Lawrence Berkeley National Laboratory
LBL Publications

Title
A Historical Extreme Cold Events Dataset for Building Energy and Resilience Modeling Across the United States

Permalink
https://escholarship.org/uc/item/5m9363m6

Journal
Scientific Data, 13(1)

ISSN
2052-4463

Authors
Krelling, Amanda F
Malik, Jeetika
Hong, Tianzhen

Publication Date
2026-06-01

DOI
10.1038/s41597-026-07717-y

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A Historical Extreme Cold Events Dataset for Building Energy and Resilience Modeling Across the United States

Amanda F. Krelling, Jeetika Malik & Tianzhen Hong

Extreme cold snaps pose significant risks to buildings, infrastructure, energy systems, and occupants, yet standardized climatic datasets tailored for resilience-focused building performance modeling remain limited. This study presents a methodology and corresponding dataset of cold snap events for 217 U.S. cities, derived from 24 years of historical hourly temperature data obtained from the NASA POWER project. Cold snaps were detected using a percentile-based, location-specific threshold that identifies periods of "abnormal cold" with additional constraints to ensure that events reflect meaningful differences from local winter conditions. Each event was characterized using a suite of metrics, including event duration, heating degree hours, and overcooling degree. Events were further classified into four categories based on the mean outdoor air dry-bulb temperature, analogous to intensity scales used in other hazard domains. A selection procedure was applied to ensure that each city is represented by a small set of short, medium, and long-duration events, resulting in a curated dataset of 880 cold snaps suitable for building energy simulations and resilience assessments. The dataset is provided as EnergyPlus Weather (EPW) files accompanied by a summary spreadsheet containing all events and their metrics. This dataset supports the systematic evaluation of building performance under extreme cold weather conditions and provides a foundation for thermal and energy resilience modeling across the U.S. climates.

Background & Summary

Extreme cold events are a major but often underestimated public health concern. Evidence from large-scale epidemiological analyses shows that cold temperatures account for a much greater share of temperature-related mortality than heat, largely because prolonged exposure to moderately cold conditions contributes to cardiovascular and respiratory stress over extended periods[1]. Although average temperatures have risen in recent decades, severe cold outbreaks continue to occur with significant social and economic consequences. The 2023 cold event in East Asia, for example, caused widespread disruption to transportation, power systems, and agriculture, despite occurring during one of the warmest years on record. Large-scale circulation anomalies, such as the Warm Arctic–Cold Eurasia pattern, continue to drive extreme cold events even in historically warm years, which evidences the need to evaluate their risks to infrastructure, energy systems, and public health[2].

Robust monitoring of temperature-related risks is fundamental to support early warning systems, cold resilience planning, and long-term mitigation investments. However, limited information is available to guide such efforts. Innovative approaches to assess resilience are essential to support decision making effectively across different populations[3]. Building energy modeling (BEM) could help bridge this gap by simulating the most diverse thermal conditions that people could be exposed to where they spend about 90% of their lives[4], inside built environments[5]. In this context, an intersectional challenge is the definition of the extreme events[6], because resilience assessments are strongly linked to specific—or a combination of—events. These are no ordinary events, but rather carefully defined conditions that stress the indoor thermal environment, such as cold snaps. Standardized methods to curate these events for resilience simulations are still limited, which consequently also limits the availability of comprehensive datasets of extreme events to support these efforts. One of the main

Building and Industrial Energy Systems Division, Lawrence Berkeley National Laboratory, California, USA. ✉e-mail: thong@lbl.gov

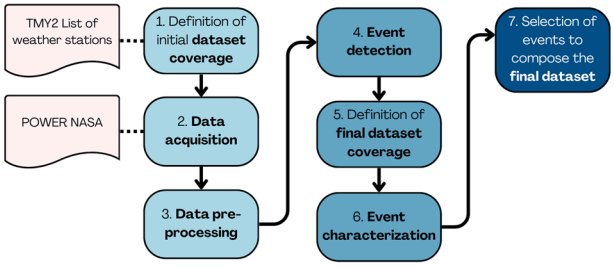


Fig. 1 Workflow for curating a U.S. cold snap dataset for resilience assessments.

datasets currently available for thermal resilience assessment purposes was published by Machard *et al.*[7], who developed future typical meteorological years (TMY) and heat wave years across 15 major cities worldwide, as part of the research done within the international Annex 80 project: Resilient Cooling of Buildings (https://annex80.iea-ebc.org/). Though valuable for the field, this dataset is limited in geographic coverage and does not consider extreme cold events.

In addition to TMY-based datasets, extreme weather files have been developed to assess building performance under more severe conditions, such as Design Summer Years (DSY)[8] for overheating risk. However, while DSY methodologies are relatively well established, equivalent approaches for extreme cold conditions remain less standardized. Some efforts have addressed this gap. For example, Nik[9] proposes the synthesis of an Extreme Cold Year (ECY) from regional climate model outputs by selecting months that capture the most severe cold conditions while preserving realistic temporal variability. Still, these approaches typically condense extremes into a single representative year, which may not fully capture the temporal continuity, duration, and sequencing of real cold events. Preserving these characteristics is critical for resilience analyses, as building performance depends not only on peak conditions but also on how long adverse conditions persist and how systems respond over time. This limitation evidences the need for alternative datasets that explicitly represent individual extreme events, such as cold snaps, to better support resilience-focused analyses.

This article presents the development of a standardized dataset[10] of cold snap events for 217 U.S. cities, designed to support building performance simulations under severe cold conditions. The scope includes defining a consistent method for detecting, characterizing, and quantifying cold snaps using multi-decadal historical hourly outdoor air temperature records; generating event-level metrics such as heating degree hours, overcooling degree, event duration, and event category; and organizing these outputs into a format suitable for integration with building energy and resilience modeling. The article also shows how the dataset enables systematic assessments of occupant and building survivability requirements during low-temperature extremes and potential power disruptions, providing a foundation for comparative resilience analysis across climates, regions, and building types.

While this dataset[10] is grounded in historical observations, it is intended to provide a consistent and physically realistic basis for resilience assessments under extreme cold conditions. The focus on observed events ensures that the identified cold snaps reflect conditions that have occurred in practice, preserving the temporal continuity, duration, and intensity characteristics that are critical for evaluating building performance and occupant safety. We recognize that, in the context of variable climate conditions, understanding how the frequency and severity of such events may evolve is of paramount importance. However, analyses of future conditions require additional assumptions[11], including the selection of scenarios (e.g., Shared Socioeconomic Pathways[12]), downscaling approaches, and other modeling choices that are often highly context-specific. As such, this work does not prescribe a framework for future projections, but instead establishes a robust methodological foundation and baseline dataset that can be extended in forthcoming studies to incorporate future climate conditions and support forward-looking resilience planning.

The primary stakeholders of this dataset[10] are building and urban energy modelers, who can directly integrate the standardized cold snap events and associated metrics into building energy modeling, urban-scale simulations, and thermal resilience assessments under severe cold conditions. Beyond direct users, the dataset also benefits decision-makers and practitioners, including planners, policymakers, utilities, and public health officials, by enabling analyses that translate modeling results into actionable insights. These downstream applications can inform cold resilience planning, infrastructure investment priorities, and emergency preparedness. The methodology and the curated dataset also support the international Annex 93 project: Energy Resilience of Buildings in Remote Cold Region (https://annex93.iea-ebc.org/).

Method

To construct a consistent and scalable dataset of historical extreme cold events for the U.S., we developed a workflow illustrated in Fig. 1. Each of these steps is described in Sections 2.1 through 2.7. The method follows two guiding principles: (1) detection of events based on "abnormal cold" metrics specific to each location, and (2) characterization of events using thresholds associated with cold-stress conditions.

Dataset coverage. The dataset[10] covers 217 U.S. cities selected to align with the national distribution of TMY weather stations[13], which have historically served as a standardized reference for climatic representation in building performance analysis. Starting from the original list of 242 TMY2 locations (see Fig. 2), cities

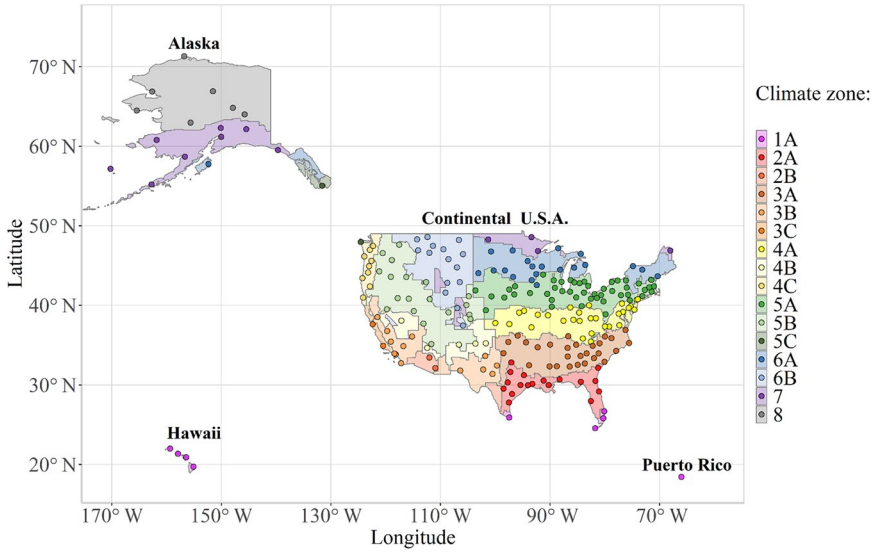


Fig. 2 Location of weather stations across the continental United States, Alaska, Hawaii, and Puerto Rico colored according to 2021 International Energy Conservation Code (IECC) climate zones[20].

were screened based on data availability, completeness of the historical temperature record, and suitability for long-term event detection, resulting in a final set that preserves broad geographic and climatic diversity across the contiguous United States, Alaska, and U.S. territories where applicable. This approach ensures that the cold snap dataset maintains consistency with established weather-data conventions in the building simulation community while providing extensive spatial coverage sufficient for regional comparisons, national-scale studies, and cross-climate resilience assessments.

Data acquisition. We compiled 24 years of historical hourly meteorological records for each city using the NASA POWER project database[14], which provides consistent, gap-free coverage across the United States. Our goal was to use at least 20 years of data to ensure a sufficiently long period for detecting and characterizing rare cold events. After assessing data availability across all candidate locations, we determined that a 24-year period representing the longest continuous interval for which complete hourly temperature data were available for every city.

Data from NASA POWER is publicly available and can be extracted manually through the Data Access Viewer (Single Point) (https://power.larc.nasa.gov/data-access-viewer/) or through a dedicated API. If the former is adopted, the following information should be provided:

- Data Resolution: Standard Resolution
- User Community: Sustainable Buildings
- Temporal Level: Hourly
- Location: Latitude and Longitude of selected location
- Time Extent: Start and End dates for the period considered
- Data Download: EPW

Data pre-processing. Data pre-processing involved combining the 24 individual years of hourly temperature data into a single continuous time series for each city, ensuring that all records followed a consistent chronological structure. From this merged dataset, average daily outdoor air dry-bulb temperatures were calculated to support event detection.

Event detection. Cold snap events were identified using a threshold-based method designed to capture periods of "abnormal cold" for each specific location. For each city, we calculated the 2.5th percentile of the full distribution of daily average dry-bulb temperatures and used this value as the preliminary event threshold (see example for the city of Chicago in Fig. 3). To prevent the detection of unrealistically warm "extremes" in mild climates, any percentile value above 0 °C was replaced with a fixed threshold of 0 °C. A cold snap was then defined as any period of at least two consecutive days with the daily average temperatures at or below this city-specific threshold. This approach ensures that events reflect locally significant cold conditions while maintaining a minimum standard that distinguishes genuinely extreme cold from normal seasonal variability.

Daily mean outdoor air temperature was used to define cold snap thresholds to ensure that detected events reflect sustained periods of cold conditions rather than short-lived temperature extremes. In particular, relying on daily minimum temperature could lead to the identification of events driven by brief, instantaneous drops (often occurring at night) that do not represent persistent cold stress over the full day. On the other hand, daily mean temperature requires that cold conditions be maintained over an extended period, making

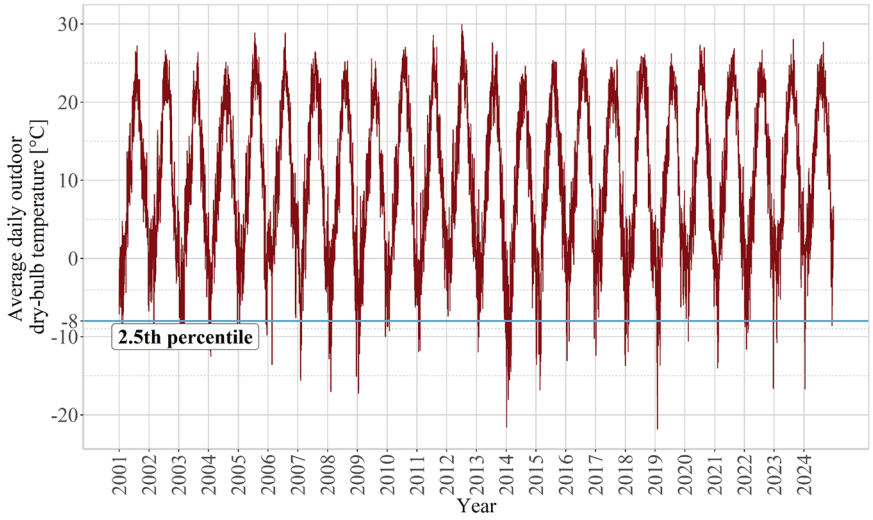


Fig. 3 Average daily outdoor air dry-bulb temperature in Chicago across 24 years of data.

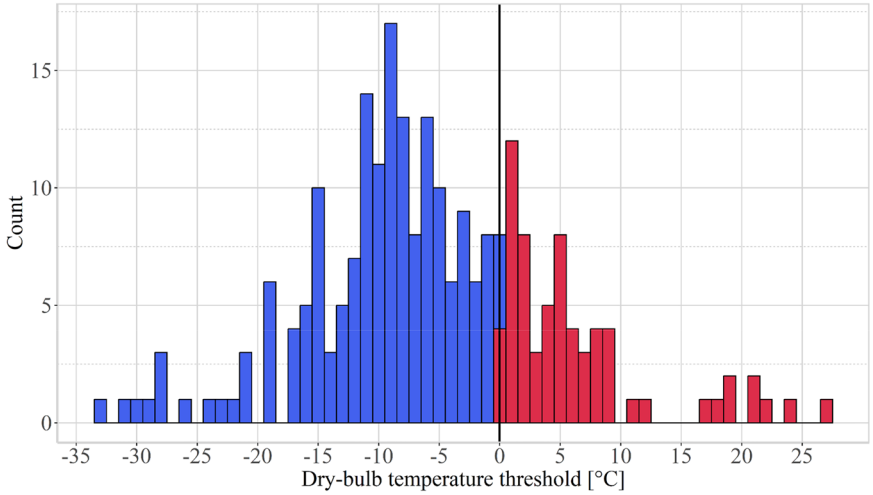


Fig. 4 Distribution of threshold values across the U.S. cities.

it more representative of the sustained exposure that affects building performance and resilience. This choice is particularly relevant for building analyses, where thermal stress depends on prolonged exposure rather than instantaneous minima. The use of daily mean temperature is also consistent with prior studies on extreme event detection in similar contexts[7,15].

Furthermore, events separated by fewer than two days were combined into a single event. This merging procedure prevents fragmentation of extended cold periods caused by short-lived threshold exceedances and ensures a more consistent representation of sustained cold stress conditions.

Figure 4 shows the distribution of these cold-snap thresholds across all cities. Colder locations cluster on the left, with much lower percentile-based thresholds, while milder climates appear on the right, where thresholds approach or meet the 0 °C cap. The vertical line marks this imposed 0 °C limit, illustrating how the method prevents unrealistically warm thresholds in mild regions.

The use of the 0 °C cap reflects a deliberate choice to focus on conditions that clearly represent cold stress from a building performance and resilience perspective. While relatively cool temperatures above freezing may still be associated with health impacts, particularly in milder climates, the distinction between cold discomfort and cold stress becomes less well defined in these ranges. By imposing a lower bound at 0 °C, the method ensures that detected events correspond to conditions that are unequivocally cold and more likely to impose sustained thermal stress on both occupants and building systems. This approach aligns with the primary objective of the dataset, which is to support analyses of building performance under severe cold conditions, rather than to comprehensively capture all cold-related health risks.

A sensitivity analysis was conducted to evaluate the impact of the percentile threshold on cold snap detection (see Section 4.1). Results show that while lower thresholds (e.g., 0.5th percentile) restrict detection to a limited number of rare and highly extreme events, higher thresholds (e.g., 5th percentile) substantially increase the

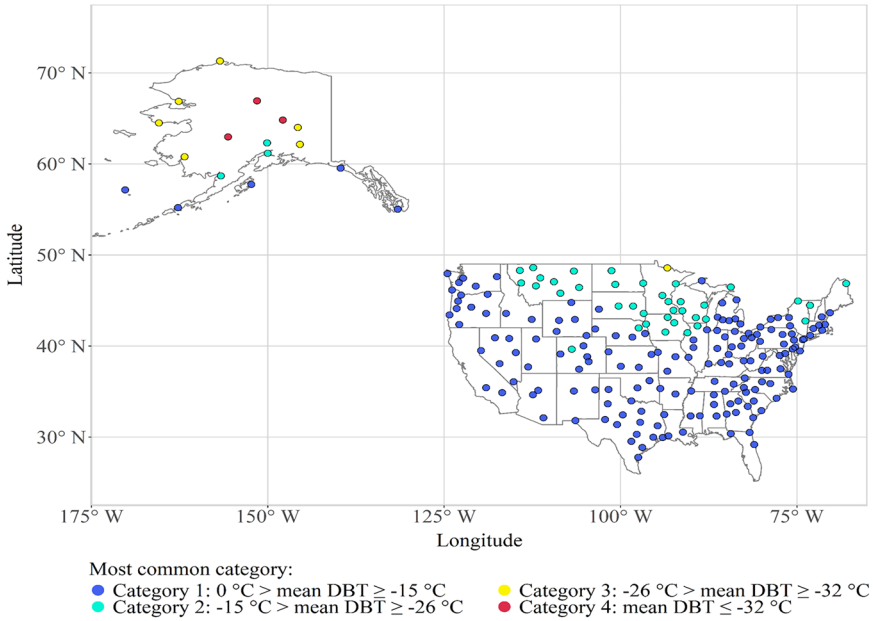


Fig. 5 Location of all selected U.S. cities to compose the dataset, colored according to the most common event category.

number of detected events and include milder conditions, leading to both more frequent short events and a more pronounced tail of longer-duration events. The 2.5th percentile provides an intermediate and balanced representation, capturing sufficiently severe cold conditions while maintaining a diverse set of events in terms of duration and intensity. This balance is particularly important for resilience analyses, which require both extreme and representative conditions to assess building performance under varying levels of thermal stress.

Final dataset coverage. After applying the event detection criteria, the final dataset[10] includes only the cities that exhibited at least one cold snap during the 24-year analysis period, totaling 217 locations. This filtering step ensures that the dataset represents locations where extreme cold conditions are both observable and climatically meaningful, preventing the spurious identification of events in regions with consistently mild winter temperatures. The resulting coverage spans a broad range of U.S. climates, capturing significant diversity in winter severity and event characteristics (Fig. 5).

Event characterization. Event characterization focused on quantifying the severity of each detected cold snap in a consistent manner across all cities. Each cold snap was assigned to one of four categories based on the mean outdoor air dry-bulb temperature averaged over the entire event (DBT_{mean}):

- Category 1: $0\,°C > DBT_{mean} \geq -15\,°C$
- Category 2: $-15\,°C > DBT_{mean} \geq -26\,°C$
- Category 3: $-26\,°C > DBT_{mean} \geq -32\,°C$
- Category 4: $DBT_{mean} \leq -32\,°C$

The temperature thresholds used to define these categories were selected to provide a simple and interpretable severity scale for building performance and resilience analyses, rather than to represent formal health warning levels. The selected values were informed by a combination of infrastructure-relevant thresholds, physiological response studies, and operational guidance, but are used here as indicative reference points rather than prescriptive limits.

The use of 0 °C as the upper boundary reflects the transition to freezing conditions, which are critical from an infrastructure perspective, as prolonged exposure around or below this point can cause pipe freezing[16], mechanical failures, and increased vulnerability during power outages. The −15 °C boundary reflects findings from a human cold-exposure study with healthy young adults[17] indicating that, at this temperature, finger temperatures can fall below 20 °C within 15 minutes—resulting in severe dexterity loss and heightened frostbite risk—while ear pain typically develops within 12 minutes. Exposure to −15 °C low temperature has been linked to elevated cardiovascular strain (e.g., blood pressure increases of roughly 16 mmHg), a sharp drop in heart rate, brief respiratory slowing, and early-onset judgment impairment within the first few minutes, even though core temperature generally remains stable. The thresholds at −26 °C and −32 °C are aligned with temperature ranges commonly used by the National Weather Service (NWS) to signal hazardous cold[18]. Although criteria vary by region, many areas issue Cold Weather Advisories when air temperature or wind chill values fall below −15 °F (−26 °C), indicating conditions that can become dangerous even without significant wind. Extreme Cold

Warnings are typically issued when apparent temperatures reach −25 °F (−32 °C) or lower, marking the onset of conditions in which frostbite and hypothermia risks rise sharply.

These categories should be interpreted as a practical severity scale to distinguish progressively more intense cold conditions for building analysis, enabling consistent comparison across events and climates while avoiding direct interpretation as health-based warning categories. The structure of this scale is analogous to intensity classifications used for hurricanes, serving as a simple but informative way to distinguish different levels of cold stress. By grouping events into clearly defined temperature bands, the categorization helps stakeholders quickly assess the relative severity of cold snaps, prioritize resource allocation, and support planning of building- and infrastructure-related resilience measures, such as backup power strategies, heating system capacity, and emergency operational planning. Figure 5 shows the most common categories across the dataset.

To further characterize the thermal severity of each cold snap, two complementary metrics were calculated using the 0 °C threshold. Heating degree hours (HDH, given in °C-hours) quantify the cumulative temperature deficit over the duration of an event, capturing both how cold it was and how long the cold condition persisted. Larger HDH values indicate events that impose greater heating demand and potential stress on building systems. Overcooling degree (given in °C), defined as HDH divided by the number of hours in the event, represents the average temperature deficit below the threshold, isolating the intensity of cold stress independent of event duration. Equations for both HDH and overcooling degree are presented below, where $T(t)$ is the outdoor air temperature at hour t; T_{thr} is the cold threshold temperature (°C); and N represents the total number of hours in the cold snap event.

$$HDH = \sum_{t=1}^{N} \max(0,\ T_{thr} - T(t))$$

$$Overcooling\ degree = \frac{HDH}{N}$$

These metrics were not used in the event selection process to create the final dataset[10], but rather to provide complementary information about the nature of the detected events and support interpretation and selection of cold snaps for different applications. In particular, HDH communicates the overall severity of an event in a way that may resonate with users in the building performance and energy simulation fields, where cumulative heating demand metrics are commonly used. The overcooling degree provides a clearer indication of how thermally extreme conditions are on average relative to the reference threshold.

In addition to temperature-based metrics, the duration of each cold snap was determined using the start and end dates of the consecutive-day event window. Capturing duration is essential because the impacts of extreme cold extend beyond temperature severity alone: longer events place greater stress on building envelopes, heating systems, and backup power resources. Extended cold periods can also increase the likelihood of pipe freezing, accelerate indoor temperature decay during power outages, and heighten risks to occupants who may face prolonged exposure to unsafe indoor condition[19]. Longer durations also complicate emergency response, evacuation, and recovery by prolonging access limitations, infrastructure damage, and system restoration.

Final dataset. After detecting all cold snap events across the full 24-year period, the final dataset[10] was assembled by selecting representative events for each climate and category. For cities with multiple events within the same category, event selection was based on the distribution of event durations: one short event (corresponding to the 2.5th percentile), one medium-duration event (the median), and one long event (the 97.5th percentile). This approach ensures that each climate is represented by a concise set of events that capture the full range of temporal variability observed within that category. For climates that exhibited more than one severity category, the same procedure was applied separately to each category. When only one event was available within a given climate and category, it was retained directly. For groups with two or more events, the percentile-based procedure was applied. When two or three events were available, this procedure resulted in all available events being retained. This selection method produced a final dataset containing 880 events across 217 climates in the United States.

Table 1 provides eight examples of cold snaps, four detected in Chicago (IL) and four in Fairbanks (AK), demonstrating the case of cities with multiple events and different characteristics. In Chicago, selected events were classified with categories 1 or 2, both with short (2 days as the 2.5th percentile) and long durations (12 days as the 97.5th percentile). Fairbanks follows a similar pattern, but with events in categories 3 and 4 and lasting up to 13 consecutive days. The structure of Table 1 mirrors the one adopted in a spreadsheet that accompanies the dataset, which is further described in Table 2.

Data Records

The complete dataset[10] is available at https://doi.org/10.5281/zenodo.20453235.

The dataset is provided as a collection of whole-year weather files in EnergyPlus Weather (EPW) format downloaded from the NASA POWER project database, with each file corresponding to a specific cold snap event for a particular U.S. city. Accompanying the EPW files is a spreadsheet (in CSV format) that consolidates the characteristics of all selected events across the 217 cities, including the parameters described on Table 2. To use the dataset, researchers should first consult the spreadsheet to identify an event of interest and then retrieve the corresponding EPW file to conduct building performance simulations for the exact period in which the cold snap occurred. The appropriate EPW file to be used for a specific event corresponds to the file covering the same year in which the event occurred. This can be done by checking the year in the name of the EPW file, which follows the pattern "STATE_City name_YEAR."

State, City	Start [month/ day/ year]	End [month/ day/ year]	Dura- tion [days]	Cate- gory	Mean dry-bulb temperature [°C]	Maximum dry-bulb temperature [°C]	Minimum dry-bulb temperature [°C]	HDH [°C-hours]	Over- cooling degree [°C]
IL, Chicago	12/21/ 2008	12/22/ 2008	2	1	−14.0	−3.9	−18.6	672.2	14.0
	12/26/ 2017	1/6/2018	12	1	−10.7	−7	−16.1	3,069.8	10.7
	2/10/ 2008	2/11/ 2008	2	2	−15.3	−8.7	−19.4	732.8	15.3
	1/6/2014	1/8/2014	3	2	−16.8	−9.7	−24.4	1,212.8	16.8
AK, Fairbanks	11/26/ 2003	11/27/ 2003	2	3	−31.8	−19.2	−40.7	1,528.7	31.8
	12/30/ 2011	1/4/2012	6	3	−31.9	−29.7	−34.6	4,592.6	31.9
	1/24/2002	1/25/ 2002	2	4	−36.2	−29.1	−41.1	1,735.3	36.2
	1/22/ 2024	2/3/2024	13	4	−35.4	−26.8	−44.9	11,056.7	35.4

Table 1. Examples of cold snaps detected in Chicago and Fairbanks and their related characteristics.

Parameter	Unit	Description	Column title	Column number
State, City	None	Location of the weather station	state_city	1
Start	Month / Day / Year	First day of the event, starting at 00:00 am	start__month_day_year	2
End	Month / Day / Year	Last day of the event	end__month_day_year	3
Duration	Days	Duration of the cold snap	duration__days	4
Category	None	Category of the cold snap	category	5
Mean dry-bulb temperature (DBT_{mean})	°C	Mean of hourly outdoor dry-bulb temperatures over event duration	dbt_mean__c	6
Maximum dry-bulb temperature (DBT_{max})	°C	Maximum of hourly outdoor dry-bulb temperatures over event duration	dbt_max__c	7
Minimum dry-bulb temperature (DBT_{min})	°C	Minimum of hourly outdoor dry-bulb temperatures over event duration	dbt_min__c	8
Heating degree hours (HDH)	°C-hours	Cumulative temperature deficit over the duration of an event	hdh__c	9
Overcooling degree	°C	Mean temperature deficit below the threshold	overcoolingdeg_c	10

Table 2. Description of event-level parameters and derived metrics for the cold snap dataset.

For cold snap events that span across calendar years (e.g., December - January), two years must be considered to fully represent the event. Accordingly, weather data from both the starting and ending years are included to ensure complete temporal coverage of these multi-year events.

Technical Validation

The distribution of the minimum outdoor air dry-bulb temperatures for all detected cold snap events shows that the most extreme conditions within each event remain well aligned with the long-term temperature characteristics of their respective cities (see Fig. 6). Across climates, event minima (boxplots) consistently fall below the city-specific cold-snap detection thresholds (red points), as expected from the percentile-based method used to identify unusually cold periods. At the same time, these event minima do not exceed the historical cold extremes represented by the lowest hourly temperatures observed over the full 24-year record (blue points), indicating that the selected events remain within the bounds of climatologically plausible conditions rather than reflecting rare outliers. These patterns confirm that the detected cold snaps occupy a reasonable portion of each city's temperature distribution and that the methodology produces events with intensities that are both meaningful and consistent with long-term climatic behavior.

The distribution of cold snap durations (Fig. 7) reveals that the vast majority of detected events are short-lived, typically lasting between two and four days, with progressively fewer events extending beyond a week. We analyzed duration primarily to verify that the detection criteria were not identifying unrealistically long periods that might represent an entire season rather than a distinct cold snap. The histogram confirms that such cases are rare, and nearly all events fall within a duration range consistent with meaningful cold anomalies. Figure 7 is therefore presented as a diagnostic check of the event detection method, illustrating the characteristics of the full set of detected events prior to any selection. It is important to note that this distribution reflects all detected events across all selected U.S. cities and the full 24-year record, rather than the subset of events included in the final curated dataset (shown in Fig. 8).

The distribution of event durations in the final dataset (Fig. 8) shows a similar right-skewed pattern to the full set of detected events, but with a more balanced representation across different durations due to the selection procedure. Because each city contributes up to three representative events per event category—corresponding to short, medium, and long durations—the final dataset includes a broader spread of durations than the raw detection output. While short events of two to three days remain the most common, the selected set includes a meaningful number of longer events, extending beyond one week, which are important for evaluating building resilience under more persistent cold stress. The comparison between Figs. 7 and 8 shows how the selection procedure preserves the overall characteristics of detected events while intentionally enhancing the representation of longer-duration events relevant for resilience analyses.

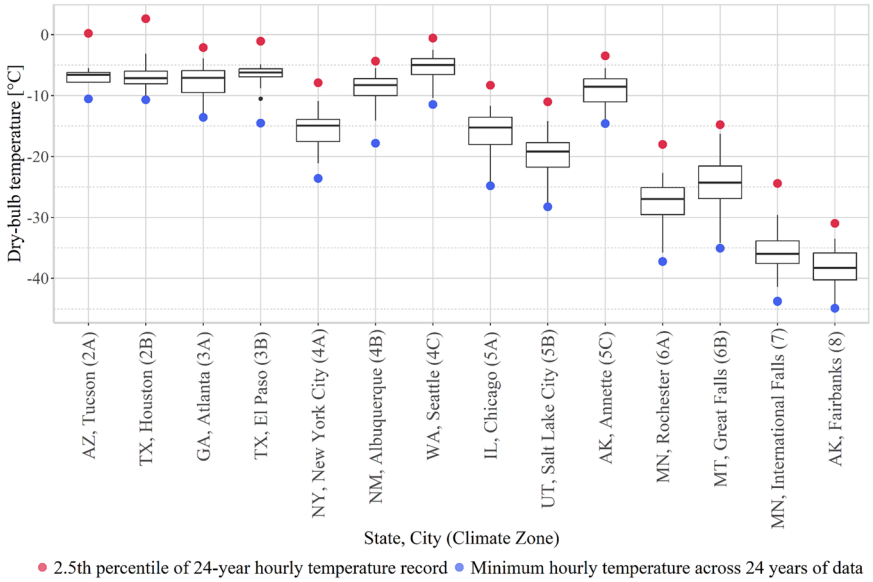


Fig. 6 Minimum hourly outdoor air dry-bulb temperatures during cold snaps compared with long-term cold extremes for representative cities across climate zones.

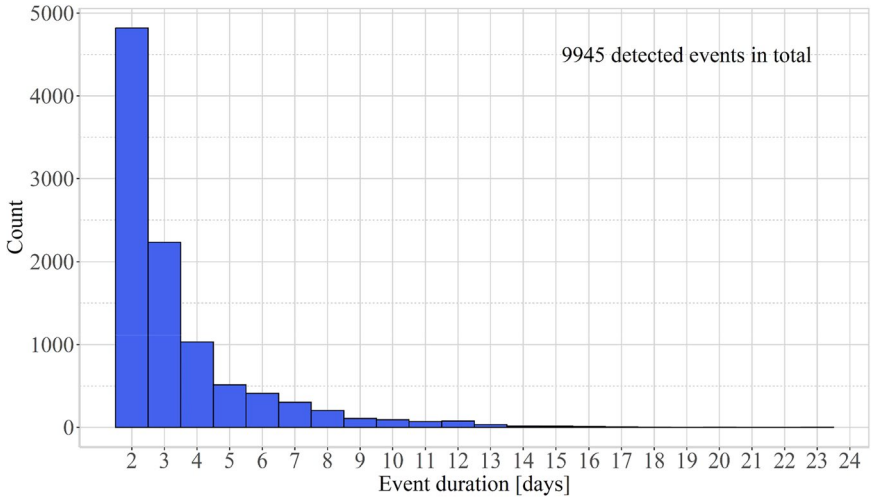


Fig. 7 Distribution of all cold snap event durations across all selected U.S. cities.

Figure 9 displays the mean outdoor air dry-bulb temperature during each selected cold snap event, ordered from the warmest to the coldest, illustrating the full range of event severity captured in the final dataset. The distribution shows a smooth gradient across events, with clear transitions at the category thresholds (0 °C, −15 °C, −26 °C, and −32 °C), which divide the events into Categories 1 through 4. Most events fall within Categories 1 and 2, while progressively fewer appear in the colder Categories 3 and 4, reflecting the climatic rarity of extremely severe cold snaps.

Heating degree hours, calculated relative to the 0 °C threshold, provide a measure of the cumulative thermal demand imposed by each cold snap event. When examined across event categories (Fig. 10), clear differences emerge in the magnitude and variability of heating demand. Category 1 events show relatively low heating degree hours, consistent with milder temperatures and shorter durations, while Categories 2 and 3 display progressively higher values and wider spreads as conditions intensify. Category 4 events exhibit both the highest median heating degree hours and the greatest variability, reflecting the extreme cold characteristic of the most severe events. To further isolate the intensity of cold stress independent of event duration, we calculated the overcooling degree (Fig. 11), defined as the heating degree hours divided by the number of hours in the event. This metric represents the average temperature deficit below the 0 °C threshold during each cold snap. When grouped by category, the differences in thermal severity become even more pronounced.

It is important to note that, because event categories are defined based on mean outdoor air temperature, both heating degree hours and overcooling degree are inherently correlated with the categorization criteria. As

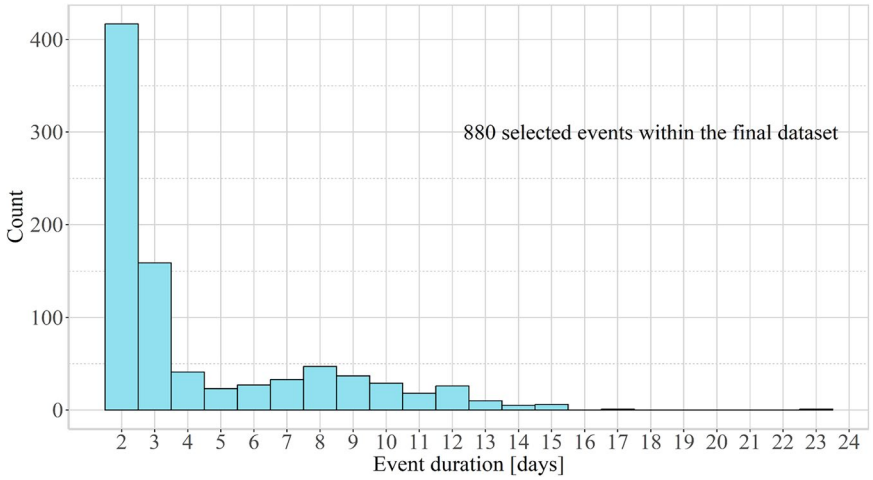


Fig. 8 Distribution of cold snap event durations across all selected U.S. cities within the final dataset.

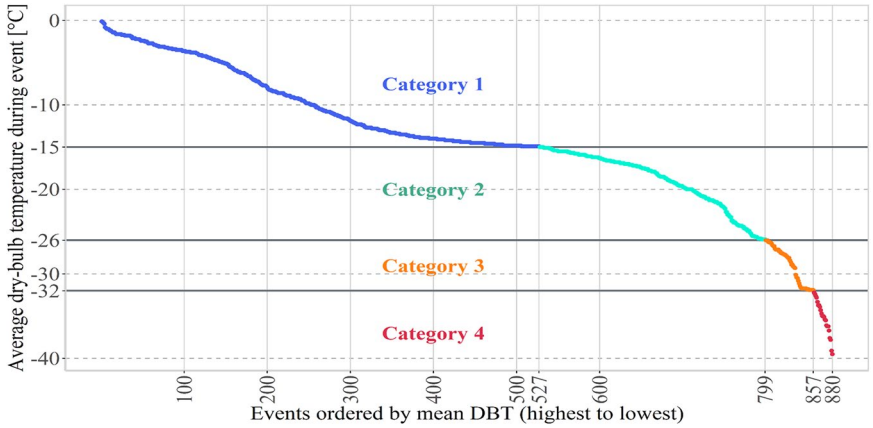


Fig. 9 Events ordered by the average outdoor air dry-bulb temperature with category thresholds.

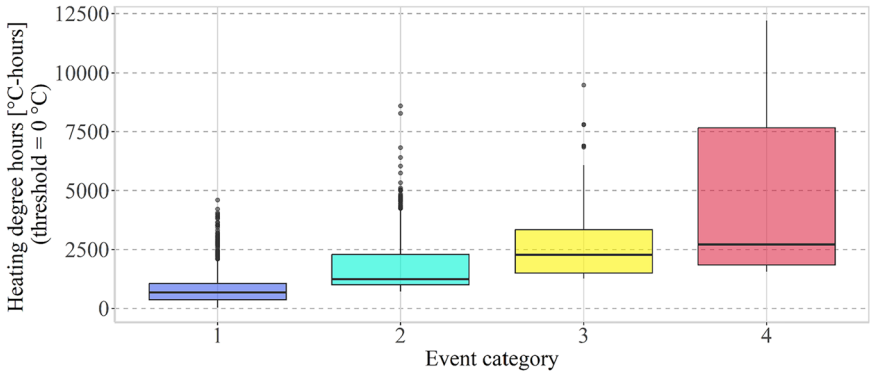


Fig. 10 Heating degree hours across cold snap categories in the final dataset.

such, these metrics are not intended to serve as independent validation of the category framework. Instead, they are provided as descriptive indicators to support interpretation of the dataset, offering complementary perspectives on cumulative thermal demand (HDH) and average cold intensity relative to the threshold (overcooling degree). Thus, these patterns illustrate how the defined categories correspond to differences in thermal stress, rather than validating the categorization itself, and help users compare events for building performance and resilience analyses.

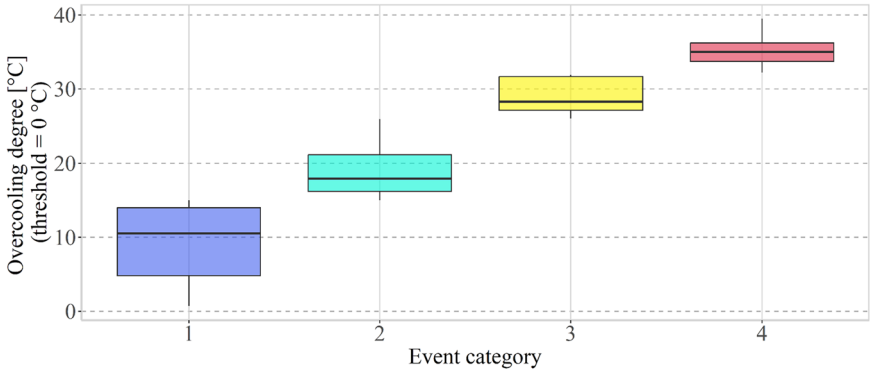


Fig. 11 Overcooling degree across cold snap categories in the final dataset.

Sensitivity analysis of percentile threshold selection for cold snap detection. To evaluate the robustness of the cold snap detection method to the choice of percentile threshold, a sensitivity analysis was conducted using three thresholds commonly adopted in extreme event detection: the 0.5th, 2.5th, and 5th percentiles of the distribution of daily average outdoor air dry-bulb temperature. These thresholds are consistent with those used in heatwave studies, where different percentile levels (e.g., 90th, 95th, and 97.5th percentiles). A notable example is Ouzeau *et al.*[15], who used a combination of these three thresholds. Machard *et al.*[7]. applied these thresholds to a larger dataset considering the climates of 15 major cities worldwide.

The comparison shows that the overall structure of the detected events is preserved across thresholds, but the characteristics of the resulting event sets vary systematically. Importantly, all results presented in this section are based on the full set of detected events for each threshold, prior to the event selection procedure described in Section 2.7.

The analysis revealed notable differences in the total number of detected events across thresholds. Using the 0.5th percentile results in a substantially smaller set of events (2,330 events), reflecting the identification of only the most rare and extreme cold conditions. In contrast, the 5th percentile yields a significantly larger number of detected events (16,609 events), including both a higher frequency of short-duration events and a greater number of longer-duration events due to the inclusion of milder conditions and the increased likelihood of merging adjacent cold periods. The 2.5th percentile produces an intermediate number of events (9,945 events), balancing event frequency and severity. As shown in Fig. 12, while all thresholds exhibit a right-skewed distribution dominated by short events, the 5th percentile results in a more pronounced tail of longer events, whereas the 0.5th percentile restricts detection to fewer, more extreme cases.

The distributions of overcooling degree (Fig. 13) show that all thresholds preserve the expected separation across event categories, but the 0.5th percentile emphasizes extreme outliers, particularly within category 4. The 5th percentile shifts the distributions toward milder conditions. Still, median results are close among all thresholds.

The 2.5th percentile provides an intermediate and balanced representation of cold snap events. It captures sufficiently severe conditions while maintaining a diverse set of events with varying durations and intensities. This balance is particularly important for resilience analyses, which require not only extreme cases but also a representative range of conditions to evaluate building performance under different levels of thermal stress.

Based on these results, the 2.5th percentile was selected as the default threshold in this study. However, the method is flexible and can be adapted to other percentile thresholds depending on the specific objectives of the analysis.

Usage Notes

The cold snap dataset[10] can be used directly in building energy simulation tools that accept weather files in EnergyPlus EPW format. Tools exist to convert EPW weather files to other formats used in major building simulation tools. Users should note that each EPW file contains a full year of weather data, but only a subset of this period corresponds to the cold snap event. Simulations should therefore be configured to run over the specific start and end dates listed in the accompanying spreadsheet. When integrating these files into modeling workflows, researchers may need to adjust simulation settings related to warm-up periods, plant operation, or HVAC schedules to ensure consistent behavior during the event window.

The code used to detect and characterize cold snap events is provided alongside the dataset and can be adapted for use with other locations or time periods, provided that sufficiently long and continuous temperature records are available. The scripts implement the full processing workflow—from dataset assembly to event detection, threshold calculation, and event characterization—and are written to allow modification of key parameters such as percentile thresholds, minimum event duration, and metric definitions. Researchers interested in extending the dataset to additional cities, updating it with future climate records, or applying the method to other regions can reuse or adapt the code with minimal changes.

Data availability

The dataset is publicly available at https://doi.org/10.5281/zenodo.20453235.

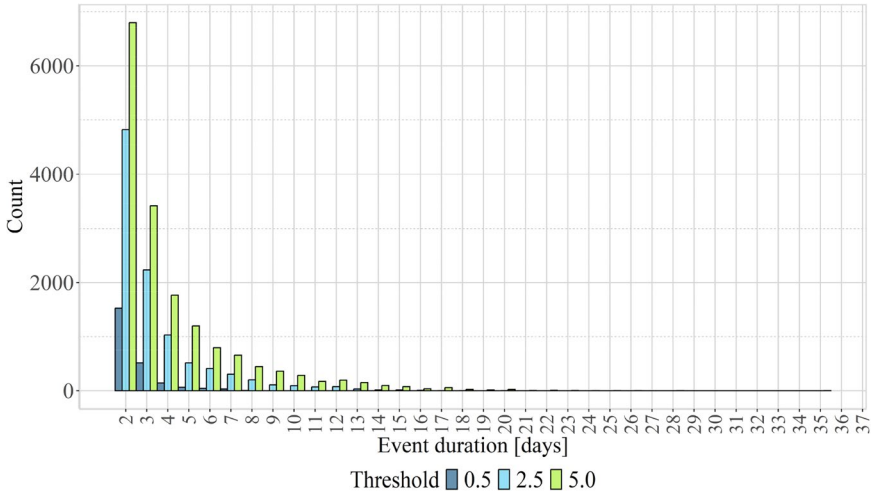


Fig. 12 Distribution of all cold snap event durations across all selected U.S. cities considering three detection thresholds: 0.5th, 2.5th, and 5th percentiles.

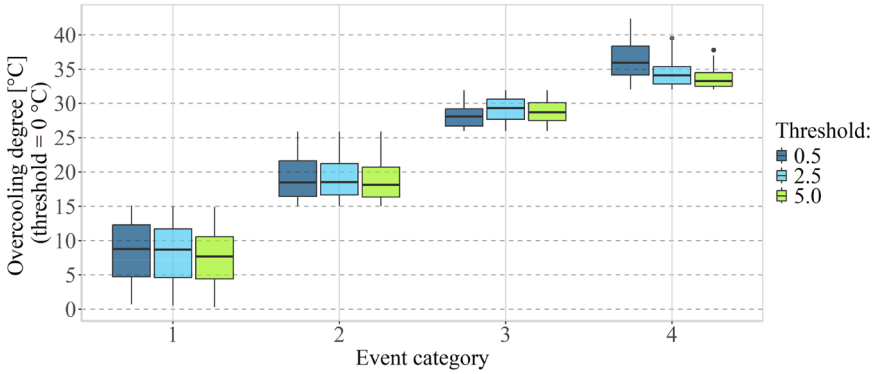


Fig. 13 Overcooling degree across cold snap categories, considering all detected events and three detection thresholds: 0.5th, 2.5th, and 5th percentiles.

Code availability

The scripts used for data preprocessing, event detection, and event characterization are publicly available at https://github.com/amandakrelling/extreme-cold-events-dataset.

Received: 7 March 2026; Accepted: 22 June 2026;
Published online: 27 June 2026

References

1. Gasparrini, A. *et al*. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet* **386**, 369–375 (2015).
2. Ye, Y. *et al*. Attribution of a record-breaking cold event in the historically warmest year of 2023 and assessing future risks. *Npj Clim. Atmospheric Sci.* **8**, 14 (2025).
3. Jay, O. *et al*. Reducing the health effects of hot weather and heat extremes: from personal cooling strategies to green cities. *The Lancet* **398**, 709–724 (2021).
4. Klepeis, N. E. *et al*. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J. Expo. Sci. Environ. Epidemiol.* **11**, 231–252 (2001).
5. Hong, T. *et al*. Ten questions concerning thermal resilience of buildings and occupants for climate adaptation. *Build. Environ.* **244**, 110806 (2023).
6. Yan, D., Wu, Y., Malik, J. & Hong, T. Ten questions on future and extreme weather data for building simulation and analysis in a changing climate. *Build. Environ.* **269**, 112461 (2025).
7. Machard, A. *et al*. Typical and extreme weather datasets for studying the resilience of buildings to climate change and heatwaves. *Sci. Data* **11**, 531 (2024).
8. CIBSE. *TM49 Design Summer Years for London*. 29, https://www.cibse.org/knowledge-research/knowledge-portal/technical-memorandum-49-design-summer-years-for-london-2014-pdf (2014).
9. Nik, V. M. Making energy simulation easier for future climate – Synthesizing typical and extreme weather data sets out of regional climate models (RCMs). *Appl. Energy* **177**, 204–226 (2016).
10. Krelling, A. F., Malik, J. & Hong, T. Extreme cold events dataset across 217 U.S. cities. *Zenodo* https://doi.org/10.5281/zenodo.20453235 (2026).

11. Bracht, M. K. *et al*. Multiple regional climate model projections to assess building thermal performance in Brazil: Understanding the uncertainty. *J. Build. Eng.* **88**, 109248 (2024).
12. IPCC. *AR6 Climate Change 2021: The Physical Science Basis*. 3949, https://www.ipcc.ch/report/ar6/wg1/#FullReport, https://doi.org/10.1017/9781009157896 (2021).
13. National Laboratory of the Rockies. NSRDB: National Solar Radiation Database. https://nsrdb.nrel.gov/data-sets/archives.
14. NASA Langley Research Center. NASA POWER Prediction Of Worldwide Energy Resources (POWER). https://power.larc.nasa.gov/ (2025).
15. Ouzeau, G., Soubeyroux, J.-M., Schneider, M., Vautard, R. & Planton, S. Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. *Clim. Serv.* **4**, 1–12 (2016).
16. Gordon, J. R. *On Establishing a Temperature Alert Threshold for Freezing Water Pipes*. 1–4 (1995).
17. Wu, J. *et al*. Human physiological responses of exposure to extremely cold environments. *J. Therm. Biol.* **98**, 102933 (2021).
18. National Weather Service. Understanding Wind Chill. https://www.weather.gov/safety/cold-wind-chill-chart.
19. Krelling, A. F., Jiang, Y., Sun, K., LaFrance, M. & Hong, T. High-performance windows improve thermal survivability of occupants during cold snaps. *Energy Built Environ*. S2666123325000509, https://doi.org/10.1016/j.enbenv.2025.05.014 (2025).
20. Antonopoulos, C., Gilbride, T., Margiotta, E. & Kaltreider, C. *Guide to Determining Climate Zone by County: Building America and IECC 2021 Updates*. PNNL-33270, 1893981, https://www.osti.gov/servlets/purl/1893981/ (2022).

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Critical Minerals & Energy Innovation (CMEI) under Contract No. DE-AC02-05CH11231. Authors thank the support and guidance from Christopher Perry of the Building Technologies Office.

Author contributions

Amanda Krelling led data curation, contributed to the method, and wrote the manuscript. Jeetika Malik contributed to the method and reviewed and edited the manuscript. Tianzhen Hong proposed the method, supervised the research effort, reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026