Actionable Intelligent Visual Analytics of Ensembles

Kristin Potter, National Renewable Energy Laboratory (Principal Investigator)
Hanqi Guo, The Ohio State University (Co-Investigator)
Franck Cappello, Argonne National Laboratory (Co-Investigator)

Ensemble simulations are needed in DOE science applications—including cosmology, Earth systems, and energy systems—to understand model sensitivities and to quantify model uncertainties. Ensembles are a sparse sample of possible outcomes of simulation parameters, and the high volumes, velocities, and varieties of ensemble simulation outputs are a grand challenge for the comprehensive understanding of simulation parameter spaces. Driven by three DOE applications, namely Earth systems, energy systems, and cosmology, our research will consist of three tightly integrated pillars: visualization surrogates, variation models, and actionable visualization outlined below.

Visualization surrogates will focus on creating AI models to relate the input parameter space to generated visualizations in image space, thus providing a comprehensive set of desirable future states. Such models will mitigate the high cost of executing full simulations for the visual analytics of parameter spaces. Specifically, we will research both image- and data-based surrogate models; the former directly predicts 2D visualization images, and the latter synthesize 3D grid/particle data for scenarios that require 3D interactions and feature extraction. The outcome of the models will allow interactive exploration and enable efficient sampling of the parameter space for deriving statistics and supporting decision-making, respectively, in Thrusts II and III.

Variation models will derive principal characteristics of the simulation parameter spaces. The surrogate models will allow efficient sampling of the parameter space for characterizing and ranking the impact of individual parameters for the user-specified objectives. Methodologies developed in this thrust will suggest high-sensitive parameter(s), the direction(s) for parameter-tuning, and the confidence of the action based on existing samples. Thrust III will further incorporate humans in the loop to support interactive decision-making.

Actionable visualization will leverage surrogate and varation models to establish an actionable paradigm for human-in-the-loop decision-making for optimizing input parameters. Specifically, it will bridge parameter space and decision space through the visualizations and features derived in Thrusts I and II to help decision-making by translating principal characteristics into interactive visual metaphors. Rather than computing expensive inverse models from decision space back to parameter space, we will store the mapping between parameter space and decision space and use human-guided interactive visual exploration.

Cover Page

Proposal

FOA Number: DE-FOA-0002726
DOE/SC Program Office: Advanced Scientific Computing Research (ASCR)
DOE/Office of Science Program Office Technical Contact: Dr. Margaret Lentz
PAMS Preproposal tracking number: PRE-0000030870
Priority Research Directions: 1, 2, and 5 (primary motivation is PRD #1)

Actionable Intelligent Visual Analytics of Ensembles

Lead Principal Investigator

Kristin Potter, Senior Visualization Scientist, Group Manager
National Renewable Energy Laboratory
Street Address: 16253 Denver West Parkway, Golden, CO 80401
Postal Address: 15013 Denver West Parkway, Golden, CO 80401
Phone 303-275-4122, Email Kristi.Potter@nrel.gov

Co-Principal Investigators

Hanqi Guo, Associate Professor, The Ohio State University
Franck Cappello, Senior Computer Scientist, Argonne National Laboratory

Senior Personnel

Han-Wei Shen, Professor, The Ohio State University
Tom Peterka, Computer Scientist, Argonne National Laboratory
Julie Bessac, Assistant Computational Statistician, Argonne National Laboratory
Sam Molnar, Visualization Scientist, National Renewable Energy Laboratory

Administrative Point of Contact

Ray Grout, National Renewable Energy Laboratory
Phone: 303-384-6447, Email: ray.grout@nrel.gov

This proposal is part of a collaborative proposal led by National Renewable Energy Laboratory

Cover Page Supplement

Collaborating Institutions and Institution PIs

National Renewable Energy Laboratory: Kristin Potter
The Ohio State University: Hanqi Guo
Argonne National Laboratory: Franck Cap pello

Lead PI for the Combined Research Activity

Kristin Potter, National Renewable Energy Laboratory

Leadership Structure of the Collaboration

PI Potter will oversee and organize the entire project and lead Thrust III on actionable ensemble visualization at the National Renewable Energy Laboratory.
Co-I Guo will lead Thrust I on surrogate visualization models at Ohio State University.
Co-I Cappello will lead Thrust II on variation models at Argonne National Laboratory.

Description of Facilities, Equipment, and Resources Available to the Collaborative Group

All of the facilities, equipment, and resources listed in Appendix 4 will be made available to the collaboration.

Student and Early-Stage Researcher Training and Mentoring by Senior Researchers

The detailed plan for recruitment and retention of students and early-stage researchers appears in Appendix 7.

Annual Budget ($ in thousands)

Name	**Institution**	**Year 1**	**Year 2**	**Year 3**	**Total Project**
Kristin Potter	National Renewable Energy Laboratory	$300	$300	$300	$900
Hanqi Guo	The Ohio State University	$300	$300	$300	$900
Franck Cappello	Argonne National Laboratory	$300	$300	$300	$900
Total		$900	$900	$900	$2,700

Contents

1 Introduction

DOE science applications——including Earth systems, cosmology, and energy systems——widely use ensemble simulations for understanding by analyzing model sensitivities, quantifying model uncertainties, and exploring the range of possible realizations. However, it is extremely difficult to make effective decisions using current visualization tools for ensemble analysis due to the complexity of ensemble data sets, the computational expense of running individual realizations, and the lack of tools that integrate parameter and uncertainty investigations with effective approaches for decision making. The goal of this research is to design a decision-making methodology that identifies the subset and range of the simulation parameters producing the desired outcomes of an end-user, stakeholder, or policymaker. For example, NREL's reV model [31] can be used to calculate renewable energy capacity, generation, and cost based on geospatial intersection with grid infrastructure and land-use characteristics; however, stakeholders must make sense of parameters relating to the technology, such as power density or solar multiplier to make a decision on the efficacy of the placement of a wind or solar farm.

Our proposed research is driven by two questions: how do visualizations of ensembles change with respect to input simulation parameters, and in what direction(s) should one change the input simulation parameters to achieve desired results? These fundamental questions drive our investigation of a visual system that provides a direct link between simulation parameters and a final, desired future output state, rather than simply exposing a range of output possibilities. We seek to enable robust decisions by users who are not experts in an underlying model and its parameters, which are often complex and unintuitive. By eliminating parameters that do not contribute to a desired result, exposing critical parameters that are necessary for a specific result, bounding the ranges of contributing parameters, and reducing the computational cost and cognitive burden of the process, our methodology will make it easier to infer parametric meaning, provide context and other metadata for decision-making, and thus lower the barriers of entry for users at all technical levels, helping them make more robust decisions.

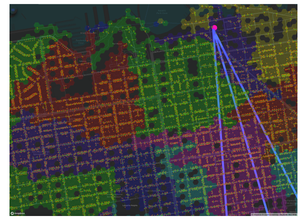
This research will establish a novel paradigm to help users understand the outcome of ensemble simulation parameter spaces and make robust decisions to achieving desired outcomes. To these ends, we will explore the use of artificial intelligence (AI) and statistical models to efficiently predict outcome visualizations for exploring parameter spaces. We will also investigate ways to highlight the most influential parameters, suggest directions to change parameters, and display the confidence of the change. This research will also build visual analytics methodologies and tools to support decision-making.

Falsifiable Scientific Hypothesis: The visual exploration of ensemble simulations paired with predictive models of parameter spaces will enable more robust decision-making, particularly with users who are not intimately familiar with the underlying model but are concerned with the outcomes. We will test this hypothesis using DOE-relevant applications and evaluations with users ranging from domain experts and model developers to decision and policy makers.

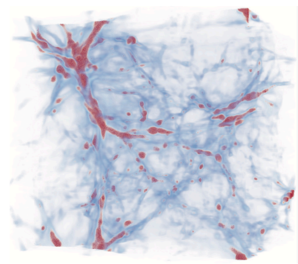
1.1 Driver Applications

We motivate this research with our extensive experience working with a wide range of DOE applications.

Earth Systems. A key challenge in understanding Earth system simulations is the exploration of multidimensional parameter spaces. For example, in the research of Atlantic Meridional Overturning Circulation (AMOC), scientists would like to understand how greenhouse gas concentrations impact the strength of AMOC and climate response [30]. To this end, scientists need to simulate different socioeconomic pathways (e.g., the percentage of yearly increments of CO_2 emissions) and different internal parameters (e.g., critical bulk Richardson number, the Gent McWilliams mesoscale eddy parameterization, and the horizontal viscosity) to understand AMOC. However, the exact value of each parameter is unknowable and requires tuning, but such tuning is usually prohibitive and may require exhausting the parameter space by running an ensemble of computationally expensive simulations.



Energy systems. Understanding how to transform our current energy system to a resilient, robust, and sustainable future is an extremely important and hard problem. The grid infrastructure is an aging system with increasing failures that can pose significant hazards including sustained outages or even wildfires. Population growth and the electrification of our transportation networks is increasing energy demand. More extreme weather events highlight the need for carbon-neutral energy sources. Solving these problems requires modeling and simulation at varying scales and resolutions, both spatially and temporally, as well as understanding the connections between energy demands, generation, transmission and distribution, while keeping an eye on the needs of the human population the power grid is serving. This broad scope and the multitude of models used to understand the problem is compounding the already hard problem of large-data analysis.



Cosmology. The origin and nature of dark matter and dark energy in the universe is one of the grand challenges of scientific inquiry. Cosmologists and astrophysicists compute very large-scale models of the evolution of the universe and compare those computational results with sky surveys. There are simulation parameters that control how the digital universe evolves and corresponds to the real one. The key challenge is the decision-making for parameter selection, because it is difficult to derive and visualize the response of cosmological outputs (e.g., halo mass density and power spectrum) to input parameters (e.g., proportion of matter, mass fluctuations, and spatial curvature). In addition, parameter tuning usually requires full-scale ensemble simulations, which are very costly to run.

1.2 Proposed Research

Driven by the application needs, we propose a novel visual analytics paradigm to allow users to understand the outcome of simulation parameter spaces and lead to robust decision-making for achieving desired outcomes. To accomplish this goal, we will investigate internal representations and models in three research thrusts, mapping between several different spaces, as shown in Figure 1 and outlined below.

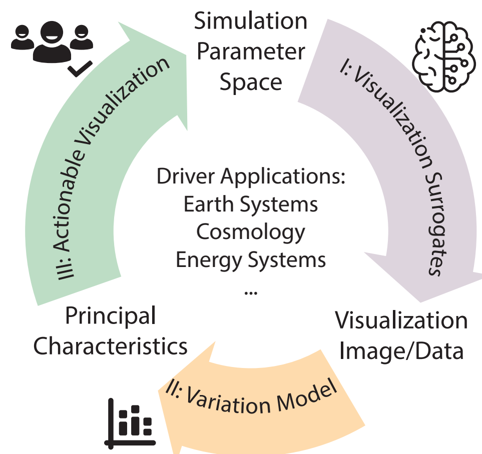


Figure 1: Overview of the proposed research framework.

Thrust I: Visualization surrogates will focus on creating AI models to relate the input parameter space to generated visualizations in image space, thus providing a comprehensive set of desirable future states. Such models will mitigate the high cost of executing full simulations for the visual analytics of parameter spaces. Specifically, we will research both image- and data-based surrogate models; the former directly predicts 2D visualization images, and the latter synthesize 3D grid/particle data for scenarios that require 3D interactions and feature extraction. The outcome of the models will allow interactive exploration and enable efficient sampling of the parameter space for deriving statistics and supporting decision-making, respectively, in Thrusts II and III.

Thrust II: Variation models. Having established the surrogate visualization models, we will research deriving the *principal characteristics* of the simulation parameter spaces. The surrogate models will allow efficient sampling of the parameter space for characterizing and ranking the impact of individual parameters for the user-specified objectives. Methodologies developed in this thrust will suggest highly-sensitive parameter(s), the direction(s) for parameter-tuning, and the confidence of the action.

Thrust III: Actionable visualization will leverage Thrusts I and II to establish an actionable paradigm for human-in-the-loop decision-making for optimizing input parameters. Specifically, it will bridge parameter and decision space through the visualizations and features derived in Thrusts I and II by understanding the sensitivities of parameters and principal characteristics and translating those relationships into interactive visual metaphors. Rather than computing expensive inverse models from decision space back to parameter space, we will store the mapping between parameter space and decision space for the inferences of the forward models in Thrusts I and II and use human-guided interactive visual exploration to implicitly solve an inverse problem.

1.3 Strategic Alignment

The proposed research aligns with multiple priority research directions (PRDs) identified by the ASCR visualization research community. First, we will address PRD 1 to advance visualization theory and techniques for understanding complex data; new methodologies will be developed for ensemble and uncertain data. Second, we will address PRD 2 to support scientific workflows; we will develop new approaches to represent statistical characteristics in ensemble simulation data (Thrusts II and III). Third, this research will address PRD 5 to develop intelligent approaches for scientific data visualization; the actionable ensemble visualization (Thrust III) will enhance the user's experiences in analyzing and optimizing input parameters while gaining scientific insights in ensembles.

1.4 Potential Benefits and Impact

The effective and efficient analysis of ensemble data has been identified as a grand challenges in visualization research, and the proposed research would benefit the ASCR visualization research community, DOE domain scientists, and the general visualization research community as a whole. Our methodology will enable scientists to gain intuition across simulation ensemble parameter spaces for domain sciences. Through AI-assisted visual analytics, the proposed research will potentially reduce computational cost while enhancing understanding and decision-making of the complex simulation parameter spaces.

1.5 Team

This research brings together seven diverse researchers from two national laboratories and one university to tackle crucial ensemble data analysis and visualization problems. Lead-PI **Kristin Potter** is a recognized leader in uncertainty and ensemble visualization, will oversee the entire project at NREL, and lead Thrust III. She is a senior computer scientist at NREL and a group manager for the data, analysis and visualization group within the computational science center. Potter is currently an associate editor for IEEE Transactions on Visualization and Computer Graphics, and a general chair of the 2024 IEEE Visualization conference. She is also very active in supporting early career researchers, having helped to develop a junior faculty "summer camp" for visualization academics, pushing for an early career talk track in the DOE Computer Graphics Forum workshop, and helping organize an early career seminar series at NREL. Co-I **Hanqi Guo** is a DOE early-career awardee in parameter-space exploration and will lead Thrust I on surrogate visualization models. Guo combines a unique background in scientific visualization, applied mathematics, high-performance computing, and AI to solve grand challenges in data analytics and visualization. He has also received best paper awards in premier conferences including IEEE VIS and IEEE PacificVis. Co-I **Franck Cappello** is internationally known for his fundamental contributions to high-performance parallel computing and to lossy reduction techniques preserving data characteristics. He is a Computer Science Leader at Argonne National Laboratory and will lead Thrust II. **Han-Wei Shen** has nearly three decades of experience in large-scale data visualization and analysis and will co-lead Thrust I with Guo. Shen was the PI for two previous Scientific Data Management and Analysis at Extreme Scale projects on information theoretic frameworks for visualization and distribution-based data analytics, a co-PI for the SciDAC RAPIDS Institute, and the recipient of 2003 DOE Early Career and 2004 NSF Career awards. He is currently an Associate Editor-in-Chief for IEEE

Transactions on Visualization and Computer Graphics, and the chair of the IEEE SciVis Steering Committee. **Tom Peterka** is a DOE early-career winner in multivariate functional data analysis and will co-lead Thrust II with Cappello. He has won five best paper awards, currently leads several DOE-funded projects and has published over 100 peer-reviewed papers in conferences and journals, including ACM/IEEE SC, IEEE IPDPS, IEEE VIS, IEEE TVCG, and ACM SIGGRAPH. **Julie Bessac** is a leader in spatio-temporal statistical modeling and will contribute to Thrust II. She is an assistant computational statistician at Argonne National Laboratory. **Sam Molnar**, a visualization scientist at NREL with rich expertise and experience working with power system dynamics, will co-lead Thrust III with Potter.

2 Surrogate Models for Ensemble Visualization

To model scientific phenomena with a wide range of possible outcomes, scientists often run simulations with different inputs to identify the parameters that can generate results comparable with empirical observations. Running many simulations is expensive, however, because both computational time and storage for the output can be prohibitively large. Recent developments in deep learning methods offer an opportunity for scientific applications to adopt novel approaches to parameter space exploration. With the assistance of deep learning methods, the exploration of parameter space can be modeled as a generative or regression problem.

In this research thrust, we will explore two distinct types: image-based and data-based surrogate models. Image-based surrogates directly predict 2D visualization images, while data-based surrogates synthesize 3D visual data such as volumes and particles. Image-based surrogates are typically trained with preset visual parameters (e.g., view angles and visual mappings) and cost relatively less to train than do data surrogates. Data surrogates allow for flexible 3D interactions and postprocessing operations such as isosurfacing and feature extraction, but they are inherently more costly than image-based surrogates. Both image- and data-based surrogates are further used in Thrusts II and III for decision-making.

2.1 Image-Based Surrogates

To allow for efficient exploration of simulation parameters, we will develop deep visualization surrogates to predict the visualization of ensemble simulation output with any given configuration of input parameters. In contrast to the more traditional simulation surrogate research, the main innovation of our approach is to skip generating simulation data and instead directly produce visualizations that will allow the scientist to obtain a quick preview of their simulations, in order to determine more salient parameters. The major benefit of our approach is to significantly reduce both the computation and storage costs for ensemble parameter exploration.

The team made the first effort, InSituNet [21], to use deep surrogate models for synthesizing visualization results conditioned with different simulation parameters. The inputs to the deep neural network DNN model included the simulation parameters and visualization parameters, such as values for isosurfaces, volume rendering transfer functions, and camera positions and directions. The output of the model was the visualization of the corresponding simulation in the form of images or features. To create the surrogate, we took advantage of both convolutional neural networks (CNNs) and generative adversarial networks (GANs). First, we trained a deep regression model called a *regressor* based on the visualization collected in situ, which included simulation parameters P_{sim}, visualization parameters P_{vis}, and the corresponding visualization images or features I. The regressor then learned a function F that took the parameters as inputs and generated the corresponding visualization. The regressor adjusted the weights iteratively with respect to the difference between the predicted visualization and the ground truth visualization. To improve the accuracy and fidelity of the prediction, we used the adversarial theory of GANs [15]. Specifically, we trained another network called a *discriminator* along with the regressor to differentiate the prediction from the ground truth, in order to stimulate the regressor to generate better results. With the trained model, we were able to infer the visualization results for arbitrary parameters within the parameter space through forward propagation. Figure 2 demonstrates the initial results of our experiments. In (a) and (b), the images predicted by the trained regression models are similar to the ground truth images for the Nyx [4] and MPAS-Ocean [46] simulations, respectively. In (c), we demonstrate that the trained model can be used to calculate the sensitivity of subregions of an image to the simulation parameters.

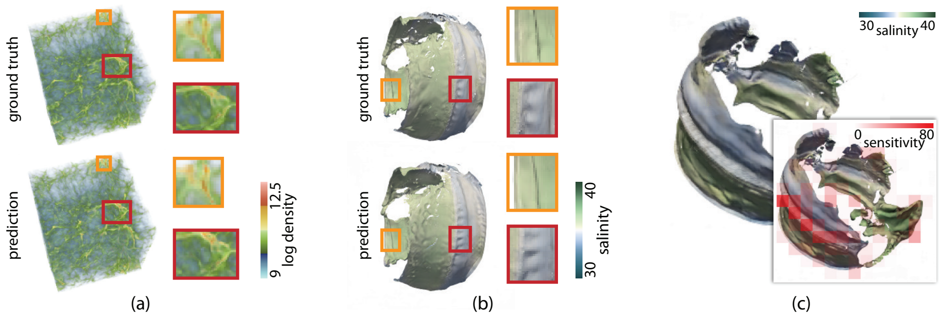


Figure 2: Comparisons of the predicted images with the ground truth, (a) and (b). (c) Sensitivity of subregions of an image computed with backpropagation.

2.5D visualization surrogates. While our initial attempt to create a visualization surrogate that generates images is encouraging, a major limitation of the work is the limited flexibility for scientists to explore the output. Instead of creating images, far more desirable is for the surrogate model to generate visualization features such as isosurfaces, flowlines, or domain-specific features such as halos or vortices. In order to do so, visualization features need to be extracted at simulation time, serving as the training data. We will extend the loss functions used in the image-based surrogate model to include feature-specific error measurements. First, we plan to extend the surrogate to output 2.5D images, that is, images with 3D attributes such as normals or depths. The 2.5D output produced will allow us to perform deep deferred shading [37] and will allow for more flexible interaction, such as rotation and relighting. The inferred data from our surrogate can also be vectors that allow image-space seeding [28] and flowline computations on the cross section [51]. We will further improve the surrogate to create 3D objects such as isosurfaces and flowlines. The ultimate goal of this research is to develop a visual analytics system that will allow scientists to explore the simulation parameter space in the forms of images, features, and higher-dimensional visualization objects.

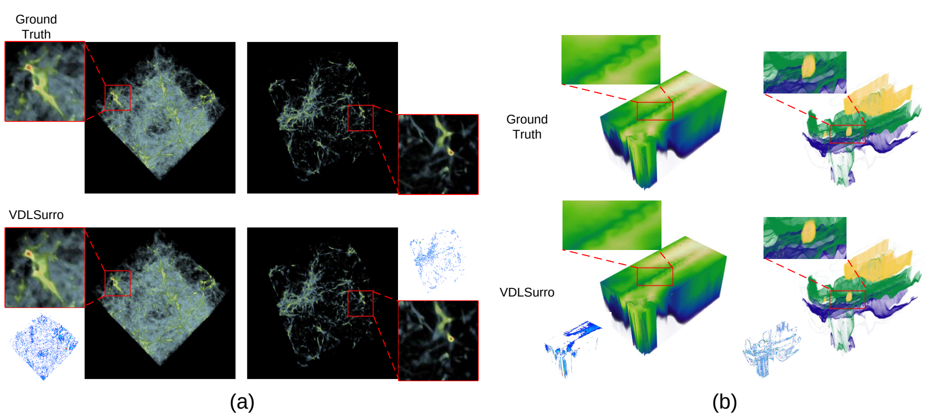


Figure 3: Comparisons of the images generated using VDL-Surrogate with the ground truth images.

View-dependent visualization surrogates. Overall, the pure image-based visualization surrogate, InSituNet [21], has three limitations: (1) low resolution of output images; (2) low accuracy of output images; and (3) poor flexibility when exploring different visual mapping parameters. The main reason is that besides simulation parameters, InSituNet encodes visual mappings and viewpoint parameters, a process that leads to two outcomes. First, it dramatically increases the difficulty of network training because the variance between different visualization results is much bigger than different data-space ensemble members. Second, because of the large joint space of all simulation, visual mapping, and viewpoint parameters, current work can encode only a limited number of visual mappings and thus constrains scientists from exploring the data with different visual mapping parameters. If we restrict our model to encode only simulation parameters, then the learning task would be much easier. With computational resources, however, it is difficult to directly learn a surrogate model that can produce previews with sufficient resolution for visualization and analysis. To improve the efficient use of computational resources, we plan to use a view-dependent neural network latent-based surrogate model called VDL-Surrogate. We can perform ray casting from different viewpoints to collect samples and produce compact latent representations, and this latent encoding process will reduce the cost of surrogate model training while maintaining the output quality.

We plan to develop a view-dependent visualization surrogate comprising three components. The first component is **view-dependent latent generation**. To ensure good coverage of the domain given the predetermined image resolution, we will select a diverse set of view directions. For a simulation output and a selected viewpoint, we perform ray casting and use a neural network-based autoencoder called **R**ay **A**uto**E**ncoder (RAE) to encode samples along each ray with a latent representation by a new information-driven weighted $L1$ loss. We will train three corresponding RAEs for the three selected viewpoints. Each RAE is trained given the stored simulation outputs from a few selected simulation runs. For other simulation runs, we will use the trained RAE encoder and generate the view-dependent latent representation from the ray samples in situ. The second component is **offline training of our simulation surrogate**. For a selected viewpoint, given the simulation parameters and view-dependent latent representation pairs, we will train a convolutional model called **V**iew-**D**ependent **L**atent Predictor (VDL-Predictor). Similar to RAE, three VDL-Predictors will be trained for three selected viewpoints. The third component is **post hoc visualization and exploration**. Scientists can use the trained VDL-Predictor to predict the view-dependent latent representations, decode the latent representations by the RAE decoder, and perform visualization using user-specified visual mappings. We have prototyped the view-dependent surrogate model on small-scale ocean and cosmology simulations (see Figure 3). We will generalize the approach to much more complex Earth and energy systems and cosmology simulations.

Research Tasks: Image based visualization surrogates

2.1.a Develop 2.5D surrogate visualization models that generates depth information
2.1.b Research view-dependent surrogate models that produces arbitrary viewpoint visualization

2.2 Data-Based Surrogates

We will explore data-based visualization surrogates, as opposed to image-based approaches for complex visualization and analysis tasks. While image-based surrogates are useful in synthesizing visualization results for exploring parameter spaces, a key difference is that image-based surrogates produce images as opposed to spatial data. For applications that require additional filtering such as isosurfacing, feature extraction, and statistics, we will directly synthesize 3D data (e.g., mesh data and particle data) for visualizations. Driven by our applications, we will investigate surrogate models for unstructured mesh data and particle data and further explore parallelization strategies for training data-based surrogates.

Surrogate visualization models for unstructured mesh data and graphs. Both driver applications—Earth systems and energy systems—produce spatial data represented by graphs. In Earth system applications, for example, the Model for Prediction Across Scales (MPAS) [46] features unstructured Voronoi meshes and the C-grid discretization. In energy system applications, the models outputs include connections between energy demands, generation, transmission, and distribution. However, existing data-based visualization surrogates focus on regular grids [3, 20], which do not directly work for graph data.

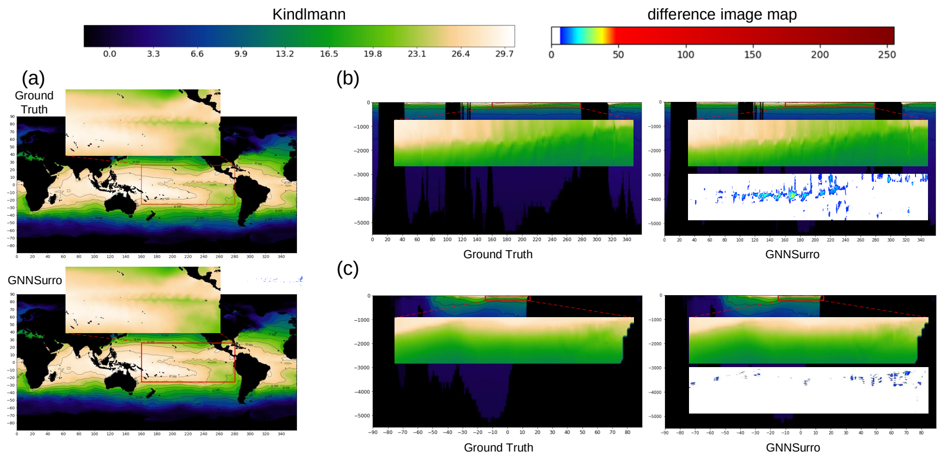


Figure 4: (a) Comparison of the sea-level temperature maps generated using GNN-Surrogate with the ground truth maps. Comparison of the vertical cross-sections at (b) the equator (c) $75°E$ generated using GNN-Surrogate with the ground truth cross-sections.

To this end, we will explore a graph neural network (GNN) method to predict data on unstructured grids with given simulation parameters. We outline the plan based on MPAS-Ocean simulations; the same technique will apply to other Earth and energy system applications. Specifically, we model MPAS-Ocean unstructured meshes as a graph and propose a GNN-Surrogate that supports graph operations to learn from MPAS-Ocean simulation output. We will build a graph that captures the vertex connectivity and distance information, one of the common choices to represent unstructured meshes [10, 39, 40]. For efficient training, we will generate hierarchical graphs, where coarse graphs help GNN-Surrogate capture the global phenomena quickly. To generate hierarchical graphs, we will use a graph-coarsening algorithm. Furthermore, we will cut the graph hierarchy to reduce the I/O and training computation cost. Specifically, with the graph hierarchy cutting, the GNN-Surrogate adaptively decides which resolutions to use at different locations, depending on the complexity of the phenomena at various locations. The simulation outputs in the training dataset will be represented by adaptive resolutions that supervise the GNN-Surrogate training. Given the hierarchical graphs, GNN-Surrogate is an upsampling-convolution generator. The graph convolution can refine the feature map represented on graphs. Moreover, its local connectivity and weight sharing scheme allow GNN-Surrogate to avoid overfitting. Figure 4 shows the initial results we generated with a simplified MPAS-Ocean test case (with 30–60 km resolution) with a few input parameters [48]. We will investigate graph simplification strategies to scale GNN-surrogate models for much higher resolutions and generalize GNN-surrogate to time-varying data by introducing 4D convolutions.

Surrogate visualization models for particle data. Driven by cosmology applications, we propose a convolutional neural network-based particle surrogate. The challenge here is designing an efficient convolution kernel. For particle data, the efficiency of the regular convolution kernel relies on finding neighbors in $O(1)$ time with the help of the regular meshed grids. Because the particle data does not have a mesh structure, however, using the straightforward method, for each particle we will need to scan all other particles and compare the distance to obtain its neighbor, a process that is $O(N)$ complexity, where N is the number of particles. To overcome this challenge, we will utilize an octree index and construct it as follows. Given a list of particles, each leaf node in the tree contains particles no more than a given threshold. The leaf nodes of the octree form a graph; and for a target particle, its neighbors are highly possible inside its residing graph node and the graph node's 1-hop neighbor. By following this assumption, we can perform the approximate nearest-neighbor search more efficiently than the straightforward neighbor-finding approach. Specifically, our workflow will involve three steps. First, given the pair of simulation parameters and construed octrees, we will design an octree prediction network (OPN) to predict octrees. Second, given the pair of simulation parameters and particles, we will design a particle prediction network (PPN) based on the ground truth octree tree to predict particles. Third, having the trained OPN, we will predict octrees based on simulation parameters; and with the predicted octrees, we will further fine-tune the PPN. The first two steps will be processed in parallel with two graphics processing units (GPUs).

Parallel graph models using DIY We plan to leverage our DIY infrastructure [34] for parallel graph neural networks because most other parallel artificial intelligence frameworks rely on data parallelism, but our large graphs require model parallelism. DIY is designed to be a generic parallel programming model, and previous results show good strong scaling when applied to parallel graph construction of a Morse-Smale Complex using up to 32K processes [17] and to graph connected components up to 8K processes [35, 57]. We will study three main aspects of high-performance parallel GNNs.

Subgraph decomposition and assignment: We will study how to determine the appropriate size and number (granularity) of subgraphs and appropriate overlap between subgraphs. This impacts the ability to balance the computational and communication load between subgraphs. *Highly irregular asynchronous communication patterns:* The concept of a local communication neighborhood can be generalized to communicate among subgraphs of a GNN. We will investigate local communication neighborhoods linked by edges in the original graph and how information can flow efficiently across multiple neighborhoods. We will utilize recent advances in asynchronous iterative communication developed by our team [35]. *Parallelizing over emerging hardware accelerators:* We will extend the DIY local neighborhood communication to communicate directly between GPUs, bypassing the host CPU. To do so, we will build on recent MPI advances in direct GPU communication. We will then extend this work to communicate among parallel tensor cores and other hardware dedicated to graph neural network algorithms.

Research Tasks: Data-based surrogates

2.2.a Research novel surrogate visualization models for unstructured mesh data
2.2.b Explore novel surrogate visualization models for particle data
2.2.c Parallelize data-based visualization surrogate models

3 Variation Models for Ensemble Visualization

Having established the surrogate models for visualizing ensembles, the next step on our roadmap is to derive variations from the models in order to assist interactive decision-making. Specifically, the objective is to characterize influences of input parameters to the model outputs in order to guide parameter selection and refinement. To this end, we propose two distinct paradigms to assist parameter tuning:

- A local approach to quantify variations and sensitivities of individual input parameters based on neural networks
- A global approach to recommend top influential parameters, directions to change parameters, and the confidence of the recommendation based on statistics

The purpose of having two different approaches is to serve different use cases: the local approach characterizes model/parameter sensitivities while the global approach helps users optimizing parameters. Outputs from both local and global approaches will be used in Thrust III to close the visual analytics loop.

3.1 Deriving Variations from Visualization Surrogates

To evaluate the predicted visualization and select salient simulation parameters, one needs to analyze the uncertainty and sensitivity of the visualization surrogate mentioned above. In general, there are sources of uncertainty for our DNN model. Understanding the behavior of the visualization surrogate with respect to its input and output can help analyze the uncertainty of our model and identify salient simulation parameters. Below we describe our approaches in detail.

Quantifying variations via dropout. Traditionally, dropout layers in neural networks are used as regularizers in the training phase to avoid overfitting of training data by randomly ignoring the neuron activations at different layers of the network. Gal et al. [13] showed that if we apply dropout during prediction, we get slightly varying predictions every time the network is run with the same inputs. By observing the variations of the predicted results, we can quantify the uncertainty associated with the predicted results. In our research we will experiment with different dropout strategies and visualize their effects on the quantification and control of model uncertainty. Starting from a random dropout of neurons, we will develop approaches to analyze the variation of results in high dimensional space, where the dimensionality is controlled by the numbers of dropped and retained neurons. We will also systematically experiment with skipping neurons at different layers so as to identify the uncertainty at different feature scales. For example, in CNNs, different layers of neurons represent different feature levels; hence, the degree of uncertainty with respect to the dropout neurons indicates the robustness of features identified from the respective layers. Since there are a number of candidate neurons for dropout, effective visualization strategies need to be developed to understand the cause and effect of dropout and uncertainty. Alternatively, we will explore the use of LassoNet [27], which combines a neural network representation with a linear combination of input features and a sparsity penalization of type LASSO [19]. The $\mathcal{L}^1$-LASSO penalization is applied in a structured way so that features with the least importance in the linear part of the model have a reduced contribution to the neural networks via the weights of their first layers.

Visualizing the sensitivity of simulation parameters. Sensitivity analysis of a neural network corresponds to computing the partial derivative of the outputs with respect to the inputs. Given the ith neuron in the output layer, we can predict the function $f_i(\mathbf{x})$ for an n-dimensional input vector $\mathbf{x} \sim \{x_0, x_1, \ldots, x_n\} \in \mathbb{R}^n$. The sensitivity of f_i with respect to the jth input parameter of the simulation can be denoted as $\left(\frac{\partial f_i}{\partial x_j}\right)^2$. A high-sensitivity value corresponds to the fact that a small change in the value of the input x_j causes a significant change in the output value of $f_i(.)$. The architecture of a neural network is such that the output of every neuron in the network is completely differentiable with respect to its inputs. As a result, we can easily compute the required partial derivatives for sensitivity analysis via a chain rule using backpropagation.

In this research we will perform parameter sensitivity analysis by visualizing the process of backpropagation. In essence, two directions can be taken to visualize the parameter sensitivity. One is to visualize how small changes in the predicted visualization are propagated back to the simulation input. Another direction is to determine how a change in simulation input will affect the visualization output and the weights of DNN links. To perform both analyses at the same time, one needs to visualize and analyze data in a very high-dimensional space, where the data is the predicted visualization outputs and the gradient of which with respect to the simulation inputs. To visualize the complex model behavior, we will sample the data (visualization output and gradients) from the high-dimensional parameter space in the DNN model and develop effective high-dimensional data visualization and query algorithms. Recently we have demonstrated an initial success in our NNVA system [20] for a small-scale DNN-based biological simulation with 35 input parameters, where salient parameters can be obtained via interactive visualization. In this research, tailored to the visualization surrogate that we will develop, interactive visual analytics systems will be developed to allow flexible exploration of the parameter space for ensemble simulations. Our goal is to obtain a better understanding of the underlying parameter sensitivity of simulations.

Research Tasks: Deriving variations from visualization surrogates

3.1.a Derive variation models of ensembles via visualization surrogates
3.1.b Derive parameter sensitivities via visualization surrogates

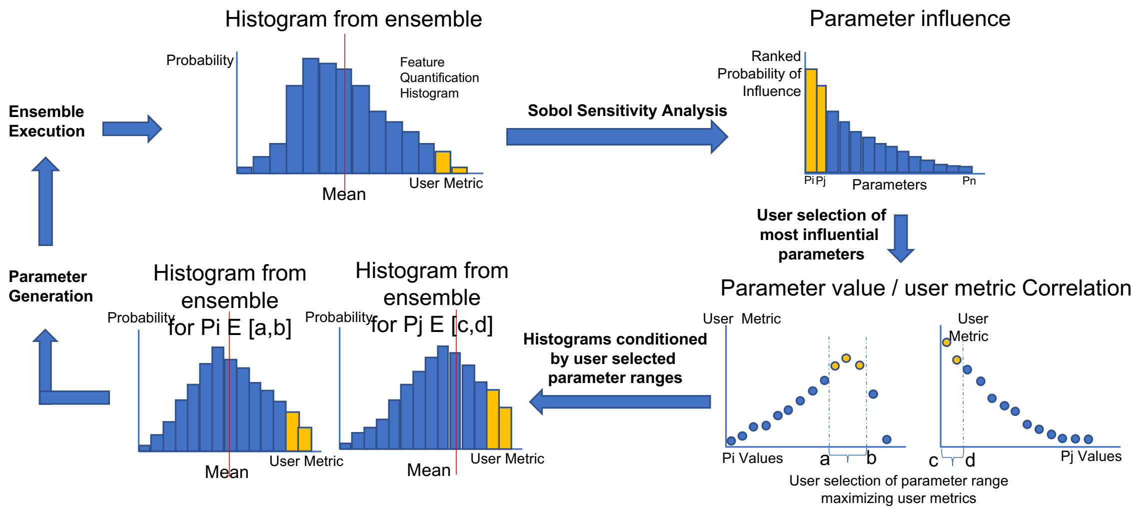


Figure 5: Overview of how parameter variation and influence will be used to modify the ranges of influential parameters.

3.2 Deriving Variations from Principal Characteristics

Given surrogate models that produce images and data for visualization, we will investigate the effect of simulation input parameters on quantities of interest to the user: what we call principal characteristics. These can be anything from a single one-dimensional feature (e.g., a streamline) to an entire high-dimensional time-varying field (e.g., a flow map) and include positions, values, quantities of interest, statistics, distributions, geometric features, topological features, and derived quantities. This correlation between simulation input parameters and principal characteristics will be explored iteratively and progressively, in conjunction with user input, to reduce and refine the parameter space and produce a small set of actionable recommendations that fit user needs. We will conduct sensitivity analyses quantifying the (potentially combined) influence of each input parameter on principal characteristics and will characterize the principal characteristic variation, producing distributions and rankings of input parameter influence. Users will visualize the ensemble data, the distributions of the principal characteristic variation, ranking of parameter influence, and ranges of selected parameters to guide them to further narrow the parameter space to an actionable region. Our team has extensive experience in global sensitivity analysis for likelihood of statistical models [18] and in computing moment-independent global sensitivity to quantify and analyze parameters in ensembles of time-varying weather simulations [5].

Approximation of parameter influence. We propose to compute and display the global sensitivity [24] of principal characteristics on simulation input parameters through an extension of Sobol indices [49]. Alternatives for computing global sensitivity rely on Fourier amplitudes [47], change in output distribution [6], derivatives [26], and parameter screening [36]. Sobol indices were recently applied to time-varying fields of scalar values and computed progressively [50]. We will apply univariate and multivariate [14] Sobol indices augmented by methods based on variograms for sensitivity analysis of surface outputs [44]. Compared with using variance-based sensitivity measures alone, our proposed method will use additional distribution characteristics (higher-order moments and other statistics of the distributions characterizing skewness or tails) to refine our understanding of sensitivity to certain input parameters. Additionally, we will explore the direct computation of distances between conditional and unconditional distributions, between parts or entire characteristic distributions [29, 56], with appropriate transformations and distances, as an indication of what direction and how much to tune each parameter to influence each principal characteristic. The resulting ranking of parameter influence, as well as other statistics on the parameter distribution, will inform the modification of influential parameters.

Modifying influential parameters. Figure 5 illustrates how we propose to use the sensitivity analysis described above to narrow down the range of simulation input parameter values to eventually arrive at an actionable subset of the parameter space. The implementation of these concepts will be described in Thrust III; here we describe the statistics needed. Figure 5 shows the results of the first round of sampling the ensemble simulation space. This first round will sample automatically, for example, using Latin hypercubes or other sampling strategies. The histogram in the upper left of Figure 5 is the distribution of some desired principal characteristic (e.g., number of halos in a cosmology ensemble) from all the samples seen so far. Assume that the user wishes to know what regions of parameter space would increase the number of samples marked yellow at the right-hand tail of the distribution. The Sobol indices of the parameters contributing to the principal characteristic are shown in the upper right. Assume that the two most influential parameters, p_i and p_j, are selected for adjustment. The lower right plots show the samples seen so far, indicating the effect of each of these two parameters on the characteristic. Based on these plots, the user adjusts the range for p_i and p_j, and the new resulting distributions are predicted in the lower left diagrams, one for p_i and the other for p_j, based on available statistics of the current iteration. If the predicted changes are accepted, the next iteration is executed; otherwise the user can reject the proposed change and select other parameters to explore.

In each iteration, more samples in the desired parameter ranges are generated and added to the existing samples, increasing the sample size and accuracy of the statistics. Once the output distribution has sufficient sampling near the desired principal characteristic, we can be confident that the space has been adequately sampled, and the resulting parameter ranges are actionable.

For simplicity, we illustrate the idea with a single principal characteristic, but we will extend the approach to multiple characteristics. Moreover, input parameters are not necessarily mutually independent. Our approach resembles the accept-reject step that is the basis of rejection sampling [9] as well as of Markov chain Monte Carlo techniques [8]. The goal of rejection sampling is to generate samples from a target distribution that cannot easily be sampled, by instead drawing samples from a simpler candidate distribution. A rule has to be built and applied to reject some of the proposed samples so that accepted samples approximate the target distribution. Our approach borrows some aspects of these techniques, with the user iteratively performing accept-reject sampling based on their judgement.

Additional areas of study involve (i) deriving explicit mathematical rules from the user's choices of accepting or rejecting a set of samples and (ii) using other automated strategies such as Bayesian optimization [16] used in hyperparameter optimization are . However, a key difference in our framework compared with fully automated search and optimization approaches is that a human expert remains in the loop in order to guide the search in directions that only the expert can evaluate. Thus we propose a system to assist humans rather than replace them. Real-life decision-making consists of complex cost functions that are difficult or impossible to express analytically and can change during the course of exploration. Outcomes can be unpredictable or unexpected, causing expectations to evolve. Moreover, human oversight is a powerful diagnostic tool to validate how a decision was reached so that it can be explained to others. This is especially important when the response of principal characteristics is discontinuous, categorical, or sparsely defined.

Research Tasks: Deriving variations from ensemble features

3.2.a Research the quantification of principal characteristic variation
3.2.b Develop an approximation of parameter influence
3.2.c Research the derivation of directions for modifying parameters

4 Actionable Ensemble Visualization for Decision Making

Ensemble visualization is an important and challenging problem [25], particularly when using ensembles for decision making [33]. Many tools use multiple windows to show different aspects of the multidimensional data or insight into the parameter space [42]; however, they are designed primarily for users with intimate knowledge of the generating model. Because of the complexity of the relationships between parameters and outputs, successful tools must link these spaces in a meaningful way.

For this work we draw inspiration from the "Is it better to rent or buy?" dashboard from the New York Times [7]. As shown in Figure 6, the user is presented with a single decision, to rent or to buy a home, and parameters impacting that decision are shown as interactive sliders with labels and captions describing their meaning. The dashboard is simple enough to use for anyone with basic knowledge on home buying by abstracting the relationships between many parameters into a single decision metric. Moving the sliders, the user learns how a parameter affects the monthly housing payment and the decision left to the user is "Can you rent a house for less than this price?" This visualization is both simplistic and informative, abstracting a decision-making process into a series of meaningful parametric settings. As a more complex example, the team designed a visual analysis framework for a yeast cell polarization model, which visually guides users to discover new parameter configurations [45]. We plan to build on these works by incorporating novel uncertainty and parameter visualizations, user interactions, and provenance capture, which together create a robust decision-making tool for diverse sets of users.

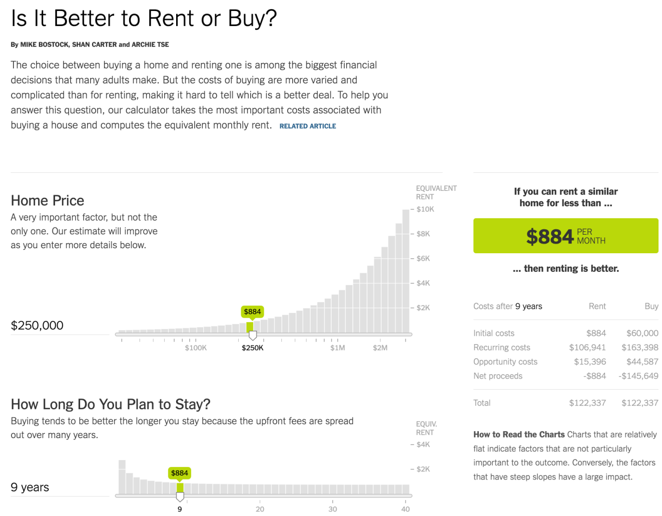


Figure 6: Rent or buy decision-making tool [7].

4.1 Actionable Decision Framework

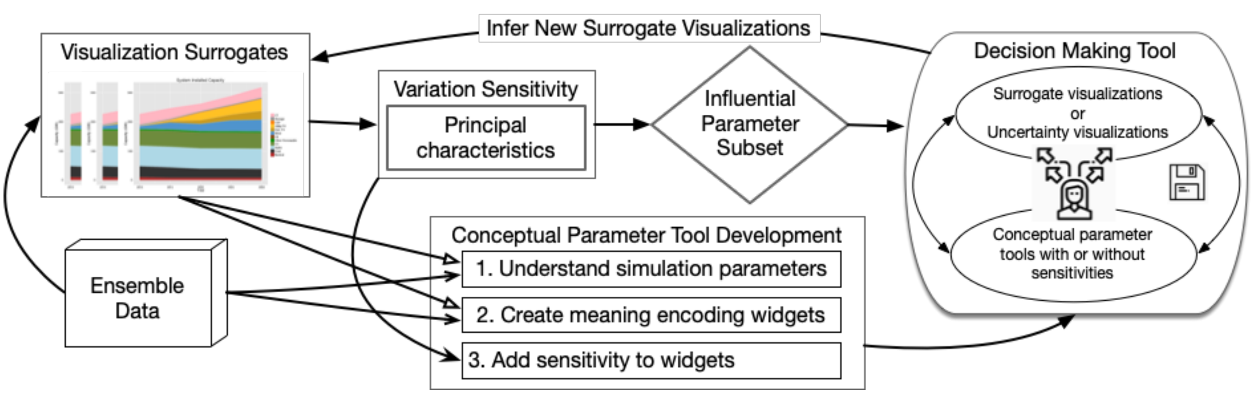


Figure 7: Decision-making framework for bringing together surrogate visualizations, sensitivity analysis of principal characteristics, tools for a conceptual understanding of parameters, and uncertainty visualizations. The person icon shows where user input will be obtained, and the file icon shows our initial plan for recording provenance.

To design a framework for real decision making using complex models, we have to change the way we think about decision making. Scientific investigations are typically formulated to understand the science domain, and the results of those inquiries are not easily accessible to lay people or policy makers [52]. Rather than providing a user with a tool that simply reflects changes in parameter settings, we look to the deep uncertainty conceptual model [53] to reframe the goals of our tool to be *How do we get near a future outcome that is robust to parameter sensitivities?* or, rather, *How can we select a subset of the parameter space that results in a future state close enough to the desired outcome state?*

Figure 7 shows the notional framework for our system. Simulation-generated ensemble data is used to create surrogate models for visualization (Thrust I), an initial understanding of parameters, and the development of meaningful parameter-manipulation widgets (this thrust). Activities in Thrust II provide sensitivities of the system that inform this thrust in the development of uncertainty visualizations and scented parameter widgets. A robust decision comes from a user choosing parameter settings that capture a desired outcome with acceptable levels of variation. Actions made by the user are recorded for future use; and depending on the perceived quality of the final decision, the user may decide to get more simulation data, thus causing the system to retrain the surrogates, while maintaining provenance of the session.

A conceptual understanding of the parameter space. Simulation models use parameters that have various meanings in the physics equations and computational algorithms encoding those equations. Abstracting those meanings into concepts understandable by a decision maker will require strong collaboration with modelers and a good understanding of the science underlying the study. We will start this research task by categorizing the interpretability of parameters, that is, how hard it will be for a decision maker to relate the impact of this parameter to the results. We will have three levels of interpretability. First are parameter meanings that can be directly interpreted as physical concepts that make sense to a user without needing much explanation, such as the slope of the terrain on which a model might want to install a wind farm (hint: not on the side of a mountain). Second are parameters that may be less familiar to users but still have a meaning that could be interpreted with help from definitions, descriptions, or annotations (for example, tooltips that are available in the '‘Rent or Buy?” tool that describe more arcane tax details). Third are parameters that will be hard for decision makers to relate to their decisions but have impact on simulation results, such as computational thresholds. For each category we will build visualization tools to help in understanding the parameter itself, for example, a map of terrain slope, an annotated visualization explaining a parameter, or simple charts showing parameter value ranges for the least interpretable parameters. These initial, descriptive visualizations will be used as the basis for building widgets, in collaboration with Thrust I, and designed to help users navigate the space of visualization surrogates. The widgets may be simplified versions of the descriptive visualizations (like a reduced boxplot), or we may need to provide tooltips or even pop-out modals for users to better understand how to best set a parameter. Once we have decided on an appropriate widget for a parameter, we will modify it with indications of sensitivities from Thrust II. We will look to work such as scented widgets [55] to guide users to parameter settings and help users understand the uncertainties in a system from the parameter point of view.

Design of uncertainty visualizations. The power of the visualization surrogates in Thrust I is the ability to quickly generate a new visualization based on a novel set of parameter settings. The sensitivity analyses in Thrust II can help us understand the extent to which a visualization or principal characteristic can change based on parameters. However there is no direct visual connection between a surrogate visualization that shows a single instantiation of parameter settings or the emergence of a principal characteristic with the range of uncertainty that spans the parameter space. In this task we will look to the uncertainty visualization literature and our previous work in the area to design visualizations that capture the range of possibilities. These visualizations will be primarily for the principal characteristics and look to the literature for a particular type of characteristic, such as probabilistic streamlines [12], or distribution visualizations [41]. We may also find the need to annotate or embed uncertainty into the visual representations coming from the surrogate models, at which point we will augment image-based or data-based visualization surrogate models from Thrust I. In essence, uncertainty visualizations of principal characteristics will show how much characteristics can change and encompass a number of surrogate visualizations. Uncertainty visualizations in the space of surrogate visualizations can capture variations across many principal characteristics.

Linking to create a robust decision-making tool. This research task will link the surrogate and uncertainty visualizations with the conceptual parameter space tools to facilitate exploration, learning, and decision making. This is the final stage seen in Figure 7 and will entail an iterative process of navigating surrogate visualizations and interrogating uncertainty visualizations using conceptual parameter widgets. Our aim is to create a robust decision-making tool that allows a user to implicitly solve the inverse problem by guiding the user to ideal outcome states and then providing the parameter ranges that define that ideal state. We define a robust decision-making tool as one that results in a decision that withstands future state fluctuations given parameter changes [53].

The primary activities for this task will be in studying how to link the parameter space and uncertainty visualizations developed in the two prior tasks in a way that allows a user to (1) explore the output space using visualization surrogates, (2) navigate to a potential ideal future state, (3) explore the uncertainties around the principle characteristics that define that state, and (4) return to the parameters that produce that uncertainty and make informed choices on parameter settings that represent both definitive requirements of a decision as well as risk tolerances and compromise. We plan to draw inspiration from tools that allow users to interact with reduced output spaces to understand parameter spaces [38]. We envision a tool with multiple views into the dataset, with each window specifically designed for each step described above, providing the right tool at the right time, along with context and annotations that we discover are helpful, and possibly necessary, for a decision.

Provenance capture and surrogate retraining. The tool will facilitate continued discovery by collecting provenance information about user actions. Provenance enables users to explain, validate, and communicate decision-making steps, building confidence and trust in the decision. Furthermore, the provenance data will allow other decision makers to recreate the analysis leading to a decision for reproducibility, for updates to the simulation data, or for any other events that might require the retraining of the surrogate models. Because Thrusts I and II provide extensive, fast computational support, we will start with just recording the series of parameter settings and allow users to "bookmark" settings to save outcomes of interest. This feature will provide undo/redo functionality, allowing users to return to a decision-making session and to share their pathway. We will study whether the series of parameter settings are enough to enable provenance; if not, we will study what other actions should be recorded, including any computational aspects from Thrusts I and II.

While interacting with our tool, the user likely will come to the limits of what our surrogate models know. Our tool will analyze the sensitivities of the surrogates described in Thrust I, and it will indicate to the user when the surrogates have a fuzzier understanding of principal characteristics or parameter subspaces that require either retraining or more simulation data. This is valuable information because it allows users to determine what new simulations to run with better precision on the parameter ranges that produce their desired outputs.

Research Tasks: Actionable decision framework

4.1.a Develop visual metaphors for conceptual parameter space understanding
4.1.b Design uncertainty visualizations to indicate system sensitivities
4.1.c Create a robust decision-making tool linking the parameter space to uncertainty visualizations
4.1.d Devise provenance capture, and construct a workflow for data updates or retraining

5 Project Evaluation and Management

5.1 Use Cases and Evaluation

We plan to evaluate the proposed methods with three DOE science applications: cosmology, Earth systems, and energy systems. We chose these three applications because they represent diverse types of ensemble data—scalar/vector fields, structured/unstructured meshes, graphs, and particles. Our understanding of the domain background based on previous collaboration with scientists will strengthen the development and evaluation of the proposed work.

Cosmology. We will use our surrogate models to enable the parameter space exploration and local variations to help scientists tune simulation parameters with visual analytics tools. Previously, Mathuriya et al. used a convolutional neural network to estimate three cosmological parameters [32], extending the earlier work of Ravanbakksh et al. [43] to larger scale and demonstrating up to three times the accuracy of earlier statistical methods. In our proposed research we will investigate how to use deep regresssors derived from GANs to act as surrogates, as well as extending our earlier work on InSituNet by incorporating domain-specific properties into the surrogate.

Earth System. We will enable fast preview visualizations for parameter space exploration. This application will build on our research on visualization surrogates for preset visualization parameters and individual timesteps. We will investigate capabilities for the exploration of time-varying ocean simulations with volume rendering and flow visualization features. We will also derive and visualize parameter sensitivities and principal characteristics.

Energy systems. Each power grid has different challenges that require robust and actionable decision-making tools. In this application area we will explore several different problems in energy systems that have wide application across this domain, such as which distribution system upgrades will increase the penetration levels of EVs, possible placement and control strategies of inverters to reach desired frequency performance targets in a transmission system, or altering the density of ozone geographically to get generation levels that could produce that.

Evaluation of visualization systems is notoriously hard [23], and the complexity and uncertainty inherent in the systems the deal with ensemble datasets compound the problem. As a first step in evaluation, we will develop our systems in close collaboration with domain scientists and will routinely request and incorporate expert feedback. As appropriate, we will perform mathematical validations and accuracy comparisons, do code reviews and software releases, and provide exemplary datasets. We will also fully document all data transformations, design decisions, and assumptions to explain our choices and allow for future updates if and when assumptions or best practices change.

An interesting artifact of our provenance collection is the ability for reproduciblilty testing. With records of user action and decisions, one can compare how different users interact with the system, explore where they make different choices, and see how close their final decisions are. An interesting study will be to see whether users come to the same decision, within a range of sensitivity. Such a study not only will test whether users can make similar decisions despite or because of the uncertainty in the system but also will help us refine the system if we see users having trouble navigating the space or coming to a final decision.

5.2 Staffing and Coordination

Potter will oversee the research and development of the entire project and lead the efforts on an actionable decision-making tool, together with Molnar. Guo and Shen will lead the thrust on surrogate models for ensemble visualization. Cappello and Peterka will lead the thrust on deriving variations. This project will also recruit three graduate students at OSU and one working year-round at NREL, and the students will participate in NREL or ANL summer programs each year to work closely with the scientists. During the project execution, monthly all-hands teleconferences will be used to report progress, present research ideas, and prioritize plans. Individual research tasks will have more regular meetings with involved staff. We will hold annual PI meetings with the kick-off meeting to be held at NREL, followed by ANL and OSU in the second and third years. We will set up a website, code repository, wiki, and mailing list for this project. We will also coordinate with the DOE SciDAC portfolio. All source code and publications will be available to the public online (see Appendix 6 for details on data management plan).

Table 1: Timeline (KP: Potter, HG: Guo, FC: Cappello, HS: Shen, TP: Peterka, JB: Bessac, SM: Molnar.)

<table>
<tr><th>Task</th><th>Y1Q1</th><th>Y1Q2</th><th>Y1Q3</th><th>Y1Q4</th><th>Y2Q1</th><th>Y2Q2</th><th>Y2Q3</th><th>Y2Q4</th><th>Y3Q1</th><th>Y3Q2</th><th>Y3Q3</th><th>Y3Q4</th></tr>
<tr><td colspan="13">Thrust I: Visualization Surrogates</td></tr>
<tr><td>2.1.a: 2.5D surrogate</td><td colspan="3">HS+TP</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2.1.b: View-dependent surrogate</td><td></td><td></td><td></td><td colspan="5">HS+JB</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2.2.a: Unstructured data</td><td colspan="5">HG+KP</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2.2.b: Particle data</td><td></td><td></td><td></td><td></td><td></td><td colspan="5">HG+FC</td><td></td><td></td></tr>
<tr><td>2.2.c: Parallelization</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">TP+HS</td></tr>
<tr><td colspan="13">Thrust II: Variation Models</td></tr>
<tr><td>3.1.a: Surrogate variations</td><td colspan="3">HS+TP</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>3.1.b: Parameter sensitivities</td><td></td><td></td><td></td><td></td><td colspan="4">TP+HG</td><td></td><td></td><td></td><td></td></tr>
<tr><td>3.2.a: Principal characteristics</td><td colspan="5">FC+JB</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>3.2.b: Parameter influence</td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">FC+JB+KP</td><td></td><td></td></tr>
<tr><td>3.2.c: Direction derivation</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">FC+JB+SM</td></tr>
<tr><td colspan="13">Thrust III: Actionable Visualization</td></tr>
<tr><td>4.1.a: Parameter space visualizations</td><td colspan="4">KP+SM+TP</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>4.1.b: Uncertainty visualizations</td><td></td><td></td><td colspan="5">KP+SM+JB</td><td></td><td></td><td></td><td></td></tr>
<tr><td>4.1.c: Decision-making tool</td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="5">KP+SM+TP</td><td></td></tr>
<tr><td>4.1.d: Provenance</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">KP+SM+HG</td></tr>
</table>

Table 2: Potential risks and pitfalls.

Probability	Impact	Risk	Mitigation
Medium	Medium	Inaccurate prediction of physical properties from the surrogates	Incorporate physics-based loss functions in training and increase training data size
Medium	Low	Insufficient training data for surrogates	Request from scientists or simulate more data with various parameters
Low	Medium	Unstable sensitivity output with dropouts	Introduce sparsity L-1 LASSO penalization terms
Low	Low	Inaccuracy of Sobol indices induced by surrogate models	Incorporate variation from surrogate model in quantifying Sobol indices
Medium	Medium	Complex parameters interaction	Separate parameter influence, and progressively add interactions
Medium	Low	Steep learning curve for actionable decision-making	Develop example use cases to ease understanding the tool

5.3 Milestones, Deliverables, and Risk Mitigation

Tables 1 and 2 list the timeline and the potential risks and mitigation plan of the project, respectively. We will disseminate our results to researchers in DOE ASCR and the broader community by integrating our models with production visualization tools including ParaView and Cinema [2]. For example, our simulation surrogate can be directly integrated with Cinema databases to enable parameter space exploration. In addition, we will integrate the developed models into the EDDA (Extreme-scale Distribution Data Analysis) software suite [1], which has been developed by PI Shen for visualizing ensemble simulations. Our tools will also be published with example data and pretrained models. All software deliverables will be open source and publicly available for download. Details on code hosting, version controlling, and software licensing are in Appendix 6.

The PIs will engage in activities beyond academic publication and software dissemination with the goal of benefiting the wider community of scientists and engineers. The PIs will organize tutorials on ensemble data analysis and visualization at annual IEEE VGTC-sponsored visualization conferences, such as the IEEE VIS and IEEE Pacific Visualization conferences. Previously, PI Potter has organized workshops on uncertainty visualization at IEEE VIS and Dagstuhl. PIs Guo and Shen have offered multiple tutorials on statistical data representation, visualization, and feature tracking, subtopics of this research at IEEE VIS 2018 and 2019, and they will continue their effort.

Appendix 1: Biographical Sketches

Attached please find biographical sketches for the following team members:
Kristin Potter (PI)
Hanqi Guo (Co-PI)
Frank Cappello (Co-PI)
Han-Wei Shen
Tom Peterka
Julie Bessac
Sam Molnar

NSF BIOGRAPHICAL SKETCH

NAME: Potter, Kristi

POSITION TITLE & INSTITUTION: Group Manager, Data, Analysis, and Visualization , National Renewable Energy Laboratory

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
University of Oregon	Eugene, Oregon	Computer Science	BS	2000
Univerity of Oregon	Eugene, Oregon	Fine Arts	BS	2000
University of Utah	Salt Lake City, Utah	Computer science - visualization	PHD	2010
University of Utah	Salt Lake City, Utah	Computer Science, Computer Graphics	MS	2023
University of Utah	Salt Lake City, Utah	Postdoctoral Researcher, Uncertainty visualization	Postdoctoral Fellow	2010 - 2011

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2021 - present	Group Manager, Data, Analysis, and Visualization , National Renewable Energy Laboratory, Computational Science Center, Golden, CO
2017 - 2021	Senior Data Visualization Scientist, National Renewable Energy Laboratory, Computational Science Center, Golden, CO
2015 - 2017	Manager, Research Support Service, University of Oregon, College of Arts and Sciences IT, Eugene, OR
2013 - 2016	Courtesy Research Associate, University of Oregon, Eugene, OR
2013 - 2015	Scientific Software Consultant, University of Oregon, College of Arts and Sciences IT, Eugene
2013 - 2014	Adjunct Graduate Faculty, Boise State University, Boise, ID
2011 - 2013	Research Computer Scientist, University of Utah, Scientific Computing and Imaging Institute, Salt Lake City, UT

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Potter K, Gerber S, Anderson EW. Visualization of uncertainty without a mean. IEEE Comput Graph Appl. 2013 Jan-Feb;33(1):75-9. PubMed PMID: 24807884.
2. Potter K, Rosen P, Johnson CR. From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches. IFIP Adv Inf Commun Technol. 2012;377:226-249. PubMed Central PMCID: PMC4319674.
3. Potter K, Kirby RM, Xiu D, Johnson CR. INTERACTIVE VISUALIZATION OF PROBABILITY AND CUMULATIVE DENSITY FUNCTIONS. Int J Uncertain Quantif. 2012;2(4):397-412. PubMed Central PMCID: PMC3609671.

4. Potter K, Wilson A, Bremer P, Williams D, Doutriaux C, Pascucci V, Johnson C. Ensemble-vis: A framework for the statistical visualization of ensemble data. 2009 IEEE International Conference on Data Mining Workshops. 2009; :233-240. isbn: 1424453844
5. Bush B, Brunhart-Lupo N, Bugbee B, Krishnan V, Potter K, Gruchalla K. Coupling visualization, simulation, and deep learning for ensemble steering of complex energy models. 2017 IEEE Workshop on Data Systems for Interactive Analysis (DSIA). 2017; :1-5. isbn: 1538621983

Other Significant Products, Whether or Not Related to the Proposed Project

1. Genton MG, Johnson C, Potter K, Stenchikov G, Sun Y. Surface boxplots. Stat (Int Stat Inst). 2014;3(1):1-11. PubMed Central PMCID: PMC4484867.
2. Potter K, Kniss J, Riesenfeld R, Johnson C. Visualizing summary statistics and uncertainty. Computer Graphics Forum. 2010; 29(3):823-832. issn: 0167-7055
3. Potter K, Wilson A, Bremer P, Williams D, Doutriaux C, Pascucci V, Johhson C. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. Journal of Physics: Conference Series. 2009; 180(1):012089. issn: 1742-6596
4. Weirs V, Fabian N, Potter K, McNamara L, Otahal T. Uncertainty in the Development and Use of Equation of State Models. International Journal for Uncertainty Quantification. 2013; 3(3). issn: 2152-5080
5. Bugbee B, Bush B, Gruchalla K, Potter K, Brunhart-Lupo N, Krishnan V. Enabling immersive engagement in energy system models with deep learning. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2019; 12(4):325-337. PMID: 1932-1864

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Symposium Chair, Large Scale Data Analysis and Visualization (LDAV) Workshop, IEEE Visualization Conference (2021, 2022).
2. Organizer, Dagstuhl Workshop on Visualization and Decision Making Design Under Uncertainty (August 2022).
3. Program committee, IEEE Visualization Conference (2015-2017, 2020-2022).
4. Associate Editor, Transactions on Visualization and Computer Graphics (2022).
5. General Chair, IEEE Visualization Conference (2024).

NSF BIOGRAPHICAL SKETCH

NAME: Guo, Hanqi

POSITION TITLE & INSTITUTION: Associate Professor, The Ohio State University

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Beijing University of Posts and Telecommunications	Beijing, Beijing	Mathematics and Applied Mathematics	BS	2009
Peking University	Beijing, Beijing	Computer Science	PHD	2014
Argonne National Laboratory	Lemont, IL	Data analysis and visualization	Postdoctoral Fellow	2014 - 2017

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2022 - present Associate Professor, The Ohio State University, Department of Computer Science and Engineering, Columbus, OH

2021 - 2022 Computer Scientist, Argonne National Laboratory, Mathematics and Computer Science Division, Lemont, IL

2019 - 2022 Scientist At-Large, University of Chicago, Consortium for Advanced Science and Engineering, Chicago, IL

2017 - 2022 Fellow, Northwestern University, Northwestern University Argonne National Laboratory Institute for Science and Engineering, Evanston, IL

2017 - 2021 Assistant Computer Scientist, Argonne National Laboratory, Mathematics and Computer Science Division, Lemont, IL

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. He W, Guo H, Shen H, Peterka T. eFESTA: Ensemble Feature Exploration with Surface Density Estimates. IEEE Transactions on Visualization and Computer Graphics. 2019; :1-1. Available from: https://ieeexplore.ieee.org/document/8525340/ DOI: 10.1109/TVCG.2018.2879866
2. He W, Wang J, Guo H, Wang K, Shen H, Raj M, Nashed Y, Peterka T. InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations. IEEE Transactions on Visualization and Computer Graphics. 2019; :1-1. Available from: https://ieeexplore.ieee.org/document/8805426/ DOI: 10.1109/TVCG.2019.2934312
3. Guo H, He W, Peterka T, Shen H, Collis S, Helmus J. Finite-Time Lyapunov Exponents and Lagrangian Coherent Structures in Uncertain Unsteady Flows. IEEE Transactions on Visualization and Computer Graphics. 2016; 22(6):1672-1682. Available from: http://ieeexplore.ieee.org/document/7422124/ DOI: 10.1109/TVCG.2016.2534560
4. Guo H, Lenz D, Xu J, Liang X, He W, Grindeanu I, Shen H, Peterka T, Munson T, Foster I. FTK: A Simplicial Spacetime Meshing Framework for Robust and Scalable Feature Tracking. IEEE Transactions on Visualization and Computer Graphics. 2021; 27(8):3463-3480. Available from: https://ieeexplore.ieee.org/document/9405464/ DOI: 10.1109/TVCG.2021.3073399

5. Hanqi Guo, Xiaoru Yuan, Jian Huang, Xiaomin Zhu. Coupled Ensemble Flow Line Advection and Analysis. IEEE Transactions on Visualization and Computer Graphics. 2013 December; 19(12):2733-2742. Available from: http://ieeexplore.ieee.org/document/6634188/ DOI: 10.1109/TVCG.2013.144

Other Significant Products, Whether or Not Related to the Proposed Project

1. Hanqi Guo, Ningyu Mao, Xiaoru Yuan. WYSIWYG (What You See is What You Get) Volume Visualization. IEEE Transactions on Visualization and Computer Graphics. 2011 December; 17(12):2106-2114. Available from: http://ieeexplore.ieee.org/document/6064975/ DOI: 10.1109/TVCG.2011.261
2. Hanqi Guo, He Xiao, Xiaoru Yuan. Scalable Multivariate Volume Visualization and Analysis Based on Dimension Projection and Parallel Coordinates. IEEE Transactions on Visualization and Computer Graphics. 2012 September; 18(9):1397-1410. Available from: http://ieeexplore.ieee.org/document/6171180/ DOI: 10.1109/TVCG.2012.80
3. Guo H, He W, Seo S, Shen H, Constantinescu E, Liu C, Peterka T. Extreme-Scale Stochastic Particle Tracing for Uncertain Unsteady Flow Visualization and Analysis. IEEE Transactions on Visualization and Computer Graphics. 2019; 25(9):2710-2724. Available from: https://ieeexplore.ieee.org/document/8419325/ DOI: 10.1109/TVCG.2018.2856772
4. Xu J, Guo H, Shen H, Raj M, Wang X, Xu X, Wang Z, Peterka T. Asynchronous and Load-Balanced Union-Find for Distributed and Parallel Scientific Data Visualization and Analysis. IEEE Transactions on Visualization and Computer Graphics. 2021; 27(6):2808-2820. Available from: https://ieeexplore.ieee.org/document/9409642/ DOI: 10.1109/TVCG.2021.3074584
5. Zhang Y, Guo H, Shang L, Wang D, Peterka T. A Multi-branch Decoder Network Approach toAdaptive Temporal Data Selection andReconstruction for Big Scientific Simulation Data. IEEE Transactions on Big Data. 2021; :1-1. Available from: https://ieeexplore.ieee.org/document/9464670/ DOI: 10.1109/TBDATA.2021.3092174

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Awardee, DOE Early Career Research Program (2022)
2. Best Paper Award, IEEE VIS 2019
3. Outstanding Postdoctoral Performance Award in Basic Research, Argonne National Laboratory, 2017
4. Program committee members for several premiere visualization conferences including IEEE VIS
5. Reviewer for several distinguished journals, including IEEE Transactions on Visualization and Computer Graphics

NSF BIOGRAPHICAL SKETCH

NAME: Cappello, Franck

POSITION TITLE & INSTITUTION: Senior Computer Scientist, Argonne National Laboratory

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
University of Paris Sud (Now called Paris Saclay)	Orsay	Computer Science (MS was called DEA in France at that time)	MS	1989
University of Paris Sud (Now called Paris Saclay)	Orsay	Computer Science	PHD	1994
University of Paris Sud (Now called Paris Saclay)	Orsay	Computers Science (HDR level: 7 years after Ph. D.)	OTH	2001

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2013 - present Senior Computer Scientist, Argonne National Laboratory, Lemont, IL
2013 - present Adjunct Research Professor , University of Illinois at Urbana Champaign, Urbana, IL
2009 - 2013 Visiting Research Professor, University of Illinois at Urbana Champaign, Urbana, IL
2003 - 2013 Senior Researcher (Directeur de Recherche), Inria, Orsay
1994 - 2003 Junior Researcher (Charge de Recherche), CNRS, Orsay

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Cappello F, Di S, Li S, Liang X, Gok A, Tao D, Yoon C, Wu X, Alexeev Y, Chong F. Use cases of lossy compression for floating-point data in scientific data sets. The International Journal of High Performance Computing Applications. 2019 July 09; 33(6):1201-1220. Available from: http://dx.doi.org/10.1177/1094342019853336 DOI: 10.1177/1094342019853336
2. Cappello F, Di S, Gok A. Fulfilling the Promises of Lossy Compression for Scientific Applications. Smoky Mountains Computational Sciences and Engineering Conference (SMC), Springer International Publishing; 2020. Available from: http://dx.doi.org/10.1007/978-3-030-63393-6_7 DOI: 10.1007/978-3-030-63393-6_7
3. Liang X, Guo H, Di S, Cappello F, Raj M, Liu C, Ono K, Chen Z, Peterka T. Toward Feature-Preserving 2D and 3D Vector Field Compression. Pacific Visualization Symposium (PacificVis); 2020; IEEE; c2020. Available from: http://dx.doi.org/10.1109/pacificvis48177.2020.6431 DOI: 10.1109/pacificvis48177.2020.6431
4. Di S, Cappello F. Fast Error-Bounded Lossy HPC Data Compression with SZ. International Conference on Parallel and Distributed Systems; 2016; IEEE; c2016. Available from: http://dx.doi.org/10.1109/ipdps.2016.11 DOI: 10.1109/ipdps.2016.11
5. Tao D, Di S, Guo H, Chen Z, Cappello F. Z-checker: A framework for assessing lossy compression of scientific data. The International Journal of High Performance Computing Applications. 2017 November 15; 33(2):285-303. Available from: http://dx.doi.org/10.1177/1094342017737147 DOI: 10.1177/1094342017737147

Other Significant Products, Whether or Not Related to the Proposed Project

1. Underwood R, Di S, Calhoun J, Cappello F. FRaZ: A Generic High-Fidelity Fixed-Ratio Lossy Compression Framework for Scientific Floating-point Data. International Parallel and Distributed Processing Symposium (IPDPS); 2020; IEEE; c2020. Available from: http://dx.doi.org/10.1109/ipdps47924.2020.00065 DOI: 10.1109/ipdps47924.2020.00065
2. Tian J, Di S, Zhao K, Rivera C, Fulp M, Underwood R, Jin S, Liang X, Calhoun J, Tao D, Cappello F. cuSZ: An Efficient GPU Based Error-Bounded Lossy Compression, Framework for Scientific Data. International Conference on Parallel Architectures and Compilation Techniques (PACT); 2020; ACM; c2020. Available from: http://dx.doi.org/10.1145/3410463.3414624 DOI: 10.1145/3410463.3414624
3. Tao D, Di S, Chen Z, Cappello F. Significantly Improving Lossy Compression for Scientific Data Sets Based on Multidimensional Prediction and Error-Controlled Quantization. International Parallel and Distributed Processing Symposium (IPDPS); 2017; IEEE; c2017. Available from: http://dx.doi.org/10.1109/ipdps.2017.115 DOI: 10.1109/ipdps.2017.115
4. Zhao K, Di S, Liang X, Li S, Tao D, Chen Z, Cappello F. Significantly Improving Lossy Compression for HPC Datasets with Second-Order Prediction and Parameter Optimization. International Symposium on High-Performance Parallel and Distributed Computing (HPDC); 2020; ACM; c2020. Available from: http://dx.doi.org/10.1145/3369583.3392688 DOI: 10.1145/3369583.3392688
5. Di S, Cappello F. Optimization of Error-Bounded Lossy Compression for Hard-to-Compress HPC Data. IEEE Transactions on Parallel and Distributed Systems. 2018 January 01; 29(1):129-143. Available from: http://dx.doi.org/10.1109/tpds.2017.2749300 DOI: 10.1109/tpds.2017.2749300

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Director of the INRIA-UIUC-ANL-BSC-JSC-Riken-UTK Joint Laboratory on Extreme Scale Computing, 2009 -- https://jlesc.github.io
2. IEEE CS Fellow Evaluation Committee, 2017-2020
3. Conference organization: many, including IEEE/ACM SC 2020 (Tech Paper Chair)
4. Member of the Editorial Board of IEEE Transactions on Parallel and Distributed Computing (2017-2019) and IEEE Transactions on Computers (2019--)
5. Steering committees of IEEE CCGRID and ACM HPDC (2014-2020)

NSF BIOGRAPHICAL SKETCH

NAME: Shen, Han-Wei

ORCID: 0000-0002-1211-2320

POSITION TITLE & INSTITUTION: Full Professor, The Ohio State University

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
National Taiwan University	Taipei	Computer Science	BS	1988
Stony Brook University	Stony Brook, New York	Computer Science	MS	1992
University of Utah	Salt Lake City, Utah	Computer Science	PHD	1998

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2012 - present Full Professor, The Ohio State University, Columbus, OH

2005 - 2012 Associate Professor, The Ohio State University, Columbus, OH

1999 - 2005 Assistant Professor, The Ohio State University, Columbus, OH

1996 - 1999 Research Scientist, NASA Ames Research Center, Mountain View, CA

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Shi N, Xu J, Wurster SW, Guo H, Woodring J, Van Roekel LP, Shen HW. GNN-Surrogate: A Hierarchical and Adaptive Graph Neural Network for Parameter Space Exploration of Unstructured-Mesh Ocean Simulations. IEEE Trans Vis Comput Graph. 2022 Jun;28(6):2301-2313. PubMed PMID: 35389867.
2. Li H, Shen HW. Local Latent Representation based on Geometric Convolution for Particle Data Feature Exploration. IEEE Trans Vis Comput Graph. 2022 Mar 15;PP PubMed PMID: 35290186.
3. Xu J, Guo H, Shen HW, Raj M, Wurster SW, Peterka T. Reinforcement Learning for Load-balanced Parallel Particle Tracing. IEEE Trans Vis Comput Graph. 2022 Feb 7;PP PubMed PMID: 35130159.
4. An Y, Shen HW, Shan G, Li G, Liu J. STSRNet: Deep Joint Space-Time Super-Resolution for Vector Field Visualization. IEEE Comput Graph Appl. 2021 Nov-Dec;41(6):122-132. PubMed PMID: 34270416.
5. Dutta S, Chen CM, Heinlein G, Shen HW, Chen JP. In Situ Distribution Guided Analysis and Visualization of Transonic Jet Engine Simulations. IEEE Trans Vis Comput Graph. 2017 Jan;23(1):811-820. PubMed PMID: 27875195.

Other Significant Products, Whether or Not Related to the Proposed Project

1. He W, Wang J, Guo H, Wang KC, Shen HW, Raj M, Nashed YSG, Peterka T. InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations. IEEE Trans Vis Comput Graph. 2020 Jan;26(1):23-33. PubMed PMID: 31425097.
2. Hazarika S, Li H, Wang KC, Shen HW, Chou CS. NNVA: Neural Network Assisted Visual

Analysis of Yeast Cell Polarization Simulation. IEEE Trans Vis Comput Graph. 2020 Jan;26(1):34-44. PubMed PMID: 31425114.

3. Wang J, Gou L, Shen HW, Yang H. DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks. IEEE Trans Vis Comput Graph. 2018 Sep 5; PubMed PMID: 30188823.
4. Wang J, Gou L, Yang H, Shen HW. GANViz: A Visual Analytics Approach to Understand the Adversarial Game. IEEE Trans Vis Comput Graph. 2018 Jun;24(6):1905-1917. PubMed PMID: 29723140.
5. Biswas A, Lin G, Liu X, Shen HW. Visualization of Time-Varying Weather Ensembles across Multiple Resolutions. IEEE Trans Vis Comput Graph. 2017 Jan;23(1):841-850. PubMed PMID: 27875198.

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Associate Editor-in-Chief, IEEE Transactions on Visualization and Computer Graphics 2019-Present; (Will be the Editor-in-Chief starting January 2023)
2. IEEE VGTC Visualization Academy inductee
3. IEEE Visualization SciVis Paper co-Chair, 2013, 2014, 2020
4. Chair, IEEE SciVis Steering Committee, 2018 – 2020
5. IEEE Visualization Executive Committee, 2016 - 2020

NSF BIOGRAPHICAL SKETCH

NAME: Peterka, Tom

POSITION TITLE & INSTITUTION: Computer Scientist, Mathematics and Computer Science Division, Argonne National Laboratory

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
University of Illinois at Chicago (UIC)	Chicago, IL	Computer Science Engineering	BS	1987
University of Illinois at Chicago (UIC)	Chicago, IL	Computer Science Engineering	MS	2003
University of Illinois at Chicago (UIC)	Chicago, IL	Computer Science Engineering	PHD	2007

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2015 - present Computer Scientist, Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL

2018 - present Scientist, University of Chicago Consortium for Advanced Science and Engineering, Chicago, IL

2014 - present Fellow, Northwestern University Argonne National Laboratory Institute of Science and Engineering, Evanston, IL

2009 - 2018 Adjunct Assistant Professor, Electronic Visualization Laboratory, University of Illinois at Chicago, Chicago, IL

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Peterka T, Nashed Y, Grindeanu I, Mahadevan V, Yeh R, Tricoche X. Foundations of multivariate functional approximation for scientific data. 2018 IEEE 8th Symposium on Large Data Analysis and Visualization (LDAV). 2018; :61-71. isbn: 1538668734
2. Grindeanu I, Peterka T, Mahadevan V, Nashed Y. Scalable, High-Order Continuity Across Block Boundaries of Functional Approximations Computed in Parallel. 2019 IEEE International Conference on Cluster Computing (CLUSTER). 2019; :1-9. isbn: 1728147344
3. Peterka T, Nashed Y, Grindeanu I, Mahadevan V, Yeh R, Lenz D. In Situ Visualization for Computational Science. Childs H, Garth C, Bennett J, editors. (to appear): Springer; 2022. Multivariate Functional Approximation of Scientific Data.
4. Lenz D, Yeh R, Mahadevan V, Grindeanu I, Peterka T. Adaptive Regularization of B-Spline Models for Scientific Data. Proceedings of International Conference on Computational Science. 2022.
5. Sun J, Lenz D, Yu H, Peterka T. MFA-DVR: Direct Volume Rendering of MFA Models. arXiv:2204.11762. 2022. Available from: https://arxiv.org/abs/2204.11762#

Other Significant Products, Whether or Not Related to the Proposed Project

1. Morozov D, Peterka T. Block-parallel data analysis with DIY2. 2016 IEEE 6th Symposium on

Large Data Analysis and Visualization (LDAV). 2016; :29-36. isbn: 1509056599
2. Nashed Y, Peterka T, Mahadevan V, Grindeanu I. Rational approximation of scientific data. International Conference on Computational Science. 2019; :18-31.
3. Yeh R, Nashed Y, Peterka T, Tricoche X. Fast Automatic Knot Placement Method for Accurate B-spline Curve Fitting. Computer-Aided Design. 2020; 128:102905. issn: 0010-4485
4. Lenz D, Marin O, Mahadevan V, Yeh R, Peterka T. Fourier-Informed Knot Placement Schemes for B-Spline Approximation. arXiv preprint arXiv:2012.04123. 2020.
5. Guo H, Lenz D, Xu J, Liang X, He W, Grindeanu I, Shen H, Peterka T, Munson T, Foster I. FTK: A Simplicial Spacetime Meshing Framework for Robust and Scalable Feature Tracking. IEEE Transactions on Visualization & Computer Graphics. 2021; (01):1-1. issn: 1077-2626

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Research collaborations and joint appointments with the leading universities in the Chicago area.
2. Multiple award recognitions for early-career research from program sponsors and academic conferences.
3. Services to the professional community organizing conferences and serving on technical program committees.
4. Organizing strategic planning workshops and generating reports proposing priority research directions for program sponsors.
5. Mentoring early-career researchers including postdocs and students.

NSF BIOGRAPHICAL SKETCH

NAME: Bessac, Julie

POSITION TITLE & INSTITUTION: Assistant Computational Statistician, Argonne National Laboratory

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Universite de Rennes 1	Rennes, Brittany	Mathematics	BS	2008
Universite de Rennes 1	Rennes, Brittany	Probability and Statistics	MS	2011
Universite de Rennes 1	Rennes, Brittany	Applied Mathematics	PHD	2014
Argonne National Laboratory	Lemont, IL	Applied Mathematics and Statistics	Postdoctoral Fellow	2014 - 2017

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2017 - present Assistant Computational Statistician, Argonne National Laboratory, Lemont, IL

2018 - present Scientist at large, Consortium for Advanced Science and Engineering of The University of Chicago, Chicago, IL

2012 - 2014 Teaching Fellow, ENSAI, Bruz

2011 - 2012 Teaching Fellow, Universite de Rennes 1, Rennes

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Bessac J, Constantinescu E, Anitescu M. Stochastic simulation of predictive space-time scenarios of wind speed using observations and physical model outputs. Annals of Applied Statistics. 2018; 12(1):432-458.
2. Bessac J, Naveau P. Forecast score distributions with imperfect observations. Advances in Statistical Climatology, Meteorology and Oceanography. 2021; 7(2):53-71. Available from: https://ascmo.copernicus.org/articles/7/53/2021/
3. Constantinescu E, Petra N, Bessac J, Petra C. Statistical treatment of inverse problems constrained by differential equations-based models with stochastic terms. SIAM/ASA Journal on Uncertainty Quantification. 2020; 8(1):170-197.
4. Hart J, Bessac J, Constantinescu E. Global sensitivity analysis for statistical model parameters. SIAM/ASA Journal on Uncertainty Quantification. 2019; 7(1):67-92.
5. Rudi J, Bessac J, Lenzi A. Parameter Estimation with Dense and Convolutional Neural Networks Applied to the FitzHugh–Nagumo ODE. Mathematical and Scientific Machine Learning. 2020; 145:1-28. Available from: https://proceedings.mlr.press/v145/rudi22a.html

Other Significant Products, Whether or Not Related to the Proposed Project

1. Bessac J, Ailliot P, Monbet V. Gaussian linear state-space model for wind fields in the North-East Atlantic. Environmetrics. 2015; 26(1):29-38.

2. Bessac J, Monahan A, Christensen H, Weitzel N. Stochastic Parameterization of Subgrid-Scale Velocity Enhancement of Sea Surface Fluxes. Monthly Weather Review. 2019; 147(5):1447-1469. Available from: https://doi.org/10.1175/MWR-D-18-0384.1 DOI: 10.1175/MWR-D-18-0384.1
3. Bessac J, Christensen H, Endo K, Monahan A, Weitzel N. Scale-aware space-time stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. Journal of Advances in Modeling Earth Systems. 2020; 13(4). Available from: https://doi.org/10.1029/2020MS002367
4. Krasowska D, Bessac J, Underwood R, Calhoun J, Di S, Cappello F. Exploring Lossy Compressibility through Statistical Correlations of Scientific Datasets. 2021 7th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-7). 2021; :47-53. Available from: doi: 10.1109/DRBSD754563.2021.00011
5. Krock M, Bessac J, Stein ML, Monahan AH. Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails. Weather and Climate Extremes. 2022 June; 36. Available from: https://doi.org/10.1016/j.wace.2022.100438

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Organizing committee SIAM Conference on Mathematics of Planet Earth (SIAM-MPE22).
2. CyberTraining on ``Big Data + High-Performance Computing + Atmospheric Sciences" (NSF Initiative on Workforce Development for Cyberinfrastructure, selective participation) organized by the University of Maryland in Baltimore County, 2020
3. HPC Innovation Excellence Award from Hyperion Research - "Risk and Resiliency of Infrastructure, Southeastern USA, for AT&T", 2019
4. Symposia organization for SIAM Conference on Uncertainty Quantification 2020 and 2022
5. Teaching Bootcamp "Introduction to the Statistics of Spatial Data", The University of Chicago, 2018.

NSF BIOGRAPHICAL SKETCH

NAME: Molnar, Samantha

POSITION TITLE & INSTITUTION: Visualization Scientist , National Renewable Energy Laboratory

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
University of Denver	Denver, CO	Physics	BS	2015
University of Colorado, Boulder	Boulder, CO	Computer Science	MS	2018
University of Colorado, Boulder	Boulder, CO	Computer Science	PHD	2021

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2022 - present Visualization Scientist , National Renewable Energy Laboratory, Golden, CO
2017 - 2022 Graduate Research Assistant, National Renewable Energy Laboratory, Golden, CO
2017 - 2017 Research Assistant, University of Colorado, Boulder, Boulder, CO
2014 - 2015 Research Assistant, University of Denver, Denver, CO
2014 - 2014 Undergraduate Research Assistant, University of Colorado, Boulder, Boulder, CO
2012 - 2014 Research Assistant, University of Denver, Denver, CO

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Molnar S, Gruchalla K. Visualizing Electrical Power Systems as Flow Fields. Proceedings of the Workshop on Visualisation in Environmental Sciences. 2018.
2. Sanderson N, Shugerman E, Molnar S, Meiss J, Bradley E. Computational Topology Techniques for Characterizing Time- Series Data. The Sixteenth International Symposium on Intelligent Data Analysis. 2017.

Other Significant Products, Whether or Not Related to the Proposed Project

1. Molnar S, Bradley E, Gruchalla K. Oscillatory spreading and inertia in power grids. Chaos. 2021 Dec;31(12):123103. PubMed PMID: 34972338.
2. Vann JM, Molnar SL, Calbi MM. Equilibration processes during gas uptake inside narrow pores. Phys Chem Chem Phys. 2015 May 21;17(19):13021-7. PubMed PMID: 25913800.

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Supervisor for four student research projects, topics involving information theory and complexity, development of economic model for demand response strategies, and developing benchmarks for visualization applications on HPC systems.
2. Development of user studies that assess visualization techniques and interactions.
3. Presenter at multiple conferences, including IEEE's Conference on Visualization, EuroVis.

Appendix 2: Current and Pending Support

Attached please find current and pending support for the following team members:
Kristin Potter (PI)
Hanqi Guo (Co-PI)
Frank Cappello (Co-PI)
Han-Wei Shen
Tom Peterka
Julie Bessac
Sam Molnar

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Potter, Kristi

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: Data Driven Mobility Equity Evaluation Platform

   Proposal/Award Number (if available):

   Source of Support: NREL Internal LDRD

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2021

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $26,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.1

   Overall Objectives: Seed funding to explore an equity metric for transportation mobility.

   Statement of Potential Overlap: This work looks at equity metrics in transportation modeling and while this is not directly related to this proposal, equity is an important aspect to decision making, paricularly in the energy sector.

2. Project/Proposal Title: ESIF Kestrel Preparation

   Proposal/Award Number (if available):

   Source of Support: NREL Indirect (EERE)

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2021

   Project/Proposal Support End Date (if available): 09/2023

   Total Award Amount (including Indirect Costs): $100,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed

Year	Person-months per year committed
2022	0.5

Overall Objectives: Prepare the datacenter's newest supercomputer for test and use.

Statement of Potential Overlap: No overlaps or synergies.

3. Project/Proposal Title: DAV Group Manager

   Proposal/Award Number (if available):

   Source of Support: NREL Internal Overhead

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 06/2021

   Project/Proposal Support End Date (if available):

   Total Award Amount (including Indirect Costs): $1,000,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	3.6
2023	3.6
2024	3.6
2025	3.6

Overall Objectives: Development of a workplan for the data, analysis, and visualization group.

Statement of Potential Overlap: No overlaps or synergies.

4. Project/Proposal Title: HPC Data Centric Analytics

   Proposal/Award Number (if available):

   Source of Support: NREL Indirect Investments

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2020

   Project/Proposal Support End Date (if available): 09/2025

   Total Award Amount (including Indirect Costs): $1,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	2.4
2023	1.25
2024	1.25
2025	1.25

Overall Objectives: Improving integration of HPC and Cloud in the NREL datacenter.

Statement of Potential Overlap: No overlaps or synergies.

5. Project/Proposal Title: Research Data Initiative

   Proposal/Award Number (if available):

   Source of Support: NREL Indirect Investments

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2020

   Project/Proposal Support End Date (if available): 09/2025

   Total Award Amount (including Indirect Costs): $2,000,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.8
2023	0.25
2024	0.25
2025	0.25

Overall Objectives: Design, support, and operations of a research data infrastructure pipeline.

Statement of Potential Overlap: Overlap with this proposal includes the management of research data.

6. Project/Proposal Title: Circular Economy Life Cycle Assessment and Visualization

   Proposal/Award Number (if available):

   Source of Support: NREL Internal LDRD

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available): 10/2019

Project/Proposal Support End Date (if available): 09/2022

Total Award Amount (including Indirect Costs): $900,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.5

Overall Objectives: This project investigates how wind turbines can be recycled to bring about a more circular economy for wind farms.

Statement of Potential Overlap: My role on this project is to design visualizations to both understand our modeling results as well as for communication to stakeholders. There is no direct overlap with this proposal, however my experience in this area will be intrinsically informative in future visualization design projects.

7. Project/Proposal Title: Behind the Meter Storage and Battery

   Proposal/Award Number (if available):

   Source of Support: DOE Vehicle Technologies Office

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2018

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $600,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.8

Overall Objectives: Thousands of simulations for understanding behind the meter storage for battery and PV adoption.

Statement of Potential Overlap: This work involves modeling and analyzing behind the meter storage for energy efficient buildings. There is no direct overlap to this proposal, however knowledge gained from this activity may result in synergistic activities.

8. Project/Proposal Title: ESIF Data Visualization

   Proposal/Award Number (if available):

Source of Support: NREL Indirect (EERE)

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available):

Project/Proposal Support End Date (if available):

Total Award Amount (including Indirect Costs): $5,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.1
2023	0.1
2024	0.1
2025	0.1

Overall Objectives: Building visualization capabilities for the ESIF Insight Center.

Statement of Potential Overlap: This work includes the software engineering of visualization tools for deployment in our visualization ecosystem. Some of these may be used in this proposal and work from this proposal may be generalized for broader use in the Insight Center.

9. Project/Proposal Title: ESIF Visualization AOP Management

Proposal/Award Number (if available):

Source of Support: NREL Indirect (EERE)

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available):

Project/Proposal Support End Date (if available):

Total Award Amount (including Indirect Costs): $5,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	2.5
2023	2
2024	2

Year	Person-months per year committed
2025	2

Overall Objectives: Development of a workplan for the data, analysis, and visualization group.

Statement of Potential Overlap: No overlaps or synergies.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

   Proposal/Award Number (if available):

   Source of Support: DOE Advanced Scientific Computing Research Program

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 08/2022

   Project/Proposal Support End Date (if available): 08/2025

   Total Award Amount (including Indirect Costs): $2,700,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.1
2023	2.4
2024	2.4
2025	2.3

   Overall Objectives: This objective of this project is to research how to use uncertainty visualization techniques that have been proven to be perceptually sound on simplistic data to complex scientific data relevant to the DOE mission space.

   Statement of Potential Overlap: This proposal.

2. Project/Proposal Title: Visualizing and Communicating Uncertainty in Complex Scientific Data

   Proposal/Award Number (if available):

   Source of Support: DOE Advanced Scientific Computing Research Program

   Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.1
2023	2.4
2024	2.4
2025	2.3

Overall Objectives: This proposed research investigates novel paradigms to address the challenges in using ensembles for decision making by creating linkages between the output and parameter spaces that specifically enable decisions.

Statement of Potential Overlap: There is no research overlap with this proposal, however research findings on uncertainty visualization best practices and techniques may be synergistic.

PI/co-PI/Senior Personnel: Guo, Hanqi

NSF ID: 000787814@nsf.gov

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: Multidimensional Parameter-Space Feature Tracking, Analysis, and Visualization

   Proposal/Award Number (if available):

   Source of Support: DOE-ECRP

   Primary Place of Performance: The Ohio State University

   Project/Proposal Support Start Date (if available): 08/2022

   Project/Proposal Support End Date (if available): 07/2027

   Total Award Amount (including Indirect Costs): $750,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	1
2024	1
2025	1
2026	1
2027	1

   Overall Objectives: The proposed research will establish a paradigm for understanding parameter space and redefining feature tracking in multidimensional parameter spaces.

   Statement of Potential Overlap: There is no overlap with the current proposal.

2. Project/Proposal Title: Modern Data Analytics for the Large Gamma-Ray Spectrometers: GRETINA/GRETA and Gammasphere via Machine Learning and Optimization

   Proposal/Award Number (if available): 0000261364

   Source of Support: DOE-NP

   Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2021

Project/Proposal Support End Date (if available): 09/2023

Total Award Amount (including Indirect Costs): $1,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1.2
2023	1.2

Overall Objectives: The objective of this project is to attack two open and connected problems in the utilization of large gamma-ray arrays, namely (1) to make significant improvements in gamma-ray tracking as it applies to GRETINA/GRETA, and (2) to develop a new framework for the construction of nuclear level schemes which are the major end products of these world-class spectrometers.

Statement of Potential Overlap: There is no overlap with the current proposal.

3. Project/Proposal Title: RAPIDS2: A SciDAC Institute for Computer Science, Data, and Artificial Intelligence

Proposal/Award Number (if available): 34367

Source of Support: DOE-ASCR

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2020

Project/Proposal Support End Date (if available): 09/2025

Total Award Amount (including Indirect Costs): $28,750,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	1.2
2022	1.2
2023	1.2
2024	1.2

Overall Objectives: The objective of the RAPIDS institute is to assist Office of Science application teams in overcoming computer science and data challenges in the use of DOE supercomputing resources to achieve scientific breakthroughs.

Statement of Potential Overlap: There is no overlap with the current proposal.

4. Project/Proposal Title: III: Medium: Collaborative Research: Deep Learning for In Situ Analysis and Visualization

   Proposal/Award Number (if available): 1955764

   Source of Support: NSF

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 06/2020

   Project/Proposal Support End Date (if available): 05/2022

   Total Award Amount (including Indirect Costs): $225,316

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2020	1.2
2021	1.2
2022	1.2

   Overall Objectives: This project researches deep learning algorithms for future-generation visualization techniques when the size of simulation output exceeds the storage capacity.

   Statement of Potential Overlap: There is no overlap with the current proposal.

5. Project/Proposal Title: Nested Task-Parallel Workflows for Scientific Applications

   Proposal/Award Number (if available): 57L93

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 09/2017

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $750,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	2.4

   Overall Objectives: Developing a task-parallel in situ workflow system that will expose a task-parallel user interface

Statement of Potential Overlap: There is no overlap with the current proposal.

6. Project/Proposal Title: Coupling Approaches for Next-Generation Architectures (CANGA)

   Proposal/Award Number (if available): 27176

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 07/2017

   Project/Proposal Support End Date (if available): 07/2022

   Total Award Amount (including Indirect Costs): $1,637,125

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2018	1.8
2019	1.8
2020	1.8
2021	1.8
2022	1.2

   Overall Objectives: The CANGA project is an integrated approach to developing new coupler services for the Earth system.

   Statement of Potential Overlap: There is no overlap with the current proposal.

7. Project/Proposal Title: CODAR: Co-Design Center for Online Data Analysis and Reduction at the Exascale

   Proposal/Award Number (if available): 1006243

   Source of Support: DOE-ECP

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 11/2016

   Project/Proposal Support End Date (if available): 06/2023

   Total Award Amount (including Indirect Costs): $8,471,854

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2018	6
2019	6
2020	6
2021	7.2
2022	4.8

Overall Objectives: This project is a co-design center focused on online data analysis and reduction at the exascale that includes approximation, reduction, assimilation, calibration, data mining, and statistical analysis.

Statement of Potential Overlap: There is no overlap with the current proposal.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Femtoscale Imaging of Nuclei Using Exascale Platforms

   Proposal/Award Number (if available):

   Source of Support: DOE-SciDAC-NP

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 10/2022

   Project/Proposal Support End Date (if available): 09/2027

   Total Award Amount (including Indirect Costs): $6,250,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	1.8
2024	1.8
2025	1.8
2026	1.8
2027	1.8

Overall Objectives: To extract the maximum amount of information on the quark and gluon tomography of nucleons and nuclei from the high-energy scattering data obtained at existing and future facilities.

Statement of Potential Overlap: There is no overlap with the current proposal.

2. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

Proposal/Award Number (if available):

Source of Support: DOE-ASCR

Primary Place of Performance: The Ohio State University

Project/Proposal Support Start Date (if available): 09/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1
2023	1
2024	1

Overall Objectives: Establish a new paradigm to help understand ensemble simulation data from various DOE science applications.

Statement of Potential Overlap: This proposal.

3. Project/Proposal Title: Implicit Continuous Representations for Visualization of Complex Data

Proposal/Award Number (if available):

Source of Support: DOE-ASCR

Primary Place of Performance: The Ohio State University

Project/Proposal Support Start Date (if available): 09/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1

Year	Person-months per year committed
2023	1
2024	1

Overall Objectives: The proposed research investigates how to accurately and reliably visualizecomplex data consisting of multiple non-uniform domains and/or data types.

Statement of Potential Overlap: There is no overlap with the current proposal.

4. Project/Proposal Title: Improving Projections of AMOC and Collapse Through advanced Simulations (ImPACTS)

   Proposal/Award Number (if available):

   Source of Support: DOE-SciDAC-BER

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 08/2022

   Project/Proposal Support End Date (if available): 07/2027

   Total Award Amount (including Indirect Costs): $14,055,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2
2024	2
2025	1.8
2026	1.85
2027	1.85

   Overall Objectives: Improved representations of the Atlantic Meridional Ocean Circulation (AMOC), including their properties, dynamics, variability, and coupling to other processes in the E3SM

   Statement of Potential Overlap: There is no overlap with the current proposal.

5. Project/Proposal Title: Modeling the dynamics of QBO and developing Observational Diagnostic of QBO and its effect on tropospheric hydrological cycle

   Proposal/Award Number (if available):

Source of Support: DOE-SciDAC-BER

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 07/2027

Total Award Amount (including Indirect Costs): $15,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	**Person-months per year committed**
2023	2.79
2024	2.68
2025	2.58
2026	2.48

Overall Objectives: Develop the gravity wave drag (GWD) representation in the E3SM model.

Statement of Potential Overlap: There is no overlap with the current proposal.

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Cappello, Franck

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: ROCCI: Integrated Cyberinfrastructure for In Situ Lossy Compression Optimization Based on Post Hoc Analysis Requirements

   Proposal/Award Number (if available): ZUE9HKT2CLC9

   Source of Support: 040100 NSF RESEARCH & RELATED ACTIVIT

   Primary Place of Performance: University of Chicago

   Project/Proposal Support Start Date (if available): 09/2021

   Project/Proposal Support End Date (if available): 08/2024

   Total Award Amount (including Indirect Costs): $319,998

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	0.5
2022	0.5
2023	0.5
2024	0.5

   Overall Objectives: The objective is to develop an integrated cyberinfrastructure for in situ lossy compression optimization based on post hoc analysis requirements

   Statement of Potential Overlap: none.

2. Project/Proposal Title: Exploration of Lossy Data Compression for Seismic Imaging Application

   Proposal/Award Number (if available): PRJ1008127

   Source of Support: ARAMCO

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 12/2019

   Project/Proposal Support End Date (if available): 06/2022

Total Award Amount (including Indirect Costs): $1,026,853

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	3.1
2022	3.1

Overall Objectives: Develop lossy compressors for seismic data.

Statement of Potential Overlap: none.

3. Project/Proposal Title: CODAR: Co-Design Center for Online Data Analysis and Reduction at the Exascale

   Proposal/Award Number (if available): PRJ1006243

   Source of Support: ECP

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 11/2016

   Project/Proposal Support End Date (if available): 06/2023

   Total Award Amount (including Indirect Costs): $8,471,854

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	2.4
2022	2.4
2023	2.4

   Overall Objectives:

   Statement of Potential Overlap:

4. Project/Proposal Title: VeloC: Very Low Overhead Transparent Multilevel Checkpoint/Restart/Sz: Fast, Effective, Parallel Error-bounded Exascale Lossy Compression for Scientific Data

   Proposal/Award Number (if available): PRJ1007629

   Source of Support: ECP

   Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 10/2016

Project/Proposal Support End Date (if available): 06/2023

Total Award Amount (including Indirect Costs): $3,833,595

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	4.8
2022	4.8
2023	4.8

Overall Objectives:

Statement of Potential Overlap:

5. Project/Proposal Title: Exasky Computing the Sky at Extreme Scales

Proposal/Award Number (if available): PRJ1006083

Source of Support: ECP

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 02/2016

Project/Proposal Support End Date (if available): 06/2023

Total Award Amount (including Indirect Costs): $7,705,142

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1.2
2023	1.2

Overall Objectives: Develop cosmology codes.My role: develop compression schemes for cosmology.

Statement of Potential Overlap: None

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

Proposal/Award Number (if available):

Source of Support: DOE Office of Science

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	**Person-months per year committed**
2022	1.1
2023	1.1
2024	1.1

Overall Objectives: Establish a new paradigm to help understand ensemble simulation data fromvarious DOE science application

Statement of Potential Overlap: none

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Shen, Han-Wei

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: AI institute for intelligent cyberinfrastructure with computational learning in the environment (ICICLE)

   Proposal/Award Number (if available):

   Source of Support: NSF

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 11/2021

   Project/Proposal Support End Date (if available): 10/2026

   Total Award Amount (including Indirect Costs): $20,000,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.5
2023	0.5
2024	0.5
2025	0.5
2026	0.5

   Overall Objectives: The goal of this project is to develop a cyber infrastructure to support plug and play AI.

   Statement of Potential Overlap: None

2. Project/Proposal Title: RAPIDS2: A SciDAC institute for computer science, data, and artificial intelligence

   Proposal/Award Number (if available):

   Source of Support: US Department of Energy

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 09/2020

Project/Proposal Support End Date (if available): 09/2025

Total Award Amount (including Indirect Costs): $550,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1
2023	1
2024	1
2025	1

Overall Objectives: The goal of this project is to support DOE applications with a wide spectrum of computer science research including data understanding, I/O, machine learning, and performance modeling.

Statement of Potential Overlap: Minimum as RAPIDS has a much broader goal and scientific mission.

3. Project/Proposal Title: III: Medium: Collaborative Research: Deep learning for in situ analysis and visualization

Proposal/Award Number (if available):

Source of Support: NSF

Primary Place of Performance: OSU

Project/Proposal Support Start Date (if available): 06/2020

Project/Proposal Support End Date (if available): 05/2024

Total Award Amount (including Indirect Costs): $715,314

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1
2023	1
2024	1

Overall Objectives: This NSF grant is to develop visualization surrogate to support ensemble simulations.

Statement of Potential Overlap: none

4. Project/Proposal Title: Visual analytics for large scale scientific ensemble datasets

   Proposal/Award Number (if available):

   Source of Support: Los Alamos National Lab

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 02/2018

   Project/Proposal Support End Date (if available): 02/2023

   Total Award Amount (including Indirect Costs): $600,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1
2023	1

   Overall Objectives: This project is to develop visual analytics methods for analyzing ensemble scientific data.

   Statement of Potential Overlap: none

5. Project/Proposal Title: RIDIR: Survey data recycling: New analytic framework, integrated database, and tools for cross-national social, behavioral and economic research

   Proposal/Award Number (if available):

   Source of Support: NSF

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 09/2017

   Project/Proposal Support End Date (if available): 08/2022

   Total Award Amount (including Indirect Costs): $1,402,259

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.1

   Overall Objectives: This project is to develop a web base data portal of survey data for sociologists.

   Statement of Potential Overlap: none.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Uncertainty-Aware Multiscale Analysis and Visualization of Time-Varying Multivariate Particle Data

   Proposal/Award Number (if available):

   Source of Support: DOE

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 10/2022

   Project/Proposal Support End Date (if available): 09/2025

   Total Award Amount (including Indirect Costs): $600,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	1
2024	1
2025	1

   Overall Objectives: The objective of this proposal is to develop an uncertainty-aware time-varying multivariate particle data analysis and visualization pipeline based on deep learning neural networks.

   Statement of Potential Overlap: none

2. Project/Proposal Title: Visual analytics of query-based document retrieval to support information seeking for evidence-based practice

   Proposal/Award Number (if available):

   Source of Support: Washington University (NIH)

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 09/2022

   Project/Proposal Support End Date (if available): 08/2027

   Total Award Amount (including Indirect Costs): $1,018,538

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed

Year	Person-months per year committed
2023	1.35
2024	1.35
2025	1.35
2026	1.35
2027	1.35

Overall Objectives: This project is to develop an information retrieval system for biomedical text analysis and search.

Statement of Potential Overlap: none

3. Project/Proposal Title: Modeling the dynamics of QBO and developing Observational Diagnostic of QBO and its effect on tropospheric hydrological cycle

   Proposal/Award Number (if available):

   Source of Support: Argonne National Lab

   Primary Place of Performance: OSU

   Project/Proposal Support Start Date (if available): 09/2022

   Project/Proposal Support End Date (if available): 08/2026

   Total Award Amount (including Indirect Costs): $400,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.55
2024	0.73
2025	0.64
2026	0.55

Overall Objectives: Develop further the gravity wave drag (GWD) representation in the E3SMmodel.

Statement of Potential Overlap: none

4. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

   Proposal/Award Number (if available):

Source of Support: DOE (NREL lead)

Primary Place of Performance: OSU

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	**Person-months per year committed**
2023	1
2024	1
2025	1

Overall Objectives: The objective of this proposal is to develop visualization surrogates to support decision making and parameter exploration for ensemble simulations.

Statement of Potential Overlap: none

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Peterka, Tom

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: Triple Convergence of HPC, BD, and AI through ASCR In Situ Workflow Tools

   Proposal/Award Number (if available):

   Source of Support: DOE Office of Science

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 10/2020

   Project/Proposal Support End Date (if available): 09/2023

   Total Award Amount (including Indirect Costs): $2,400,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	1.8
2022	1.8
2023	1.8

   Overall Objectives: This project will deliver workflow tools that combine the triple convergence of HPC, big data, and AI for the in situ management, analysis, and preservation of scientific data.

   Statement of Potential Overlap: None

2. Project/Proposal Title: CODAR: A Co-design Center for Online Data Analysis and Reduction at the Exascale

   Proposal/Award Number (if available):

   Source of Support: DOE Office of Science

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 10/2017

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $16,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2018	1.2
2019	1.2
2020	1.2
2021	1.2
2022	1.2

Overall Objectives: CODAR is a co-design center focused on online data analysis and reduction at the exascale that includes approximation, reduction, assimilation, calibration, data mining, and statistical analysis.

Statement of Potential Overlap: None

3. Project/Proposal Title: MFA: Multivariate Functional Approximation

Proposal/Award Number (if available):

Source of Support: DOE Office of Science

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2017

Project/Proposal Support End Date (if available): 08/2022

Total Award Amount (including Indirect Costs): $2,500,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2018	6
2019	6
2020	6
2021	6
2022	6

Overall Objectives: This project explores a fundamentally different kind of data model that will conserve resources while improving data understanding and sharing. The approach being taken redefines scientific data in a mathematical model that will be more efficient to communicate, store, and analyze than the original form.

Statement of Potential Overlap: None

4. Project/Proposal Title: HEP Data Analytics in HPC

Proposal/Award Number (if available):

Source of Support: DOE Office of Science

Primary Place of Performance: Fermi National Accelerator Laboratory

Project/Proposal Support Start Date (if available): 08/2017

Project/Proposal Support End Date (if available): 07/2022

Total Award Amount (including Indirect Costs): $10,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2018	3
2019	3
2020	3
2021	3
2022	3

Overall Objectives: This project addresses the science needs of the HEP community through the development and deployment of new HPC tools and analysis algorithms for HEP.

Statement of Potential Overlap: None

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Implicit Continuous Representations for Visualization of Complex Data

Proposal/Award Number (if available):

Source of Support: DOE Office of Science

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	5
2024	5
2025	5

Overall Objectives: The proposed research investigates how to accurately and reliably visualize complex data consist- ing of multiple non-uniform domains and/or data types.

Statement of Potential Overlap: None

2. Project/Proposal Title: MINCE: Model Integration and Numerical Coupling in E3SM

Proposal/Award Number (if available):

Source of Support: DOE SciDAC-5

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2022

Project/Proposal Support End Date (if available): 06/2027

Total Award Amount (including Indirect Costs): $15,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2.4
2024	2.4
2025	2.4
2026	2.4
2027	2.4

Overall Objectives: Our principal objective is to significantly improve the physical and numerical coupling in E3SM by developing new schemes to reduce the coupling error from multiple sources and increase the stability of the fully coupled system. These new schemes will be based on mathematically rigorous analysis of the error propagation and numerical stability of various temporal integration and coupling schemes in the presence of spatial remapping errors for current and future configurations of E3SM.

Statement of Potential Overlap: None

3. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

Proposal/Award Number (if available):

Source of Support: DOE Office of Science

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	**Person-months per year committed**
2023	3.6
2024	3.6
2025	3.6

Overall Objectives: The goal of this research is to design a decision-making methodology that identifies the range of the simulation parameter space producing desired outcomes of an end-user, stakeholder, or policymaker.

Statement of Potential Overlap: None

4. Project/Proposal Title: Advanced Feedback-Driven Uncertainty Quantification for Neutrino Science on HPC Platforms

Proposal/Award Number (if available):

Source of Support: DOE SciDAC-5

Primary Place of Performance: Fermi National Accelerator Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 07/2027

Total Award Amount (including Indirect Costs): $10,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	**Person-months per year committed**
2023	1.2
2024	1.2

Year	Person-months per year committed
2025	1.2
2026	1.2
2027	1.2

Overall Objectives: We propose to develop new “end-to-end” loop-capable infrastructure to support this effort for systematic variation of experimental models and data.

Statement of Potential Overlap: None

5. Project/Proposal Title: Next-Generation Ecosystem for Real-Time Digital Twins of Particle Accelerators

Proposal/Award Number (if available):

Source of Support: DOE SciDAC-5

Primary Place of Performance: SLAC National Accelerator Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 07/2027

Total Award Amount (including Indirect Costs): $10,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	1.2
2024	1.2
2025	1.2
2026	1.2
2027	1.2

Overall Objectives: Develop a next-generation software ecosystem for producing and deploying digital twins for particle accelerators that can be evaluated in real time.

Statement of Potential Overlap: None

6. Project/Proposal Title: Stratosphere/Troposphere Interactions and Dynamics in E3SM via RRM (STrIDER)

Proposal/Award Number (if available):

Source of Support: DOE SciDAC-5

Primary Place of Performance: Sandia National Laboratories

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 07/2027

Total Award Amount (including Indirect Costs): $14,100,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	3.6
2024	3.6
2025	3.6
2026	3.6
2027	3.6

Overall Objectives: The specific mechanisms responsible for the observed teleconnections between the QBO and MJO are not known. The proposed research will advance the building of E3SM into an invaluable tool for studying the QBO-MJO teleconnection and its impacts for weather and climate.

Statement of Potential Overlap: None

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Bessac, Julie

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: Develop an Artificial Intelligence Based Risk Analysis Tool with Future Extreme Weather Prediction to Analyze its Impact on Buildings and Occupants

   Proposal/Award Number (if available): PRJ1006714

   Source of Support: Argonne LDRD

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 01/2022

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $50,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.2

   Overall Objectives: Extracting changing weather and community characteristics leading to high-risk building occupancy.

   Statement of Potential Overlap: No overlap.

2. Project/Proposal Title: Impacts of Climate Change on Intensity, Duration, and Frequency of Drought Events at Different Geographic Scales

   Proposal/Award Number (if available): PRJ1009776

   Source of Support: Argonne LDRD

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 09/2021

   Project/Proposal Support End Date (if available): 09/2023

   Total Award Amount (including Indirect Costs): $250,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed

Year	Person-months per year committed
2021	2.4
2022	2.4

Overall Objectives: Detection and characterization of spatiotemporal extreme events such as drought.

Statement of Potential Overlap: No overlap.

3. Project/Proposal Title: Machine-Learning Accelerated Simulations for Forecasting, Data Assimilation, and Extreme Events

   Proposal/Award Number (if available):

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 09/2021

   Project/Proposal Support End Date (if available): 09/2024

   Total Award Amount (including Indirect Costs): $2,059,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	4.2
2023	4.2
2024	4.2

Overall Objectives: Scientific machine learning embedding physics knowledge and statistics for data assimilation with extreme events.

Statement of Potential Overlap: No overlap.

4. Project/Proposal Title: Frameworks, Algorithms and Scalable Technologies for Mathematics (FASTMath) SciDAC Institute

   Proposal/Award Number (if available): PRJ1008604

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 09/2020

Project/Proposal Support End Date (if available): 09/2025

Total Award Amount (including Indirect Costs): $4,055,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	2.4
2022	0.6
2023	0.6
2024	0.6
2025	0.6

Overall Objectives: FASTMath institute.

Statement of Potential Overlap: No overlap.

5. Project/Proposal Title: Climate Resiliency and Adaptation

   Proposal/Award Number (if available): PRJ1006756

   Source of Support: Private sponsor: AT&T Inc.

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 08/2020

   Project/Proposal Support End Date (if available): 12/2022

   Total Award Amount (including Indirect Costs): $1,350,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	1.8
2022	1.8

   Overall Objectives: Scientific and statistical assessment of climate-related risks for telecommunication infrastructures.

   Statement of Potential Overlap: No overlap.

6. Project/Proposal Title: VeloC: Very Low Overhead Transparent Multilevel Checkpoint/Restart/Sz: Fast, Effective, Parallel Error-bounded Exascale Lossy Compression for Scientific Data

Proposal/Award Number (if available): PRJ1007629

Source of Support: DOE-ECP

Primary Place of Performance: Argonne National Laboratory

Project/Proposal Support Start Date (if available): 09/2016

Project/Proposal Support End Date (if available): 09/2022

Total Award Amount (including Indirect Costs): $3,833,595

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2021	2.4
2022	2.4

Overall Objectives: Multi-level checkpoint-restart runtime for HPC supercomputing infrastructures and large-scale data centers.

Statement of Potential Overlap: No overlap.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: LODES: Learning and Optimization for Decarbonized and Resilient Energy Systems

   Proposal/Award Number (if available):

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 10/2022

   Project/Proposal Support End Date (if available): 09/2027

   Total Award Amount (including Indirect Costs): $15,000,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2.4
2024	2.4

Year	Person-months per year committed
2025	2.4
2026	2.4
2027	2.4

Overall Objectives: Optimal decisions for networked systems subject to exogenous nonstationaryuncertainties in a scalable manner, by using both data-driven and physics-basedmodels, under the scarcity of observables and historical data, and with various physicalconstraints.

Statement of Potential Overlap: There is no overlap with existing and pending fundings.

2. Project/Proposal Title: SEnsor-Enabled Data-driven Multifaceted Optimization for REsilient integrated systems (SEEDMORE)

Proposal/Award Number (if available):

Source of Support: DOE-ASCR

Primary Place of Performance: University of Michigan

Project/Proposal Support Start Date (if available): 10/2022

Project/Proposal Support End Date (if available): 09/2027

Total Award Amount (including Indirect Costs): $5,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2.4
2024	2.4
2025	2.4
2026	2.4
2027	2.4

Overall Objectives: The objective of the proposed SEEDMORE center is to develop data-driven multifaceted machinelearning (ML) and optimization tools and approaches for complex system design and operations,needed in the development of next-generation resilient integrated infrastructure systems.

Statement of Potential Overlap: No overlap

3. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

   Proposal/Award Number (if available):

   Source of Support: DOE-ASCR

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 08/2022

   Project/Proposal Support End Date (if available): 08/2025

   Total Award Amount (including Indirect Costs): $2,700,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2
2024	2
2025	2

   Overall Objectives: Establish a new paradigm to help understand ensemble simulation data fromvarious DOE science applications.

   Statement of Potential Overlap: There is no overlap with existing or pending proposals.

4. Project/Proposal Title: Femtoscale Imaging of Nuclei using Exascale Platforms

   Proposal/Award Number (if available):

   Source of Support: DOE-ASCR-NP

   Primary Place of Performance: Argonne National Laboratory

   Project/Proposal Support Start Date (if available): 08/2022

   Project/Proposal Support End Date (if available): 09/2027

   Total Award Amount (including Indirect Costs): $13,750,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	2.4
2024	2.4
2025	2.4

Year	Person-months per year committed
2026	2.4
2027	2.4

Overall Objectives: Develop numerical codes for a variety of theoretical frameworks for hadrontomography that specializes in utilizing state-of-the-art QCD theory to convert QCFs into event level data using accelerated computing environments.

Statement of Potential Overlap: No overlap

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Molnar, Samantha

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: C2C Data Pipeline

   Proposal/Award Number (if available):

   Source of Support: NREL Internal Investments

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 04/2022

   Project/Proposal Support End Date (if available): 03/2023

   Total Award Amount (including Indirect Costs): $75,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.3

   Overall Objectives: Development of a multi-faceted data pipeline service that supports data and information collaboration among diverse stakeholders in integrated modeling projects.

   Statement of Potential Overlap: Developing tools and techniques for easier collaboration and exploration of data in large modeling projects for decision making at all projects levels from modelers to key project stakeholders.

2. Project/Proposal Title: Collaborative Visualization and Analytics

   Proposal/Award Number (if available):

   Source of Support: NREL Internal LDRD

   Primary Place of Performance: National Renewable Energy Laboratory

   Project/Proposal Support Start Date (if available): 10/2020

   Project/Proposal Support End Date (if available): 09/2022

   Total Award Amount (including Indirect Costs): $500,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed

Year	Person-months per year committed
2022	2.5

Overall Objectives: Develop collaborative visualization tools to support control room operations.

Statement of Potential Overlap: Identifying the strengths and limitations of various visual representations used in power systems to build and expand upon visualization theory in the power system application space.

3. Project/Proposal Title: HPC Data Centric Analysis

Proposal/Award Number (if available):

Source of Support: NREL Internal Investments

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available): 10/2020

Project/Proposal Support End Date (if available): 09/2025

Total Award Amount (including Indirect Costs): $1,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	1
2023	1
2024	1
2025	1

Overall Objectives: Improved integration of HPC and Cloud for the ESIF Data center.

Statement of Potential Overlap: Developing analysis and visualization workflows to increase access and provide better services for decision-making at scale.

4. Project/Proposal Title: ESIF AOP Visualization

Proposal/Award Number (if available):

Source of Support: NREL Internal (EERE)

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available):

Project/Proposal Support End Date (if available):

Total Award Amount (including Indirect Costs): $5,000,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	4.5
2023	4.5
2024	4.5
2025	4.5

Overall Objectives: ESIF user facility visualization management and operations.

Statement of Potential Overlap: Designing and implementing heterogeneous visualization environments using novel technologies to support decision making for diverse stakeholders ranging from scientists to policy makers.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Actionable Intelligent Visual Analytics of Ensembles

Proposal/Award Number (if available):

Source of Support: DOE Office of Science ASCR

Primary Place of Performance: National Renewable Energy Laboratory

Project/Proposal Support Start Date (if available): 08/2022

Project/Proposal Support End Date (if available): 08/2025

Total Award Amount (including Indirect Costs): $2,700,000

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2022	0.15
2023	3.6
2024	3.6
2025	3.45

Overall Objectives: This objective of this project is to research how to use uncertainty visualization techniques that have been proven to be perceptually sound on simplistic data to complex scientific data relevant to the DOE mission space.

Statement of Potential Overlap: This proposal.

Appendix 3: Bibliography and References Cited

References

[1] Gravity visualization database. https://sites.google.com/site/gravityvisdb/edda.

[2] James P. Ahrens, Sébastien Jourdain, Patrick O'Leary, John Patchett, David H. Rogers, and Mark R. Petersen. An image-based approach to extreme scale in situ visualization and analysis. In Trish Damkroger and Jack J. Dongarra, editors, *International Conference for High Performance Computing, Networking, Storage and Analysis, SC 2014, New Orleans, LA, USA, November 16-21, 2014*, pages 424–434. IEEE Computer Society, 2014.

[3] Kieran Alden, Jason Cosgrove, Mark Coles, and Jon Timmis. Using Emulation to Engineer and Understand Simulations of Biological Systems. *IEEE/ACM transactions on Computational Biology and Bioinformatics*, 17(1):302–315, 2018.

[4] Ann S. Almgren, John B. Bell, Mike J. Lijewski, Zarija Lukić, and Ethan Van Andel. Nyx: A massively parallel AMR code for computational cosmology. *The Astrophysical Journal*, 765(1):39, 2013.

[5] A. Biswas, G. Lin, X. Liu, and H. W. Shen. Visualization of time-varying weather ensembles across multiple resolutions. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016.

[6] Emanuele Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.

[7] Mike Bostock, Shan Carter, and Archie Tse. Is it better to rent or buy. *The New York Times*, 2014.

[8] George Casella and Christian P Robert. Monte carlo statistical methods, 1999.

[9] George Casella, Christian P. Robert, and Martin T. Wells. Generalized accept-reject sampling schemes. *Lecture Notes-Monograph Series*, 45:342–347, 2004.

[10] Michaël Defferrard, Nathanaël Perraudin, Tomasz Kacprzak, and Raphael Sgier. DeepSphere: Towards an Equivariant Graph-Based Spherical CNN. *arXiv preprint arXiv:1904.05146*, 2019.

[11] Kjiersten Fagnan, Youssef Nashed, Gabriel Perdue, Daniel Ratner, Arjun Shankar, and Shinjae Yoo. Data and models: A framework for advancing AI in science. Technical report, US DOE Office of Science (SC) (United States), 2019.

[12] Florian Ferstl, Kai Bürger, and Rüdiger Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, 2015.

[13] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 1050–1059. JMLR.org, 2016.

[14] Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agnès Lagnoux. Sensitivity indices for multivariate outputs. *arXiv preprint arXiv:1303.3574*, 2013.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[16] Robert B. Gramacy. *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC, Boca Raton, Florida, 2020. http://bobby.gramacy.com/surrogates/.

[17] Attila Gyulassy, Valerio Pascucci, Tom Peterka, and Robert Ross. The parallel computation of morse-smale complexes. In *2012 IEEE 26th International Parallel and Distributed Processing Symposium*, pages 484–495. IEEE, 2012.

[18] Joseph L Hart, Julie Bessac, and Emil M Constantinescu. Global sensitivity analysis for statistical model parameters. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):67–92, 2019.

[19] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[20] Subhashis Hazarika, Haoyu Li, Ko-Chih Wang, Han-Wei Shen, and Ching-Shan Chou. NNVA: neural network assisted visual analysis of yeast cell polarization simulation. *IEEE Trans. Vis. Comput. Graph.*, 26(1):34–44, 2020.

[21] Wenbin He, Junpeng Wang, Hanqi Guo, Ko-Chih Wang, Han-Wei Shen, Mukund Raj, Youssef S. G. Nashed, and Tom Peterka. InSituNet: Deep image synthesis for parameter space exploration of ensemble simulations. *IEEE Trans. Comput. Graph.*, 26(1):23–33, 2020.

[22] Michael A Heroux, Lois McInnes, David Bernholdt, Anshu Dubey, Elsa Gonsiorowski, Osni Marques, J David Moulton, Boyana Norris, Elaine Raybourn, Satish Balay, et al. Advancing scientific productivity through better scientific software: Developer productivity and software sustainability report. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2020.

[23] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 25(1):903–913, 2018.

[24] Bertrand Iooss and Paul Lemaître. A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems*, pages 101–122. Springer, 2015.

[25] Aasim Kamal, Parashar Dhakal, Ahmad Y Javaid, Vijay K Devabhaktuni, Devinder Kaur, Jack Zaientz, and Robert Marinier. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, 24(5):861–890, 2021.

[26] Sergei Kucherenko and Shugfang Song. Derivative-based global sensitivity measures and their link with sobol'sensitivity indices. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 455–469. Springer, 2016.

[27] Ismael Lemhadri, Feng Ruan, and Rob Tibshirani. Lassonet: Neural networks with feature sparsity. In *International Conference on Artificial Intelligence and Statistics*, pages 10–18. PMLR, 2021.

[28] Liya Li and Han-Wei Shen. Image-based streamline generation and rendering. *IEEE Trans. Vis. Comput. Graph.*, 13(3):630–640, 2007.

[29] Luyi Li, Zhenzhou Lu, and Danqing Wu. A new kind of sensitivity index for multivariate output. *Reliability Engineering & System Safety*, 147:123–131, 2016.

[30] Wei Liu, Alexey V Fedorov, Shang-Ping Xie, and Shineng Hu. Climate impacts of a weakened atlantic meridional overturning circulation in a warming climate. *Science advances*, 6(26):eaaz4876, 2020.

[31] Anthony Lopez, Billy Roberts, Donna Heimiller, Nate Blair, and Gian Porro. Us renewable energy technical potentials. a gis-based analysis. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2012.

[32] Amrita Mathuriya, Deborah Bard, Peter Mendygral, Lawrence Meadows, James Arnemann, Lei Shao, Siyu He, Tuomas Kärnä, Diana Moise, Simon J Pennycook, et al. CosmoFlow: Using Deep Learning to Learn the Universe at Scale. In *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 819–829. IEEE, 2018.

[33] Barbara Millet, Alberto Cairo, Sharanya J Majumdar, Carolina Diaz, Scotney D Evans, and Kenneth Broad. Beautiful visualizations slain by ugly facts: Redesigning the national hurricane center's 'cone of uncertainty' map, 2020.

[34] Dmitriy Morozov and Tom Peterka. Block-parallel data analysis with diy2. In *2016 IEEE 6th Symposium on Large Data Analysis and Visualization (LDAV)*, pages 29–36. IEEE, 2016.

[35] Dmitriy Morozov, Tom Peterka, Hanqi Guo, Mukund Raj, Jiayi Xu, and Han-Wei Shen. Iexchange: Asynchronous communication and termination detection for iterative algorithms. In *11th IEEE Symposium on Large Data Analysis and Visualization, LDAV 2021, New Orleans, LA, USA, October 25, 2021*, pages 12–21. IEEE, 2021.

[36] Max D Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[37] Oliver Nalbach, Elena Arabadzhiyska, Dushyant Mehta, Hans-Peter Seidel, and Tobias Ritschel. Deep shading: Convolutional neural networks for screen space shading. *Comput. Graph. Forum*, 36(4):65–78, 2017.

[38] Daniel Orban, Daniel F. Keefe, Ayan Biswas, James Ahrens, and David Rogers. Drag and track: A direct manipulation interface for contextualizing data instances within a continuous parameter space. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):256–266, 2019.

[39] Nathanaël Perraudin, Michaël Defferrard, Tomasz Kacprzak, and Raphael Sgier. DeepSphere: Efficient Spherical Convolutional Neural Network with HEALPix Sampling for Cosmological Applications. *Astronomy and Computing*, 27:130–146, 2019.

[40] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning Mesh-Based Simulation with Graph Networks. *arXiv preprint arXiv:2010.03409*, 2020.

[41] Kristin Potter, Joe Kniss, Richard Riesenfeld, and Chris R Johnson. Visualizing summary statistics and uncertainty. In *Computer Graphics Forum*, volume 29, pages 823–832. Wiley Online Library, 2010.

[42] Kristin Potter, Andrew T. Wilson, Peer-Timo Bremer, Dean N. Williams, Charles M. Doutriaux, Valerio Pascucci, and Chris R. Johnson. Ensemble-Vis: A framework for the statistical visualization of ensemble data. In *Proceedings of IEEE International Conference on Data Mining (ICDM) Workshops, Miami, Florida, USA, 6 December 2009*, pages 233–240, 2009.

[43] Siamak Ravanbakhsh, Junier B Oliva, Sebastian Fromenteau, Layne Price, Shirley Ho, Jeff G Schneider, and Barnabás Póczos. Estimating cosmological parameters from the dark matter distribution. In *ICML*, pages 2407–2416, 2016.

[44] Saman Razavi and Hoshin V Gupta. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. theory. *Water Resources Research*, 52(1):423–439, 2016.

[45] Marissa Renardy, Tau-Mu Yi, Dongbin Xiu, and Ching-Shan Chou. Parameter uncertainty quantification using surrogate models applied to a spatial model of yeast mating polarization. *PLOS Computational Biology*, 14(5):1–26, 05 2018.

[46] Todd Ringler, Mark Petersen, Robert L. Higdon, Doug Jacobsen, Philip W. Jones, and Mathew Maltrud. A multi-resolution approach to global ocean modeling. *Ocean Modelling*, 69:211–232, 2013.

[47] Andrea Saltelli, Stefano Tarantola, and KP-S Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

[48] Neng Shi, Jiayi Xu, Skylar W Wurster, Hanqi Guo, Jonathan Woodring, Luke P Van Roekel, and Han-Wei Shen. Gnn-surrogate: A hierarchical and adaptive graph neural network for parameter space exploration of unstructured-mesh ocean simulations. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[49] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

[50] Théophile Terraz, Alejandro Ribes, Yvan Fournier, Bertrand Iooss, and Bruno Raffin. Melissa: Large scale in transit sensitivity analysis avoiding intermediate files. In *Proceedings of the international conference for high performance computing, networking, storage and analysis*, pages 1–14, 2017.

[51] Jarke J. van Wijk. Image based flow visualization. *ACM Trans. Graph.*, 21(3):754, 2002.

[52] Detlof Von Winterfeldt. Bridging the gap between science and decision making. *Proceedings of the National Academy of Sciences*, 110(Supplement 3):14055–14061, 2013.

[53] Warren E Walker, Robert J Lempert, and Jan H Kwakkel. Deep uncertainty. *Delft University of Technology*, 1(2), 2012.

[54] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9, 2016.

[55] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE transactions on visualization and computer graphics*, 13(6):1129–1136, 2007.

[56] Sinan Xiao, Zhenzhou Lu, and Pan Wang. Multivariate global sensitivity analysis for dynamic models based on energy distance. *Structural and multidisciplinary optimization*, 57(1):279–291, 2018.

[57] Jiayi Xu, Hanqi Guo, Han-Wei Shen, Mukund Raj, Xueyun Wang, Xueqiao Xu, Zhehui Wang, and Tom Peterka. Asynchronous and load-balanced union-find for distributed and parallel scientific data visualization and analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):2808–2820, 2021.

Appendix 4: Facilities and Other Resources

National Renewable Energy Laboratory

NREL is home to an array of readily available facilities and resources for researchers. The following resources will be utilized for the proposed work:

High Performance Computing and Capabilities

The High Performance Computing User Facility within NREL's Energy Systems Integration Facility is home to NREL's high performance computing system, high performance computing data center, and the Insight Visualization Center. This facility is the first and one of the only megawatt-scale research facilities in the United States, enabling integration studies at full power and actual load levels in real-time simulation.

High Performance Computing System Eagle is NREL's current high-performance computing system dedicated to advancing renewable energy and energy efficiency technologies. As a replacement for NREL's prior supercomputer, Peregrine, the new system was put into production use in January 2019. The system is a Linux cluster that uses a fast InfiniBand network. It is composed of 2,114 interconnected "compute nodes" with 4,228 Intel Skylake processors and a total of 76,104 cores—along with 14 petabytes of high-speed data storage. The peak performance of Eagle is approximately 8 petaflops, or 8 million billion floating point (mathematical) operations per second. Eagle was designed and built by Hewlett-Packard Enterprise (HPE) and has an innovative warm water liquid-cooling system that allows waste heat to be captured for reuse.

Kestrel is NREL's newest HPC and is expected to come online during fiscal year 2023. The system is built with 2436 total compute nodes comprised of: 2304 "general compute nodes" of type dual socket Intel Xeon Sapphire Rapids each with 52-core processors 256 GB of DDR5 memory and 256 of those nodes with 1.92 TB local storage, 132 "accelerated compute nodes" of dual socket AMD Genoa (96-core) processors with 384GB of memory, 4 NVIDIA H100 SXM GPUs each with 80 GB Memory, and 2 x 1.6 TB NVMe local storage, 10 "fat memory nodes of dual socket Intel Xeon Sapphire Rapids (52-core) processors with 2 TB DDR5, and 10 "data analysis and visualization (DAV) nodes" made up of Dual socket Intel Xeon Sapphire Rapids (52-core) processors each with 256 GB DDR5 memory, 2 NVIDIA A40 GPUs, and 2 x 3.84 TB NVMe local storage. This system will be connected using HPE's high-performance Slingshot 11 fabric in a dragonfly topology and in the case of the accelerated nodes, each node will have two 200 Gb/s Network Interface Cards (NICs) to the interconnect. Red Hat Enterprise Linux will be used for login and data analysis and visualization nodes and Rocky Linux for the compute nodes. The ClusterStor Lustre file system will be used for the Parallel File System (PFS) and scratch, with PFS having 68 PB and 200 GB/s of IOR bandwidth and scratch using a hybrid flash-disk configuration providing a total of 27 petabytes (PB) of capacity with 354 gigabytes (GB)/s of IOR bandwidth. The home file system will also use CluterStor with 1.2 PB of capacity.

Insight Visualization Center NREL's Insight Center at the Energy Systems Integration Facility combines state-of-the-art visualization and collaboration tools to promote knowledge discovery for energy systems. The Insight Center comprises two spaces: the Collaboration Room and the Visualization Room. Altogether, five graphics servers with 268-1024 GB of RAM and Quadro-class GPUs. These servers are connected to the NREL HPC data center with 10GbE connections. The insight servers also have 40GbE connections to a dedicated 84TB Insight Center storage system.

The Collaboration Room provides multiple workspaces in which researchers and partners from all disciplines of science and engineering can interactively visualize highly complex, large-scale data, systems, and operations in real time. Simulation data may span from the atomistic scale of new materials to the atmospheric boundary scales used in multi-turbine array simulations. Observational data span from the nanostructures of biomass pretreatments to the complexity of the North American electrical power grid. The Collaboration Room utilizes a variety of immersive three-dimensional visualization and interaction technologies that bring these scientific data to the human scale. These technologies include a custom-built, state-of-the-art tracked stereoscopic immersive virtual environment composed of six projectors that illuminate two surfaces–a wall and a floor–creating a 10 megapixel (5m ¿ 2.5m ¿ 2m) immersive volume. The use of a 4K 240Hz projector allows for a 40 megapixel display which can track two users simultaneously. In addition, this space includes a suite of VR head-mounted displays including HTC Vives and XTAL 5K, a suite of AR head-mounted displays including Hololens, MagicLEAP, and ODG headsets, and a commercial desktop VR system including a 27” head-tracked passive-stereo display with a 3D stylus.

Complementing the Collaboration Room, the Visualization Room boosts the exchange of ideas among NREL researchers and their collaborating partners through a collection of 2D display paradigms. The primary visualization format is a large-scale, 100 megapixel 2D multi-touch display wall which allow for unlimited, simultaneous touches from fingers, hands, pens, and Codice 2D markers. In addition, the room contains an HD, multi-touch display wall which allows for unlimited, simultaneous touches. Large-scale, high-resolution two-dimensional visual imagery can be used to effectively convey information and illustrate research findings to stakeholders and visitors. Using the displays available in the Visualization Room, researchers and others now have the visual real estate to lay out significant amounts of data to analyze large-scale simulations, ensembles of simulations, and highly detailed visual analytics.

Cloud Services NREL’s Advanced Computing Operations group facilitates Cloud services via Amazon Web Services across the laboratory. This allows PIs to have ready access to project-specific cloud architectures, data streaming environments, Edge device support, collaborative environments, environments for API publication, data migration and data pipelines, machine learning pipelines and automation workflows, large scale computing for modeling and simulation, web app deployment, and other opportunities.

AWS Cloud services readily available to researchers include Athena, S3, and Glue Analysis databases, Quantum Computing through Bracket service, Elastic Map Reduce service, GreenGrass for IoT, specialized databases, and other cutting edge services. In addition to these, Cloud storage, backups, and archives are available for PIs to store, replicate, and manage their data. NREL’s mass storage system has shifted entirely to AWS Glacier which offers long-term storage of data associated with projects running on HPC and/or Cloud. Data is quickly accessibly as well as stored securely and efficiently.

On-Premises Cloud, Vermilion Vermilion provides an OpenStack based on-premises Cloud using familiar tools and environments such as the OpenHPC operating environment and SLURM scheduling. Additionally, Vermillion offers capabilities geared toward more modern, web-scale, and Internet of Things (IoT) models such as Hive, Spark, Hadoop, Kubernetes, and longer running analysis tools that do not fit in a classic HPC environment. The system is composed of 92 ”compute nodes” with 184 AMD EPYC Rome processors and a total of 11,776 cores and 8 “accelerated nodes” with 16 AMD EPYC Rome processors and one Nvidia A100 GPU with 420 TB of Ceph based storage.

Databases and Other Resources

NREL Data Catalog The NREL Data Catalog is where descriptive information (i.e., metadata) is maintained about public data resulting from federally funded research conducted by the National Renewable Energy Laboratory (NREL) researchers and analysts. Catalog submissions for each dataset will then be distributed to highly visible, public data platforms such as osti.gov, data.gov, and other scientific data hubs to ensure NREL research data is widely available to others. Available: https://thesource.nrel.gov/data-management/data-storage-sharing.html

The Ohio State University

The critical requirement will be high performance computing resources to run fine-grained simulations and per- form analyses, including on accelerators like GPUs, and for certain studies, visualization devices. The PI has access to resources in the department of Computer Science and Engineering (CSE), Mechanical and Aerospace Engineering (MEA), the Ohio Supercomputing Center (OSC), as well as major NSF and DOE resources.

The relevant resources at OSC include: (1) the Oakley HP Intel Xeon cluster, which features total 8,328 cores on 694 nodes with 12 cores and 48 GB DRAM per node, and (2) Glenn IBM Cluster 1350, which consists of 658 System x3455 compute nodes with dual socket, quad core 2.5 GHz Opteron and 24 GB DRAM per node. Many of the nodes at both the clusters have different types of GPUs, and plans are to purchase MIC nodes soon.

We can also leverage two research clusters provided through NSF Research Infrastructure (RI) program. One cluster resides in the Department of Computer Science and Engineering. It consists of 64 nodes with dual Intel EM64T processors on the 32 nodes and dual AMD Opteron processors on the other 32 nodes. The nodes are connected through InfiniBand DDR. The other cluster resides in the Department of Biomedical Informatics. It consists of 64 nodes with AMD dual Opteron processors, which are connected via InfiniBand SDR. 8 nodes in the first cluster and 32 nodes in the second cluster have GPUs. Another relevant resource within CSE is in Shen's visualization research lab, which, besides workstations and the GPU, includes a 4x4 display panel wall, where each panel is a 40-inch flat LCD screen. Through PI's projects with DOE and NSF, students either already have or will likely be able to access resources from XSEDE, as well as the supercomputers at Argonne National Laboratory, Oak Ridge National Laboratory, and Lawrence Berkeley National Laboratory.

Argonne National Laboratory

Argonne National Laboratory has access to three main facilities for computing resources; data and networking; and data analytics and visualization.

ALCF. ALCF resources include leadership-class supercomputers, visualization clusters, advanced data storage systems, high-performance networking capabilities, and a wide variety of software tools and services to help facility users achieve their science goals.

Theta is an 11.69-petaflops supercomputer based on Intel processors and interconnect technology, an advanced memory architecture, and a Lustre-based parallel file system, all integrated by Cray's HPC software stack. Theta is helping bridge the gap between Mira and Argonne's next extreme-computing system, Aurora.

Cooley, the ALCF's visualization cluster, enables users to analyze and visualize large-scale datasets. Equipped with state-of-the-art graphics processing units (GPUs), Cooley helps users gain deeper insights into simulations and data generated on the facility's supercomputers.

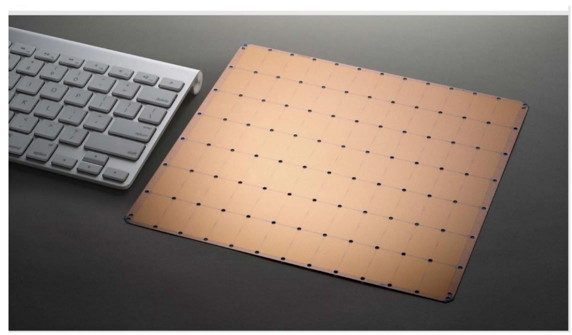


Figure 8: Cerebras wafer-scale AI accelerator.

The Argonne AI-Testbed provides an infrastructure of next-generation AI-accelerator machines. It aims to help evaluate usability and performance of machine learning based high-performance computing applications running on these accelerators. Currently, the AI-Testbed consists of Cerebras, SambaNova, GraphCore, and Groq systems. See https://ai.alcf.anl.gov/#systems. Argonne operates a **Cerebras CS-1 system**, which builds on the pioneering Wafer Scale Engine, the largest and fastest AI processor ever built. With 1.2 trillion transistors and over 400,000 cores, the Cerebras system enables 100–1000-fold improvements over conventional GPUs, depending on workload. Access to these systems will enable project investigators to examine how algorithms and methods need to evolve to benefit from such next-generation systems. Early experiments with GraphCore, for example, have demonstrated order-of-magnitude performance improvements over NVIDIA GPUs.

The ALCF's storage comprises both disk and tape. In early 2021 the ALCF is also deploying two new storage systems, **Grand** and **Eagle**, providing a total of 200 PB of storage capacity. The systems initially will connect to existing ALCF supercomputers: Theta, based on the Cray XC40-AC, and Cooley, based on the Cray CS-300.

Polaris is a 35—45-petaflop CPU/GPU hybrid resource developed in collaboration with Hewlett Packard Enterprise (HPE). It will allow users to prepare and scale their codes, and ultimately their science for future exascale systems. Polaris will comprise of more than 500 nodes of 4 GPUs each. The system will be fully integrated with the 200-petabyte file system ALCF deployed in 2020, with increased data sharing support.

Figure 9: Aurora exascale supercomputer.

Aurora is Argonne's first exascale computer. Installation of Aurora began at Argonne in mid-2022. Aurora combines more than 10,000 Intel-outfitted blades into an HPE Cray EX supercomputer. Each compute blade has two Sapphire Rapids Xeon CPUs and six Ponte Vecchio GPUs, integrated into HPE's Cray EX architecture with Slingshot networking. Aurora's revolutionary architecture will support machine learning and data science workloads alongside traditional modeling and simulation workloads.

Currently, the ALCF provides resource allocations on its resources, based on competitive proposals through the INCITE program, the ASCR Leadership Computing Challenge, the ALCF Director's Discretionary program, the ALCF Data Science Program, and the Aurora Early Science Program.

JLSE. The Joint Laboratory for System Evaluation (JLSE) is a collaboration between the Mathematics and Computer Science (MCS) Division and the Argonne Leadership Computing Facility (ALCF) with the aim of evaluating future high-performance computing platforms, developing system software, and measuring power/energy. JLSE hosts more than two dozen different cutting-edge hardware platforms, including Intel development GPU cards (code names XeHP and DG1), as well as NVIDIA A100 and RTX8000 cards. The post-Moore architecture lab in the Mathematics and Computer Science Division, which is part of JLSE, hosts an FPGA testbed.

LCRC. Apart from the ALCF facilities, Argonne also hosts several other resources in the Laboratory Computing Resource Center (LCRC) (https://www.lcrc.anl.gov). **Bebop** is the newest addition to the computational power of LCRC. It has 1,024 public nodes, with 128 GB (Intel Broadwell)/96 GB (Intel Knights Landing) of memory on each node. Blues, the second computing cluster, has approximately 350 public nodes, with 64 GB (Intel Sandy Bridge)/128 GB (Intel Haswell) of memory on each node.

Appendix 5: Equipment

Relevant equipment is included in the discussion of Facilities and Other Resources (Appendix 4).

Appendix 6: Data Management Plan

We recognize the value of preserving a record of our research and sharing the software and data products that result. We believe that well-crafted data management is essential to advancing the quality and pace of scientific research, and it is an integral part of our planning.

We will use public web-based tools designed primarily for software development to manage both software and data artifacts. High-quality systems for building software, managing dependencies, controlling versions, tracking issues, scripting, plotting, building wikis, project pages, and writing documentation are all readily available today. The use of these tools is essential for guaranteeing reproducible results, provided that they are designed into the project from the outset and not as an afterthought. The automation of many of these tasks not only helps others validate our research, it streamlines and organizes our own work and saves us time as well. Sound data management is truly an integral part of an organized scientific method.

This section outlines our strategy for using modern web tools for curating data and software. Preserved artifacts entail source code, documentation, publications, plotting scripts, utilities, test data sets, metadata and notes, and software build scripts. The content, format, and tools managing these data products throughout their life cycle are described below.

Availability. All of the following will be freely available to the public, except in cases where proprietary, business confidential, or personally identifiable information must be protected.

Data products. All source code will be developed and hosted in publicly available repositories such as github, bitbucket, or gitlab. During development, a simple public-domain copyright statement will accompany the software, and when an official release is deemed appropriate, a Berkeley Source Distribution (BSD) style open source license will be sought in accordance with ASCR and Argonne policy. We will use Cmake to maintain the building and installation of our software and to install any needed dependencies. Documentation of the source code will be generated by using Doxygen and will be hosted alongside the code in a public location. Github, for example, provides public web page hosting for this purpose.

Publications in portable document format (PDF) will also reside alongside the software and documentation. Scripts in R or Python to generate plots in publications and the datasets needed to execute them will be similarly curated. Shell commands to generate plots, including their input arguments, will also be stored in Bash or Python scripts, automating their execution and ensuring their reproducibility. We have found Ipython Notebook to also be extremely useful for managing Python scripts for this purpose. Papers that are in preparation will not be publicly available until they have been cleared by the laboratory and submitted for review or have been accepted for publication or have appeared in print or online in final or preprint form.

Additional data sets used in our research consist of small test datasets for validating proof of concept and larger test data for scalable performance testing. In the first case, we will include such sample datasets in the software repository to the extent that there is sufficient space available. In the second case, we will host datasets using laboratory or computing facility resources such as network-mounted and parallel file systems that are maintained by the individual facilities.

Project metadata consists of meeting notes, design documents, project reports, and research highlights. We will use the wiki feature of the software repository to maintain such documents.

Software. All source code will be developed and hosted in publicly available repositories such as GitHub, Bitbucket, or GitLab. During development, a public-domain copyright statement will accompany the software. When an official release is deemed appropriate, a BSD-style open-source license will be sought in accordance with ASCR and Argonne policy. We will use Cmake and Spack to maintain the building and installation of our software and to install any needed dependencies. Any and all dependencies on external third-party software will be clearly listed in the README title page of the software repository. Documentation of the source code will be generated using Doxygen and Sphinx and hosted alongside the code in a public location. GitHub, for example, provides public web page hosting for this purpose. Continuous integration testing will be employed for all software in a "release" category of deployment. In all stages of the software lifecyle—from planning through development to testing and maintenance—best practices will be employed as outlined in the Better Scientific Software (BSSw) initiative [22].

Publications. Publications in portable document format (pdf) will also reside alongside the software and documentation. Scripts to generate plots in publications and the datasets needed to execute them will be similarly curated. Shell commands to generate plots, including their input arguments, will also be stored in scripts or electronic notebooks, automating their execution and ensuring their reproducibility. Papers in preparation will not be publicly available until they have been cleared by the laboratory and submitted for review or have been accepted for publication, or have appeared in print or online in final or preprint form. To accelerate dissemination, we will make manuscripts submitted for peer review available as preprints on Argonne's website or on the https://arxiv.org/ archive when publishers allow. Any peer-reviewed publications produced by this project will be made available in a publicly accessible, open-access database freely available to the public via the Department of Energy's Office of Scientific and Technical Information (OSTI) (https://www.osti.gov/). This database is permanent, open access, free archive for the scientific output of DOE laboratories.

Protection. We will ensure that none of the data contains personally identifiable, proprietary, or confidential information. No copyrighted information will be shared without written consent of the copyright holder. No excerpts of scientific data from our collaborating researchers will be shared without their written consent. In all cases, laboratory review and release processes will be followed to ensure that privacy is protected.

FAIR Digital Objects. Wilkinson et al. argue that FAIR (findable, accessible, interoperable, and reusable) [54] principles ought to be applied to all digital objects—algorithms and workflows—in additional to conventional datasets. Fagnan et al. [11] extend this definition to include AI models and the data needed to train them. We intend that our data management plan as outlined above exemplifies FAIR principles for our software, implicit continuous models, test and training datasets, publications, and other metadata related to reproducing the results of our publications and experiments. Specifically:

- Findable: All datasets, models, and software will be accompanied by metadata and documentation to allow their use by others, and will be located in public repositories that are searchable online.
- Accessible: Datasets and models will be available in standard formats readable by publicly available software tools. Trusted mainstream repositories will be used for access. Software will be written in standard languages with any dependencies clearly indicated and installable through easy to use tools such as PyPI and Spack.
- Interoperable: Data, metadata, models, and software are retained in broadly accessible formats and languages. Publicly available converters exist to convert formats when necessary.
- Reusable: All data, models, and software are versioned with clear provenance. Permissive open-source software licenses and open access of unlicensed models and data will be employed.

Appendix 7: Recruitment and Retention of Students and Early-Stage Investigators

PI Potter has an extensive track record of recruiting and retaining young investigators as demonstrated by the large percentage of junior-level staff under her purview as group manager, and has been very active in the development of novel activities to support career development. A major endeavor includes the Visualization Summer Camp, co-created by Potter. The summer camp is a small workshop that brings together pre-tenure faculty to create a cohort of people in the same stage of their career. The camp is designed to create a peer-mentoring group that will help junior faculty navigate the challenges of moving from life as graduate students or postdocs to faculty positions that require not only success in research, but also teaching, managing staff and students, working with budgets, and navigating politics and other roadblocks in getting tenure. In addition to supporting junior faculty, Potter has worked to create similar support mechanisms within the DOE; PI Potter has helped create an early-career track at the DOE Computer Graphics Forum, and hopes to develop more early-career activities once the forum is back in person, and she is helping to develop PI training seminars within NREL to demystify the process of landing funding at the lab. Potter also regularly mentors student interns at all levels and, along with Molnar, mentor the graduate students supported by this proposal during the summer months. Specifically, over the summer, graduate students will embed with the visualization team and work closely with domain experts to ensure the research developed herein will be directly applicable to the aforementioned DOE science problems. PI Potter will also mentor junior-level staff member Molnar on managing graduate students, career development, and the granting processes.

Ohio State has a strong track record of mentoring under- and graduate students. Up to date Shen has graduated 26 Ph.D. and numerious MS students. Through his mentorship, his students have worked with application scientists and visualization researchers in various national laboratories and federal agencies. Through summer internships and project collaborations, this research will create opportunities for graduate and undergraduate students to participate in key research initiatives with top scientists. We will also broaden participation of underrepresented groups through our various activities in Ohio. Both Shen and Guo at Ohio State have been actively involved in various diversity and outreach activities in the past. These activities include joining summer programs to educate pre-college students and advise them in research, giving guest lectures on data science and visualization to teachers from local school districts and mentoring female and ethnic minority students in cooperation with established diversity and outreach channels. We will continue this effort in this project.

Argonne employs over 900 undergraduate and graduate students each year. Argonne hires undergraduate students for part-time and temporary assignments, and the lab also offers a variety of graduate student research appointments to fit the needs of graduate students.

Argonne's postdoctoral program provides early-career professionals with the opportunity to conduct research in multiple disciplines, routinely working together to solve large, complex problems. In addition to research, postdocs engage with both the Argonne and outside communities through various outreach and educational activities. Postdocs at Argonne are supported through career development, mentoring, and a vibrant postdoctoral community. Each postdoc pairs up with a staff mentor outside of the postdoc's direct supervision chain, and mentors meet with postdocs on a regular schedule to track career development. Argonne also has a vibrant postdoctoral society, hosts an annual career fair, annual postdoctoral symposium, and monthly meetings.

Another way that Argonne maximizes its recruitment and retention of students and early-stage investigators is through a structured program of diversity, equity, and inclusion (DEI). We actively recruit students and early-career researchers whose cultural and intellectual backgrounds equip them to look at a problem from a variety of viewpoints. Diversity in thought, background, and approach results in the development of unique and creative, pioneering solutions.

Appendix 8: Other Attachments–Letters of Collaboration

Table of Attachments

Letter	Author
Letter Of Collaboration	Dr. Nadia Panossian
Letter of Collaboration	Dr. Paul Ullrich

Dear Kristin Potter:

If your proposal entitled, “Actionable Intelligent Visual Analytics of Ensembles,” is selected for funding under the DOE Office of Science Advanced Scientific Research Program, it is my intent to collaborate in this research by advising on the domain of energy systems and the challenges that exist in the analysis and decision making in this space. Thank you for the opportunity to participate.

Sincerely,
Nadia Panossian Ph.D.
Electrical Engineering Researcher
National Renewable Energy Laboratory
nadia.panossian@nrel.gov

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO • SANTA BARBARA • SANTA CRUZ

DR. PAUL ULLRICH
DEPARTMENT OF LAND, AIR AND WATER RESOURCES
251 HOAGLAND HALL

ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-8627

PHONE: (530) 400-9817
FAX: (530) 752-1793
EMAIL: paullrich@ucdavis.edu
WEB: http://climate.ucdavis.edu

16 June 2022

Dear Kristin Potter,

If your proposal entitled, "Actionable Intelligent Visual Analytics of Ensembles," is selected for funding under the DOE Office of Science Advanced Scientific Research Program, it is my intent to collaborate in this research by evaluating the effectiveness of the approach for understanding climate simulations. Thank you for the opportunity to participate.

Sincerely,

Paul A Ullrich

Dr. Paul A. Ullrich
Professor of Regional Climate Modeling
University of California, Davis