2021

A Surrogate Weather Generator for Estimating Natural Gas Design Day Conditions

David J. Kaftan

George Corliss

Richard J. Povinelli

Ronald H. Brown

energies

MDPI

Article

A Surrogate Weather Generator for Estimating Natural Gas Design Day Conditions

David Kaftan *, George F. Corliss, Richard J. Povinelli and Ronald H. Brown

Marquette Energy Analytics, Marquette University, Milwaukee, WI 53202, USA;
george.corliss@marquetteenergyanalytics.com (G.F.C.); richard.povinelli@marquettte.edu (R.J.P.);
ronald.brown@marquetteenergyanalytics.com (R.H.B.)
* Correspondence: david.kaftan@marquetteenergyanalytics.com

Abstract: Natural gas customers rely upon utilities to provide gas for heating in the coldest parts of winter. Heating capacity is expensive, so utilities and end users (represented by commissions) must agree on the coldest day on which a utility is expected to meet demand. The return period of such a day is long relative to the amount of weather data that are typically available. This paper develops a weather resampling method called the Surrogate Weather Resampler, which creates a large dataset to support analysis of extremely infrequent events. While most current methods for generating weather data are based on simulation, this method resamples the deviations from typical weather. The paper also shows how extreme temperatures are strongly correlated to the demand for natural gas. The Surrogate Weather Resampler was compared in-sample and out-of-sample to the WeaGETS weather generator using both the Kolmogorov–Smirnov test and an exceedance-based test for cold weather generation. A naïve benchmark was also examined. These methods studied weather data from the National Oceanic and Atmospheric Administration and AccuWeather. Weather data were collected for 33 weather stations across North America, with 69 years of data from each weather station. We show that the Surrogate Weather Resampler can reproduce the cold tail of distribution better than the naïve benchmark and WeaGETS.

Keywords: weather generator; design day conditions; extreme cold temperatures

check for updates

Citation: Kaftan, D.; Corliss, G.F.; Povinelli, R.J.; Brown, R.H. A Surrogate Weather Generator for Estimating Natural Gas Design Day Conditions. *Energies* **2021**, *14*, 7118. https://doi.org/10.3390/en14217118

Academic Editor: Carlos Guedes Soares

Received: 1 September 2021
Accepted: 25 October 2021
Published: 1 November 2021

1. Introduction

This manuscript presents a novel extreme weather generator. Our Surrogate Weather Resampler (SWR) provides substantially more examples of very cold days by a translation method that transforms near extreme cold days to have a similar statistical distribution as the coldest days. This is performed by examining weather conditions during the whole of winter. We compared our SWR against a naïve benchmark that estimates the coldest days without examining the entire winter and against the WeaGETS weather simulator [1]. The results show that the new SWR outperforms both. The contributions of this manuscript are the novel SWR, a comparison of the three weather generators on 33 weather stations each with 69 years of data, and a description of how such weather is used by natural gas utilities in determining design day conditions, which is the coldest day on which a utility is obligated to meet demand [2]. While weather generators are commonly tested in their ability to reproduce extremes, this is, to the best of our knowledge, the first work comparing methods for determining 1-in-N conditions for the natural gas industry.

As a motivation for this work, we need only recall the cold wave of early 2019 that broke temperature records across the American Midwest. This resulted in record-breaking natural gas demand [3] that put stress on the gas utility infrastructure, i.e., their ability to provide enough gas to their customers during such extreme weather conditions. In the winter of 2021, some utilities in Texas were unable to provide enough gas during another cold event resulting in electric utilities losing their ability to run their natural gas-fired plants [4]. Being able to have the appropriate infrastructure requires accurate estimates of design day conditions.

The gas supply planning organization within a utility is responsible for ensuring that sufficient natural gas can be delivered to meet the customers' demands, especially critical on extremely cold days when the customers' needs are high. Utilities and regulatory commissions must agree on the weather characteristics of the design day—known as the design day conditions. Natural gas infrastructure is expensive, so determining the extremity of weather for which a utility must prepare requires balancing the cost of the infrastructure and risk of demand exceeding capacity. Once a level of risk is agreed upon, utilities must determine the design day conditions that correspond to that risk. Determining the design day conditions is difficult because the likelihood of them occurring (i.e., once in 30 years) is very small compared to the amount of data available. The SWR generates data that accurately represents the extremely cold weather of an area allowing for better estimates of design day conditions.

To that end, the paper defines design day conditions and describes the current methods for determining the likelihood of such days. It continues to propose a novel method for creating larger weather datasets and evaluates our method's impact on choosing the extreme weather used for the design day.

The remainder of the paper is organized as follows. The next section provides background on weather generators and design day conditions. Section 3 describes the Surrogate Weather Resampler (SWR) method. In Section 4, we present the dataset, discuss preprocessing, and compare the SWR method against the naïve approach and WeaGETS. Section 5 discusses the results for both in-sample and out-of-sample experiments, and Section 6 provides a conclusion.

2. Background

2.1. Generating Extreme Weather

There are several ways to generate larger weather datasets. In Navigant's design day study for Enbridge Gas Distribution [5], they modeled wind speed and heating degree day (HDD), which is a nonlinear transformation of temperature. Let T be the temperature and T_{ref} be the reference temperature, then

$$HDD^{T_{ref}} = \max\left(0, T_{ref} - T\right) \tag{1}$$

Typically, T_{ref} is set to 65° F, which is the approximate temperature at which natural gas customers will turn on their furnaces. Navigant built three parameter linear models for wind speed and HDD. Each model had an intercept term, an autoregressive term, and a monthly dummy variable. They fit a log-normal distribution to the error of these models then sampled from the error distribution using Oracle's Crystal Ball Monte Carlo simulation software [6]. Using the Monte Carlo error samples, they generated HDD and wind speed distributions using their linear models.

Semenov uses a stochastic weather generator (LARS-WG) to build 300-year datasets to which he fit the generalized extreme value distribution [7]. Semenov examines the 10- and 20-year return periods for three weather conditions: maximum temperature, heatwave duration, and rainfall. He looks at 20 weather stations of various climate zones in Europe (10), North America (8), Australia (1), and New Zealand (1). His method fit the average maximum temperature, rainfall, and heatwave duration well. The 10 and 20 returns fit reasonably well but with greater variance.

Tebaldi et al. use nine atmosphere–ocean generalized circulation models to simulate weather data [8]. They identified five indicators of temperature-related extremes and five indicators of precipitation extremes. They examined the trends from 1960 to 2000 and then used the generalized circulation models to simulate weather for the 21st century predicting that extreme cold events would increase.

Chen et al. simulated minimum and maximum temperatures and daily precipitation using linear autoregressive models [1]. Their weather generator (WeaGETS) outperforms popular WGEN and CLIGEN weather generators in reproducing temperature data. However, WeaGETS inaccurately recreates cold quantiles of temperature datasets and struggles to reproduce the minimum temperature of a 60-year dataset. In an experiment with two weather stations, WeaGETS simulated extreme cold temperatures that were much colder than occurred.

WeaGETS is one of the methods to which we compare our SWR approach. To that end, we next describe in detail how WeaGETS generates temperatures. The WeaGETS algorithm uses a linear autoregression model to generate daily minimum and daily maximum temperature. First, residual temperature time series were created by subtracting daily mean temperature from the maximum and minimum temperature time series and dividing by the standard deviation of the temperature. A linear autoregressive model was fit to the residual time series.

The minimum and maximum temperatures are generated using the following equations. Let $T_{\max}$ and $T_{\min}$ be the generated maximum and minimum temperatures, respectively. Let $\mu_{\min}$ and $\sigma_{\min}$ be the mean and standard deviation of the minimum temperatures. Similarly, let $\mu_{\max}$ and $\sigma_{\max}$ be the mean and standard deviation of the maximum temperatures. Finally, let $r_{\min}$ and $r_{\max}$ be the residual minimum and maximum temperature generated from the linear autoregressive model, respectively. If $\sigma_{\max} \geq \sigma_{\min}$, then

$$T_{\min} = \mu_{\min} + \sigma_{\min} r_{\min} \tag{2}$$

$$T_{\max} = T_{\min} + (\mu_{\max} - \mu_{\min}) + \sqrt{\sigma_{\max}^2 - \sigma_{\min}^2} \times r_{\max} \tag{3}$$

If $\sigma_{\max} < \sigma_{\min}$ then

$$T_{\max} = \mu_{\max} + \sigma_{\max} r_{\max} \tag{4}$$

$$T_{\min} = T_{\max} - (\mu_{\max} - \mu_{\min}) - \sqrt{\sigma_{\min}^2 - \sigma_{\max}^2} \times r_{\min} \tag{5}$$

Next, low-frequency variability is modeled and removed using a power spectral approach. For more details on the WeaGETS method, refer to [1].

2.2. Design Day Conditions

Oliver et al. [9] set out three ways that utilities define design day conditions: the quantifiable metric of extreme weather (i.e., temperature); the odds of the event occurring (i.e., 1-in-30 years); and the period over which the extreme weather occurs (i.e., day, week).

The metric used for design day conditions must be predictive of gas demand and understandable by a utility's customer base. A predictive metric gives utilities an idea of how much gas will be demanded on design day conditions. On the other hand, an understandable metric allows utilities to make transparent obligations to their customers about how much demand they will be able to fill. Temperature is such a metric with both predictive and understandable properties. The relationship between temperature and gas demand is well documented [10–12], and the public understands the utility's commitment to heat their homes down to a certain temperature.

In [10], Sarak and Satman calculated HDD with reference temperatures of 15, 17, and 18.3 °C to estimate the natural gas demand needed for residential heating in Turkey. The temperature of 18.3 °C corresponds to 65 °F. Let U be the heat transfer coefficient, H the fuel heating value, η the heating system efficiency, HDY be the heating degree year, which is the sum of the heating degree days for a year, and n the number of residences. They modeled the energy consumption, Q, for heating as

$$Q = n \frac{U}{H\eta} \text{HDY} \tag{6}$$

Assuming 100% saturation of natural gas use, their forecast for total residential natural gas use in Turkey was 14.9 Gm^3 in 2023.

Aras and Aras build monthly natural gas demand models using a heating degree month (HDM), which is calculated as the sum of heating degree days in a month [11]. They used different nonlinear models for the heating season and non-heating seasons. Let d_t be the demand for month t, R_t a residual component then they model heating season as

$$d_t = e^{\beta_0} e^{\beta_1 t} e^{\beta_2 \mathrm{HDM}_t} e^{R_t} \tag{7}$$

For the non-heating season, they model demand as

$$d_t = \beta_0 t^{\beta_1} \mathrm{HDM}_t^{\beta_2} e^{R_t} \tag{8}$$

They achieved a yearly MAPE of 0.16% using the two nonlinear models.

In [12], Vitullo et al. showed the relationship between temperature and natural gas demand. They built both neural network and linear models for forecasting natural gas demand using temperature as an input. Their linear model includes terms for HDD at reference temperatures of 55 °F and 65 °F, cooling degree days, and change in HDD between days. Let T_t be the temperature at time t, T_{ref} be the reference temperature, then the cooling degree day is defined as

$$CDD_t^{T_{ref}} = \max\left(T_t - T_{ref}, 0\right) \tag{9}$$

Let $HDD_t^{T_{ref}}$ be the HDD at T_{ref}, then the change in HDD is

$$\Delta HDD_t^{T_{ref}} = HDD_t^{T_{ref}} - HDD_{t-1}^{T_{ref}} \tag{10}$$

Let d_t be the daily demand for day t; then, Vitullo et al. proposed a five-parameter model.

$$d_t = \beta_o + \beta_2 HDD_t^{65} + \beta_2 HDD_t^{55} + \beta_3 \Delta HDD_t^{65} + \beta_4 CDD_t^{65} \tag{11}$$

Additional components of more complicated versions of their model include wind adjusted HDD, day of the week indicators, and sine and cosine of the day of the week. Their neural network is ensembled with the linear model to produce the final forecasts.

Temperature can also be adjusted to be more predictive of gas demand. HDD, see Equation (1) above, is a temperature adjustment that takes advantage of the linear relationship of gas demand and temperatures below 65 °F. Temperature is highly correlated with natural gas consumption below 65 °F, as illustrated in Figure 1 below. The portion of the graph with temperatures less than 65 °F represents the effects of space heating or the heat load. The portion of the graph greater than 65 °F represents the baseload, where the baseload includes water heating and industrial use of natural gas. The use of natural gas for heating above 65 °F is minimal. For temperatures below 65 °F, there is a linear trend with an r^2 of -0.98, which indicates that as temperature decreases, natural gas consumption increases. Other weather variables such as wind do not have as high a correlation with natural gas consumption with an r^2 of 0.22. However, the wind can be used as an adjustment for temperature [12], as can prior day temperature [13], and solar radiation [9]. Figure 2 below illustrates the lack of data at extremely low temperatures and flows, as seen on the right-hand side of the figure. This dearth of data at the extremes, which is exactly where the design day conditions occur, further indicates the need for weather generators such as our SWR.

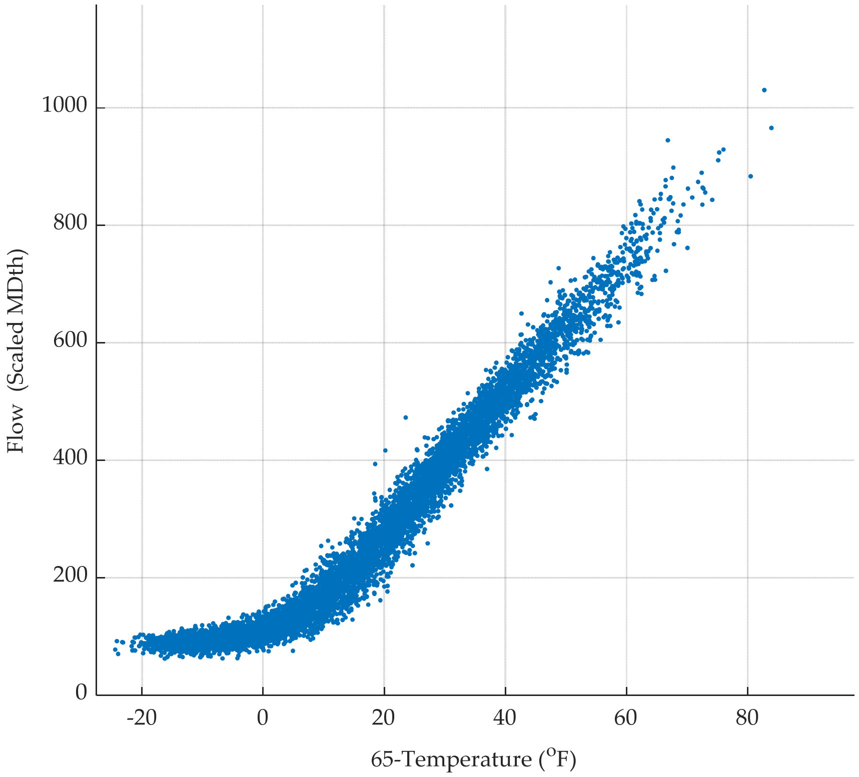

Figure 1. The blue dots represent the temperature at a particular flow. The 65-Temperature (°F) is related to the HDD but without the maximization term.

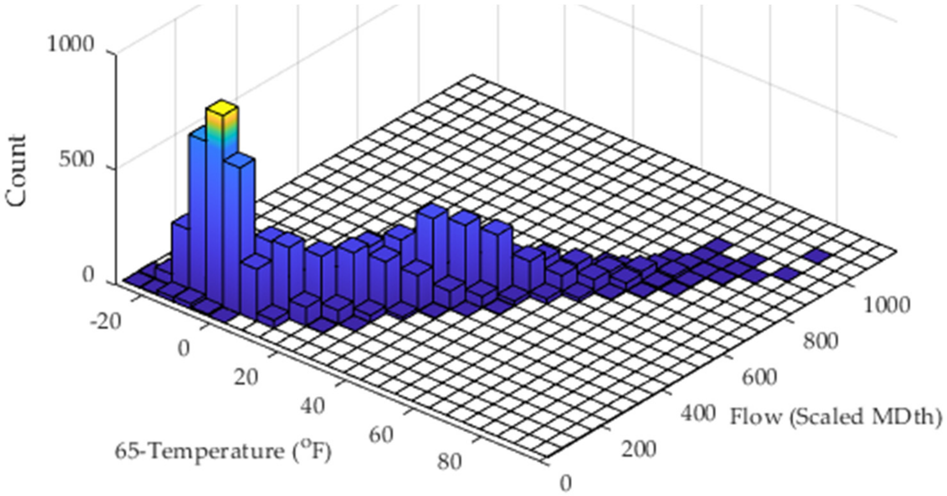

Figure 2. The number of occurrences at a given flow and temperature.

The probability of design day conditions occurring is characterized in one of two ways. Some utilities consider an event that has a 1/N probability of being below a given temperature (one or more times) in a given year [14]. Other utilities consider an event that is expected to be exceeded one time in N years [15]. Due to autocorrelation in weather temperature, these two probabilities yield different results. Table 1, below, describes the 1-in-N year probability used by different utilities according to four surveys. The Oliver survey counts the number of European countries using each range of N [9]. The other three surveys account for utilities in the United States. Integrated Resource Plans (IRPs) are utility reports of their design day conditions [14–18]. The American Gas Association (AGA) survey notes that eight additional utilities chose values of N between 25 and 100 but did not specify the exact values [19].

Table 1. The 1-in-N year probability that defines an extreme event varies from utility to utility. Below is the choice of N across several surveys of utilities.

Survey	N < 16	16 <= N < 26	26 <= N < 36	N >= 36
Oliver [9]	0	2	0	3
Navigant [5]	3	3	2	1
AGA [19]	5	10	27	0
IRPs [14–18]	1	0	2	2
Total	9	15	31	6

Finally, design day conditions are defined by their duration. Five days of cold weather stress a utility differently from one day of extreme cold. For the sake of simplicity, this paper focuses on single-day events.

2.3. Methods for Determining Design Day Conditions

The methods utilities use for determining design day conditions can be categorized into three approaches with increasing complexity: (1) choose the coldest day in the last N years; (2) fit a distribution to historical temperature and calculate the temperature with the return period of N years; and (3) generate a large weather dataset, then repeat (2).

The first approach is to set the design day conditions as the coldest recorded day in the last N years [16,17,19]. Not only does this lack statistical rigor, but it causes a serious logistical problem for utilities. When the coldest historic day falls out of the previous N years' window, the design day conditions can change dramatically. Since these conditions are used for long-term planning, a large change in conditions from one year to the next can have serious consequences. Some utilities simply choose the coldest day on record [15,18,19]. This avoids the aforementioned logistic problem. However, the likelihood of such a day occurring is no longer linked to a likelihood factor-such as once in N years. Rather it is arbitrarily tied to the length of the available weather dataset.

The second approach fits a distribution to historical weather [9,14,15]. These methods can use the entire history of recorded weather to determine the design day conditions. These methods are often derived from Extreme Value Theory [20], which has precedence as a tool for modeling meteorological extremes [7,21]. However, the datasets that are used to fit these distributions are limited—the National Oceanic and Atmospheric Administration (NOAA) has data back to 1973 for many weather stations [22]. Therefore, the extreme quantiles-such as 1-in-30 years-are being estimated from relatively small datasets.

The final approach is to use a weather simulator to create a large dataset of extreme events. The second approach is then applied to this larger dataset to fit a distribution to it. Section 2.1 discusses example weather generators. This manuscript presents a novel alternative to simulation methods called the Surrogate Weather Resampler (SWR). In contrast to the previously mentioned methods, the SWR focuses on reproducing the extreme cold tail of temperature to best aid in determining design day conditions.

3. Surrogate Weather Resampler Method

The SWR is inspired by the question: what if an autumn cold snap occurred in the winter? To answer this question, the SWR removes the seasonality of the autumn weather from the cold snap and applies the seasonality of winter weather. To that end, the SWR is split into two steps. First, a model is fit to temperature data with respect to the season. Second, the residuals from the model are resampled. The resampled residuals are input into the temperature model to generate surrogate temperatures.

3.1. Modeling Temperature

The SWR uses the time (t) dependent temperature model published by Breinl et al. [23]. Let T_t be the temperature, N_t be the normal temperature (the average temperature for that day in the year), and σ_t be the standard deviation for the normal temperature.

$$T_t = N_t + \sigma_t \varepsilon_t \tag{12}$$

N_t and σ_t are deterministic, seasonal, and repeated annually. The scaled deviation from normal temperature, ε_t, is the stochastic component of the model that is resampled, as explained below in Section 3.2. N_t and σ_t are calculated from a temperature time series. Given a long temperature time series, generally exceeding 50 years, the mean temperature for each day of the year (i.e., 30 December) is calculated as shown in Figure 3 below. This mean temperature is smoothed using a fifth-order Fourier series as a low pass filter to produce N_t. Such Fourier series smoothing of mean temperature to yield normal temperature is a common technique [23–25].

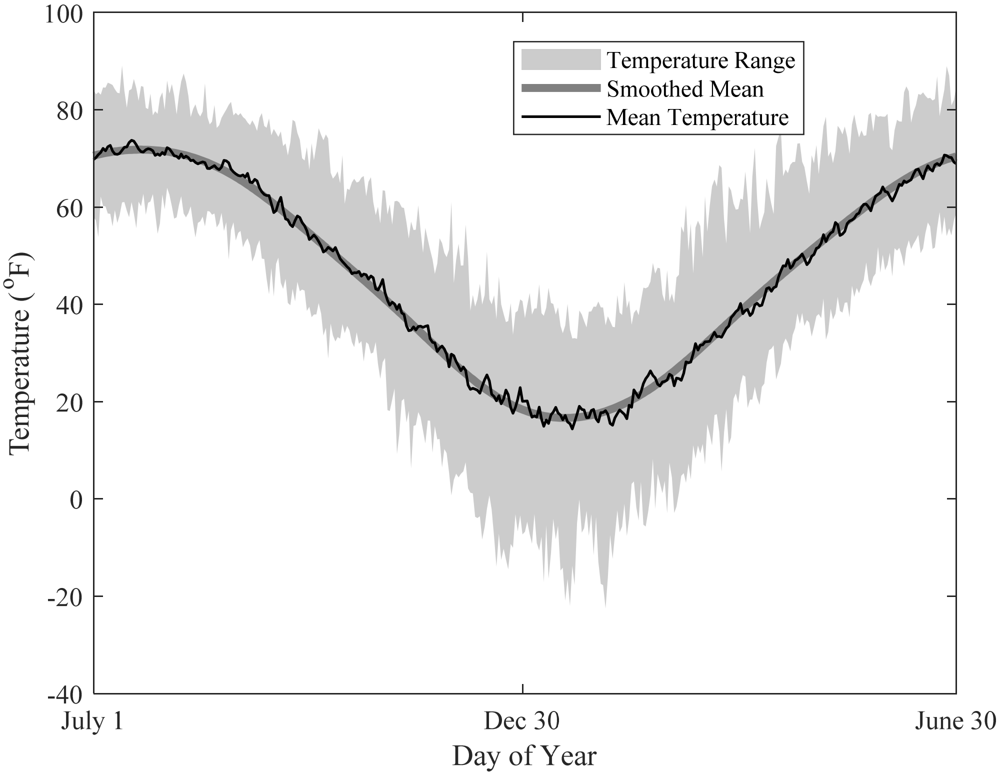

Figure 3. The light grey color, which shows temperature range, illustrates all the temperatures on a given day of the year. The thin black line is the mean temperature for each day of year. The thick grey line is the smoothed mean or normal temperature. The illustration is centered around December 30th, as the winter in the northern hemisphere contains the extreme cold temperatures in which we are interested.

The term σ_t is analogous to the daily temperature standard deviation but has three important transformations from the daily temperature standard deviation. First, the mean for each day of the year is replaced by N_t, which is the smoothed mean. Second, only temperatures colder than N_t are used in the calculation. As the temperature range in Figure 3 demonstrates, the variance of colder-than-normal temperatures is not necessarily equal to the variance of warmer-than-normal temperatures, which is particularly true in the winter. Moreover, the design day must occur on a day colder than normal, so the "lower" deviations are of primary interest. Let T_t be the temperatures on the t^{th} day of the year, τ_t be those temperatures less than N_t, $\overline{\tau}_t$ be the mean of τ_t, and n_t be the number of temperatures from the t^{th} day of the year, then

$$\sigma_{t,raw} = \sqrt{\frac{\sum\limits_{\tau_t \in (T_t < N_t)} (\tau_t - \overline{\tau}_t)^2}{n}} \tag{13}$$

Third, as performed in previous temperature generation studies [23], we smoothed the raw standard deviations ($\sigma_{t,raw}$) using a fifth-order Fourier series as a low pass filter to produce σ_t. This process is illustrated in Figure 4.

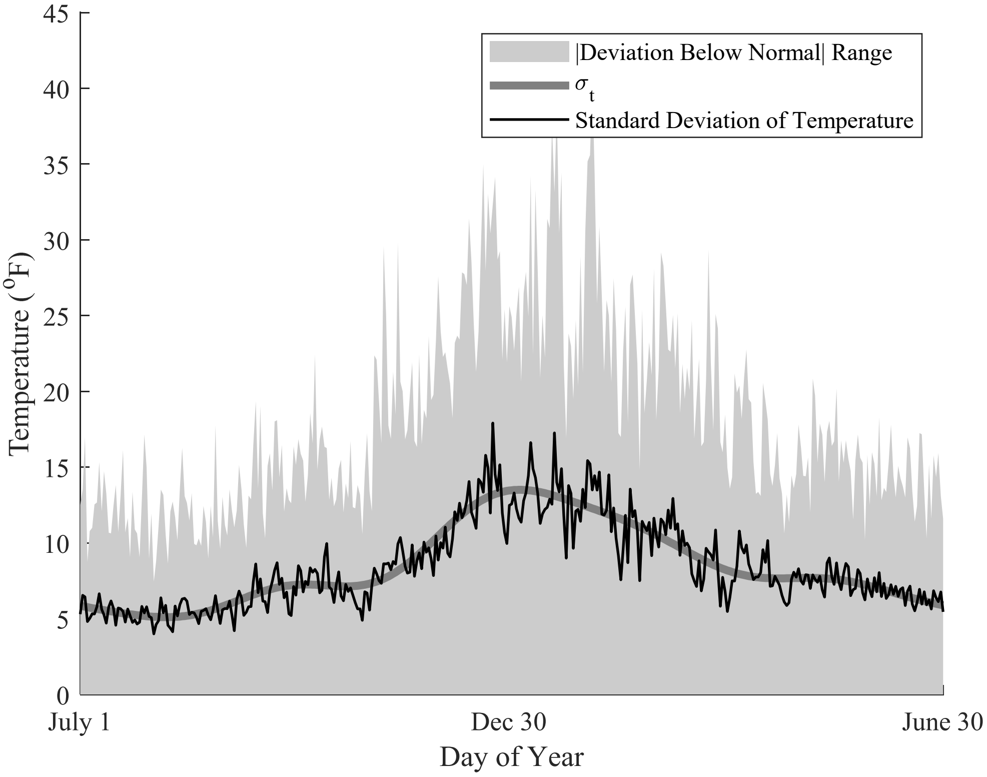

Figure 4. The process of calculating σ_t is illustrated above. The absolute value of the temperatures less than N_t (the smoothed mean or normal temperature) is plotted in light grey. Typically, as seen in this example, cold deviations from normal have higher variability in the winter. The standard deviations ($\sigma_{t,raw}$) of these temperatures are plotted in black. The $\sigma_{t,raw}$ are smoothed to yield σ_t illustrated by the thick grey line.

3.2. *Resampling the Scaled Deviation from Normal Temperature*

In a typical simulation method, the scaled deviation from normal temperature (ε_t) would be modeled and sampled. Instead, this paper proposes resampling ε_t from the empirical distribution. The SWR takes advantage of the non-seasonality of ε_t by introducing a time lag. The new model becomes

$$T_t = N_t + \sigma_t \varepsilon_{t-lag} \tag{14}$$

Lagging ε_t answers the following question: what if the weather patterns driving ε_t occur later in the year? By solving T_t with lag set to 1, a surrogate temperature dataset is created. Many surrogate datasets can be created by introducing different lags. By using every integer between −45 and 45 as lags, 90 surrogate datasets are created, increasing the amount of data by a factor of 90. This form of resampling preserves the autocorrelation of ε_t.

4. Experiments—Comparison of SWR, Naïve Benchmark, and WeaGETS

This section compares the SWR approach to a naïve resampling method and WeaGETS [1]. The naïve benchmark method is equivalent to the SWR method but without lags. We analyzed the SWR's ability to produce sensible 1-in-N conditions across several weather stations in the United States and Canada. We compared the SWR and WeaGETS to using raw weather data in determining 1-in-N conditions. We performed an in-sample test and an out-of-sample cross-validation test, and we considered these methods in the context of design day planning.

4.1. Dataset

Thirty-three stations across the United States and Canada were examined for testing the performance of the weather generators (see Table 2 and Figure 5). The stations were chosen for geographic diversity and dataset length. All stations recorded data from 1950 to 2018. Station data are acquired primarily through NOAA [21] and is available publicly. The NOAA data are supplemented with data from AccuWeather, where NOAA lacks history.

Table 2. The 33 stations used in this study span across North America, Alaska, and Hawaii. They are identifiable in NOAA datasets by their callsign.

Station Location	Callsign	Station Location	Callsign
Ottawa, ON	CYOW	Kansas City, MO	KMCI
Regina, SK	CYQR	Memphis, TN	KMEM
Vancouver, BC	CYVR	Miami, FL	KMIA
Winnipeg, MB	CYWG	Minneapolis, MN	KMSP
Calgary, AB	CYYC	New Orleans, LA	KMSY
Amarillo, TX	KAMA	New York, NY	KNYC
Aspen, CO	KASE	Pittsburg, PA	KPIT
Bakersfield, CA	KBFL	Pueblo, CO	KPUB
Nashville, TN	KBNA	Raleigh/Durham, NC	KRDU
Boston, MA	KBOS	San Antonio, TX	KSAT
Brownsville, TX	KBRO	Louisville, KY	KSDF
Corpus Christi, TX	KCRP	Seattle, WA	KSEA
Dallas-Fort Worth, TX	KDFW	Salt Lake City, UT	KSLC
Evansville, IN	KEVV	Tulsa, OK	KTUL
Fort Smith, AR	KFSM	King Salmon, AK	PAKN
Hays, KS	KHYS	Honolulu, HI	PHNL
Jackson, MS	KJAN		

The weather variable of interest is the average daily temperature. Ideally, this could be calculated by the average of hourly temperature readings—this method correlates best with natural gas demand [26] However, WeaGETS simulates daily minimum and maximum temperatures. Therefore, the average daily temperature is calculated by the average of the minimum and maximum for the purpose of comparison.

4.2. Data Preprocessing

Data are originally obtained in NOAA's ISH format. In general, NOAA's ISH data are recorded at an hourly frequency, close to on-the-hour. For ease in data processing, the data are normalized to always be at an hourly frequency, occurring on the hour. The ISH data are prepared by rounding each data point to the nearest hour. For example, a meteorological measurement made at 6:05 PM is moved to 6:00 PM. In cases where there are several records made within an hour, the record closest to the hour is chosen. For example, if two meteorological measurements were made at 6:05 PM and 6:10 PM, respectively, the 6:05 PM measurement is moved to 6:00 PM, and the measurement at 6:10 PM is discarded.

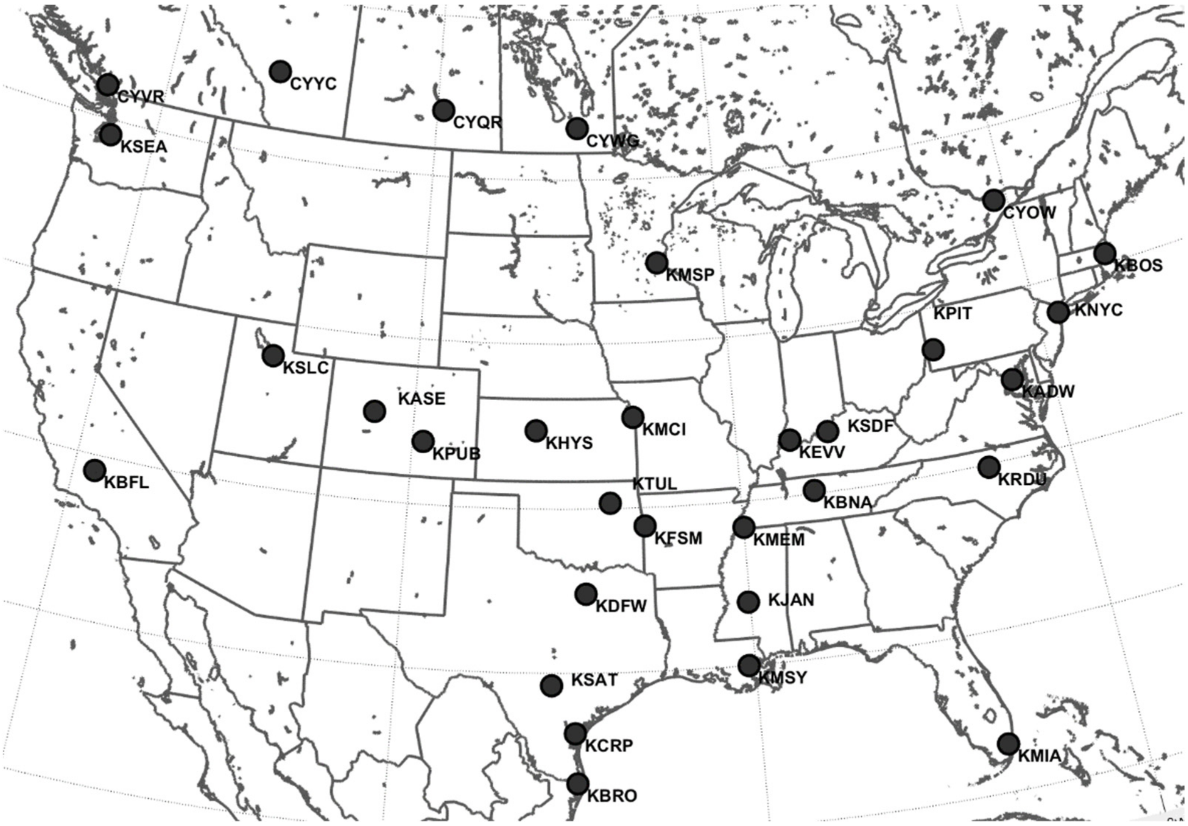

Figure 5. The stations were chosen with the intent of evenly spanning North America under the constraint of data availability. Not shown are PAKN in King Salmon, Alaska and PHNL in Honolulu, Hawaii.

After the dataset is normalized to hourly, daily summary statistics are calculated. In particular, the daily minimum and maximum are taken by choosing the minimum and maximum hourly temperature in a day. Ideally, the average daily temperature would be calculated by averaging all the hourly temperatures for a day. However, because WeaGETS generates minimum and maximum daily temperatures, daily average temperatures are estimated using the average of the daily maximum and minimum temperatures.

Rather than cleaning bad data points, station datasets were chosen based on their data quality. Station temperature data were graphed, and stations with obvious data quality issues were not included. The stations listed in Table 2 are the stations with acceptable quality. Special attention was paid to cold temperature data. If a cold outlier was found in the graph, that day's hourly temperature was analyzed; often, a cause of cold outliers is missing temperature data imputed as zero. Stations with this characteristic are not used.

In general, data are sourced as much as possible from NOAA ISH files, as they are publicly available. However, most NOAA ISH datasets do not have history back to 1950. In cases where an NOAA ISH dataset is otherwise usable but does not have history back to 1950, the data are supplemented by AccuWeather. This was an important compromise to achieve the geographic diversity shown in Figure 5, while maintaining a consistent time frame across stations.

4.3. In-Sample Experiment

It is common in weather generation studies to analyze the generated weather in-sample. In other words, the generated temperature is compared to the original temperature data used to train the generator. An in-sample study provides an important sanity check; the generated weather comes from the same distribution as the original weather.

Using the SWR for each station, 90 sets of surrogate daily average temperature data are generated (using time lags from −45 to +45). This amounts to roughly 6000 years of weather data per station: 90 sets of surrogate weather × 69 years of weather + the original dataset. By using WeaGETS, the equivalent number of years is generated. WeaGETS generates daily high and low temperatures. Each day's high and low temperatures are averaged, resulting in roughly 6000 years of daily average temperatures for each station.

The order statistic "coldest daily average temperature of each winter" was first calculated for each of the generated datasets by taking the minimum temperature of each of the generated winters. This was compared to the raw, measured order statistic using a two-sample Kolmogorov–Smirnov test (KS-test). The KS-test is a common tool in testing the validity of generated temperature datasets in-sample [27,28]. H_0 for the KS-test states that the generated order statistic and the original order statistic come from the same distribution-it is ideal not to reject H_0. Conveniently, using the annual minimum temperatures addresses Semenov's concerns for using the KS-test for generated weather data, as there is no annual periodicity nor daily autocorrelation represented in the annual minima [29].

1-in-N conditions are estimated by fitting a kernel density function to the generated weather for each station. The inverse cumulative density function was evaluated at $1/(30 \text{ years} \times 365)$ to determine the 1-in-30 condition; according to Table 1, 30 is a common choice of N. We compared the actual exceedances of the 1-in-30 condition to the expected exceedances (69 years in sample$/(30$ years $) = 2.3$ exceedances).

4.4. Out-of-Sample Experiment

In practice, the 1-in-N condition is used as a basis for planning for the next several years. It is therefore important to evaluate techniques for determining 1-in-N conditions out-of-sample. To this end, we conducted an experiment where we generated weather based on a subset of temperature data and compared that weather to the actual weather in the held-out years.

For each station, a 50-fold Monte Carlo cross-validation was performed [30]. Each fold randomly samples 30 years (without replacement) for the test set. The remaining years were used for the training set. An example of how the train and test data might be split for the first four folds is shown in Figure 6. Ideally, we would be able to perform cross-validation using a causal schema (the test set is chronologically after the training set). However, due to the small amount of data present relative to the extreme conditions being estimated, we used Monte Carlo cross-validation; it allowed us to perform enough folds to achieve a statistically significant result.

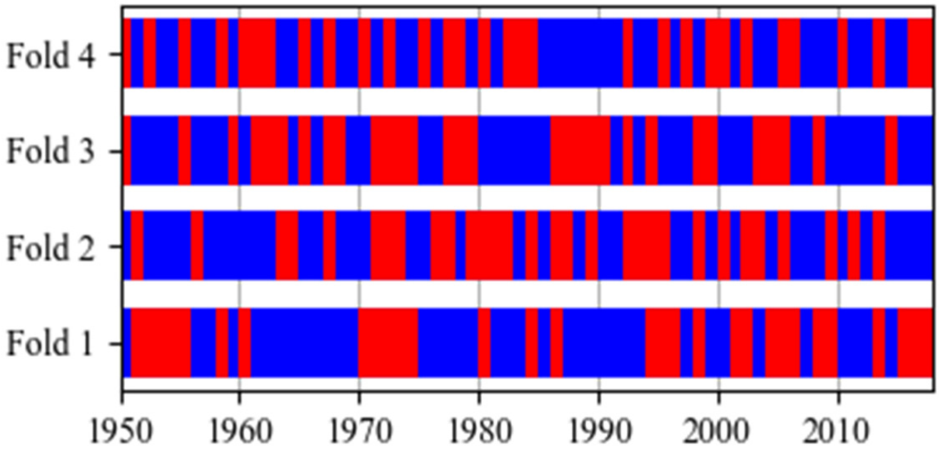

Figure 6. Example train-test splits of the cross-validation scheme are shown above. The blue bars represent years of data used for training. The red bars represent the years used for the testing set. The folds are chosen randomly; the figure above shows examples of what they might look like.

WeaGETS and the SWR were used to generate weather data from the training dataset. For each fold, 3367 years of weather are generated. The out-of-sample experiment allows an additional, naïve weather generator: resample the raw training data to have 3367 equivalent years of data. This can also be thought of as the SWR with 0 lags (as opposed to the 90 used for the SWR). Therefore, comparing the SWR to this naïve model asks the question: what impact does the novel resampling have on the 1-in-N conditions? This method is referred to as the naïve benchmark.

In the out-of-sample experiment, the same tests were performed: KS-test and exceedance test. The only difference is the data on which the weather generators were trained and tested. For the KS-test, all methods generate weather data using only information from the training set. The coldest annual temperatures of those generated sets were compared to the coldest annual temperatures from the held-out set, asking the question: how well do the generated datasets represent out-of-sample cold temperatures? For the exceedance test, we again generated weather using only information from the training set. We fit a kernel density function to the generated data, and from the kernel density function, we determined the temperature threshold expected to be exceeded once every 30 years. We counted how many times in the test set the threshold is exceeded.

5. Results

The results section is split into two parts. First, we analyze the in-sample experiment. Second, the out-of-sample experiment is examined.

5.1. In-Sample Results

The in-sample results are analyzed in two parts. First, we examined how well WeaGETS and the SWR reproduce the coldest annual temperatures. This provides a sanity check to show if the generated weather comes from the same distribution as the actual weather. Second, we performed a practitioner-based test, observing how frequently the 1-in-N temperatures estimated from each generated dataset exceeded.

5.1.1. Comparing Distributions of Cold Temperatures

At $p < 0.05$, SWR correctly did not reject H_0 (the generated and original datasets come from the same distribution) 33 out of 33 times on the weather stations shown in Table 2. This shows that SWR accurately recreates the distribution of the coldest annual temperatures. On the other hand, WeaGETS correctly accepts the H_0 only 1 out of 33 times. This result shows that our novel SWR method generates cold tails in better agreement with the empirical distributions than WeaGETS. In other words, the coldest temperatures generated by SWR align with what has happened over the past 69 years. Since the design day conditions are estimated from the generated datasets, it is essential the SWR generates cold weather in line with reality.

As previously described, the two-sample KS-test compares the cumulative density functions (cdf) of different empirical distributions. Figure 7, below, shows the cumulative density plots for the generated and measured data for the KFSM weather stations, which was chosen because it illustrates typical results. For the KFSM weather station, the cdf of the SWR (black) sits on top of the cdf of the measured temperature data (grey). The KS-test between the SWR and the measured temperature data was accepted for the KFSM weather station. However, the cdf generated by WeaGETS (gold) differs significantly from the measured temperature data for KFSM. The corresponding KS-test was rejected. Figure 8 illustrates weather station KMIA, which is the sole occurrence where the WeaGETS (gold) cdf does not differ significantly from the measured data (grey).

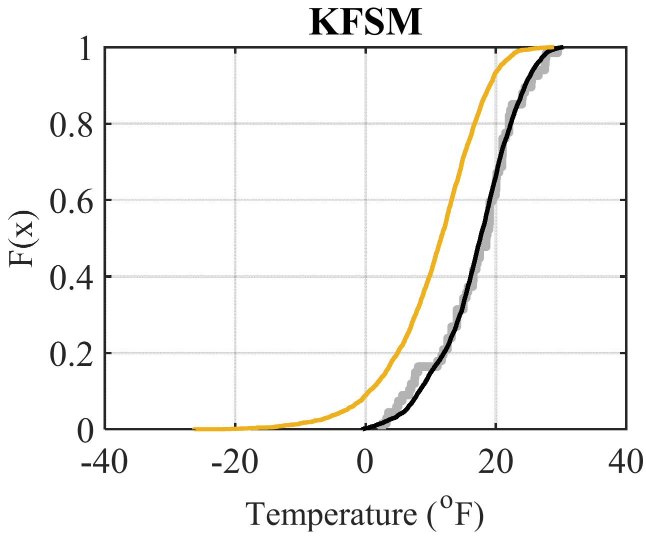

Figure 7. CDF of the annual coldest temperatures for station KFSM. The measured annual coldest temperature cdf (grey) is compared to the coldest annual temperatures generated by SWR (black) and WeaGETS (gold). For SWR, the KS-Test accepts that the measured and generated distributions are the same. This is observed by the close correspondence between measured coldest temperature cdf (grey) and the coldest annual temperatures generated by SWR (black). WeaGETS (gold) fails to capture the coldest temperature cdf, and the KS-Test H_0 is rejected for WeaGETS. i.e., the generated data do not match the measured distribution.

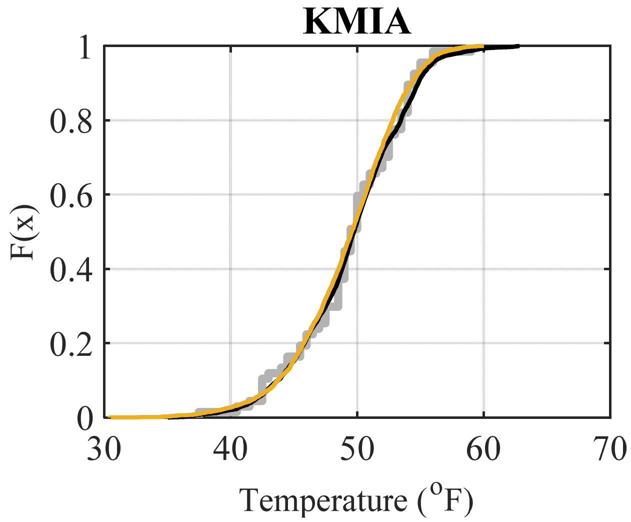

Figure 8. CDF plot of annual coldest temperatures for station KMIA. The measured annual coldest temperature cdf (grey) is compared to the coldest annual temperatures generated by SWR (black) and WeaGETS (gold). KMIA is the only station for which the KS-Test is accepted by both methods, i.e., the generated data are not significantly different from the empirical.

5.1.2. Frequency of Exceedance Test

While the KS-Test is a standard method for evaluating the performance of generated weather, it might mean little to a utility planning their design day conditions. A utility would ask the question, "what is my design day temperature, and how many times has it been exceeded in history?" Table 3 answers this question.

With 69 years of data, the threshold should be exceeded around two times per station. In general, our SWR method produces a temperature that is exceeded as often as expected, while WeaGETS does not. When aggregated across all stations, the proposed method is exceeded 17 fewer times than expected; the estimated conditions are slightly biased cold.

In comparison, the temperature estimated from the WeaGETS dataset is exceeded 48 fewer times than expected—the estimated conditions are consistently biased cold.

Table 3. The number of times a threshold generated by either the SWR or WeaGETS is exceeded. This is compared to the expected number of exceedances and summarized across stations.

Station	Expected	SWR	WeaGETS	Station	Expected	SWR	WeaGETS
CYOW	2.3	3	0	KMCI	2.3	2	0
CYQR	2.3	0	0	KMEM	2.3	2	0
CYVR	2.3	2	6	KMIA	2.3	2	1
CYWG	2.3	1	0	KMSP	2.3	0	0
CYYC	2.3	0	0	KMSY	2.3	5	0
KAMA	2.3	1	0	KNYC	2.3	2	0
KASE	2.3	2	3	KPIT	2.3	3	0
KBFL	2.3	2	0	KPUB	2.3	0	0
KBNA	2.3	1	1	KRDU	2.3	3	0
KBOS	2.3	1	0	KSAT	2.3	3	0
KBRO	2.3	2	0	KSDF	2.3	2	1
KCRP	2.3	2	0	KSEA	2.3	0	12
KDFW	2.3	1	0	KSLC	2.3	1	3
KEVV	2.3	1	0	KTUL	2.3	1	0
KFSM	2.3	3	0	PAKN	2.3	3	0
KHYS	2.3	1	0	PHNL	2.3	2	0
KJAN	2.3	4	0	Sum	74.9	58	27

While this test lacks precedence in the literature, it is important for communication to the public. If a utility can say, for example, that the estimated 1-in-30 condition has been exceeded twice in the past 60 years, even people without statistical backgrounds probably would consider such a design day condition to be reasonable.

5.2. Out of Sample Results

5.2.1. Out of Sample KS-Test

For the out-of-sample test, we analyzed the conditions generated by the SWR, WeaGETS, and a naïve benchmark. In contrast to the KS-Tests used in the in-sample experiment, 50 KS-Tests were performed per station: one for each fold. The count of those tests for which H_0 was not rejected is shown in Table 4. The SWR rejects fewer KS-Tests than any other method-in 1392 out of 1650 tests, H_0 was correctly not rejected. The naïve benchmark rejects H_0 92 more times than the SWR. This difference is significant in a one-tailed proportions test ($p < 0.0001$); the naïve baseline is significantly more likely to reject the KS-Test. WeaGETS consistently rejects H_0. It is worth noting that the KS-Tests fail more frequently in the out-of-sample test than the in-sample test. Where the in-sample test acted as a sanity check, the out-of-sample test evaluated each method's ability to generalize data the generators had not yet seen; this makes it a more difficult test.

5.2.2. Out of Sample Frequency of Exceedance Test

Table 5 below shows the average number of exceedances for each station across all 50 folds. The expected number of exceedances for each station is 1. Therefore, across all 33 stations, there should be an average of 33 exceedances per fold. Across all stations and folds, the naïve benchmark performs best, averaging 39.42 exceedances per fold—6.42 more exceedances than expected, meaning the actual weather in the test set is colder than the naïve baseline expects. The SWR performed slightly worse, averaging 40.26 exceedances per fold. Again, the weather in the test set is colder than expected by the SWR. However, the difference in the total count of exceedances between the SWR and naïve benchmark is not significant (one-tailed proportions test with $p = 0.25$). WeaGETS performs poorly on this task by consistently setting thresholds too cold, consistent with its original paper's findings [1].

Table 4. KS-Test results for out-of-sample tests. The values in the table represent the number of folds for which H_0 is not rejected. The SWR failed the KS-Test 258 out of 1650 times, compared to 350 failures for the naïve benchmark. These proportions are significantly different in a one-tailed proportions test ($p < 0.0001$). WeaGETS fails the KS-Test in every fold.

Station	Naïve	SWR	WeaGETS	Station	Naïve	SWR	WeaGETS
CYOW	40	37	0	KMCI	42	40	0
CYQR	40	44	0	KMEM	38	40	0
CYVR	40	45	0	KMIA	43	44	0
CYWG	42	45	0	KMSP	32	39	0
CYYC	39	40	0	KMSY	40	39	0
KAMA	39	42	0	KNYC	40	39	0
KASE	42	43	0	KPIT	41	43	0
KBFL	42	44	0	KPUB	41	46	0
KBNA	40	42	0	KRDU	40	42	0
KBOS	44	48	0	KSAT	39	39	0
KBRO	38	46	0	KSDF	40	40	0
KCRP	37	42	0	KSEA	41	42	0
KDFW	43	44	0	KSLC	41	44	0
KEVV	37	40	0	KTUL	35	42	0
KFSM	36	38	0	PAKN	37	44	0
KHYS	36	46	0	PHNL	39	41	0
KJAN	36	42	0	Sum	1300	1392	0

Table 5. Average number of exceedances for each station. This table is the out-of-sample analog to Table 3. For each station, the average number of exceedances across all folds was generated.

Station	Naïve	SWR	WeaGETS	Station	Naïve	SWR	WeaGETS
CYOW	1.02	1.08	0.00	KMCI	1.50	1.20	0.00
CYQR	0.04	0.10	0.00	KMEM	0.92	1.64	0.00
CYVR	3.38	1.56	3.40	KMIA	0.80	1.22	0.50
CYWG	0.78	0.52	0.00	KMSP	0.32	0.08	0.00
CYYC	0.36	0.00	0.00	KMSY	1.40	2.46	0.10
KAMA	0.84	1.00	0.00	KNYC	1.00	0.94	0.00
KASE	1.46	1.32	1.44	KPIT	1.42	1.86	0.00
KBFL	1.46	1.04	0.00	KPUB	1.02	0.40	0.58
KBNA	1.42	1.90	0.54	KRDU	0.56	1.48	0.00
KBOS	1.10	1.02	0.00	KSAT	0.90	1.44	0.18
KBRO	1.46	1.80	0.18	KSDF	1.22	1.20	0.30
KCRP	0.92	1.52	0.08	KSEA	1.62	0.52	4.64
KDFW	1.78	1.80	0.00	KSLC	1.98	1.48	1.48
KEVV	0.92	1.00	0.14	KTUL	0.64	0.66	0.00
KFSM	1.02	1.86	0.00	PAKN	1.22	1.28	0.00
KHYS	1.40	0.78	0.00	PHNL	1.90	1.74	0.06
KJAN	1.64	2.36	0.00	Sum	39.42	40.26	13.62

One of the motivations for the out-of-sample test is to demonstrate the difference between the naïve benchmark and the SWR. In other words, what is the impact that the SWR has on the 1-in-N temperature? To this end, the differences between the 1-in-N temperatures generated by the SWR and the naïve benchmark were calculated for all folds and all stations. The histogram of these differences is displayed in Figure 9. The standard deviation is 1.9 °F, which can make a large difference in design day planning. For example, the naïve benchmark and the SWR estimated 1-in-30 conditions with an average difference of 1.4 °F for KSAT. This roughly corresponds to a 3% difference in heat load. Therefore, estimating the 1-in-30 conditions using temperatures generated by the SWR does make a substantial difference.

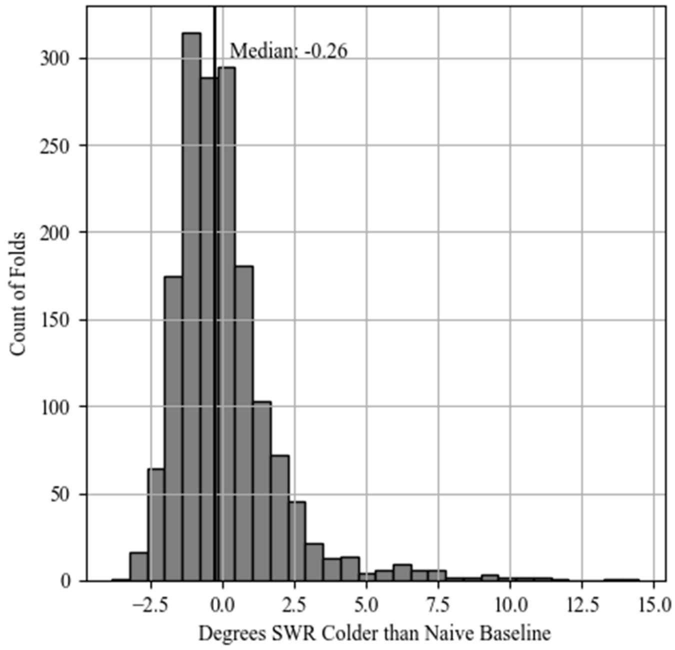

Figure 9. Difference between conditions generated by naïve benchmark and SWR.

It is important for 1-in-N conditions to be relatively stable over time—it is difficult to make long-term plans based on conditions that frequently change. Because each fold determines the 1-in-30 threshold based on a random subset of data, the variance of thresholds across difference folds represents each method's sensitivity to changing data. In practice, one year of data are added to the training set each year—this test will exaggerate the volatility of threshold estimation compared to how thresholds will change in practice.

Figure 10 shows the spread of variances across different stations. WeaGETS varies the least over time, indicating it is not overly sensitive to new data. The naïve benchmark and the SWR vary more over time, indicating they are more sensitive to new data.

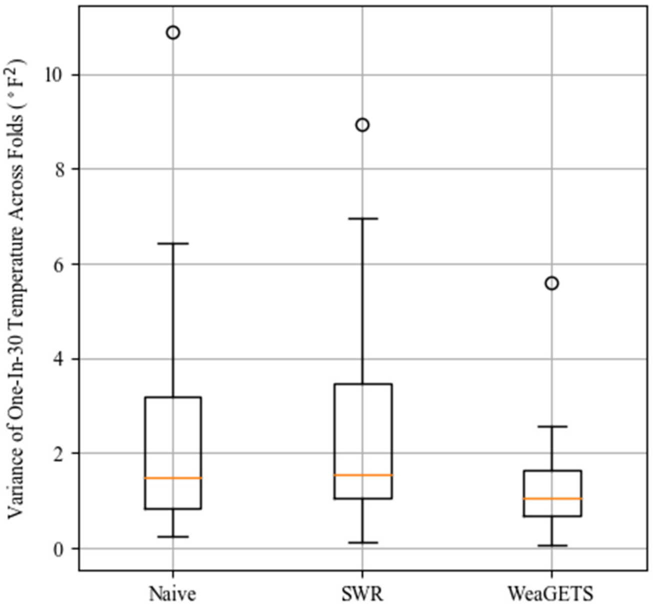

Figure 10. Variance of 1-in-30 temperatures across folds. For each station, the variance of the 1-in-30 temperatures is calculated. Variance represents sensitivity to new data. The naïve benchmark and the SWR have the largest spread with a mean variance of 2.4 °F^2. WeaGETS has the lowest variance at 1.3 °F^2.

6. Conclusions

The SWR accurately reproduces the cold tail of the temperature distribution. The two sample KS-test shows that the SWR generates temperature data that have the same annual minima characteristics as the measured temperature data. For the in-sample test, the KS-test is not rejected for any weather station data generated by the SWR, demonstrating the method's effectiveness across different climates in North America in comparison to WeaGETS. For the out-of-sample test, the SWR fails the KS-test significantly less frequently than WeaGETS and the naïve benchmark, indicating the weather generated by the SWR generalizes well to out-of-sample data. This is important as the SWR is used in design day studies. These studies determine the necessary infrastructure needed to deliver natural gas to customers on extreme temperature days. The out-of-sample tests give a good indication of how the estimated design day conditions will perform in the future.

The SWR reasonably estimates design day conditions as shown by the KS-test. Both the SWR and the naïve benchmark perform equivalently well on the exceedance test. The purpose of generating validated weather data is to use the generated dataset to determine the design day conditions with a return value of 1-in-N. Design day conditions estimated from the SWR and the data more closely agree with historical data than conditions estimated from the WeaGETS. The fact that the naïve benchmark also outperforms WeaGETS in estimating 1-in-30 conditions demonstrates the care that needs to be taken when generating weather. It is possible using the little data available might be better than an abundance of non-representative generated data.

The SWR allows for interpretable results with respect to the definition of odds occurring. As discussed in Section 2.2, utilities define the odds of occurring in one of two ways. (Case 1) Some utilities want to estimate the annual minimum temperature that will be exceeded, on average, in one out of N years. (Case 2) Other utilities want to estimate the temperature that is expected to be exceeded once every N years. Due to autocorrelation in temperatures, the 1-in-N temperature in Case 2 will likely be colder than that of Case 1. Our results provide validation for both cases. Case 1 is addressed in Sections 5.1.1 and 5.2.1 by comparing the distributions of the annual minima. Case 2 is addressed in Sections 5.1.2 and 5.2.2 by comparing the total number of exceedances, including multiple exceedances per year, to the expected number of exceedances under the Case 2 definition of likelihood.

WeaGETS performed poorly across all metrics except for its variance of thresholds across folds. This metric is important to practitioners who use design day conditions. It is important that design day condition estimates are stable; if they change much over time, it would be difficult to make long-term plans. Future analysis of WeaGETS could provide useful insights about why it is a more stable estimator than the SWR and the naïve benchmark.

The SWR can be applied to future work in energy forecasting. For example, Hong discusses different forms of temperature scenario generation for probabilistic peak load forecasting in electricity [31]. In fact, the shifted-date method evaluated by Hong is similar to the SWR without any treatment of seasonality. The SWR is, therefore, a straightforward improvement on current methods in this area.

Accurate design day condition estimates are essential for utilities. This paper aids utilities with two contributions. First, this paper details different methods for characterizing design day conditions. Second, the SWR generates data that aligns with historical cold weather outperforming both the naïve benchmark and the WeaGETS method. Utilities can use the generated data to estimate design day conditions that can be justified by the frequency at which those conditions have historically been exceeded. Accurately estimated design day conditions allow utilities to balance the risk of not meeting extreme demand with the cost of capacity.

Author Contributions: Conceptualization, R.H.B.; methodology, R.H.B.; software, D.K. and R.H.B.; validation, D.K. and R.H.B.; formal analysis, D.K.; investigation, D.K.; resources, R.H.B.; data curation, R.H.B. and D.K.; writing—original draft preparation, D.K.; writing—review and editing, R.J.P. and G.F.C.; visualization, D.K.; supervision, R.H.B., G.F.C. and R.J.P.; project administration, R.H.B. and R.J.P.; funding acquisition, R.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Marquette Energy Analytics, LLC.

Data Availability Statement: The data used in this study are available from the National Oceanic and Atmospheric Administration (NOAA) and can be retrieved from https://www.ncdc.noaa.gov/cdo-web/datasets (accessed on 1 October 2021). Data missing from NOAA were patched using data from AccuWeather (https://dev-business.accuweather.com/for-business/services/, accessed on 1 October 2021).

Acknowledgments: We wish to thank the anonymous reviewers for their insight and suggestions, which have greatly improved this manuscript.

Conflicts of Interest: The authors are employees of Marquette Energy Analytics. Marquette Energy Analytics uses the SWR approach in its design day studies.

References

1. Chen, J.; Brissette, F.P.; Leconte, R. WeaGETS—A Matlab-Based Daily Scale Weather Generator for Generating Precipitation and Temperature. *Procedia Environ. Sci.* **2012**, *13*, 2222–2235. [CrossRef]
2. American Gas Association. Glossary-D. Available online: https://www.aga.org/natural-gas/glossary/d/ (accessed on 11 August 2021).
3. Lee, A. Extreme Cold in the Midwest Led to High Power Demand and Record Natural Gas Demand. Available online: https://www.eia.gov/todayinenergy/detail.php?id=38472 (accessed on 11 August 2021).
4. Regulators Knew of Freeze Risk to Texas' Natural Gas System. It Still Crippled Power Generation. Available online: https://www.houstonchronicle.com/business/energy/article/freeze-risk-texas-natural-gas-supply-system-power-16020457.php (accessed on 21 September 2021).
5. Navigant Consulting. *Analysis of Peak Gasday Design Criteria*; Navigant Consulting: Toronto, ON, Canada, 2011.
6. Oracle Crystal Ball. Available online: https://www.oracle.com/middleware/technologies/crystalball.html (accessed on 21 September 2021).
7. Semenov, M.A. Simulation of Extreme Weather Events by a Stochastic Weather Generator. *Clim. Res.* **2008**, *35*, 203–212. [CrossRef]
8. Tebaldi, C.; Hayhoe, K.; Arblaster, J.M.; Meehl, G.A. Going to the Extremes: An Intercomparison of Model-Simulated Historical and Future Changes in Extreme Events. *Clim. Chang.* **2006**, *79*, 185–211. [CrossRef]
9. Oliver, R.; Duffy, A.; Enright, B.; O'Connor, R. Forecasting Peak-Day Consumption for Year-Ahead Management of Natural Gas Networks. *Util. Policy* **2017**, *44*, 1–11. [CrossRef]
10. Sarak, H.; Satman, A. The Degree-Day Method to Estimate the Residential Heating Natural Gas Consumption in Turkey: A Case Study. *Energy* **2003**, *28*, 929–939. [CrossRef]
11. Aras, H.; Aras, N. Forecasting Residential Natural Gas Demand. *Energy Sources* **2004**, *26*, 463–472. [CrossRef]
12. Vitullo, S.R.; Brown, R.H.; Corliss, G.F.; Marx, B.M. Mathematical Models for Natural Gas Forecasting. *Can. Appl. Math. Quat.* **2009**, *17*, 807–827.
13. Ishola, B. Improving Gas Demand Forecast during Extreme Cold Events. Master's. Thesis, Marquette University, Milwaukee, WI, USA, 2016.
14. Columbia Gas of Pennsylvania. *2015 Summary Report*; Columbia Gas of Pennsylvania: Canonsburg, PA, USA, 2015.
15. Intermountain Gas Company. *Integrated Resource Plan 2019–2023*; Intermountain Gas Company: Boise, ID, USA, 2019.
16. Cascade Natural Gas Corporation. *2016 Integrated Resource Plan*; Cascade Natural Gas Corporation: Kennewick, WA, USA, 2017.
17. Vermont Gas Systems. *Integrated Resource Plan 2017*; Vermont Gas Systems: South Burlington, VT, USA, 2017.
18. Avista. *2018 Natural Gas Integrated Resource Plan*; Avista: Spokane, WA, USA, 2018.
19. American Gas Association. *Winter Heating Season Energy Analysis*; American Gas Association: Wasington, DC, USA, 2014.
20. de Haan, L.; Ferreira, A.F. *Extreme Value Theory: An Introduction*; Springer: New York, NY, USA, 2006.
21. Gross, J.; Heckert, A.; Lechner, J.; Simiu, E. Novel Extreme Value Estimation Procedures: Application to Extreme Wind Data. In *Extreme Value Theory and Applications*; Galambos, J., Lechner, J., Simiu, E., Eds.; Springer: Boston, MA, USA, 1994; pp. 139–158. [CrossRef]
22. Smith, A.; Lott, N.; Vose, R. The Integrated Surface Database: Recent Developments and Partnerships. *Bull. Am. Meteorol. Soc.* **2011**, *92*, 704–708. [CrossRef]
23. Breinl, K.; Turkington, T.; Stowasser, M. Simulating Daily Precipitation and Temperature: A Weather Generation Framework for Assessing Hydrometeorological Hazards. *Meteorol. Appl.* **2015**, *22*, 334–347. [CrossRef]

24. Richardson, C.W. Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. *Water Resour. Res.* **1981**, *17*, 182–190. [CrossRef]
25. Underwood, F.M. Describing Seasonal Variability in the Distribution of Daily Effective Temperatures for 1985–2009 Compared to 1904–1984 for De Bilt, Holland. *Meteorol. Appl.* **2013**, *20*, 394–404. [CrossRef]
26. Mcmenamin, J.S. Defining Normal Weather for Energy and Peak Normalization. San Diego, CA, USA, 2008. Available online: www.itron.com/PublishedContent/DefiningNormalWeatherforEnergyandPeakNormalization.pdf (accessed on 21 September 2021).
27. Qian, B.; Gameda, S.; Hayhoe, H.; De Jong, R.; Bootsma, A. Comparison of LARS-WG and AAFC-WG Stochastic Weather Generators for Diverse Canadian Climates. *Clim. Res.* **2004**, *26*, 175–191. [CrossRef]
28. Tayebiyan, A.; Mohammad, T.A.; Ghazali, A.H.; Malek, M.A.; Mashohor, S. Potential Impacts of Climate Change on Precipitation and Temperature at Jor Dam Lake. *Pertanika J. Sci. Technol.* **2016**, *24*, 213–224.
29. Semenov, M.A.; Welham, S. Comments on the Use of Statistical Tests in the Comparison of Stochastic Weather Generators by Qian et al. *Clim. Res.* **2004**, *28*, 83–84. [CrossRef]
30. Xu, Q.; Liang, Y.-Z.; Du, Y.-P. Monte Carlo Cross-Validation for Selecting a Model and Estimating the Prediction Error in Multivariate Calibration. *J. Chemom.* **2004**, *18*, 112–120. [CrossRef]
31. Xie, J.; Hong, T. Temperature Scenario Generation for Probabilistic Load Forecasting. *IEEE Trans. Smart Grid* **2018**, *9*, 1680–1687. [CrossRef]