David Sellers

Learning About TMY

BY DAVID SELLERS, P.E., MEMBER ASHRAE

Outdoor environmental conditions are major drivers for building heating and cooling loads. The assessment techniques I discussed in my two most recent columns use outdoor air temperature (OAT) as the independent variable. I have been interested in meteorology for most of my life and thus have learned much about climate data acquisition and reporting in real time. However, I was not as familiar with how the normalized (for example, typical meteorological year or TMY) data sets we use are developed. But thanks to a recent project, I was able to explore that topic and learned some surprising things. So, I decided to use this column to share some of the insights I've gained.

Some of the figures that follow are fairly "intense." The idea is to give you a general visualization of a concept by allowing you to generally compare things rather than read specific data, i.e., contrast the shape of different colored lines and shapes in the context of the discussion referencing the figure vs. actually reading the data. But if you want to look at them in detail, you will find higher resolutions versions at https://tinyurl.com/TMIAboutTMY.

Real-Time Climate Data vs. Normalized Climate Data

There are two different ways I look at weather data, depending on if I am trying to understand what happened in the past vs. trying to predict what might happen in the future.

Looking Backward

Frequently, during existing building commissioning or ongoing commissioning, we are diagnosing why a building system performed in a certain way. Since OAT can be a big driver behind system performance, we frequently consider what happened in that context.

I've found the National Oceanic and Atmospheric Administration's (NOAA) Online Weather Data (NOWData) to be a helpful way to target days I would like to look at for diagnostics, as illustrated in *Figure 1* and *Figure 2*.†

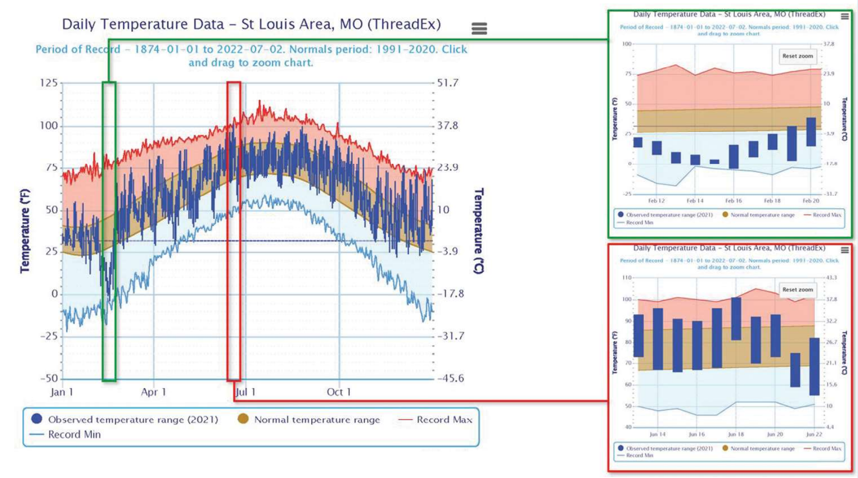
NOWData presents a picture of the climate for the location you select by plotting what happened on a given day against the normal range and the extremes on record. This makes it easy to spot a day when the systems were subjected to the extreme heating condition or the extreme cooling condition that could push them to (or beyond) the limits of their design envelope. Obtaining trend data from those days can be particularly insightful.

Even better, perhaps, is data from a day when the facility had to deal with an extreme swing in temperature, as illustrated in *Figure 2*. I was commissioning a surgery system on a day like that in the late 1980s, and the system really was put through its paces. We left the site the next morning feeling pretty happy because everything worked. But if the design had been less well thought out and the equipment less robust, it could

†Learn how to access this data here: https://tinyurl.com/NOWData

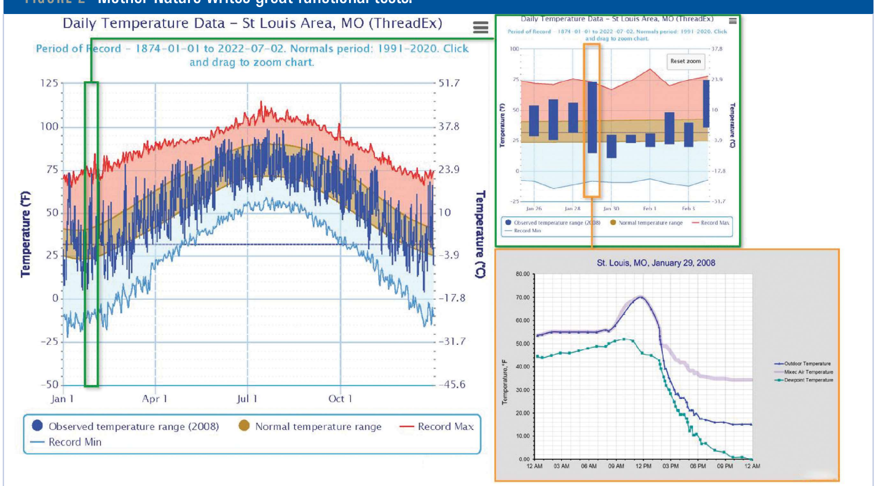
David Sellers, P.E., is senior engineer at Facility Dynamics Engineering's office in Portland, Ore.

FIGURE 1 Using NOWData to focus trend analysis.



Reviewing trend data for systems as they react to extreme days can provide a lot of insight regarding their peak load capability relative to the design requirements. As a frame of reference, the St. Louis 0.4% cooling design condition is 95.6°F DB/76.8°F WB. The 99.6% heating design condition is 4.1°F/50% RH.

FIGURE 2 Mother Nature writes great functional tests.



Days with large diurnal swings can reveal a lot about a system's ability to turn down (or the lack thereof). They can also provide a reasonable basis for a regression analysis of the system's performance.

have been a nightmare.

Data from a day or several days with a large diurnal swing is very helpful if you are doing a regression‡ to project what might happen over the course of a year. That is because even though you did not see a lot of hours at any given condition, you did see what happens over a wide range of conditions. That can be much more insightful than trend data collected over a day or week in which the OAT did not vary significantly.

While the NOWData can lead you to the days you want to consider for analysis, it is not granular enough to support the analysis. In other words, it shows the range for the day, but not what happened every hour.

However, NOWData is typically derived from Automated Surface Observing Systems (ASOS), and the Iowa State University provides a great resource for obtaining data from ASOS sites around the world: https://tinyurl.com/IowaStateASOS.

Looking Forward

Assuming accurate data and valid assessment techniques, the results you generate using ASOS and NOWData along with various HVAC relationships to project what happened on a given day at a given time are pretty firm—perhaps even facts. But if you develop a regression based on those facts and want to use it to predict what might happen in the future, then ASOS data may not be the best choice for the projections you make. More on this will follow.

The NOWData is still useful since, in addition to showing you what happened, it illustrates the norm and the extremes. But evidence suggests that the norm and extremes are probably shifting.

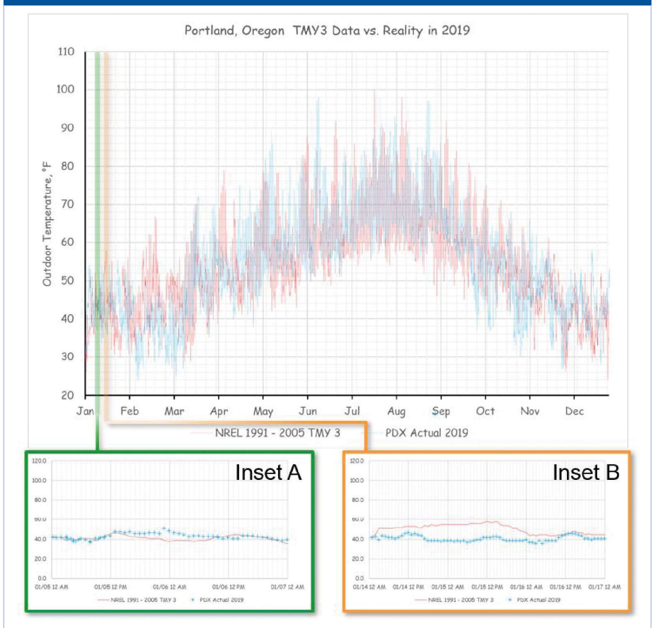
Still, if you have trend data to help you understand how your system worked on an extreme day now, it is likely useful for you in terms of understanding how it will react on an extreme day in the future and for devising strategies to mitigate the impact of the extremes.

For example, based on observations of this type, you might elect to:

- Provide tables in your direct digital control (DDC) graphics that give the operating team insight into the basic operating parameters for a facility's systems and

‡Regression: a scary sounding word that generally means "predicting what you think might happen based on what you know did happen." Information and examples here: https://tinyurl.com/DoingRegressions.

FIGURE 3 TMY 3 (normalized) data vs. reality for a year in Portland, Ore.



subsystems in real time.

- Develop logic that allows the operators to selectively adjust system and subsystem setpoints to mitigate the impact of extreme conditions based on experience and functional tests of the mitigation strategies. For instance:
- A hospitality facility could write logic that temporarily shifts cooling capacity priority from the guest rooms to the meeting rooms during the day, when the guests are likely in the meeting rooms and then back to the guest rooms when events have ended.
- A health-care facility could write logic that temporarily resets leaving air temperatures for systems serving non-patient care areas like administration and the lobby in favor of maintaining targeted leaving air conditions for the surgery, ICU and ER.

The logic sequences could be made accessible via buttons on the summary tables associated with the relevant systems. The specific parameters invoked by the logic can be based on functional tests that are executed during non-critical hours.

Bottom line, extremes will happen. The real questions:

- How prepared are you?
- How bad might it get?
- How often might it happen?

ASOS and NOWData can help you answer these questions so you can be proactive vs. reactive when the inevitable occurs.

What ASOS and NOWData are less adept at helping us understand is what the benefit of a proposed improvement might be, especially in terms of saving energy. The NOWData is simply not granular enough, and the ASOS data is very specific to the realities of a given day rather than the average conditions.

For instance, if we were to project energy savings for a proposed improvement based on the hourly data for a year that had more extreme events than normal, we might overstate the benefit. But if we use data for a year that had fewer extreme events than normal, we might understate the benefits.

That is where the normalized weather data files can provide benefit.

Normalized Climate Data vs. Extremes

Normalized climate data endeavors to looks at what happened on average. That seems reasonable if one does not want to over- or understate the benefit of making an improvement. For a given (unknown future) year, if you project using normalized data, then "sometimes you win, sometimes you lose (and sometimes the blues just gets a hold on you)," but on average you will be correct.

Common wisdom says that using normalized data "tends to eliminate or flatten out the extremes." But when I took a closer look, I decided I needed to add the word "kinda" to that statement, as can be seen from *Figure 3*.

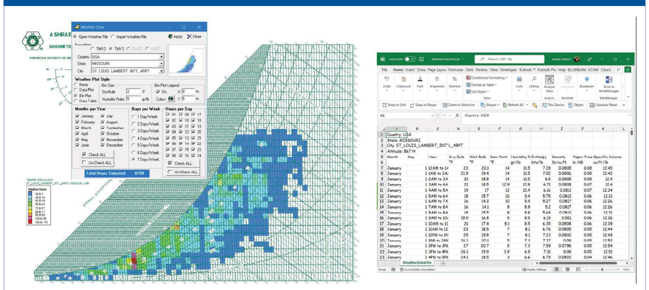
I say "kinda" because (to my way of thinking) you would expect normalized data to look more like the brown bands in *Figure 1* and *Figure 2*, i.e., the TMY (normalized) data would ride between the extremes. However, as can be seen from *Figure 3*, inset A, sometimes the TMY data does in fact seem to generally follow the actual conditions. But other times the TMY data seems to be the extreme (*Figure 3*, inset B).

Taking a Closer Look

While there have been numerous climate data normalization approaches used over the years,§ three seem the most common for use in spreadsheets and modeling programs. The bin data plot and export feature found in many electronic psychrometric charts (including ASHRAE's) use data from one of these sources, allowing you to get "a visual" of the climate on the chart or export

§Many are documented in footnotes II, II II, and ‡‡‡. I will touch on some here, but will focus on TMY data, which is very commonly applied.

FIGURE 4 ASHRAE Psych Chart Bin Data Plot and Export. This feature is found in the "Analysis" drop-down menu under "Open Weather Data File..."



the data for use in a spreadsheet (*Figure 4*).

Bin Data

In the olden days, when spreadsheets were literally sheets of paper that were manually filled out using a pencil to document the results of calculations performed with a slide rule or calculator, normalized data sets called bin data were often used for energy calculation (*Figure 5*).

This was a convenient way to accumulate data (bin-hours) that could be graphically displayed and analyzed in an analog manner.

To apply the data, you did a calculation for each row. For example, if I wanted to calculate the load on a 100% OA preheat coil for the highlighted bin in *Figure 5*, it might look something like *Figure 6*.

When I started using computers, I applied the same technique in electronic spreadsheets. But at some point, I realized that the "work" was in writing the relationships you needed into columns associated with each row. Once you had figured out the relationships, you could copy and paste them into the 8,760 rows in an hourly weather data file just as easily as the 22 to 30 rows in a typical bin file.

The granularity of the hourly data allows me to:

- Provide a schedule, or
- Use a seasonal or hourly flow profile, or
- Use a reset schedule.

This can then be applied to each hour of the year vs. prorating bin data to reflect a metric that spans the boundaries of a bin. As a result, my projections contain fewer assumptions and, thus, are more accurate.

TMY Data

TMY stands for typical meteorological year and has its roots in a data set produced by Sandia National Laboratories in 1978 for 248 locations using long-term weather and solar data from the 1952–1975 Solar Meteorological (SOLMET) database. This original

FIGURE 5 Typical bin data.

Source - NOAA Engineering Weather Manal
ST. LOUIS/LAMBERT MO WMO No. 724340
Dry-Bulb Temperature Hours For An Average Year (Sheet 5 of 5)
Period of Record = 1973 to 1996

Bin			Hours in Bin				Mean
Upper End, °F	Lower End, °F	Mid Range, F	00 - 08 Hours	09 - 16 Hours	17 - 24 Hours	Total Hours	Coincident Wetbulb
109	105	107	0	0	0	0	78.0
104	100	102	0	7	2	9	77.4
99	95	97	0	36	11	47	76.9
94	90	92	0	123	49	172	75.4
89	85	87	11	233	128	371	73.0
84	80	82	82	280	216	578	70.6
79	75	77	208	280	273	761	68.2
74	70	72	301	252	296	849	65.1
69	65	67	285	207	245	737	60.5
64	60	62	269	199	227	695	56.1
59	55	57	233	188	208	629	51.4
54	50	52	210	173	194	577	46.8
49	45	47	211	173	192	576	42.6
44	40	42	228	181	200	609	38.5
39	35	37	230	182	200	612	34.1
34	30	32	245	157	188	590	29.9
29	25	27	149	95	109	353	25.0
24	20	22	92	60	67	219	20.3
19	15	17	60	41	49	150	15.6
14	10	12	49	27	34	110	10.9
9	5	7	29	14	16	59	6.3
4	0	2	17	7	8	32	1.3
-1	-5	-3	7	3	3	13	-2.9
-6	-10	-8	5	2	2	9	-7.1
-11	-15	-13	2	1	1	4	-11.9
-16	-20	-18	1	0	0	1	-16.2

Bin data derives its name from the way the data is organized. In this example, the data is in "bins" that are 5°F by 8 hours. For example, the highlighted row contains three 8-hour bins that span the 30°F to 34°F temperature range. In a normalized year for St. Louis, based on data recorded between 1973 and 1996, there were:

- 245 hours in the temperature range between midnight and 8:59 a.m., and
- 157 hours in the temperature range between 9 a.m. and 3:59 p.m., and
- 188 hours in the temperature range between 4:00 p.m. and 11:59 p.m.

The mean coincident wet bulb when the dry-bulb temperature was in this range was 29.9°F.

FIGURE 6 Example bin calculation.#

Source - NOAA Engineering Weather Manal
ST. LOUIS/LAMBERT MO WMO No. 724340
Dry-Bulb Temperature Hours For An Average Year (Sheet 5 of 5)
Period of Record = 1973 to 1996

Bin			Hours in Bin				Mean
Upper End, °F	Lower End, °F	Mid Range, F	00 - 08 Hours	09 - 16 Hours	17 - 24 Hours	Total Hours	Coincident Wetbulb
34	30	32	245	157	188	590	29.9

Preheat Coil Load Calculation						
Flow Rate, cfm	Leaving Temp, °F	Q = 1.08 x cfm x Δt, °F	00 -08 Btu	09 - 16 Btu	17 - 24 Btu	TOTAL Btu
10,000	58	280,800	68,796,000	44,085,600	52,790,400	165,672,000

To develop the total energy consumption for this preheat coil for the normalized year, you would simply perform the calculations in the columns with blue headers for each bin and then add them up.

#More on bin data along with a source: https://tinyurl.com/BinData

TMY database was an effort to improve upon the Test Reference Year (TRY) database, which was one of the earliest building energy simulation databases.[||]

National Renewable Energy Laboratory (NREL) produced the TMY 2 and TMY 3 data sets by updating the Sandia data using the National Solar Radiation Database (NSRDB). Recently, gridded TMY data has become available that is based on satellite readings taken on a 4 km by 4 km (13,123 ft) grid from 1998–2017.

TMY2 vs. TMY3

The TMY3 data sets are hourly values of solar radiation and meteorological information for a one-year period based on records from 1991–2005. TMY2's are similar, but for the years 1961–1990.[**, ††] The TMY3 data set contains data for 1,020 locations vs. 239 for the TMY2 data set.

Gridded TMY Data

Gridded data is amazing, at least to me. You can click on any point in the colored area on the National Solar Radiation Database (https://nsrdb.nrel.gov/) and once you set up your data query, obtain a data set for the 4 km (13,123 ft) square that your selected point falls into. A similar website developed by the European Commission, Joint Research Centre (Photovoltaic Geographical Information System) is here: https://tinyurl.com/EuropeanPVData.

The data is derived from readings taken by satellites,[‡‡] which (I think) is the reason for a hard line with no data north of the arctic circle; the satellite viewing angle is skewed enough so that above that latitude the data is unacceptable for analysis purposes.

CWEC Data

CWEC stands for Canadian Weather Year for Energy Calculation and includes 492 data sets created from 30 years of CWEEDS (Canadian Weather Energy and Engineering Datasets) data.[§§] The data may not be totally in compliance with the TMY standard (more on that in a minute) but has been coordinated with ASHRAE.

IWEC Data

The ASHRAE IWEC 2 database contains weather files for 3,012 locations outside the United States and Canada.[###] The International Weather for Energy Calculations files are derived from Integrated Surface Hourly (or similar) weather data originally archived at the National Climatic Data Center based on periods of record for at least 12 years up to 25 years. This data also may not be fully compliant with the official TMY standard.

A Closer Look at TMY Data

If you look at the actual data sets, they all have different numbers of variables in different columns. So, while conceptually similar, they are not identical regarding format or the details of the normalization algorithms.

The ISO Standard Behind TMY Data

An ISO standard defines how one goes about putting a TMY data set together.[||||] There is a very significant aspect of this standard that I feel one needs to be aware of. To quote the TMY3 user's manual (emphasis is mine):

Because the TMY algorithm assigns priority to the solar radiation elements, *the selected months may or may not be typical for other elements. Cloud cover, which correlates well with solar radiation, is probably reasonably typical. Other elements are not related to the elements used for selection; consequently, their values may not be typical.*

Maximum and minimum temperature, dewpoint and wind data are given the least weight and solar data is heavily weighed. The means are daily means calculated from the hourly values for each month and year in the data set. A statistical analysis technique is used to assess all the months in the data set to determine the month

[||]Crawley, D. 1998. "Which weather data should you use for energy simulations of commercial buildings." *ASHRAE Transactions* 104(2):498–515.

[**]Source: "Users Manual for TMY3 Data Sets" from NREL, a great resource for understanding TMY data: https://tinyurl.com/NRELTMY3Manual

[††]Downloads here from the National Solar Radiation Database's Archived Data: https://tinyurl.com/TMYDataArchive

[‡‡]This paper, "Assembling Typical Meteorological Year Data Sets for Building Energy Performance Using Reanalysis and Satellite-Based Data," describes how this works: https://tinyurl.com/TMYFromSatellites

[§§]CWEC Engineering Climate Data Sets downloads here: https://tinyurl.com/CWECData

[###]Download the International Weather for Energy Calculations files here for a fee: https://tinyurl.com/IWECData

[||||]ISO 15927-4—Hygrothermal Performance of Buildings—Calculation and Presentation of Climatic Data, Part 4: Hourly Data for Assessing the Annual Energy Use for Heating and Cooling

FIGURE 7 Portland Data Comparison – Part 1.

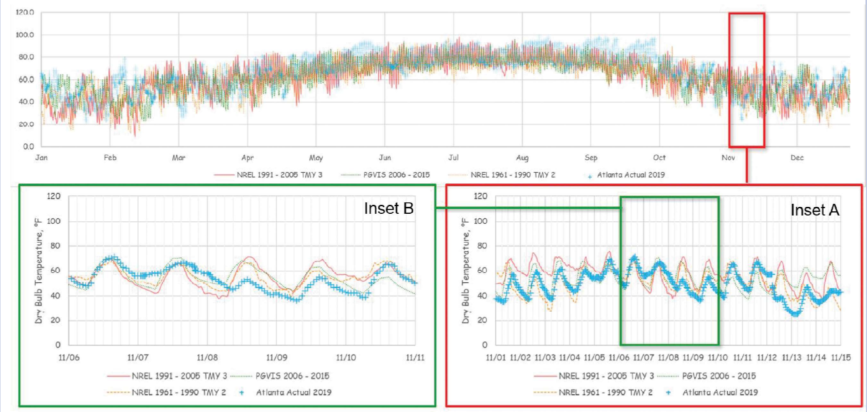
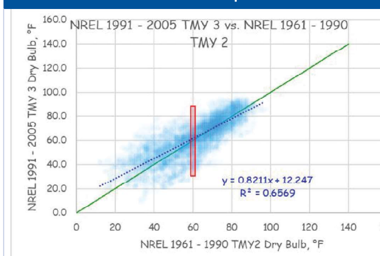
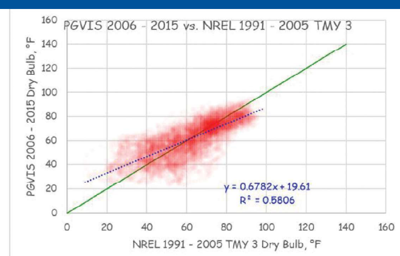
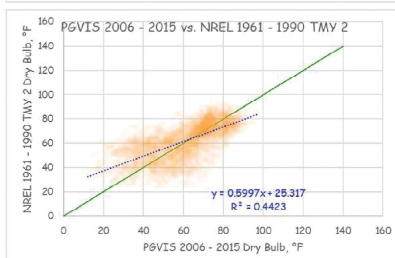


FIGURE 8 Portland Data Comparison – Part 2.







This is another way of contrasting the data sets.

If the data sets were identical, all of the dots would fall on the green straight line. In other words, for the upper left chart, if all dates and times when the NREL TMY 3 database had a temperature of 60°F the NREL TMY 2 database also had a temperature of 60°F, then all of the dots would be on the green line at the intersection of an *x* (horizontal axis) value of 60 with the *y* (vertical axis) value of 60.

But as you can see, there are times when, for instance, the TMY 2 data reported a value of 60 when the TMY 3 data would have reported value anywhere between about 37°F and about 84°F (the red highlighted area).

that is the most "typical."***

As a result of this, if you look at a TMY data file, you discover that the months are from different years. For instance, in the TMY3 file behind *Figure 3*, the data's from:

- 1976 (January and December);
- 1977 (February);
- 1980 (July and September);
- 1986 (November);
- 1988 (March);
- 1989 (May and June);
- 1996 (April);
- 2000 (October); and
- 2001 (August).

A smoothing algorithm is applied to the six hours at the end and beginning of the months to make the data seem more seamless.

Contrasting TMY Data Sets

All that made me curious, and I ended up comparing TMY2 with TMY3 data, satellite reanalysis TMY data and a year of actual data for five different locations. Portland's result is illustrated in *Figures 7* and *8*.†††

Before doing all the research, I expected the TMY data sets to lie on top of each other, while the real data set would match sometimes but deviate from the TMY data other times. Or perhaps, if the climate is warming, I would see the TMY 3 data consistently above the TMY 2 data. Instead, I observed no consistent pattern, which I realize now is likely due to weighting factors and the date range for the weather data used to build the files.

Conclusions

So, what does all this mean? My thoughts.

1. Most of the math we do for building systems is an *estimate*, not an *exactimate*, especially for existing buildings. Meaning, for the bin in the example in *Figure 6*, the total Btu for the bin is 165,672,000 *±10% or so* and should be represented as such. And it certainly is not 165,672,078.93721651.
2. If the climate is changing, then projecting what we think will happen in the future based on data we have collected from the past has its limitations. *An Assessment of Typical Weather Year Data Impacts vs. Multi-year Weather Data on Net-Zero Energy Building Simulations*‡‡‡ provides

***The paper "Assembling Typical Meteorological Year Data Sets for Building Energy Performance Using Reanalysis and Satellite-Based Data" provides a summary of the data processing (on page 5), which, as you probably suspect, involves using the Finkelstein–Schafer statistic: https://tinyurl.com/ReanalysisTMY

†††If you want a closer look at this data or any of the other data sets (Atlanta, Bremerton, Wash., Honolulu and Phoenix), spreadsheets are here: https://tinyurl.com/TMIAboutTMY

some thought-provoking insights into this in the context of residential building loads. Having said that:

- I think residential loads may be driven more by outside air (OA) than commercial, institutional and industrial building loads because generally there are not core areas (areas with no exposure to the envelope) in houses. And in terms of total energy use, I imagine lighting and equipment loads are generally lower in a home on a per square foot basis. Thus, the impact of the TMY data anomalies I have discussed may be more significant for residential building energy projections.
- On the other hand, for buildings with 100% OA systems, the load on the coils is an OA load. Bear in mind that air-handling systems that have an integrated economizer cycle can be 100% OA systems for a significant portion of their operating cycle. So, the central plant loads for many of our buildings are not as divorced from OA as you might think at first blush.

3. Typically, the spreadsheet style modeling many of us do with weather data is targeted at diagnostics or projecting relatively near-term cost benefit for existing building commissioning and ongoing commissioning projects. The systems and equipment targeted will have a life cycle significantly longer than the owner's targeted simple payback or return on investment window, which is often in the two-to-five-year range. Thus:

- The impact of long-term change on the accuracy of normalized data will be less significant than it would be for someone projecting the impact of my proposed modification for 20 or 30 years from now.
- Given that I am using a spreadsheet, once it is set up, it is possible to quickly use NOWdata to find some warmer and cooler years and then load ASOS data into the spreadsheet to see how more extreme conditions would impact my projections.

Bottom line, I believe what matters is that we recognize potential limitations of the data we use. In doing that, we can use a better resource if/when it becomes available. But lacking a better resource, by understanding the resource and its constraints, we can make informed decisions when we interpret and present results to clients. ■

‡‡‡https://tinyurl.com/TMYvsRealResidential