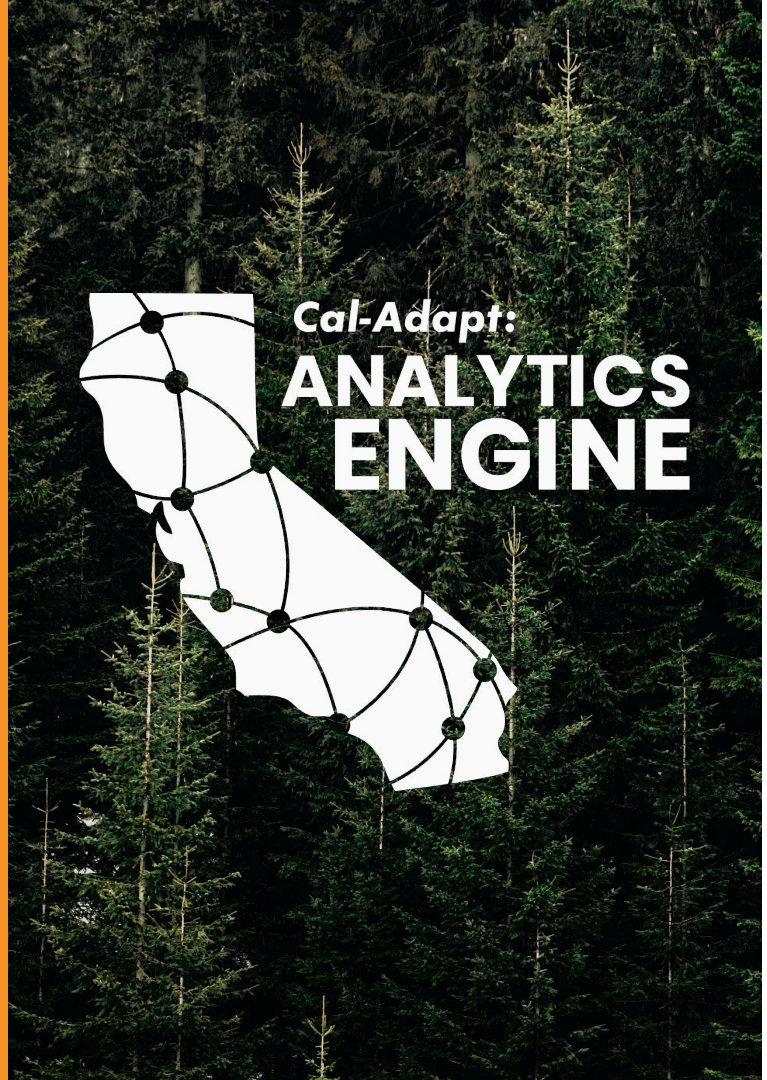
Annual Report 2022

Cal-Adapt: Analytics Engine

Progress update



Open access to data, models, and analytics related to energy resilience supporting California's needs.

Primary Funding Provided by

California Energy Commission – EPIC Program (EPC-20-007)

The Cal-Adapt: Analytics Engine will guide users to customized and curated localized climate data to make informed decisions based on the best science to improve electricity sector resilience. This platform will provide users with customized data, advanced analytics, and powerful cloud computing resources, allowing users to perform high-level analysis without needing to download massive localized climate datasets.

analytics.cal-adapt.org/

Where We Started

Cal-Adapt:
ANALYTICS
ENGINE

Climate Data and Analytics Workshop

In September 2021 our team conducted a workshop titled “Climate Data and Analytics for California’s Ambitious Climate and Energy Goals: A Virtual Workshop for Potential Users.” This three-day workshop raised awareness of various state-supported climate data efforts, solicited feedback on critical use cases for the energy sector, and spurred collaboration and alignment among creators and users of climate data.

Cal-Adapt:
ANALYTICS
ENGINE

analytics.cal-adapt.org/

Data Applications

Stakeholder engagement with key representatives of the energy sector began in April 2021. This engagement drove prioritization for this project of five key data applications within the energy sector.

In Progress

- Threshold-based analytics for asset-by-asset vulnerability assessments and updating design standards
- “Hourly climate profiles” (for future time periods) as inputs into production cost, energy load forecasting, and other models

Forthcoming

- Distribution of extreme temperature events to inform peak load, demand forecasts, and other applications
- Climate metrics and analytics to support long-range wildfire planning and management
- Using climate data to examine impacts on renewable energy generation and operations

EAGLE ROCK ANALYTICS

GEOSPATIAL INNOVATION FACILITY
Cutting-Edge Mapping Technology at UC Berkeley

UCLA Center for Climate Science

Spatial Informatics Group

Energy+Environmental Economics

EARTH & ENVIRONMENTAL SCIENCES

Where We Started

5 Data Applications

- Prioritized from an original list of 15

3 Day Workshop

- Sept 2021
- Detailing the 5 priority data applications with stakeholders
- Stakeholder deep dive

6 Collaborating Institutions

- A team of academic, technical, and policy leaders are building the Cal-Adapt: Analytics Engine

Where We Are Now

Cal-Adapt: ANALYTICS ENGINE

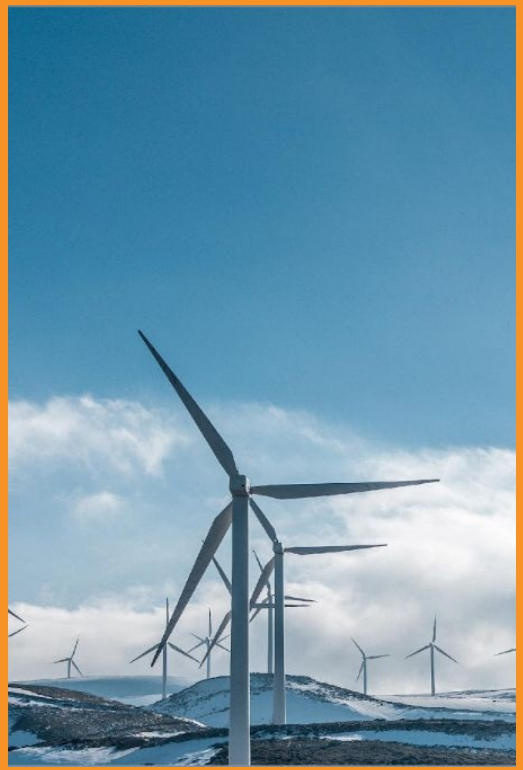
CA's Fifth Climate Change Assessment

"A new and critical addition to California's Fifth Climate Change Assessment is the Cal-Adapt: Analytics Engine, which reflects the California Energy Commission's intention to move beyond providing data, and to provide support in effective application of that data in California's resilient clean energy transition."

– *Susan Wilhelm, California Energy Commission*

Cal-Adapt:
ANALYTICS
ENGINE

analytics.cal-adapt.org/



Transforming Climate Data

The Analytics Engine JupyterHub is an interactive cloud-based environment providing access to localized California specific climate data, analytical toolkits, and cloud compute resources. Users access Jupyter notebooks which provide the capacity to find data, process or select data, perform analysis and visualizations, and download smaller subsets of data into local machines using the open-source Python package *climakitae*. Current analytic functionality includes time series, threshold-related analysis, average hourly profiles, and changes contingent on warming levels.

User Personas

The Analytics Engine identified key user personas to better understand potential platform users and uses.

Technical User

… uses tools and data to refine information and help shape it into something that suits their technical and decision needs. We can find our technical user at a utility company or a government entity that regulates those utility companies.

Policy User

… is in a decision making position in which they can impact the behavior and outcomes within the energy sector. Their biggest motivations include lowering risk and uncertainty, lowering costs, and protecting public health and safety.

Researcher / Academic

… is interested in what the Analytics Engine is working on and are likely to use the information we offer for their own scientific purposes. We can usually find this group at universities doing research or in the private sector doing consulting.

Public Consumer

… is curious about climate change and the natural sciences and aims to better understand what's happening. They may have personal experiences in or a personal passion for this kind of work and are looking to be more informed.

User Testing

The team developed usability testing protocol for the first iteration of Jupyter notebook and website content to gather rapid feedback from potential users. Notebook and website content revisions and expansions incorporate this feedback. Additional user testing is underway.

Cal-Adapt:
ANALYTICS
ENGINE

Where We Are Now

35 Datasets

- Future climate projections and historical climate from Global Climate Models downscaled using the Weather Research and Forecasting model by UCLA
- More data from Localized Constructed Analogs statistical downscaling coming soon from UCSD Scripps Institution of Oceanography

4 Working Group Meetings

- Dec 2021, Feb 2022, June 2022, & Aug 2022
- Enabling co-production of the platform and analytics with stakeholders

6 Jupyter Notebooks

- Step-by-step functionality to access, analyze, and plot climate data available on the Analytics Engine
- Additional notebooks currently in development

SY32D-0639: Beyond Data Repositories and Visualizations: Supporting Evidence-Based Decision-Making for Resilience Planning in the Energy Sector via the Cal-Adapt: Analytics Engine

Grace J Di Cecco[1], Owen M Doherty[1], Andrew D Jones[1,4], Naomi L Goldenson[3], Nancy Thomas[1], Kripa Jagannathan[1,4], Brian Galey[2], Nancy L Freitas[1], Mark D Koenig[1], Justine Bui[2], Susan Wilhelm[1], Michael Mastrandrea[2], Elizabeth Elliott McClenny[1], Mithra Moezzi[2], Ashley Conrad-Saydah[2], Victoria Ford[3], Nicole Keeney[3], Cora Kingdon[1,4] and Smitha Buddhavarapu[1]

From Data to Decisions

Localized climate data is massive, too big for decision makers to use. Desired metrics are too complex for simple visualizations and tools. The Cal-Adapt: Analytics Engine provides a flexible environment for analyzing and processing climate data on the cloud, making the data available to decision makers in the form desired for their application.

Existing challenges

1. California public agencies and departments invest billions of dollars in integrated climate action: mitigation, adaptation and resilience

2. Analysts and decision-makers need access to vetted, reliable, and relevant data that climate scientists have deemed appropriate for use

3. Inter-agency data sharing is challenging and inconsistent; data storage is expensive; centralized repository is needed

Our approach

We are building a **cloud-based analytics** platform through **co-production** to transform petabytes of data into useful and accessible data products.

Co-production develops trusted relationships through facilitated conversations between scientists, developers, and energy-sector practitioners, bridging climate expertise and grid resilience expertise.

Stakeholder Requested Features

Stakeholders, scientists and developers convened at a series of meetings and working groups and collectively identified the following priorities:

1. Threshold & Asset Vulnerability

Decision makers require specific information to understand how climate change will impact built infrastructure. Frequency of extreme events exceeding a known threshold, or other metrics of interest, are calculated at a specific location.

2. Extreme Heat & Electricity Demand

Climate change perturbs peak loads and typical consumption profiles. Characterizing the distribution of extreme events that impact electricity for cooling buildings or pumping groundwater.

3. Hourly Climate Profiles for Energy Modeling

Energy system models often parameterize sub-daily weather impacts. The Analytics Engine instead pre-processes data and creates outputs customized to support production cost and capacity expansion models used in developing integrated resource plans and assessing electricity sector reliability.

4. Climate and Wildfire

Wildfire mitigation planning and resilience (e.g. hardening upgrades) planning require specific formulations of wildfire risk. Analytics Engine converts gridded fields to application-specific information.

5. Supporting a Low-Carbon Grid

A high-renewables grid requires understanding of how zero-carbon energy resources (e.g. solar, wind and hydro) co-vary and behave during periods of extremes in electricity demand. Specific formulations of climate informatics are needed for energy modeling and planning.

The Analytics Engine

What's in the Analytics Engine?

1. Code
Jupyter Notebooks pre-loaded with energy sector applications; an open-source code repository full of established and innovative evidence-based climate and energy functions.

2. Cloud Compute
AWS cloud compute resources tailored and ready to execute on climate data

3. Climate Data
A massive, open data repository full of CEC-funded climate data localized to California (WECC-wide) with high spatial and temporal coverage.

Climate Data Availability
The Analytics Engine hosts the foundational climate data for California's Fifth Climate Change Assessment. This include CMIP6 downscaled climate data at daily (LOCA) or hourly (WRF) temporal resolution via two downscaling methods:
Statistical downscaling via LOCA (1) at 3-km over California
Dynamical downscaling via WRF (2) at 3-km over California, 9-km and 45-km over broader regions (Figure 1). Downscaling was conducted by Scripps Institution of Oceanography and UCLA.

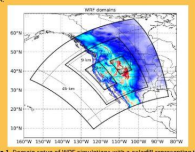


Figure 1. Domain setup of WRF simulations with a colorfill representing the complex topography across the Western US. Source: Stefan Rahimi, UCLA

Stakeholder uses of climate data are highly varied. Some applications require consideration of multiple ensemble members; others require consideration of multiple future climate scenarios. The need for bias-corrected and non-bias-corrected data is also driven by diverse utilization. Models in the data catalogue are prioritized by their skill in capturing California climate (3). The wide range of applications, and in turn data, necessitates a large data catalogue, exceeding 1.75 PB in size.

Building the Analytics Engine

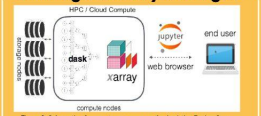


Figure 2. Schematic of compute resources on the Analytics Engine, from https://pangeo.io/architecture.html#architecture.

The Build of the Analytics Engine
Meeting stakeholder needs requires providing users with computational capacity, scientific expertise and data storage – accessible from their own computers.
Our software ecosystem builds upon Pangeo's Docker image (4) and AWS cloud computing resources to provide:
- Ability to use high-level data models (Xarray)
- Ability to work interactively in computing environment, with containerized applications through Kubernetes (JupyterHub)
- Ability to leverage distributed parallel computing on cloud computing systems automatically, no specialized code required from users (Dask)

Data storage is provided by AWS Open Data Program, which donates 1.75 PB of S3 storage. Stored data is considered analysis-ready, cloud-optimized (ARCO) compliant, as it is stored in Zarr format.

Linking Stakeholders, Science and Programming
Open-source analytics and tools are available for use with data on the Analytics Engine, in the form of a python library *climakitae* (5) and publicly available Jupyter Notebooks (6) providing guidance and examples for identified applications.

Notebooks and analytical tools are developed and improved through iterative methods including working groups and usability testing sessions with stakeholders.

Example Notebook Topics
Getting started: introduction to the data on the Analytics Engine
Timeseries example: perform common timeseries analyses with climate data
Extreme events: calculate frequencies of exceedance events and return values and return periods for user-defined thresholds
Warming levels: evaluate models across scenarios using global warming levels
Average Meteorological Year: produce synthetic hourly annual timeseries that represent average historical or future conditions

Stakeholders & Users

The co-production process for the Cal-Adapt Analytics Engine involves many different stakeholders from organizations across public and private sectors, with an emphasis on applications in the electricity sector.

Technical Users
- Users of our platform include technical and scientific staff at Investor-Owned Utilities and Publicly Owned Utilities
- Technical staff at various government agencies
- Researchers in the private sector and at academic research institutions

Many of our technical users are not climate scientists by training, so guidance and best practices for using climate data is embedded within our analytical tools and Jupyter Notebooks accessed through the JupyterHub

Semi-Technical & Policy Users
We provide a public-facing web presence and data catalog to easily communicate information about data availability and intended uses

The Cal-Adapt Analytics Engine will also host data to support research across different disciplines for **California's Fifth Assessment.**

Trustworthy Data

We have developed metadata standards to determine which datasets to host on the platform via neutral and standardized process: datasets must be CF compliant (7), follow CMIP6 naming conventions when possible (8) and will be evaluated using a *Data Quality Agnostic Test* based on FAIR principles (9).

F – Findable
- Dataset conforms to Cal-Adapt Metadata Standards
- Dataset assigned persistent and unique identifier

A – Accessible
- Metadata are open and free
- Metadata and data do not require specialized software to access

I – Interoperable
- Metadata use an accessible and broadly applicable language
- Metadata documentation is easily findable by users

R – Reusable
- Metadata describe input/source data
- Metadata are CF compliant

We will provide ratings for datasets based on completeness of Data Quality Agnostic Test criteria.

Contact us!

analytics@cal-adapt.org

Scan the QR code below to visit us at analytics.cal-adapt.org

Thank you to our funders

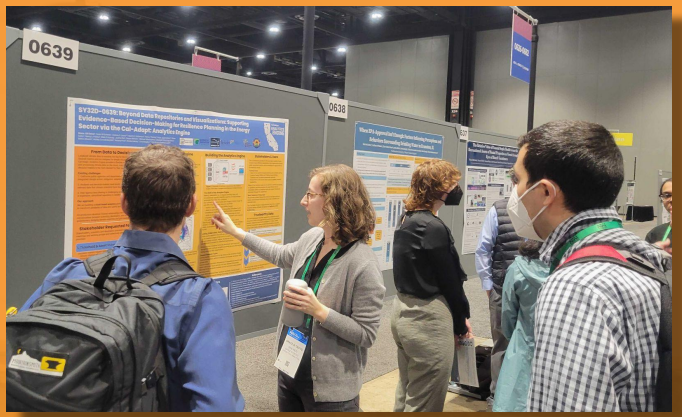
FUNDING PROVIDED BY THE CALIFORNIA ENERGY COMMISSION

California Energy Commission – EPIC Program (EPC-20-007)

References

1. Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using Localized Constructed Analogs (LOCA). Journal of Hydrometeorology, volume 15, page 2558-2585.
2. Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., ... Huang, X. -yu. (2021). A Description of the Advanced Research WRF Model Version 4.3 (No. NCAR/TN-556+STR). doi:10.5065/1dfh-6p97
3. Krantz, W., Pierce, D., Goldenson, N., and Cayan, D (2021) Memorandum on Evaluation Global Climate Models for Studying Regional Climate Change in California, EPC-20-006.
4. https://pangeo.io/
5. https://github.com/cal-adapt/climakitae
6. https://github.com/cal-adapt/cae-notebooks
7. https://cfconventions.org/
8. https://pcmdi.llnl.gov/CMIP6/Guide/dataDescriptors.html
9. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Cal-Adapt: ANALYTICS ENGINE

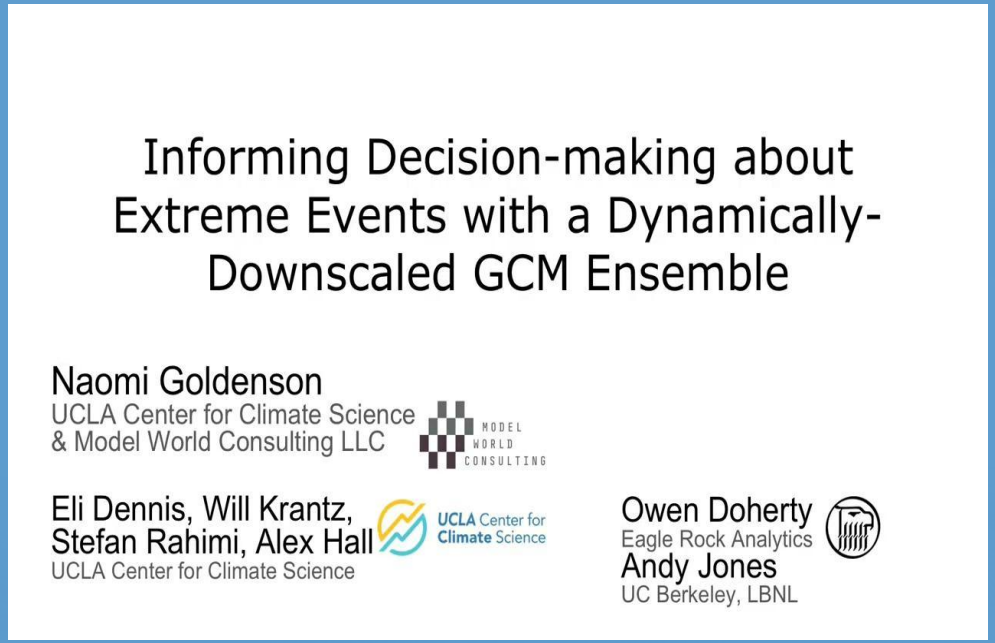


Poster presented by Dr. Grace Di Cecco at AGU 2022

Informing Decision-making about Extreme Events with a Dynamically-Downscaled GCM Ensemble

Presented by Dr. Naomi Goldenson at AGU 2022

Click here to be directed to the video!



What's Next

Cal-Adapt: ANALYTICS ENGINE

Potential Extensions

Heat and Human Health Impacts

Calculations of human focused heat indices and impacts

Scenario and Statistical Planning for Hydrological Modeling

Water focused climate efforts

Smoke Emissions and Transport

Smoke estimated needed for solar energy modeling, human health, and regional planning purposes

Climate Indicators

For standardized reports and commonly used climate metrics and indicators in more general support of California's Fifth Climate Change Assessment

Consulting and Vulnerability Assessment for IOUs

Operationalizing thoughtful population and asset vulnerability studies

Regional Assessment Support

Regional and local climate assessment reports and programmatic goals

In addition to various support for specific applications and connectivity with other state-funded projects.

Coming Soon

Cal-Adapt: ANALYTICS ENGINE

- Upcoming notebooks
  - Understanding Climate Uncertainty
  - New Climate Data for IEPR
- You can contact us at: analytics@cal-adapt.org
- For more information, check out our website: https://analytics.cal-adapt.org/