Automating Climate Scenario Creation for Wildfire Modeling

A Capstone Project submitted in partial satisfaction of the requirements for the degree of
Master of Environmental Data Science
for the
Bren School of Environmental Science & Management

by

VICTORIA CUTLER
ERICA DALE
MALLORY GIESIE
LEWIS WHITE

Committee in charge:
NAOMI TAGUE
RUTH OLIVER

Technical Documentation Signature Page

Automating Climate Scenario Creation for Wildfire Modeling

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principle of the School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions.

The Capstone Project is required of all students in the Master of Environmental Data Science (MEDS) Program. The project is a six-month-long activity in which small groups of students contribute to data science practices, products or analyses that address a challenge or need related to a specific environmental issue. This MEDS Capstone Project Technical Documentation is authored by MEDS students and has been reviewed and approved by:

____

Victoria Cutler

____

Erica Dale

____

Mallory Giesie

____

Lewis White

____

Date

This Capstone Technical Documentation is authored by MEDS students and has been reviewed and approved by:

____

Naomi Tague

____

Ruth Oliver

____

Date

Table of Contents

1.0 Abstract

California's increasing fire risk due to climate change requires strategies to mitigate fire severity and help communities adapt. While climate models provide valuable insights into climate trends and future projections, they don't capture the complete spectrum of future climate scenarios for a given region. Therefore, global climate models have limitations when used in conjunction with fire models to capture the various possibilities of fire regimes and their associated risks. Thus, it is essential to create a diverse suite of climate scenarios in order to better prepare for the full range of possible climate futures. This project tackles this issue and contributes to a larger initiative, *Building Resilience to Wildfire,* by presenting a process for creating these unique climate scenarios by stitching together pieces of existing climate model projections. The segments, defined by season or year, are filtered to match specified climate characteristics of interest, stitched together to make a continuous time-series, and converted into a format that can be easily imported into an ecohydrological wildfire model. This process enables researchers to create a variety of scenarios that meet their specific research objectives. To further simplify the process of creating climate scenarios, an interactive tool allows scientists to efficiently extract pre-constructed scenarios or construct their own, download an accompanying metadata summary that provides information about the inputs and conditions used to generate the climate scenario, and generate RHESSys-fire inputs for their selected scenarios. Additionally, visualizations designed for community outreach were developed to illustrate the changing climate in Santa Barbara. These visualizations highlight both the trends seen in the data as well as the inherent uncertainty that is associated with climate projections.

2.0 Executive Summary

Climate-related environmental change in California is causing increased fire risk with significant impacts to our natural environment and communities (Goss et al., 2020). As catastrophic wildfires continue to become more prevalent, strategies to reduce the severity of these wildfires and help communities adapt to changing fire regimes are necessary. The *Building Resilience to Wildfires* initiative, which uses the eco-hydrological model RHESSys-fire[1], is currently underway to model fire regimes and target adaptation strategies. This MEDS project makes an important contribution to the larger initiative by generating, analyzing, and visualizing potential climate change scenarios that will be used to estimate future fire risk and the likely effectiveness of different fuel treatment options.

As the *Building Resilience to Wildfires* initiative aims to model fire regimes and assess potential adaptation strategies, it is crucial to consider the impact of climate change on future fire risk. However, creating climate scenarios from climate projections poses several challenges. There are many appropriate climate models with various emissions assumptions that can lead to different predictions for the climate (About Climate Projections and Models, 2018). Additionally, each climate model output is a sample of possible long-term climate trajectories and not a prediction of specific day-to-day weather in the future. Furthermore, the spatial resolution of these climate models is often fairly coarse and must be downscaled to specific regions, resulting in data that might not capture the full climate variability in a region. Since climate models do not capture the full range of possible climate futures and downscaled data may not fully capture regional variability, it is important to create a wider variety of scenarios to ensure that communities and decision-makers can prepare for a range of potential future conditions. This will enable more effective planning and implementation of adaptive measures, reducing the risk of being unprepared for unique and unforeseen scenarios.

Researchers using RHESSys-fire already explore a range of possible climate trajectories as inputs into the model; however, to simulate these scenarios, they often adjust one climate model's predictions and change individual variables (e.g. to see what a scenario would look like if it was 2°C warmer or if there was a couple more inches of rain in the winter). While this approach allows researchers to investigate simple climate warming, it does not take full advantage of climate model outputs and ignores interactions among climate variables and long term trends. Moreover, this process is time consuming and is not easily reproducible for other areas of interest. A method that allows researchers to use climate model output to create and analyze specific scenarios (e.g. most probable, high fire risk) that fit the researcher's particular objectives and quantifies the respective probabilities of these scenarios would be ideal.

Here, we demonstrate a process for creating a range of unique climate scenarios by stitching together pieces of existing climate model projections. The segments can be defined as either years or seasons (based on wet or dry periods). To create each segment, we filter the climate model

[1] https://github.com/RHESSys/RHESSys. RHESSys-fire integrates fire regime modeling into the eco-hydrological model, RHESSys.

projections to identify seasons/years that match specified climate characteristics of interest. If multiple years/seasons meet the criteria, a random sample of the available segments is selected, allowing for a variety of scenarios that fit the same baseline criteria. This sampling and reshuffling process is repeated several times, often with varying climate criteria per segment, to create a number of individual time segments. The resulting segments are then joined together to form a continuous time series with daily values for the climate variables of interest. The resulting time series for each climate variable (min/max temperature, precipitation, wind speed, and min/max humidity) are then output into the format required by RHESSys. Once formatted, the time-series output of these scenarios can be used as inputs for RHESSys-fire to compile climate model estimates with changing hydrology, vegetation growth and future fire regimes.

This method allows for greater flexibility in generating climate scenarios that meet specific research objectives (e.g. a researcher would be able to create a scenario with two wet years followed by three years of drought with high wind) and can provide a more accurate representation of the possible range of future conditions. A literature review showed that previous research has used this stitches approach and deemed it appropriate for generating a variety of climate scenarios (Tebaldi et al., 2022). Expert climate scientists at UCSB[2] were also consulted to ensure that the project followed key principles of climate science.

Along with a workflow that outlines the process for creating scenarios, an interactive tool allows scientists to efficiently extract or construct climate scenarios, download an accompanying scenario `receipt'' for describing climate criteria and segment details, and generate RHESSys inputs for their selected scenarios. As a separate deliverable, graphs intended for community education were created. These visualizations illustrate how the climate in Santa Barbara is expected to change and emphasize the uncertainty of day-to-day weather predictions of the future.

As our clients are primarily focused on the central coast of California, daily climate predictions from Cal-Adapt models[3] were used to create scenarios for this project; however, our approach and the workflow implemented in our R Markdown could be expanded globally.

---

[2] Thank you Dr. Sam Stevenson and Dr. Max Moritz!

[3] Cal-Adapt includes a data catalog with a suite of Global Climate Models (GCMs). For this project, we utilized the LOCA Downscaled CMIP5 Climate Projections. While the projection includes 9 GCMs, we focused on the 4 priority models, as these included daily values for our climate variables of interest. These 4 priority models also included two different emission assumptions, RCP 4.5 (moderate emission scenario) and RCP 8.5 (high emission scenario).

For more information, head to https://cal-adapt.org/data/download/ to see where we downloaded our data.

3.0 Problem Statement

Catastrophic wildfires are becoming increasingly frequent, and their occurrence is dependent on a multitude of variables—an important one being climate (Kennedy et al., 2021). While models that predict future climate trends exist, the future of climate is inherently uncertain (Hallegatte, 2009). Even according to these climate models, climate predictions vary widely and do not provide a comprehensive picture of potential future scenarios. Moreover, a climate model represents only one realization of a possible future and the current range of climate scenarios fails to capture the full spectrum of possibilities.

In order to address this issue, specific scenarios can be created by intentionally sampling and reshuffling seasons and years within each climate model to produce an ensemble of new time-series data. By failing to include a diverse range of climate scenarios in fire regime models, we risk losing essential insights into potential future wildfire regimes. Addressing this gap is necessary for developing effective wildfire management strategies that mitigate the impact of wildfires and support community adaptation to changing fire regimes (Sample et al., 2022).

4.0 Specific Objectives

The objectives of this project are as follows:

1. Synthesize Cal-Adapt models to summarize future trends for the Santa Barbara region
    a. Summarize changing climate variables and identify climate conditions that make up extreme fire-risk scenarios for the Santa Barbara region
    b. Visualize Santa Barbara future climate for community education

2. Create a method to construct customized climate scenarios[4] using climate projection data
    a. Outline and define criteria (e.g. building by season/year long segments, sampling window[5], emission assumption, climate criteria statistics, etc) for creating a climate segment[6]
    b. Stitch together climate segments to create climate scenarios and format the outputs to conform to the RHESSys input requirements

3. Build an interactive web application that allows users to easily create unique climate scenarios from Cal-Adapt data

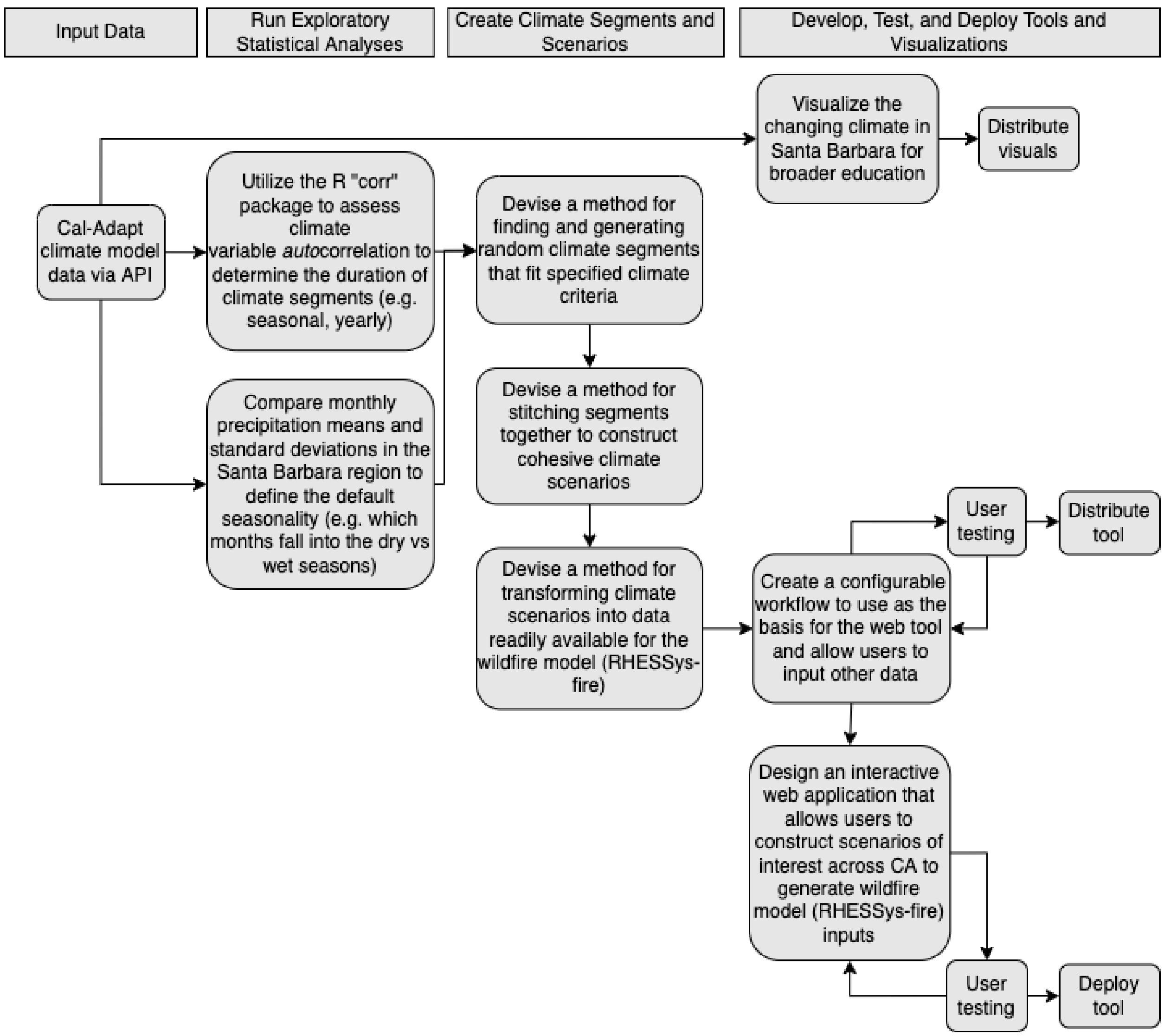
[4] In a literal sense, we define climate scenarios as the methodological aggregation of climate segments, defined below.
[5] We define a sampling window as the "window" of climate segments (as defined below) before and after the year of interest available for sampling. For example, if a user wants to build a hypothetical year 2025, if the user specifies a sampling window of 12 years, the user will sample from years 2025+/- 6 years (i.e. 2019 - 2031).
[6] We define a climate segment as the time series data for all climate variables (e.g. precipitation, maximum temperature), over a certain time period (e.g. season, year).

a. This interactive application takes user-defined climate segments, outputs time-series of climate variables in RHESSys accepted format, and provides the ability to download relevant metadata to serve as a scenario "receipt" for reproducibility

5.0 Summary of Solution Design

The high level solution design for this project is summarized, below, in Figure 1. Further details are provided in Sections A-D, below.

Figure 1. High Level Solution Design



A. Input Data:

Cal-Adapt, a data platform that hosts LOCA Downscaled CMIP5 Climate Projections at a resolution of 1/16° (divided into cells of approximately 6 km by 6 km), was used as input data, and accessed on an ad-hoc basis via the caladaptR API. Specifically, four priority GCMs and two RCP emission assumptions were used, as detailed in Table 1, below.

Table 1. Cal-Adapt GCMs of Interest

Model	Description	Available RCPs
HadGEM2-ES	A warmer/dryer simulation	4.5, 8.5
CNRM-CM5	A cooler/wetter simulation	4.5, 8.5
CanESM2	An average simulation	4.5, 8.5
MIROC5	A simulation that is most unlike the first three for the best coverage of different possibilities	4.5, 8.5

B. Segment Exploratory Statistical Analyses

To explore autocorrelation across time for each climate variable (e.g. evaluating autocorrelation for precipitation across seasons, years, and climate patterns such as El Niño or La Niña), autocorrelation plots were scrutinized using the `acf ( )` function. The plots include lag-wise 95% confidence intervals centered at zero, which can be used to determine statistical significance of the autocorrelation. It was found that, in the Santa Barbara region, individual years were not correlated and therefore could be treated independently, validating that it is reasonable to stitch together individual years without having to consider cyclical patterns. This autocorrelation analysis also confirmed that defining years by the water year (October 1st – September 30th), rather than calendar year, improves scenario accuracy by avoiding jumps in the time series at the point where segments are stitched together.

In addition, to examine seasonality, precipitation data was grouped by month and analyzed via distributional boxplots. This analysis informed the months assigned to the wet season and dry season for Santa Barbara and therefore the default seasonality for future users.

C. Create Climate Segments and Scenarios

Since the underlying physics between each GCM differs, we did not stitch together segments from multiple models. Instead, we kept all models intact and output an ensemble representing the desired scenario, for each climate model, separately. We also sampled climate variables for a given year together, as sampling individual variables and stitching them together would ignore the correlations between these variables. Given long-term trends for some of the climate variables, we recognized that it wasn't practical to sample from a distant year as if it would occur in the near future (e.g. sample a warm year from 2100 and create a scenario where this climate occurs in 2025). Thus, we determined an appropriate sampling window for each segment.

We then stitched climate segments together (while considering leap years) to make cohesive climate scenarios for the desired duration of interest (5 and 20 years). These climate scenarios were then translated into RHESSys-fire format.

D. Develop, Test, and Deploy Tools and Visualizations

Customizable Workflow:

This workflow (1) provided the flow that was then incorporated into the user-friendly interactive web application; and (2) serves as an alternative method outside of our interactive web application that allows users with other data (data besides Cal-Adapt) and/or with niche specifications to create climate scenarios of interest

Currently, the workflow is defaulted to an example use-case. The first step in this use-case is to either (a) input data through the caladaptR API or (b) input a sample local file of other data.

The user then begins to add climate criteria for each segment in the scenario, for any climate variables of interest (see Table 2, below). In the use-case, the first segment uses climate criteria pertaining to precipitation, and requests that the segment that falls into the 0th to 20th percentile, thereby requesting a dry segment (e.g. dry year). For faster input, the user can utilize a custom function to repeat segment climate criteria as many times as desired. The user then inputs the segment type (year or season), the scenario start date, and the sampling window. The default specification of wet/dry season is tailored to the central coast; however, this is customizable within the workflow. The workflow then finds existing segments that match each segment climate criteria, and returns the number of matches to the user. The user can then adjust any climate criteria. Once the climate criteria are finalized, the workflow then randomly selects a matching segment to include in the final scenario.

Each selected segment is then stitched together to form one continuous time series of daily climate variable data. To handle leap years, the extra day (February 29th) was removed and added back in every four years. Daily data for the leap year was constructed by averaging temperature, wind, and relative humidity data from February 28th and March 1, while setting precipitation to 0. The time series data is then available for plotting and exploration within the workflow, and downloaded with accompanying metadata.

Tague Team Lab members ran through this use-case in the customizable workflow and provided valuable feedback. All feedback was incorporated into the final customizable workflow. Once finalized, the lab was notified via email.

Table 2. Climate Variables and their Corresponding Units + Method of Summarization

Climate Variables	Measures & Units	Aggregation Method
Temperature	minimum and maximum °C	Average temperature per time segment
Wind	m/s	Number of days above high wind threshold per time segment
Relative Humidity	minimum and maximum as a decimal	Average relative per time segment
Precipitation	m	Total Precipitation per time segment

Interactive Web Application:

The interactive web application was created using Shiny Dashboard and serves a user-friendly version of the customizable workflow, described above.[7] By running the application, users can easily construct their own climate scenarios and download RHESSys-fire inputs for their specified region of interest. The files configured for RHESSys-fire are accompanied by a metadata file with the details and visualizations of the climate scenario downloaded. In essence, this metadata file acts as a scenario "receipt" that can help users recall the steps used to create the scenario. This 'receipt' will include the following information from the following user inputs: GCM, RCP emission assumption, spatial extent, start date, scenario duration, sampling window, segment type (year/season), climate criteria for each segment in the scenario, the number of matches for each segment, and any notes that the user added while specifying their climate criteria.

To ensure that the interactive web application was user-friendly and intuitive, Tague Team Lab members tested the app by building actual scenarios of interest. All feedback was then incorporated either into the app, itself, or into the accompanying demonstrational video. Once finalized, the application was distributed for deployment.

Educational Visualizations:

Per request from our clients at the Moore Foundation, we created a series of visualizations that depict the changing climate trends in the Central Coast. The goal of these visualizations is to emphasize the unpredictability and extremes of day-to-day weather in the future, while highlighting that trends still exist.

Once finalized, the visualizations were distributed to all stakeholders.

6.0 Products & Deliverables

6.1 R Markdown notebook documenting workflow

[7] The customizable workflow does allow for more user-input data functionality than its web application counterpart. The interactive web application only currently allows for Cal-Adapt input data.

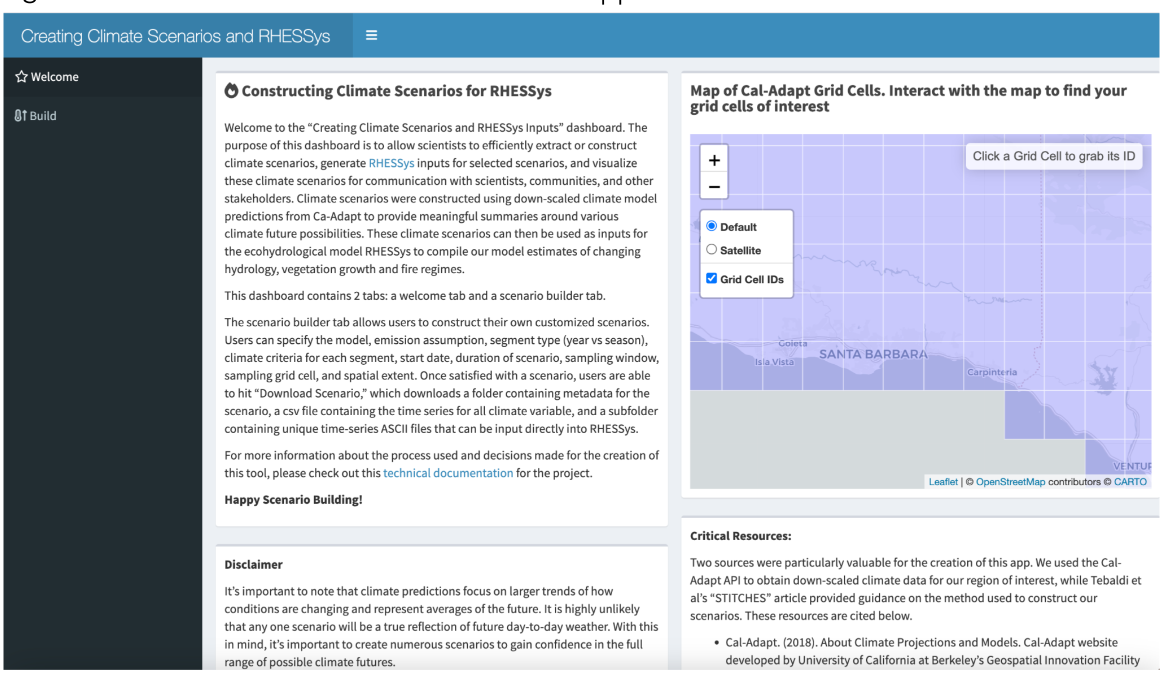
a. Provides thorough documentation of functions and detailed commenting to enable users to understand the process and replicate it using Cal-Adapt data or their own
b. Offers more flexibility for scenarios than the interactive web application
    i. Allows users to specify their own definition of season
    ii. Can be used with any climate data that includes daily values for min/max temperature, precipitation, wind, and min/max relative humidity
    iii. Can define their own filtering criteria for climate variables (e.g. change the threshold for windy days or use maximum wind)

6.2 Interactive Shiny dashboard for obtaining and creating climate scenarios

A demonstrational video was also created that walks users through the scenario building process on the app. The video can be accessed here. We've also outlined the structure of the dashboard below.

Tab 1: Background
a. Outlines the objectives of the tool and provides rationale for its development
b. Provides a written description of how to use the tool
c. Includes a hyperlink to this technical documentation is case users wish to learn more about the project

Figure 2: Welcome Tab of the Interactive Web Application



Tab 2: Custom Scenario Builder:
a. Allows users to specify their own criteria (below) for each season/year long segment in their customized scenario. Definitions of these terms are included in a glossary in the Appendix section of this document.
    i. GCM model

ii. RCP emission assumption
iii. Sampling grid cell
iv. Full spatial extent
v. Segment type (year vs season)
vi. Start date
vii. Duration of scenario
viii. Sampling Window
ix. Climate criteria

b. Visualizes the distributions for each climate variable within the sampling window so users have a sense for how many segments will match their criteria. Users will also be able to specify their preferred climate criteria for each segment in their scenario by adjusting an interactive table. The application will state how many segments within the sampling window fit the designated climate criteria.
c. Creates a time-series visualization so users can see what their final scenario looks like across each climate variable
d. Allows users to easily download a folder for their scenario that includes:
   i. A CSV that includes the time series for each climate variable
   ii. A metadata file with information about the criteria used to create the downloaded scenario, visualizations of the downloaded scenario, and any notes that users wrote while creating the scenario
   iii. Individual time series files for each climate variable that can be used as input for RHESSys

Figure 3: Build Tab of the Interactive Web Application

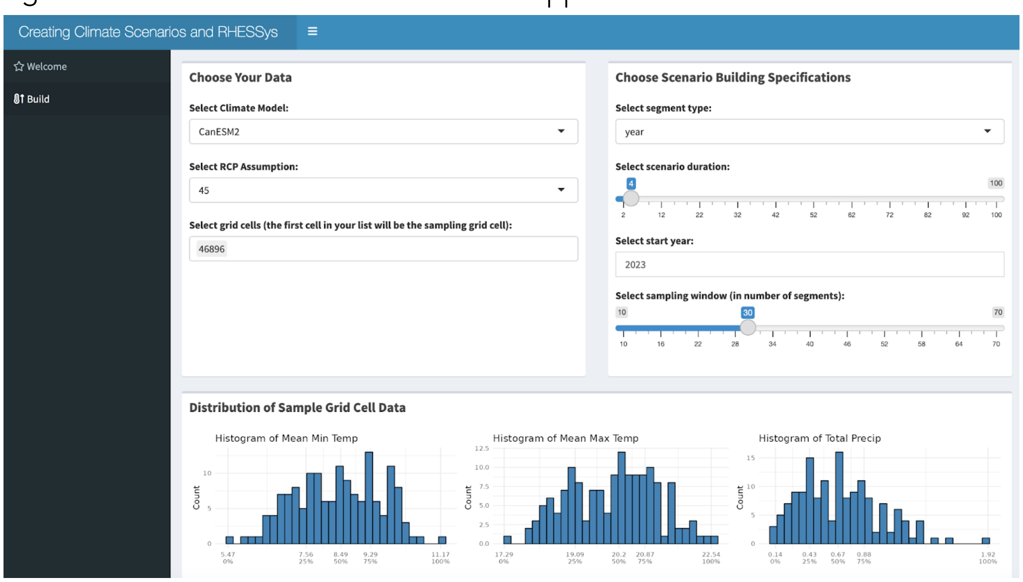


Figure 4: Build Tab of the Interactive Web Application (cont.)

Select climate criteria for each segment by clicking into this table. To apply changes across rows or columns, drag values.

	notes	min_temp_l	min_temp_u	max_temp_l	max_temp_u	precip_l	precip_u	wind_l	wind_u	min_humidity_l	min_humidity_
1	Likely segment	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.0
2	Likely segment	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.0
3	Likely segment	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.0
4	Likely segment	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.0

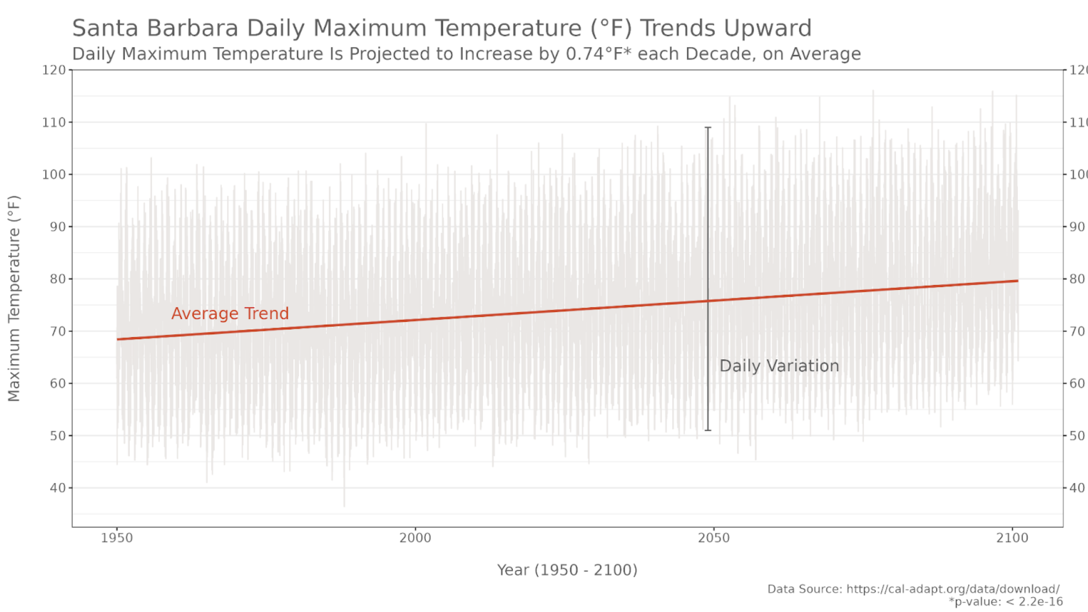
Number of available matches in each segment

Count
30
30
30
30

6.2 Visualizations Depicting the Changing Climate in the Santa Barbara Region

a. Includes a graph for each climate variable of interest that depict how climate change is changing in the Santa Barbara region
b. Intended to be used by the Moore Foundation and Tague Team Lab for community outreach

Figure 5: Example Visualization



7.0 Summary of Testing

Throughout this project, user interviews were conducted to provide guidance and obtain feedback on plans for the deliverables. Members of the Tague Team Lab—Dr. Tague, Janet Choate, Ning Ren, Rachel Torres, Louis Graup, and Christopher Heckman—were interviewed to build user personas and inform what features of the deliverable would be most important to include. This feedback was synthesized and instrumental in the creation of an initial prototype for the interactive web application. Rachel and Louis also completed user experience interviews and shared important feedback on the interactive web application mock-up.

7.1 Workflow

To test the functionality of the workflow, we utilized a monkey test approach by trying out all combinations of users inputs and confirming whether each step in our workflow was operating as expected. We examined outputs and graphed resulting time series to ensure that the stitching process worked as expected.

Along with providing high level guidance throughout the project, Dr. Tague and Janet Choate thoroughly reviewed our workflow code to ensure that it met quality standards to be included in the RHESSys GitHub wiki. Through conversations with Dr. Tague, we aligned on definitions for critical terms in our project (e.g. 'segment', 'sampling window', etc) and improved the overall clarity of our workflow document.

7.2 Interactive Web Application

In terms of functionality, a monkey test approach was utilized to ensure that no combination of user inputs could break the application. Before a user specifies their spatial extent, retrieving the data from an API, all charts that rely on this data provide a message reminding the user to select their region of interest.

User testing was also conducted to improve the intractability, layout design, and general experience of the tool. Before the tool was created, user interviews with PhD students Rachel Torred and Louis Graup were conducted to obtain feedback on the expected layout and capabilities of the application. Feedback from these interviews was incorporated into the final structure of the application. Once the application was developed, Dr. Tague and Janet Choate thoroughly tested the functionality of the application and provided suggestions for improvement, and this feedback was incorporated into our final application design.

7.3 Climate Change Visualizations for the Community

Interviews with our clients from the Moore Foundation, Marion Wittmann and Sarah Anderson, were conducted to learn about what visualizations may be helpful for community outreach. We met with Kevin Varga and Anne-Marie Parkinson from the Santa Barbara FireSafe council to learn more about how our project can help with their community outreach plans as well.

Through conversations with all of these clients, we ultimately focused our visualizations on how the climate is changing across our climate variables of interest for the Santa Barbara region. It was particularly important to showcase trends in the data while also emphasizing the vast daily variation seen throughout the projected climate future. Our clients also reviewed these visualizations and provided feedback to ensure that the graphs met their needs.

8.0 User Documentation

The primary goal of this section is to help facilitate a smooth handover of this project to Dr. Naomi Tague and the Tague Team Lab.

This entails summarizing the overarching structure of our repository, detailing our workflow for creating climate scenarios, explaining how to use our interactive web application, providing a comprehensive rationale for important decisions made during the project, and outlining the proper procedures for using and maintaining the repositories for this project.

The README file in our GitHub repository also contains a link to this technical documentation so individuals outside of the Tague Team Lab can benefit from in-depth explanations and guidelines.

8.1 Repository Information

How to access the repository

Our code and documentation is version controlled and saved on GitHub at the following repository: https://github.com/fire-futures/Automating-Climate-Scenarios.

Our repository contains 3 folders corresponding to each of our main deliverables:

- climate-scenarios
    - An R Markdown file that provides step by step instructions for creating climate scenarios and exporting them into RHESSys formatted time series.
    - A folder containing all necessary functions to run through the workflow
    - Example data used to show how user-input data should be formatted
- shiny-app
    - A README that summarizes the purpose of the automated app and links out to a tutorial of how to use it.
    - Files and subdirectories associated with the interactive application. The folder scenario-building-app contains all of the relevant content required to run the application.
- educational-visualizations
    - An R Markdown file that includes the code utilized to create the visualizations
    - 6 images that depict how each climate variable of interest in changing over time in the Santa Barbara area

8.2 How to use workflow

1. Clone the repository and open the climate-scenarios directory

To use the workflow, fork and clone the entire repository before opening the repository as an R Project in RStudio. This workflow will not run outside of the project, as it relies on relative file paths and sourced functions specific to the project's structure.

2. Access data via API or download, read in, and format your own data

The workflow provides an example using data—accessed via API—from Cal-Adapts 4 priority models in the Santa Barbara region. It is possible to load in other data and adjust the formatting to match that of Cal-Adapt. To guide users in achieving the desired final data format, a template is included as part of the workflow. This template serves as a reference and helps users ensure that their data is properly formatted to meet the required specifications.

3. Follow instructions to construct climate scenario

Detailed instructions are provided that walk users through each step in the scenario building process. Users will need to make a number of specifications when going through the workflow, including:

- The start date of the scenario
- The duration of the scenario
- The sampling window to be used
- Whether to build the scenario by year or season length segments
- The climate specification criteria for each year/season segment in their scenario
- If using Cal-Adapt data, the grid cell that should be used to sample and build the time series as well as the entire spatial extent of interest

8.3 How to use the Interactive Web Application (Shiny Dashboard)

The Shiny Dashboard can be launched by opening any of the server.R, global.R, or ui.R files in the scenario-building-app directory in RStudio and clicking "Run App" in the top right of the window.

8.3.1 The Automated-App

This dashboard contains 2 tabs: a "welcome" tab and a "build" tab.

The welcome tab describes the motivation/background for the project and includes a description of how to use the tool. It also provides a link that will open a screen recorded demo of how to use the application to download pre-generated scenarios and/or build a scenario from scratch.

The scenario builder tab allows users to construct their own customized scenarios. Users can specify the model, emission assumption, segment type (year vs season), start date, duration of scenario, sampling window, sampling grid cell, and spatial extent. For the climate criteria, there are interactive distribution charts that help users determine the appropriate climate specification criteria. To specify the climate criteria, users can manually build a table of values in the app or upload a csv file with appropriate column names and cell values. Once satisfied with a scenario, users are able to hit "Download Scenario," which downloads a folder containing metadata for the scenario, a csv file containing the time series for all climate variable, and a subfolder containing unique time-series ASCII files that can be input directly into RHESSys.

It is important to note that this app relies on the Cal-Adapt API for data, and thus is not compatible with outside data.

8.4 Support for key decisions in projects

1. Not allowing users to create a scenario that includes data from multiple climate models

Using time segments from multiple different global climate models (GCMs) and stitching them together is not recommended because different global climate models (GCMs) make different assumptions about the underlying physical processes that drive climate, and these assumptions may not be compatible with each other. As a result, combining outputs from multiple GCMs can lead to inconsistencies and inaccuracies in the resulting climate scenarios. By restricting scenarios to data from a single GCM, the resulting projections are more likely to be internally consistent and better reflect the characteristics of the underlying model, leading to more reliable decision-making.

2. Concluding independence between seasons and years based on autocorrelation

The central coast of California is subject to El Niño and La Niña cycles. To check whether it was necessary to account for these cycles in our workflow, we constructed ACF autocorrelation plots. While there was some autocorrelation at the 4 year lag, the effect was not statistically significant at the seasonal or yearly segment length, allowing us to ignore El Niño and La Niña patterns when building our scenario building workflow.

3. Defining wet / dry season in the interactive application

In the interactive "automated-app" application, users are not able to customize the definition of wet and dry seasons. As a default, we defined wet season as November to April and dry season as May to October, based on an analysis of monthly precipitation data. We created boxplots to examine the distribution of total monthly precipitation across all combinations of climate models and emission assumptions. We found that the total precipitation was near zero for May to October, making it suitable to stitch together climate segments on these breaks, as they are unlikely to cause spurious jumps in precipitation. We also examined temperature plots, which showed a gradual, continuous fluctuation. By using these methods, we were able to define the wet and dry seasons in a way that is consistent across all climate models and emission assumptions used in the workflow.

4. Using the number of days above a wind speed threshold rather than average wind speed.

Wind is an important factor in determining how quickly a fire is able to spread. The National Weather Service (US Department of Commerce, n.d.) defines "red flag warning" weather criteria for wildfire as:

- Relative humidity is 15% or less with either sustained winds of 25 mph or greater or frequent gusts of 35 mph or greater for a duration of 6 hours or more.
- Relative humidity is 10% or less with either sustained winds of 15 mph or greater or frequent gusts of 25 mph or greater for a duration of 6 hours or more.

When creating scenarios segmented by season or year, knowing the average wind speed is less useful than knowing the number of days that exceed these thresholds. Our data contained daily average wind values, which tend to be lower due to wind peaking midday. To account for this, we chose a high wind threshold value of 15 mph (6.5 m/s) average wind speed per day (disregarding relative humidity) and counted the number of days in a segment that exceeded this threshold. This method allows us to estimate the number of high wind days in a given segment.

5. Strategy for leap years

When creating scenarios, it's important to note that stitching together years without considering leap years can lead to issues preserving the seasonality of the climate, particularly for scenarios of substantial duration. For example, a 100 year scenario built solely using segments without leap years will be off by 25 days in the 100th year of the scenario. To account for this, first every instance of February 29th is deleted, leaving each year with 365 days. Then, an extra day is added as a row into the data frame every 1460 days (4 years) in the time series to create artificial leap years. To fill in the new row for precipitation, 0 is input as rain is uncommon in the central coast region. For all other climate variables, the day before and day after are averaged to obtain the value for this proxy leap day. This approach allowed preservation of the seasonality of the climate data while accounting for the impact of leap years.

9.0 Data Documentation

The goal of the user documentation is to provide background, instructions, manuals, and videos to support potential users while using these products. In addition to user documentation, comprehensive metadata and data management plans are in place.

9.1 Metadata

Metadata is provided alongside any input or output data within the workflow document and interactive web application. The goal of this metadata is to provide users with context and support reproducibility.

1. Interactive Web Application

The end goal of the Interactive Web Application is to download the daily time series data associated with the customizable climate scenarios. To ensure the user has a record of the criteria used to create the climate scenario, a metadata file will be automatically downloaded alongside this data. An example of this metadata file is found below, in Tables 3.1 and 3.2.

To ensure users have a clear understanding of the definitions behind each column header within the metadata - definitions are provided in the shiny application. Definitions are also found in the Appendix.

Table 3.1 Downloadable metadata.csv File (columns 1–13)

notes	min_temp_l	min_temp_u	max_temp_l	max_temp_u	precip_l	precip_u	wind_l	wind_u	min_humidity_l	min_humidity_u	max_humidity_l	max_humidity_u
Wet segment	0	1	0.3	1	0.5	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Warm segment	0.5	1	0.5	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1
Likely segment	0	1	0	1	0	1	0	1	0	1	0	1

Table 3.2 Downloadable metadata.csv File (columns 14–22)

gcm	rcp	sample_cell	spatial_extent	start_date	segment_type	duration	sample_window	number_of_matches
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	11
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	8
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30
CanESM2	45	46896	46896,46897	2023-10-01	year	10	30	30

2. Customizable Workflow

Within the Customizable Workflow, users are interacting with the Cal Adapt API, or, alternatively, inputting desired data of their own. Extensive documentation in the form of comments are included within the workflow to ensure that the user understands the full context behind the input data, the user inputs, and the output data.

3. Educational Visualizations

To create the educational visualizations, data was obtained through the Cal-Adapt API. The graphs were created using the CanESM2 global climate model under the high emission (RCP 8.5) scenario. The R Markdown used to create the visualizations clearly indicates what data was used, and provided flexibility for users to create similar visualizations under different model assumptions for regions across the Cal-Adapt extent. A graph displaying the range of daily variation along with the trend line has been created for each climate variable. All trends are statistically significant ($p < 0.05$). The goal of these visualizations is to communicate that there is a trend expected for the future climate of Santa Barbara, while a wide range of daily weather patterns and extremes can still be experienced.

9.2 Data Licensing and Storage

1. Licensing

All data utilized in this project is from Cal-Adapt, which is open source with the right to redistribute. There is no disclosure risk or limitations due to licensing or access.

2. Storage

Since all data is accessed via API (or, by user input via the Manual Workflow), there is no need for long term data storage. All applications are stored within the Tague Team Lab Server, where the products will be stored and updated as needed.

10.0 Archive Access

After inspection for quality assurance, the interactive web application will be integrated into the RHESSys GitHub Wiki. Moving forward, the Tague Team Lab will undertake the responsibility of its maintenance, ensuring the application's sustained functionality and performance.

11.0 Limitations and Future Work

11.1 Limitations

The scenarios we've created in our project use modeled LOCA Downscaled CMIP5 Climate Projections which come with limitations to note. Because this is modeled data, it is hard to ensure complete understanding of the climate system and therefore doesn't guarantee to include the same nuances as ground truth data. In addition, the data for the California region was downscaled and uses historical data to establish links between the climate in a smaller region and those same variables in a larger global climate model. This can become problematic if the historic data doesn't capture the ways climate change is affecting the region. One final limitation regarding the Cal-Adapt data is its spatial extent. It's coarse and limits the conclusions that can be made regarding areas smaller than a Cal-Adapt grid cell.

The data provided serves as a means of generating a range of future climate scenarios, and it is important to note that these scenarios are not definitive predictions of the future weather. While we have labeled some scenarios as "probable," it is highly unlikely that any of them will perfectly match the actual climate conditions that will occur. Rather, they represent a range of possibilities for broad changes in climate. In the context of using these scenarios with the eco-hydrological model RHESSys-fire, the primary goal is to explore the potential fire regimes under various scenarios, rather than making specific predictions about the exact timing or location of future fires.

11.2 Future Work

Detrend Climate Variables: In our workflow we use a sampling window in order to not erroneously sample from years so far in the future or past that we are ignoring how climate change is affecting our data. However, it may be interesting, and helpful to detrend our data. Detrending our data would help us better understand the natural variability in our data, while minimizing the influence of underlying trends, like warming temperature or increased precipitation. This would allow us to sample from the entire historic and future data that Cal-Adapt provides (1950-2100) and could allow us to more easily find extremes matching our criteria.

Three-Day Moving Wind Window: In our workflow, we create climate scenarios by summarizing and filtering days within our segments where daily wind values exceed a threshold. After reviewing literature on fire regimes and individual wildfires, we believe it would be insightful to identify the number of days within a segment when the wind remains high for 3+ consecutive days. This is because sustained high winds can increase both the likelihood and intensity of wildfires.

Finer Resolution Data: As previously mentioned, the Cal-Adapt data has a coarse resolution, which makes it challenging to draw conclusions about smaller regions or microclimates. Our workflow is designed to accept data other than Cal-Adapt, provided it is in tabular format. This allows for the integration of finer resolution data if formatted correctly. Future adaptations of the process could explore compatibility with other file formats, specifically NetCDF, as it is supported by RHESSys.

Adapting Seasons: Our current workflow assumes the user is working within a tropical climate, characterized by two seasons per year. The seasons selected for the Santa Barbara region can be easily adjusted for other areas with different wet and dry seasons. Future work could focus on adapting the process to account for temperate climates featuring four distinct seasons. Prior to making such adaptations, research should be conducted to investigate autocorrelation, as was done in our project.

12.0 Acknowledgments

We'd like to firstly acknowledge and thank Dr. Naomi Tague for guiding us through this process. Her advice and wisdom were essential to the success of the project. We also thank Dr. Ruth Oliver for her support and encouragement throughout the Capstone process. Additionally, we'd like to extend thanks to everyone at the Tague Team lab and Moore foundation who supported this project—their input and expertise greatly contributed to the success of this project.

13.0 References + Bibliography

13.1 References

Cal-Adapt. (n.d.). Retrieved May 5, 2023, from https://cal-adapt.org/help/glossary/

Cal-Adapt. (2018). About Climate Projections and Models. Cal-Adapt website developed by University of California at Berkeley's Geospatial Innovation Facility under contract with the California Energy Commission. Retrieved 10 February 2023, from https://cal-adapt.org/help/get-started/about-climate-projections-and-models/.

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., & Diffenbaugh, N. S. (2020). Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environmental Research Letters*, *15*(9), 094016. https://doi.org/10.1088/1748-9326/ab83a7

Hallegatte, S. (2009). Strategies to adapt to an uncertain climate change. *Global Environmental Change*, *19*(2), 240–247. https://doi.org/10.1016/j.gloenvcha.2008.12.003

Sample, M., Thode, A. E., Peterson, C., Gallagher, M. R., Flatley, W., Friggens, M., Evans, A., Loehman, R., Hedwall, S., Brandt, L., Janowiak, M., & Swanston, C. (2022). Adaptation Strategies and Approaches for Managing Fire in a Changing Climate. *Climate*, *10*(4), Article 4. https://doi.org/10.3390/cli10040058

Tebaldi, C., Snyder, A., & Dorheim, K. (2022). STITCHES: Creating new scenarios of climate model output by stitching together pieces of existing simulations. *Earth System Dynamics*, *13*(4), 1557–1609. https://doi.org/10.5194/esd-13-1557-2022

US Department of Commerce, N. (n.d.). *Fire Weather Criteria*. NOAA's National Weather Service. Retrieved May 5, 2023, from https://www.weather.gov/gjt/firewxcriteria

13.2 Bibliography

Critical Fire Weather | NWCG. (n.d.). Retrieved May 5, 2023, from https://www.nwcg.gov/publications/pms437/weather/critical-fire-weather

Duine, G.-J., Carvalho, L. M. V., & Jones, C. (2022). Mesoscale patterns associated with two distinct heatwave events in coastal Santa Barbara, California, and their impact on local fire risk conditions. *Weather and Climate Extremes*, *37*, 100482. https://doi.org/10.1016/j.wace.2022.100482

Wildfire climate connection. (n.d.). Retrieved May 5, 2023, from https://www.noaa.gov/noaa-wildfire/wildfire-climate-connection

Zeise, L., Garcia, Y., & Newsome, G. (n.d.). *Indicators of climate change in California 2022*. Retrieved May 5, 2023, from https://www.preventionweb.net/publication/indicators-climate-change-california-2022

Zigner, K., Carvalho, L. M. V., Jones, C., Benoit, J., Duine, G.-J., Roberts, D., Fujioka, F., Moritz, M., Elmquist, N., & Hazard, R. (2022). Wildfire Risk in the Complex Terrain of the Santa Barbara Wildland–Urban Interface during Extreme Winds. *Fire*, *5*(5), 138. https://doi.org/10.3390/fire5050138

Appendix: Glossary of Terms

Climate Scenario (Scenario)

In a literal sense, climate scenarios are the aggregation of climate scenario segments, as defined below, for the specified duration.

Climate Scenario Segment (Segment)

A climate scenario *segment* is the daily time series data for all climate variables (e.g. precipitation, maximum temperature), over a certain time period (e.g. season, year).

Segment Type

Segment type represents the type of time frame surrounding a segment. This project is currently configured for two segment types: year and season.

Climate Variables

The climate variables are defined as the variables available for download. The current scope includes minimum temperature, maximum temperature, minimum relative humidity, maximum relative humidity, wind speed, and precipitation.

Global Climate Model (GCM)

GCMs are mathematical models that depict the climate on Earth.[8] While there are many GCMs, four priority models are deemed as most suitable for California climate: HadGEM2-ES, CNRM-CM5, CanESM2, MIROC5. This project scope includes these four models.

Representative Concentration Pathway (RCP)

In order for GCMs to predict climate, assumptions are made around greenhouse gas emissions and thus future climatological warming. RCPs represent these plausible future warming scenarios. This project includes the two emissions assumptions used by California's Fourth Climate Change Assessment, RCP 4.5 and RCP 8.5. RCP 4.5 represents the medium emissions scenario in which average annual temperature in California is estimated to rise from 2 - 4 degrees Celsius by the end of 2100. RCP 8.5 represents the high emissions scenario in which the average annual temperature in California is estimated to rise from 4 - 7 degrees Celsius by the end of 2100.[9]

Grid Cell (Cell)

This project uses data from LOCA Downscaled CMIP5 Climate Projections found on Cal Adapt. The finest resolution for these data is available in 6 km by 6 km cells, which this project refers to as grid cells. Users specify the grid cell(s) of interest and download climate scenarios for all grid cells of interest, as in reference to the sample cell, as defined below.

Sample Cell

A sample cell is defined as the grid cell that the user will use to query data according to the specified climate criteria, as defined below. Since grid cells are linked in time and space, only one grid cell can be used to form a scenario, and all other grid cells of interest are stitched together by the same GCM, RCP, and segment as the sample cell. For example, after entering climate criteria, if the first segment of the sample cell (say, year 2030) is selected from the GCM HadGEM2-ES and

[8] https://cal-adapt.org/help/glossary/
[9] ibid.

the RCP 4.5 from the year 2050, all other grid cell scenarios will build from this same GCM, RCP, and segment.

Scenario Start Date (Start Date)
The start date of the entire scenario.

Scenario Duration (Duration)
The length of the output time series, in years.

Segment Climate Variable Criteria (Climate Criteria)
Climate criteria represent the lower and upper percentiles that a user may include to find a segment of interest. For example, if a user is interested in finding a segment of high precipitation, the user may want to filter the available segments for the 95th to 100th percentile in precipitation. That way, this segment will be randomly selected according to these criteria. Users may include many filters on climate variables, simultaneously.

Season Definitions
The season segment type includes two seasons: the wet season and the dry season. The months that make up each season are configurable for the user, since seasonality may change over time.

Wet Season
This project defaults the wet season as November through April.

Dry Season
This project defaults the dry season as May through October.

Available Segment Re-Samples (Available Re-Samples)
Users can build scenarios by inputting specific climate criteria filters for each segment. The user will also be shown the available re-samples, which represent the number of segments that fit the specified climate criteria, and therefore the number of segments that fit this criteria. It is possible that this number can equal 0, which means that there are no segments that fit the specified climate criteria, and the user would adjust the climate criteria filters accordingly. When there is at least one or more available segment re-samples, these segments are randomly sampled from to select the segment that is used as part of the scenario.

Requested Scenario Re-Samples (Requested Re-Samples)
Users may request re-samples of entire scenarios when more than one realization of the built scenario is desired. Users should consider that available segment re-samples may limit the uniqueness of the requested scenario re-samples.

Segment Sampling Window (Sampling Window)
A sampling window represents the “window” of climate segments before and after the segment of interest that are available for sampling. For example, if a user builds a segment for year 2025, and specifies a sampling window of 20 years, this means that years 2025+/- 10 years (i.e. 2015 - 2030)

will be used to sample from for the segment in 2025. Similarly, for a segment in 2026, years 2016 - 2036 will be used for sampling.

Historical Extremes

When building an extreme scenario, users may want to include extreme segments in the past (e.g. a 100 drought year). While historical data may be included in the available segments, contingent on the scenario start date and segment sampling window, including specific segments classified as historically extreme is currently out of scope for this project. Including these types of segments are recommended to include for future work.