World Climate Service U.S. Regional Degree Days

June 2021

1. Introduction

Degree day indices are widely used in the U.S. energy industry for estimating and predicting energy demand. The value of this approach stems from the rather high correlation between degree days and energy consumption, both for heating (heating degree days, HDDs) and cooling (cooling degree days, CDDs). Degree days are therefore a simple but powerful example of a weather "impact variable" in which the raw weather data has been transformed into an index that is more useful for specific applications.

The World Climate Service (WCS) has developed subseasonal and seasonal dynamical model forecasts of HDDs and CDDs for U.S. regions, with spatial weighting proportional to population and residential fuel use classifications. The purpose of this document is to describe the methods we used to develop our degree day values; our goal is to enable others to reproduce our historical degree day numbers and to understand our approach to the forecast component of the WCS Trading Markets product.

2. Historical Aggregation

The development of a degree day forecast product rests on the foundation of a historical degree day database. Here we describe our historical calculations, beginning with monthly data and then downscaling to daily data.

A. Monthly Data

For a historical baseline, we rely on the monthly HDD and CDD values available in the longstanding climate division data (nClimDiv) from NOAA NCEI. Monthly HDD and CDD values (65°F base) are provided for 344 climate divisions in the contiguous U.S. (CONUS), with data back to 1895, and monthly updates occur around the 5th of the month. While we could have produced our own historical baseline of degree day numbers from other sources of temperature data, we feel that the NCEI benchmark provides a good foundation for our work[1].

---

[1] A minor shortcoming of the NCEI data set is that the February HDD and CDD values appear to be calculated for 28-day months in leap years. This may simply be because February 29 was excluded, but we are unsure of the precise NCEI methodology, and therefore we multiply the NCEI February HDD and CDD values by 29/28 in leap years.

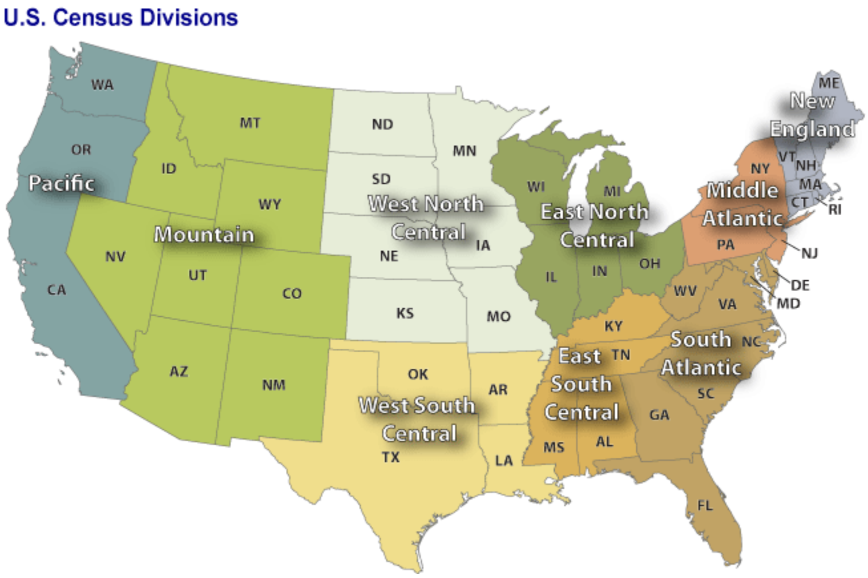
The first step in building our regional degree day history is to aggregate the monthly NCEI climate division HDD and CDD data into

(i) population-weighted state numbers

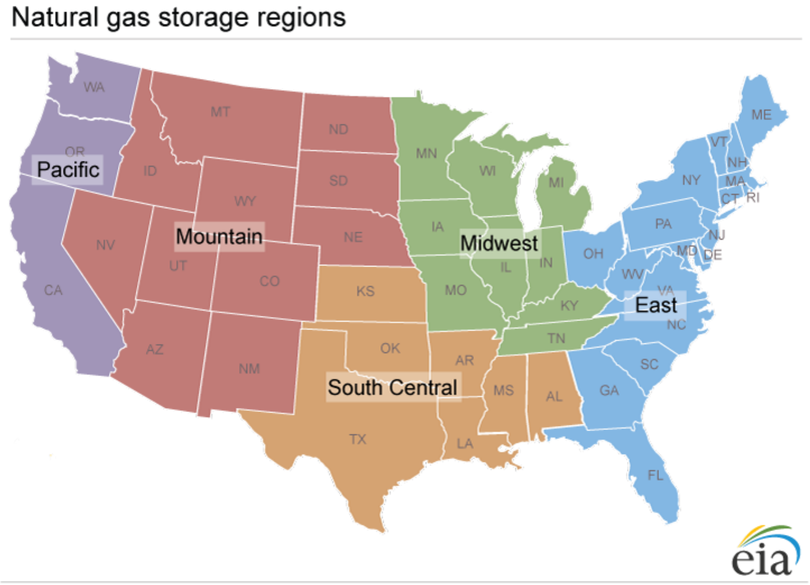
(ii) regional population-weighted and energy-weighted numbers.

Three sets of regions are produced, as illustrated below:

(i) U.S. census regions



(ii) EIA gas storage regions



(iii) electric power markets

For the census regions and EIA regions, the aggregation process is straightforward, as the regions are simply aggregations of whole states. First, we use the 2010 census populations within each climate division, as provided by NOAA's Climate Prediction Center (CPC)[2]. Second, we use the state-level populations within each region, both for total population and for the following residential fuel use classifications:

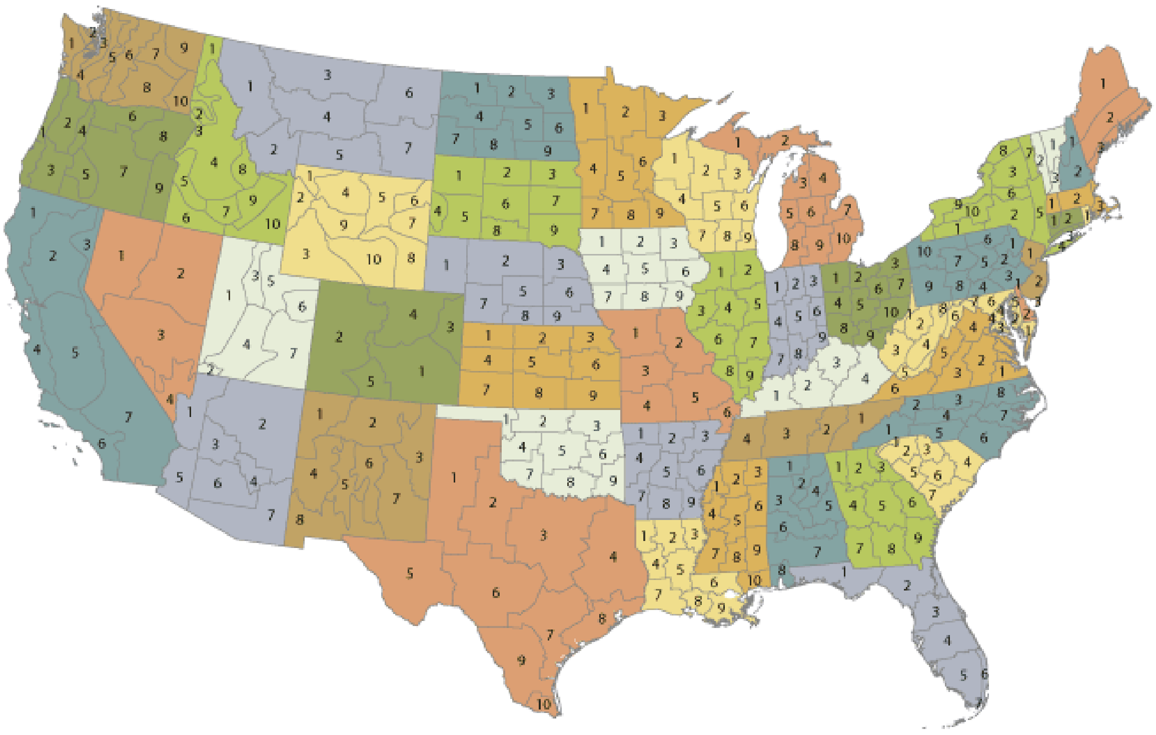
(i) natural gas
(ii) heating oil

The state-level fuel use numbers are also available from CPC. For example, the gas-weighted HDD value for the Pacific EIA region is the weighted average of the state-level HDD values in California, Oregon, and Washington, with weights equal to the state populations that rely on natural gas for heating.

The process for the electric power regions is complicated by the fact that the regional boundaries do not coincide with state boundaries. First we assign each climate division to the power region that intersects the largest area within the climate division. We then aggregate all climate divisions within each power region, using population values at the climate division scale for weights. For the energy-weighted aggregations, we assume that the fraction of each state's population within each fuel class is constant across the state. As an example, the northwesternmost climate division in Texas (see below) is assigned to the Southwest Power Pool (SPP, see above), and the energy weightings applied to that climate division are obtained from the Texas statewide fuel use information.

---

[2] https://ftp.cpc.ncep.noaa.gov/htdocs/degree_days/weighted/daily_data/populations/



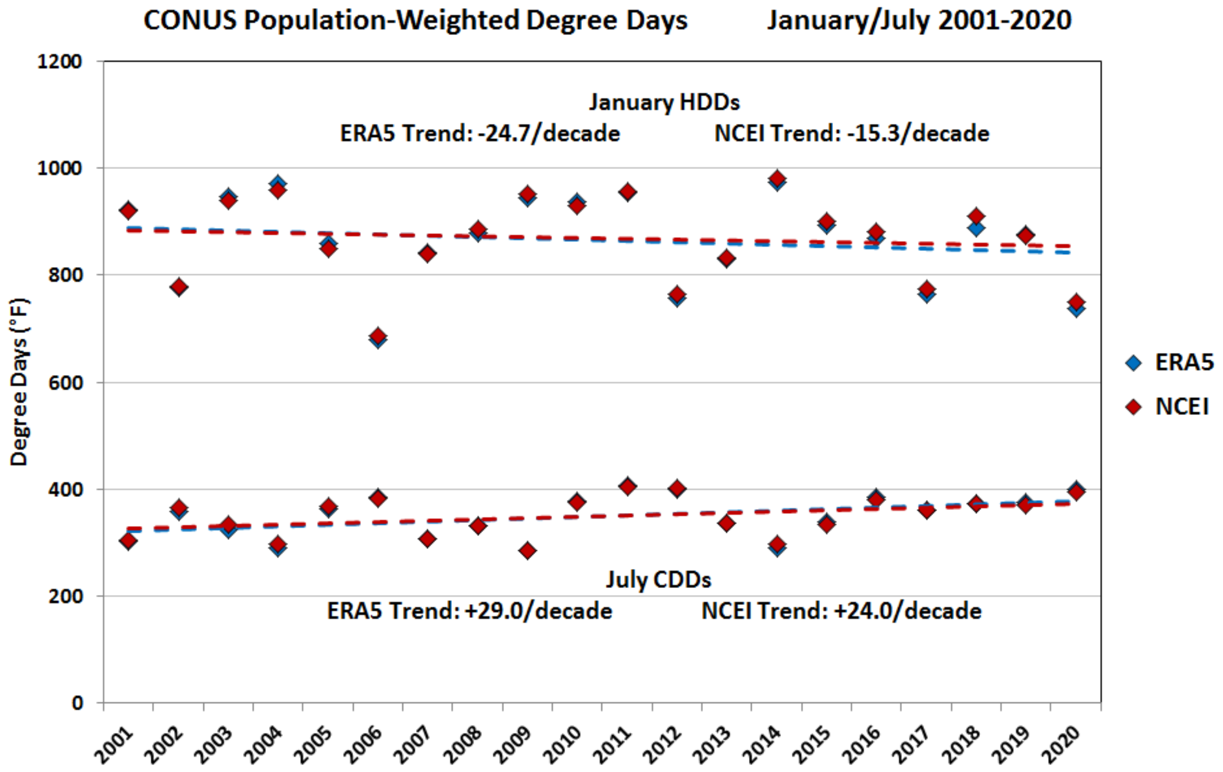
U.S. Climatological Divisions

B. Daily Data

The second step in the historical calculations involves downscaling the monthly degree day data to daily values. We use the 0.25-degree gridded ERA5 reanalysis for this purpose, as follows.

(1) We calculate a 2001-2020 daily climatology of daily mean temperature from both ERA5 and the climate division data, using the same method; in both cases we start with monthly mean temperature and then fit three harmonics to describe the annual cycle.
(2) We then find all of the ERA5 grid cells whose central points lie within each climate division, and we take a simple average of these grid cells to obtain a 2001-2020 history of ERA5 daily mean temperature for each climate division. Note that the ERA5 daily values are calculated for the 24 hours ending at 06 UTC, which is close to midnight in the CONUS.
(3) We then apply a bias correction to the raw ERA5 values by subtracting the difference between the ERA5 and climate division climatologies for the appropriate day of the year.
(4) Finally, we aggregate the bias-corrected daily values using the same state and regional aggregation process that we used for the monthly NCEI data.

A basic check on the quality of the daily ERA5 data is to compare the monthly HDD and CDD totals to the baseline values derived from NCEI. The chart below shows a comparison of CONUS population-weighted January HDDs and July CDDs, and while the interannual correlation is very high, it is apparent that ERA5 has a stronger warming trend, especially in winter. This is an important consideration when using ERA5 for near-realtime monitoring, and we discuss a trend adjustment below. However, for the historical data we eliminate the problem by scaling all of the ERA5 daily values so that the monthly totals are equal to the baseline (NCEI) values. This is done with a multiplicative method; for example, if the ERA5 monthly total is 890 and the baseline value is 930, then all ERA5 daily values within the month are multiplied by 930/890. In cases where the ERA5 monthly total is zero and the baseline value is non-zero, a small discrepancy remains; this occurs in 0.7% of months for HDDs, with a maximum monthly difference of 1.4 HDDs. For CDDs, more frequent small discrepancies occur in winter, mainly in the western CONUS, where ERA5 often has zero CDDs while NCEI has monthly CDDs between 0 and 10. Discrepancies of this magnitude will not materially impact estimates of regional energy demand.



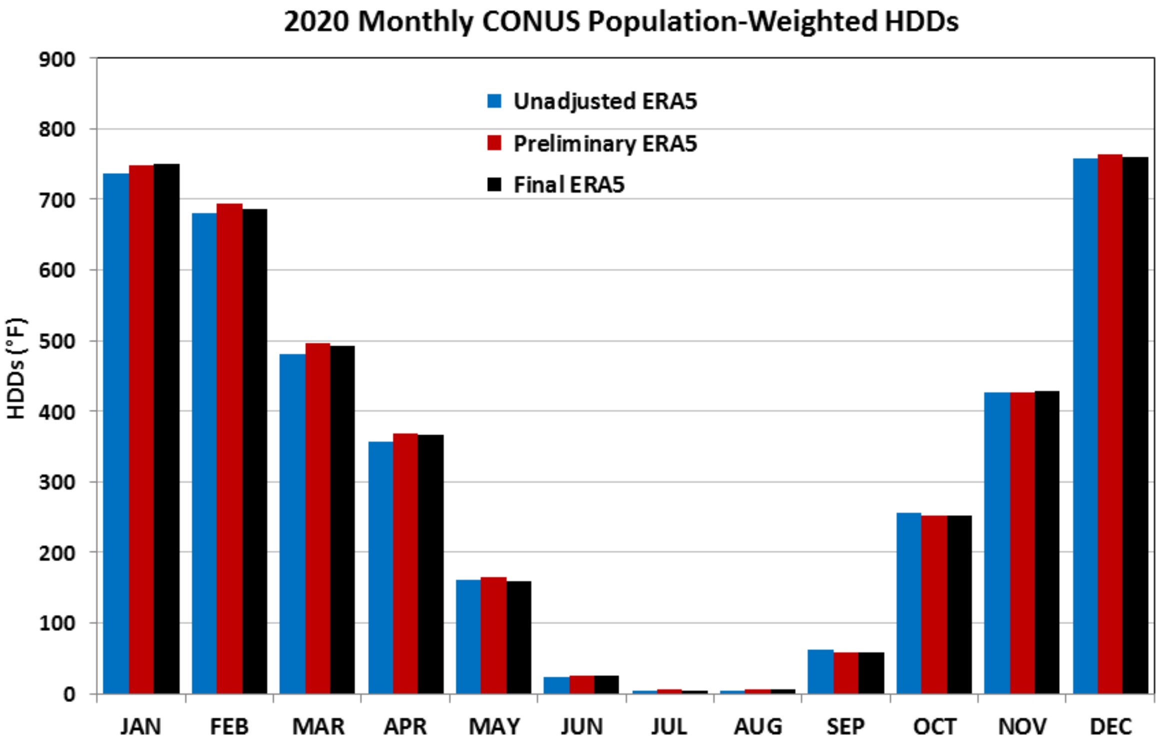
3. Realtime Updates

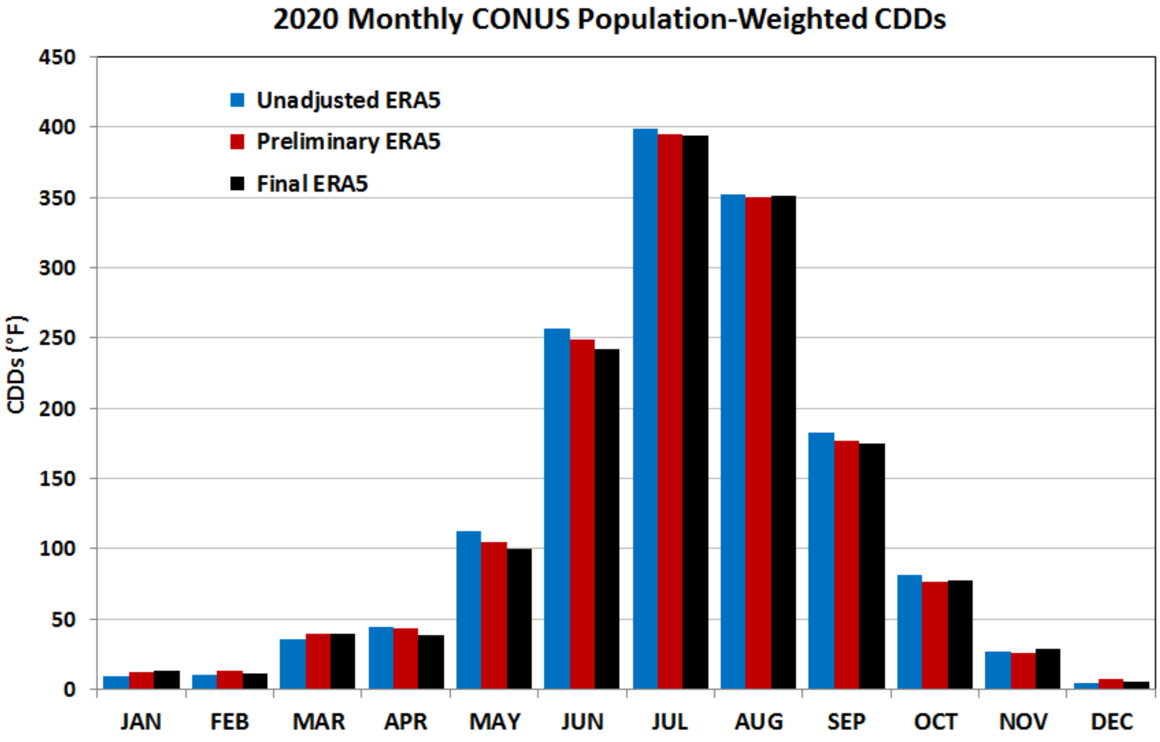
The NCEI climate division data updates once per month, allowing us to update the monthly and daily regional degree day history at the same time. However, it is useful to have preliminary updates on a daily basis, and for this purpose we use both daily-updating ERA5 and ECMWF model analyses.

A. Near-Realtime ERA5 Updates

The near-realtime ERA5 data has a lag of approximately five days, and we calculate preliminary daily degree day values accordingly. The values are "preliminary" because the monthly NCEI data are not yet available to perform the scaling, and so the preliminary ERA5 monthly totals do not equal the NCEI values. However, we minimize this problem by using an additional bias adjustment based on the difference between the scaled and non-scaled ERA5 values in the 2001-2020 history. Furthermore, we calculate the 2001-2020 linear trend in these differences and remove this trend from the preliminary daily values; this corrects for the fact that ERA5 has a stronger warming trend than NCEI.

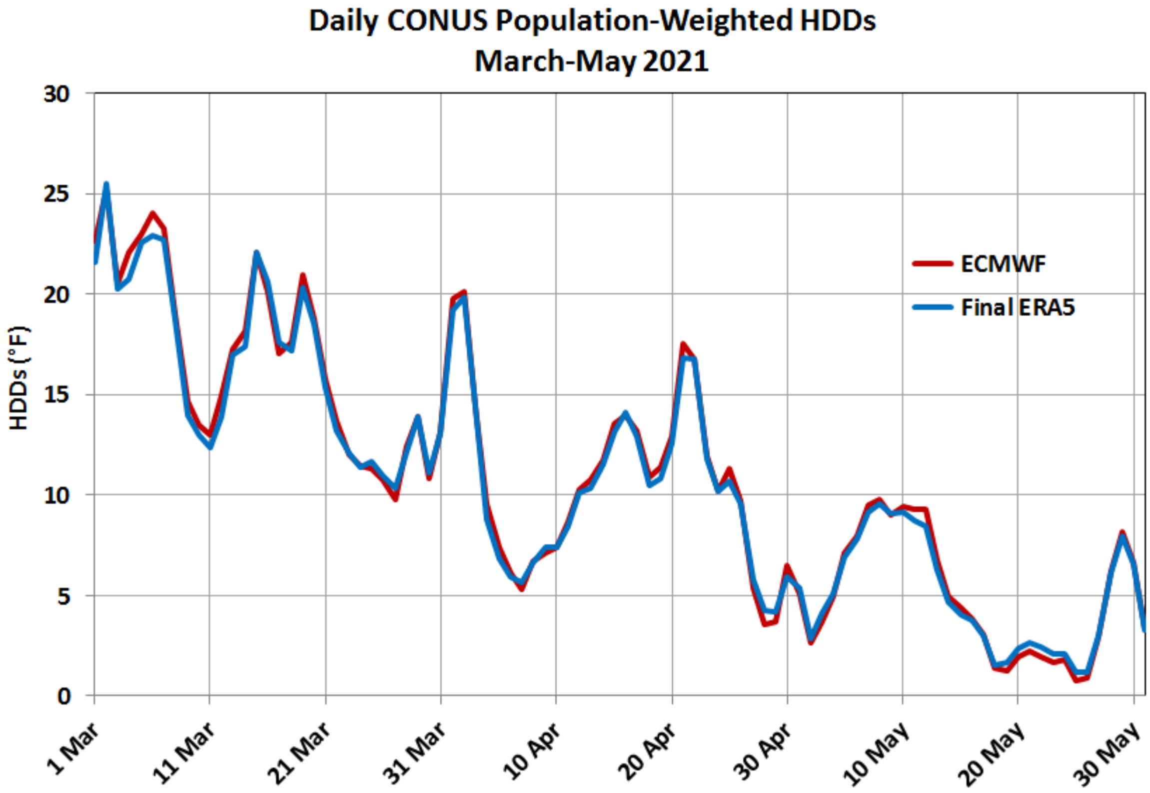
After performing the additional bias-correction and trend-adjustment steps, the preliminary ERA5 monthly totals are usually closer to the eventual NCEI values than the unadjusted ERA5 values. For example, the charts below show 2020 monthly CONUS population-weighted HDDs and CDDs from the unadjusted (blue), preliminary (red), and final (black) methods. The preliminary values are those obtained in near-realtime from ERA5, prior to the NCEI monthly update.





B. Realtime ECMWF Updates

To fill in the 5-day ERA5 lag, we use temperature grids from the 00 UTC and 12 UTC high-resolution ECMWF deterministic model forecasts. The calculation is exactly as described for the ERA5 data in Section 2(B), with the gridded daily ERA5 climatology serving as the model baseline. However, to remove ECMWF bias relative to ERA5, we also apply a running 30-day bias correction to the regional HDD and CDD values. The figure below shows a comparison of 2021 daily CONUS HDDs from ECMWF and (final scaled) ERA5.



C. Operational Realtime Update Schedule

The data availability described above necessitates the following schedule for daily updates:

- Initial update: the previous day's values are calculated from ECMWF data around 09 UTC each day
- Preliminary revision: daily values are revised after a 5-6 day lag when ERA5 data becomes available
- Final revision: the previous month's monthly and daily data is finalized when the NCEI climate division data updates on about the 5th of the month. Note that NCEI often performs small additional adjustments to the historical data in subsequent months and years, and we re-calculate the entire history with each monthly update.

4. Subseasonal and Seasonal Forecasts

The subseasonal and seasonal dynamical model forecast data is processed in essentially the same way as the ERA5 data to obtain daily regional weighted degree day forecasts. A key component of the process is that we use a long-term climatology of each dynamical model's temperature forecasts to remove the systematic difference (bias) between model temperatures and climate division temperatures. This is critical for two reasons:

(i) the long-range model forecast grids are relatively coarse (typically about 1 degree latitude/longitude), and very large temperature biases can occur in coastal regions or over complex terrain

(ii) model forecasts often exhibit systematic drift or lead-dependent bias that becomes significant beyond lead times of 10-15 days, when the ensemble mean signals are usually small. We use a lead-dependent bias correction to prevent the model drift from overwhelming the marginal long-lead model signals.

The dynamical model climatologies are obtained from retrospective forecasts provided by the modeling centers. For ECMWF, a 20-year set of retrospective forecasts is provided with each realtime subseasonal update, ensuring that the model climatology remains relevant to the latest model version. For GEFS, a static set of retrospective forecasts is available from 2000-2019, and subseasonal CFSv2 forecasts (both retrospective and realtime) are available since 2000. Seasonal forecast histories are also available for both ECMWF and CFSv2.

For each model, degree day calculations are performed separately for each ensemble member, and then the ensemble mean degree day forecast is obtained. The degree day forecast cannot be obtained directly from the ensemble mean temperature forecast, because the algorithm is non-linear; the differences can be substantial when the temperature range within the ensemble crosses 65°F.

Finally, it is worth noting that we aggregate the subseasonal CFSv2 ensemble members over a 2-day initialization window, because the CFSv2 daily subseasonal ensemble only includes 16 members. This is in contrast to the GEFS (31 members) and ECMWF (51 members). The small size of the CFSv2 ensemble means that it is more prone to “jump” from day to day, which is an undesirable feature for long-lead forecasts in which slowly-varying boundary conditions provide a significant part of the predictability. For the seasonal CFSv2 forecast, we aggregate members over a 5-day initialization window, because only 4 seasonal members are available each day.

5. Data Availability

The 2001-2020 daily history of regional degree day values is available at the following URL:

https://s2s.worldclimateservice.com/wcs/regional_degree_day_history_daily.csv

Realtime daily updates and subseasonal/seasonal forecasts are available in the World Climate Service Trading Markets product. Contact us for a trial today.

6. Conclusion

The World Climate Service has prepared a historical database of regional daily degree day data to support the development of models for predicting weather-based energy demand. The historical database also provides the foundation to support the subseasonal and seasonal HDD and CDD forecasts that are now available in the WCS Trading Markets product. By using consistent methods for both historical and forecast components, the WCS degree day forecasts have a robust historical context and provide customers with an information advantage for long-lead planning and decision processes.