Undergraduate Forestry Data Science Climate Dashboard Technical Report and User Guide

Julian Schmitt*[1], Maxwell J.D. VanLandschoot*[2], Kelly McConville[3], Kate Hu[4], Holly Munro[5], and Stephen Prisley[5]

[1]Harvard College
[2]Reed College
[3]Department of Statistics, Harvard University
[4]HSPH Department of Biostatistics, Harvard University
[5]National Council for Air and Stream Improvement, Inc. (NCASI)

October 20, 2022

Contents

1 Introduction

Climate change is expected to alter temperature and precipitation conditions non-uniformly across the globe and, relevantly, in regions of the United States which contain millions of acres of both publicly and privately managed forests. The frequency of extreme events is predicted to increase [1]. Forests are likely to be particularly affected, as some tree species survive in narrow climate ranges and are susceptible to extreme conditions [2, 3].

Forest landscape models (FLMs) project that, in all, these changes have mixed positive and negative effects, with increasing temperature improving tree growth up to a threshold and lessening precipitation generally having a negative effect [2]. Exact tree response, however, is species dependent. Complex relationships exist between environmental factors like temperature, precipitation, local soil composition, and CO_2 concentration, and shift the limiting reagent for tree health and growth and are markedly hard to study at even the local scale [3]. Thus, any attempt to model potential tree growth across the continental US on a per species basis using only aggregated climate projections would be inaccurate and misleading. We instead focus on presenting climate and forestry data in a way that both depicts current climate metrics and density and productivity by tree species that can be leveraged as an empirical proxy for species' habitable ranges.

1.1 Project Goals

Our overarching goal is to present information about future climate and its corresponding uncertainty in ways that are helpful to foresters, forest stakeholders, and the National Council for Air and Stream Improvement, Inc. (NCASI) member companies. Building on the Climate Projection Analysis Tool (CPAT) developed by Dr. Stephen Prisley, we have developed a RShiny dashboard tool to convey climate metrics that provide a balanced picture of how tree species may shift given future climatic conditions and enhanced interactivity of the application. We also considered who our users are and how they will be interacting with our work, leading us to create a dashboard that is statistically rigorous, while still being approachable to an audience of varying statistical acumen.

2 Data and Methodology

2.1 The Coupled Model Intercomparison Project

The climate data used in our analysis were primarily derived from phase 5 of the Coupled Model Intercomparison Project (CMIP5) and the Parameter-elevation Regressions on Independent Slopes Model (PRISM dataset). The CMIP is a collaboration among climate science research groups around the world and is considered the scientific community's current consensus view on climate change. Alongside the Intergovernmental Panel on Climate Change (IPCC), the CMIP's goal is to understand how climate responds to both natural and anthropogenic forcings. Below, we have written an additional summary of climate modeling more broadly, which provides information about large ensemble climate modeling, RCP pathways, and contrasts between findings in the CMIP5 and the more recent CMIP6 reports and models. The PRISM project provides three data products: 30-year normal (averages, currently 1981-2010), daily time steps, and monthly time steps at high resolution (~800m and ~4km) across the US. For this project, we used the 4 km^2 resolution data on the main map feature, averaging the monthly data between January 2000 and December 2009.

2.2 CMIP5 Ensemble Means vs SPEAR Ensemble Members

The CMIP5 multi-model ensemble was released alongside the IPCC's 2013 Annual Report (AR5) [4]. Another model we considered was the SPEAR large ensemble (Seamless System for Prediction and Earth System Research), which was first published in Delworth et al. (2020) and was included as one of the CMIP6 multi-model ensembles [5]. Each ensemble run contained between 30 and 100 ensemble members, and CMIP5 contained roughly 70 models. Figure 1 contains a visualization of the differences between the two datasets. For this project, both datasets had their own advantages and disadvantages. CMIP5 allowed us to capture four different concentration pathway scenarios (RCP 2.6, 4.5, 6.0, 8.5) to SPEAR's RCP 4.5 and 8.5. Furthermore, the CMIP5 multi-model ensemble contained 70 ensemble means generated by climate research centers distributed globally, which allowed for a broader set of modeling assumptions and techniques. SPEAR on the other hand, was a more up to date model, meaning it contained the latest scientific understanding and utilized more advanced computing. Furthermore, because the SPEAR dataset contained 30 ensemble members as opposed to means, using this dataset allowed us to develop additional measures of uncertainty. While the dashboard is ultimately implemented using only the CMIP5 dataset, as the larger number of available concentration pathways was deemed more useful for users, it's still important to discuss model advantages and disadvantages to spur future application refinement.

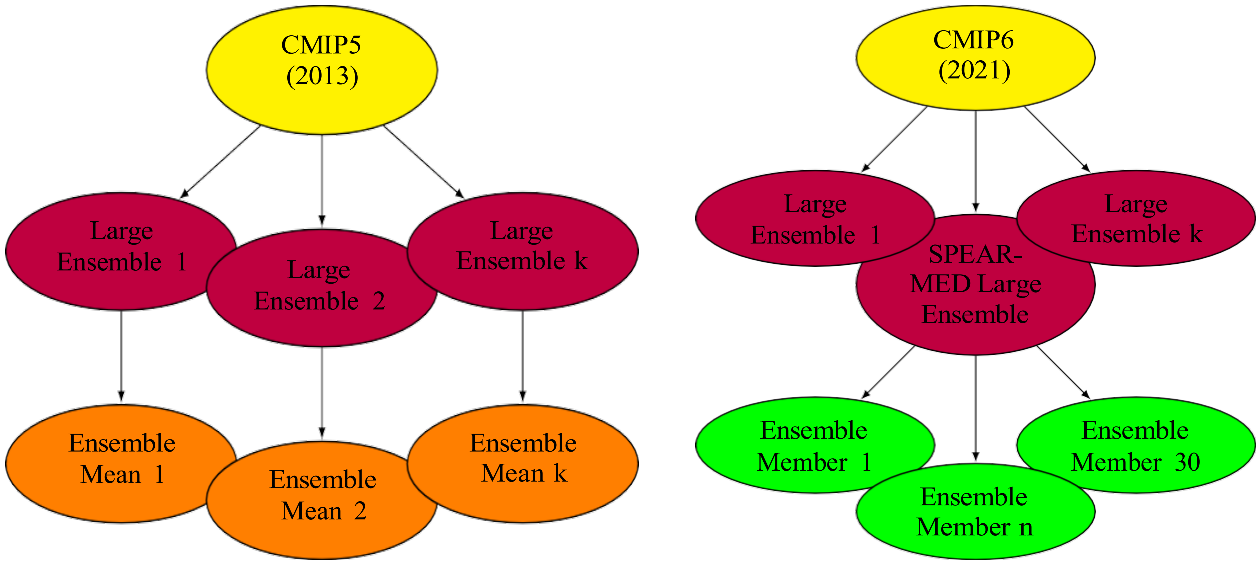

Figure 1: Relationship between multi-model ensembles (yellow), large ensemble climate models (red), ensemble members (green) and ensemble means (orange). We used the CMIP5 ensemble means (orange) and the CMIP6 ensemble members (green). The CMIP5 ensemble means allowed us to use information across ensembles produced by different research facilities, while the CMIP6 SPEAR-MED large ensemble allowed us to accurately represent climate variability and extreme conditions drawn from a single model.

2.2.1 Model Trade-Offs

CMIP6's SPEAR model was developed almost a decade after CMIP5 models, which has some distinct advantages. First, advances in scientific understanding and computing mean that CMIP6 models are up to date with the latest scientific understanding and are run at higher resolution than earlier CMIP5 models. This allows for better prediction at local scales. A further difference between CMIP5 and CMIP6 models is that CMIP6 models, on average, found a higher equilibrium climate sensitivity (ECS), meaning that for the same Representative Concentration Pathway (RCP), CMIP6 models project more warming than their CMIP5 counterparts. The CMIP5 ensemble means still retain some advantages over the CMIP6 data, however, namely that we had information from all the large ensembles used, leveraging a broader range of assumptions and modeling techniques. Furthermore, the CMIP5 ensemble means tracked four distinct emission pathways —RCP 2.6, 4.5, 6.0, 8.5— as opposed to SPEAR's RCP 4.5 and RCP 8.5.

2.3 Feature Implementation

Below, we detailed how each of the Climate Application features were constructed and what aspects of climate change they were expected to influence.

2.3.1 Summary Table

The summary table, shown in Figure 2, is intended to give the most high-level overview of the projected effects of climate change to the region of interest. Let k index across the CMIP5 model ensemble means, j indexes the year (2000-2100), and r indexes the RCP pathway. From the selected RCP and year, the statistics calculated are: 'Yearly Mean Temperature' ($T_{mean}^{(j,r)}$), 'Yearly Minimum Temperature' ($T_{min}^{(j,r)}$), 'Yearly Maximum Temperature' ($T_{max}^{(j,r)}$), and 'Yearly Mean Precipitation' ($P_{mean}^{(j,r)}$). Then the summaries are calculated by first computing an average, maximum, or minimum for each ensemble mean for each of the selected years, and then computing a mean across ensemble mean members. As our original data were on the monthly resolution, the 'Yearly Mean Temperature' is an average across all months and ensemble observations for a given year. 'Yearly Mean Precipitation' is defined analogously, while 'Yearly Minimum Temperature' and 'Yearly Maximum Temperature' are calculated by first taking either the maximum or minimum across months for the selected year and then averaging across the ensemble mean members. This allows us to capture average low January temperatures and high July temperatures. Questions that a potential user might be able to answer with the summary table include:

1. Is the temperature in my region expected to increase or decrease over time?
2. Will there be a more rapid increase in maximum or minimum temperatures?
3. Is the expected range of temperatures going to increase or decrease over time?

Table | Yearly | Monthly | Density | Lookup Table

Summary Table for Red Maple Under RCP 8.5 from 2022 to 2080

	Metric	Unit	2022	2080	Difference
1	Yearly Mean Temp	°C	9.68	11.66	1.98
2	Yearly Average Max Temp	°C	29.62	31.85	2.23
3	Yearly Average Min Temp	°C	-9.9	-7.25	2.65
4	Annual Precipitation	mm/yr	1201.72	1251.48	49.76

Figure 2: Summary table displaying the absolute change for key climate variables under the selected RCP pathways and years as an example.

2.3.2 Annual Conditions Over Time

Filling in the gaps of the summary table, we tracked climate variables across the 21st century by plotting the summary metrics for each year of interest and adding in uncertainty in Figure 3. Before deriving confidence intervals, we first averaged the climate data to the yearly level for all four RCP pathways. Uncertainty for the CMIP5 models was generated by taking an empirical 90% confidence interval a cross all 70 CMIP5 ensemble means, capturing a broad range of methods and frameworks. Focusing on yearly high and low temperatures as opposed to averaging monthly high and low temperatures demonstrates seasonal extremes. We note that, as our CMIP5 dataset only had ensemble means, extracting extreme heat and cold event frequencies is impossible, as the ensemble means themselves already represent the climatology and thus are not representative of true yearly extremes. For example, if 30°C is reported as an upper bound for the yearly high average monthly temperature, it refers only to the ensemble mean values and not the individual ensemble members. This can be broadly interpreted as follows: "We expect that years around 2100 will have average summer daily high temperatures no greater than 30°C. Though, we cannot conclude that an individual year will not have higher temperatures." With these plots, users can begin to ask questions related to broad level climatology such as:

1. How extreme will average summer heat become over time? Does precipitation also increase or will drought risk worsen or develop?
2. Will a tree species that is particularly sensitive to high summer heat or variable precipitation conditions be able to survive at the end of the century?
3. What weather conditions might foresters have to confront in the field?

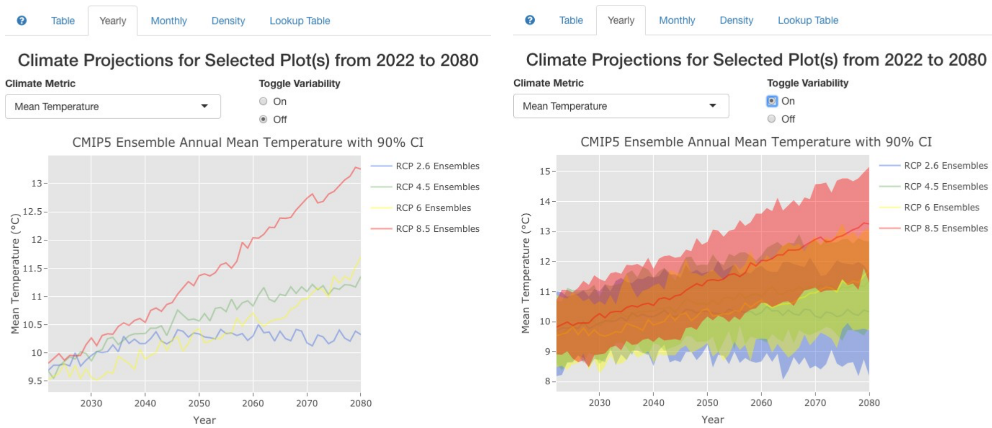

Figure 3: Plotting conditions by a yearly average across the 21st Century. Both the time ranges and climate metric can be toggled. These sample graphics display mean temperature for the selected region from 2022 to 2080.

2.3.3 Conditions by Month

While the yearly RCP plot shows conditions over time, we visualize how monthly conditions are expected to shift by plotting the climatology by month, where the difference in conditions is measured by the difference among the curves (Figure 4). Thus, the vertical distance shows the difference in conditions by month (e.g., increased May temperatures by 2°C). Horizontal distance between the curves represents shifting seasons (i.e., similar minimum temperatures are expected in 2080 in mid-March as mid-April in 2022 under RCP 8.5). By pairing the RCP pathways, we can explore how conditions at the selected year are predicted to change. Currently, there are no measures of uncertainty in this graphic, as the dimensionality of our data required sacrificing some measures in favor of more important features. Questions which can be answered with this graphic might include:

1. How will the growing season shift in the future? For example, here we see that average temperatures above freezing are predicted to be experienced roughly a month earlier towards the end of the century.
2. Will pests be more virulent in the future? Pests, such as the emerald ash borer, don't tolerate extreme cold well (-20F or below), so changing winter lows has implications for how aggressive these invasive species will be in the future.

2.3.4 Modeling Changing Tree Species Environmental Niches

To capture how climate change might affect a particular tree species for the selected region, we compared conditions in the selected location to the nationwide distribution of where the tree species lives. In particular, modeling the nationwide density of said species across the temperature and precipitation space allowed us to track how climate conditions at the location of interest influenced species habitability. This figure can answer questions like:

1. Does the selected region typically support this tree species?
2. Will this selected region continue to be suitable for this tree species in the future under the RCP pathway of interest?
3. Does the selected region have new potential to support this type of tree in the future?

Note that these recommendations will solely be based on the selected climate variables, and do not include other factors that govern species habitability, such as elevation and soil properties.

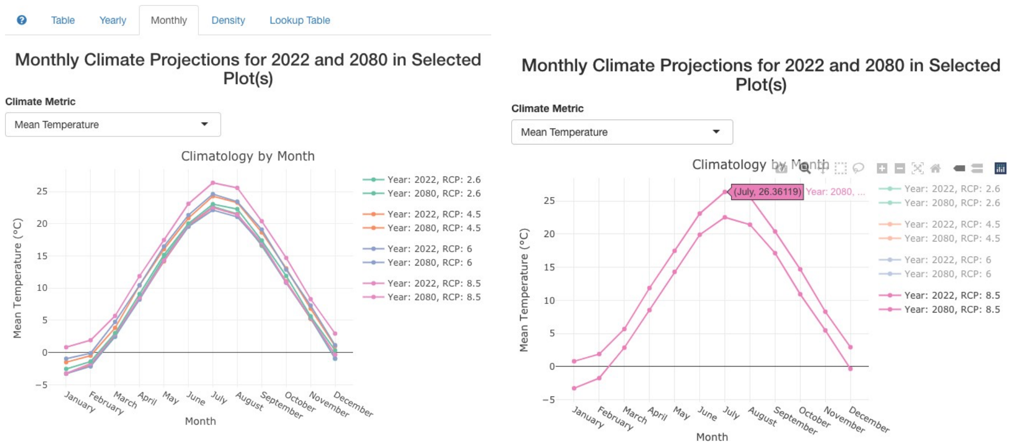

Figure 4: Climatology for a selected region and metric between 2022 and 2080. The image on the right shows all climate scenarios by color, while the graphic at left displays one selected pathway (RCP 8.5). As the plots are interactive, the year can be differentiated using the hover feature of Plotly graphics.

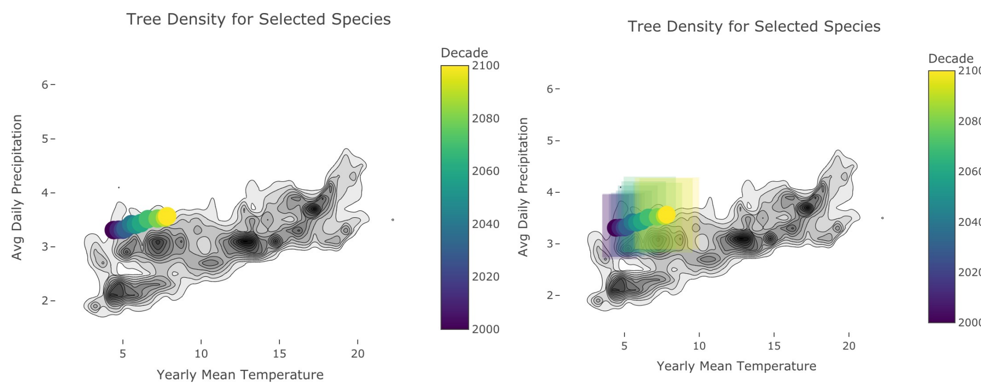

Figure 5: Current nationwide density of the selected species and climate metrics shaded by tree density (in grey). The climate metrics' decadal average for the selected region of interest is scattered on top. Uncertainty, when selected, is generated from 90% confidence intervals (CI) on the ensemble means.

3 User Guide

Below, we provide a brief user guide for our ShinyApp, which provides instructions for selected features. As one of our project goals was updating the current NCASI CPAT, we focused on detailing features that have changed. An overview of the application screen can be found in Figure 6 and is comprised of a draggable background map that can also be zoomed in and out, a control panel that can be used to select different tree species, year ranges, and RCPs, and a dynamic information panel. The entire workflow can be found in our publicly hosted GitHub Repository.

3.1 Adding Graphical Plot Selection

The updated CPAT tool can extract data for a specific location or selected region on the background screen. Instead of entering latitude and longitude coordinates, a user can click on the location of interest to obtain relevant climate information. Furthermore, using the pentagon button, at right of the control panel, a user can now draw a polygon over a region of interest to extract averaged data to be used analyses.

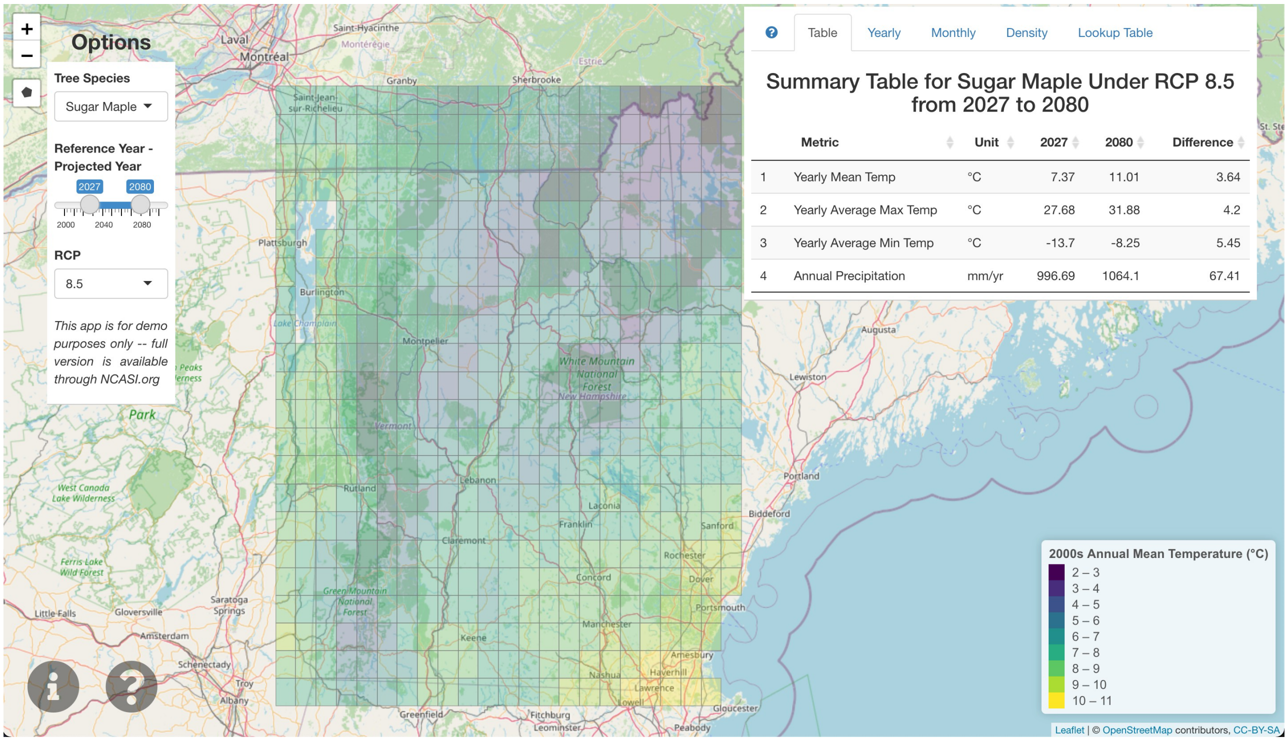

Figure 6: Main Screen of the new CPAT. The main features include an interactive background map, control panel (upper left), and information panel (at right). Parameters such as tree species, RCP, and range of interest can be updated from the control panel, while the information panel contains the climate and tree species summary tables and graphs relevant to the selected plot or region.

3.2 Implementing Uncertainty

Uncertainty in climate projections adds another layer of understanding about how climate is expected to change. Uncertainty in climate models is typically broken down into "reducible" and "irreducible" uncertainty. "Irreducible" uncertainty comes from inherent variability in the climate system - e.g., what ranges of scenarios are actually possible - while "reducible" uncertainty is generated as an artifact from large ensemble climate modeling and can be reduced by model improvements. By using the entire CMIP5 multi-model ensemble, we are theoretically able to reduce the "reducible" error on our climate projections as much as possible given current understanding of the climate system to produce an accurate mean trajectory. In the updated CPAT tool, wherever we added uncertainty into our projections, we added an option to toggle uncertainty "on" and "off". Because the uncertainty is generated from ensemble means, it can be understood as uncertainty in the underlying climatology and should not be used to bound expected temperatures or precipitation values for any given year or month.

3.3 Help Infographic and Information Panel

To help guide users on how to select plots and interact with the control and information panel, we included a visualization GIF which provides a quick snapshot of how to use the application. It is available by clicking on the "?" at the bottom left of the screen (Figure 6). Adjacent to the help infographic is the information "i" button, which pulls up information regarding the app creation and people involved with the project.

3.4 R-Plotly Figures

One significant update to the CPAT tool was re-designing the figures with Plotly, an interactive graphing package. Plotly allows users to zoom in on regions of interest in a figure, hover to view specific measurements from points of interest, and the option to download graphics directly to their personal devices, which can be used for presentations and reports. We have also used Plotly to allow users to select any subset of RCP pathways to highlight for comparison. This can be accessed by clicking on a pathway in the legend to deselect or double-clicking a pathway of interest to deselect all other pathways.

4 Discussion

The goal of this project was to continue to advance the user experience, develop measures of uncertainty, and design new climate features for the NCASI CPAT. By creating a flexible sample product for Vermont and New Hampshire, our team created an exciting new user experience and updated visualization tools. The updated CPAT tool focused on interactivity, placing a background map of the United States at the center of the design, allowing users to select a specific location or region of interest graphically. The figures have been redesigned with the goal of improved data access and transparency in addition to presenting uncertainty in the climatology.

4.1 Scaling to the United States

To capture uncertainty, we had to creatively design ways to collapse the 25 GB raw CMIP5 dataset into a size that could be uploaded and used on the web. By calculating uncertainty windows and saving the modified dataset, we were ultimately able to shrink our dataset by a factor of 75 to about 350 MB. While this demo can easily be deployed to even a free account on shinyapps.io, we estimate that a dataset for the entire US's nearly ten trillion square meter land-area would be close to 60 GB, far exceeding even a paid shinyapps.io account. Perhaps the easiest way to scale to this level and shrink the dataset would be to aggregate the yearly monthly data to a decadal timescale, which would not significantly change the results of the climate dataset and could shrink the data by a factor of three as it is by far the largest dataset required. Uncertainty is an essential component of the climate story and expending additional time and resources to ensure its inclusion will lend significant weight to the results.

4.2 Next Steps

While a sleek user experience and added uncertainty lends confidence to the results, there are several additional features we believe could be invaluable to future CPAT development. As noted in Section 2.1, there are distinct advantages to using CMIP6 data. Furthermore, developing features using ensemble members as opposed to means would allow for analyses that report extremes. For example, we assess changes in drought frequency and severity alongside changes in extreme temperature, which can threaten tree survival over shorter timescales.

References

1. Portner, H.O., D.C. Roberts, H. Adams, C. Adler, P. Aldunce, E. Ali, …, N. Stevens. 2022. Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*.

2. Ryan, M. Temperature and tree growth. 2010. *Tree physiology* **30,** 667–8.

3. Gustafson, E.J., B.R. Miranda, A.M. De Bruijn, B.R. Sturtevant, and M.E. Kubiske. 2017. Do rising temperatures always increase forest productivity? Interacting effects of temperature, precipitation, cloudiness and soil texture on tree species growth and competition. *Environmental Modelling Software* **97,** 171–183.

4. Taylor, K.E., R.J. Stouffer, and G.A. Meehl. 2012. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society.* **93,** 485-498.

5. Delworth, T.L., W.F. Cooke, A. Adcroft, M. Bushuk, J. Chen, K.A. Dunne, P. Ginoux, R. Gudgel, R.W. Hallberg, L. Harris, …, M. Zhao. 2020. SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multi- decadal Prediction and Projection. *Journal of Advances in Modeling Earth Systems.* **12,** e2019MS001895.