Received: 22 April 2021 | Revised: 13 September 2021 | Accepted: 23 December 2021
DOI: 10.1002/cpe.6843

WILEY

SPECIAL ISSUE PAPER

The Hawai'i Groundwater Recharge Tool

Jared McLean[1,2] | **Sean B. Cleveland**[1,2] | **Kolja Rotzoll**[3] | **Scot K. Izuka**[3] | **Jason Leigh**[1,4] | **Gwen A. Jacobs**[1,2] | **Ryan Theriot**[4]

[1]Hawai'i Data Science Institute, University of Hawai'i, Hilo, Hawai'i, USA

[2]Cyberinfrastructure, University of Hawai'i, Hilo, Hawai'i, USA

[3]Pacific Islands Water Science Center, U.S. Geological Survey, Hilo, Hawai'i, USA

[4]Laboratory for Advanced Visualization and Applications, University of Hawai'i at Manoa, Hilo, Hawai'i, USA

Correspondence
Jared McLean, Hawai'i Data Science Institute, University of Hawai'i, 1659 Oneawa Pl, Hilo, HI 96720, USA.
Email: mcleanj@hawaii.edu

Funding information
National Science Foundation, Grant/Award Numbers: ACI #1450413, CSSI #1931439, CSSI #1931575, OIA #1557349

Abstract

This article discusses the design and implementation of the Hawai'i Groundwater Recharge Tool, an application for providing data and analyses of the impacts of land-cover modifications and changes in precipitation on groundwater-recharge rates for the island of O'ahu. This application uses simulation data based on a set of 29 land-cover types and 2 precipitation conditions to provide users with real-time recharge calculations for interactively defined land-cover modifications. The tool provides two visualizations, representing the land cover for the island and the resultant groundwater-recharge rates, and a set of metrics indicating the changes to groundwater recharge for relevant areas to present a set of easily interpretable outcomes based on user-defined scenarios. Users have varying degrees of control over the granularity of data input and output, allowing for the quick production of a roughly defined scenario, or more precise land-cover definitions. These modifications can be exported for further analysis. Heuristics are used to provide a responsive user interface and performant integration with the database containing the full set of simulation data. This tool is designed to provide user-friendly access to the information on the impacts of land-cover and precipitation changes on groundwater-recharge rates needed to assist in making data-driven decisions.

KEYWORDS

groundwater recharge, hydrology, land cover, sustainability

1 | INTRODUCTION

The Hawai'i EPSCoR Track 1 project, named 'Ike Wai (https://www.hawaii.edu/epscor/ike-wai-project/), has a mission to help ensure Hawai'i's future water security through an integrated program of research, education, community engagement, and decision support. 'Ike Wai's vision is to promote water resource management in Hawai'i that is sustainable, responsible, and data driven. Scientific, cultural, and social dimensions to the problem of water security are integrated in a transparent, stakeholder-driven, and rigorous water research enterprise. In order to support this mission, a robust cyberinfrastructure solution, such as a science gateway, was required.

As cloud-based technologies become more prevalent, science gateways, and science-gateway frameworks have become popular for providing research infrastructure. Organizations such as the Extreme Science and Engineering Discovery Environment (XSEDE)[1] and the Science Gateways Community Institute (SGCI)[2-4] help to support the development of these gateways. Several "as a service" gateways, providing users with remote scientific computing functionality, have been developed such as WS-PGRADE/gUSE project,[5] The HUBzero project,[6] Apache Airavata,[7,8] Galaxy,[9] and Globus.[10] These gateways have a cloud-style model for managing data transfers and sharing data. Existing hydrology gateways and tools, including The CUAHSI Hydrological Information System (HIS),[11] Hydroshare,[12] and Virtual Observatory and Ecological Informatics System (VOEIS),[13] provide server-side support, data entry, and quality control client tools. However, these gateways lack tools for managing and validating multi-domain metadata, integration for multiple data stores, and non-integration of compute resources or flexibility in accessing and selecting those resources.

To address these challenges, Tapis,[14] an open source, scientific application programming interface (API) middleware for hybrid cloud computing and data management that powers a number of current community science gateways, was chosen as the framework to build the ʻIke Wai Gateway. In particular Tapis was selected due to the wide range of functionality that meets the above needs and The University of Hawaiʻi's (UH's) partnership with the Texas Advanced Computing Center (TACC) that supports it. The ʻIke Wai Gateway (http://ikewai.org) currently supports research in hydrology and water management, providing tools to address questions of water sustainability in Hawaiʻi. The Hawaiʻi Groundwater Recharge Tool is one of these tools consisting of a web-based interface on the ʻIke Wai Gateway available to researchers, stakeholders, resource managers, and other interested parties.

The Hawaiʻi Groundwater Recharge Tool[15] (available at https://recharge.ikewai.org) is designed as a publicly available tool to facilitate rapid assessment of the impacts of changing land-cover and precipitation conditions on groundwater recharge, utilizing data collected and processed by researchers at the United States Geological Survey (USGS) and UH. Land-cover changes can have a major impact on groundwater-recharge rates, which in turn can have a major impact on the freshwater availability of an area. The availability of groundwater is vital for meeting human and ecological needs.[16] For example, the island of O'ahu has around 70% of the population of the state of Hawaiʻi and groundwater from O'ahu's aquifers provides about 89% of the freshwater used.[16]

The overall objective of the Hawaiʻi Groundwater Recharge Tool is to provide researchers, stakeholders, and resource managers the ability to rapidly quantify the effects of user-defined land-cover and precipitation changes on groundwater-recharge rates. The tool provides easy access to recharge data via an interactive web interface providing pre-computed groundwater-recharge rates for a set of 29 land covers over the island of O'ahu for two precipitation conditions.[16,17] O'ahu was chosen for the pilot application because input files to calculate recharge were available[18] and it aligns with ʻIke Wai's study sites. Users are able to define changes in land cover and precipitation and retrieve a set of updated groundwater-recharge values for the selected conditions. This article describes the development of the tool to provide a set of versatile and easily interpreted visualizations and metric-based analyses for the potential impacts of simulated changes in land-cover and precipitation conditions on groundwater-recharge rates.

2 | BACKGROUND

Land-cover types used by the Hawaiʻi Groundwater Recharge Tool are based on those identified by studies from Engott et al.[18] and Izuka et al.[16] and are referenced by unique identification numbers. The baseline land-cover scenario provided by the application is based on land-cover data for the island of Oʻahu as of 2010; however, this baseline land cover can be redefined by the user if they have a more preferred or up to date data source.

Precipitation is commonly the largest source of groundwater recharge and influences other hydrological processes used in determining groundwater-recharge rates. Recharge data are provided for two different precipitation conditions:

1. Average precipitation for 1978–2007.[19]
2. Projected precipitation for 2041–2071.[20]

The default precipitation condition provides a baseline-precipitation estimate based on data for the island of O'ahu during 1978–2007.[19] The alternate precipitation condition uses a statistically downscaled projection of rainfall during 2041–2071 for a Representative Concentration Pathway warming scenario with total radiative forcing of 8.5 Watts per square meter by the year 2100 (RCP 8.5).[20] Although other precipitation projections are available, the RCP8.5 2041–2071 scenario was chosen for the pilot application, because it represents a condition drier than conditions during 1978–2007 and the mid-century time frame is generally consistent with the planning horizon of resource managers.

2.1 | SWB modeling

The Soil Water-Balance (SWB) program developed by Westenbroek and others[18] was used to simulate groundwater-recharge rates for the entire range of possible land-cover conditions and for each of the two precipitation conditions. The SWB program was chosen to estimate recharge rates over water-balance models traditionally used in Hawaiʻi, because the SWB output is grid-based, simplifying data handling and presentation in this application. The SWB program uses gridded precipitation, soil, land-cover, and other data to estimate the distribution and timing of net infiltration out of the soil root zone, known as groundwater recharge. The SWB program accounts for water entering, leaving, and being stored within the soil on a daily basis. The SWB program uses the estimated capacity of soil-water retention, antecedent soil-moisture content, and loss through runoff and evapotranspiration—the combined effects of evaporation of water from plant and ground surfaces and loss of water through plant stomata

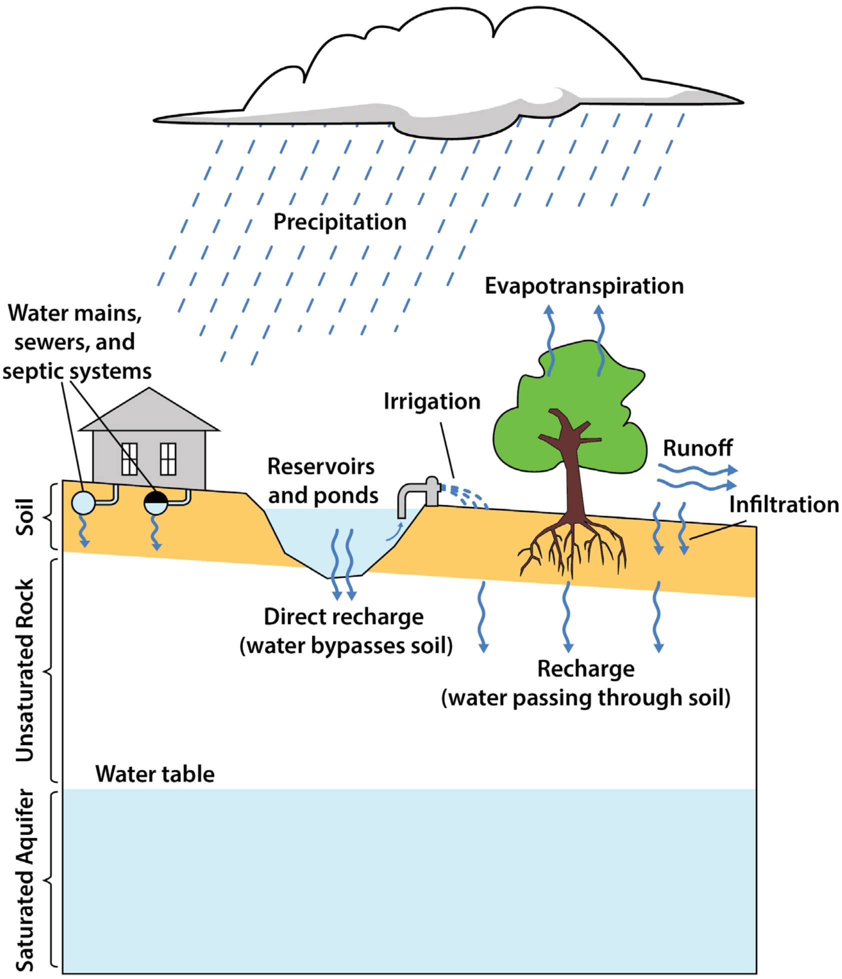

FIGURE 1 Conceptual diagram of soil-water balance processes[21,22]

(transpiration)—to estimate the amount of water that penetrates through the soil and into the groundwater system (Figure 1). Additionally, water that leaks from some land-cover types such as reservoirs and ponds bypass the soil root-zone layer and contributes to groundwater recharge directly without losses from evapotranspiration.[18]

The SWB input datasets were based on spatial datasets from Engott et al.[18] and Izuka et al.[16] SWB uses a grid-based value system whereas these datasets were polygon based; therefore, these datasets were rasterized onto a grid of 689 × 887 cells over the island of O'ahu at 75 m × 75 m resolution. The Hawai'i Groundwater Recharge Tool adopted this gridding system for the land-cover and groundwater-recharge values.

The recharge values of the SWB output are compiled in a network common data format (NetCDF) file in four dimensions—x and y gridded geospatial values using a Universal Transverse Mercator (UTM) coordinate system, land-cover type, and precipitation condition. The Tapis platform, used to store the data for the application and to provide API endpoints, utilizes MongoDB, a no-SQL (does not use structured query language (SQL)), JavaScript object notation (JSON) based database, as a metadata storage system.[14] For compatibility with this platform, the SWB output was reserialized into a set of JSON documents split out by geospatial position. More information on this process is covered in Section 3.

While a statically defined set of groundwater-recharge estimations based on pre-computed recharge rates has limitations, the large amount of computational overhead associated with performing dynamic groundwater-recharge simulations would not allow for real-time data pullback and analysis. The goal of this application is to allow for the rapid development, testing, and comparison of land-cover scenarios and outcomes.

2.2 | Limitations and responsible usage

Proper use of the Hawai'i Groundwater Recharge Tool and interpretation of its results requires that users have a conceptual understanding of how the web application works and a full understanding of its limitations. The application will not prevent a user from attempting unrealistic scenarios.

The website has a "Background" page, available at http://recharge.ikewai.org/#/background, that describes how the application assesses recharge when a user changes land cover or precipitation conditions and how these values were computed using SWB. The background page also includes a list of limitations. These limitations stem in part from the original computations of the soil water-balance model and in part from the goal to provide users with a responsive, real-time interface by using a statically defined model.

Before a user is allowed to do any analysis on the web application, the user must agree to "Conditions of Use" that states that the user is responsible for understanding the background and limitations of the application.

3 | ARCHITECTURE AND CYBERINFRASTRUCTURE

The Hawai'i Groundwater Recharge Tool was built using the Angular framework, a TypeScript-based framework for building modular web applications, and is built on Tapis[14] as part of the 'Ike Wai Gateway[23] suite of tools. This section presents the underlying technologies that support the storage and retrieval of recharge data for the application.

3.1 | Tapis

Tapis is an open source, platform-as-a-service for hybrid cloud computing, data management, and reproducible science. It uses standards-based technologies and community promoted best practices to enable users to run code, manage data, collaborate meaningfully, and integrate anywhere. Tapis serves as a middleware powering a number of community science gateways. It is a multi-tenant, cloud-native distributed system.

Tapis services run as Docker containers, orchestrated as a single microservice architecture. The platform contains three logical tiers: platform services, science APIs, and support services. Platform services contain services providing identity and API management, client registration, tenant administrative services, and documentation. Science APIs contain the primary Science as a Service (ScaaS) functionality such as data and job management, application publishing, and notifications used to power science gateways. This can be subdivided into a frontend collection of loosely coupled services, microservices, and backend service workers. Microservices serve the user Hypertext Transfer Protocol (HTTP) requests to the representational state transfer (REST) APIs and service workers handle processing of asynchronous requests such as data movement and application publishing. Finally, support services include databases, message queues, caches, object stores, service discovery, notification relays, and websocket proxies.

Each instance of Tapis can support one or more tenants, a group of users, applications, and entitlements. Individual tiers or microservices can be replicated, configured, and scaled in support of a specific tenant, or shared to better leverage a single, consolidated resource footprint.[24]

3.2 | Spatial support for data queries

Tapis provides a metadata service backed by a MongoDB engine for storing metadata objects as JSON documents. MongoDB provides geospatial query support for GeoJSON objects, a format for encoding geographic data into JSON. UH extended the Tapis metadata service with a geospatial index by generating a 2dsphere index on the location property of records in the database.[22] This allows for any metadata records containing a field named "loc" in its value, with a GeoJSON object as its key, to be efficiently checked by any of MongoDB's geospatial queries. This capacity is leveraged by the Hawai'i Groundwater Recharge Tool to make client-side data retrieval as efficient as possible. Additionally, other tools and services utilizing the Tapis metadata database can benefit from accelerated spatial queries.

3.3 | Serializing and storing model outputs

A requirement of the Hawai'i Groundwater Recharge Tool was to support users' ability to select regions of the island of O'ahu and modify how that region was used by defining its land-cover type out of the 29 defined types. Once the land-cover type is changed for a region, the corresponding groundwater-recharge values for the area must be updated. The set of pre-computed groundwater-recharge values for each combination of land cover and precipitation were provided for a grid of 689×887 cells that are $75 \text{ m} \times 75 \text{ m}$ each and cover all of O'ahu. Having 29 land-cover types and 2 precipitation conditions results in 58 total recharge values for each grid cell. Data processing efforts output the data in a NetCDF format containing 58 matrices of recharge values with latitude and longitude data in a separate matrix.[22] The application was enabled to access this information with Tapis by storing the model output data within the metadata database and querying the appropriate grid cells with spatial queries. To store the model output data, an ingestor application was developed (https://github.com/UH-CI/landuse_ingestor) that transformed the NetCDF model output to a set of JSON documents for each cell that holds geospatial information such as UTM coordinates, (x, y) indices relative to the data grid, and land-cover to recharge value mappings for each scenario.[22] These JSON documents were then added to the Tapis metadata store and accessed via spatial queries wrapped in RESTful API requests to the Tapis metadata endpoint by the web interface.

4 | APPLICATION FEATURES AND DESIGN

The Hawai'i Groundwater Recharge Tool provides a workflow for users to construct and visualize custom land-cover scenarios and view the potential outcomes of these scenarios relative to groundwater-recharge rates. Scenario construction can be done using the in-application tools or by importing pre-created land-cover or other geospatial data into the application. Users can also select the precipitation condition they want to use for the scenario. Land-cover values that differ from the baseline land usage trigger the application to retrieve the new groundwater-recharge values based on the SWB simulation values for that land-cover type, geospatial position, and precipitation condition. A visualization and set of metrics of the estimated differences in groundwater recharge are then presented to the user. The simulated land cover, groundwater-recharge values, and metrics can then be exported from the application for further analysis, or imported back into the application. The implementation and usage of the core application features are covered in the following sections.

4.1 | Data views

The application uses a multiple-document interface allowing for simultaneous scenarios and comparison of different land-cover patterns. The application holds a set of elements containing independent maps of land-cover and recharge data that can be freely positioned within the browser (Figure 2). The primary workspace is designed to be expanded allowing windows to be placed outside of the immediate reference frame of the browser. The user can add or remove these maps to adjust the set of simulations being tracked. Since these map instances are designed to be

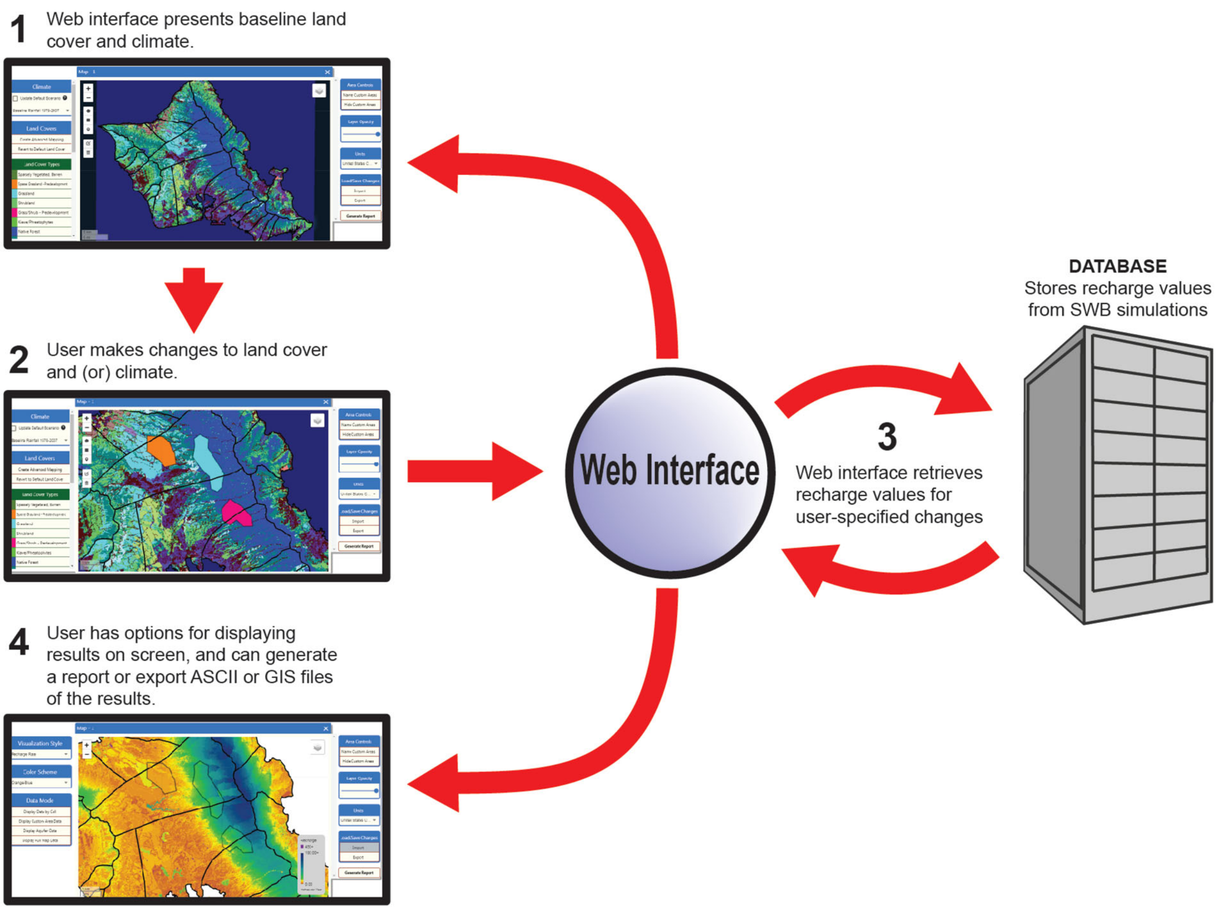

FIGURE 2 The general workflow established by The Hawai'i Groundwater Recharge Tool[21]

independently manipulated, further discussion of the data schema and control mechanisms for the application may be considered as local to an individual map instance unless stated otherwise.

Two primary visualization instances are provided to the users, based on the two types of data tracked by the application: a visualization of the land cover and a visualization of the groundwater-recharge rates. A foundational map of the island of O'ahu is set using Leaflet, an interactive map library for JavaScript. Each visualization provides a raster overlay of the respective dataset. Data for the application are provided as a grid of values at 75 m × 75 m resolution defined by cell centroids using the World Geodetic System from 1984 (WGS84) geodetic datum. Although the raster datasets for this application are in a UTM projection, Leaflet uses unprojected latitude-longitude values for positioning map data. To work around this issue, proj4, a JavaScript package for transforming geographical coordinates between coordinate systems, is used to translate between UTM and unprojected coordinates as needed.

Each value is colored using a categorical color scheme, for the land-cover visualization, or using a continuous color scheme, for the groundwater-recharge visualization—the generation of these color schemes will be discussed in the next section. The land-cover visualization serves as the default view of the map; the groundwater-recharge visualization provides a view of the outcomes of modifications made to the land-cover values. The land-cover visualization uses a modified version of Leaflet's drawing tools to allow users to draw polygons on the map representing areas for land-cover updates to be applied. In addition to the built-in polygon and rectangle drawing tools, a tool for selecting a single map grid cell was added for making fine adjustments to an area. Checks were also added to the shape-editing feature to limit the number of shapes that could be modified at a time to prevent performance issues if a very large number of shapes are being tracked. Drawn areas can be selected or deselected for receiving updates and are tracked in the output as individual metrics. Controls are provided for modifying the land-cover types for the selected areas. Land-cover values in a selected area can all be updated to one of the available land-cover types, or a mapping of the currently represented land-cover values to new ones can be constructed. For example, a selected area that both contains grassland and shrubland could be updated to golf course or mapped to golf course and diversified agriculture, respectively. An option is provided to change between the two included precipitation conditions, which is applied globally to the user scenario. Further, an option is provided to modify the baseline data to this new precipitation condition; thus, a user can evaluate the impacts of land-cover and precipitation condition changes separately or in combination.

The groundwater-recharge visualization reflects the simulated land-cover modifications, providing a view of the resultant groundwater-recharge rates for the island and the metrics associated with these modifications. The user can change between modes that display various metrics for the entire island, a single selected data point, a set of selected aquifer systems, or a set of selected user-defined areas. These metrics display the groundwater-recharge rates at baseline and for the current modifications, the area, and the volumetric difference and percentage change in recharge rate respective to baseline. A graph of the baseline and current scenario values is also provided. Additionally, the data overlay can be changed to represent the groundwater-recharge values, percentage change, or difference from baseline for each value. Figure 3 demonstrates the workflow of the application and its data retrieval. A report containing a comprehensive summary of these metrics for all relevant areas can be brought up in a separate document within the page for side-by-side analysis, and can be downloaded in Portable Document Format (PDF). All land-cover, groundwater-recharge, and user-defined area data can be exported from the application for further analysis or to restore the application state at a later time.

To allow for an intuitive control schema for each of the visualization styles, the set of user controls is broken into two parts: one which is unique to each visualization, and one which contains global controls for the entire map instance, such as the ability to change unit types or hide user-defined areas. The unique controls are placed in a collapsible menu on the left side of the map and the global controls in a collapsible menu on the right. The metrics view for the groundwater-recharge visualization is displayed in an additional collapsible panel along the bottom of the map.

4.1.1 | Additional visualization integrations

The ability to construct and compare multiple scenarios on-the-fly also makes this application a good use case for interactive collaboration technologies such as SAGE2 (Scalable Amplified Group Environment). SAGE2 is a web-based platform that allows multi-user interaction within a shared environment to enhance research collaboration.[25] As a web application, this application is supported natively by SAGE2 and is easily accessible. Integration with technologies such as SAGE2 have the potential to greatly enhance its usefulness by allowing multiple researchers to collaboratively interact with and construct land-cover scenarios and compare their projected outcomes.

Additionally, an adaptation of this project has been created to utilize ProjecTable, a physical, 3D, geographical visualization table developed by the Laboratory for Advanced Visualization and Applications (LAVA) at UH Mānoa. This visualization platform projects geographic visualizations onto high definition three-dimensional physical terrain models and provides a hands-on user interface for manipulating the visualization view.[26] By presenting geographical data projected into a physical 3D environment, information corresponding to terrain-based variables is significantly easier to establish than it may otherwise be in a two-dimensional projection on a computer monitor. This may be especially useful for the land-cover component of this application, since certain land-cover scenarios may be infeasible as a result of terrain-based factors.

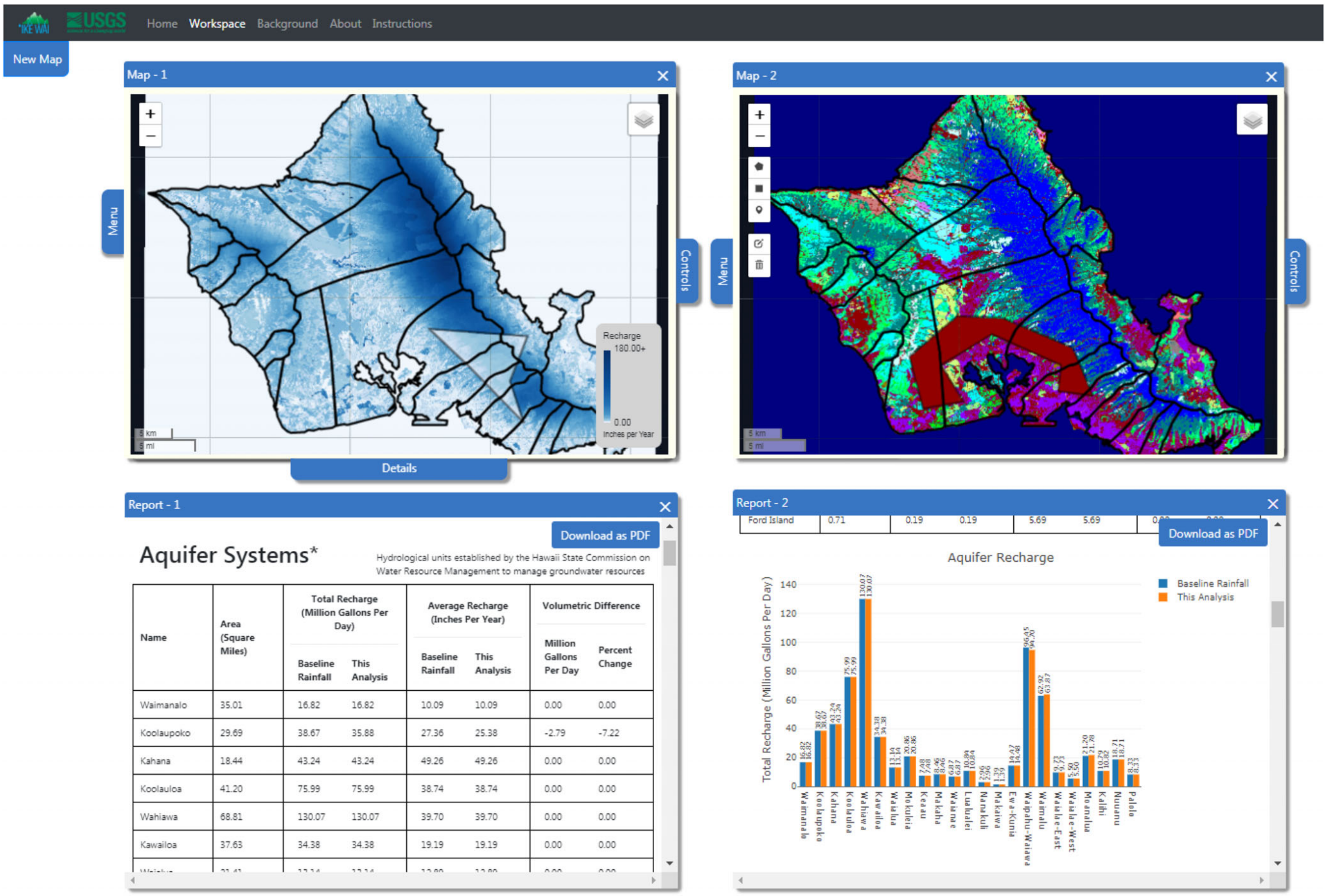

FIGURE 3 The Hawai'i Groundwater Recharge Tool interface displaying multiple maps and generated data reports

4.2 | Data coloring

The map-based visualization component of the application relies heavily on coloring methods to present the data in an understandable fashion. Generating a reasonable color scheme for each set of data being presented is an important factor in the efficacy of these visualizations. This section will cover the challenges encountered and the principles used for generating these color schemes.

4.2.1 | Land-cover visualization

Due to the relatively large number of potential land-cover types, displaying each as its own categorically distinguishable color is challenging. A color palette of as few as seven different colors is estimated as the maximum for quick distinguishability when colors are not distributed in a regular manner.[27] Since users may be working with an arbitrary number of the potential land-cover types at a given time it is also difficult to limit the working set of land covers. It is therefore acknowledged that any potential color scheme used for the land-cover raster will result in sets of colors difficult to quickly distinguish. The primary drawback of this is in the difficulty in determining land cover via a color legend. Identifying a plot of land's constituent cover types using this mapping would be a slow and difficult process where sets of colors that are not easily distinguishable are involved. To work around this issue, a mechanism was implemented to quickly identify the land-cover type associated with a particular spot on the map. Hovering over a point on the map will highlight the cell under the pointer—the 75 m × 75 m grid cell centered nearest the pointer location—and display the set land cover for that cell via a popup (Figure 4). Additionally, this on-hover function displays both the currently set land-cover value and the baseline land-cover value if the land cover for the cell has been changed.

To produce a set of colors as distinct as possible, the color scheme was produced via a permutation of subdivided color channels. Excluding black and white, the number of potential colors for a set of red, green, and blue (RGB) color channel values is equal to the product of produced values for each color channel minus two, assuming the two extreme values are included. This means that, for the 29 required colors, the number of subdivisions in the color channel must be composed of the set {3, 3, 4}. Green-blue color combinations were found to be the most difficult to

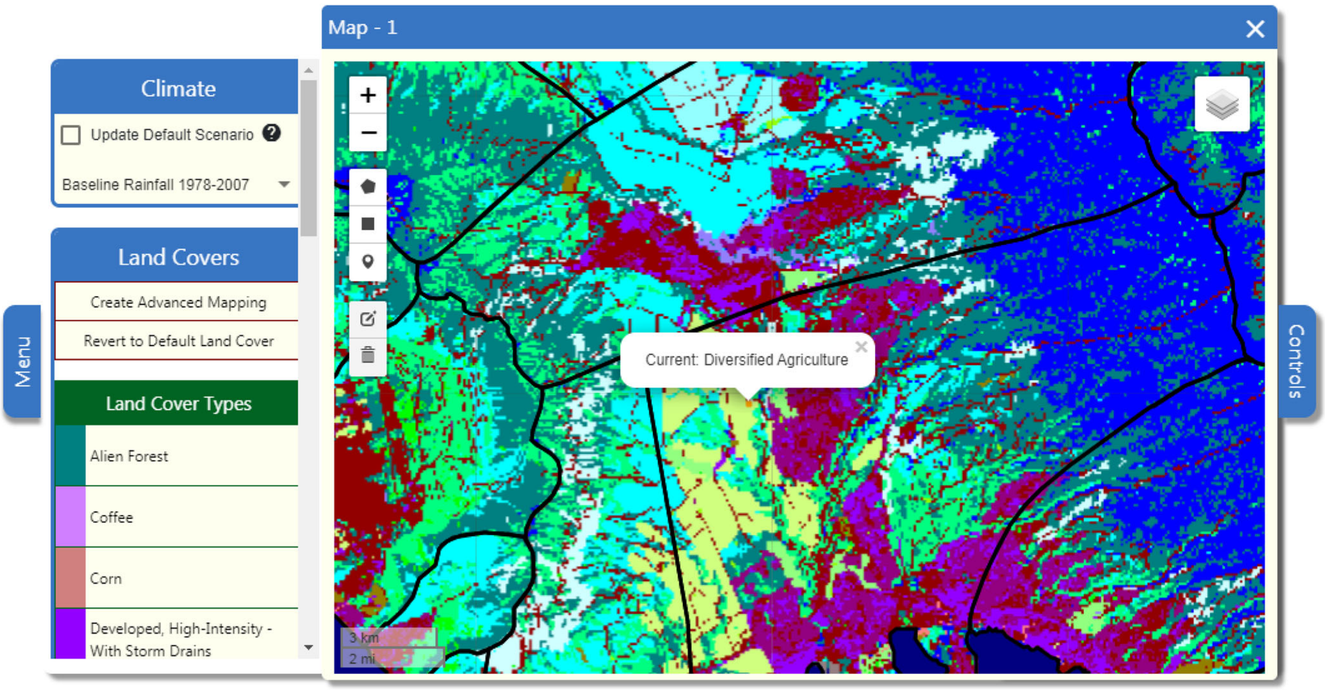

FIGURE 4 The land-cover visualization with the generated set of categorical colors. Hovering over an area displays the land-cover type

differentiate, so the red color channel was selected for the larger division set. Divisions were made using linear RGB color space for the red color channel, and standard RGB color space for the green and blue color channels. Linear RGB color space corrects for the differences in computational and human perceived brightness; however, due to the overwhelming effect of green in comparison to blue on relative luminance, colors brighter in the green and blue color channels were more difficult to differentiate than those with lower values. Green–blue color combinations with higher green values will have more similar relative luminance values which relates to lower contrast.[28] The darkening effect of proportionally dividing these channels in standard RGB color space was used to offset this, whereas the red color channel was scaled on perceptual brightness using linear RGB color space.

4.2.2 | Groundwater-recharge visualization

Color scaling for the groundwater-recharge visualization was provided in two different color schemes: rainbow and monochromatic. The rainbow color scheme, a polychromatic color scheme scaling through the standard rainbow progression (red, orange, yellow, green, blue, indigo, violet), is provided due to its prevalence in water science and related fields. The target audience of the application is likely to be familiar with this color scheme; as such, it is provided to increase user comfort with the interface. Research has shown, however, that this color scheme has the potential to be misinterpreted due to human perception of the colors in the progression. Colors such as yellow and green have a relative luminance greater than that of the surrounding colors, creating perceptual artifacts in the color mapping since these colors are prone to drawing attention.[29] In contrast, monochromatic color scales are more perceptually consistent and can be translated to grayscale without loss of data.[29] To produce the monochromatic scaling, a sequential color scheme was selected using Colorbrewer 2.0, a web-based tool providing concise color schemes for cartographic visualizations.[27] Colorbrewer 2.0 provided a nine-part single-hue blue color scheme, which was then scaled to an exponential gradient in Lab color space. Lab color space is designed to yield a perceptually consistent color scale[30] and creates a consistent transition between the colors. An exponential scaling is employed for groundwater-recharge data since most of the values fall in lower ranges; however, the entire range of values spans a couple orders of magnitude. Without an exponential scaling, a majority of the features in the visualization would be lost. Recharge values were capped at 180 inches per year, with higher values taking on the maximum color (Figure 5).

The color scheme for the final two visualization components, the difference and percentage change mappings, was constructed using a diverging color scheme between red and blue. The difference scale was capped at $+/-10$ inches per year and the percentage change scale at $+/-100\%$. This color scheme was similarly adapted from a color mapping provided by Colorbrewer. To create a smooth color gradient, the two color extrema for an eleven-class diverging color scale were taken, and each extremum scaled to white using Lab color space. A 200-part total mapping was used to provide a smooth transition between the colors.

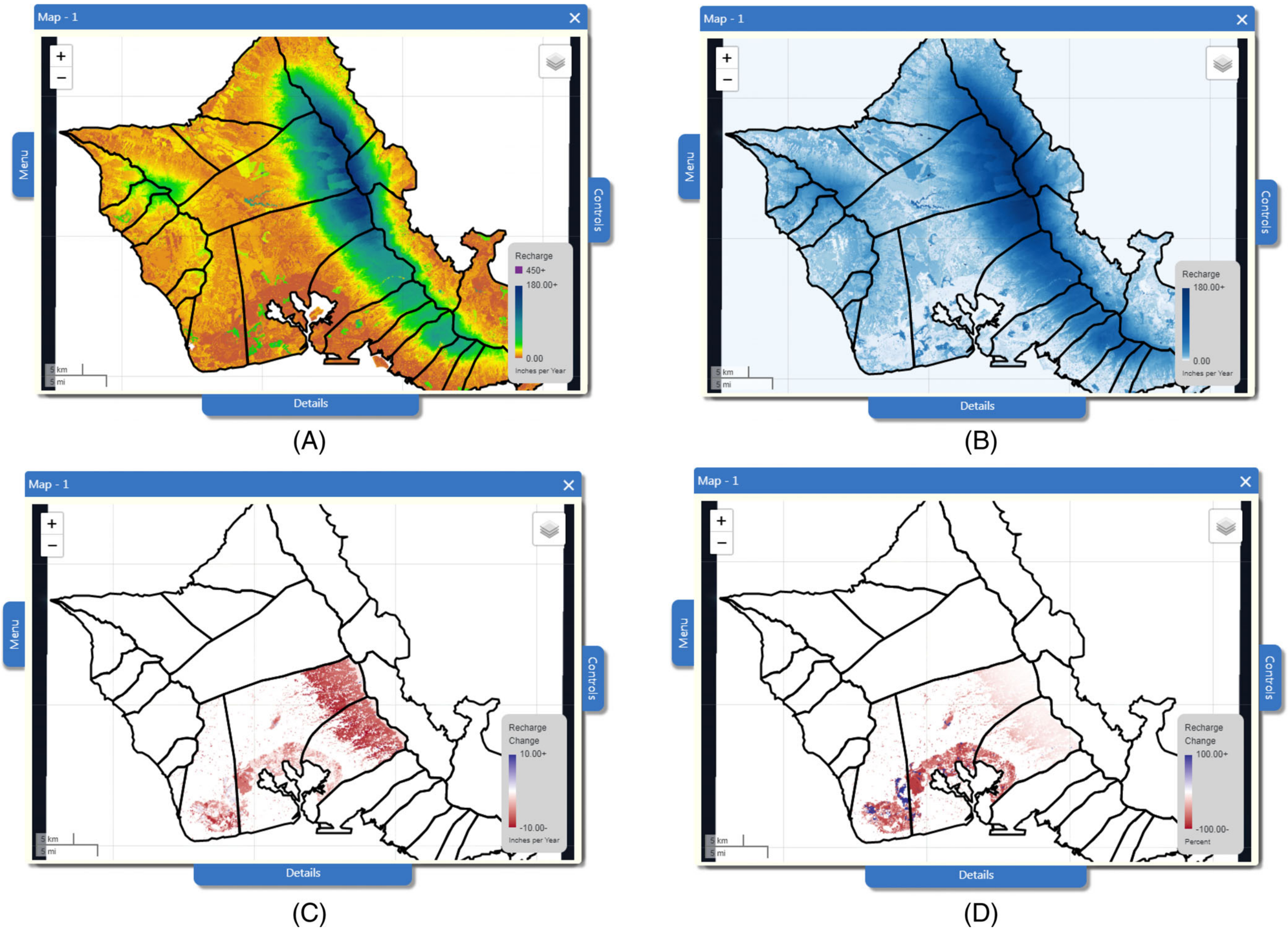

FIGURE 5 Groundwater-recharge visualization color schemes. (A) Rainbow color gradient for groundwater-recharge rates. (B) Monochromatic color gradient for groundwater recharge rates. (C) Diverging color gradient for groundwater-recharge difference map. (D) Diverging color gradient for groundwater-recharge percent change map

4.3 | Data importing and exporting

To improve the versatility of the data input and analysis for the application, several options are provided for importing and exporting data. Data can be imported either using a georeferenced land-cover raster, provided as some gridded subset of land-cover values defined by the application, or as a shapefile, an ESRI geospatial data storage standard.[31] Land-cover rasters can be in American Standard Code for Information Interchange (ASCII) or coverageJSON[32] formats. ASCII datasets are a space-separated array of land-cover codes, preceded by a six-row header with the number of columns, number of rows, lower left x and y UTM coordinates, the size of the cells, and a no-data value. The no-data value will be ignored, and the pre-existent data used for any cells marked as such. For cell sizes that do not match the 75 m × 75 m resolution of the application, the resolution will be adapted by combining raster cells into the most dominant value for the given position or cells will be skipped for higher resolutions. Additionally, in the case of grids misaligned with that provided by the application, the imported grid will be shifted to align with the nearest cell. This provides a potential spatial error of up to 37.5 m for this alignment. The coverageJSON format provides a JSON-based geospatial raster with fields indicating the set of values and positions.[32] Uploads in this format should have a set of coordinates consistently spaced on the resolution of the data.

For uploaded shapefiles, three different formats are available. The shapefile can be uploaded as a set of user defined areas. These areas are treated the same as areas defined in the application using the provided drawing tools and will have individually defined metrics associated with them for generated reports. The shapefile can be uploaded as a reference layer, which provides a toggleable overlay of the defined areas. Reference layers are not used by the application for analysis, but can be any map, such as property or planning maps, that help users define areas of interest. Finally, shapefiles can be provided purely as land-cover updates. In any of the provided upload types, a property can be attached to the shapefile that defines an associated land-cover code. The land-cover code can be used to update any internal data values to this value. Imported shapefiles specified as purely land-cover updates will not have their constituent defined areas displayed to the map, but any relevant values will be updated.

Any modifications to land cover can also be specified as a new set of base data values. If this option is selected, any additional changes will be compared with the imported dataset rather than the application-defined default dataset. This feature can be used for easier direct comparison of two potential scenarios or for providing simple user-defined updates to the default dataset.

Data exports are provided in the same three formats as allowed for imports. Three geo-registered data types are available for export: a raster of land-cover values, a raster of groundwater-recharge values, and the set of user-defined areas. The raster datasets are provided in either the ASCII or coverageJSON formats. The user-defined areas are provided as a shapefile. To avoid the potential ambiguity of multiple defined land-cover types within a given user-defined area, the set of properties provided with exported shapefiles are limited to names provided to the areas and exclude their land covers. The data exported from the application can, however, be used to restore the state of a given map by exporting the land-cover raster and set of user-defined areas, then re-importing these data to the application.

4.4 | Database interaction and index packing

Having a large set of total simulation data—data for the island of O'ahu at 75 m × 75 m resolution over 29 land-cover types for 2 precipitation conditions—only the current working set of groundwater-recharge values, made up of the baseline values and updated values based on modifications to the land cover for each precipitation condition, are tracked by the application. Any updates to land cover trigger a request to the MongoDB database holding the primary simulation data. A spatial query, containing the geometry of the updated area, is used to fetch the appropriate data. Two primary limitations are imposed on constructed queries that must be handled by the application: a maximum of 10,000 values can be returned by any single query and, having the geometry encoded into the request uniform resource identifier (URI), the request must be limited to a safe number of characters handleable by the browser and database API. To this end, modifications that contain a large number of data points or that contain very complex geometries must be repackaged into a query that falls within these limitations. The repacking process employed constructs a bounding box for the set of values that must be included in the query. This bounding box is then broken into a set of subgrids containing no more than 500 values. Once the set of subgrids is computed, a bounding box of the modified values in each subgrid is computed. This set of sub-bounding boxes is then used as the set of objects constructed into queries (Figure 6). This method does not necessarily construct a tight boundary for the modified values but it provides a quick estimate that is guaranteed to contain the full set of modified values. Using rectangular geometries also reduces the overhead required to verify the query length and the complexity of query parsing. Limiting the maximum subgrid size to 500, as opposed to the API maximum values, presents a few major benefits. The smaller subdivisions reduce the number of superfluous values included in

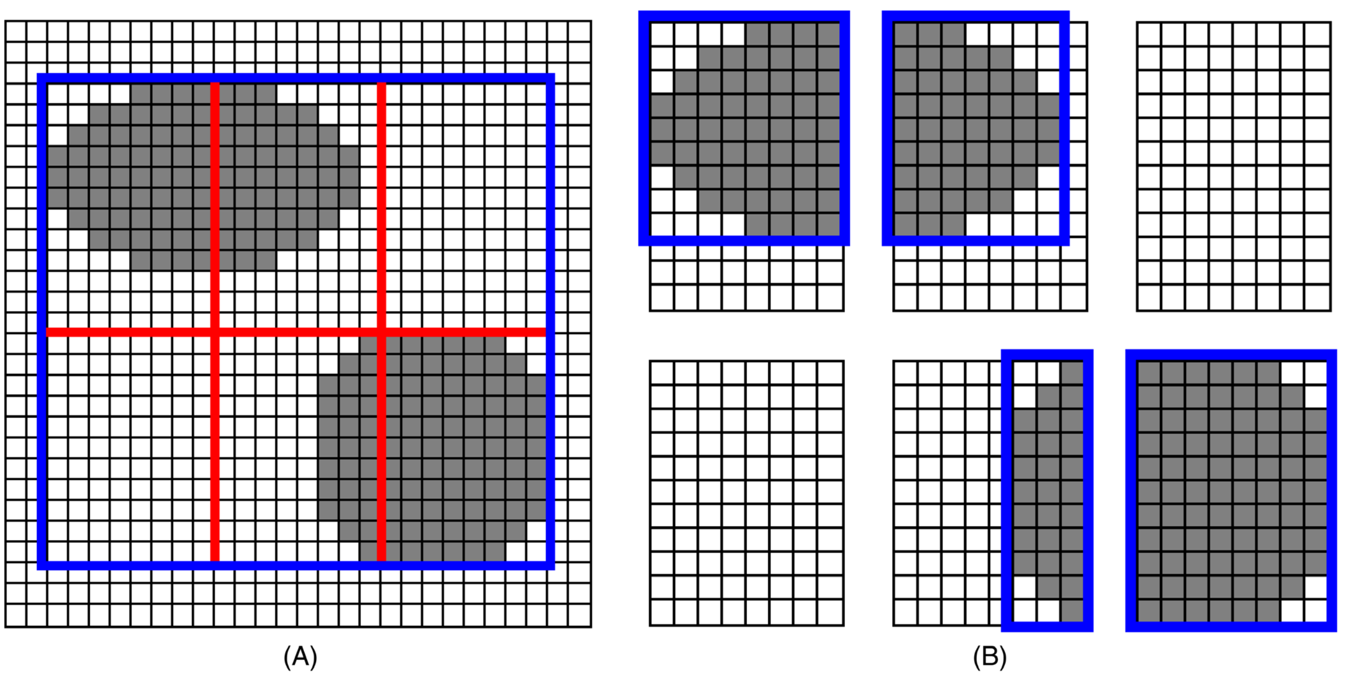

FIGURE 6 An example of the repacking algorithm used (not to scale). Modified cells are colored in gray. (A) A bounding box (blue) is created for the set of modified cells. This bounding box is then divided into a set of subgrids (red) containing no more than 500 data cells each. (B) Bounding boxes (blue) on the set of modified cells for each constructed subgrid. These geometries would be used to construct the query. Empty subgrids are ignored

the query geometry and the download size of each query result. Additionally, while this results in a larger number of total requests, the processing overhead on the API is reduced, resulting in faster response times and improved scalability for large numbers of users performing concurrent requests.

When a land-cover update is triggered for a given area, the set of internal values is computed using the bounding box of the updated area and a raycasting algorithm. In the case of a set of multiple features being updated at once, a determination must be made of whether to perform local or global value repacking when a feature violates the query limitations. Queries are constructed from each feature geometry; however, performing local value repacking, that is, repacking only features that require it, could result in overlapping indices being queried if features are close to each other due to the bounding box heuristic utilized by the repacking algorithm. Global repacking solves this issue by repacking every modified value if a single feature in the current update fails the validation check. Global repacking incurs extra overhead if values that would not need to be included in the query under local repacking end up within the repacked geometry set. This would likely be the case if some set of small features that did not require repacking were sufficiently far away from a set of features that did. Weighing the pros and cons of each of these methods with the likely use cases of this application, global repacking of values was employed. In general, a set of areas being modified within a single query instance are reasonably likely to be in a geographically similar area. Additionally, due to the bounding boxes of each individual subgrid being computed to get a reasonably tight bound of the updated values, and the fact that features that do not require repacking may not be particularly large or intricate, it is likely that the additional overhead of unnecessary values will be small.

A further situation where this method must be employed is the case where an imported dataset contains a set of updated land-cover values without a bounding geometry. Although an index-based query can be constructed from the updated values, this is typically infeasible due to the limitations on query length and the resolution of the dataset, which results in a typically large set of individual values being updated. It is more efficient to pack the updated values into a set of bounding geometries than to perform a set of satisfactory index queries that could be on the order of hundreds or thousands of subqueries.

5 | TOOL USAGE AND SCALABILITY

As of the time of writing, analytics on the tool have shown around 700 user sessions over the previous year. To evaluate the scalability of the tool, Locust was used to stress test the Tapis metadata endpoint that services geospatial queries. Locust is an open source load testing tool for python that will simulate concurrent users by submitting API requests to an endpoint at a defined rate. An API request generated by the application pulling a rectangular area containing 483 cells was used for testing. Locust was configured to submit groups of 25 queries per user with a pause of 1 s before submitting a new batch of queries to simulate a heavy application user interaction.

Multiple instances of the Tapis metadata service can be run to parallelize request processing. Tests were executed using one, two, and three instances of the metadata service. Three instances of the service is currently the maximum the system running the metadata services and other Tapis core API services could handle, based on the resource utilization when under the 200 user Locust load—based on a CentOS 7 virtual system with 6 CPU cores and 10GB of memory. Additional cores and memory would allow the application to scale further. During initial testing of the system at 4 cores and 8 GB of memory it was only able to handle up to 100 concurrent users.

Tests were performed maintaining a user load of 50–200 users in increments of 50. A ramp up rate of 5 user per second was used at initialization. That is, 5 simulated users were established per second up to the maximum number of concurrent users. Each test was executed for 5 min. The failure rate, average response time, and 90 percentile response time of each test are provided in Table 1. One additional test of 500 concurrent users was executed for each configuration, but this resulted in impractically high failure rates of around 10%.

While the deployment of multiple instances of the metadata service would be expected to provide better overall response times, the average response times were relatively consistent between each number of instances, scaling primarily based on the number of concurrent users, with minor benefits only appearing at high user loads. However, having multiple instances of the service running does provide some additional benefit by introducing redundancy in the event of an instance of the service failing.

6 | FUTURE OPPORTUNITIES

To broaden usage and the impact of the tool, the addition of islands other than O'ahu and other future precipitation projections can be explored as interest in expansion grows. Additionally, a feature to ingest new recharge datasets for other areas to make the tool more adoptable for the broader community can be undertaken.

The interface was designed to support similar datasets for other geographic locations. Reducing the memory utilization of the application may be required for larger datasets. Improvements to the internal data representation to streamline the mapping of values to the data raster would help with this. Removing background data values from the set of cached data—these values will always be zero—and maintaining a list only of the values that have been modified for tracking changes to the dataset should improve the memory utilization.

TABLE 1 Tabulation of the response times and failure rates for sets of 50–200 simulated concurrent users with one, two, and three instances of the metadata API

API instances	Users	Failures (%)	Average response (ms)	90 percentile response (ms)
1	50	0.058	1192	1900
1	100	0.15	2488	4400
1	150	0.38	3844	6700
1	200	0.64	5231	9000
2	50	0.041	1182	1900
2	100	0.27	2489	4400
2	150	1.0	3765	6600
2	200	0.76	5222	9000
3	50	0.34	1203	1900
3	100	0.78	2531	4500
3	150	0.45	3798	6600
3	200	0.53	5120	8700

Since the application is able to handle multiple map instances, a mechanism to clear memory from inactive map instances would be useful for maintaining a reasonable level of memory usage with multiple instances of data of a higher resolution or covering a larger geographical area. This should maintain performance for active map instances and greatly reduce the memory overhead of maintaining multiple instances in exchange for some delay when swapping between instances while the data are re-loaded.

The windowing interface used by the application is extensible to other types of visualization elements and interface designs by replacing the items displayed in window panels and data interaction. The coloring and interaction structure used by the application should also be applicable to other raster-based datasets.

7 | CONCLUSION

The Hawai'i Groundwater Recharge Tool attempts to overcome some of the challenges that are present in estimating the effects of land-cover and climate changes on groundwater-recharge rates with currently available data sources. Improving the accessibility and interpretability of these data should reduce the overhead required for researchers working with groundwater-recharge analyses. Presenting a set of concise metrics and allowing for the exportation of data for more specialized analysis, in a user-friendly and responsive interface, should allow for more rapid construction of models for the impacts of simulated changes on groundwater-recharge rates. Furthermore, the ability to extend this project and its visualization format to a wide variety of additional visualization tools and platforms can further enhance the ability of researchers to make meaningful use of the data provided.

ACKNOWLEDGMENTS

This work was made possible by grant funding from the following:

1. RII Track 1: 'Ike Wai Securing Hawai'i's Water Future NSF OIA #1557349
2. Project Tapis: Next Generation Software for Distributed Research NSF CSSI #1931439, #1931575
3. Collaborative Research: SS2-SSI: The Agave Platform: An Open Science-As-A-Service Cloud Platform for Reproducible Science NSF ACI #1450413

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

The data used by this study are openly available through the Ike Wai Gateway at recharge.ikewai.org.

ORCID

Jared McLean https://orcid.org/0000-0002-6347-0734
Sean B. Cleveland https://orcid.org/0000-0002-7130-3434

REFERENCES

1. Dahan M, Towns J, Cockerill T, et al. XSEDE: accelerating scientific discovery. *Comput Sci Eng*. 2014;16(5):62-74. doi:10.1109/MCSE.2014.80
2. Wilkins-Diehr N, Crawford TD. The science gateways community institute and the molecular sciences software institute. *Comput Sci Eng*. 2018;20(5):26-38. doi:10.1109/MCSE.2018.05329813
3. Gesing S, Lawrence K, Dahan M, Pierce ME, Wilkins-Diehr N, Zentner M. Science gateways: sustainability via on-campus teams. *Future Gener Comput Syst*. 2019;94:97-102. doi:10.1016/j.future.2018.09.067
4. Gesing S, Wilkins-Diehr N, Dahan M, et al. Science gateways: the long road to the birth of an institute. Proceedings of the 50th Hawaii International Conference on System Sciences; 2017
5. Kacsuk P, Farkas Z, Kozlovszky M, et al. WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities. *J Grid Comput*. 2012;10(4):601-630. doi:10.1007/s10723‐012‐9240‐5
6. Mclennan M, Kennell R. HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Comput Sci Eng*. 2010;12(2):48-53. doi:10.1109/MCSE.2010.41
7. Pierce M, Marru S, Gunathilake L, et al. Apache Airavata: design and directions of a science gateway framework. *Concurr Comput Pract Exper*. 2015;27(16):4282-4291. doi:10.1109/IWSG.2014.15
8. Marru S, Gunathilake L, Herath C, et al. Apache airavata: a framework for distributed applications and computational workflows. Proceedings of the 2011 ACM Workshop on Gateway Computing Environments; 2011:21-28.
9. Afgan E, Goecks J, Baker D, Coraor N, Nekrutenko A, Taylor J. Galaxy: a gateway to tools in e-science. *Guide e-Science*. 2011;145-177. doi:10.1007/978‐0‐85729‐439‐5&uscore;6
10. Ananthakrishnan R, Chard K, Foster I, Tuecke S. Globus platform-as-a-service for collaborative science applications. *Concurr Comput Pract Exper*. 2014;27(2):290-305. doi:10.1002/cpe.3262
11. Tarboton DG, Horsburgh JS, Maidment DR, et al. Development of a community hydrologic information system. Proceedings of 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation; 2009:988-994.
12. Tarboton D, Idaszak R, Horsburgh J, et al. HydroShare: advancing collaboration through hydrologic data and model sharing. Proceedings of the 7th International Congress on Environmental Modelling and Software; 2014; San Diego, California, International Environmental Modelling and Software Society (iEMSs).
13. Mason SJ, Cleveland SB, Llovet P, Izurieta C, Poole GC. A centralized tool for managing, archiving, and serving point-in-time data in ecological research laboratories. *Environ Model Softw*. 2014;51:59-69. doi:10.1016/j.envsoft.2013.09.008
14. Stubbs J, Cardone R, Packard M, et al. Tapis: an API platform for reproducible, distributed computational research; 2020.
15. McLean J, Cleveland S, Rotzoll K, et al. The 'Ike Wai Hawai'i groundwater recharge tool. *Proc Gateways*. 2019.
16. Izuka SK, Engott JA, Rotzoll K, et al. Volcanic aquifers of Hawai'i–Hydrogeology, water budgets, and conceptual models. Scientific investigations report 2015-5164, U.S. Geological Survey; 2018
17. Engott JA, Johnson AG, Bassiouni M, Izuka SK, Rotzoll K. Spatially distributed groundwater recharge for 2010 land cover estimated using a water-budget model for the Island of O'ahu, Hawai'i. Scientific investigations report 2015-5010, U.S. Geological Survey; 2017
18. Westenbroek SM, Engott JA, Kelson VA, Hunt RJ. SWB Version 2.0–A soil-water-balance code for estimating net infiltration and other water-budget components. U.S. Geological Survey Techniques and Methods; Vol. 6, 2018:118.
19. Frazier AG, Giambelluca TW, Diaz HF, Needham HL. Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian Islands. *Int J Climatol*. 2016;36(3):1459-1470. doi:10.1002/joc.4437
20. Elison Timm O, Giambelluca TW, Diaz HF. Statistical downscaling of rainfall changes in Hawai'i based on the CMIP5 global model projections. *J Geophys Res Atmosph*. 2014;120(1):92-112. doi:10.1002/2014JD022059
21. McLean JH, Rotzoll K, Cleveland SB, Izuka SK. *The Hawai'i groundwater recharge tool. Accessed May 6, 2020*. https://recharge.ikewai.org
22. Cleveland SB, McLean JH, Rotzoll K, Izuka SK, Jacobs GA. Enabling real-time user interaction for decision support. *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*. ACM; 2019.
23. Cleveland S, Geis JA, Jacobs G. The 'Ike Wai gateway–a science gateway for the water future of Hawaii - science gateways 2018 paper. *SGCI. Gateways*. IEEE; 2018.
24. Cleveland SB, Dooley R, Perry D, Stubbs J, Fonner JM, Jacobs GA. Building science gateway infrastructure in the middle of the pacific and beyond. Proceedings of the Practice and Experience on Advanced Research Computing; 2018; ACM.
25. Marrinan T, Aurisano J, Nishimoto A, et al. SAGE2: a new approach for data intensive collaboration using scalable resolution shared displays. Proceedings of the 10th IEEE International Conference on Collaborative Computing: Networking, Applications, and Worksharing; 2014; IEEE.
26. Kirshenbaum N, Hutchison J, Theriot R, Kobayashi D, Leigh J. Data in context: engaging audiences with 3D physical geo-visualization; 2020; ACM, CHI EA.
27. Harrower M, Brewer CA. ColorBrewer.orgAn online tool for selecting colour schemes for maps. *Cartograph J*. 2003;40(1):27-37. doi:10.1559/152304003100010929
28. Web content accessibility guidelines (WCAG) 2.0. guideline, W3C; 2008.
29. Borland D, Ii RMT. Rainbow color map (Still) considered harmful. *IEEE Comput Graph Appl*. 2007;27(2):14-17. doi:10.1109/MCG.2007.323435
30. Colorimetry – Part 4: CIE 1976 L*a*b* colour space. Standard 11664-4, ISO; 2019.
31. ESRI shapefile technical description. Standard, ESRI; 1998.
32. Riechert M, Blower J. The CoverageJSON format specification. Standard 0.2-draft; 2016.

How to cite this article: McLean J, Cleveland SB, Rotzoll K, et al. The Hawaiʻi Groundwater Recharge Tool. *Concurrency Computat Pract Exper*. 2022;e6843. doi: 10.1002/cpe.6843