An Approach to Synthetic Future Climate Hourly Profiles for Power System Modeling

Summary

Power system modeling tools typically rely on input data in the form of 8760 timeseries to capture the intra-annual conditions (including weather) that influence electricity supply and demand. However, nearly all global climate model (GCM) projections are limited to daily temporal resolution which presents a data challenge for incorporating their projections of future climate changes directly into power system modeling. This research presents an innovative approach developed by EPRI to create hourly weather timeseries for future climates at the local-level. A *monthly quantile anomaly* mapping technique is used to shift historical profiles according to the seasonal climatological shift being projected by an individual or ensemble of climate models. This method preserves important, real-world characteristics from the historical record that is otherwise missing from climate model output. Specifically, this approach captures important information from the historical record, such as locationally-specific extremes which can be missing from coarse climate projections, natural variability which isn't always well represented in the climate models, and important joint correlations among physically-linked variables such as wind, solar, and temperature. This method has many potential applications in the power sector, where 8760 timeseries are needed for simulation modeling, as well as for resource adequacy assessments that require many realizations (e.g., a sample of 100s or 1000s) to identify possible extremes for stress-testing a future year of interest.

Introduction

Many electricity system capacity planning and operational modeling tools are designed to use hourly timeseries, or profiles, as input data. These tools typically use 8760 timeseries of historical meteorological data or synthetic profiles (e.g., a typical meteorological year) to capture the intra-annual weather conditions that influence power supply and demand. Increasingly, power system planners are interested in accounting for climatic trends and potential extreme events in the meteorological inputs to their simulation models[1]. However, nearly all global climate model (GCM) projections are limited to daily temporal resolution which presents a data challenge for incorporating their projections of future changes directly into power system modeling.

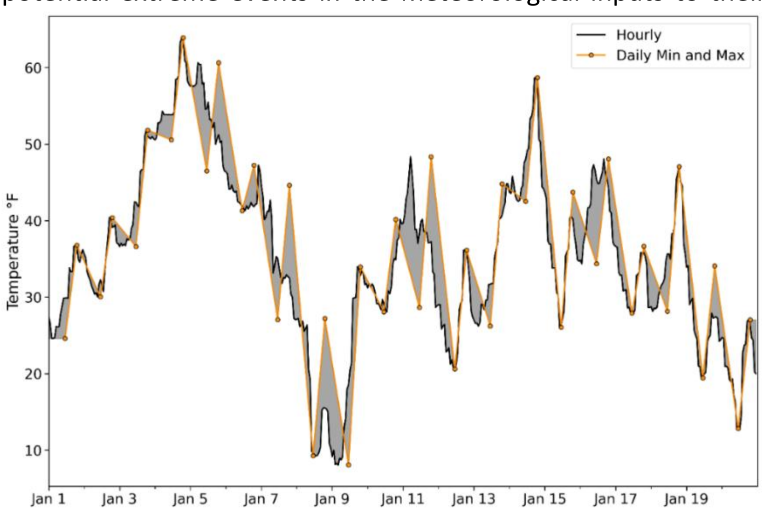

Figure 1: Hourly temperature vs daily max and min from Jan 1-20, 1950. For the winter season this illustration assumes the typical pattern of daily max temperature at 2 pm (19z) and daily min at 6 am (11z). Gray shading highlights differences between the daily and hourly lines.

Because of this limitation, an approach that transforms daily climate model projections into hourly timeseries is needed. One such 'temporal downscaling' approach is to interpolate between the daily values to fill in the missing hours, but this simple method could miss important diurnal fluctuations that have material consequences on electricity demand or design thresholds. Figure 1 illustrates this interpolation approach for the temperature variable, which is typically reported from climate models with a daily mean, maximum and minimum value; most other variables are limited to a single daily average value. Another approach is to conduct customized dynamical downscaling of climate model output to produce hourly resolution data, but this is both computationally and monetarily expensive. Between these two approaches, there is a range of potential methods worthy of evaluation. For example, a hybrid approach could combine machine learning with dynamical relationships to model the gaps between daily values; however, this approach can be computationally expensive and requires advaned expertise.

Here we present an innovative approach to develop hourly weather profiles for future climates based on the historical record at the local level combined with the climatological shift being projected by an individual or ensemble of climate models using a monthly quantile anomaly technique. This method preserves important information from the historical record, primarily in terms of variability, which itself isn't well represented in the climate models. Climate models are also not necessarily meant to capture extreme events at the local scale as much as they are meant to capture long term, large-scale changes. However, variability and individual extreme events are critical characteristics to consider for power system reliability. Furthermore, leveraging the historical record preserves the physical link between variables, such as wind, solar, and temperature, and this synchronicity may be critical for power system modeling. This physical link is difficult to maintain when using climate model data because many of the needed variables are not outputs of all models, requiring the end-user to pull additional variables from different models. This method leverages the important characteristics of both historical and projection datasets as described in Table 1. Additionally, this method can overcome the limitation of only having a single representation of annual variability, because all historical years since 1950 can be detrended and used to create an extended dataset with intrinsic variability over 72 years, yielding a vast array (e.g., hundreds) of realistic synthetic hourly timeseries that can be used to represent a future climate.

Table 1: Important characteristics from historical data and climate projections.

Historical Data	Climate Projections
Hourly data	Daily data
Realistic variability - Scales of weeks, months, & years from 72 years of historical weather (1950-2021)	Limited variability - Variability is constrained to the underlying physical model; typically not well-captured
Historical years only - Can't represent weather extremes that haven't happened	**Future years + historical simulations** - Can capture how the climate will change - Can represent weather that has never happened
Preserves physical link between variables - Variables are dynamically consistent since they come from the same dataset (ERA5)	Projection data lacks variables at hourly resolution - Physical link is absent when interpolating daily data or using variables from different sources
All variables available - i.e., 10 m & 100 m wind speeds	Limited number of variables - i.e., 10 m wind speeds only

Important or desired characteristic

Applications

While this approach was specifically designed to meet the needs of power system modelers, it can be applied more broadly to other economic sectors or end-users in need of hourly timeseries: including those focused on risk analysis, system planning, load projections, line ratings, asset/engineering design standards, among others. A couple of power-system specific applications are shown below:

<u>Application example 1: Load Projection or Capacity Expansion</u>

- Need: 8760 profiles of temperature, wind, and solar for planning years through 2050
- Approach: treating different GCMs/RCPs via sensitivity analysis

<u>Application example 2: Resource Adequacy</u>

- Need: 100s of realizations of a future year of interest to get a better understanding of the likelihood of specific extreme events
- Approach: sample to identify dozens of possible extremes that can be used to stress test the risk model

Data and Methods

This approach builds upon well-established methods from the climate science community to create a novel method for generating synthetic hourly profiles from historical and projected climate data.

For the historical source to form the underlying hourly timeseries, we use ERA5 gridded reanalysis data, from the European Center for Medium Range Weather Forecasting (ECMWF)[2], though this method could be applied to other sources of hourly historical data such as station observations. ERA5 data has a spatial resolution of 31 km x 31 km and is spatially and temporally complete, available from 1950 to present. For the projection source of future climatology, we use five CMIP6 climate models from the Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP). Two climate scenarios from each model, SSP126 (lower emission scenario) and SSP370 (higher emission scenario), are used to create a 10-member projection sample (5 models x 2 scenarios).

A high-level overview of each step is show in Figure 2 with a more detailed explanation below:

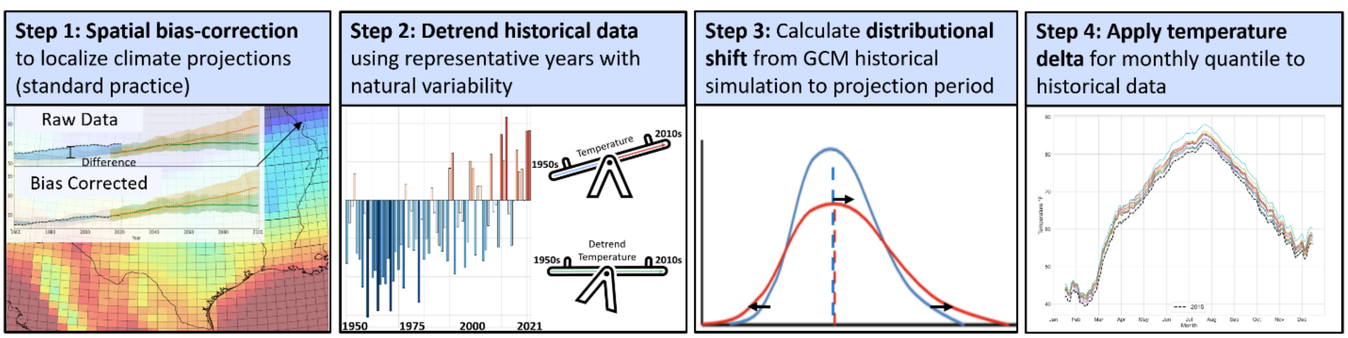

Figure 2: High-level overview of each of the 4 steps used to create future hourly synthetic climate projections.

Step 1: Spatial bias-correction to localize the climate projections

Spatial bias-correction is conducted for the GCM data using historical data (ERA5) as a reference, a standard best practice. Spatial bias-correction helps reduce biases in the climate models by comparing the climate models historical simulation with the historical data. Adjustments needed to make the historical climate model simulation more like the historical data are quantified and then applied to the future projections to remove known historical biases. Bias-correction may be done using lower-level statistics (i.e., mean) however, quantile mapping approaches have been shown to outperform the more simple approaches as they are able to also correct variance in the distribution[3,4]. While spatial bias-correction can remove some bias, it is not a perfect solution and can impact trends in the model and introduce inconsistencies[4]. While not done in this approach, limiting inconsistencies can be achieved through multivariate and spatially cohesive bias-correction[5].

Step 2: Detrend historical data using representative years with natural variability

Removing the warming trend from the historical data creates 72 equally representative 8760s but preserves the natural variability from the past 72 years. The natural variability that plays out from one year to the next or one decade to the next is critical to capture for resource adequacy assessments.

Step 3: Calculate distributional shift from the GCM historical simulation to the projection period

Cumulative distribution functions (CDFs) are created from the historical climate model simulations (and projections for 2015 – 2020) for the current climate normal period (1991 – 2020) while the projected CDF is created for a 30-year period centered on a specific future year, like 2050 (2036 – 2065). This results in 10 projected CDFs for each for each GCMxSSP scenario (e.g., 2036-2065). The temperature change (i.e., delta or anomaly) for 20 quantiles per month is calculated[3].

Step 4: Apply the temperature change (i.e., delta or anomaly) for each month's quantiles to historical data

The temperature delta for each of the 20 monthly quantiles (12 months x 20 quantiles) is applied to the detrended historical data (ERA5). This results in 72 years of synthetic 8760s for each climate model and scenario (72 years x 10 scenarios = 720 synthetic 8760s) for any future year. The number of synthetic 8760s can easily be increased by adding more models or emissions scenarios.

Results

The resulting synthetic future climate data retains many similar properties to the historical data, but with a seasonally varying shift based on the climate model and scenario. This preserves realistic weather variability, but the resultant shift turns the data into a theoretically more plausible future scenario. For example, by shifting 2015 to 2050, the overall 8760 profile is similar but warmer with the largest shift in the summer and fall and a smaller shift in the spring (Figure 3). However, there is a relatively large spread across the 10 future scenarios which includes 5 models and 2 emissions scenarios, as would be expected. In fact, this variation across the ensemble of climate projection data being used is desired; it demonstrates an approach to accounting for uncertainty by incorporating a range of potential future profiles when possible.

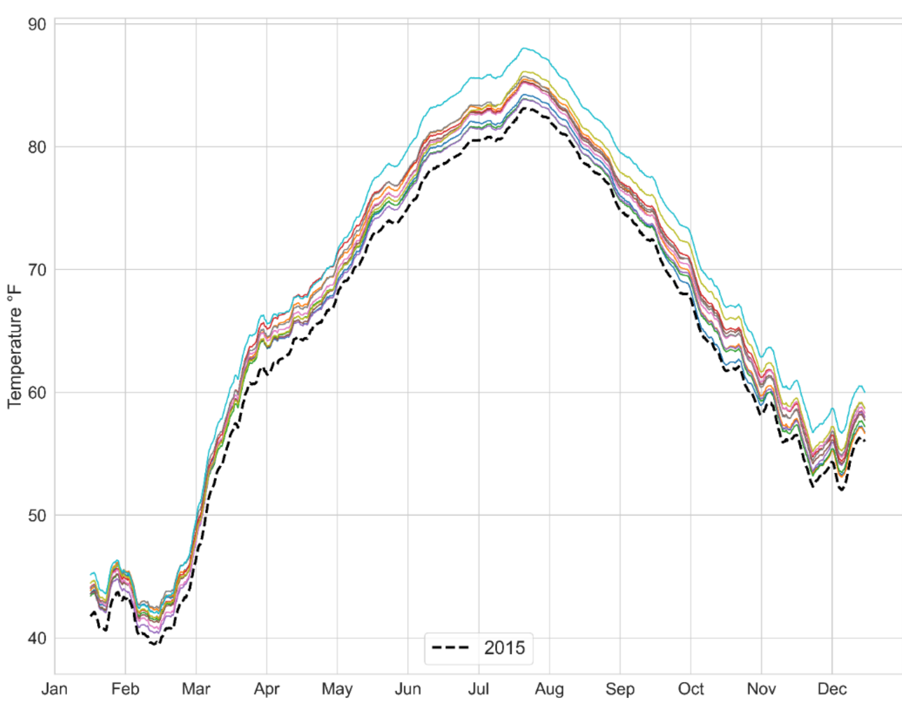

Figure 3: 2015 temperature (actual profile) compared to 2050 climate (synthetic profiles) shifted for each of the 5 climate models and 2 climate scenarios. The 8760s are smoothed for easier comparison between each synthetic profile.

To better illustrate the shift in the historical data by month, the 8760 from 1980 for Atlanta, GA is shifted to 2050 for a single climate model and two climate scenarios (Figure 4). The higher climate scenario (SSP370) results in an expectantly larger shift than the lower climate scenario (SSP126), however both scenarios result in a significant warming of the mean from 1980 to 2050. The extremes also tend to warm between 1980 and 2050 outside of the months of March and December. It should be noted that the most extreme cold, which occurs in March, remains in both the lower and higher climate scenario for 2050. This is likely due to the seasonally varying delta described in step 4 and preservation of these cold extremes is important for stress testing power systems during winter months.

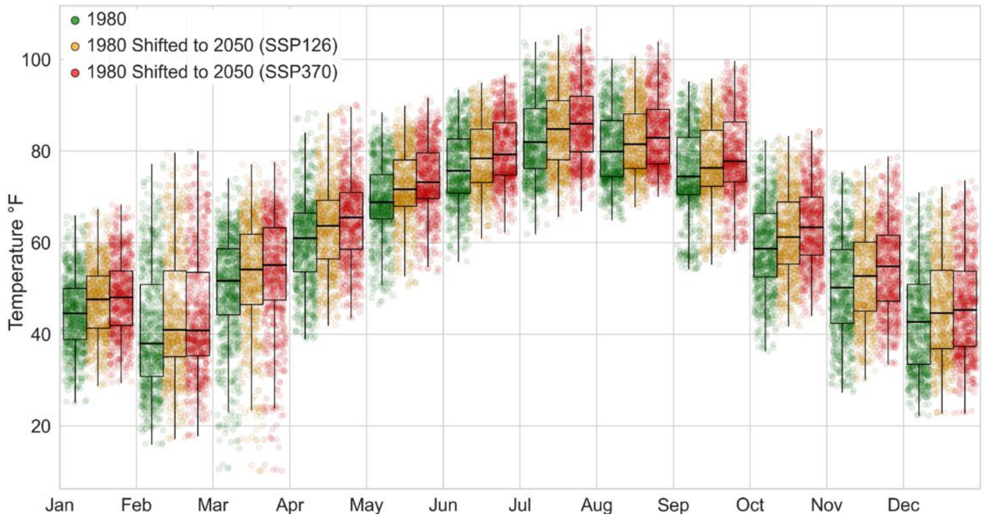

Figure 4: Comparison of 1980 shifted to 2050 for a single model (MRI model) and two different climate scenarios (SSP126 and SSP370) for Atlanta, GA.

Validation

Validation of this method is a critical step prior to application, to demonstrate confidence in the resulting profiles (since power system decisions may be made from models based on these synthetic profiles). Below, a validation exercise shows that the synthetic data for Denver, CO matches the benchmark climate model distribution very closely, with the largest differences being near the middle of the distribution (Figure 5). The tails of the distribution are particularly well captured, which is critical for most future power system applications.

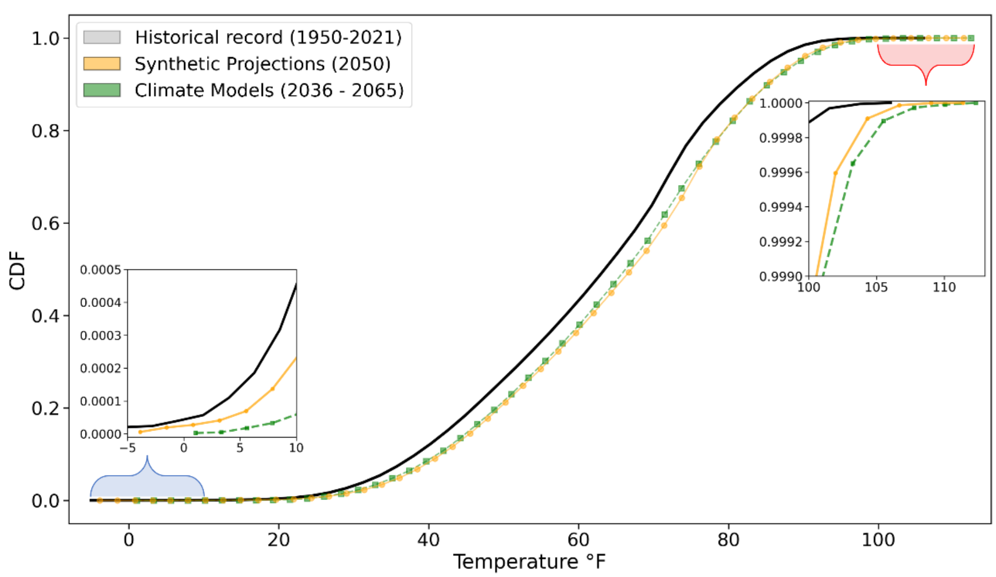

Figure 5: Comparison of the cumulative distribution functions (CDF) for the historical data (1950 – 2021), climate model projections (2036 – 2065) and 720 synthetic profiles for 2050 for Atlanta, GA.

A slightly different look at how the synthetic profiles compare to climate models is shown in Figure 6 for Atlanta, GA. The boxplots for the synthetic profiles generally match the climate models very well from December through May, but the mean is lower than the models from June through November and the warm extremes during the summer months aren't as well captured as for the Denver, CO example. However, the cold extremes are much more

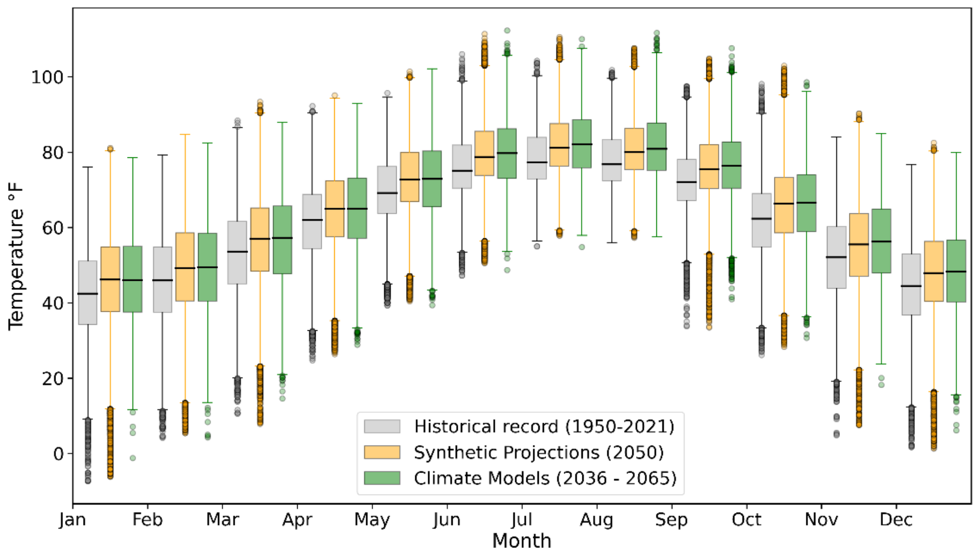

Figure 6: Comparison between the historical distribution (1950 – 2021), climate model projections (2036 – 2065) and 720 synthetic profiles for 2050 for Atlanta, GA.

prominent in the synthetic data than the climate model data and for many power system applications this offers the advantage of not overlooking the possibility of extreme cold events. While not all locations will have hot or cold tails in the synthetic profiles that match the climate models, small tweaks in the bias-correction technique could improve the representation of extremes. This preliminary validation suggests this approach could be very useful when needing hourly timeseries for future years.

Discussion and Conclusion

Climate projections are available at a daily resolution, but they are inherently probabilistic. This means we shouldn't take the output of a climate model for the year 2050 and assume that particular scenario will play out as shown by the model. Rather, multiple models, emissions scenarios, and years surrounding 2050 should be examined to get a better idea of what a year in the 2050 timeframe might look like. Because of the probabilistic nature of projections, the information that is often most critical to capture from the models are the long-term trends and changes in the tails of the distribution, not the actual values on any given day.

While GCM projections lack the hourly resolution that is often necessary for company planning, they can be leveraged to create realistic synthetic future scenarios that preserve the critical characteristics of the data, namely the trends and extremes. There are several benefits to doing this in that it captures real-world variability from a long historical record and creates potentially 1000s of realistic climate-adjusted profiles. This method also preserves the physical link between synchronous meteorological variables which is critical as many hazards that pose a risk to the system are comprised of multiple variables. While all variables are generally available in the historical period, climate models are limited in the number of output variables which means end-users need to pull additional variables from different models. Lastly, it can include historical years in future scenarios as a lower bound for risk assessments particularly concerned with the impact of extreme cold on the power system.

There are limitations to this approach that should be considered. Firstly, this method is best suited for single points as spatial coherence may be impacted by the point-based approach. It remains to be shown whether the spatial continuity among grid cells is sufficiently preserved when the quantile-delta mapping is conducted at the individual cell-level. Alternative approaches could be explored to consider the distribution of surrounding cells to better reflect the spatial relationship. Secondly, this approach is also applied to individual variables which could impact the relationship between variables. Because not all variables are expected to change significantly, like wind speeds, and not all variables are represented by climate models, like 100 m wind speeds, these variables can be left as they are in the historical period. However, as other variables change, the relationship between variables may also change. In this case, multivariate bias correction may be more appropriate.

Future work will focus on validation of these synthetic profiles against the climate models as well as other approaches. Specifically, a more detailed comparison of the statistical properties (e.g., variability and extreme event characteristics) of these synthetic profiles versus the native daily GCM projections, as well as a deeper dive into how well the synthetic profiles represent extremes like heat waves, cold events, and droughts, is needed. It will also be beneficial to explore alternative approaches to help identify when different methods are more appropriate for the application at hand. Some of these approaches are rather simple, like applying a fixed stationary adder (e.g., +5°F) to all historical 8760s to represent a future year when temperatures are projected to be 5°F higher on average, or temporal downscaling by linear interpolation between the climate models' daily min, max, and mean values to create hourly data. Others are more complex, such as the method outline in this article, or the use of regional climate models (RCM) initialized with daily climate model data to produce output variables with hourly resolution.

Technical Contact

For more information, contact Erik Smith (esmith@epri.com) or Delavane Diaz (ddiaz@epri.com).

References

1. Craig, M. T. *et al.* Overcoming the disconnect between energy system and climate modeling. *Joule* **6**, 1405–1417 (2022).
2. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
3. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J. Clim.* **28**, 6938–6959 (2015).
4. Maraun, D. *et al.* Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change* **7**, 764–773 (2017).
5. Vrac, M. & Friederichs, P. Multivariate—Intervariable, Spatial, and Temporal—Bias Correction. *J. Clim.* **28**, 218–237 (2015).

Appendix

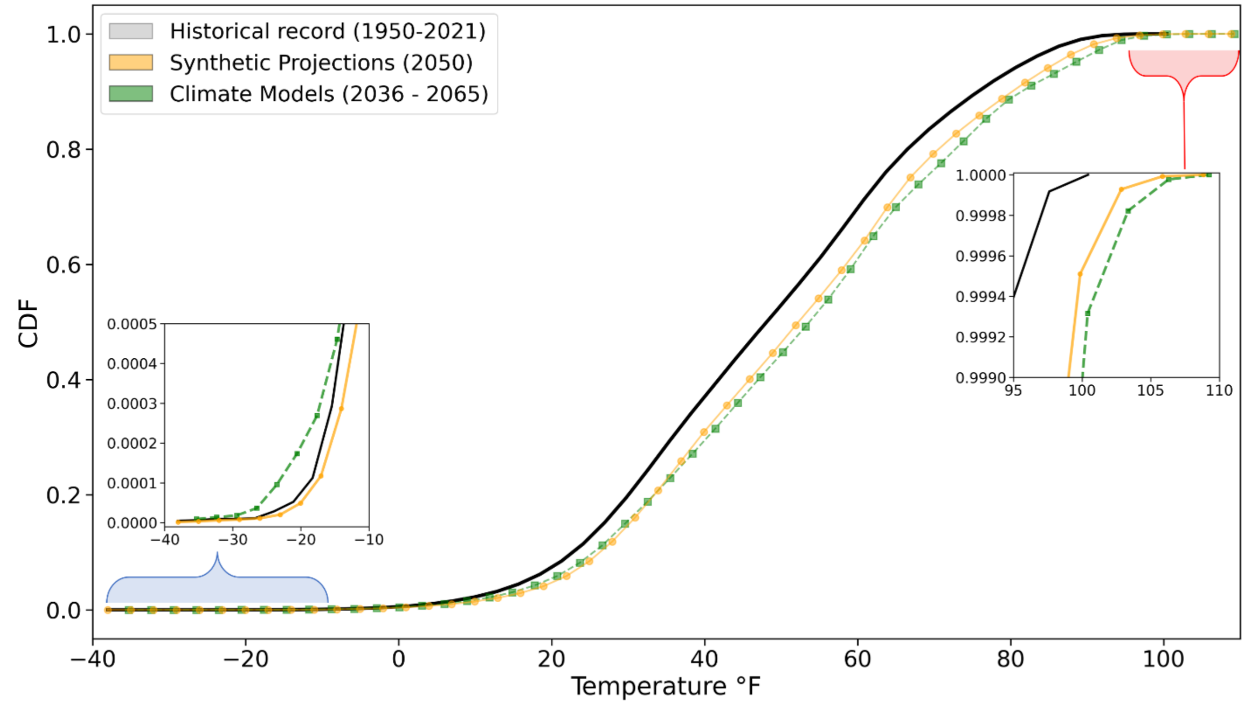

Figure A1: Comparison of the cumulative distribution functions (CDF) for the historical data (1950 – 2021)*, climate model projections* (2036 – 2065) *and* 720 *synthetic profiles for* 2050 for *Denver, CO.*

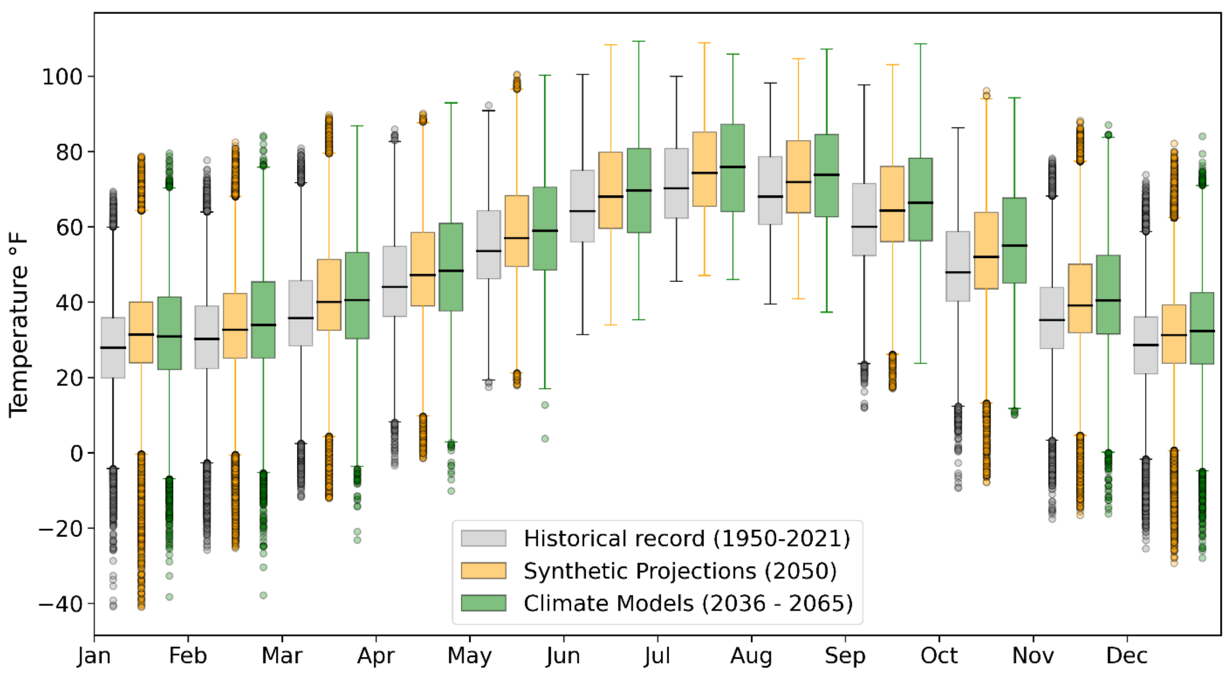

Figure A2: Comparison between the historical distribution (1950 – 2021)*, climate model projections* (2036 – 2065) *and* 720 *synthetic profiles for* 2050 for *Denver, CO.*