INSTITUTE FOR DEFENSE ANALYSES

Climate Data Standards and Tool Usability Assessments for ESTCP Climate Resilience

Jennifer L. Bewley
Gifford J. Wong
Isaac S. Chappell
Zachary M. Berzolla

March 2023

IDA Document D-33445
Log: H 23-000100

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305-3086

About This Publication

This work was conducted by the IDA Systems and Analyses Center under contract HQ0034-19-D-0001, project AM-2-5090, "Technical Assessment of New Methods for Enhancing DoD Installation Climate Resilience," for the Executive Director, Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP), under the Deputy Assistant Secretary of Defense, Environment and Energy Resiliency. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

For More Information

Jennifer L. Bewley, Project Leader
jbewley@ida.org, 703-845-2390

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org, 703-578-2800

Copyright Notice

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-33445

Climate Data Standards and Tool Usability Assessments for ESTCP Climate Resilience

Jennifer L. Bewley
Gifford J. Wong
Isaac S. Chappell
Zachary M. Berzolla

Executive Summary

The Strategic Environmental Research and Development Program (SERDP) and the Environmental Security Technology Certification Program (ESTCP) are Department of Defense (DoD) environmental research programs with the shared mission to improve the DoD's environmental performance, reduce costs, and enhance and sustain mission capabilities. SERDP and ESTCP accomplish this mission using the latest scientific and technological advancements. With the effects of a changing climate and extreme weather events on DoD infrastructure threatening its ability to execute its mission, the Biden administration is requiring climate considerations to be an integral part of all DoD strategy, planning, and programming activities, as stated in Executive Order 14008, Tackling the Climate Crisis at Home and Abroad.[1] To support this mission, ESTCP has developed a new focus area in Climate Resilience (CR), which will invest in innovative technologies and methodologies that address climate resilience issues facing DoD installations today and in the near future.

The goal of this paper is to further investigate the use of climate data standards and tool usability assessments to help improve proposal responses for ESTCP CR fiscal year 2024 (FY24) proposals and beyond. For the DoD to successfully adapt and be resilient to climate change, a thorough understanding of the changing climate and resulting vulnerabilities is key. Climate models are a tool that can be used to better understand the ways in which the climate is changing for a region. While global climate models (GCMs) can provide information at a large scale, more localized information is needed for installation-specific assessments. Dynamic downscaling (DD) and statistical downscaling (SD) are two promising methods for providing more localized climate information from GCMs. The biggest limitations for DD are the computational time and power requirements, as well as the storage of the outputs and the need for bias correction. While SD requires significantly less computational power than DD and less storage, some of the challenges include the assumption of stationarity in climate, the compounding nature of errors in simulations over time, and the difficulty in evaluating the accuracy of the outputs. Several publicly available authoritative data sources for climate information exist and should be leveraged by the DoD when possible. However, if currently available climate information sources are deemed inadequate, then researchers should identify the gaps and explain how

[1] Executive Order 14008, Tackling the Climate Crisis at Home and Abroad, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/executive-order-on-tackling-the-climate-crisis-at-home-and-abroad/

the proposed method will improve upon current data. If a new dataset is generated and deemed skillful, it should then be made available to others in the community through a secure website or cloud provider.

Similarly, new decision-making tools are likely to be needed for decision makers across the DoD. However, because installation-level decision makers are already overloaded with tools, it is ideal to incorporate climate information seamlessly into existing tools when possible. If it is not possible to incorporate the new climate information into an existing tool, objective usability testing (e.g., NASA Task Load Index, System Usability Scale, User Engagement Scale, Technology Acceptance Model, Computer System Usability Questionnaire, Usability Metric for User Experience) should be completed to ensure the tool generates value-add for the end users. Proposals should also clearly state who the intended end users are and the benefits of the proposed tool.

As the DoD moves towards incorporating climate-informed decision-making into its current operating procedures, there is likely going to be a strong demand signal for finer-resolution climate projections and tools that produce actionable information. To be successful at building climate resilience, the DoD must plan for the right future. This can be achieved through developing best practices for handling climate information and incorporating uncertainties into the decision-making process, while leveraging the best available science.

Contents

1. Introduction

The Strategic Environmental Research and Development Program (SERDP) and the Environmental Security Technology Certification Program (ESTCP) are Department of Defense (DoD) environmental research programs with the shared mission to improve the DoD's environmental performance, reduce costs, and enhance and sustain mission capabilities. SERDP and ESTCP accomplish this mission using the latest scientific and technological advancements. With the effects of a changing climate and extreme weather events on DoD infrastructure threatening its ability to execute its mission, the Biden administration is requiring climate considerations to be an integral part of all DoD strategy, planning, and programming activities, as stated in Executive Order 14008, Tackling the Climate Crisis at Home and Abroad.[2] To support this mission, ESTCP has developed a new focus area in Climate Resilience (CR) which will invest in innovative technologies and methodologies that address climate resilience issues facing DoD installations today and in the near future.

In early 2022, ESTCP released fiscal year 2023 (FY23) funding opportunities in the area of climate resilience and the Institute for Defense Analyses (IDA) served as a guest member of the CR Technical Committee whose responsibilities include participating in multiple levels of proposal review. A summary of results from the FY23 review process published by IDA highlighted trends in the proposals received (Bewley and Runkel 2022). The main themes for the lessons learned for future calls for proposals included creating climate data standards, ensuring tool usability, and providing a survey of available DoD tools and data to prospective principle investigators. The goal of this paper is to further investigate the themes of climate data standards (Chapter 2) and tool usability (Chapter 3) to help improve proposal responses for FY24 proposals and beyond. The primary audience for this paper is the ESTCP CR program manager and the CR Technical Committee members.

[2] Executive Order 14008, Tackling the Climate Crisis at Home and Abroad, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/executive-order-on-tackling-the-climate-crisis-at-home-and-abroad/

2. Climate Projections

Climate change is already affecting DoD operations globally and will have an increasing effect in the coming decades. Planning for these changes will enable the Services to minimize the impacts on their operations and ultimately be better equipped to carry out their missions. From more powerful and frequent extreme weather events such as wildfires, hurricanes, and heat waves, to long-term shifts in local climates such as temperature, winds, precipitation or sea level rise, the DoD must adapt its planning to be able to operate in a changed climate.

To successfully adapt and be resilient to change, a thorough understanding of the changing climate is key. Climate models are a tool that can be used to better understand the way in which the climate is changing for a specific region. For example, climate models can help installation managers better understand where there might be long-term reductions in precipitation that can lead to droughts that jeopardize the water supply to an installation and its surrounding community (Department of the Navy 2022). They can also help identify the risks of extreme precipitation that can lead to flooding events, causing billions in damage and interrupting an installation's ability to carry out its mission effectively (Department of the Navy 2022). Furthermore, climate models can help understand long-term increases in temperature that can affect the DoD's ability to supply remote arctic bases. Extreme heat today or a long-term upward temperature trend can lead to excess energy consumption to air conditioned spaces as well as limit the ability of an installation to conduct outdoor training and maintenance (Garfin 2017). Additionally, sea level rise might render low-lying airfields unusable and lead to instability within allied nations, especially in U.S. Indo-Pacific Command (USINDOPACOM) (Department of Defense 2022). Climate models can be used to identify where these changes might occur and help create a proactive response.

The risks of climate change are not new; the U.S. government already requires new buildings to be "responsive to any government provided projections of climate change and determination of acceptable risk," and this principle needs to be applied across existing DoD installations too (Department of Defense 2020). The key challenge, however, is providing those projections at a scale and certainty that provides a path to action for the DoD and other stakeholders. Climate modeling techniques and tools have been rapidly developing to meet this need, but efforts have not always been aligned to create a reliable, verifiable, and reproducible modeling approach that can be scaled to users across the Department. In this section, we identify the current state of climate modeling efforts, their applicability to DoD installations, and provide recommendations on how to best facilitate solicitations that involve climate modeling going forward to ensure it meets the aforementioned criteria.

A. Global Climate Models

1. Definition

The National Oceanic and Atmospheric Administration (NOAA) defines a global climate model (GCM) as "a complex mathematical representation of the major climate system components (atmosphere, land surface, ocean, and sea ice), and their interactions" (NOAA Geophysical Fluid Dynamics Laboratory n.d.). GCMs divide the globe into a three-dimensional grid of cells that cover the surface as well as extend into the lower atmosphere. The size of the grid cells is dependent upon the power of the computer available to solve these equations. Finer, more detailed resolution requires a bigger and faster computer—and likely more time as well—to perform the simulation.

There are two types of processes within climate models used today: simulated and parameterized. Simulated processes are larger than grid scale and their dynamics, which are governed by physical principles including conservation of mass, conservation of energy, and conservation of momentum of air in three directions, can be described as equations. By solving the equations, climate models can simulate climate variables such as temperature, pressure, or water vapor content in three dimensions—including how different parts exchange fluxes of heat, water, and momentum—and in time.

Other variables that affect the Earth's climate may be complex and smaller than grid scale, which means they cannot be physically represented in the model. GCMs will use parameterizations (i.e., simplified equations) to represent these processes. While guided by fundamental physical principles, these simplifications also make use of observational data (e.g., cloud and aerosol composition). However, parameterizations are one of the main sources of uncertainty in climate models (Jakob 2010).

2. Benefits and Limitations of GCM for DoD Installations

The DoD's 500+ installations around the world range from small single building properties (e.g., National Guard armories) to sprawling test ranges that cover hundreds of square miles and several geographically dispersed sub-sites (Department of Defense 2018). GCMs can help predict large-scale change in climate trends that might affect individual installations, such as changing weather patterns that result in whole regions being in drought or sea level rise threatening low-lying installations. While these big picture trends are important for some planning purposes, more detailed information is needed to best plan for resilience[3] in a changing climate. Even the highest resolution GCMs—which use 50 km x 50 km grids—will only have a few grid areas that are relevant to an individual installation (Kotamarthi et al. 2017). This means extreme weather events such as extreme precipitation and weather patterns that are affected by small-scale geographic changes such as mountains or marine boundaries will often not be captured by a GCM (Trzaska and Schnarr 2014). To better simulate climate change at the installation level, finer scale climate models are needed.

GCMs are important tools that allow the scientific community to improve the understanding and prediction of atmosphere, ocean, and climate behavior. GCMs provide researchers a way to explore climate sensitivities with experiments that cannot be physically conducted. GCMs can also be used to detect statistically significant changes in climate and attribute observed changes to physically plausible causes (e.g., the role of anthropogenic forcing in 20th century climate change). While powerful, GCMs usually have resolutions of hundreds of kilometers, which means they can only show climate trends on a very large scale. As such, GCMs are mainly used by the DoD to inform long-term trend data and as inputs for downscaling efforts.

B. Regional Climate Models and Statistical Downscaling

Identifying climate impacts that are relevant to DoD installations requires outputs at scales much smaller than the 50–250 km^2 resolution of most GCMs. Downscaling was developed in the broader field to provide locally relevant information from GCM outputs. There are two main types of downscaling: dynamic downscaling (DD), often called Regional Climate Models (RCMs), and statistical downscaling (SD), often called Empirical Statistical Downscaling (ESD) (Kotamarthi et al. 2016). To date, no one approach has been identified as applicable for all use cases. There are benefits and limitations depending on the scenario being investigated and the available data. The following sections describe each type of downscaling and the associated benefits and limitations. Appendix A provides example SERDP-ESTCP projects that used dynamic and statistical downscaling methods and summarizes the main findings of each and Appendix B provides example DoD tools that use climate projections.

1. RCMs (Dynamic Downscaling)

a. Definition

DD runs simulations using many of the same physical processes used in GCMs but focuses on smaller sections of the world. To do this, boundary conditions appropriate for

[3] Resilience is the ability to anticipate, prepare for, and adapt to changing conditions and withstand, respond to, and recover rapidly from disruptions (Department of Defense 2016).

the area of interest are defined, usually using the GCM outputs (Lanzante et al. 2018). Then, the climate models are run with smaller grid spacing to simulate climate and weather patterns that happen at more local levels such as precipitation and temperature (Demissie 2019). Models can be tested using the "perfect model" approach where the boundary conditions from the GCMs are instead provided by using measured weather data and the simulations run for a historical time period for comparison against measured data (Karlicky 2013). This model check is a necessary but not sufficient conditions for validating an RCM model. One of the most commonly used RCMs in the United States is the Weather Research and Forecasting (WRF)[4] model developed collaboratively by the National Center for Atmospheric Research (NCAR), NOAA (represented by the National Centers for Environmental Prediction (NCEP) and the Earth System Research Laboratory), the U.S. Air Force, the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA). RCMs such as the WRF are focused on accounting for mesoscale processes such as the effects of snow on ground albedo, geographic changes in precipitation, and other factors that are smoothed out by the GCM's large grid areas (Salathé et al. 2010).

b. Benefits and limitations for DoD Installations

The WRF has been shown to be able to distinguish between microclimatic differences in temperature between water and inland weather stations in historical tests, something that GCMs and many ESD models have failed to do and could be particularly valuable for coastal installations (Alessi and DeGaetano 2021). The biggest limitation of DD for DoD installations is the computational time and power required, as well as storage for the outputs. Running an RCM typically involves whole swaths, if not all of the United States, and as a result many terabytes of data are produced and hundreds of thousands of CPU hours are consumed (Kotamarthi et al. 2016). Consequently, it is prohibitively expensive to carry out DD for all 100 GCMs in the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project version 6 (CMIP6) across several different climate scenarios at local resolutions. Thus, oftentimes only one or two climate change scenarios are simulated for any given study instead of all of them, which narrows the potential climate futures available for end users (Kotamarthi et al. 2016). One of the challenges with RCMs is that by their nature they must be constrained within boundary conditions (Salathé et al. 2010). The boundary conditions are usually taken from the parent GCM but are hard to validate and not always correct (Kotamarthi et al. 2016). Furthermore, RCMs still need bias correction (using the difference between the historical RCM and observations to correct the historical and future RCM projections) to ensure that the models do not violate the boundary conditions (Kotamarthi et al. 2016).

[4] https://www.mmm.ucar.edu/models/wrf

2. Statistical Downscaling

a. Definition

Statistical downscaling (SD) is an empirical approach to downscaling climate models that uses two datasets: real-world historical observations and the outputs of a GCM or RCM. Most modern SD is essentially a machine-learning method to train a model on the connection between historical climate data and the GCM or RCM predictions for that time. The trained model can then be used to project future climate at the finer scale. One of the biggest pitfalls with training a model only on historical data is that historical relationships may not be valid in a changed climate (Lanzante et al. 2018). To resolve some of the issues with using historical data to predict future climate, bias-correction or cross-validation is often used (Lanzante et al. 2018). These correction methods go by many names including asynchronous regional regression model (ARRM), bias-corrected spatial disaggregation (BCSD), bias-corrected constructed analog (BCCA), and multivariate adaptive constructed analog (MACA) (Lanzante et al. 2018). While biases can be corrected, the underlying question of whether stationarity is applicable remains. SD assumes stationarity in climate (i.e., past weather and patterns will continue to apply in the future) (Kotamarthi et al. 2016). This assumption is critical because SD is built around the use of historical data to project future changes. If future climate falls outside of the realm of previous experiences, it may be hard for SD to predict (Lanzante et al. 2018).

b. Benefits and Limitations for DoD Installations

Stationarity is particularly an issue for studies trying to evaluate the impacts of extreme weather. This means if a high or low temperature or other variable is outside of the historical record (training set), it will not be forecasted accurately from the trained model (Lanzante et al. 2018). This limitation was shown by Guentchev et al. when looking at future climate projections (2016). In their study, days over 35°C are underestimated compared to overall counts in larger GCM models (Guentchev et al. 2016). The aforementioned bias correction can help with this to some extent, but still results in an under prediction which could have lasting consequences, particularly if climate adaptation measures are not resilient enough to protect DoD installations.

SD is the method of choice for most researchers who need to understand a problem at a level of detail a GCM cannot provide because it is computationally much easier and cheaper to carry out than DD. SD modeling at 1 km^2 resolution can require 100 times less computational power than DD and requires a lot less storage space and input data (Alessi and DeGaetano 2021). Using SD allows for flexibility in the research approach, as many different scenarios can be studied even with limited resources. Whether the use of SD is appropriate for the variables being studied, however, will depend on the specifics of the scenario.

Marine boundary layers are often a challenge in SD models (and are outright ignored in GCMs) because of the complicated interactions of several different earth systems (Feldman 2019). Oceans provide temperature regulation to the coasts and thus in models the warming can be smoothed out, which may under or overstate warming for specific areas (Lanzante et al. 2018). This issue also presents itself with mountainous areas as the GCM will not capture the behavior well and the historical data are also likely lacking, so the predicted behavior will likely be inaccurate (Lanzante et al. 2018).

One of the biggest challenges with SD is the compounding of errors in simulations out to the end of the 21st century. If the model is wrong in 2050, it gets worse by 2100 (Feldman 2019). Furthermore, SD has inherent data smoothing in it. This has implications in many different geographic regions but can also affect predictions during seasonal transitions. For example, one study that used SD in the Washington, DC, area found that early season hot days are not well accounted for when compared back to historical records (Guentchev et al. 2016).

Additionally, it is hard to test whether the SD model is "correct." GCMs do not provide enough detailed information. Measured weather data only represent climate in the past and may not be representative of future climates, and DD models have their own limitations which might make them inaccurate too. In particular, testing an SD model using historical records is fraught because it favors models that are tuned to predict historical but not necessarily future climate.

Overall, the utility of SD for climate projections for DoD installations is uncertain (see Appendix A for example SD projects for the DoD). While it requires less computational power than DD, it may not work for all climate hazards and the results may become uncertain the further the projection extends into the future. In the worst case, a climate projection generated through SD could miss the important weather extremes that have major impacts on installations and the surrounding communities.

C. Authoritative Data Sources

While generating new high-resolution climate projections is time consuming and computationally intensive, there are several authoritative data sources that are publicly available for use which may limit the need for datasets generated specifically for the DoD. Before using any climate projection, regardless of the method of generation, it is important for users to thoroughly understand the limitations of each dataset and how uncertainties may impact the output of decision support tools or investment decisions.

1. GCMs

There are more than two dozen scientific institutions that develop GCMs, including the UK Met Office Hadley Centre "HadGEM3" family of models, the NOAA Geophysical

Fluid Dynamics Laboratory "GFDL ESM2M" model, and the NCAR "Community Earth System Model" (CESM). Having multiple GCMs to compare simulations and results can help inform which parameterizations the community has confidence in, where GCMs agree, and where the community may be less certain, and where GCMs disagree (Carbon Brief 2018).

CMIP is a framework for climate model experiments—running GCMs with a common set of input parameters and sharing the output—that allows scientists to analyze, validate, and improve their climate models in a systematic way (Copernicus Programme 2021). The most recent coordinated effort, CMIP6, will consist of model "runs" from over 100 models being produced across 49 different modeling groups (Hausfather 2019). Importantly, CMIP6 is informing the latest Intergovernmental Panel on Climate Change (IPCC) assessment report (AR6) (Arias et al. 2021). NOAA provides an easy viewer to see projections at varying geographic scales based on the CMIP6 dataset.[5] Previous versions of CMIP (CMIP3 and CMIP5) models are available from various sources, including Lawrence Livermore National Lab.[6]

2. RCMs and ESDs

The North American Regional Climate Change Assessment Program (NARCCAP) was an effort led by the NCAR to create several North American RCMs at a 50 km^2 resolution, based on several different GCMs (Mearns et al. 2009). The dataset includes several different runs, was last updated in 2014, and was succeeded by the Coordinated Regional Climate Downscaling Experiment (CORDEX).[7] CORDEX is a global effort to evaluate the regional climate modeling effort and has traditionally been the home for DD efforts, but it is also incorporating some ESD efforts as well.[8] CORDEX has sub-areas for various regions, including North America.[9] All currently available models and their data are available through the CORDEX website.

The High Resolution Model Intercomparison Project (HighResMIP) is an outgrowth of this effort, with investigations into how to achieve native high-resolution (~50 km^2) projections directly out of CMIP6 with no downscaling required (Haarsma et al. 2016). This effort grew out of concern that downscaling ignores some high-level atmospheric circulations that can only be represented by GCM. The resulting dataset is publicly

5 https://psl.noaa.gov/ipcc/cmip6/

6 https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/#Projections:%20Complete%20Archives

7 https://cordex.org/ for more information.

8 https://cordex.org/wp-content/uploads/2017/08/CORDEX_ESD_Reference_Document.pdf

9 https://na-cordex.org/

available through Lawrence Livermore National Laboratory.[10] The newest version of the protocol for HighResMIP will be released in 2023.[11]

The WRF model is another commonly used RCM. NCAR publishes the source code for the WRF model, and it is used in many of the DD studies.[12] The U.S. Geological Survey Geo Data Portal also has several climate datasets covering some or all of the contiguous United States.[13] This portal contains both dynamically downscaled data as well as statistically downscaled models that are freely accessible (Blodgett et al. 2011). Some datasets are bias-corrected and some are raw data. A dataset created from 20 CMIP5 GCMs and downscaled with the MACA SD method is available in the Climate Futures Toolbox, developed with the National Park Service's Climate Change Response Program.[14,15] Recently, the Climate Mapping for Resilience and Adaptation (CMRA)[16] website was launched as part of an interagency partnership under the leadership of the U.S. Global Change Research Program (USGCRP). This website houses climate hazard information, an assessment tool, and several climate datasets including localized constructed analogs (LOCA)-downscaled CMIP5 models for variables used in the Fourth National Climate Assessment.

The Parameter-elevation Relationships on Independent Slopes Model (PRISM) Climate Group maintains a U.S. Department of Agriculture-supported database of climate observations (Daly et al. 2008). They have observations down to an 800 m^2 grid, with 4 km^2 freely available to all, with some data ranging back to 1895.[17] This dataset provides a good baseline for testing all models.

Before researchers embark on creating new climate projections for assessing the vulnerability of DoD installations, one should evaluate if any of the publicly available datasets mentioned above fill the desired need. If the currently available data are deemed inadequate, then researchers should explain where the gaps are and how they could be filled. Once a new dataset is generated and skill is assessed, it should be made available for others in the community to access through a secure (e.g., Common Access Card (CAC)-enabled) website to ensure consistency across installations and to conserve resources. It is also important to note that an ESTCP solicitation for demonstration projects that assess statistical and dynamical downscaling of climate-related data applied to CMIP6 was

[10] https://esgf-node.llnl.gov/search/cmip6/

[11] https://highresmip.org/

[12] https://www.mmm.ucar.edu/models/wrf

[13] https://labs.waterdata.usgs.gov/gdp_web/

[14] https://www.earthdatascience.org/cft/index.html

[15] https://climate.northwestknowledge.net/MACA/

[16] https://resilience.climate.gov/

[17] https://prism.oregonstate.edu/

released for FY22.[18] The resulting reports from projects funded under this solicitation may provide additional useful insight into the best approaches for DD and SD methods for DoD installations.

---

[18] https://www.serdp-estcp.org/workingwithus/callforproposal?Id=5918bad7-4f5e-45c8-b465-f585d559856a&slug=climate-model-comparative-assessment-for-dod-infrastructure-applications

3. Tool Usability

Planning for climate resilience at DoD installations requires tools for installation-level decision makers. These tools need to provide actionable information derived from complicated climate projections that decision makers and their staff can use to identify the climate hazards important to their installation, and invest in the appropriate climate adaptations. Many ESTCP proposals include tool development as a component of the work. These tools will never be effective toward driving change at installations if they are not used by their intended audience. There is already a perception that there are too many tools that are too specialized for any one staff member at an individual installation to effectively leverage. It is thus critical that new tools are only developed when the functionality cannot be integrated into existing workflows or tools. Furthermore, most previous proposals for new tool development have included qualitative ways of collecting feedback on the tool, when quantitative metrics may be more useful and objective. An objective evaluation is key to ensuring that the tools being used provide actionable value-adds to the personnel who are going to be on the front lines of adaptation at the individual installations. In the following section, we outline quantitative usability scales that could help with objective evaluations of new tools. We then present some key takeaways on tool usability in the findings section.

A. Review of Usability Scales

1. Definition of Usability

Usability is defined by ISO-9241-11 as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (9241-11). In the DoD context, this means personnel at the installation and command level must be able to clearly identify the actions they must take to adapt to climate hazards and ensure resilient operation of the DoD's myriad of installations. To evaluate usability, many different scales and indexes have been developed; a few key ones are catalogued below.

2. Likert Scale

Before describing some of the various measures of usability for computer programs, it is worth explaining a very common scale used by human factors experts. The Likert Scale is a measurement scale commonly used for questionnaires to determine the strength or intensity of a person's feeling or attitude about a survey item. It uses either a 5-answer or

7-answer qualitative scale rather than ranges across a spectrum. Most people are familiar with the 5-answer Likert Scale of Strongly Agree, Somewhat Agree, Neutral, Somewhat Disagree, and Strongly Disagree. The basis of the Likert Scale is the assumption that the strength of a person's feelings towards a subject or question is linear between extremes. Numbers can be associated with the values with the understanding that they should be normalized and interpreted. The Likert Scale can also be used to gauge other measures besides agreement such as frequency, importance, likelihood, and quality (McLeod 2019).

3. NASA TLX (NASA Task Load Index)

The NASA TLX questionnaire (see example in Figure 1) is "a subjective workload assessment tool which allows users to perform subjective workload assessments of operator(s) working with various human-machine interface (HMI) systems" (So 2020). Developed by NASA over multiple years in the 1980s by the Human Performance Group at NASA's Ames Research Center, it was designed to measure how well an HMI system is working for the user. It has six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. It uses a 7-answer scale further broken down into high, medium, and low subscales such that there are 21 gradations for each class. An acceptable score is between 39 and 61 (Favre-Félix et al. 2022). NASA has a pen and paper version as well as a mobile phone app available through the Apple App Store.

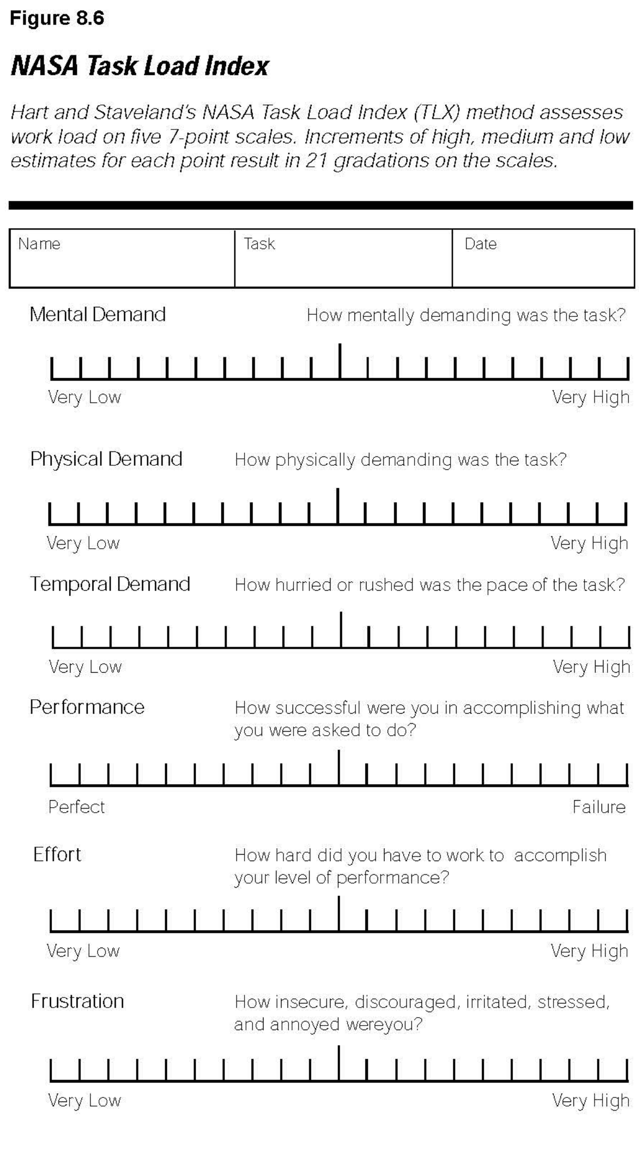

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 1. Example of the NASA TLX Questionnaire.

4. SUS (System Usability Scale)

The System Usability Scale (or SUS, as it is more commonly known,) is a common usability scale used for a large number of systems. It is a 10-question/item form (see Figure 2) using a 5-answer Likert Scale focusing on agreement. It was developed by the Digital Equipment Corporation in the late 1980s (Brooke 1996). The SUS looks at three main areas of usability:

- Effectiveness (can you complete the objective)
- Efficiency (how much effort is required)

- Satisfaction (how do you feel about the experience)

A system is considered acceptable if the total score is above 70 and unacceptable if it is below 50. If the score is greater than 85, it is considered excellent. The average score for the SUS is 68 so anything above that is considered above average (Sauro 2018).

System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Figure 2. Example SUS Form.

5. UES (User Engagement Scale)

The UES (User Engagement Scale) measures user engagement (UE) for human-computer interactions (HCI) (O'Brien and Toms 2010). The original questionnaire had 31 items that looked at 6 different dimensions of engagement: aesthetic appeal, focused attention, novelty, perceived usability, felt involvement, and endurability. A shorter questionnaire, called the Short Form (UES-SF), was developed in 2018 to provide similar information from fewer items (O'Brien et al. 2018). The UES data are analyzed statistically using a procedure developed by the original scientists. Figure 3 and Figure 4 show examples for both the regular and short versions.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	2	3	4	5

User Engagement Scale Long Form (UES-LF).

FA.1	I lost myself in this experience.
FA.2	I was so involved in this experience that I lost track of time.
FA.3	I blocked out things around me when I was using Application X.
FA.4	When I was using Application X, I lost track of the world around me.
FA.5	The time I spent using Application X just slipped away.
FA.6	I was absorbed in this experience.
FA.7	During this experience I let myself go.
PU.1	I felt frustrated while using this Application X.
PU.2	I found this Application X confusing to use.
PU.3	I felt annoyed while using Application X.
PU.4	I felt discouraged while using this Application X.
PU.5	Using this Application X was taxing
PU.6	This experience was demanding.
PU.7	I felt in control while using this Application X.
PU.8	I could not do some of the things I needed to do while using Application X.
AE.1	This Application X was attractive
AE.2	This Application X was aesthetically appealing
AE.3	I liked the graphics and images of Application X.
AE.4	Application X appealed to be visual senses.
AE.5	The screen layout of Application X was visually pleasing.
RW.1	Using Application X was worthwhile
RW.2	I consider my experience a success.
RW.3	This experience did not work out the way I had planned.
RW.4	My experience was rewarding.
RW.5	I would recommend Application X to my family and friends
RW.6	I continued to use Application X out of curiosity.
RW.7	The content of Application X incited my curiosity.
RW.8	I was really drawn into this experience.
RW.9	I felt involved in this experience.
RW.10	This experience was fun.

Figure 3. Example UES-LF.

FA-S.1	I lost myself in this experience.
FA-S.2	The time I spent using Application X just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated while using this Application X.
PU-S.2	I found this Application X confusing to use.
PU-S.3	Using this Application X was taxing.
AE-S.1	This Application X was attractive.
AE-S.2	This Application X was aesthetically appealing.
AE-S.3	This Application X appealed to my senses.
RW-S.1	Using Application X was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

Figure 4. Example UES-SF.

6. TAM (Technology Acceptance Model)

The TAM (Technology Acceptance Model) measures user acceptance of an application. By looking at two factors, perceived usefulness and perceived ease of use, the TAM relates these factors to the user's attitude towards the application and "intensity" of application usage. A questionnaire (see Figure 5) is used to measure variables such as EOU (Ease of Use), USEF (Perceived Usefulness), ATT (attitude toward using), and USE (intensity of usage). The questionnaire uses a 7-point Likert Scale on agreement. Statistical tests are used to test specific hypotheses about the relationship between the variables (Davis 1993).

USER ACCEPTANCE 487

PERCEIVED EASE OF USE OF ELECTRONIC MAIL

	Strongly agree			Neutral			Strongly disagree
1. I find the electronic mail system cumbersome to use.	1	2	3	4	5	6	7
2. Learning to operate the electronic mail system is easy for me.	1	2	3	4	5	6	7
3. Interacting with the electronic mail system is often frustrating.	1	2	3	4	5	6	7
4. I find it easy to get the electronic mail system to do what I want it to do.	1	2	3	4	5	6	7
5. The electronic mail system is rigid and inflexible to interact with.	1	2	3	4	5	6	7
6. It is easy for me to remember how to perform tasks using the electronic mail system.	1	2	3	4	5	6	7
7. Interacting with the electronic mail system requires a lot of mental effort.	1	2	3	4	5	6	7
8. My interaction with the electronic mail system is clear and understandable.	1	2	3	4	5	6	7
9. I find it takes a lot of effort to become skillful at using electronic mail.	1	2	3	4	5	6	7
10. Overall, I find the electronic mail system easy to use.	1	2	3	4	5	6	7

Figure 5. Example TAM Questionnaire.

7. CSUQ (Computer System Usability Questionnaire)

The CSUQ (Computer System Usability Questionnaire) is a tool to evaluate the usability of an application. It is related to the PSSUQ (Post Study System Usability Questionnaire), only differing in that the PSSUQ's questions are past tense and the CSUQ items are present tense. The questionnaire (see Figure 6) has 16 questions/items using a 7-answer Likert Scale on agreement with an additional "N/A" answer for each question. The 16 questions are grouped into three subareas: System Usefulness, Information Quality, Interface Quality. It was developed internally by IBM in the 1980s and published externally in the 1990s (Lewis 2018). The CSUQ is scored similar to the SUS with higher scores being better, but exact thresholds for rating usability are not provided (Lewis 2018).

The Computer System Usability Questionnaire
Version 3

		Strongly Agree 1	2	3	4	5	6	Strongly Disagree 7	NA
1	Overall, I am satisfied with how easy it is to use this system.	O	O	O	O	O	O	O	O
2	It is simple to use this system.	O	O	O	O	O	O	O	O
3	I am able to complete my work quickly using this system.	O	O	O	O	O	O	O	O
4	I feel comfortable using this system.	O	O	O	O	O	O	O	O
5	It was easy to learn to use this system	O	O	O	O	O	O	O	O
6	I believe I became productive quickly using this system.	O	O	O	O	O	O	O	O
7	The system gives error messages that clearly tell me how to fix problems.	O	O	O	O	O	O	O	O
8	Whenever I make a mistake using the system, I recover easily and quickly.	O	O	O	O	O	O	O	O
9	The information (such as online help, on-screen messages and other documentation) provided with this system is clear.	O	O	O	O	O	O	O	O
10	It is easy to find the information I needed.	O	O	O	O	O	O	O	O
11	The information provided with the system is effective in helping me complete my work.	O	O	O	O	O	O	O	O
12	The organization of information on the system screens is clear.	O	O	O	O	O	O	O	O
13	The interface* of this system is pleasant.	O	O	O	O	O	O	O	O
14	I like using the interface of this system.	O	O	O	O	O	O	O	O
15	This system has all the functions and capabilities I expect it to have.	O	O	O	O	O	O	O	O
16	Overall, I am satisfied with this system.	O	O	O	O	O	O	O	O

** The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).*

Figure 6. Example CSUQ.

8. UMUX (Usability Metric for User Experience)

The UMUX (Usability Metric for User Experience) was developed in 2010 as a tool for assessing the perceived usability of a program or application. It is a 4-item questionnaire (see Figure 7) using a 7-point Likert Scale on agreement. It looks at the same subareas as the SUS (effectiveness, efficiency, and satisfaction) (Lewis 2018). Similar to the SUS, a higher score is regarded as a better user experience, but exact numbers are not provided (Lewis 2018).

	The Usability Metric for User Experience Version 1		Strongly Disagree 1	2	3	4	5	6	Strongly Agree 7
1	This system's capabilities meet my requirements.		O	O	O	O	O	O	O
2	Using this system is a frustrating experience.		O	O	O	O	O	O	O
3	This system is easy to use.		O	O	O	O	O	O	O
4	I have to spend too much time correcting things with this system.		O	O	O	O	O	O	O

Figure 7. Example UMUX Questionnaire.

4. Findings

When reviewing proposals for the ESTCP CR FY23 funding process, it was noted that several projects recommended using existing, yet mostly different, climate datasets to support their analysis, while others proposed to generate new downscaled datasets which are computationally intensive and costly. Several other proposals were related to tool development or required installation-specific data (e.g., energy and water consumption). In some cases, proposals suggested developing new tools that did not have a clearly defined end user in mind and proposed qualitative metrics for evaluating the success of the tool. This paper explored these topics further and the results are described below.

A. Climate Information for DoD Installations

Climate information is needed to enable climate-informed decision-making at the installation level including investments in adaptation measures. As mentioned previously, most climate datasets and models used for simulation are open-source and publicly available. Principle Investigators (PIs) should be encouraged to use these authoritative data sources. ESTCP could encourage the use of such datasets by providing a list of recommended climate data sources or repositories. If a PI proposes the generation of a new climate dataset (e.g., a new downscaled dataset), their proposal should describe why the currently available data are inadequate for their study and show how the new dataset will address the identified gaps and improve on what is currently available. For any study, it is critical that the methods used are clearly documented: which GCMs were used, what downscaling techniques, which years of historical data were used for bias-correction or verification, etc. For the approach to be validated by others, the whole chain of steps to arrive at the downscaled data must be clearly presented. ESTCP and/or the research community should identify a standard process to describe the inputs and models used or make these data easily accessible to all. Creating a secure, DoD-specific repository for standardized climate change projections (e.g., CAC-protected like the DoD Climate Assessment Tool (DCAT)), will help ensure that the relevant data are widely available but secure. It is also critical that limitations of the climate information are transparent and uncertainties are incorporated into the final outputs

B. Tool Usability Assessments

While the need to incorporate climate-informed decision-making at DoD installations will likely result in the creation of new decision-making tools, it is recommended that PIs first consider how climate information can be incorporated into already existing tools and processes to reduce the workload on installation staff and ease the transition. ESTCP could encourage this behavior by providing PIs with a survey of available DoD tools. If a PI proposes the generation of a new tool, their proposal should clearly state who the intended end user is, why the tool is needed and cannot be incorporated into an already existing tool, and the benefits to the end user. In addition to outlining the design of the tool, a plan should be generated to quantitatively and objectively assess the usability of the proposed tool to ensure that the new tool provides actionable value-adds to installation personnel. Section 3.A provides a summary of eight potential methods for quantitatively assessing the usability of new tools.

5. Conclusions

The goal of this paper was to further investigate the use of climate data standards and tool usability assessments to help improve proposal responses for ESTCP CR FY24 proposals and beyond. For the DoD to successfully adapt and be resilient to climate change, a thorough understanding of the changing climate and resulting vulnerabilities is key. Climate models are a tool that can be used to better understand the way in which the climate is changing for a region. While GCMs can provide information at a large scale, more localized information is needed for installation-specific assessments. DD and SD are two promising methods for providing more localized climate information from GCMs. The biggest limitations for DD are the computational time and power requirements, as well as the storage of the outputs and the need for bias correction. While SD requires significantly less computational power than DD and less storage, some of the challenges include the assumption of stationarity in climate, the compounding nature of errors in simulations over time and the difficulty in evaluating the accuracy of the outputs. Several publicly available authoritative data sources for climate information exist and should be leveraged by the DoD when possible. However, if currently available climate information sources are deemed inadequate, then researchers should identify the gaps and explain how the proposed method will improve upon current data. If a new dataset is generated and deemed skillful, it should then be made available to others in the community through a secure website or cloud provider.

Similarly, new decision-making tools are likely to be needed for decision makers across the DoD. However, because installation-level decision makers are already overloaded with tools, it is ideal to incorporate climate information seamlessly into currently existing tools when possible. If it is not possible to incorporate the new climate information into an already existing tool, objective usability testing should be completed to ensure the tool generates value-add for the end users. Proposals should also clearly state who the intended end users are and the benefits of the proposed tool.

As the DoD moves towards incorporating climate-informed decision-making into its current operating procedures, there is likely going to be a strong demand signal for finer-resolution climate projections and tools that produce actionable information. To be successful at building climate resilience, the DoD must plan for the right future. This can be achieved through developing best practices for handling climate information and incorporating uncertainties into the decision-making process, while leveraging the best available science.

Appendix A.
Example SERDP-ESTCP Projects

This appendix briefly describes past SERDP-ESTCP research projects that used dynamic and statistical downscaling methods and summarizes the results of each.

Dynamic Downscaling (DD)

DD approaches have typically been time and computationally expensive, meaning they are only used in ESTCP solicitations that absolutely necessitate them. Project 2205, “Assessing Climate Change Impacts for DoD Installations in the Southwest U.S. During the Warm Season,” is one such project (Castro 2017). The researchers used CMIP3 and CMIP5 models and dynamically downscaled them to understand the change in warm season extreme weather events. The actual model used an ensemble of five RCMs and the CORDEX DD dataset to study the impact of monsoons on extreme precipitation across several DoD ranges in the Southwest. This study used the WRF model with Community Climate System Model (CCSM) bias-correction as their “baseline” model. They found that their RCM ensemble model and the CORDEX DD dataset performed worse than the WRF model with CCSM bias-correction. But, the downscaled CMIP data did agree with historical data for monsoon precipitation. They showed that DD is important in the monsoon context because it better captures local extremes in precipitation, which is one of the sources of the flooding hazards to DoD installations in the area.

Project 2514, “Linked Rainfall and Runoff Intensity-Duration-Frequency in the Face of Climate Change and Uncertainty,” also used DD (Demissie 2019). In this study, the authors used a WRF model downscaled to 12 km^2 to understand how the intensity-duration-frequency curves of storms are affected by climate change. Specifically, they were interested in whether stationarity is preserved and how this affects flooding events for DoD installations. They argue that several studies have shown that climate change has led to non-stationarity and thus SD methods cannot be used because the historical weather is not a good predictor of future trends (Demissie 2019). Furthermore, they argued that SD does not successfully capture the local weather patterns and storms that cause extreme precipitation. The downside with the DD approach taken in this study is that it took 800,000 core hours to simulate a model of parts of the United States at the 12 km^2 scale, a resolution that works well enough for flooding concerns across watersheds and large installations, but may not be fine enough for smaller installations (Demissie 2019). They used quantile-mapping bias correction on their DD data to reduce low-frequency, high precipitation events that were forecast by the DD WRF that are well outside the mean plus 2.563 times the standard deviation (Demissie 2019). They found that although overall climatic changes out to 2065 were only statistically significant for a handful of the 13 DoD installations they reviewed, precipitation levels had a statistically significant increase in over 50% of them (Demissie 2019). While their analysis of historical data found that extreme flooding is not always caused by extreme precipitation, as it usually has some co-factor such as snowmelt or overall ground saturation, it warrants careful consideration. Furthermore, much of the stormwater planning today uses even older data, and thus the impacts of not accounting for climate change for extreme weather planning could be substantial.

Statistical Downscaling (SD)

The IDA team reviewed two relevant ESTCP reports on SD, 2232, "Assessing Climate Change Risk, Lessons Learned from the DoD" and 18-1577, "Determining the Temporal and Spatial Scales of Nonstationary in Temperature and Precipitation Across the Continental United States for a Given Emissions Scenario" (Garfin 2017; Feldman 2019). Garfin used CMIP5 data with MACA correction (2017). They found this model worked well compared to historical data and helped them analyze the risks of high heat, wildfires, and sea level rise leading to coastal erosion for DoD installations (Garfin 2017). The report itself was focused more broadly on DoD engagement, but the need to build working relationships and trust in the modeling was a key takeaway.

Feldman (2019) used an ensemble of 10 CMIP5 models and downscaled using LOCA techniques. They compared several different downscaling techniques: WRF with DD, WRF-LOCA SD, and using historical station data with LOCA SD (Livneh method) (Feldman 2019). They found that the WRF-LOCA SD methodology damps out extreme days by finding the closest historical days and constructing a weighted average for the future (Feldman 2019). They also noted that changes in snow and coastal weather shifts violate stationarity assumptions for using SD, which concerned them (Feldman 2019). They pointed out that the stationarity assumption in SD continues to be untested and further research is needed to gain confidence in using downscaled methods.

While both of these studies focused on climate projections from CMIP5, it's important to note that there was an FY22 ESTCP solicitation for demonstration projects that assess statistical and dynamical downscaling of climate-related data applied to CMIP6. The resulting reports from projects funded under this solicitation may provide additional useful insight into the best approaches, but the question of verifying the validity of the model outputs for either DD or SD is difficult.

Appendix B.
Example DoD Applications that Use Climate Projections

DoD Climate Assessment Tool (DCAT)

DCAT[19,20] is a CAC-enabled tool that helps installations assess their exposure to natural hazards impacted by climate change across 1,391 DoD sites (Pinson et al. 2021). The eight climate hazards are: drought, coastal flooding, riverine flooding, heat, energy demand, wildfires, land degradation, and historical extreme weather (Pinson et al. 2021). The tool covers both lower and higher emission scenarios in two time periods: 2035–2065 and 2070–2100. Each installation needs to take the exposure information from the tool and understand its individual vulnerability extreme weather events in the future (Pinson et al. 2021). DCAT uses CMIP5 GCM data and RCP 8.5 and 4.5 scenarios to address the spread of possible climate outcomes to 2100 (Pinson et al. 2021). For most of the parameters studied, the GCM data were downscaled to provide more local impact information (Pinson et al. 2021). Outside the United States, temperature and precipitation data were downscaled from CMIP5 ensembles (Pinson et al. 2021) using bias-corrected spatial disaggregation (BCSD) while in the United States, a 32 CMIP-5 ensemble was downscaled with LOCA. DCAT does not provide a detailed analysis of hazards that are specific to local conditions at specific installations (Pinson et al. 2021).

DCAT is designed to provide an initial high-level assessment of exposure to different hazards at many installations (Pinson et al. 2021). Once the hazards have been identified and priorities created, each individual installation will need adaptation plans which might require more detail than DCAT can provide. To create these plans, installation managers will need actionable information at a local level that will be the focus for their climate resilience efforts. Some of these capabilities already exist for certain hazards within the DoD, for instance with coastal flooding and sea level rise using the DoD Sea Level Rise Database (DRSL). Other hazards will need further analysis using commercially available tools, such as the infrastructure changes required by an increasing energy demand, which can be identified through energy modeling of installations. Others will require climate modeling with localized downscaling, a capability that users DoD-wide do not yet have access to. The exact downscaling methods used should be chosen based on the hazards

---

19 DCAT (CONUS/AK/HI): https://dodclimate.sec.usace.army.mil/ords/f?p=118:1::::::

20 DCAT (International): https://dodclimate.sec.usace.army.mil/ords/f?p=119:1::::::

identified and the needs of the installation, and ideally, created from a standard set of scenarios where the user can just ask for the more specific data for their location and run the models using best practices. For example, if riverine flooding is the biggest hazard, an RCM that can resolve local weather patterns and account for watersheds might be most appropriate.

DoD Regional Sea Level Rise Database (DRSL)

DRSL was developed to help understand the impacts of regional sea level rise for 1,177 coastal DoD sites around the world (Hall et al. 2016). It provides sea level projections over a range of five different global sea level rise scenarios for three future time horizons (2035, 2065, 2100). For each scenario, storm event probabilities (e.g. 10 year, 50 year, 100 year flood) are provided that show the impacts of coastal flooding on top of the sea level rise (or fall if land mass uplifts). There are two versions of the tool: a publicly available version[21] and a CAC-enabled version.

[21] https://drsl.serdp-estcp.org/site

References

9241-11. 2018. "Ergonomics of human-system interaction." Available online at https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en, checked on 1/26/2023.

Alessi, Marc J., and Arthur T. DeGaetano. 2021. "A comparison of statistical and dynamical downscaling methods for short-term weather forecasts in the US Northeast." *Meteorological Applications* 28 (1). https://www.doi.org/10.1002/met.1976.

Arias, P. A., N. Bellouin, E. Coppola, et al. 2021. "Technical Summary." In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https:www.doi.org/10.1017/9781009157896.002.

Bewley, Jennifer, and Sara Runkel. 2022. *Summary of the Results of the Evaluation and Selection Process for the FY23 ESTCP Climate Resilience Broad Agency Announcement*. IDA Document NS D-33321. Alexandria, VA: Institute for Defense Analyses.

Blodgett, David L., Nathaniel L. Booth, Thomas C. Kunicki, Jordan I. Walker, and Roland J. Viger. 2011. Description and Testing of the Geo Data Portal: A Data Integration Framework and Web Processing Services for Environmental Science Collaboration. *U.S. Geologic Survey* Open-File Report 2011–1157. https://pubs.usgs.gov/of/2011/1157/.

Brooke, John. 1996. "SUS: A quick and dirty usability scale. Chapter Twenty One." In: *Usability Evaluation in Industry*, 189–194. https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke..

Carbon Brief. 2018. "Q&A: How do climate models work? Who does climate modelling around the world." January 15, 2018. https://www.carbonbrief.org/qa-how-do-climate-models-work/.

Castro, Christopher L. 2017. "Assessing Climate Change Impacts for DoD Installations in the Southwest United States During the Warm Season." In *SERDP Final Report* RC-2205.

Copernicus Programme. 2021. "Latest projections of future climate now available." March 23, 2021. https://climate.copernicus.eu/latest-projections-future-climate-now-available.

Daly, Christopher, Michael Halbleib, Joseph I. Smith, Wayne P. Gibson, Matthew K. Doggett, George H. Taylor, et al. 2008. "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United

States." *International Journal of Climatology.* 28 (15), 2031–2064. https://www.doi.org/10.1002/joc.1688.

Davis, Fred D. 1993. "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts." *International Journal of Man-Machine Studies* 38 (3), 475–487. https://www.doi.org/10.1006/imms.1993.1022.

Demissie, Yonas. 2019. "Linked Rainfall and Runoff Intensity-Duration-Frequency in the Face of Climate Change and Uncertainty." In *SERDP Final Report* RC-2514.

Department of Defense, Office of the Under Secretary of Defense (Acquisition, Technology and Logistics). 2016. *Climate Change Adaptation and Resilience*. DoD Directive 4715.21. Washington, DC: Department of Defense.

Department of Defense, Office of the Under Secretary of Defense (Acquisition, Technology and Logistics). 2018. *Climate-Related Risk to DOD Infrastructure Initial Vulnerability Assessment Survey (SLVAS) Report*. Washington, DC: Department of Defense.

Department of Defense. 2020. UFC 1-200-02 High Performance and Sustainable Building Requirements with Change 2. Washington, DC: Department of Defense.

Department of Defense, Office of the Under Secretary of Defense (Acquisition and Sustainment). 2022. *Department of Defense Climate Adaptation Plan 2022 Progress Report*. Report Submitted to National Climate Task Force and Federal Chief Sustainability Officer. Washington, DC: Department of Defense.

Department of the Navy. 2022. *Climate Action 2030*. Washington, DC: Department of the Navy. https://www.navy.mil/Portals/1/Documents/Department%20of%20the%20Navy%20Climate%20Action%202030%20220531.pdf?ver=3Q7ynB4Z0qUzlFg_2uKnYw%3d%3d&timestamp=1654016322287.

Favre-Félix, J., M. Dziadzko, C. Bauer, A. Duclos, J. J. Lehot, T. Rimmelé, and M. Lilot. 2022. "High-Fidelity Simulation to Assess Task Load Index and Performance: A Prospective Observational Study." *Turkish Journal of Anesthesiology and Reanimation* 50(4):282-287. https://doi.org/10.5152/TJAR.2022.21234. PMID: 35979975; PMCID: PMC9524413.

Feldman, Daniel. 2019. "Determining the Temporal and Spatial Scales of Nonstationarity in Temperature and Precipitation across the Continental United States for a Given Emissions Scenario." In *SERDP Final Report* RC18-1577.

Garfin, Gregg. 2017. "Climate Change Impacts and Adaptation on Southwestern DoD Facilities." In *SERDP Final Report* RC-2232.

Guentchev, Galina S., Richard B. Rood, Caspar M. Ammann, Joseph J. Barsugli, Kirstie Ebi, Veronica Berrocal, et al. 2016. "Evaluating the Appropriateness of Downscaled Climate Information for Projecting Risks of Salmonella." *International Journal of Environmental Research and Public Health* 13 (3). https://doi.org/10.3390/ijerph13030267.

Haarsma, Reindert J., Malcom J. Roberts, Pier Luigi Vidale, Catherine A. Senior, Alessio Bellucci, Qing Bao, et al. 2016. "High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6." *Geoscientific Model Development* 9 (11), 4185–4208. https://www.doi.org/10.5194/gmd-9-4185-2016.

Hall, J. A., S. Gill, J. Obeysekera, W. Sweet, K. Knuuti, and J. Marburger. 2016. *Regional Sea Level Scenarios for Coastal Risk Management: Managing the Uncertainty of Future Sea Level Change and Extreme Water Levels for Department of Defense Coastal Sites Worldwide*. U.S. Department of Defense, Strategic Environmental Research and Development Program. https://drsl.serdp-estcp.org/Docs/CARSWG_SLR.pdf.

Hausfather, Zeke. 2019. "CMIP6: the next generation of climate models explained." Carbon Brief. https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained/.

Jakob, Christian. 2010. "Accelerating Progress in Global Atmospheric Model Development through Improved Parameterizations." *Bulletin of the American Meteorological Society* 91 (7), 869–876. https://www.doi.org/10.1175/2009BAMS2898.1.

Karlicky, J. (Ed.) 2013. Regional Climate Simulations with WRF Model. 22nd Annual Conference of Doctoral Students, WDS'13 "Week of Doctoral Students 2013." Prague, Czech Republic, June 4–7, 2013.

Kotamarthi, Rao, Linda Mearns, Katharine Hayhoe, Christopher L. Castro, and Donald Wuebbles. 2016. "Use of Climate Information for Decision-Making and Impacts Research: State of our Understanding." In *SERDP Final Report*.

Kotamarthi, Rao, Jiali Wang, Zach Zoebel, Donald Wuebbles, Katharine Hayhoe, Michael Stein, and David Changon. 2017. "Climate Change Impacts at Department of Defense Installations." *Argonne National Laboratory* (ANL/EVS-17/9).

Lanzante, John R., Keith W. Dixon, Mary Jo Nath, Carolyn E. Whitlock, and Dennis Adams-Smith. 2018. "Some Pitfalls in Statistical Downscaling of Future Climate." *Bulletin of the American Meteorological Society* 99 (4), 791–803. https://www.doi.org/10.1175/BAMS-D-17-0046.1.

Lewis, James R. 2018. "Measuring Perceived Usability: The CSUQ, SUS, and UMUX." *International Journal of Human–Computer Interaction* 34 (12), 1148–1156. https://www.doi.org/10.1080/10447318.2017.1418805.

Mearns, Linda O., William Gutowski, Richard Jones, Ruby Leung, Seth McGinnis, Ana Nunes, and Yun Qian. 2009. "A Regional Climate Change Assessment Program for North America." *Eos, Transactions, American Geophysical Union* 90 (36), 311. https://doi.org/10.1029/2009EO360002.

McLeod, Saul. 2019. "Likert Scale Definition, Examples and Analysis." Simply Psychology. Updated on March 8, 2023. https://www.simplypsychology.org/likert-scale.html.

NOAA Geophysical Fluid Dynamics Laboratory. n.d. "Climate Modeling." Accessed January 23, 2023. https://www.gfdl.noaa.gov/climate-modeling/.

Pinson, A.O., K. D. White, E. E. Ritchie, H. M. Conners, and J. R. Arnold. 2021. "DoD Installation Exposure to Climate Change at Home and Abroad." Washington, DC: U.S. Army Corps of Engineers.

O'Brien, Heather L., Paul Cairns, and Mark Hall. 2018. "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form." *International Journal of Human-Computer Studies* 112, 28–39. https://www.doi.org/10.1016/j.ijhcs.2018.01.004.

O'Brien, Heather L., and Elaine G. Toms. 2010. "The development and evaluation of a survey to measure user engagement." *Journal of the American Society for Information Science and Technology* 61 (1), 50–69. https://www.doi.org/10.1002/asi.21229.

Salathé, Eric P., L. Ruby Leung, Yun Qian, and Yongxin Zhang. 2010. "Regional climate model projections for the State of Washington." *Climatic Change* 102 (1-2), 51–75. https://www.doi.org/10.1007/s10584-010-9849-y.

Sauro, Jeff. 2018. "5 Ways to Interpret a SUS Score." Measuring U. September 19, 2018. https://measuringu.com/interpret-sus-score/#:~:text=Acceptable%20corresponds%20to%20roughly%20above,in%20our%20curved%20grading%20scale.

So, Phil. 2020. "NASA TLX: Task Load Index." Edited by Brian Gore. https://humansystems.arc.nasa.gov/groups/tlx/.

Trzaska, Sylwia, and Emilie Schnarr. 2014. "A Review of Downscaling Methods for Climate Change Projections." USAID: African and Latin American Resilience to Climate Change Project. https://www.climatelinks.org/sites/default/files/asset/document/Downscaling_CLEARED.pdf.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
March 2023	FINAL	

4. TITLE AND SUBTITLE	
Climate Data Standards and Tool Usability Assessments for ESTCP Climate Resilience	**5a. CONTRACT NUMBER** HQ0034-19-D-0001
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER** AM-2-5090
Bewley, Jennifer L. Wong, Gifford J. Chappell, Isaac S. Berzolla, Zachary M.	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Institute for Defense Analyses Systems and Analyses Center 730 East Glebe Road Alexandra, VA 22305-3086	IDA Document D-33445
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
DASD, Environment and Energy Resiliency Strategic Environmental Research and Development Program and Environmental Security Technology Certification Program 4800 Mark Center Drive, Suite 16F16 Alexandria, VA 22350	SERDEP/ESTCP **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimited (6 July 2023).

13. SUPPLEMENTARY NOTES

14. ABSTRACT

With the effects of a changing climate and extreme weather events on Department of Defense (DoD) infrastructure threatening its ability to execute its mission, the Biden administration is requiring climate considerations to be an integral part of all DoD strategy, planning, and programming activities. To support this mission, the Environmental Security Technology Certification Program (ESTCP) developed a new focus area in Climate Resilience (CR) which will invest in innovative technologies and methodologies that address climate resilience issues facing DoD installations today and in the near future. The goal of this paper is to further investigate the use of climate data standards and tool usability assessments to help improve proposal responses for ESTCP CR fiscal year 2024 proposals and beyond. Climate models are a tool that can be used to better understand the ways in which the climate is changing for a region. While global climate models (GCMs) can provide information at a large scale, more localized information is needed for installation-specific assessments. Dynamic downscaling (DD) and statistical downscaling (SD) are two promising methods for providing more localized climate information from GCSMs. Similarly, new decision-making tools are likely to be needed for decision makers across the Department. However, because installation-level decision makers are already overloaded with tools, climate information should be incorporated seamlessly into existing tools when possible. If it is not possible, objective usability testing (e.g., NASA Task Load Index, System Usability Scale, User Engagement Scale) should be completed to ensure the tool generates value-add for the end users. Proposals should also clearly state who the intended end users are and the benefits of the proposed tool.

15. SUBJECT TERMS

changing climate; Climate Change; climate change adaptation; climate resilience; dynamic downscaling; Environmental Security Technology Certification Program (ESTCP); Global Climate Model (GCM); objective usability testing; statistical downscaling

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Preston, Kurt
Unclassified	Unclassified	Unclassified	Same as Report	36	**19b. TELEPHONE NUMBER** *(Include area code)* 502-528-0796

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18