

Choosing Temporal and Spatial Scales

Cal-Adapt

Published: 2024-01-01

Last modified: 2026-10-11

For attribution, please cite this work as:

Cal-Adapt. (2026). *Choosing Temporal and Spatial Scales*. Cal-Adapt.

<https://analytics.cal-adapt.org/scientific-guidance/temporal-spatial-scales.html>

Temporal and spatial scale choices have big implications for the size of the data and the computational time required for an analysis. This page covers the key temporal and spatial choices available in the Analytics Engine and how to approach them.

What temporal scale should users consider?

Temporal scale considerations for users often include four key choices:

- **Temporal resolution** of the data (hourly or daily)
- **Sampling window** (number of years of analyses)
- **Temporal aggregations** (whether and how to average across years of analyses)
- **Timeframe** (historical, earlier-21st, mid-century, end-century)

The data and tools in the Analytics Engine allow a user to work with different options for these four choices. Note that these choices have big implications for the size of the data and computational time required to perform analyses. Therefore choices need to be made thoughtfully based on the type of application and potential tradeoffs, i.e. choosing the finest temporal resolution of data might limit the ability to work with longer sampling windows or a larger number of models (if data size is an issue). A user can also conduct preliminary analysis within the Engine's server to examine the implications of these different choices before a decision is made on the temporal scale of the data to be downloaded for further use.

Temporal resolution

In the Analytics Engine, depending on the type of data and variable of interest, a user can select between hourly or daily resolution data or pre-aggregated data at daily or monthly scales (see the [Data Catalog](#) for details). Users' choice about the temporal resolution of the data should be informed by the temporal nature of the climate hazard of interest and how it interacts with the user's system and planning process.

Overall, the broad recommendation is for users to identify specific temporal aspects of their application, start coarsely with some preliminary analyses, and then work towards larger sets of data with finer resolutions. For instance, if users are working with broader trend analysis, changes in central tendency, or seasonal questions, where very specific temporal characteristics are less relevant, they may be able to work with data at daily scales or even pre-aggregated monthly, seasonal, or annual scale data. However, if the application or process involves a daily or hourly model, or if an 8760 analysis is needed, then the user will likely need to work with the higher



resolution data. Generating a *Typical Meteorological Year (TMY)* is another example where hourly data may be required for a long climatological period (see the [TMY methodology notebook](#)).

Additionally, if a user is looking at applications or hazards where a certain temporal characteristic (e.g. time of day or the diurnal pattern of data) is of critical importance, then the choice of daily versus hourly data can make a big difference to the results. For instance, examining wildfire risk in the afternoons, hourly exceedances for parking lot design, storm surges, or Santa Ana winds may need hourly or sub-hourly data, whereas examining the recurrence of long-term droughts might not require as fine resolution data (i.e. it may be okay to trade off finer resolution data in exchange for the ability to include a larger number of years in the sampling window).

Sampling window

Often the choice of an appropriate sampling window is determined by the type of planning or modeling processes that a user is interested in. **It is important to note that longer time windows (i.e. 30+ years) are better able to sample key modes of variability that affect California**, e.g. a 30 year sampling window is needed to capture a reasonable number of El Niño or La Niña cycles, as well as positive and negative Pacific Decadal Oscillations. In addition, some applications like [Extreme Value Analyses](#) require at least 30 years of simulation to appropriately pull event maxima. While it is possible to select and examine shorter sampling windows within the Analytics Engine, users need to be aware that these windows might not adequately represent the modes of variability, and therefore may not be fully representative of the various hazards that might impact their systems (i.e. there is a possibility of missing out on key hazards).

If users have to choose a sampling window that is less than 30 years, it is recommended that they use as many models and ensemble members as possible in their analysis. Using data with many different initial conditions and perturbations enables a better representation of different forms of natural variation over a shorter sampling window, than a single model and ensemble member (see [Model Selection](#))

Temporal aggregations

When analyzing data at certain resolutions and sampling windows, some applications may require further temporal aggregations (e.g., for input into other models, or for developing summary statistics or graphs for reports). The tools in the Analytics Engine allow users to analyze events or metrics at a finer scale within the engine's servers, and then aggregate and download (or further analyze) the data at a coarser scale. For instance, users can evaluate certain heat wave metrics at a fine scale within the engine (e.g. use hourly data to examine total hours of heat index above 90 degrees F). Users can then choose to temporally aggregate the resultant derived heat wave data more coarsely before downloading (i.e., total daily/weekly hours of heat index above 90 degrees F, or daily max heat index) without losing the quality of information. In this manner, the Analytics Engine enables users to examine heat waves at a fine scale using hourly data, without having to download the entire suite of hourly data from every model. Some fine-scale metrics (such as heat index) are already pre-aggregated in the Analytics Engine (some of these pre-aggregated metrics can be found [here](#)).

Broadly, it is recommended that the temporal averaging (if necessary) be done at the end, once the temporal variability and distribution have been examined and considered within the Analytics Engine server. For example, upon examining hourly versus more temporally aggregated data, it was found that for energy demand forecasts, it is more appropriate to calculate demand



for each hour and then average or summarize the demand outputs across several years. This was found to be more appropriate than calculating energy demand from the average over a day or over several years, since energy demand varies non-linearly with temperature and other factors. Examples of how to work with temporal aggregations of hourly data depending on desired metric are demonstrated in the Analytics Engine's [Heat Index](#) notebook and the [Annual Consumption](#) notebook.

Timeframe

How to choose a timeframe for analysis (whether to analyze climate impacts for mid-century, 2070, end-of-century, or any other period) is one of the most common questions that users have. Users often need to align their timeframe choices with their planning horizon and/or with the lifetime of the asset/infrastructure under consideration. However, climatic changes, as well as the impacts and risks that stem from these changes, could be non-linear and increase at a more rapid rate towards the end of the century. Hence, near-term timeframes might not always provide a holistic picture of long-term risk and impacts, particularly for the higher SSPs. For instance, if peak load forecasting is analyzed only until 2050, it might not provide a good understanding of the risk that might ensue post 2060.

Therefore, when feasible, users should consider the evolution of risk over longer periods of time (e.g. 2070, end of century, etc), and then break up the analysis into shorter time windows that better align with their planning horizons, asset lifetimes, or other application-specific considerations. Such longer-term analyses may be essential to better understand how decisions or actions may need to scale over time, particularly for climate impacts that are likely to change at a much faster rate after mid-century (e.g. sea level rise). It can be noted here that using [Global Warming Levels](#) (GWL) as an alternate framework can help to overcome some of these challenges, as this framework is more suited to adaptive planning.

What spatial scale should users consider?

Spatial scale considerations for users often include three key choices:

- **Spatial resolution** of data (the size of grid cell or data at a point location)
- **Spatial extent** (the area over which data is analyzed)
- **Spatial aggregations** (whether and how to aggregate over a larger geographic area, such as counties or hydrological units, HUCs)

Similar to temporal scales, the data and tools in the Analytics Engine offer different options for these choices. Given that these choices have big implications for the size of the data and computational time required to perform analyses, **it is recommended that users conduct preliminary analyses within the Analytics Engine server to examine these implications before deciding on the spatial scale of the data to be downloaded or synthesized for further use.**

Spatial resolution

The first spatial scale choice that climate data users often need to make is whether they need data at a point location or if they are able to work with gridded data. In general, if users want to examine a hyper-local issue (e.g. a point-based asset such as a substation) or if they are working with a model or workflow that is parameterised or dependent on a particular weather station, then they



may need to use point location data. However, if users are looking at regional climate responses, then gridded data might be most appropriate.

Point location data

Climate models do not automatically output data at point locations. However, the Analytics Engine has developed a Weather Station [Localization](#) notebook that allows users to generate data at particular locations for which the user has historical data. Hourly temperature data is also available on the Analytics Engine at 32 select weather stations that were identified as important for the energy sector for the LOCA2-Hybrid dataset. If users are looking to examine a point location that does not have available historical data, then they would need to conduct a qualitative assessment using the finest resolution gridded data for the nearest grid cells to the point location to determine how representative the grid cell average is of the specific point location.

Gridded data

If users choose the LOCA2-Hybrid dataset, this is offered only at 3km x 3km resolution. However if users choose the WRF dataset, this data is available at 9km x 9km as well as 3km x 3km resolution (see [WRF vs. LOCA2-Hybrid](#)). **When choosing a spatial resolution, the broad recommendation is that users should first identify the specific spatial aspects of their application, and then conduct preliminary analyses in the Analytics Engine servers to compare between different resolutions.**

Users should evaluate the tradeoffs (information that is gained or lost) by choosing different resolutions. For instance, if users have data size or computational constraints, choosing very fine resolution data may restrict their ability to examine a greater number of model runs or larger spatial extents. It can also be noted that higher resolution spatial data is not always better or necessary for every analysis. Many informed decisions and investments can be made by using lower resolution data that considers more models and plausible future outcomes.

When conducting preliminary analyses, users must also be aware of the spatial heterogeneity/variability (topographic, climatic, or other) within the region that they want to assess. Finer resolution data could provide greater benefits for regions that are spatially heterogeneous (e.g. areas of complex terrain or regions that transition quickly from water to land) as compared to regions that are more spatially homogeneous (e.g. the Central Valley).

It can be noted that many areas in California are characterized by a high degree of spatial variability in weather, topography, and the human dimensions of infrastructure and demographics. Therefore, depending on the region and climate variable under consideration, the spatial heterogeneity within a grid cell (and also across adjacent/nearby grid cells) might be large. Overall, the goal should be to select the most appropriate spatial resolution for the context, that also allows for a broad range of model runs to be included.

Spatial extent

The choice of spatial extent is often dependent on the risk or climate hazard that users want to analyze, and/or the question they are trying to answer. The spatial extent needs to be large enough to properly analyze the climate issue under consideration. For instance, to analyze the impacts of a heat event on system-level planning, a user may need to examine the entire regional extent



where the impacts of the heat events may occur. For other specialized analyses, such as evaluation of streamflow, sector-specific domains (or spatial extents) such as watersheds can be applied to gridded climate data. Here, care must be taken to create a rigorous workflow which carefully considers edge cases where a grid cell may be only partially included in a polygon area (In the Analytics Engine, functions such as `rioxarray clip()` have been used to better examine some of these issues).

Users should also note that conducting hyper-local spatial analyses (e.g. using census tracts in urban areas) may require specialized approaches that are highly dependent upon a given question or application. Please see the [Census Tracts](#) page for more information.

Spatial aggregations

Some applications may require aggregating data across a larger spatial area (such as a county or a watershed) before it is downloaded or examined for further use (e.g. for developing synthesis graphs or summary numbers for reports). However, users should note that some of the quality of the information may be lost when spatially aggregating across different locations or grid cells. For instance some aggregations (such as averaging over large spatial areas) could lead to missing out on specific distributional impacts for different communities, which might have equity implications. Therefore, users need to be aware of the types of impacts that they might miss out on when aggregating across grid cells.

If aggregation is necessary, it is recommended that users do not default to spatial averaging. **Instead, users should first examine climate impacts without aggregation, and then aggregate at the end, once the spatial variability and distribution have been examined within the Analytics Engine server.** For example, finer scale analyses of population-weighted heat exposure can be first assessed at the grid cell level and then aggregated to a county level for reporting, after the variability has been examined. In such cases, users should consider reporting other distributional statistics over the aggregated areas, such as the spatial variability within the county.

The homogeneity of the climate hazard over the spatial area that a user is aggregating over is also very important. For instance, aggregating heat metrics over a largely homogenous plain area might be less challenging than aggregating precipitation over a topographically diverse area. In addition, users must exercise further caution when aggregating over grid cells for assessments of extremes, such as for [extreme value analysis](#). In extreme value analyses, if multiple grid cells from a climatologically diverse area are averaged, there is a chance of accidentally biasing the results to the most extreme value in the area and/or missing an extreme of interest. As in several of the previous questions, users will need to make an informed decision about the right balance between performing a good quality and rigorous analysis versus one that is computationally light and technically feasible within their context.

