Utilizing Historical Hourly Observations to Analyze Climatological Air Temperature Distributions

Logan K. McLaurin

North Carolina State University

04/27/2023

Introduction

In recent decades, compelling scientific evidence for anthropogenic climate change has been observed. Key publications such as the Intergovernmental Panel on Climate Change (IPCC) 2014 Synthesis Report and the Fourth National Climate Assessment thoroughly present background information regarding these shifts in the climate system by drawing on historical weather data (IPCC et al., 2014; Wuebbles et al., 2017) . Typically, these and similar studies present their findings through the lens of illustrating changes in averages, daily highs, and daily lows that are derived from daily observations (Guo et al., 2021). The Fourth National Climate Assessment for instance highlights variations in annual average temperatures, seasonal average air temperatures, and daily streaks of warm and cold spells (Wuebbles et al., 2017).

While daily temperature metrics provide significant insights, they fail to capture the full story. The biological effects of weather and climate changes have been shown to be much better understood by analyzing hourly changes directly as opposed to average values (Sheldon et al., 2016; Camacho et al., 2015). People, plants, animals, and buildings all experience air temperature stresses on an hour by hour basis. For instance, a high temperature of 95°F maintained over six hours during a day will have a markedly different effect than if that high temperature was recorded for just one hour. Thus, the societal significance of analyzing air temperatures on a finer temporal resolution such as hourly data cannot be overstated. This is particularly relevant with regards to the heat stress impacts on the general public as the occurrence of potentially fatal heatwaves is likely to rise under most climate change scenarios and is expected to be associated with reductions in work capacity and labor productivity (Matthews et al., 2017; Kjellstrom et al., 2018). Local considerations prove to be critically important to consider as the urban heat island effect has shown to compromise overall health, increasing the incidence of heat-related illnesses and mortality rates in cities (Tong et al., 2021). The specific location dependency of climate change impacts also pertain to the agricultural and cattle industries as there are many unknowns with regards to how these industries will be harmed or influenced in the long term (Herbut et al., 2019; Zhu et al., 2021). Currently, it is crucial to gather a broader range of environmental data to assess the capacity of these industries to adjust to these changes (Herbut et al., 2019; Siebert, Ewert, 2014). Hourly temperature data analysis is also critical for the energy sector, helping to address regional differences in energy demand (Bloomfield et al., 2022; Amato et al., 2005). This approach is valuable not only for adapting to long-term trends but also for managing abrupt fluctuations in energy requirements (Perera et al., 2020). A last but crucial reason for using historical hourly data is to enhance the development of effective scientific communication strategies to engage the public about climate change. The human mind and societal structures have inherent limitations in processing the impacts of climate change over long periods (Pahl et al., 2014). Presenting air temperature changes as long-term averages may not be inherently intuitive. To overcome the perceived remoteness of climate change, scientific communicators should employ innovative storytelling techniques, reframe the issues, and ensure they resonate with the actual experiences of their audience (Pahl et al., 2014; Markowitz, Guckian, 2018).

For this report, an application of using historical hourly air temperature observations is explored in the context of building and analyzing air temperature hours climatologies over various locations in the Continental United States (CONUS) and southern Canada. Conceptually, for a single location the number of hours of recorded air temperature at each individual degree Fahrenheit can be compiled over the course of the past few decades. This would be expected to produce a climatological distribution of air temperature hours with seasonal concentrations of air temperature values depending on the specific climate of the location. Utilizing Conrad's Continentality Index, climates of high continentality or higher temperature variability, dynamic seasons, and a large diurnal range correspond to central North America at higher latitudes while climates of high oceanity or lower variability and moderated seasons and diurnal range correspond to coastal North America and lower latitudes (Stonevicius et al., 2018). Thus, it is believed continental climates would feature bimodal distributions of air temperature hours, highlighting distinct peaks in the number of air temperature hours measured during winter and during the summer. Oceanic climates, on the other hand, would likely feature unimodal distributions of air temperature hours with a single, prominent peak in the number of air temperature hours as a feature of a moderated climate. Intuitively, continental and oceanic air mass movements, positioning, and interactions would modify the hourly surface air temperatures. Thus, long term climatological shifts in these air mass positions and strengths would likely correspond with changes in the seasonal concentrations of these air temperature values. Consequently, analyzing any potential evolutions in these distributions whether from unimodal to bimodal or bimodal to unimodal could provide groundbreaking insights as these specific strategies for analyzing historical hourly temperature observations have not previously been explored in the literature. Studies addressing location based climatological air temperature changes over time have been compiled with respect to daily data. The continentality of the eastern third of North America has been shown to increase while the central and western portions of the continent have illustrated a decrease in the past few decades (Stonevicius et al., 2018). Furthermore, warming over the Arctic region is likely to influence air mass patterns, leading to lesser negative air temperature anomaly events over North America with increasing frequencies of anomalous Rossby wave amplitudes over the midlatitudes (Gervais et al., 2016). In terms of historical temperature trends, the most significant warming across North America has occurred with respect to the warming of minimum temperatures particularly during winter (Robeson, 2004). Spatial patterns are distinct across CONUS with regional variations and ocean-atmosphere variability playing a major role (Robeson, 2004; Mutiibwa et al., 2015).

Identifying regional patterns of climatological unimodal and bimodal air temperature hours distributions at specific locations as well as evaluating how these distributions are changing over time has the potential to produce an innovative synthesis of the climate change that has already occurred.

Data and Analysis Methods

This project accesses quality controlled historical hourly in-situ surface weather station data from the Integrated Surface Database Lite (ISD-Lite) product (NCEI, 2024). The ISD-Lite is a derived product from the full Integrated Surface Database provided by the National Centers for Environmental Information (NCEI) which contains a variety of combined surface hourly datasets from across the world. These are condensed into a structured attribute table for each individual station after undergoing NCEI's standardized quality control methods with air temperature being one of the parameters validated most extensively (Smith et al., 2011). Hourly air temperature observation data is pulled for the 45 year period from 1978 to 2022 for individual locations across CONUS and southern Canada. Each of these locations correspond to weather stations at airports and are labeled by International Civil Aviation Organization (ICAO) identifiers. 332 total stations are used for the following analysis as a subset of the total available stations in the ISD-Lite for the domain. This was determined by ensuring that stations used for analysis were active at both the beginning and end of the period and included a minimum of 85% of the total possible hourly air temperature observations for each year during the period. This allows for the flexibility of missing observations while maintaining a sufficient and continuous record of hourly data throughout the time span being analyzed. The native air temperature resolution of the hourly data for the stations in this domain is recorded in every whole degree Fahrenheit

For each station, hourly air temperature observations are pulled for the 1978-2022 period. The total number of hours recorded at each degree Fahrenheit are then compiled. This allows for a climatological distribution of air temperature hours to be constructed for the 45 year period that corresponds to a particular station. Stemming from this distribution, the primary objective pertains to identifying the amount of modes which correspond to the prominent peak(s) of the distribution. A model in the form of a probability density function (PDF) is fit to the distribution by using kernel density estimation (KDE) so as to apply a smoothing algorithm that can be used for understanding the shape of the distribution (Sheather, 2004). The KDE algorithm is a non-parametric method of determining the probability density function that does not assume any particular underlying distribution of the data. The “Silverman” method is used for the bandwidth parameter which dictates the smoothness of the fit (Sheather, 2004). Attempting to find the mode(s) of the raw hourly data alone for a station would be too sensitive to slight random variations whereas the PDF model incorporates smoothing that averages out these fluctuations, providing a more robust and stable estimation of the underlying distribution patterns. This approach allows for the identification of true modal values using the PDF, reflecting the most realistic climate behavior. Examples of these air temperature hours distributions and corresponding PDFs are illustrated in **Figure 1**.

There are three categories that air temperature hours distributions can be classified into: Unimodal, Bimodal, or In-Between. This determination is made using an automated, stringent methodology that filters based upon the prominence of these modes. The accuracy of this process was visually inspected and verified as sufficient for 100 of the stations. The number and locations of modes are determined by using the Scipy find_peaks() function which requires the input of the PDF evaluated over the range of air temperature values and finds the individual peak density values of this model. When the process is ran for a distribution, a maximum of two modes can be found which corresponds to the modes of the highest prominences. It is assumed at least one prominent mode will always be found. If the prominence value of a mode found is less than 2.5% of the maximum probability density of the PDF, that mode is removed from the distribution as this is a minor perturbation. After this step, a secondary, over-smoothed PDF is fit using KDE with a bandwidth value of .15, much greater than the solved bandwidths using the "Silverman" method. The purpose of this over-smoothed PDF is to serve as a quality check to ensure that a second mode detected is dominant enough to be found even in a smoothed case. It is important to note that the over-smoothed PDF would not capture the sensitivities of the modes effectively. It is only used for confirming the existence of a "true" second mode where a prominence filter was found not to be sufficient in all cases. The same methods for finding the modes for this over-smoothed PDF are also applied. If both the normal PDF and over-smoothed PDF solve for one mode, the distribution is classified as unimodal. An example case pertains to **Figure 1a** which illustrates the unimodal case for San Antonio, TX (KSAT). Similarly, if both the normal PDF and over-smoothed PDF solve for two modes, the distribution is classified as bimodal. A strong bimodal distribution is shown for Chicago, IL (KORD) in **Figure 1b**. However, if the normal PDF solves for two modes and the over-smoothed PDF solves for one mode, the distribution is classified as In-Between. This In-Between category is designed as a buffer for distributions that do not precisely fit the definition of a unimodal or bimodal distribution. **Figure 1c** illustrates an In-Between case for St. Louis, MO (KSTL) where a dominant mode on the warm side of the distribution is clearly identified, and while the subtle plateau on the cold side of the distribution is not explicitly considered a mode, it is still noted as significant relative to the climatology by the nature of the classification.

Once all of the distribution types are identified, changes to the distributions over time can be identified by following a very similar process. For each station, the time period can be subdivided into the two 15 year stints of 1978-1992 and 2008-2022. These stints capture distinct climate periods that can illustrate evolution within the total time period. The same analysis that was performed for the 45-year period is replicated for each 15 year stint. Stations with changes to their distribution classifications are noted.

Results

The 45 year climatological air temperature hours distribution type was determined for each station across CONUS and southern Canada following the automated classification process. Illustrated in **Figure 2**, there are a wide variety of regional patterns that can be identified and assessed. Primarily, the general expectations for which stations were to be classified as Unimodal versus Bimodal were relatively upheld at least for regions east of 105°W. Most locations in the southeast US featured unimodal distributions that would be characteristic of an oceanic climate whereas higher latitude locations featured a bimodal distribution more characteristic of a continental climate. The stations in the In-Between category functioned as an effective transition region between the unimodal and bimodal distributions mostly along the 38th parallel. For regions west of 105°W, the mountainous terrain of the Rockies as well as the desert climate certainly contributed to these differing results comparatively. A majority of these western stations were classified as Unimodal apart from regional pockets of In-Between and Bimodal stations in the desert southwest and the Pacific northwest. These regions help to highlight the role that local climates play in determining the air temperature distributions found, and this rationale can be extended to some of the outlying or unexpected cases such as the few In-Between stations in the southeast and the lone Unimodal and In-Between stations in Michigan.

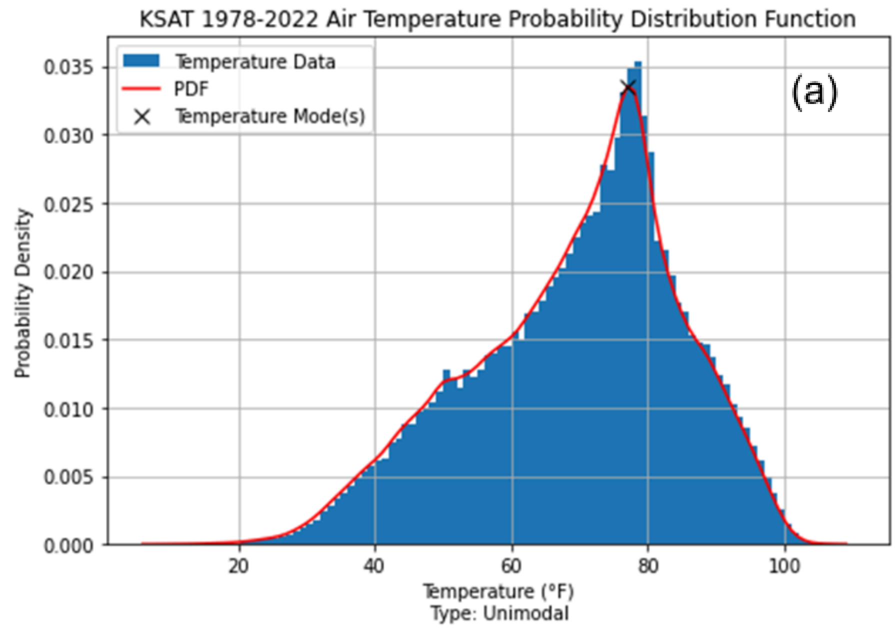
The evolution of these climatological air temperature hours distributions can also be observed by referencing **Figure 3**. For the stations that witnessed a change in distribution type, the inside station color represents the original distribution type analyzed for the 1978-1992 period while the outline station color represents the most current distribution type analyzed for the 2008-2022 period. The most apparent result pertains to the fact that the overwhelming majority of stations did not feature a change in the type of distribution across the period. No stations in the southeast US experienced distribution changes. There are two dominant regions that feature distinct distribution changes, both of which involve regions of In-Between stations. The buffer region between the Unimodal stations and the Bimodal stations illustrated in **Figure 2** surrounding the 38th parallel featured transitions towards a unimodal distribution. Most originally Bimodal stations surrounding this area transitioned to In-Between stations while originally In-Between stations transitioned to Unimodal stations. Surprisingly, there were even stations that featured a complete change from a bimodal distribution to a unimodal distribution. The second region of interest pertains to the desert southwest which appears to largely feature a transition towards a bimodal distribution from Unimodal and In-Between stations. Apart from these regions, there are a scattered few stations in the higher latitudes that feature transitions, but no discernable patterns.

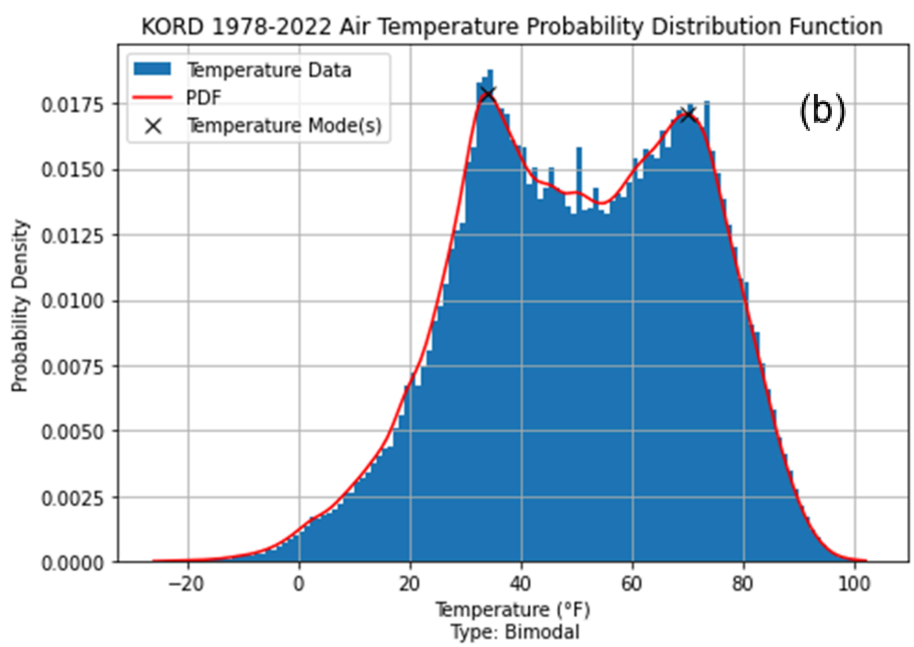
Conclusion

The analysis of air temperature hours distributions for locations across CONUS and southern Canada for the 45 year historical period provided significant climatological insights. By employing kernel density estimation to model a probability distribution function, these distributions were classified into either Unimodal, Bimodal, or In-Between categories. Most bimodal distributions were associated with inland regions and higher latitudes while unimodal distributions were associated with coastal regions and lower latitudes. The inclusion of the In-Between category was instrumental in ensuring that the classifications of unimodal and bimodal distributions could be confidently made. This category also functioned as a method of observing key transition zones over different regions as well as an important indicator with respect to how these distributions are changing over time. Considering that a vast majority of the outliers observed in **Figure 2** that did not hold to a regional pattern also did not feature a distribution change over time points to these being reliable classifications as well as to the importance that local climatology has on impacting these distributions. However, the classification methods are potentially still too subjective as qualifying the accuracy of these methods are currently based upon an eye test alone as no prior literature has covered these bases. Even so, the analysis overall could be impacted by how stringent the prominence and smoothness thresholds are for interpreting which cases fall as In-Between cases. For the regions identified that did witness changes to distribution type over time, specific examples in **Figure 4** were pulled to provide context for why this might be occurring. The distribution for Las Vegas, NV (KLAS) in **Figure 4a** originally featured a cold, prominent mode with a warm plateau during the first 15 year period. The latest period saw a warmer mode develop as concentrations of cooler temperatures weakened. This is likely the pattern of the desert southwest stations that exhibited a shift towards a bimodal distribution. In contrast, the distribution for Kansas City, MO (KMCI) in **Figure 4b** originally featured both prominent cold and warm modes. However, the latest period featured the less prominent cold mode eroding away completely as the warm mode strengthened in prominence. Furthermore, the concentrations of air temperature hours between the previous two modes increase. Stations in the Midwest region along the 38th parallel featuring a unimodal transition are likely experiencing similar trends. It is possible that the 15 year stints used for determining the distribution changes are too short in length. These periods most certainly include the confluence of impacts beyond climate change related to decadal oscillations such as the Atlantic Multidecadal Oscillation.

There are broad implications related to the field of climate analysis that can be discussed. The ability to pinpoint climate behavior on an hourly basis provides a more nuanced understanding for developing tailored climate adaptation strategies. For instance, regions transitioning from bimodal to unimodal distributions could indicate a moderating climate, which might influence agricultural planning, urban development, and public health policies. The structure of this research enables the analysis to be extended to cover many other metrics that have the potential to change over the course of the climatological period. Changes to the standard deviation, temperature mode values, skewness, and extreme values of these air temperature hour distributions could all be explored as extensions that daily data cannot offer. These findings could also work to enhance climate communication strategies by providing concrete examples of how subtle shifts in climate can have pronounced effects on local temperature patterns. This would help make the abstract concept of climate change more tangible for the public and policymakers alike.

Figures


KSAT 1978-2022 Air Temperature Probability Distribution Function
Temperature Data
PDF
Temperature Mode(s)
(a)
Probability Density
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
20
40
60
80
100
Temperature (°F)
Type: Unimodal


KORD 1978-2022 Air Temperature Probability Distribution Function
Temperature Data
PDF
Temperature Mode(s)
(b)
Probability Density
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
−20
0
20
40
60
80
100
Temperature (°F)
Type: Bimodal

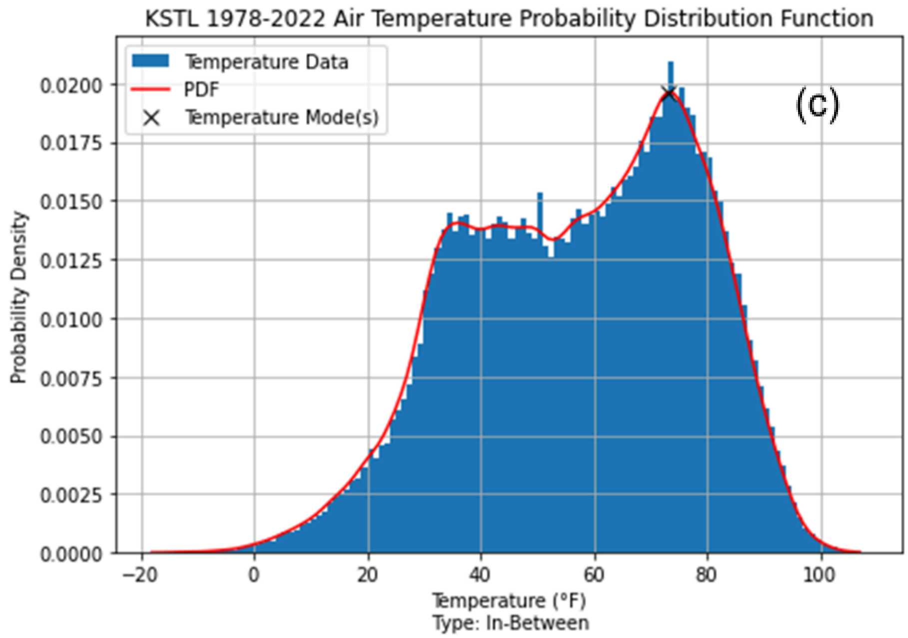


Figure 1: Climatological air temperature hours distributions from 1978-2022 illustrating the probability density at each degree Fahrenheit and corresponding probability distribution functions are shown for **(a)** San Antonio, TX (KSAT), **(b)** Chicago, IL (KORD), and **(c)** St. Louis, MO (KSTL). The temperature modes found are also illustrated, and each city represents an example case of a Unimodal, Bimodal, and In-Between distribution type respectively.

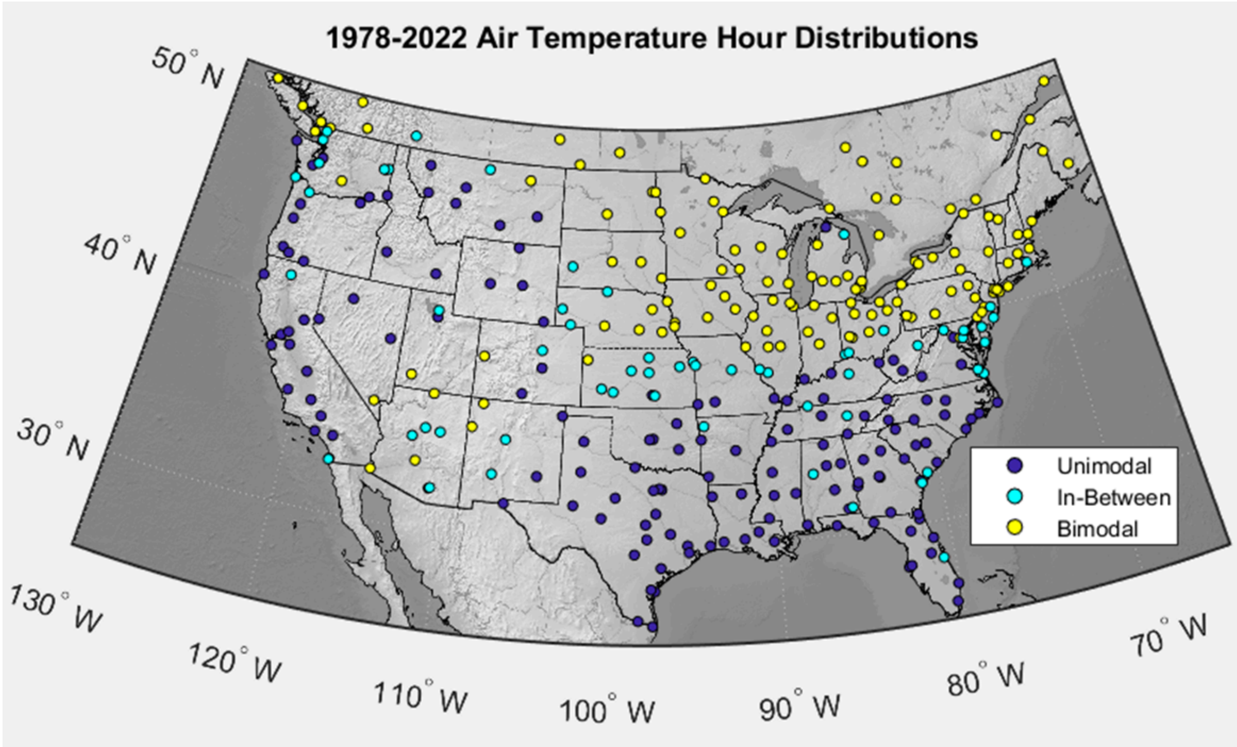


Figure 2: The climatological air temperature hours distributions from 1978-2022 for 332 stations across CONUS and southern Canada are classified into either Unimodal, Bimodal, or In-Between categories. Regional patterns can be observed in relation to different climates.

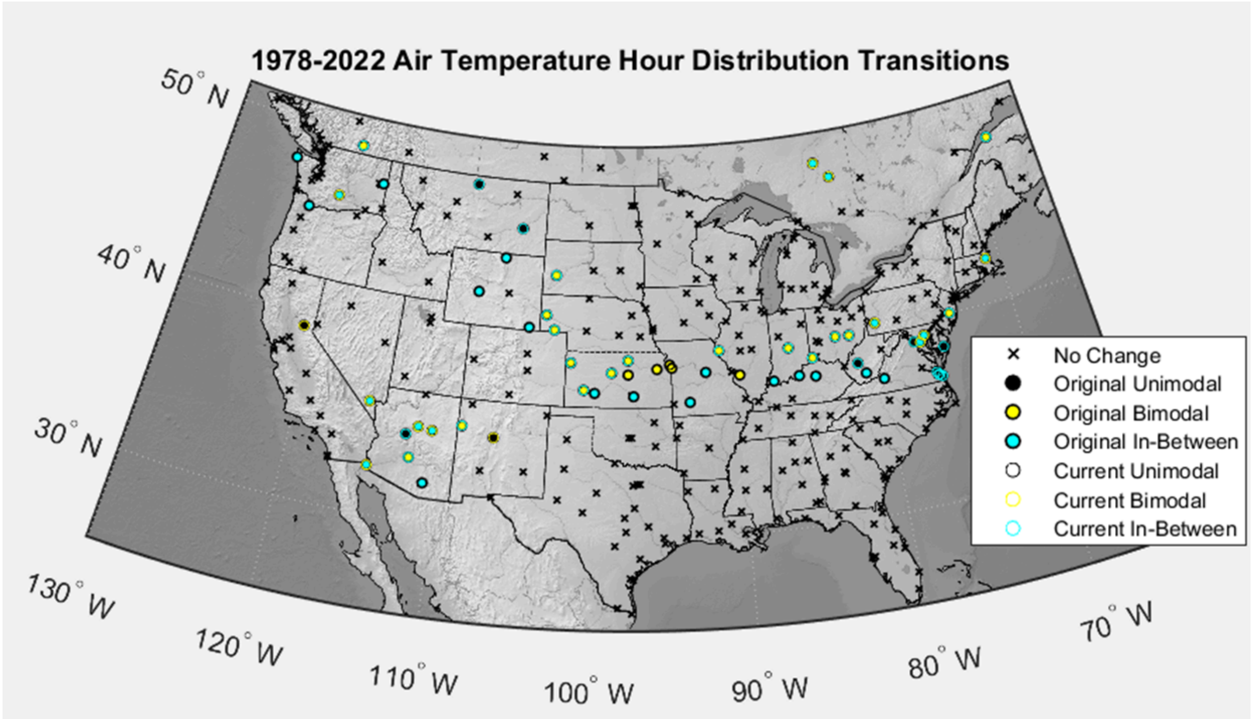


Figure 3: The air temperature hours distribution transitions from 1978-2022 for 332 stations across CONUS and southern Canada are illustrated. The "Original" classification of the distribution type was determined for the 1978-1992 period and represents the interior color. The "Current" classification of the distribution type was determined for the 2008-2022 period and represents the outline color. Transitions towards a unimodal distribution are shown for stations along the 38th parallel while stations in the desert southwest feature transitions towards a bimodal distribution.

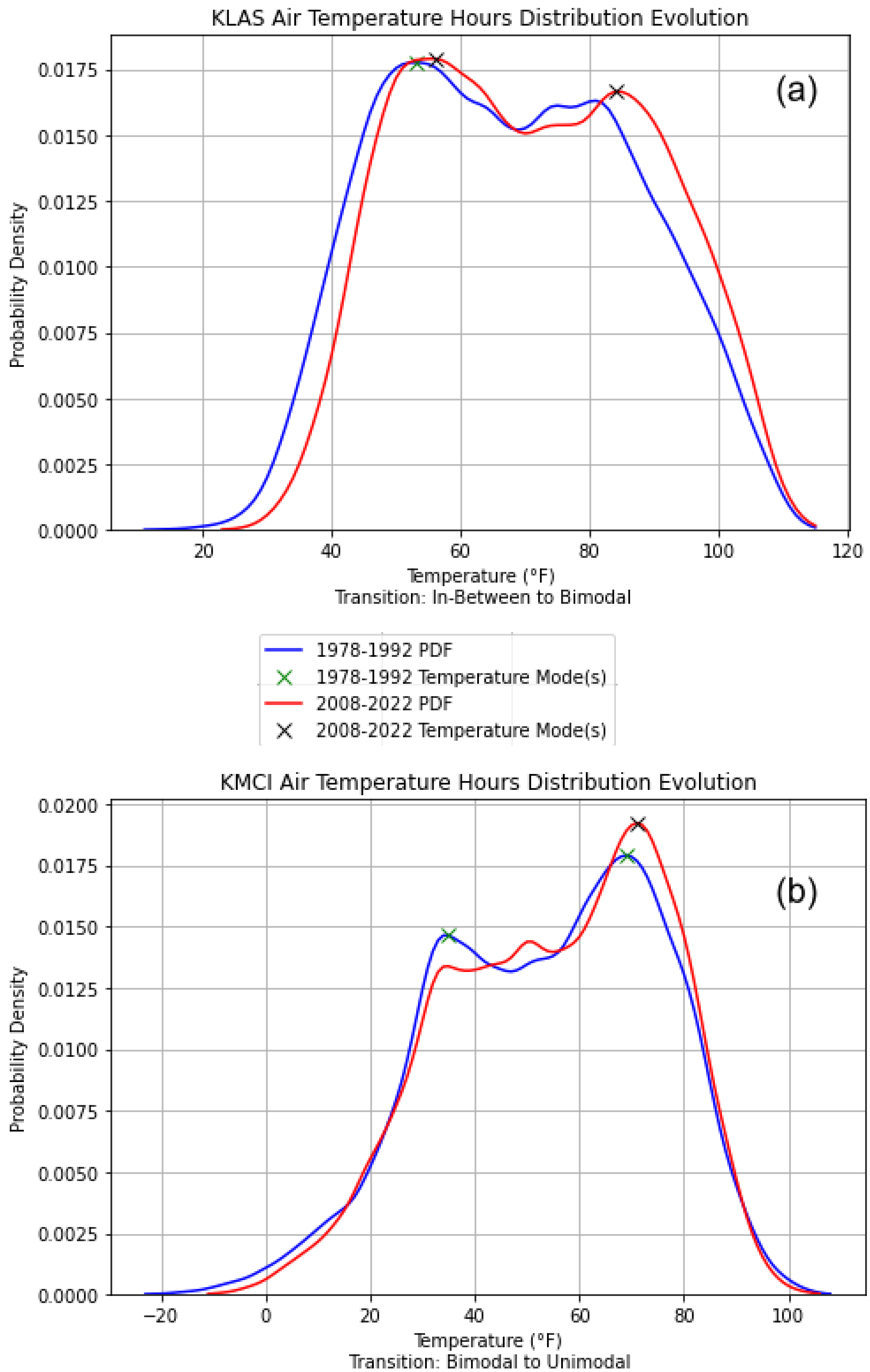


Figure 4: Air temperature hours probability density functions shown for 1978-1992 and 2008-2022 with modes for these corresponding curves identified. **(a)** Las Vegas, NV (KLAS)

features a transition from In-Between to Bimodal. **(b)** Kansas City, MO (KMCI) features a transition from Bimodal to Unimodal.

References

Markowitz, E. M., & Guckian, M. L. (2018). 3 - Climate change communication: Challenges, insights, and opportunities. In S. Clayton & C. Manning (Eds.), *Psychology and Climate Change* (pp. 35–63). doi:10.1016/B978-0-12-813130-5.00003-5

Pahl, S., Sheppard, S., Boomsma, C., & Groves, C. (2014). Perceptions of time in relation to climate change. *WIREs Climate Change*, *5*(3), 375–388. doi:10.1002/wcc.272

Bloomfield, H. C., Brayshaw, D. J., Deakin, M., & Greenwood, D. M. (2022). Hourly historical and near-future weather and climate variables for energy system modelling. *Earth System Science Data*.

Amato, A. D., Ruth, M., Kirshen, P., & Horwitz, J. (2005). Regional Energy Demand Responses To Climate Change: Methodology And Application To The Commonwealth Of Massachusetts. *Climatic Change*, *71*(1), 175–201. doi:10.1007/s10584-005-5931-2

Matthews, T. K. R., Wilby, R. L., & Murphy, C. (2017). Communicating the deadly consequences of global warming for human heat stress. *Proceedings of the National Academy of Sciences*, *114*(15), 3861–3866. doi:10.1073/pnas.1617526114

Kjellstrom, T., Freyberg, C., Lemke, B., Otto, M., & Briggs, D. (2018). Estimating population heat exposure and impacts on working people in conjunction with climate change. *International Journal of Biometeorology*, *62*(3), 291–306. doi:10.1007/s00484-017-1407-0

Herbut, P., Angrecka, S., Godyń, D., & Hoffmann, G. (2019). The Physiological and Productivity Effects of Heat Stress in Cattle – A Review. *Annals of Animal Science*, *19*(3), 579–593. doi:10.2478/aoas-2019-0011

Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., & Stewart, B. C. (2017). *Climate Science Special Report: Fourth National Climate Assessment, Volume I*. doi:10.7930/J0J964J6

Perera, A. T. D., Nik, V. M., Chen, D., Scartezzini, J.-L., & Hong, T. (2020). Quantifying the impacts of climate change and extreme climate events on energy systems. *Nature Energy*, *5*(2), 150–159. doi:10.1038/s41560-020-0558-0

Smith, A., Lott, N., & Vose, R. (2011). The Integrated Surface Database: Recent Developments and Partnerships. *Bulletin of the American Meteorological Society*, *92*(6), 704–708. doi:10.1175/2011BAMS3015.1

National Centers for Environmental Information. (2021). *Integrated Surface Database Lite (ISD-Lite)*.

Zhu, T., Fonseca De Lima, C. F., & De Smet, I. (2021). The heat is on: how crop growth, development, and yield respond to high temperature. *Journal of Experimental Botany*, *72*(21), 7359–7373. doi:10.1093/jxb/erab308

Siebert, S., & Ewert, F. (2014). Future crop production threatened by extreme heat. *Environmental Research Letters*, Vol. 9. doi:10.1088/1748-9326/9/4/041001

Tong, S., Prior, J., McGregor, G., Shi, X., & Kinney, P. (2021). Urban heat: An increasing threat to global health. *The BMJ*, *375*. doi:10.1136/bmj.n2467

Guo, F., Do, V., Cooper, R., Huang, Y., Zhang, P., Ran, J., … Fu, Z. (2021). Trends of temperature variability: Which variability and what health implications? *Science of the Total Environment*, Vol. 768. doi:10.1016/j.scitotenv.2020.144487

Gervais, M., Atallah, E., Gyakum, J. R., & Bruno Tremblay, L. (2016). Arctic air masses in a warming world. *Journal of Climate*, *29*(7), 2359–2373. doi:10.1175/JCLI-D-15-0499.1

Stonevicius, E., Stankunavicius, G., & Rimkus, E. (2018). Continentality and Oceanity in the Mid and High Latitudes of the Northern Hemisphere and Their Links to Atmospheric Circulation. *Advances in Meteorology*, *2018*. doi:10.1155/2018/5746191

Intergovernmental Panel on Climate Change. (2014). *Climate change 2014 : synthesis report : longer report*.

Sheldon, K. S., & Dillon, M. E. (2016). Beyond the Mean: Biological Impacts of Cryptic Temperature Change. *Integrative and Comparative Biology*, *56*(1), 110–119. doi:10.1093/icb/icw005

Camacho, A., Trefaut Rodrigues, M., & Navas, C. (2015). Extreme operative temperatures are better descriptors of the thermal environment than mean temperatures. *Journal of Thermal Biology*, *49–50*, 106–111. doi:10.1016/j.jtherbio.2015.02.007

Robeson, S. M. (2004). Trends in time-varying percentiles of daily minimum and maximum temperature over North America. *Geophysical Research Letters*, *31*(4). doi:10.1029/2003GL019019

Mutiibwa, D., Vavrus, S. J., McAfee, S. A., & Albright, T. P. (2015). Recent spatiotemporal patterns in temperature extremes across conterminous United States. *Journal of Geophysical Research*, *120*(15), 7378–7392. doi:10.1002/2015JD023598

Sheather, S. J. (2004). *Density Estimation*.