DCPViz: A Visual Analytics Approach for Downscaled Climate Projections

Abdullah-Al-Raihan Nayeem 1, Huikyo Lee2, Dongyun Han3, Mohammad Elshambakey4, William J. Tolone1, Todd Dobbs1, Daniel Crichton2, and Isaac Cho1,3

1College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, United States
2Jet Propulsion Laboratory, California Institute of Technology, Pasadena, United States
3Computer Science, Utah State University, Logan, United States
4City of Scientific Research and Technological Applications, Alexandria, Egypt

Abstract—This paper introduces a novel visual analytics approach, DCPViz, to enable climate scientists to explore massive climate data interactively without requiring the upfront movement of massive data. Thus, climate scientists are afforded more effective approaches to support the identification of potential trends and patterns in climate projections and their subsequent impacts. We designed the DCPViz pipeline to fetch and extract NEX-DCP30 data with minimal data transfer from their public sources. We implemented DCPViz to demonstrate its scalability and scientific value and to evaluate its utility under three use cases based on different models and through domain expert feedback.

Index Terms—climate projection analysis, spatiotemporal visualization, exploratory visual analytics, coordinated multiple-views

I. INTRODUCTION

In the United States (U.S.), the National Climate Assessment (NCA) [1] is the nation’s leading resource for exploring and addressing adaptation and mitigation questions related to climate change across the various sectors of society. To explore and assess the potential impacts of future climates, the NCA uses statistically downscaled climate projections at a high spatial resolution of 10km or finer. Statistical downscaling (SD) is a correction technique for mitigating biases in important variables of interest (e.g., temperature and precipitation). This technique uses observations to estimate the bias-corrected variables appropriately and at a higher spatial resolution than the original model outputs.

Recognizing the importance and need for bias-corrected statistical downscaling to make reliable future climate projections (e.g., [2]), NASA’s Earth eXchange project released the Downscaled Climate Projections (NEX-DCP30) at 800m resolution to support climate change research [3], [4]. The NEX-DCP30 provides climate projection data, including monthly averaged precipitation (pr), daily maximum temperature (tasmax), and daily minimum temperature (tasmin), for the contiguous U.S. for the 115 years between 1985 and 2099. The high-resolution spatiotemporal dataset is organized according to the Network Common Data Form (NetCDF) format [5] and publicly available on the Amazon Simple Storage Service (S3) [6].

There are, however, many challenges that limit the utility of the NEX-DCP30 and similar downscaled climate projection (DCP) data for climate research and relevant scientific analyses. First, due to the size of such data, it is often problematic for climate scientists to migrate data to local storage, which is often necessary for data analysis and exploration. For example, each NetCDF file contains 5 years of data over the contiguous U.S. at 800m resolution, which makes each file size approximately 2GB (or, 17TB for the complete archive) [3], [7]. Similarly, CMIP6 [8] is expected to produce more than 30 PB of simulation results through the Earth System Grid Federation. Second, existing tools for visualizing, analyzing, and exploring data like the NEX-DCP30 (e.g., Panoply [9], NCCV [4]) are limited. Existing tools normally depict only a few static plots with minimal support for interactive exploration and comparative analyses. Finally, even with prior knowledge of climate models and their associated data formats (e.g., NetCDF), it is a prohibitive and complex task to convert data into formats that are easily interpretable, understandable, and, thus, useful for decision-makers [4], [10].

In a traditional pipeline, analyzing massive climate data usually begins with users searching and locating high-resolution climate data. Users, then, must manage local storage, prepare computational resources, and request data access. Once access is granted, users must download the (usually massive) data, convert the data into the desired format, analyze the data computationally, and, visualize the converted data for scientific analyses to create, for example, performance benchmarks based on the dataset. Due to the importance of high-resolution climate data to the U.S. NCA [1] and other international climate assessments (e.g., the IPCC) [11], climate scientists and decision-makers require more innovative and interactive, visual, and cloud-based solutions to overcome the highlighted challenges and maximize the utility of climate data like the NEX-DCP30 to improve our understanding of future climate change at multiple scales (e.g., regional, national).

To address the challenges of analyzing and exploring shared high-resolution climate projection data like the NEX-DCP30, this paper introduces a novel interactive visual analytics approach, called DCPViz, that improves the efficiency and efficacy of climate projection analyses. DCPViz provides data exploration and comparative analyses with better accessibility to shared data, and a streamlined environment for climate
scientists and decision-makers to analyze climate projections data. To evaluate our approach, we demonstrated DCPViz to climate scientists, allowed the scientists to explore DCPViz (and associate climate data - i.e., NEX-DCP30), solicited expert feedback, and performed a qualitative evaluation of the system. Our findings revealed that key aspects of DCPViz - its interactivity, accessibility, and sense-making support - assisted the evaluators in climate analyses and improved overall task performance. The contributions of this paper are:

- We provide a novel cloud-based pipeline to extract high-resolution climate data (e.g., NEX-DCP30 data) for an interactive visual analytics system. The pipeline substantially reduces redundant downloads of large-scale data by moving analysis to the site of the data and migrating only the results of the analyses.
- We introduce a web-based interactive visual analytics tool, DCPViz, that helps scientists perform analysis, reasoning, and decision-making tasks with spatiotemporal climate projections.
- We present DCPViz use cases designed and recommended by climate scientists. We also present findings based on expert evaluations. These findings highlight the aspects of DCPViz that assist climate scientists to improve task performance.

II. RELATED WORK

To situate the contributions of DCPViz, we summarize related works from two perspectives: i) visualization systems for earth science domains and ii) visualization techniques for relevant data types. Our literature review is further scoped to the challenges of analyzing massive, high-resolution DCP data.

A. Visualization Systems for Earth Science

Several visualization and modeling platforms have been developed for climate data analysis in recent years. However, to our best knowledge, no previous platforms demonstrate either: i) a visual analytics pipeline with exploratory features for massive, high-resolution DCP data; or ii) sufficient interactive visualizations for exploratory and comparative analyses.

1) General Purpose Systems for Earth Science: Most general-purpose visualization systems demand data to be organized according to a well-defined format prior to visualizing, exploring, and analyzing the data. In addition, visualization systems until recently were not compatible with cloud-based architectures. Analytics systems for large-scale data such as Community Data Analysis Tool (CDAT) [12], [13], Climate Engine [14], MATLAB [15], and NCAR [16], and GrADS [17] have been widely applied to the earth science domains. The goal of these platforms is to provide data mapping and time-series visualization features to users of geospatial data. Despite comprehensive analytical features, these systems do not offer integrated visualizations for analyzing spatiotemporal patterns.

2) Visualization Systems for Climate Change: Climate scientists work with diversified data that frequently demand visualizations and interactions specifically tailored for the analysis of massive, high-resolution spatiotemporal data. Hence, most of the special purpose systems have domain experts involved in the design, development, and evaluation process [18] according to task-specific requirements such as operational weather forecasting, climate change assessments, etc.

HI-RAMA [19] and VAPOR [20] are domain-specific systems to help climate scientists predict and explore the impact of potential climate change. NCCV [4] offers exploratory visualizations for NEX-DCP30 temperature and precipitation data. To our best knowledge, NCCV is the only tool focused on NEX-DCP models, but it was developed based on Flash player which is no longer available. Similar visualization systems have been developed to facilitate deep exploration and comparative analysis of climate projections [20]–[23]. But these systems limit analytical support to specific variables or models. NASA developed an online analysis platform, Giovanni [24], that limits its study of geophysical variables including precipitation and temperature, as well, restricts users to load preferred data and processing scripts.
Several platforms have been developed (e.g., MeteoInfo [25], WebGlobe [23], Panoply [9]) to analyze the data encoded in the NetCDF format, an abstraction for storing multidimensional data, and a popular data encoding format for climate variables [20], [23], [26]. These tools are sufficient to generate a geospatial view from NetCDF data but lack interactive features for exploring spatial patterns.

B. Interactive Visualization Techniques

Climate variables are often attributed to geolocation and timestamp information. As such, climate model evaluations utilize geospatial visualizations and time charts [18]. Geospatial visualizations commonly enable data exploration based on geographic location [4], [27]. Time charts are often used to identify trends over time. Information access, navigation, user interaction, and manipulation are some common affordances of interactive geospatial visualization systems. While visualization has a lengthy history with climatology, researchers continue to search for better visual approaches to explore complex environmental information in a manner that is both understandable and useful for a broader range of tasks [27]. As such, visual analytics systems often leverage other visual artifacts to support geospatial visualization. For example, area maps are used to plot the area of modeled impact [20], [24]. Parallel coordinates are utilized to analyze the association in multivariate climate data [28]. Histograms, bar plots, and hierarchical ensemble clustering are popularly leveraged in several systems to compare the climate models in different spatiotemporal granularity [29], [30]. However, these tools and systems provide insufficient support for coordinated multivisualizations with interlinked user interactions, which are very useful for pattern detection and comparative analysis.

DCPViz, however, allows climate scientists to explore spatiotemporal patterns and covariability among projected climate variables using interactive spatiotemporal and geospatial visualization. The interactive features also allow the user to interact with its system pipeline to transform the data and link them to coordinated visualizations.

III. DESIGN METHOD & REQUIREMENTS

To design DCPViz, our team of visualization researchers and climate scientists met weekly for more than one year to define design requirements, discuss climate science workflows, and conduct several research and design interactions. In addition, we reviewed related works and the goals of NEX for the NCA [3], [31]. These goals include: a) engaging scientists in addressing climate change, b) improving the scope of climate research and analysis, and c) facilitating better accessibility to data and software tools to encourage collaboration and knowledge sharing. Based on the literature, we identified five primary design requirements:

R1 To facilitate data preparation without the overhead of massive data downloads. The system should extract spatiotemporal climate projections from NetCDF data files stored in a cloud environment efficiently and eliminate redundant downloads of large data files in order to run exploratory analyses.

R2 To enable interactive transformation and filtering to multidimensional data. The system should provide seamless accessibility to interactive features for data transformation and filtering for exploratory and sense-making tasks. Leveraging a web-based protocol for data access will also support a NEX objective by enabling better data accessibility to climate scientists.

R3 To enable climate scientists to perform visual exploratory analyses. The system should offer a web-based interactive interface that allows climate scientists to access model projections for the emission scenarios. The interface should provide climate scientists visual components to perform analysis tasks, reason over analytical results, and identify latent insights.

R4 To facilitate trend and covariance analysis in different spatiotemporal granularity. The system should offer interactive visualizations to observe underlying patterns and identify covariance among climate variables at multiple spatial (e.g., region, state, county) and temporal (e.g., annual, seasonal, monthly) granularities.

R5 To enhance collaboration and sharing among climate scientists. The system should serve as a platform for climate scientists where they can perform data exploration and analyze climate variables simultaneously. The system should facilitate further collaboration by enabling the sharing of key findings among climate scientists.

To satisfy the aforementioned requirements, we designed a novel visual analytics pipeline for DCPViz. This pipeline contains four modules: data processing, data collection, (RESTful) web API, and the UI, which provides interactive visual components. Fig. 2 illustrates the modules and process flows.

The Data Processing Module enables data transformation and analysis over distributed data by leveraging VIFI [32], [33]. This module runs on a remote server to access the NetCDF files from a mounted S3 bucket containing the DCP data (Fig. 2A and 2-3). These data are extracted from the cloud in an efficient matter to facilitate data preparation and reduce the overhead of massive data downloads prior to analysis (R1).

The Data Collection Module stores the extracted data in a shared repository and obtains the data index to facilitate accessibility ‘on demand’ (Fig. 2B and 2-4). Each index is composed of a unique code for the model name, variable name, temporal identifier, and the DCP dataset name (e.g., ‘NEX_DCP_CESM1-CAM5_pr_2021-03-01’). Each index identifies its original NetCDF file. The shared repository stores the indices for the entire dataset (e.g., spatiotemporal precipitation, maximum and minimum temperatures). When the user focuses on a certain segment in the timeline or a spatial region, the interface passes a request to the API module to fetch the required data (R2).

The Web API Module is a backend server that extracts a subset of the data from the remote site using specified filters (Fig. 2C). To eliminate the need for prerequisite knowledge
about the climate model and data format, the system automatically pre-processes the data into the requested format. This is a standalone service that provides precipitation and temperature data to the UI module and associated analysis scripts (R2).

The UI Module provides Coordinated Multi Views (CMV) to allow users to investigate climate change on a regional scale leveraging high-resolution NEX-DCP30 data [18]. This module provides not only conventional plots according to the traditional workflow but also an interactive visual interface (Fig. 2D). The interface supports exploratory analyses on future climate change over spatial and temporal domains (R3). Moreover, the interface displays climate variability and a summary of overall seasonal changes between pr and tasmax and across the 7 NCA regions in the contiguous U.S. (R4). Furthermore, DCPViz supports textual annotations in the map and matrix views to share key findings with other users on DCPViz (R5). Section IV-C describes the UI in detail.

We integrated these modules to build our proposed visual analytics system. The UI module interprets a sequence of events from the user and delegates them to the distributed modules to support requested functionalities.

IV. Visual Analytics Approach

In this section, we describe DCPViz and illustrate how it meets our design requirements.

A. The NEX-DCP30 Dataset

NEX is a collaborative platform to facilitate the analysis and forecast of climate data [3]. NEX projections provide several statistically downscaled datasets via a public S3 bucket. Among these datasets, NEX-DCP30 is the dataset with the highest spatial resolution of about 800m. Three variables are included in NEX-DCP30 (precipitation - pr (mm/day), daily maximum temperature - tasmax (K), and daily minimum temperature - tasmin (K) later converted to (°C)). They are derived from 36 climate models and observational data. We use two of these models for DCPViz: Community Atmospheric Model version 5 (CESM1-CAM5 [34]) and Goddard Institute for Space Studies Model E2 coupled with the Russell ocean model (GISS-E2-R [35]). NEX-DCP30 contains 115 years of monthly data for the contiguous U.S. Data from 1985 to 2005 are historical and data from 2006 to 2099 are projected under three RCP scenarios (2.6, 4.5, and 8.5).

B. Data Transformation & Analysis

We leverage VIFI, an open-source tool that enables the analysis of distributed fragmented data. VIFI enables data sharing by migrating analytics (often lightweight) to data locations rather than by migrating massive data [32], [33]. Users can perform a variety of analyses, specified as analytical workflows (Fig. 3), over distributed data stored at multiple locations. VIFI workflows are implemented for data-driven discoveries from distributed climate data [32]. We utilize VIFI to enable contour map generation, analyze climate variability and identify anomalous patterns from the RCP scenario projections (Fig 2A-B). Workflow execution begins with query generation based on user-selected parameters such as model name, climate variables, experiment ID, and year range. Once the backend server accesses the data corresponding to specified parameters, the Data Processing Module (Fig. 2A) triggers the VIFI workflows to perform data extraction, transformation, and analysis. To demonstrate this process in DCPViz, we provide an illustration using pr and tasmax in the NEX-DCP30 models.

Contour Map Generation. We extract NetCDF files using VIFI data orchestration to retrieve geo-coordinated monthly averaged climate projections (e.g. pr and tasmax). The workflow extracts 60 projected snapshots (one for each month) from each NetCDF file that recorded the spatiotemporal climate variables. Next, the contour generation process (Fig. 3F) transforms the geo-coordinated data to the GeoJSON format. While extracting the data remotely, the collection module attaches an index to the generated GeoJSON using the search parameters.

Spatiotemporal Granularity Extraction. We apply a series of transformations to geo-coordinated monthly projection data to extract the spatiotemporal granularity (Fig. 3). We apply an NCA's region mask on the projected data to segment
the temporal variable intensity into seven regions (Northeast, Southeast, Midwest, Southwest, Northwest, Northern Great Plains, and Southern Great Plains). Then, we utilized the segmented data to calculate seasonal and yearly regional means for the projected variables. These data are further exploited to analyze and assess climate variability in the projected models.

**Extracting Anomalous Projections.** Lastly, we derive measures to extract anomalous patterns from the climate projections. A previously calculated yearly mean is leveraged to measure the percentage change between the yearly mean and the selected decades of retrospective mean (e.g., 1985-2006), what we call ‘Relative Intensity’. Seasonal regional retrospective means are described as \(RM(t, s, r)\) in the following:

\[
RM(t, s, r) = \frac{1}{3\Delta t} \sum_{y=t_0}^{t_1} \sum_{m+s} X_r(y, m)
\]

where \(t\) represents a year range within retrospective timeline, \(s\) represents the season, and \(r\) represents the region. \(X_r(y, m)\) provides the monthly-averaged regional mean for the climate variables. \(RM(t, s)\) is leveraged in calculating the ‘Relative Intensity’, \(RI(t, y, m, r)\) for \(pr\) described as follows:

\[
RI(t, y, m, r) = \frac{1}{RM(t, [m/3], r)} X_r(y, m) - X_r(y, m) \]

The intensity fluctuation is often projected within a wide range of variations, which can overlook anomalies at certain data points. For example, a 0.5 \(mm/day\) change in \(pr\) may not be noticeable in a snapshot compared to the overall range. However, this small change could be significant if the \(pr\) trends around 2 \(mm/day\) in that month/season. As such, \(RI(t, y, m, r)\) can amplify anomalous patterns.

**The Data Collection Module** (Fig. 2B) maintains a repository of extracted and transformed data as well as analytical results. We store these data in a manner that allows for efficient filtering while exploring projections from the entire timeline. In addition, this module manages the user’s credentials and annotations.

The **Web API Module** transforms collected data as per the required structure for the interactive visualizations before responding to the requests from the UI Module. All of these processes are executed remotely to limit data migration and the need for user knowledge of data format and organization.

### C. DCPViz: Exploratory Visual Interface

The DCPViz interface presents interactive temporal and spatial visualizations for retrospective and projected climate data. The UI leverages CMV to enable rich user interaction. It consists of three main views: 1) a map view (Fig. 4C), 2) an adaptive spatiotemporal view (Fig. 4A-B), and 3) a reconfigurable summary panel (Fig. 5, 6, and 7). We implemented the interface for exploring the projection of historical \(pr\) and \(tasmax\) data across the U.S. The interface is developed using open-source web libraries.

1) **Map View:** The map view shows a geo-coordinated contour that represents the monthly-averaged intensity across the U.S. regions. The view has three map layers: a base map, a contour map, and an area border map for the contiguous U.S. The base map shows different geographical features of the U.S. such as boundaries, rivers, and highways. The geospatial contour map is plotted as a variable intensity layer. The area border map shows the borders of the seven U.S. regions.

The map view provides three interaction modes (Fig. 4i): 1) inspection, 2) selection, and 3) annotation. The inspection mode allows users to find areas of similar intensity on the contour layer. Areas are grouped by the projected intensity levels, which range from 0-14 \(mm/day\) with a yellow-green-blue scheme. Hovering over a specific location opens a tooltip that shows additional information such as the inspected region and the monthly-averaged value at the selected snapshot. Hovering also strengthens the color intensity to let the user effortlessly differentiate the focused area from other areas (Fig. 1C). The selection mode allows users to indicate attention in a specific location. Downscaled climate data are projected at 800m resolution [3], [7]. So, concentrating attention on a specific region can provide essential detail that may not be easily observed from the default view. Users can also change the view hierarchy to the entire U.S., a specific region, or a specific state on the map.

The annotation mode allows users to share findings and observations on the map view. Once the user adds an annotation, the view shows a pin to share points of interest with others for further exploration (Fig. 4h). An additional marker is added to each thumbnail in the spatiotemporal view to indicate whether the annotation is added for that snapshot (Fig. 1B). This interaction also allows users to post annotations directly without marking a specific location on the map.

2) **Spatiotemporal View:** The spatiotemporal view presents thumbnails of geospatial contours with the corresponding observations on the map view. Once the user adds an annotation, the view shows a pin to share points of interest with others for further exploration (Fig. 4h). An additional marker is added to each thumbnail in the spatiotemporal view to indicate whether the annotation is added for that snapshot (Fig. 1B). This interaction also allows users to post annotations directly without marking a specific location on the map.

---

1The 4th NCA report segments U.S. in 10 regions [36] but Alaskan, Hawaiian, and U.S. Caribbean are excluded in the NEX-DCP30 data.

2The interface is publicly available at https://esva.jpllab.net/
temporal mean from 2036 to 2099. The entire dataset for the model contains 150 years of data [7] of which we extracted remotely the most recent 1380 (monthly average for 115 years) recorded/projected snapshots. Observations reveal that it can be difficult to identify trends or mark points of anomaly by traversing snapshots using only geospatial projections on the map. To address this, we developed the spatiotemporal view to render the geospatial data against a temporal axis. We add color-coded legends (Fig. 1f) with corresponding measuring units to denote \( pr \) and \( tasmax \) in the visualizations. This view has four visual components (Fig. 1A-B): a temporal heat map, a geo-spatiotemporal contour matrix, a climate variability time series, and a regional bar chart.

**Temporal Heatmap.** The temporal heatmap shows the projected mean \( pr \) sharing the identical axis with the temporal geospatial contour (Fig. 1C). We used two different measures to set the intensity for the heatmap: 1) mean \( pr \) for each snapshot (Fig. 4A); and, 2) relative \( pr \) (Fig. 1A) derived as \( RI_{pr}(t, y, m, r) \) in section IV-B. Fig. 1e provides a control ‘Relative PR’ to switch between these measures. We also add the option to filter this view by setting upper and lower limits of \( pr \). A similar seasonal grouping is applied in the temporal geospatial thumbnail view. Fig. 4A shows an example of the Fall season where the cells in a row represent months in the corresponding year. The temporal heatmap helps users understand the progression of \( pr \) over time. The yellow-green-blue color scheme is used for monthly-averaged mean in the geospatial visualization [37]. In the ‘Relative PR’ view, we adjust the color scheme to blue-pale yellow-red because it is popular for showing anomalies [37], which in our case denotes maximum increase to maximum decrease from blue to red.

**Geo-spatiotemporal Contour Matrix.** When capturing the contour for each geo-coordinated monthly snapshot, we store them as static images. These images are essential to show a preview to help users identify the point of interest. DCPViz presents the captured geospatial snapshot against a temporal axis. Therefore, the user obtains both the time information and suggested \( pr \) intensity before deciding to explore the contour in the map view. We maintain a uniform alignment between the temporal heatmap and geospatial contour matrix to help the user identify the corresponding contours from the heatmap cell. We also group the contour images by seasons (Fig. 1B) to help users perceive the seasonal \( pr \) trend. Hovering over thumbnails shows enlarged renderings of the contour with metadata for a detailed view.

**Climate Variability Time-Series.** The line charts represent a time series of mean \( pr \) (blue) and mean \( tasmax \) (gray) (Fig. 1A). \( pr \) and \( tasmax \) are subset data for the selected region, grouped by season and rendered for each season. Each data point in the line charts presents the yearly seasonal-average \( pr \) record and their associated change in the \( tasmax \).

**Regional Bar Chart.** The regional bar presents the yearly seasonal-average \( pr \) grouped by region. The map view provides an indication of the intensity of \( pr \) in different regions. However, the map view may not be adequate for the user to compare \( pr \) among the regions. Hence, the regional bar chart view is aligned with the geospatial contour in the temporal matrix, as shown in Fig. 4B, to show a precise comparison of \( pr \) among the regions. In addition, the bar representing the selected region is highlighted to amplify the seasonal comparison among regions when scrolling up or down. A yearly-averaged regional bar chart is also included to reveal yearly changes among the seven regions.

3) **Summary Views:** Climate projections are often large and complex, and thus difficult to visualize both in totality and with granularity. As such, we provide a reconfigurable summary view to encode and visualize the data. We considered three design factors for the summary view: the visualization purposes, the supporting analysis tasks, and the data type [38]. We leverage the multi-level spatiotemporal granularity extracted from the projected data during data transformation. The summary view contains three visualizations: a time-series visualization (Fig. 5), an RCP scenario comparison (Fig. 6), and a hierarchical treemap (Fig. 7A).

**Time-Series Visualization.** The time-series visualization shows the seasonal \( pr \) and \( tasmax \) mean from 1985 to 2098 (Fig. 5). It allows users to explore the trends of projected climate variables, such as \( pr \) and \( tasmax \). It also provides...
support for direct search to help users find anomalies and patterns in the timeline. We found the time-series visualization to be a suitable candidate for identifying outliers and patterns across a large number of data points [39], [40].

RCP Scenario Comparison. A different design of a stacked bar chart is adopted for visual comparison and anomaly detection across RCP scenario projections (Fig. 6A). Each RCP scenario is distinguished by color, where overlapping areas are highlighted with strengthened intensity. This visualization presents a yearly-seasonal regional projection spread of the extreme RCP scenarios (e.g., RCP 4.5 and RCP 8.5 as compared to RCP 2.6). In Fig. 6B, we compare the scenarios for \textit{tasmax} during the summer season in the western region according to CESM1-CAM5. Lower fluctuations among different RCP scenarios are observed in earlier decades in the 21st century. However, \textit{tasmax} is projected to be significantly higher in the later years in the most extreme scenario.

Hierarchical Treemap. The treemap visualizes data with a nested structure for two selected climate variables, e.g., \textit{pr} and \textit{tasmax} (Fig. 7). The treemap is constructed maintaining the following hierarchy of data from the top: region, season, and year (Fig. 3). \textit{pr} is denoted by the dimension of the cells whereas \textit{tasmax} is denoted by the intensity using a blue-yellow-red color scheme for lowest to highest maximums. In the treemap, users can select an area of interest (Fig. 7B). The treemap presents an overview of the climate co-variability between variables to highlight, for example, which region is getting hotter & wetter or hotter & drier under a changing climate.

V. EVALUATION

Climate scientists provided a qualitative evaluation of DCPViz to determine its effectiveness at meeting its design requirements (R1-R5), characterized the scientific value of climate data analyses, and demonstrated our research contributions. In this section, we describe the DCPViz evaluation metrics, the case studies utilized, and domain expert feedback.

A. Qualitative Evaluation Metrics

Evaluations of climate science visualizations span numerous qualitative and quantitative dimensions. For our evaluation,
we focused on qualitative metrics that assess DCPViz based on its ability to reveal nuances of climate change, intricately visualize massive spatiotemporal data, and depict multivariate association factors for climate change [41]. To assess the dimensions of interlinked analysis tasks, communication, and decision-making in climate research, we used evaluation metrics from the following categories: Visualization - ability to provide inexplicable insights that reduce time and complexity of analysis, Interaction - ability to coordinate multiple views and exclude irrelevant parameters and values from the view, and Presentation - elements utilized such as visual components, colors, legends, and tooltips to present the data demonstrates self-exploratory capabilities.

B. Evaluation Procedure

We employed a survey to solicit feedback from domain experts on DCPViz. Seven climate scientists, who are experts in analyzing NEX-DCP30 data, participated in the survey. The survey contained three sections. First, participants were asked to complete the pre-questionnaire regarding their experience with climate data, analysis tasks, and the tools and data analysis libraries they utilize in their analysis. Next, participants watched a video that demonstrated DCPViz functions. Then, they interacted with the UI to explore NEX-DCP30 data. Finally, participants were asked to complete a post-questionnaire that inquires about: 1) Preference of the visualizations, 2) Features that support their analysis tasks, 3) Performance of the CMVs for explanatory capabilities and features.

C. Use Cases

We highlight the three use cases where DCPViz demonstrated the greatest value to climate scientists.

Comprehensive investigation of the spatiotemporal variability in regional precipitation (UC1). The temporal heatmap (Fig. 1A) revealed to domain experts future pr changes for each of the four seasons. For example, when the user selects the Southeast U.S. region (Fig. 4e) and enables ‘Relative PR’ (Fig. 1e), the heat map cells in the winter season are mostly rendered in red colors (Fig. 4A). This essentially indicates a relatively drier winter later (2036-2085) compared to the first three decades of the century. The contour matrix (Fig. 4B) reveals that most of the region’s pr occurs along the West Coast and Sierra Nevada. Users can move their cursors over the thumbnails to see the spatial patterns in pr. By clicking the thumbnail, the precipitation intensity is displayed on the map view with boundaries of the associated NCA region (Fig. 4C). The map view supports zooming to maximize the utility of the high spatial resolution (800m). Making annotations (Fig. 1g) facilitates a comparison of pr patterns between dry and wet months.

Analyses of long-term trends in temperature and precipitation under different emission scenarios (UC2). The summary view (Fig. 5) reveals in the climate data the overall warming trend of tasmax for all the seven NCA regions. The observed warming from 1985 to the present is well known, and the bias-corrected temperature from NEX-DCP30 is expected to represent the observed trend. Unlike the warming trends, however, observed pr trends in the climate data, as revealed by the relevant visualizations, appear not to be significant across the contiguous U.S.. By default, the line and bar charts represent CESM1-CAM5 model predictions under the RCP 8.5 emissions scenario, which represents high emissions of greenhouse gases without any mitigation efforts. Using DCPViz, domain experts were able to observe that across all seven NCA regions and four seasons, the associated warming trends are higher than those for the RCP 4.5 and 2.6 scenarios. The plots indicate that pr changes are more variable than tasmax and tasmin in all three scenarios. Also, the pr predicted by CESM1-CAM5 does not exhibit consistent sensitivity to the emission scenarios across the seven regions (Fig. 6A).

Covariability of seasonal maximum temperature and precipitation (UC3). The hierarchical treemap (Fig. 7) provides scope to observe co-variability between the climate variables. The cell dimension represents mean pr while color intensity scales temperature. The hierarchical treemap also helps domain experts answer key scientific questions related to climate projections - e.g., Are warmer and more humid conditions expected during the latter decades of the 21st century in the fall season of the Western regions? What about spring in the Northeast region?

D. Domain Expert Feedback

The domain experts who participated in the survey possessed on average 5 years of experience in climate science research. In the pre-questionnaire, they mentioned that climate data processing and visualization is a moderately to highly difficult task due to the volume and complexity of high-dimensional structure. The domain experts also expressed their desire to have interactive geospatial and time-series views, in contrast to traditional static plots (e.g., by Python or Matlab).

Visualizations. The domain experts mentioned that interactive geospatial visualization (Fig. 4C) is worthwhile to have and appreciated the annotation feature as it allows them to share the observations with other scientists. Additionally, inspect user interactions helped them identify the spatial pattern in each snapshot (UC1). The spatiotemporal view (Fig. 4A) was also identified as effective by four experts to understand the temporal progression along with the spatial distribution. The heatmap helped the domain experts quickly inspect the data as the heatmap contains the summary for the whole timeline. The temporal contour matrix aided the domain experts with a quick view of the spatial patterns and their changes over time. Along with the seasonal and regional time charts, the domain experts identified the treemap as a less common yet useful visualization for analyzing comparative patterns for multiple variables.

Interactions. The experts’ feedback suggests that the reconfigurable view for summary visualization is convenient for sensemaking long-term patterns from different aggregation perspectives (UC2, UC3). The exploratory features such as inspecting the map view from
the contour thumbnail (Fig. 4B) and switching the multiviews based on region selection helped the domain experts to identify quickly interesting facts or insights from the data (R3). Moreover, the tooltips were well appreciated as they provide comprehensive data inspection capabilities and the additional context in the visual analysis.

**Presentation.** The domain experts were able to discover their target information from the interactive visualizations while one expert stated the interface became intuitive after the initial learning process. Moreover, the experts mentioned that the treemap is a good candidate to illustrate the covariability between pr and tasmax (UC3). We also learned that the use of information icons, color codes, and legends were deemed useful by them. The domain experts also left us a few suggestions such as - 1) to include annotation for other visualizations (beyond the map), 2) to increase the use of animation in visualizing the projections, and 3) to attempt to simplify the user interface.

Based on the majority of the feedback, we understand DCPViz benefits domain experts as they perform exploratory analyses of climate data (UC1). Moreover, the use cases suggest that DCPViz is also useful in examining long-term trends within the NCA regions (UC2, UC3).

**VI. Discussion and Limitations**

We proposed a cloud-based visual analytics approach, DCPViz, for making sense of DCP datasets that benefits exploratory analysis by climate scientists. We described specific challenges and identified important design requirements for systems that support the analysis of massive high-resolution spatiotemporal data for climate change research. We hypothesized that our proposed requirements would enhance the traditional analysis workflow and support climate scientists in sense-making and decision-making processes. We described the four modules that comprise DCPViz (data processing, data collection, web API, and UI) to satisfy the design requirements of the DCPViz pipeline (Fig. 2). In addition, we demonstrated how DCPViz can enhance transitional analysis workflows and support climate scientists’ sensemaking and decision-making.

**System Scalability.** We demonstrated DCPViz with two NEX-DCP30 models (CESM1-CAM5 and GISS-E2-R) in 3 RCP scenarios. The web-based UI provides simultaneous access to climate scientists to visually explore these extracted projection data. However, climate models are continuously updated to provide high-fidelity simulations by changing parameters. With computational resources becoming more easily accessible, the climate simulations with finer spatial and temporal resolutions (e.g., NEX-GDDP [42]), daily projections are becoming increasingly available. To incorporate other DCP models and RCP scenarios, we modularized the data processing and collection modules. This allows the scientists to trigger the processing module to extract more model data and store it in the collection module (R1).

**Limitations.** To support navigating and exploring the entire model projection, we have added exploratory and explanatory visualizations in the same window. As convenient as it is to gain insight promptly, the interface looks a bit complicated and it requires an initial learning curve to use the interactions. To perform a comparative analysis between projection models and scenarios, the user needs to open two DCPViz interfaces with the participant models or scenarios. In the future, we aim to support visual analysis of multi-model comparisons in DCPViz.

DCPViz enables the exploration of large volumes of data by creating visual components. We learned from the user study that system performance is affected by client-side network status and computational capacity. Despite the clear benefits of climate science analyses, system performance improvement will be part of our future work. Implementing combinable tabs [43] and adopting the progressive visualization pipeline [44] can ease users’ analytical vision while the processing is running on background. Employing a server-side tiled map technique for the map view can reduce the rendering time for geospatial contours to enhance the overall performance.

**VII. Conclusion and Future Work**

In this paper, we introduced DCPViz, a novel visual analytics approach to analyze and explore NEX-DCP30 data. We presented three use cases provided by climate scientists while interacting with the DCPViz interface. Experts’ feedback is utilized to evaluate the proposed visual analytics approach. While the feedback we received from the experts is promising, we identified several future directions for further research. Our future work will encompass a data mining architecture that is fully integrated with a distributed framework to optimize the data transformation and analysis time. In addition, we will utilize the extracted data from multiple DCP models to assess the climate projection uncertainty. To evaluate the model projections for each climate variable, we plan to add observational datasets, more analytical workflows, comparative visualizations, and interactive features on the interface. Finally, we plan to conduct a more extensive user evaluation by gathering climate scientists in focus groups to evaluate in greater detail the usability and efficiency of DCPViz.

**Acknowledgment**

This paper was supported by the US National Science Foundations (NSF) Data Infrastructure Building Blocks (DIBBs) Program (Award #1640818).

**References**

[1] K. Jacobs, *The US national climate assessment.* New York, NY: Springer Berlin Heidelberg, 2016.

[2] A. M. Wootten, E. C. Massoud, A. Sengupta, D. E. Waliser, and H. Lee, “The effect of statistical downscaling on the weighting of multi-model ensembles of precipitation,” *Climate,* vol. 8, no. 12, 2020.

[3] R. R. Nemani, B. L. Thrasher, W. Wang, T. J. Lee, F. S. Melton, J. L. Dungan, and A. Michaelis, “NASAs Earth Exchange (NEX) Supporting Analyses for National Climate Assessments,” in *AGU Fall Meeting Abstracts,* vol. 2015, 2015, pp. GC21E-04.

[4] J. R. Alder and S. W. Hostetler, “Web based visualization of large climate data sets,” *Environmental Modelling & Software,* vol. 68, pp. 175–180, 2015.

[5] E. Edward Hartnett and R. Rew, “Experience with an enhanced netcdf data model and interface for scientific data access,” in *24th Conference on IIPS,* 2008.
“Nasa earth exchange (nex) downscaled climate projections (nex-dcp30),” https://nasedx.s3.amazonaws.com/NEX-DCP30/, accessed March 31, 2021.

H. Lee, A. Goodman, and K. Gorski, “The Big Climate Data Pipeline (BCDP): A tool to facilitate server-side processing of NASA-NEX data on Amazon Cloud,” AGUFM, vol. 2019, pp. IN12A–03, 2019.

V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, “Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization,” Geoscientific Model Development, vol. 9, no. 5, pp. 1937–1958, 2016.

J. Palenik, G. V. AlecTo, M. Ventura, J. C. Wei, and D. J. Meyer, “The Advantages of Synergy - Quantitative Earth Science Data Visualization and Analysis with Giovanni, Panopoly, and Excel,” in AGU Fall Meeting Abstracts, vol. 2018, Dec. 2018, pp. IN21B–35.

E. Davis, C. S. Zender, D. K. Arctur, K. O’Brien, A. Jelenak, D. Santek, M. J. Dixon, T. L. Whiteaker, and K. Yang, “NetCDF-CF: supporting earth system science with data access, analysis, and visualization,” AGUFM, vol. 2017, pp. IN33C–0137, 2017.

M. Stockhause, R. Matthews, A. Pirani, A. M. Treguier, and O. Yeleki, “Cmip6 data documentation and citation in ipecc’s sixth assessment report (ar6),” in EGU General Assembly Conference Abstracts, 2021, pp. EGU21–2886.

K. Potter, A. Wilson, P-T. Bremer, D. Williams, C. Doutriaux, and V. Pasjuk, “Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated vis-cdat systems,” in Journal of Physics: Conference Series, vol. 180, no. 1, IOP Publishing, 2009, p. 012089.

E. Santos, J. Poco, Y. Wei, S. Liu, B. Cook, D. N. Williams, and C. T. Silva, “Uv-cdat: Analyzing climate datasets from a user’s perspective,” Computing in Science & Engineering, vol. 15, no. 1, pp. 94–103, 2013.

J. L. Huntington, K. C. Hegewisch, B. Daudert, C. G. Morton, and J. T. Abatzoglou, D. J. McEvoy, and T. Erickson, “Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding,” Bulletin of the American Meteorological Society, vol. 98, pp. 2397–2409, 11 2017.

D. Hanselman and B. Littlefield, Mastering MATLAB: a comprehensive tutorial and reference. Prentice-Hall, Inc., 1996.

D. Brown, B. Brownrigg, M. Haley, and W. Huang, “Ncar command language (ncl),” 2012.

F. Berman, A. Chien, K. Cooper, J. Dongarra, I. Foster, D. Gannon, L. Johnsson, K. Kennedy, C. Kesselman, J. Mellor-Crumme, D. Reed, L. Torczon, and R. Wolski, “The grads project: Software support for high-level grid application development,” The International Journal of High Performance Computing Applications, vol. 15, no. 4, pp. 327–344, 2001.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. McLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbel-luca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. MeLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbel-luca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. McLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbelluca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. McLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbelluca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. McLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbelluca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.

T. Nocke, T. Sterzel, M. Böttinger, and M. Wrobel, “Visualization of climate and climate change data: An overview.”

J. H. McLean, S. B. Cleveland, M. Lucas, R. Longman, T. W. Gianbelluca, J. Leigh, and G. A. Jacobs, “The hawaii’s rainfall analysis and mapping application (hi-rama): Decision support and data visualization for statewide rainfall data,” Association for Computing Machinery, 7 2020, pp. 239–245.