Physics- and Statistics-Based Quality Control of Weather Files

Florent Herbinger[1], Michaël Kummert[1]
[1]Polytechnique Montréal, Montréal, Canada

Abstract

Dynamic building energy simulation programs require accurate weather data to produce reliable energy consumption results. Traditionally, hourly weather files have been vetted using "physical" quality control procedures (QCPs), such as wind speeds rarely changing more than 15 m/s in an hour. However, weather data that proves to be suspect upon careful inspection may pass these physical QCPs unnoticed. The quality control strategy for yearly hour weather files proposed in this paper supplements current physical QCPs in the literature with statistical QCPs based on historical local weather data, pulled from ASHRAE's Weather Data Viewer (2017). These statistical QCPs use historical data going back up to 30 years at the weather station to flag hourly magnitudes, hourly steps, daily profiles, or monthly averages of meteorological variables that fall outside what is historically expected at that weather station. This enhanced methodology successfully catches suspect weather data that would otherwise pass traditional QCPs.

Introduction

Dynamic building energy simulation (BES) programs require accurate weather data to produce reliable energy consumption results. Solar irradiance, air temperature, humidity, and wind speed are some of the meteorological variables that drive the boundary conditions in building energy models. These models can even fail completely if the weather data are not consistent or "clean"; for example, solar irradiation at night or a wet bulb air temperature above the dry bulb temperature (Barnaby & Crawley, 2011). Producing reliable BES results is important for modelers to not only predict the changes in carbon emissions from energy retrofits but also to test the building systems' reliability during extreme weather events. The European heat wave in 2003 was said to have killed more than 70,000 people, mainly the elderly and young, due to poorly conditioned spaces (Fouillet et al., 2006; Robine et al., 2008). The return period for these heat waves is likely to decrease from every 250 years, to every 50 or even 35 years owing to climate change (Charpentier, 2011).

Historically, extensive weather observation programs have served the needs of various nationally significant activities such as agriculture, weather forecasting, and aviation, but little consideration has gone into developing data collection procedures for engineering applications such as BES (Barnaby & Crawley, 2011). The largest repercussion of this neglect is the overwhelming predominance of hourly sampled weather data, which are collected in yearly 8760-hour or 8784-hour (for leap years) weather files (Hensen, 1999). While outside air temperature rarely changes more than a few degrees hour-to-hour, solar irradiance and wind speed can vary by several factors over the span of mere minutes. In a 1997 study, Janak (1997) developed a coupling method between BES and global illuminance simulation for applications such as artificial lighting; he observed that using 5-minute timesteps over hourly timesteps resulted in prediction differences of over 40%. However, considering that hourly sampled weather data is still by far the most readily available, it is imperative that they be as error-free and consistent as possible as to not throw off energy predictions even more.

There are four different kinds of errors that can be introduced in a weather file:

Random errors: an instrument can only give an approximation of the true value of a natural property. Moreover, variations in other variables can influence the measured value. As the law of large numbers governs the multitude of independent factors at play, random errors can be considered to be normally distributed around zero (Gandin, 1988).

Systematic errors: calibration errors or long-term drift in the sensors as well as erroneous data processing can introduce systematic errors. Since they are usually persistent in time and asymmetrically distributed around zero, systematic errors add a bias to the measured variable (Gandin, 1988).

Micrometeorological errors: small-scale perturbations surrounding the weather stations, such as urban heat islands or random turbulence, can result in the non-representativeness of observations when considering larger meteorological scales. Micrometeorological errors can be systematic or random in nature (Steinacker et al., 2000).

Gross errors: the malfunctioning of instruments and mistakes during data transmission, reception, and processing can cause large or gross errors in the magnitude of variable values. Gross errors strongly affect analyses and forecasts. They do not follow the Gaussian distribution law due to their rare occurrences and large magnitudes (Gandin, 1988).

The World Meteorological Organization (WMO) has produced several lengthy reports that detail the quality control procedures (QCPs) to vet for these four different kinds of errors in weather data and possibly correct them (1993). The checks for the physical elements (*i.e.* the meteorological variables) cover the following:

Gross error checks: ensure that observed values are below or above fixed limits (*e.g.* wind speed is below 60 m/s in the winter between latitudes of 45 °S and 45 °N).

Internal consistency checks: the different observed values are consistent (*e.g.* cloud cover (h) and visibility (VV) values are considered suspect when $0 \leq h \leq 1$ and ($70 \leq VV \leq 89$ or $98 \leq VV \leq 99$).

Time consistency checks: the differences between values at the observed timestamp and previous timestamp are consistent (*e.g.* the dry-bulb air temperature changes less than 4 °C in one hour or 25 °C in 12 hours).

Space/time consistency checks: observed values between geographically close weather stations that are close in time are consistent (*i.e.* each pair of spatially neighbouring observations agree, making them both correct or erroneous).

However, the physical elements related to the sun are excluded in the WMO report. ENDORSE (ENergy DOwnstReam SErvices) (2013) is a project from the European Global Monitoring for Environment and Security (GMES) aimed to promote energy services from the sun, wind, and biomass as well as electricity grid management and building engineering. They produced a report detailing several QCPs for hourly solar irradiance presented in the Quality Control Procedures section of the paper.

However, neither the WMO nor ENDORSE consider local historical data in their quality control strategies—the WMO just considers large areas such as latitudes from 45 °S and 45 °N and 45 °N to 90 °N, while ENDORSE considers no local regions at all. Incorporating statistical QCPs based on historical local weather data from the past, say 30 years, could catch errors that would otherwise pass through physical QCPs. For example, solar irradiance from a yearly weather file may pass all gross error and consistency checks, but when comparing the global horizontal irradiation (irradiance integrated over time) for the month of November to historical monthly averages and standard deviations, we might find that it is abnormally high, falling in the upper 99th percentile, or greater than the average plus ~2.326 standard deviations. Taking the dry bulb temperature as another example, its daily profile for a week in April may never fall above or below gross error checks and never vary more than a couple of degrees in an hour, thus passing all physical QCPs, but the profile could be consistently shifted 6 hours forward when comparing it to the temperatures most observed for each hour of the month of April over the past 30 years.

The quality control strategy for yearly hour weather files proposed in this project supplements current physical quality control procedures in the literature with statistical QCPs based on historical local weather data, pulled from ASHRAE's (American Society of Heating, Refrigeration, and Air Conditioning Engineers) Weather Data Viewer (2017). Close to 30 years of vetted weather data for 8118 weather stations is easily fetchable and analysable through an integrated workflow developed in Python, which takes a weather file as input and produces a report with flags for erroneous or suspect data points. A visual report with graphs for the various meteorological variables and flagged values will also be produced in future work.

Methodology

The following methodology developed in a Python package checks the validity of data points in a yearly weather file and offers a degree of quality control, while flagging erroneous, rare, and suspect data points. The color-coding of the text refers to the colors used in the QCPs below.

The data points of the weather file, which are the time-series of meteorological variables such as direct and diffuse solar irradiation, dry and wet bulb temperatures, and wind speed and direction are passed through four quality control procedures based on different temporal scales (depending on the possible physical and statistical checks as well as the historical data available):

Hourly value quality control: check for errors as well as rare and suspect data points in the magnitude of the hourly data points (*i.e.* extreme values)

Hourly step quality control: check for exceedingly rare hourly steps in the data (*e.g.* temperature difference of more than 8°C in one hour)

Daily profile quality control: check that the observed values fall during a typical time of the day (*e.g.* high air temperatures occur usually between 12:00 and 16:00 solar time based on 30 years of data)

Monthly averages quality control: check for rare months in the data (*e.g.* monthly average precipitation exceeds maximum observed in a 30-year timespan)

The QCPs above are further subdivided into physical and statistical checks, depending on the meteorological variable— the physical checks are taken from the literature, while the statistical checks are newly developed in this paper. The statistical checks are based on weather data points falling below the 1st percentile or above the 99th percentile of the 30-year distribution of data at the local weather station. These bounds are somewhat arbitrary but are still a good starting point to flag outliers.

Physical and statistical weather checks:

Impossible physics checks (erroneous): the laws of physics don't allow such values (*e.g.* solar irradiation at night)

Extreme weather checks (rare)**:** exceeds values very rarely observed across the globe (*i.e.* extreme weather events)

Historical weather checks (suspect)**:** falls in the lower 1st percentile or upper 99th percentile based on histograms or the average and standard deviation of historical local weather data

Note: the solar irradiation checks assume typical weather station environments with little or no surface reflections, and it is assumed that small irradiation measurement errors during the night (*e.g.*, $-5 \leq G \leq 5\,\frac{\mathrm{W}}{\mathrm{m}^2}$) have been removed.

Quality Control Procedures

1. Global Horizontal Irradiance (G)

G_{on} = normal extraterrestrial solar irradiance
G_o = horizontal extraterrestrial solar irradiance
θ_z = solar zenith angle

Hourly value quality control

Impossible physics [1,2,4] — (flagged as erroneous)

$G > \min(1.2G_{on},\ 1.5G_{on}\cos(\theta_z)^{1.2} + 100)$
$G \neq 0$ during nighttime

Extreme weather [2,3,4] — (flagged as rare)

$G > 1.2G_{on}\cos(\theta_z)^{1.2} + 50$
$G < 0.03G_o$

Monthly averages quality control

Historical weather — (flagged as suspect)

$G > G(99th\,\%)$
$G < G(1st\,\%)$

2. Direct Normal Irradiance (G_{bn})

Hourly value quality control

Impossible physics [2] — (flagged as erroneous)

$G_{bn} > G_{on}$
$G_{bn} < 0$
$G_{bn} \neq 0$ during nighttime

Extreme weather [2] — (flagged as rare)

$G_{bn} > 0.95G_{on}\cos(\theta_z)^{0.2} + 10$

3. Diffuse Horizontal Irradiance (G_d)

Hourly value quality control

Impossible physics [1,2,3,4] — (flagged as erroneous)

$G_d > \min(0.8G_{on},\ 0.95G_{on}\,cos(\theta_z)^{1.2} + 50)$
$G_d \neq 0$ during nighttime

Extreme weather [1,2] — (flagged as rare)

$G_d > 0.75G_{on}\cos(\theta_z)^{1.2} + 30$
$G_d < 0.03G_o$

4. Dry Bulb Temperature (T_{db})

Hourly value quality control

Impossible physics [7,8] — (flagged as erroneous)

$T_{db} > 60\ ℃$
$T_{db} < -90\ ℃$

Extreme weather [7,8] — (flagged as rare)

$T_{db} > 50\ ℃$
$T_{db} < -80\ ℃$

Historical weather — (flagged as suspect)

$T_{db} > T_{db}(99th\,\%)$
$T_{db} < T_{db}(1st\,\%)$

Hourly step quality control

Impossible physics [5,6] — (flagged as erroneous)

$T_{db,i+1} - T_{db,i} > 8\ ℃$

Daily profile quality control

Historical weather — (flagged as suspect)

$h(T_{db}) > h(T_{db},\ 99th\,\%)$
$h(T_{db}) < h(T_{db},\ 1st\,\%)$

5. Relative Humidity (Hum)

Hourly value quality control

Impossible physics [8] — (flagged as erroneous)

$Hum > 100\%$
$Hum < 0\%$

Historical weather — (flagged as suspect)

$Hum > Hum(99th\,\%)$
$Hum < Hum(1st\,\%)$

Hourly step quality control

Impossible physics [6] — (flagged as erroneous)

$Hum_{i+1} - Hum_i > 30\ \%$

6. Dewpoint Temperature (T_{dp})

Hourly value quality control

Impossible physics — (flagged as erroneous)

$T_{dp} > T_{db}$

Historical weather — (flagged as suspect)

$T_{dp} > T_{dp}(99th\,\%)$
$T_{dp} < T_{dp}(1st\,\%)$

7. Wind Speed (u_w)

Hourly value quality control

Historical weather — (flagged as suspect)

$u_w > u_w(99th\,\%)$

Hourly step quality control

Extreme weather [5,6] — (flagged as rare)

$u_{w_{i+1}} - u_{w_i} > 15\ m\ s^{-1}$

8. Precipitation Depth (d_{prec})

Monthly averages quality control

Historical weather — (flagged as suspect)

$d_{prec} > d_{prec}(99th\,\%)$
$d_{prec} < d_{prec}(1st\,\%)$

[1] (Geiger et al., 2002)
[2] (Long & Dutton, 2010)
[3] (de Miguel et al., 2001)
[4] (Muneer & Fairooz, 2002)
[5] (National Oceanic and Atmospheric Administration (NOAA), n.d.)
[6] (Vejen et al., 2002)
[7] (World Meteorological Organization (WMO), 1993)
[8] (Zahumenskỳ, 2004)

The originality of the methodology resides in the statistical checks that compare the weather file's hourly values, daily profiles, and monthly averages to past weather data. For comparative purposes, we will use three weather files for Montreal, Canada:

- A vetted typical meteorological year (TMY) file;
- A file that has been artificially tampered with to have improbable/suspect values; and
- A file for the year 2015, which is known to have strange solar irradiation.

We can compare hourly value and daily profile checks of the tampered file with the typical meteorological year, since a TMY captures both average and improbable weather in a location while hopefully remaining realistic, but the tampered file has improbable weather that is demonstrably false. Through the graphs that are produced by the Python workflow, the user will see that the suspect values that are flagged in the TMY appear correct, while those in the tampered file appear erroneous. Similarly, we can compare monthly average checks for solar irradiation between the 2015 weather file and the TMY. This comparison will be useful to demonstrate the strength of the monthly solar irradiation checks, since as will be shown, the irradiation in both the 2015 file and the TMY pass the physical irradiation checks but the 2015 file fails the statistical check.

Hourly Historical Weather Checks

For the hourly values quality control procedure, we have access to histograms from ASHRAE's Weather Data Viewer for 8118 WMO stations around the world. The histograms have bins of meteorological variable values (*e.g.* dry bulb temperature from 20 to 20.5 °C) with observed frequencies (*e.g.* 3000 counts) over the past ~30 years. The following histograms were used in the quality control methodology: dry bulb temperature, wet bulb temperature, dewpoint temperature, wind speed, and dry bulb / time of day.

The last histogram is presented as frequencies with crossed bins of dry bulb temperature and time of day (*e.g.* counts when the dry bulb temperature was between 20 to 21 °C between 9:00 and 10:00 in a particular month).

It will be interesting to study how climate change impacts weather trends and this quality control methodology. Past weather data will surely need to be restricted to more and more recent years when vetting future weather files; 30 years will most likely be too far back of a time span.

Results and Discussion

Figure 1 and 2 show the histogram for the dewpoint temperature during the month of March. The corresponding March 22 daily profile with the flagged hourly values of dewpoint temperature is shown for the TMY in Figure 1 and for the tampered year in Figure 2. The red lines represent the dewpoint temperature threshold where 1% of observed values fall below the left line or above the right line (*i.e.* in the lower 1st and upper 99th percentiles, respectively). We see in the morning between 6:00 and 9:00 of March 22 in the TMY that the dewpoint temperatures were very improbable, with less than 1% of all observed dewpoint temperatures over the past 30 years being as low. In Figure 2, between 13:00 and 14:00 in the tampered year, the dewpoint temperatures were also very improbable, with less than 1% of recorded values being as high. However, the user can quickly discern that the dewpoint temperatures at 13:00 and 14:00 are most likely erroneous due to the sharp spike in the profile, but the dewpoint temperatures between 6:00 and 9:00 appeared to just be part of a very cold and dry day. Viewing the flagged hourly values in context with neighbouring values is important to help make a final judgement on the validity of the hourly data points.

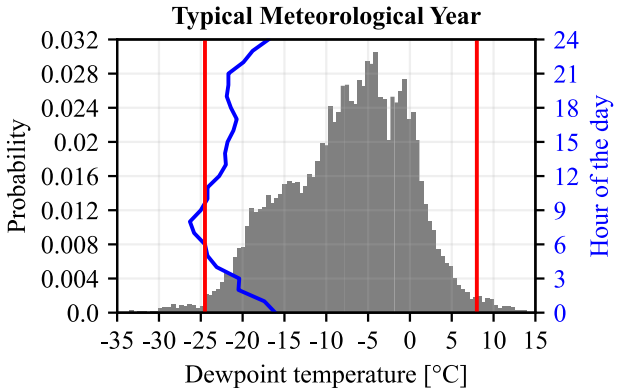

Figure 1: Histogram of all observed dewpoint temperatures in the month of March and the March 22 daily profile with the suspect dewpoint temperatures in the TMY file

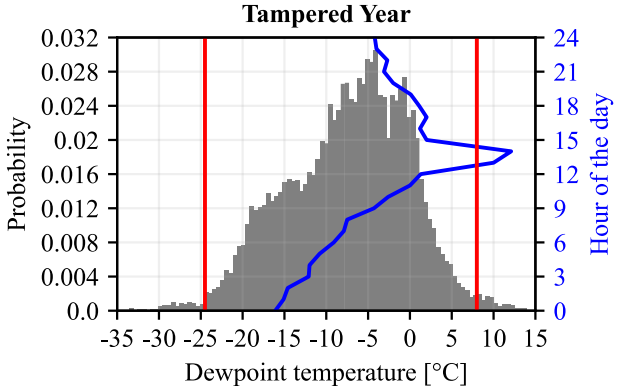

Figure 2: Histogram of all observed dewpoint temperatures in the month of March and the March 22 daily profile with the suspect dewpoint temperatures in the tampered file

Daily Historical Weather Checks

For the daily profile QCP, unfortunately we only have access to historical data for the dry bulb temperature. This data also takes the form of a histogram with counts for temperature bins for each of the 24 hours of the day. The Python algorithm flags days where six or more hours fall in either the lower 1st percentile or upper 99th percentile. In the typical meteorological year, June 6 was flagged as such a day. The following two Figures show the histogram for each hour of the month of June along with the suspect daily dry bulb temperature profile of June 6. In Figure 3, we realise that although in the middle of the night the dry bulb temperatures are very low, falling in the lower 1st percentile, the overall shape of the profile for the TMY is congruent with the histograms: temperatures falling in the night and rising in the day. However, in the tampered year in Figure 4, the daily profile was artificially shifted forward visually (or backwards in time if one prefers), and it is evident that this expected falling and rising pattern is not respected. Therefore, the user would most likely flag this day as erroneous.

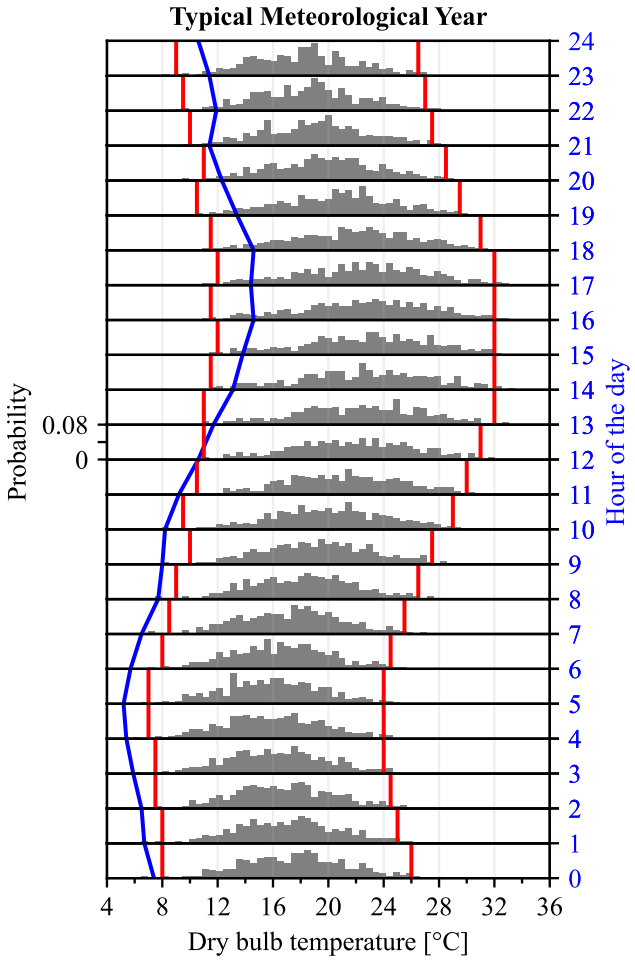

Figure 3: Hourly histograms of the dry bulb temperature for the month of June and June 6 daily profile of the typical meteorological year

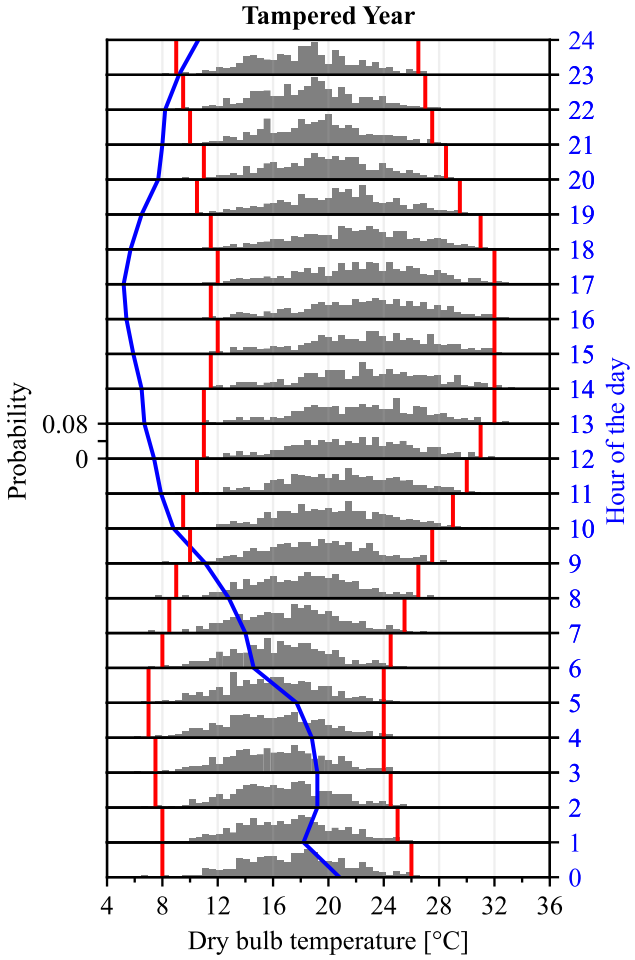

Figure 4: Hourly histograms of the dry bulb temperature for the month of June and June 6 daily profile that is visually shifted forward (or backwards in time) in the tampered year file

Monthly Historical Weather Checks

The final demonstration of the methodology addresses the monthly averages for the global horizontal irradiation. The monthly averages in ASHRAE's Weather Data Viewer are calculated as the monthly average of the daily integrated solar irradiance. The monthly averages for the typical meteorological weather file are shown in Figure 5 and the averages for the 2015 weather file with strange irradiation is shown in Figure 6. It should be noted that while both files use weather stations located in Montreal, the TMY uses the downtown McTavish station, while the 2015 weather file uses the station located at the Pierre Elliott Trudeau airport; however, the different weather stations are not pertinent to this comparison.

Since no detailed histograms are available for solar irradiation in ASHRAE's weather viewer, the 1st and 99th percentiles had to be calculated this time with the mean and standard deviation assuming a normal probability distribution of the monthly irradiation averages. In other words, the 1st percentile threshold is calculated by mean –

z score ∗ standard deviations, and the 99th percentile threshold with mean + z score ∗ standard deviations. The z score is approximately equal to 2.32635 based on a two-sided z score table.

The irradiation in the 2015 weather file (and the TMY for that matter) passes all the physical irradiation hourly value checks, so one might be led to believe that the irradiation is fine; however, when we look at monthly averages over the past ~30 years, we discover that the hourly global horizontal irradiation is most likely too high, as January, February, April, May and October have irradiation averages in the upper 99th percentile. This feature of the irradiation remained hidden from the physical checks. On the other hand, despite having periods of extreme weather, the vetted TMY passed all monthly average historical checks, thereby showing an example of how the methodology can separate good from bad files.

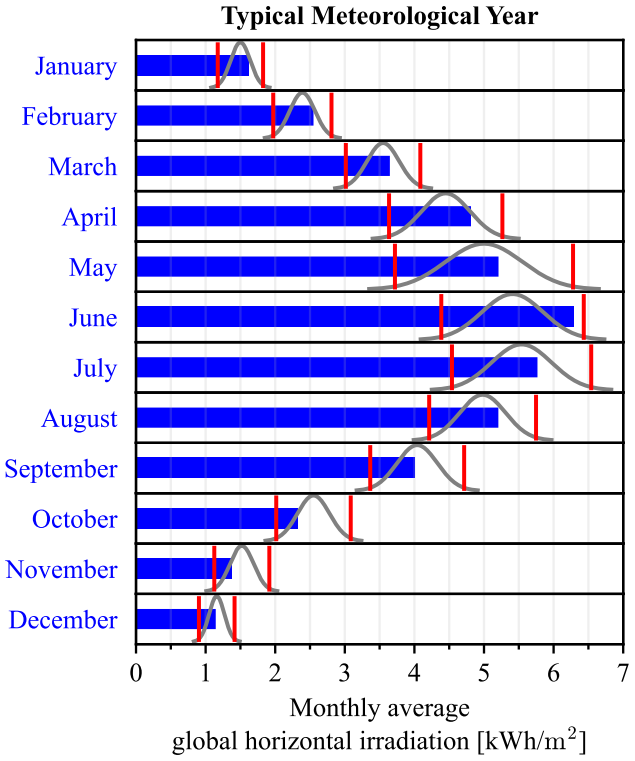

Figure 5: Assumed probably distribution functions with 1st and 99th percentiles in red bars along with monthly average global horizontal irradiation for the typical meteorological year

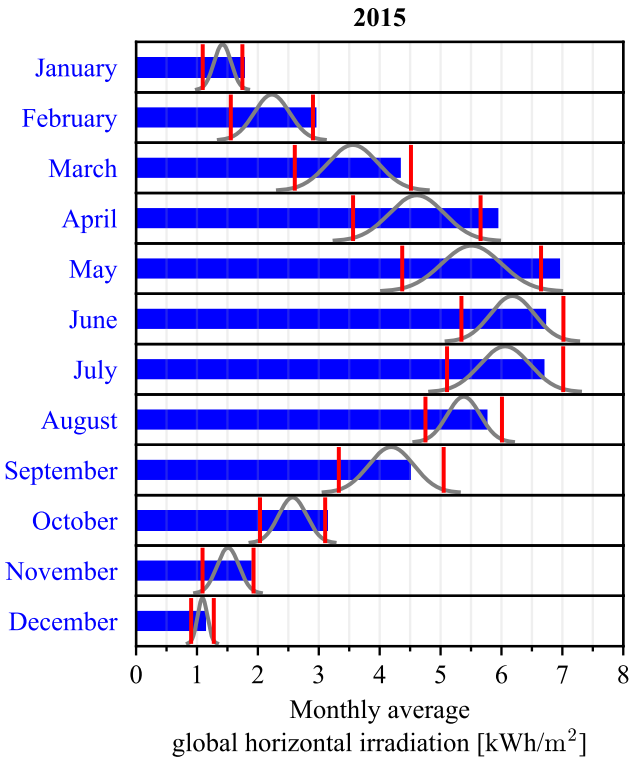

Figure 6: Assumed probably distribution functions with 1st and 99th percentiles in red bars along with monthly average global horizontal irradiation for the 2015 weather file

Upon futher inspection of the daily average irradiation profiles for each month of the year in Figure 7, we notice that the global horizontal irradiation for the 2015 file (in orange) is noticeably higher than the global horizontal irrdiation in the TMY file. When inspecting the direct and diffuse irradiation, we find the culprit. In Figure 8, the direct normal radiation appears to be much too high for the 2015 file compared to the TMY file. This manual investigation was first instigated by an automated quality control procedure in the Python workflow, and future work could allow the QCPs to even zero-in on the erroneous direct normal radiation directly.

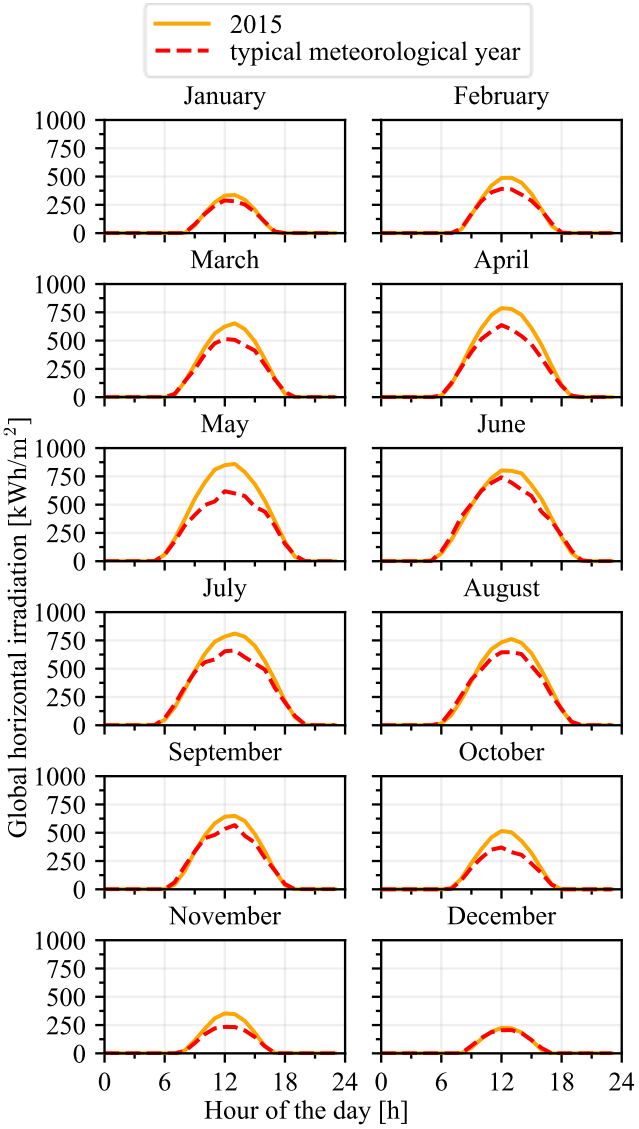

Figure 7: Daily average profiles of the global horizontal radiation for each month of the year in the 2015 and TMY files

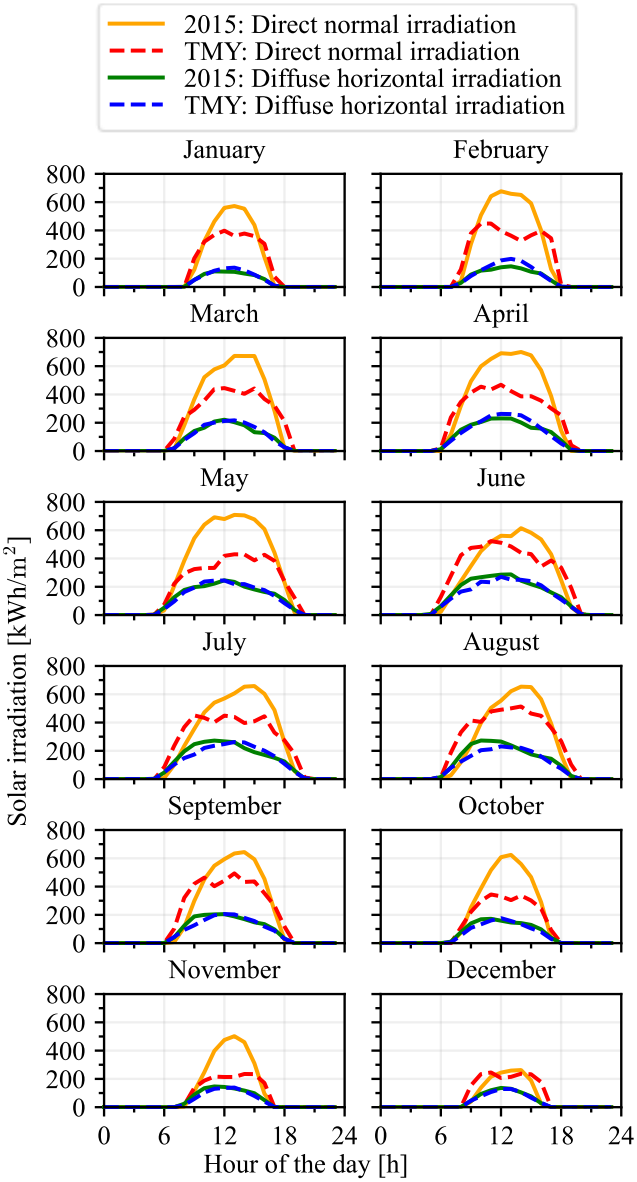

Figure 8: Daily average profiles of the direct normal radiation and diffuse horizontal irradiation for each month of the year in the 2015 and TMY files

Conclusion

Using accurate and reliable weather files for building energy simulations is primordial as the reported meteorological values in the weather files have large impacts on the outputs of the simulations and can lead to false conclusions by the building energy modeler. An original methodology for assessing the quality of hourly weather files was developed in Python that supplements the physical checks typically used in the industry—such as the dry bulb temperature cannot be above 60 °C—with statistical checks based on ASHRAE's Weather Data Viewer, holding nearly 30 years of vetted weather data for 8118 weather stations around the world. The methodology assesses the quality of the weather file through four scales of quality control procedures

(QCPs): hourly values, hourly steps, daily profiles, and monthly averages. In each of these QCPs, three different checks can be applied based on physics and the statistical data available: impossible physics, extreme weather, and historical weather. The methodology flags data points in the weather file for further inspection by the user. Particular attention was paid to the historical weather checks in this paper as they do not appear to be implemented in the industry at the same scale, or at all, compared to physical checks. This paper has shown that these historical weather checks can offer crucial supplementary information to building energy modelers assessing whether they can trust the weather files they input into their simulation programs.

Future Work

The authors will provide a free web-based tool that will allow users to upload their weather files to perform the quality control checks. For now, the python package that implements the methodology is publicly available at https://github.com/Florent-H/qcweather.

Several aspects of the methodology could still be improved however. Including monthly average historical weather checks for the direct normal and diffuse horizontal irradiation would be a useful addition; however, determining how the monthly averages for these irradiation components relate to the monthly average global horizontal irradiation is a challenge. Perhaps using the monthly clear sky optical depth and clear sky noon irradiance on the 21st day of the month, also available in the Weather Data Viewer, would be an option. Another improvement could be to implement the full range of World Meteorological Organization gross error checks as well internal, time, and space / time consistency checks to allow the Python algorithms to flag even more erroneous data points. Lastly, acquiring another easily searchable source of historical weather data could provide more statistical checks applied to more time scales, thereby providing additional quality assurance for weather files.

References

ASHRAE. (2017). *Weather Data Center*. https://www.ashrae.org/technical-resources/bookstore/weather-data-center

Barnaby, C. S., & Crawley, D. B. (2011). Weather data for building performance simulation. *Building Performance Simulation for Design and Operation*, 37–55.

Charpentier, A. (2011). On the return period of the 2003 heat wave. *Climatic Change*, *109*(3–4), 245–260.

de Miguel, A., Bilbao, J., Aguiar, R., Kambezidis, H., & Negro, E. (2001). Diffuse solar irradiation model evaluation in the North Mediterranean Belt area. *Solar Energy*, *70*(2), 143–153.

ENDORSE. (2013). *ENDORSE FP7 project | ENDORSE (ENergy DOwnstReam SErvices)*. http://www.endorse-fp7.eu/

Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneuc-Jouyaux, C., Clavel, J., Jougla, E., & Hémon, D. (2006). Excess mortality related to the August 2003 heat wave in France. *International Archives of Occupational and Environmental Health*, *80*(1), 16–24.

Gandin, L. S. (1988). Complex quality control of meteorological observations. *Monthly Weather Review*, *116*(5), 1137–1156.

Geiger, M., Diabaté, L., Ménard, L., & Wald, L. (2002). A web service for controlling the quality of measurements of global solar irradiation. *Solar Energy*, *73*(6), 475–480.

Hensen, J. L. (1999). Simulation of building energy and indoor environmental quality-some weather data issues. *Proc. Int. Workshop on Climate Data and Their Applications in Engineering*, 4–6.

Janak, M. (1997). Coupling building energy and lighting simulation. *Proc. Building Simulation*, *2*, 313–319.

Long, C. N., & Dutton, E. G. (2010). *BSRN Global Network recommended QC tests, V2. X*.

Muneer, T., & Fairooz, F. (2002). *Quality control of solar radiation and sunshine measurements–lessons learnt from processing worldwide databases, Build Serv. Eng. Res. T., 23, 151–166.*

National Oceanic and Atmospheric Administration (NOAA). (n.d.). *Handbook of Automated Data Quality Control Checks and Procedures* (p. 78).

Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P., & Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies*, *331*(2), 171–178.

Steinacker, R., Häberli, C., & Pöttschacher, W. (2000). A transparent method for the analysis and quality evaluation of irregularly distributed and noisy observational data. *Monthly Weather Review*, *128*(7), 2303–2316.

Vejen, F., Jacobsson, C., Fredriksson, U., Moe, M., Andresen, L., Hellsten, E., Rissanen, P., Pálsdóttir, Þ., & Arason, Þ. (2002). Quality Control of Meteorological Observations-Automatic Methods Used in the Nordic Countries. *Climate Report*, *8*.

World Meteorological Organization (WMO). (1993). *Guide on the Global Data-Processing System (GDPS)*. WMO.

Zahumenskỳ, I. (2004). Guidelines on quality control procedures for data from automatic weather stations, World Meteorol. *Organ. Switz*, *955*, 2–6.