of researchers to innovate and revolutionize what they do, how they do it, who participates (Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure). Data and experiment notebooks are posted online as they are collected, facilitating real-time, distributed science. To enhance climate science and the ability to meet the two critical challenges of model intercomparison and validation, the required advanced cyberinfrastructure should at least provide the following capabilities.

3.1. Multisource data discovery and multifile subsetting collection

Due to the big data challenges in climate science, there are only a few choices available for scientists to find their desired data sets and collect them as initial input data of their models. Currently, a common routine is that scientists either search in Google to find the data providers' descriptive web pages and access the data via their portal entrance link or learn about the data sets from colleagues/conferences/seminars and go directly to their websites to access the data sets. Many datasets are too big (e.g., over 100 TB) to readily transfer and store locally, requiring scientists to subset or downsample the data on site and prior to transferring the reduced data to their own facilities. However, the current cyberinfrastructure still has a lot of room for improvements (e.g., THREDDS Data Server need search function). To solve this problem, advanced cyberinfrastructure should be equipped in the data centers to further facilitate the manipulation of big data to improve climate research.

3.2. Efficient regridding

All veteran climate modelers know the sophisticated challenges raised by the gridding task (Hill et al., 2004). Scientists don't have a limitless supply of computing power at their disposal, and so it is necessary for models to divide up the Earth into grid cells to make the calculations more manageable. This means that in every step of the model takes through time it calculates quantities averaged over a grid cell that may be 100 km on a side. However, the precise definition of the grid typically varies from model to model. The proliferation of models discussed in section 2.1 means that scientists must be able to work with many different grid definitions, and direct comparison of models requires first transforming the data to a common grid. Likewise, many observational data products are either supplied on their own grids (such as reanalysis) or at point locations (such as station data), and validating models against such data again requires a regridding step. Automated, efficient conversions between different grids would lead to significant savings in time and effort in model analysis workflows.

3.3. Interactive visualization

Recent development in web map services such as Google Maps, Bing Maps, Apple Maps, has greatly reformed the habits of people examining spatial datasets. When it comes to visualize spatial datasets, people generally expect something similar to those widely-used map applications which allow users to freely zoom, pan, rotate, search, and query on maps. Rapid and interactive visualization is also a key element in the development of successful climate analysis workflows, allowing immediate and flexible feedback on the outcome of experiments and the identification of both errors and features of interest. CI tools for visualization should be intuitive and provide steeing, 3 or higher dimension mapping, as well as many other functions such as on-the-fly clicking for values, changing legends, switching projections, and adjusting opacity.

3.4. Result comparison and difference calculation

Advanced climate cyberinfrastructure should serve the capability of comparing the results from different models and data products and calculate their differences. Comparing to geographic information systems, climate scientists are more comfortable with side-by-side comparison instead of layers that are overlaid together and compared by showing and hiding each layer. Cyberinfrastructure development should honor the user habit and make the comparison function friendly enough for veteran scientists to adopt. Difference calculation should follow the community paradigm to produce the difference map and avoid potential misunderstanding caused by the changes in map production.

3.5. Hindcast validation

Hindcasting in climate science, or backtesting in many other disciplines, is one of the major validation strategies for climate models at present[2] (Sotillo et al., 2005). In hindcast experiments, researchers test the ability of a model to reproduce the known or estimated state of the climate system (e.g., reanalysis data) in the past (Chawla et al., 2013). To facilitate the hindcasting experiment, advanced cyberinfrastructure for conveniently retrieving both reanalysis data and observed data and automatically performing standard comparisons (including the necessary regridding noted in section 3.2) would be in high demand. Dealing with the twin challenges of data volume and variety are key challenges in establishing a qualified cyberinfrastructure for hindcasting.

3.6. FAIR data

To reproducibility of and confidence in the results of climate models, many[3] scientific communities and publishers have affirmed their commitment to the principle that scientific data should be findable, accessible, interoperable, and reusable (FAIR). Given the volume and complexity of climate simulations and observations, implementing FAIR principles represents a significant technical and logistical challenge. There should be cyberinfrastructure allowing climate modelers to easily publish their results in FAIR web services which could interoperate with the models of other components of the Earth system and beyond to generate a big picture of environmental current status as well as the future trends.

3.7. Automation

After years of performing challenging and complex data processing

[2] https://www.emc.ncep.noaa.gov/users/meg/fv3gfs/.
[3] https://copdess.org/enabling-fair-data-project/.

tasks (Yue, 2010), climate scientists eagerly expect automation as one of the essential features in advanced cyberinfrastructure (Sun and Di, 2019; Sun et al., 2012, 2014). Scientists spend hours and hours on a daily basis on preprocessing datasets, because one error in data processing could lead to hundreds of computation hours wasted (they must recalculate the models again) (Sun et al., 2017). A workflow management module should be provided to help scientists oversee the entire process, check provenance (Sun, 2013), quickly locate the sources of errors and decrease the resource waste caused by wrong runs of models to the minimum (Sun, 2019). In model comparison and validation, once the files on either side of the comparison/validation settings are changed, the comparison workflow should be triggered by cyberinfrastructure to automatically calculate the difference (Sun, 2016) and the results will be sent to the subscribers of the comparison instantly once the calculation is over.

4. Prototype implementation

To implement such a cyberinfrastructure, we utilized dozens of state-of-art techniques, tools, libraries and performed substantial software development work and held regular discussions with modelers. A layered architecture interoperability framework including data repositories and discovery catalogs, models, workflow environments, and analysis and visualization tools, is adopted to manage resources and provided functionalities (Fig. 1). The system is named COVALI, short for intercomparision and validation (the user interface in Fig. 2), which is implemented as a module system of the NSF-funded EarthCube CyberConnector. Different from other online web mapping systems, COVALI is customized according to the user habits of climate scientists formed in their long-term research careers. The interface is divided into two side-by-side maps, each of them has multiple view modes including 2-D and 3-D modes. There are two toolbars on the top and left side, providing redundant entries for scientists to capabilities such as data discovery, retrieval, uploading, visualization, metadata display, layer manipulation, animation, pixel value query, difference calculation, print, etc. There is a scaler for each map to indicate the current scale of the corresponding map window. A legend panel at the bottom shows the layer name and the color-value legend of the top data layer in the map. A small container on the top right area shows the coordinates of the mouse pointer to remind the real-time location scientists are pointing at. At the bottom right corner, a projection/mode selector is provided for users to switch among various projection and dimension choices. COVALI is a decentralized open-source system which can be downloaded from its Github repository (https://www.github.com/CSISS/cc) and installed on most machines including servers, clusters, cloud virtual machines, personal laptops, where the climate models and datasets reside. The normal usage of COVALI relies on no third-party services. After being installed, COVALI allows scientists to browse the data files in public folders on the host servers, interactively visualize them on the maps, intercompare the simulation of different models side by side, and validate the accuracy with observed data products ranging from gridded reanalysis data to individual station measurements of precipitation and temperature.

The development of the system combines various community-standard tools which are very familiar to climate scientists. For example, the Unidata NetCDF library (Java) is used to parse the variable metadata and visualize NetCDF files in web browser. ncWMS is used for rendering the NetCDF/GRIB files into standard web map service which can be directly integrated by many map tool providers. UCAR THREDDS Data Server is connected to provide the NEXAD radar datasets (recent two weeks). UCAR research data archive is connected to retrieve the reanalysis data products (e.g., Arctic System Reanalysis). NCO (NetCDF Operator) command lines are integrated to manipulate and analyze NetCDF data files with many powerful mathematical and statistical algorithms (difference calculation, ensemble, subsetting, average, and etc.). OpenLayers, a popular open-source JavaScript library, is used as the API for building the two rich web-based maps which can be panned, zoomed, rotated, and overlaid. CesiumJS is utilized to provide 3-D virtual globe view mode for visualizing both static and dynamic data.

5. Experiment & showcase

We have helped climate scientists use COVALI in comparing and validating their model results. They exercised COVALI to compare a series of climate model results, some of which are introduced below. However, it is important to note that while the comparisons choose the closest reanalysis products and the dates might not be strictly the same. These comparison results aim to demonstrate the capabilities of the proposed framework only.

5.1. GFS vs HRRR

The Global Forecasting System (GFS),[4] is an operational weather forecast model developed by the National Centers for Environmental Prediction (NCEP) (Whitaker et al., 2008). GFS datasets consist of dozens of variables, such as temperature, wind, precipitation, and ozone concentration and are used by the forecasters to predict weather up to 16 days in advance.

HRRR[5] short for the High-Resolution Rapid Refresh, is a NOAA real-time 3-km resolution, hourly updated, cloud resolving, convection-allowing atmospheric model. Radar data is assimilated in the HRRR every 15 min over a 1-h period adding further detail to that provided by the hourly data assimilation from the 13 km radar enhanced rapid refresh.

We compared their convective available potential energy (CAPE), a measure of the energy present in the atmosphere to fuel the development of storms, for CONUS at the same time period (Apr 22, 2019 00:00) shown in Fig. 3. The two maps use the same legend (maximum, minimum, color style, base map, scale, and etc.) and can be zoomed to specific regions for detail comparison. In COVALI, the differences of the two models are very obvious in the central plain region and across the southeast. The impact of the higher resolution in HRRR is very apparent in the more sharply-defined nature of the main feature. Conveniently, COVALI allows users to rapidly and easily adjust the focus region to compare results at various scales.

5.2. MERRA2 vs NARR

The NCEP North American Regional Reanalysis (NARR) (Mesinger et al., 2006) products are long-term (1979-present) atmospheric datasets with 3-h temporal, 32-km horizontal, and 45-layer vertical resolutions over the North American domain. NARR contains outputs of many atmospheric variables and fluxes and is nicely suited for diagnosis of synoptic and mesoscale conditions. In this portion of the study, we use NARR monthly means and the NASA Modern-Era Retrospective analysis for Research and Applications, Version 2 (**MERRA-2**) monthly data to compare their assessments of the air temperatures at 2 m above the surface (t2m) (Fig. 4a). The results above the coastal states are relatively close. However, further in-land states, particularly in the region of the Appalachian mountains, the two datasets have clear disagreements. These differences are difficult to see, however, using the default palette and value range, so to highlight these differences we use COVALI to adjust the value range of the color legend. (Fig. 4b). Now the broader area of colder temperatures near the mountains in MERRA2 relative to NARR can be idenfied clearly.

5.3. ASR vs ERA5 reanalysis

ASR (Arctic System Reanalysis) is a retrospective reanalysis of the

[4] https://www.emc.ncep.noaa.gov/GFS/.php.
[5] https://rapidrefresh.noaa.gov/hrrr/.

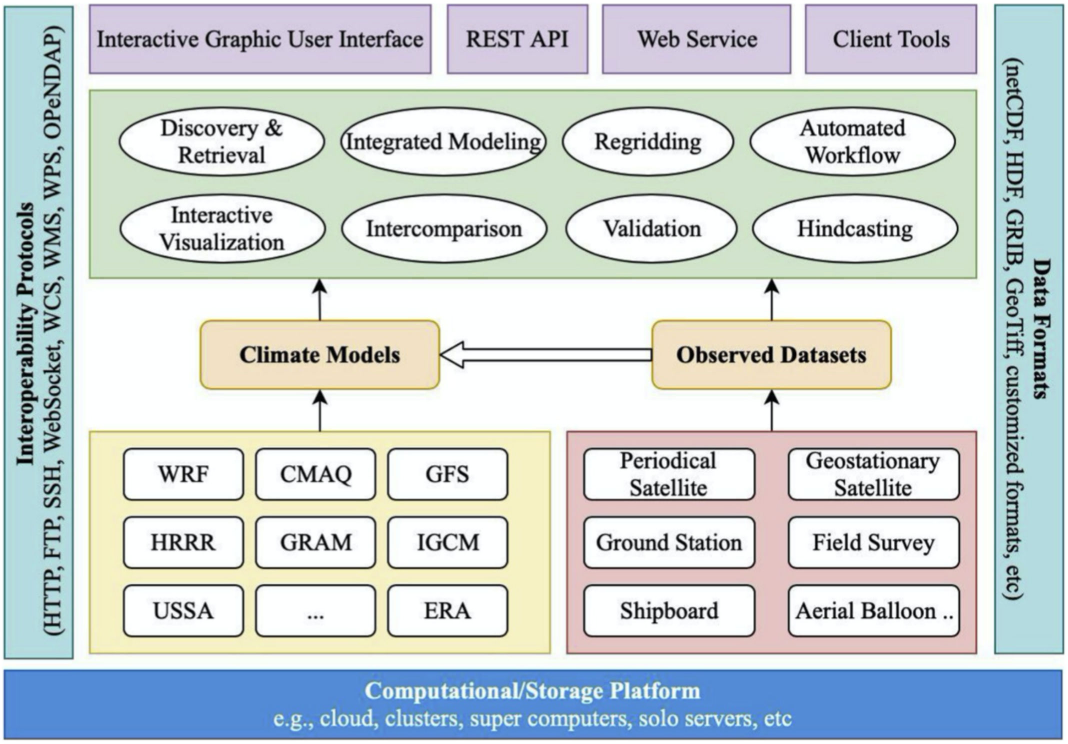

Fig. 1. Proposed CI solution.

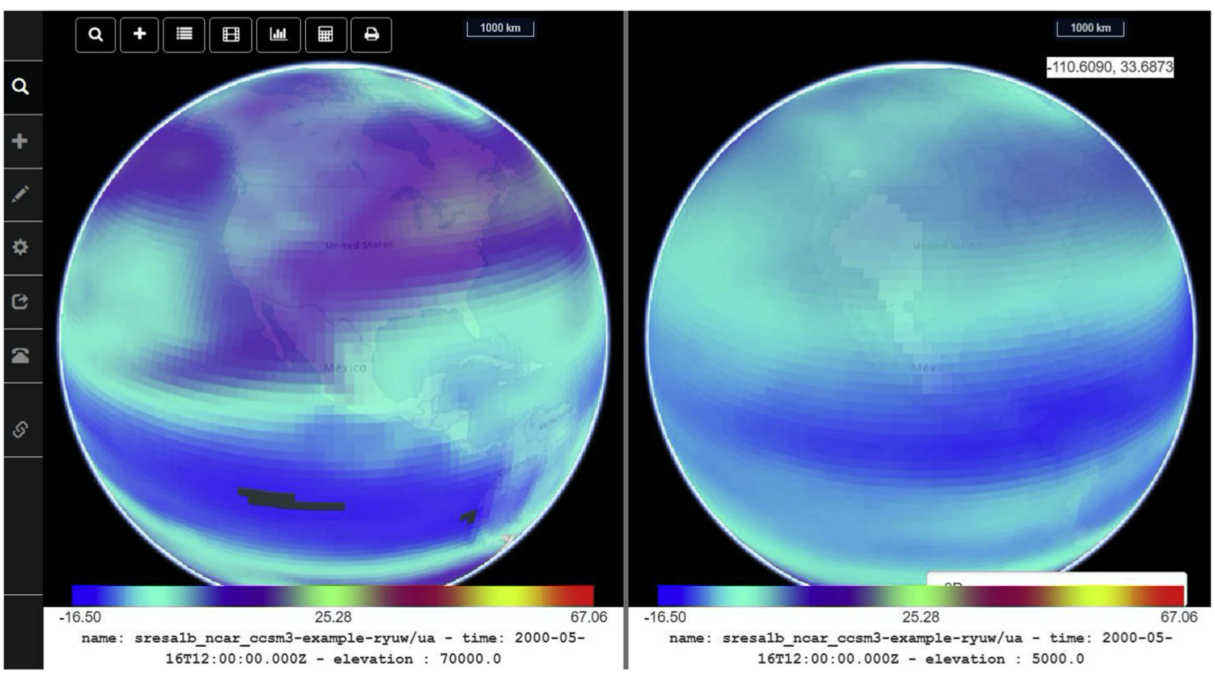

Fig. 2. COVALI user interface.

Great Arctic region produced by high-resolution versions of the Polar Weather Forecast Model (PWRF) and the WRF-VAR and High-Resolution Land Data Assimilation (HRLDAS) data assimilation systems that have been optimized for the Arctic. It covers the period of 2000–2016 at two resolutions: 15 km and 30 km. It has 29 pressure levels, 27 surface and 10 upper air analysis variables, 74 surface and 16 upper air forecast variables, and 3 soil variables.

The ERA5 dataset is the fifth generation of ECMWF (European Centre for Medium-Range Weather Forecast) reanalysis of the global climate, which started with the FGGE reanalyses produced in the 1980s, followed by ERA-15, ERA-40 and more recently ERA-Interim. ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables, and currently covers the period 1979 to within 3 months of real time. It combines vast amounts of historical observations

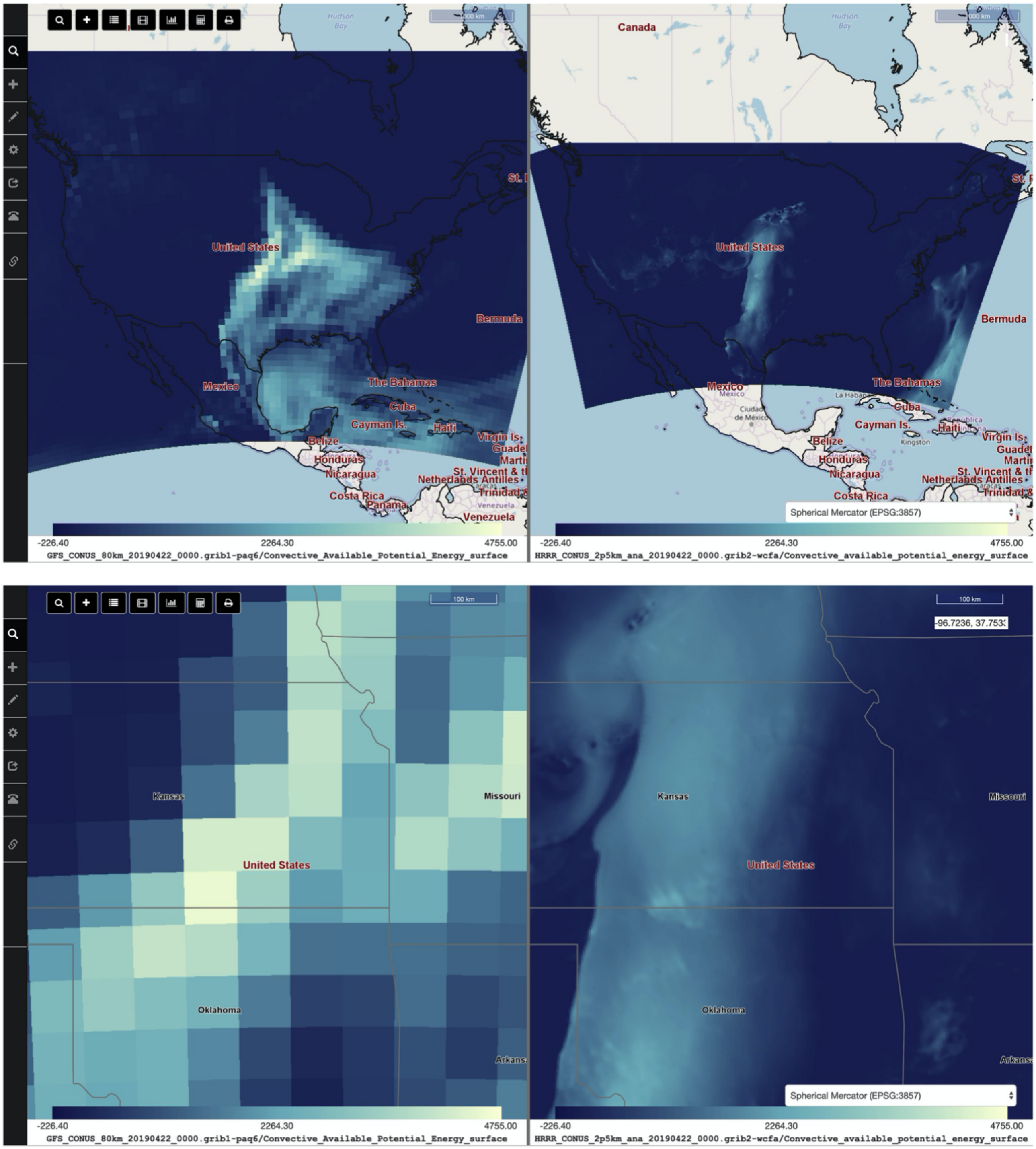

Fig. 3. Comparison of CAPE for GFS (left) and HRRR (right).

in its estimation. Specially, ERA5 includes information about its uncertainties for all the variables at reduced spatial and temporal resolutions for quality assurance. The hourly analysis and twice daily forecast parameters form the basis of the monthly means (and monthly diurnal means) in the dataset.

We used COVALI to load their estimates of the albedo in the greater Arctic. Fig. 5 shows the results with ASR on the left map and ERA5 on the right. The general distribution of high and low areas matches well, especially in the Arctic ocean, Canada, Russia, and Alaska. Differences are localized and driven by reasonable forcings. Via the comparison it could be concluded that both ASR and ERA5 produced consistent estimates in the Arctic. More improvements on each model might be made at local scale and further validated by the future datasets collected by Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC).

5.4. *Model validation with ground station measurements*

COVALI not only provides a platform for interactively visualizing the model results, but also allows directly overlaying observations collected

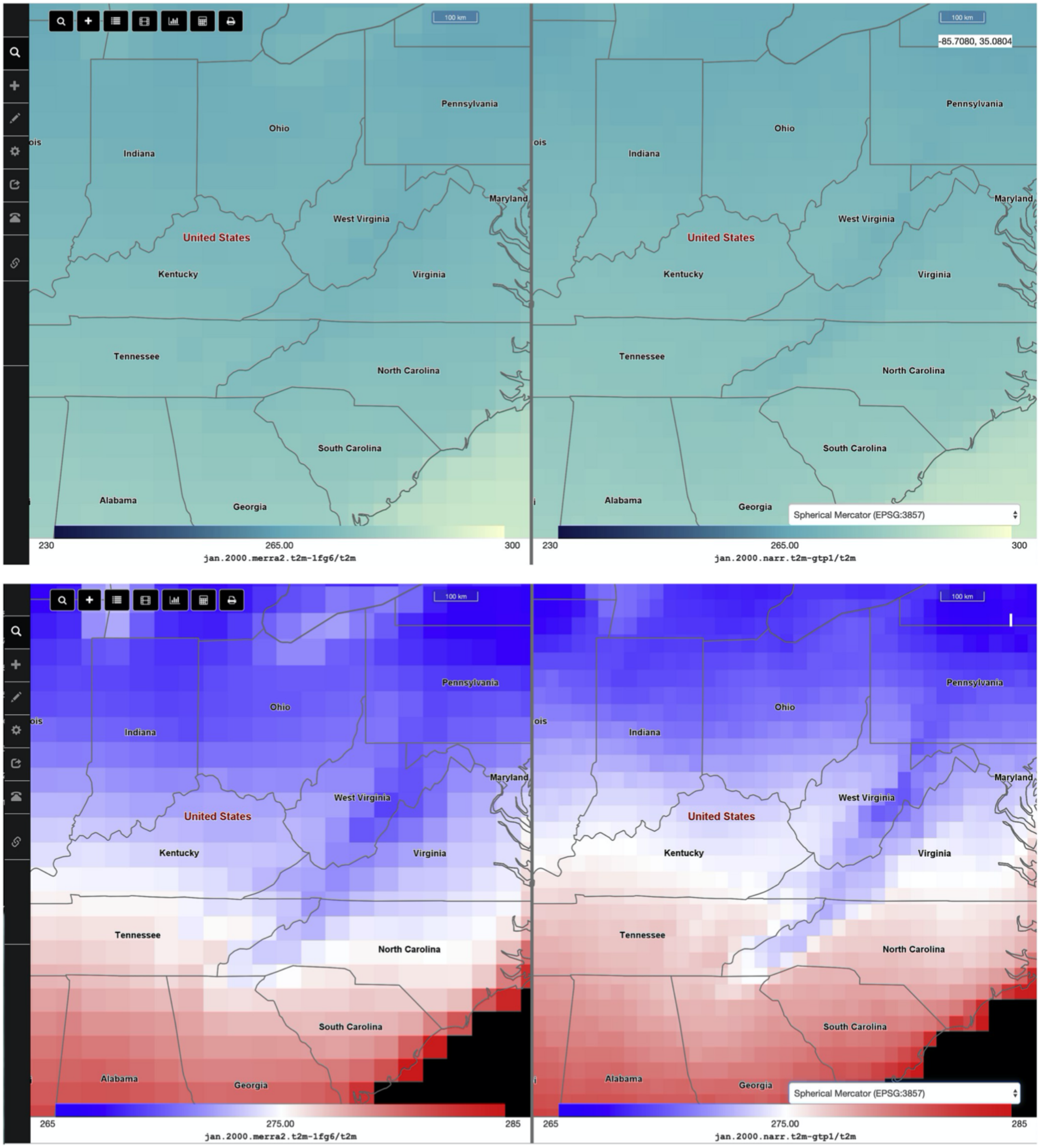

Fig. 4. Comparison of 2-m temperature (t2m) in GFS reanalysis (left) and NCEP North American Regional Reanalysis (NARR, right).

by ground stations. EarthCube CHORDS (Cloud-Hosted Real-Time Data Services for the Geosciences) is a real-time data service infrastructure that will provide an easy-to-use system to acquire, navigate, and distribute real-time data streams from 3-D printed, self-deployed in-situ sensors. The CHORDS group has deployed several sensors in Colorado which are keeping streaming the observations to their cloud-hosted databases. Using the CHORDS API, COVALI can real-time retrieve and display those observations on the maps (Fig. 6). We compared the precipitation in GFS hindcast predictions with the accurate observation from CHORDS by clicking on the maps. Scientists can check their results against the real observations to evaluate the accuracy of their models and to help in identifying sources of errors.

Besides CHORDS, COVALI can also directly load observations from IRIS station network. Although IRIS network is specifically used for seismology research, they also have some surface sensors which provide near surface measurements of variables like radiation, temperature, and etc., and can help climate models to evaluate their results at near ground surface level. Fig. 7 shows the distribution of some stations in the IRIS network. Scientists can click on the station icons to get the station name, the station network, channels, and the measurements of each channel. In

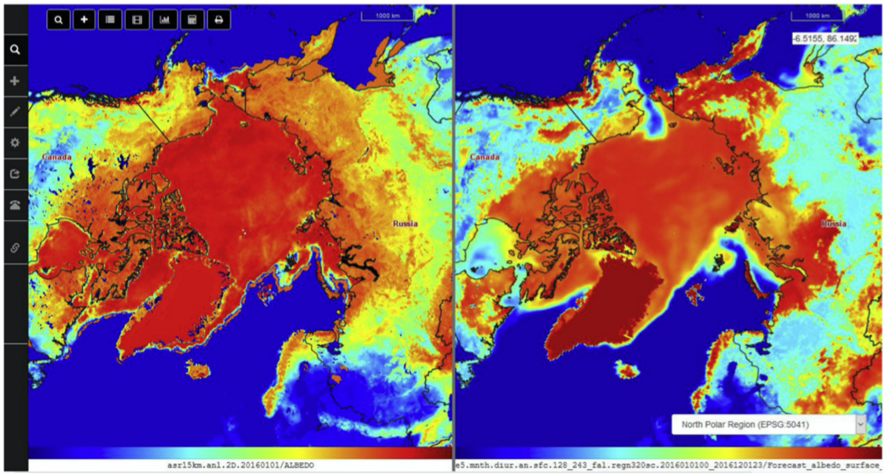

Fig. 5. Comparison of Arctic albedo between ASR (left) and ERA5 (right).

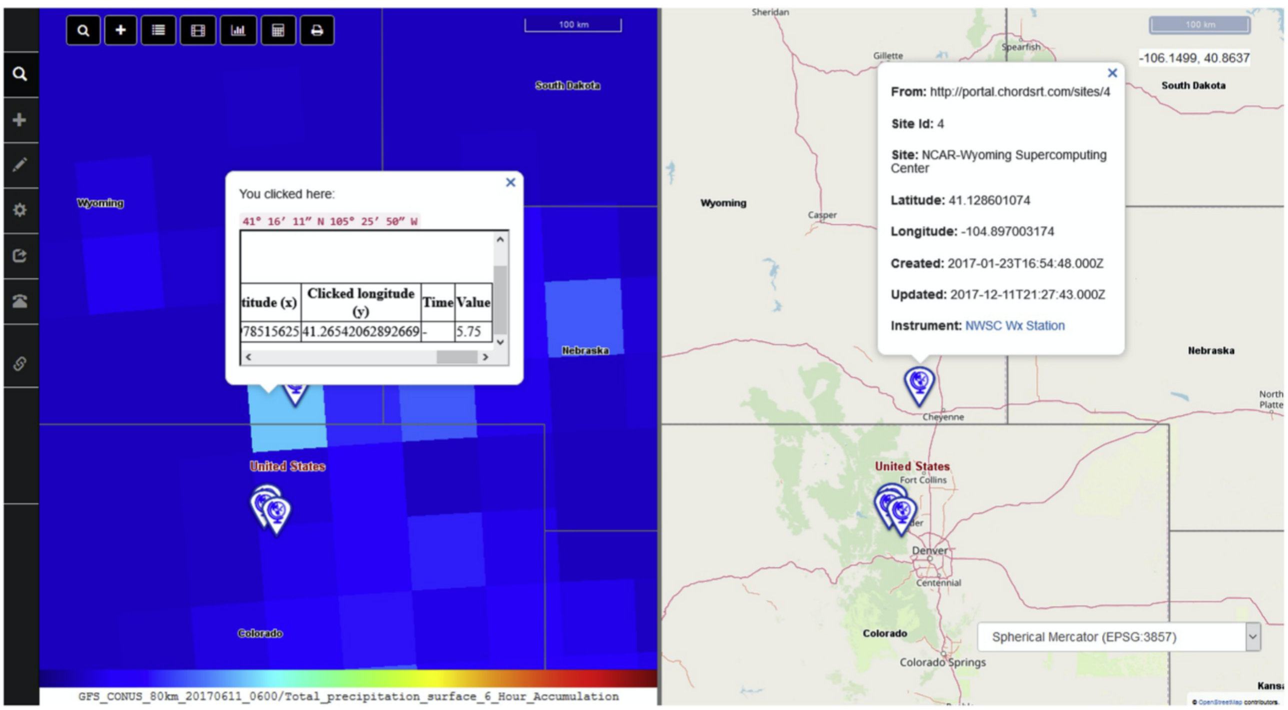

Fig. 6. Model validation with 3-D printed sensor measurements from EarthCube CHORDS.

future, more ground or remotely sensed datasets will be made available to scientists in COVALI to facilitate directly adjusting their results against the real-world datasets.

5.5. *Data processing workflow*

To assist COVALI to make model results more comparable, Geoweaver, an open-source geoscientific workflow system, is used in combination with COVALI to assist the creation and running of data processing workflow to automate the tedious daily jobs like regridding, averaging, reformatting, reprojecting, and etc. (Fig. 8). COVALI uses Geoweaver in the backend to complete the data processing tasks received through its front interface. Geoweaver is very flexible for managing resources like servers, cloud virtual machines, clusters, scripts, Notebooks, Python code, and workflows. Underneath the Geoweaver processes are the community-standard tools, such as NCO, GrADS, NCL, THREDDS Data Server, and etc., which provide the actual algorithms and commands to operate on the data files. Geoweaver wraps all the resources into automated workflows and greatly reduces the time cost of scientists spent on the frequently repeated data processing per day.

6. Discussion

One biggest advantage of COVALI is to make climate model results and observed datasets FAIRable (findable, accessible, interoperable, reusable) via a single platform. Model results can be intercompared and validated in an interactive and uniform interface instead of static and random images. Since it runs in web browsers, COVALI allows scientists to bypass many layers of protocols and connection techniques to directly

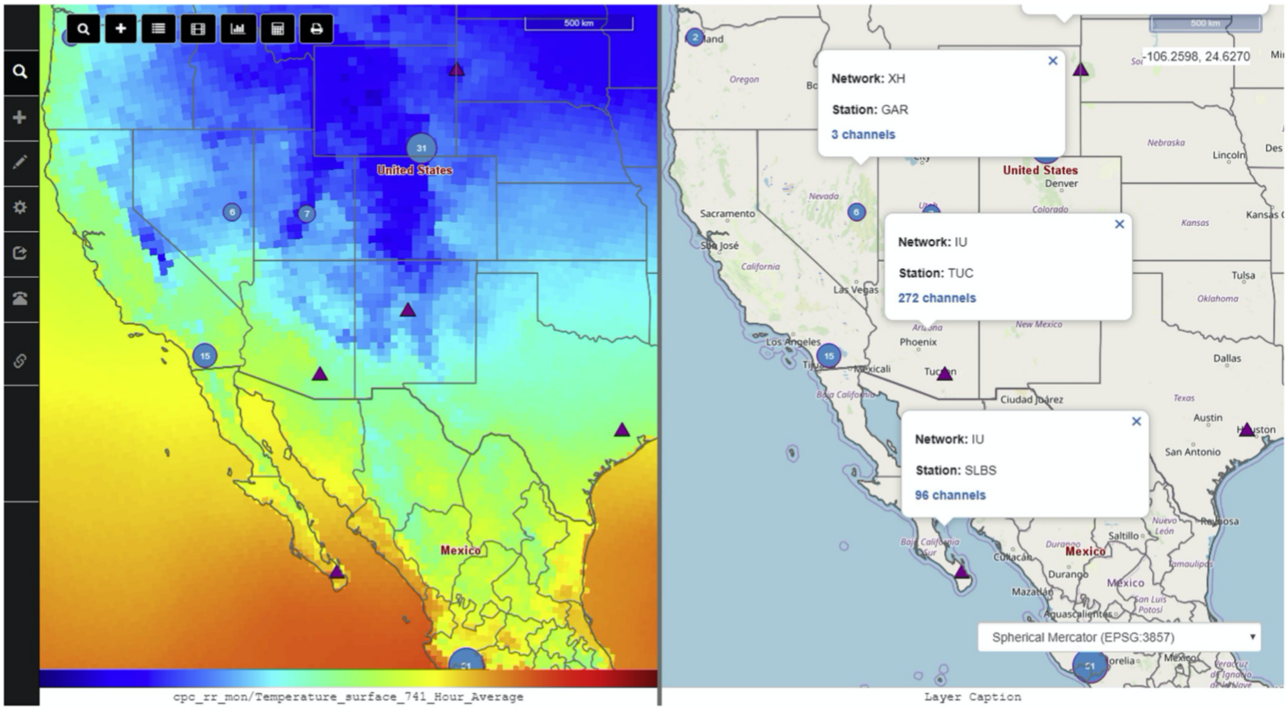

Fig. 7. Model validation with IRIS station networks.

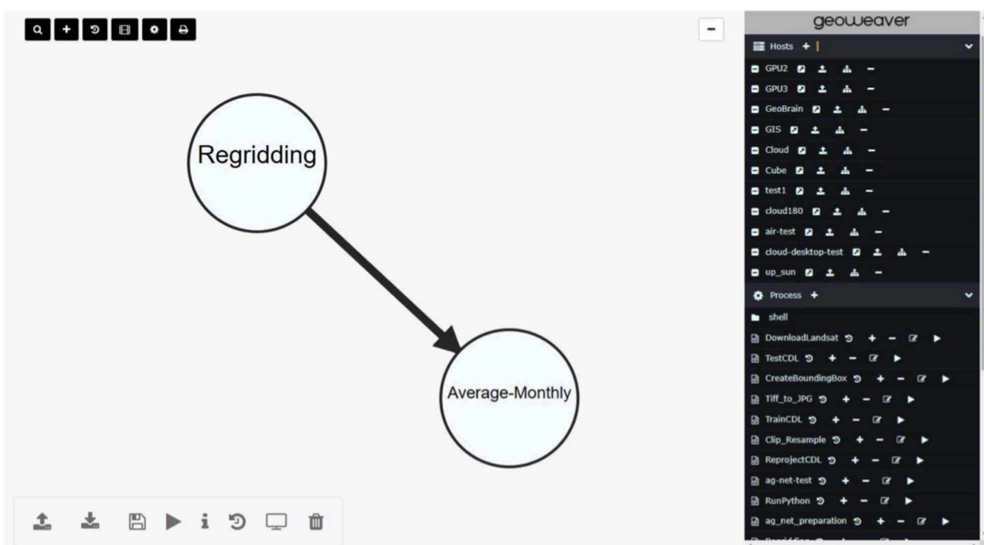

Fig. 8. Geoweaver for data processing automation (e.g., regridding, averaging).

manipulate the model results. Regarding the volume challenge of big data in climate science, COVALI provides an advanced cyberinfrastructure solution to meet the substantial problems in data retrieval, visualization, pre-/post-processing, statistics, intercomparison, and validation. The availability of COVALI could make these daily activities of climate scientists easy, straightforward, automatic, and time efficient. It can help improve climate modeling by providing objective comparison among various models and hindcasts, and also enhance integrated modeling by allowing the model results to be accessible, downloadable and reusable by other modelers. The simple integration of ground station observations will facilitate validation efforts that answer the reliability questions asked by all the stakeholders. In addition, the tedious data processing tasks which take a big portion of work time of climate scientists will be simplified and automated in COVALI using scientific workflow techniques. The software can run on Windows, Mac OS X, and Unix/Linux and others that can run Java Virtual Machine. Every scientist can install a copy on their machines to take advantage of its capabilities.

The major barrier preventing the usage of cyberinfrastructure like COVALI in real world scenarios is the long-term user habits and software practicality. It takes a while for people, not just scientists, to transfer from their familiar tools to something brand new. We are aware of this obstacle and have taken it into consideration since the beginning of the project. The interface of COVALI is made into a similar layout and style as the ones that climate scientists are familiar with. During the development, we have periodically interacted with scientists to discuss the design of user interfaces and functionalities to eliminate resistance in their minds. Currently COVALI is at the prototype phase and we will follow the best practices of standard software engineering cycle to develop it into an operational system which is robust and useable by the large climate community.

7. Conclusion and future work

We proposed an advanced cyberinfrastructure framework to enable intercomparison and validation of climate models and observed datasets via web-based systems. The proposed method will bring a change in the research routine of climate scientists and make the numeric models more intercomparable and the validation of results with ground truth data more efficient. A prototype system has been implemented using Java to realize the framework and has been used to compare the results of widely used climate data sets: WRF, GFS, HRRR, NARR, ASR, ERA5, and etc. We worked together with specialists of climate models and help them to use the developed system to compare their results with other existing products via searching and visualizing in the web browser. Validation data from weather stationary datasets, radar observation dataset, sensor networks, are indexed, discovered, and displayed in the prototype to validate the model results and calculate the model errors. The results show that the proposed framework for advanced cyberinfrastructure can greatly improve the automation and efficiency of intercomparing and validating numerical models.

In future, we will make more ground and remotely sensed datasets available in COVALI and provide more data processing and analysis functions. By reaching out to the climate community, we will solicit engagement of users and developers of different models to compare their results via COVALI. The short-term benefits of adopting the cyberinfrastructure in the climate domain include reduced time cost, standardized process, confident results, and improved reproducibility. Ultimately, COVALI will automate the model intercomparison and validation for improving the model simulation to support the requirements of operational decision making in atmosphere-related activities, such as air quality monitoring, greenhouse effect prediction, ozone monitoring, dust forecasting, atmosphere thickness monitoring, air pollutant control, etc.

Acknowledgment

This study is supported by a grant from National Science Foundation EarthCube program (grant # AGS-1740693, PI: Dr. Liping Di). The authors of this paper would also like to thank all the authors and copyright holders of the open source software used in this work and the experts of tested climate models, Dr. Sheng-hung Wang, Dr. David Bromwich, Dr. Jim Kinter and Dr. Daniel Tong for their support, and Earth Science Information Partnership Federation (ESIP) for their support on Geoweaver.

References

Abramson, D., Kommineni, J., McGregor, J.L., Katzfey, J., 2005. An atmospheric sciences workflow and its implementation with web services. Future Gener. Comput. Syst. 21 (1), 69–78.

Anderson, M., Bates, P., 2001. Model Validation. Perspectives in Hydrological Science. John Wiley & Sons.

Bromwich, D., Wilson, A., Bai, L., Liu, Z., Barlage, M., Shih, C.-F., Maldonado, S., Hines, K., Wang, S.-H., Woollen, J., 2018. The arctic system reanalysis, version 2. Bull. Am. Meteorol. Soc. 99 (4), 805–828.

Chawla, A., Spindler, D.M., Tolman, H.L., 2013. Validation of a thirty year wave hindcast using the Climate Forecast System Reanalysis winds. Ocean Model. 70, 189–206.

Craig, A.P., Jacob, R., Kauffman, B., Bettge, T., Larson, J., Ong, E., Ding, C., He, Y., 2005. CPL6: the new extensible, high performance parallel coupler for the Community Climate System Model. Int. J. High Perform. Comput. Appl. 19 (3), 309–327.

Done, J., Davis, C.A., Weisman, M., 2004. The next generation of NWP: explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. Atmos. Sci. Lett. 5 (6), 110–117.

Evangelinos, C., Hill, C., 2008. Cloud computing for parallel scientific hpc applications: feasibility of running coupled atmosphere-ocean climate models on amazons ec2. Ratio 2 (2.40), 2–34.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization. Geoscientific Model Development. Online) 9 (LLNL-JRNL-736881).

Fernández-Quiruelas, V., Fernández, J., Cofiño, A.S., Fita, L., Gutiérrez, J.M., 2011. Benefits and requirements of grid computing for climate applications. An example with the community atmospheric model. Environ. Model. Softw 26 (9), 1057–1069.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., 2014. Evaluation of Climate Models, Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 741–866.

Gimeno, L., 2013. Grand challenges in atmospheric science. Front. Earth Sci. 1, 1.

Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008. Performance metrics for climate models. J. Geophys. Res.: Atmosphere 113 (D6).

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. 47, 98–115.

Haupt, S.E., Pasini, A., Marzban, C., 2008. Artificial Intelligence Methods in the Environmental Sciences. Springer Science & Business Media.

Hersbach, H., 2018. Operational Global Reanalysis: Progress, Future Directions and Synergies with NWP. European Centre for Medium Range Weather Forecasts.

Hersbach, H., Dee, D., 2016. ERA5 reanalysis is in production. ECMWF newsletter 147 (7).

Hey, T., et al., 2005. Cyberinfrastructure for e-Science. Science 308 (5723), 817–821. https://doi.org/10.1126/science.1110410.

Hill, C., DeLuca, C., Suarez, M., Da Silva, A., 2004. The architecture of the earth system modeling framework. Comput. Sci. Eng. 6 (1), 18.

Huntington, J.L., Hegewisch, K.C., Daudert, B., Morton, C.G., Abatzoglou, J.T., McEvoy, D.J., Erickson, T., 2017. Climate Engine: cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. Bull. Am. Meteorol. Soc. 98 (11), 2397–2410.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resour. Res. 52 (10), 7548–7555.

Jones, C.D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J.G., 2016. C4MIP–The coupled climate–carbon cycle model intercomparison project: experimental protocol for CMIP6. Geosci. Model Dev. (GMD) 9 (8), 2853–2880.

Karmalkar, A.V., Bradley, R.S., Diaz, H.F., 2011. Climate change in Central America and Mexico: regional climate model validation and climate change projections. Clim. Dyn. 37 (3–4), 605.

Kennedy, A.D., Dong, X., Xi, B., Xie, S., Zhang, Y., Chen, J., 2011. A comparison of MERRA and NARR reanalyses with the DOE ARM SGP data. J. Clim. 24 (17), 4541–4557.

Kouzes, R.T., Anderson, G.A., Elbert, S.T., Gorton, I., Gracio, D.K., 2009. The changing paradigm of data-intensive computing. Computer 42 (1), 26–34.

Lopez, D., Manogaran, G., 2016. Big Data Architecture for Climate Change and Disease Dynamics. The Human Element of Big Data: Issues, Analytics, and Performance, pp. 301–331.

Lupo, A., Kininmonth, W., 2013. Global climate models and their limitations. Climate change reconsidered II: Physical science 9, 148.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E.H., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87 (3), 343–360.

Mizielinski, M., Roberts, M., Vidale, P., Schiemann, R., Demory, M.-E., Strachan, J., Edwards, T., Stephens, A., Lawrence, B., Pritchard, M., 2014. High resolution global climate modelling; the UPSCALE project, a large simulation campaign. Geosci. Model Dev. (GMD) 7 (4), 1629–1640.

Ng, E., Evans, K., Caldwell, P., Hoffman, F., Jackson, C., Van Dam, K., Leung, R., Martin, D., Ostrouchov, G., Tuminaro, R., 2017. Advances in Cross-Cutting Ideas for Computational Climate Science.

Pedersen, C.A., Winther, J.-G., 2005. Intercomparison and validation of snow albedo parameterization schemes in climate models. Clim. Dyn. 25 (4), 351–362.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., 2007. Climate Models and Their Evaluation, Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR). Cambridge University Press, pp. 589–662.

Reed, D.A., Dongarra, J., 2015. Exascale computing and big data. Commun. ACM 58 (7), 56–68.

Rood, R., 2011. The Validation of Climate Models: the Development of Essential Practice. AGU Fall Meeting Abstracts.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., 2010. The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. 91 (8), 1015–1058.

Schnase, J.L., Lee, T.J., Mattmann, C.A., Lynnes, C.S., Cinquini, L., Ramirez, P.M., Hart, A.F., Williams, D.N., Waliser, D., Rinsland, P., 2016. Big data challenges in climate science: improving the next-generation cyberinfrastructure. IEEE Geoscience and Remote Sensing Magazine 4 (3), 10–22.

Schnase, J., et al., 2016. Big Data Challenges in Climate Science: Improving the next-generation cyberinfrastructure. IEEE Geoscience and Remote Sensing Magazine 4 (3), 10–22. https://doi.org/10.1109/MGRS.2015.2514192.

Schnase, J.L., Duffy, D.Q., Tamkin, G.S., Nadeau, D., Thompson, J.H., Grieg, C.M., McInerney, M.A., Webster, W.P., 2017. MERRA analytic services: meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. Comput. Environ. Urban Syst. 61, 198–211.
Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K.-l., Sorooshian, S., 2013. Computational Earth science: big data transformed into insight. Eos, Transactions American Geophysical Union 94 (32), 277–278.
Smith, T.L., Benjamin, S.G., Brown, J.M., Weygandt, S., Smirnova, T., Schwartz, B., 2008. 11.1 CONVECTION FORECASTS FROM THE HOURLY UPDATED, 3-KM HIGH RESOLUTION RAPID REFRESH (HRRR) MODEL.
Sotillo, M., Ratsimandresy, A., Carretero, J., Bentamy, A., Valero, F., González-Rouco, F., 2005. A high-resolution 44-year atmospheric hindcast for the Mediterranean Basin: contribution to the regional improvement of global reanalysis. Clim. Dyn. 25 (2–3), 219–236.
Stewart, C.A., Cockerill, T.M., Foster, I., Hancock, D., Merchant, N., Skidmore, E., Stanzione, D., Taylor, J., Tuecke, S., Turner, G., 2015. Jetstream: a self-provisioned, scalable science and engineering cloud environment. In: Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure. ACM, p. 29.
Sun, Z., et al., 2012. A Task Ontology Driven Approach for Live Geoprocessing in a Service-Oriented Environment. Transactions in GIS 867–884. https://doi.org/10.1111/j.1467-9671.2012.01364.x.
Sun, Z., et al., 2013. GeoPWProv: interleaving map and faceted metadata for provenance visualization and navigation. IEEE Transactions on Geoscience and Remote Sensing 51 (11), 5131–5136. https://doi.org/10.1109/TGRS.2013.2248064.
Sun, Z., et al., 2016. Realizing parameterless automatic classification of remote sensing imagery using ontology engineering and cyberinfrastructure techniques. Computers & Geosciences 94, 56–67. https://doi.org/10.1016/j.cageo.2016.06.004.
Sun, Z., et al., 2019. SUIS: Simplify the use of geospatial web services in environmentalmodelling. Environmental Modelling & Software 119, 228–241. https://doi.org/10.1016/j.envsoft.2019.06.005.
Sun, Z., Di, L., 2019. Geoweaver: a Web-Based Prototype System for Managing Compound Geospatial Workflows of Large-Scale Distributed Deep Networks.
Sun, Z., Yue, P., Di, L., 2012. GeoPWTManager: a task-oriented web geoprocessing system. Comput. Geosci. 47 (0), 34–45.
Sun, Z., Peng, C., Deng, M., Chen, A., Yue, P., Fang, H., Di, L., 2014. Automation of customized and near-real-time vegetation condition index generation through cyberinfrastructure-based geoprocessing workflows. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (11), 4512–4522.
Sun, Z., Di, L., Huang, H., Wu, X., Tong, D.Q., Zhang, C., Virgei, C., Fang, H., Yu, E., Tan, X., 2017. CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models. Earth Sci. India 11 (1), 1–17.
Tsipis, K., 2019. 3Q: Machine Learning and Climate Modeling. MIT News.
Webb, M.J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C.S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J.E., 2017. The cloud feedback model intercomparison project (CFMIP) contribution to CMIP6. Geosci. Model Dev. (GMD) 359–384, 2017.
Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y., Toth, Z., 2008. Ensemble data assimilation with the NCEP global forecast system. Mon. Weather Rev. 136 (2), 463–482.
Yue, P., et al., 2010. GeoPW: Laying blocks for the geospatial processing web. Transactions in GIS 755–772. https://doi.org/10.1111/j.1467-9671.2010.01232.x.
Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E.P., Hegerl, G., Robock, A., Pausata, F.S., Ball, W.T., Bauer, S.E., 2016. The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): experimental design and forcing input data for CMIP6. Geosci. Model Dev. (GMD) 9 (8), 2701–2719.
Zhang, C., et al., 2019. Exploring cloud-based Web Processing Service: A case study on the implementation of CMAQ as a Service. Environmental Modelling & Software 113, 29–41. https://doi.org/10.1016/j.envsoft.2018.11.019.
Council, NSF Cyberinfrastructure, 2006. NSF's cyberinfrastructure vision for 21st century discovery. National Science Foundation. July 20. http://climateknowledge.org/openclimate/docref/NSF_Cyberinfrastructure_Vision_2007.pdf.
Revolutionizing science and engineering through cyberinfrastructure, 2003.