Geospatial data integration middleware for exploratory analytics addressing regional natural resource grand challenges in the U.S Mountain West

Shannon Albeke
School of Computing
University of Wyoming
Laramie, Wyoming
salbeke@uwyo.edu

Nicholas Case
School of Computing
University of Wyoming
Laramie, Wyoming
ncase2@uwyo.edu

Samantha Ewers
School of Computing
University of Wyoming
Laramie, Wyoming
sewers@uwyo.edu

Jeffrey Hamerlinck
School of Computing
University of Wyoming
Laramie, Wyoming
jeff.hamerlinck@uwyo.edu

William Kirkpatrick
School of Computing
University of Wyoming
Laramie, Wyoming
wkirkpa2@uwyo.edu

Jerod Merkle
Zoology and Physiology
University of Wyoming
Laramie, Wyoming
jmerkle@uwyo.edu

Luke Todd
School of Computing
University of Wyoming
Laramie, Wyoming
luke.todd@uwyo.edu

***Abstract*—This paper describes CyberGIS-based research and development aimed at improving geospatial data integration and visual analytics to better understand the impact of regional climate change on water availability in the U.S. Rocky Mountains. Two Web computing applications are presented. *DEVISE - Derived Environmental Variability Indices Spatial Extractor,* streamlines utilization of environmental data for better-informed wildlife decisions by biologists and game managers. The *WY-Adapt* platform aims to enhance predictive understanding of climate change impacts on water availability through two modules: "Current Conditions" and "Future Scenarios". It integrates high-resolution models of the biophysical environment and human interactions, providing a robust framework for professionals and the public to visualize and explore climate data through a user-friendly map-based interface. Both tools represent significant strides in bridging the gap between CI-enabled science and on-the-ground application, especially in the context of regional climate change, by streamlining both data integration methods and customized spatial queries.**

Keywords—CyberGIS, Web computing, visual analytics, climate data, DEVISE, Wy-Adapt, Wyoming.

I. Introduction

Geospatial information is widely recognized as critical in interdisciplinary approaches to solving societal problems [1]. Increasingly geospatial analyses have come to rely on robust supporting information and communication technologies capable of combining data from multiple and diverse sources and providing users with a unified view of the resulting synthesis of information [2]. As cyberinfrastructure modality frameworks for GIS advance, issues of scalable data integration and visual analytics remain a challenge in supporting credible spatial decision-making [3].

This paper presents ongoing CyberGIS-based research and development aimed at improving geospatial data integration and visual analytics services tied to better understanding the impact of regional climate change on water availability in coupled human and natural systems in the central Rocky Mountain region of the United States.

Water resources in Wyoming and the Rocky Mountain region are under significant threat from climate change. A shift from snow- to rain-dominated precipitation over the mountains, higher rates of evaporation, and increasing interannual variability elevate risks to watershed and ecosystem functioning and diminish the quality of natural resources that underpin social and economic stability in the region. Increasing our capacity to identify and quantify risks and predict societal consequences of shifting climate conditions in the nation's critical headwater areas are necessary.

Taking full advantage of CyberGIS frameworks can still be challenging for domain-specific end-users such as on-the-ground natural resource managers and planners because of the complex workflows and associated data skills necessary to synthesize large amounts of climate-based data. Even though the data are freely available, dynamic information displays affording exploratory analysis of multiple biogeophysical metrics over space and time requires a specific set of software tools and skills.

We are actively developing and supporting two data integration middleware applications, ***DEVISE*** (Derived

This work was supported by the Wyoming Innovation Partnership (https://wip.wyo.gov/) and by the US National Science Foundation (NSF) under award OAI-2149105.

Environmental Variability Indices Spatial Extractor; https://devise.uwyo.edu/) and ***WY-Adapt*** (current and future climate visualization tools; https://wyadapt.org/) aimed at overcoming disconnects between the cyberinfrastructure-enabled science and on-the-ground application. An overview of each application's function and workflow is described followed by a discussion of ongoing challenges and future work.

II. DEVISE

DEVISE (Derived Environmental Variability Indices Spatial Extractor; https://devise.uwyo.edu/) is an integrated data system developed in collaboration with the Wyoming Game and Fish Department user community to streamline how land managers and biologists can use environmental data. The system facilitates informed decision-making, such as where to conduct habitat restoration, how many game animals to harvest, and optimal placements of future development (e.g., wind farms) by significantly lowering the 'data access' bar. Development of DEVISE had the following objectives: 1) develop a database and analytical system to calculate derived metrics from federated data products and algorithms; 2) develop a front-end, easy-to-use web interface for users to download and visualize metrics on-demand for specific management areas; 3) develop additional tools that enable agency-provided polygons attributed with existing time-series datasets; and 4) develop new cyberinfrastructure and data systems to support automated daily maintenance of the DEVISE environment.

The DEVISE cyberinfrastructure (Fig. 1) consists of connected and interrelated servers and services, each with specific tasks for maintaining data currency. To begin, a Virtual Machine (VM) hosted on UW's Advanced Research Computing Center (ARCC) infrastructure nightly runs a set of Program R scripts (the DEVISER R package; https://gitlab.com/albekeecoinfolab/deviser) to download and process federated data sources, creating a consistently formatted group of data across all sources, which include DAYMET, PRISM, and SNODAS (see website documentation). The VM writes the processed daily data to ARCC's S3 File Server (Pathfinder Storage; Fig. 1) and a corresponding row of metadata to the MSSQL Server Database (DEVISE Derived Metrics; Fig. 1) via an ASP.NET Web API.

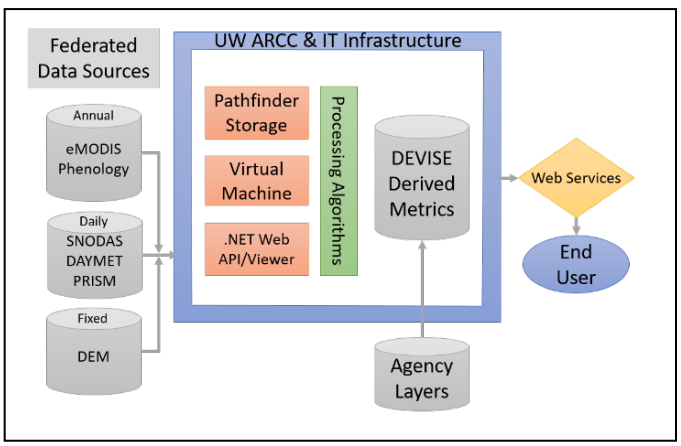

Fig. 1. **DEVISE cyberinfrastructure schematic.** We designed this system to be self-sustaining, providing users with near real-time updates for various datasets.

Concurrently, the VM runs additional jobs on specific days of the week to calculate temporal statistics of each dataset at weekly, monthly, and annual intervals. The jobs calculate Maximum, Mean, and Median values for each raster cell for each new temporal dataset. For example, if calculating the weekly datasets, we stack the appropriate daily rasters (Monday-Sunday for a single week) and determine the Maximum (97.5 percentile), Mean (average), and Median (50th percentile) of the available measurements. Finally, the VM adds metadata rows describing the additional raster datasets to the repository via the Web API.

After a new temporal dataset has been created and uploaded, a GIS extraction script attributes the polygon datasets in which we calculate the Maximum, Mean, and Median of the temporal dataset given the entirety of raster cells within each polygon. For example, the selected 'Sublette Mule Deer Herd Unit' polygon (Fig. 2) within Wyoming using all cells within its boundary to calculate the Maximum (97.5 percentile), Mean (average), and Median (50th percentile), along with confidence intervals (95% CI for the Mean, 2.5 and 97.5 percentiles for the Median) of the dataset. The resulting row of data containing these summary statistics for the 'Sublette Mule Deer Herd Unit' is then written to the MSSQL Server via the Web API

The same Web API also supports the DEVISE user interface and Open Data endpoints (Fig. 2). We developed a dynamic interface for discovering, viewing, and exporting synthesized climate datasets for predetermined spatial areas (e.g., hunt areas for a species of interest within Wyoming) using a combination of C#, JavaScript, and CSS. The interactive search (Fig. 2, Panel A) uses interconnected dropdown lists that allow users to narrow the focus of their inquiry. With each selection of available polygons, the linked map (Fig. 2, Panel B) and charting (Fig. 2, Panel C) widgets respond to user selections. Furthering the user's ability to view synthesized data on demand, we have developed additional widgets that allow grouping selected data months by year. Thus, if a user inspects maximum snow depth at the monthly temporal scale and desires to investigate data only for January, February, and March, limiting the visible data to only these months is quite simple for the user to accomplish. Further, the user is presented with a toggle to summarize data by the selected months (e.g., if toggled on, the three selected months are aggregated to represent one data point on the plot, average, or sum depending on the metric type). As an additional option for visualizing data, users can add a second dataset to the plot for on-the-fly comparison of differing metrics.

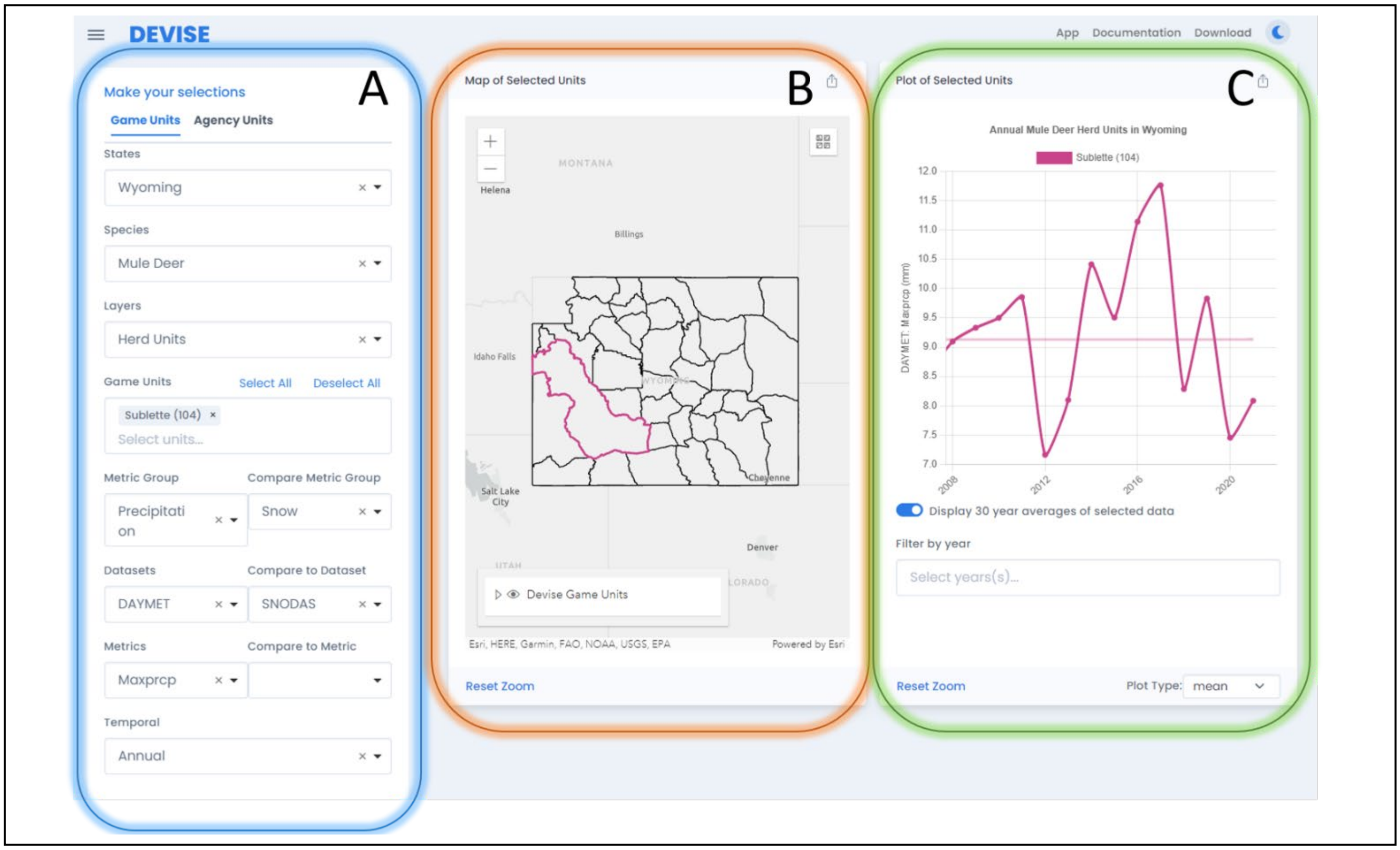

Fig. 2. **DEVISE Web Interface.** Panel A (blue) provides options to narrow data to be visualized, with filters for species, management units, etc. Panel B (orange) is an interactive map highlighting the selected 'unit' (Sublette Mule Deer Herd Unit in Wyoming) and map download capability. Panel C (green) is an interactive chart displaying the summarized climatic data for the selected 'unit(s)' (e.g., Max precipitation using DAYMET dataset). The plot is interactive, where users can select specific time periods, toggle 30-yr. averages, zoom into specific sections of the plot, and hover to view data point values. Export functionality is also supported.

III. Wy-Adapt

The WY-Adapt application suite (https://wyadapt.org/) is being built in support of the University of Wyoming's current NSF EPSCoR-funded grant *Anticipating the Climate-Water Transition and Cascading Challenges to Socio-Environmental Systems in America's Headwaters* (WyACT). Two major aspects of WyACT are the (1) the development of capabilities and infrastructure that improves predictive understanding of the coupled human-environment impacts of climate change on water availability, and (2) implementation of a transdisciplinary framework of co-production of knowledge between an interdisciplinary university research team and participants from Wyoming communities, sovereign tribes, and government.

WyACT's cyberinfrastructure effort is aimed at supporting interdisciplinary spatially informed research and providing a knowledge co-production framework to guide and build adaptive capacity within Wyoming communities in response to regional climate change. The Wy-Adapt application suite uses process-based, high-resolution models of the biophysical environment and human interactions which will be combined into multiple decision support applications to guide community planners and members of the public. Through our first year of development, WY-Adapt has focused on developing cyberinfrastructure and applications via data repositories, analytics tools, and dashboards.

The initial WY-Adapt application, 'Current Conditions', acts as a map-based reference for easily accessible climate and water data (Fig. 3) with sources originating from different networks and agencies as well as data collected by UW researchers and instruments. We are designing this interface to provide data to users with as few mouse clicks as possible. Even though many commonly used pieces of information are 'freely available' to the public, we found that many of the APIs are complex, provide data in very different formats, and are often inaccessible to the common user. Thus, streamlining and simplifying the interface aims to enable greater use by both researchers and lay users undertaking decision-making processes.

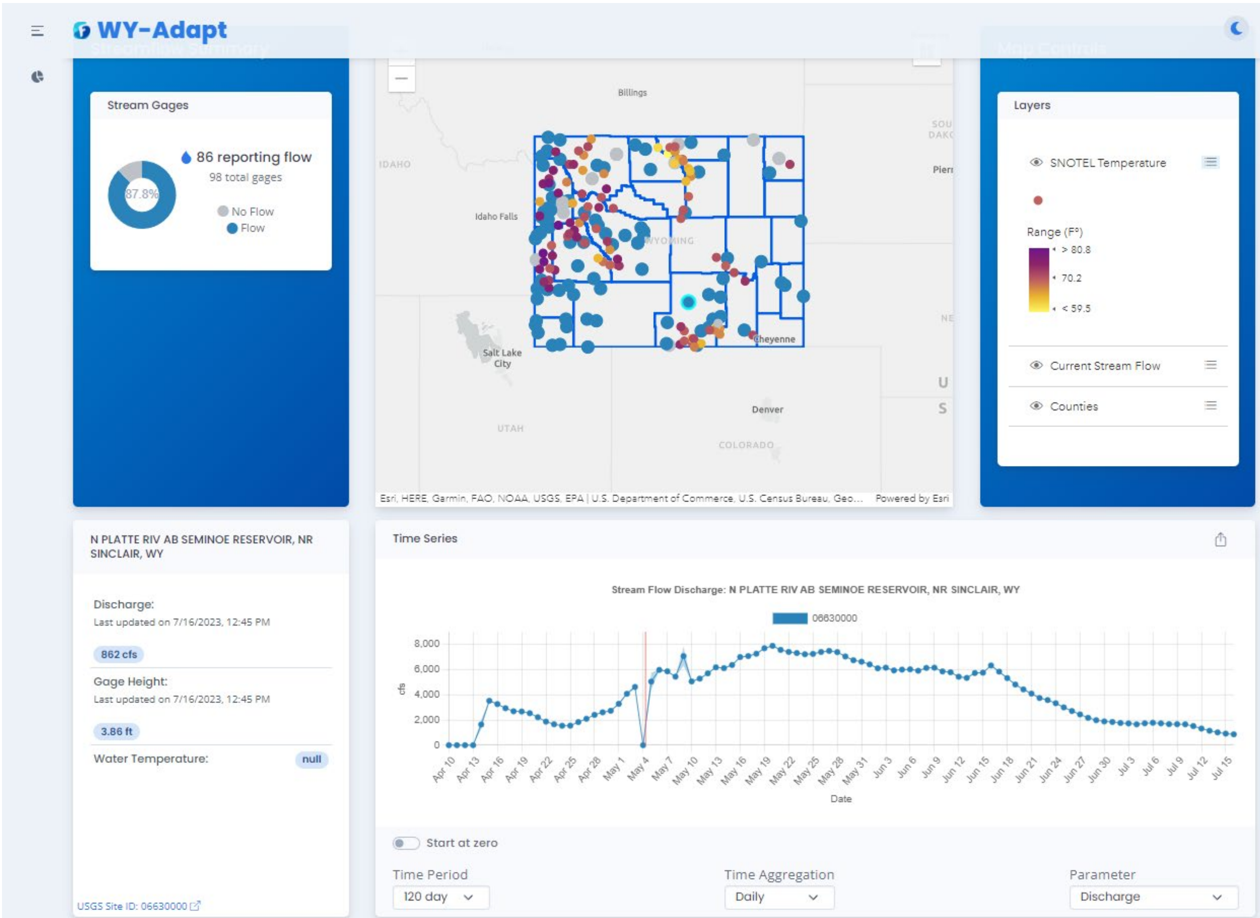

Fig. 3. **WY-Adapt Current Conditions application** where users can find real-time climate monitoring stations such as stream discharge, meteorological stations (precipitation and temperature), lake buoy stations, lake/reservoir current conditions and others. The time-series plots are enabled to provide users with interactivity to pick parameters and duration of interest as well as to download tables, maps, and plots.

The second Wy-Adapt application, 'Future Climate', aims to provide visualization of summarized future climate scenarios to users (Fig. 4). The WyACT atmospheric science team have provided a suite of nine GCMs at daily (1980-2100) 9-km resolution of precipitation, snow water equivalent (swe) and air temperature for the Wyoming region. These data, although very useful, require extremely large physical storage which are within dense NetCDF raster format and require special knowledge and software tools to read and extract this information. Thus, for the common user, this information is out of their reach. To facilitate responsive, on-demand requests of the data, we tabularize the raster data at a monthly scale for each grid cell. Using MSSQL Server relational infrastructure and C# Entity Framework design, we provide data analytics at multiple geographic and temporal scales to enable users to investigate potential future climate scenarios for their area of interest. For example, a user can select from the map Albany County, WY and in real time see a time-series plot of the annual trend of average air temperature expected within the county boundary. Concurrently, we provide a geographic representation of air temperature across all counties of Wyoming for a given decade, allowing the user to compare Albany to all other counties.

IV. Discussion

DEVISE places practical climate-data synthesis tools within easy reach of natural resource practitioners to assist in management decisions, providing resource specialists "laptop-level" on-demand access to large, distributed data sets and pre-processed information and visualization tools. The DEVISE framework reflects several significant CI-related technological advances. Creating the DEVISER R package for processing and maintaining up-to-date data that can run autonomously on a research computing facility supported VM is a significant leap forward. Secondly, we have developed an infrastructure that provides enough flexibility to allow additional data sets, both climatic and polygon analysis units, to be added with minimal effort. Finally, our approach of precompiling summary data for polygon analysis units facilitates rapid data synthesis and visualization by wrapping up multiple computational processes into a single row of tabular data.

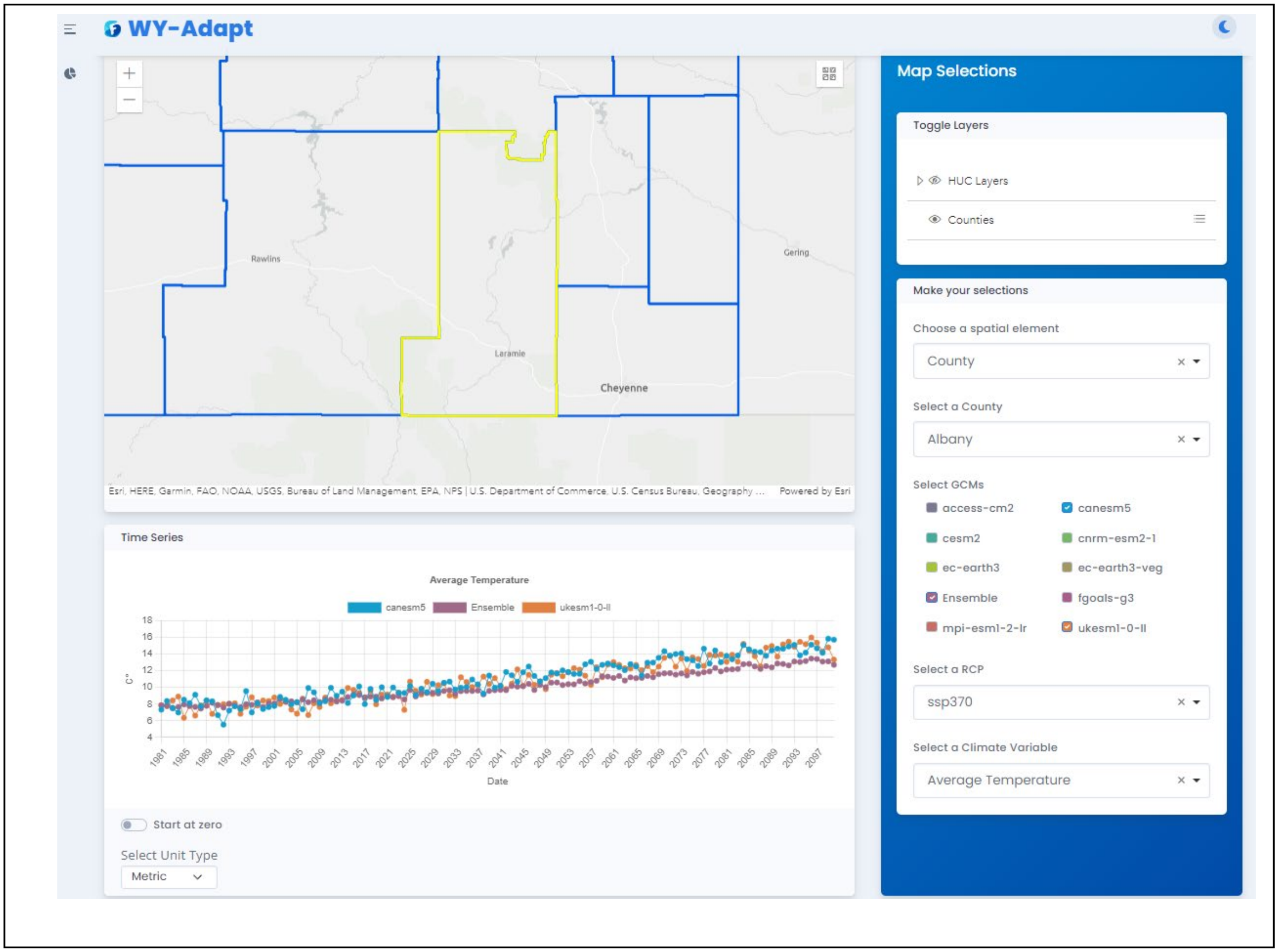

Fig. 4. **Future Scenarios application** provides on-demand access to nine GCMs and an ensemble of predictions from 1980 – 2100 for a variety of climate parameters at multiple spatial resolutions (grid cell, county, watersheds, etc.). The maps, plots and data are downloadable.

The WY-Adapt applications continue to evolve, and much of the planned functionality and analytic tools are yet to be programmed. However, we see great potential for enabling users to have the ability to ask custom exploratory queries of the data. Additionally, the WyACT atmospheric science team is developing and processing new models that will downscale results to 2-km and provide even more precise predictions for the high topographic variability that exists in the Rocky Mountain West, furthering the usefulness of this tool.

Stemming from lessons learned while developing DEVISE, WY-Adapt has further improved UW's CyberGIS capabilities through the relational tabularization of summarized raster datasets. Although some data redundancy does occur, we are finding that users benefit from, as well as expect, responsive spatial queries of time-series data at the highest spatial resolutions (i.e., grid cell). Thus, we aim to refactor DEVISE's schema to incorporate this approach and provide better analytics and flexibility for our users.

Acknowledgment

The authors appreciate the feedback provided by three anonymous reviewers, which helped improve the overall quality of the final manuscript.

References

[1] Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science*. 26: 2267-2276. https://doi.org/10.1080/13658816.2012.722637

[2] Abdalla, R. (2016). Geospatial data integration. In: *Introduction to Geospatial Information and Communication Technology* (GeoICT). Springer, Cham. https://doi.org/10.1007/978-3-319-33603-9_6

[3] Wang, S. (2019). Cyberinfrastructure. *The Geographic Information Science & Technology Body of Knowledge* (2nd Quarter 2019 Edition), John P. Wilson (Ed.). https://doi.org/10.22224/gistbok/2019.2.4