Michael Town
michaelstown@protonmail.com
Metis: Web scraping and Linear Regression Module
14 Dec 2021

Title:

Temperature time series reconstruction for the southern aspect of Mt. Baker using meteorological measurements from the northern aspect of Mt. Baker.

Upshot:

Here I share reconstructed time series of mean daily temperature from Sep 2014 – Dec 2021 for the south side of Mt. Baker using data from the north side of Mt. Baker. The reconstruction tracks general trends, but is missing extreme day-to-day variability.

Abstract:

A short time series of mean daily temperature from Schrieber’s Meadow, Mt. Baker (southern aspect, time range = 1040 m, Jul 2018 – Jul 2021) has been extended using multiple linear regressions on meteorological measurements from two stations at the Mt. Baker Ski Resort (northern aspect, Heather Meadows at 1300 m asl and Pan Dome at 1530 m asl, time range = Sep 2014 – Dec 2021). Improved power is found when categorizing the temperature relationships by season (djf, mam, jja, son), wind speed, wind direction (N-NE, N-NW, W, S) and precipitation amount (no precip, ‘drizzle’, light precip, moderate precip, heavy precip). Temperature measurements at Heather Meadows and Pan Dome are highly correlated. The temperature reconstruction reproduces mean values well, but misses extreme variability on the order of 2-5 ^{o}C.

Data:

Target variable: The target variable here is mean daily 2-m air temperature from Schrieber’s Meadow, Mt. Baker, WA USA. The mean daily temperatures are compute from 4.25-hourly 2-m air temperature collected from Jul 2018 – Jul 2021 as part of the Mt. Baker Climate Project (A Lakeside School and Ballard Data Science LLC collaboration).

Feature variables: The feature variables are collected at the Mt. Baker Ski Resort on the northern aspect of Mt. Baker, WA USA and served to the public by the Northwest Avalanche Center. All data collected from this site are They are:

1) Mean daily 2-m air temperatures computed from hourly 2-m air temperatures from the Heather Meadows site (1300 m asl).
2) Mean daily 2-m air temperatures computed from hourly 2-m air temperatures from the Pan Dome site (1530 m asl).
3) Season category (December-January-February, djf; March-April-May, mam; June-July-August, jja; September-October-November, son).
4) Precipitation category (no precip, drizzle = 0-1 mm/hr, light precip = 1-2.5 mm/hr, moderate precip = 2.5-7.5 mm/hr, heavy precip > 7.5 mm/hr) computed from hourly measurements at the Heather Meadows site.
5) Mean daily wind speed and direction computed from hourly measurements at the Pan Dome site.

Design:
Data from the Northwest Avalanche Center were webscraped using Selnium and processed with BeautifulSoup. Initial exploratory data analysis was performed via monthly histograms of all temperatures, wind speed, and wind direction data. A multiple linear regression (ordinary least squares fit) was performed on these data in various combinations. The r-squared score was used as a measure of the performance of the fit, including p-values to determine whether or not elements of the feature pool were useful to the model. A QQ plot was used to as a qualitative determination of normality of the target data. A residual vs. target scatter plot was used to determine any range dependencies of the model.

Models were built on 70% of the data and tested on 30% of the data. The test results are included in the appendix included in this package.

Discussion:

The regressions from Figures 1 and 2 demonstrate that there is a lot of power in explaining mean daily temperature across the mountain simply by using the 2-m air temperature. Mean daily temperatures from Heather Meadows and Pan Dome do not have independent power in predicting 2-m air temperature at Schrieber's Meadow, so in future regressions I left out the Pan Dome air temperature.

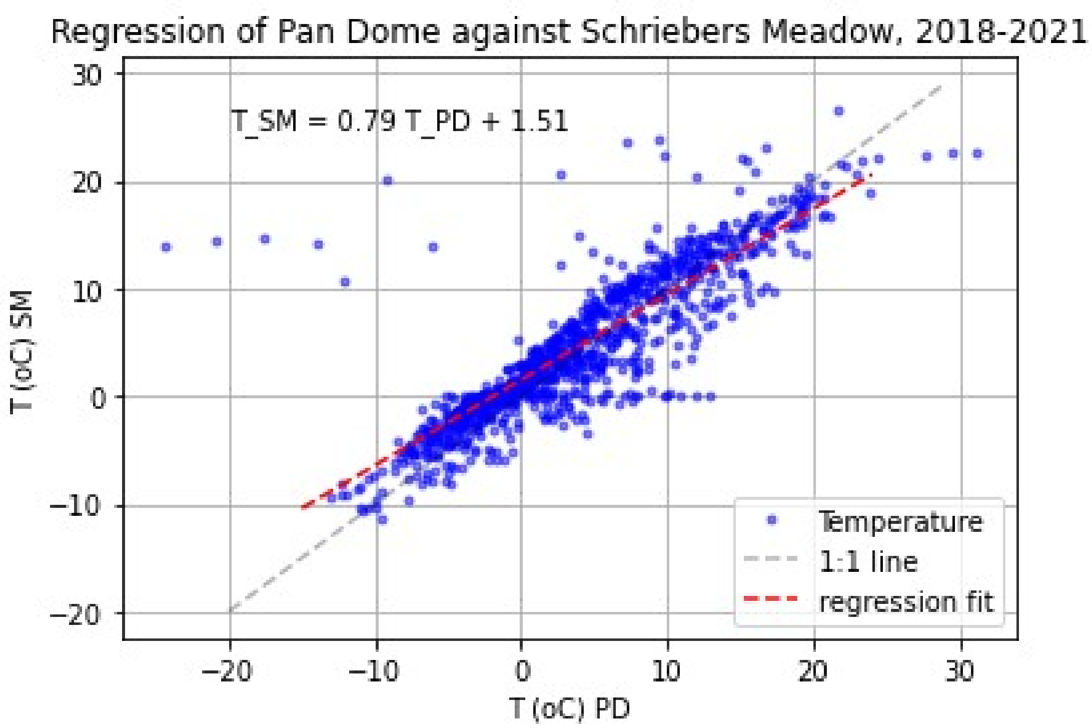

Figure 1. A linear regression of mean daily temperature from Pan Dome, Mt. Baker, WA USA against mean daily temperature from Schrieber's Meadow, Mt. Baker WA, USA. A single event caused scatter at high values. The many values spread around zero are likely when snow is covering the sensor at Schrieber's Meadow.

Figure 2 demonstrates that the linear regressions by season are all slightly different. Details can be found in the appendix. The slopes range from 0.70 – 0.80. When combined into one regression model, the March-April-May Heather Meadows temperature data have a very high p-value (0.700>p or higher depending on what else in included in the regression). This is likely because the MAM temperature range includes DJF, SON, and JJA values.

Wind speed was not an important predictor of temperature as a continuous variable. Future work might go into determining relevant a priori categories for wind speeds to improve this model. Similarly, wind direction does not do a strong job of explaining any air temperature correlations across the mountain. Wind speed and direction were ultimately removed from the final reconstruction model.

Precipitation categories of no precip/rain, drizzle, and heavy rain had p-values below 0.05, and were therefore moderately useful in explaining overall temperature trends at Schrieber's Meadow. Precipitation was included in the final model.

The residual plot (Figure 4) shows no trend in residuals, but large residuals across the range of data.

The final reconstruction is shown in Figure 5. The gray values are the measured temperatures, the blue values are the modeled values based on daily mean values from the north side of Mt. Baker, WA USA. The model reproduces mean trends, but misses extreme values in mean daily temperature. See Table 1 below for the evolution of R2 and p-values through the model experiments run.

This reconstructed data set will be useful in understanding mountain-wide trends, but more work in characterizing residuals or finding more feature variables will have to be done to explain the extreme values of mean daily temperature observed at Schrieber's Meadow.

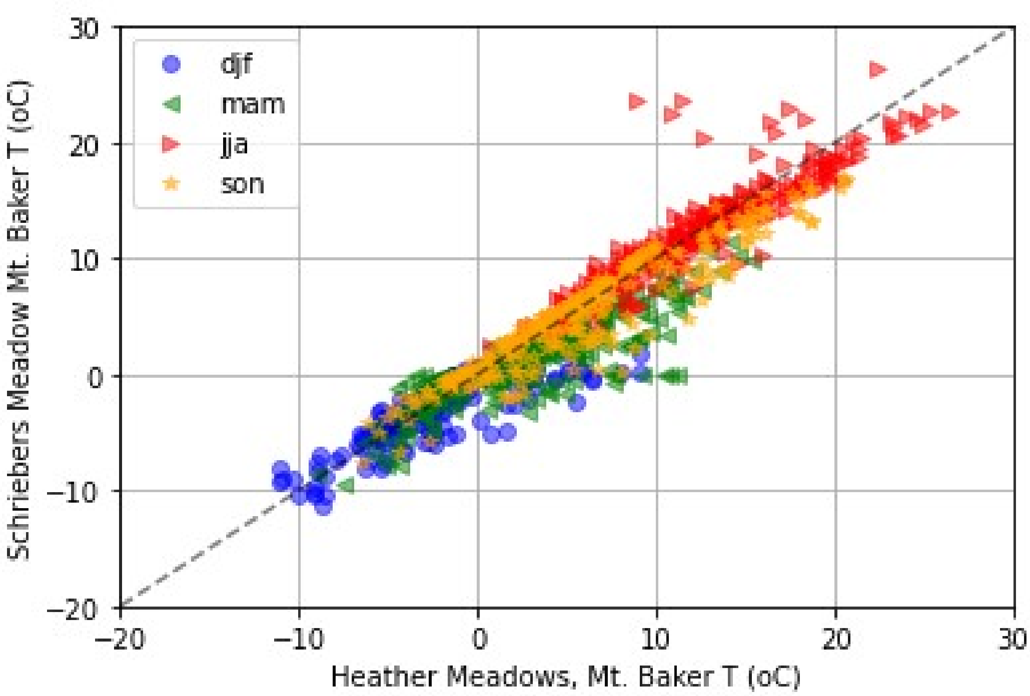

Figure 2. A linear regression of mean daily temperature from Heather Meadows, Mt. Baker, WA USA against mean daily temperature from Schrieber's Meadow, Mt. Baker WA, USA. These data are composited by season. A single event caused scatter at high values. The many values spread around zero are likely when snow is covering the sensor at Schrieber's Meadow.

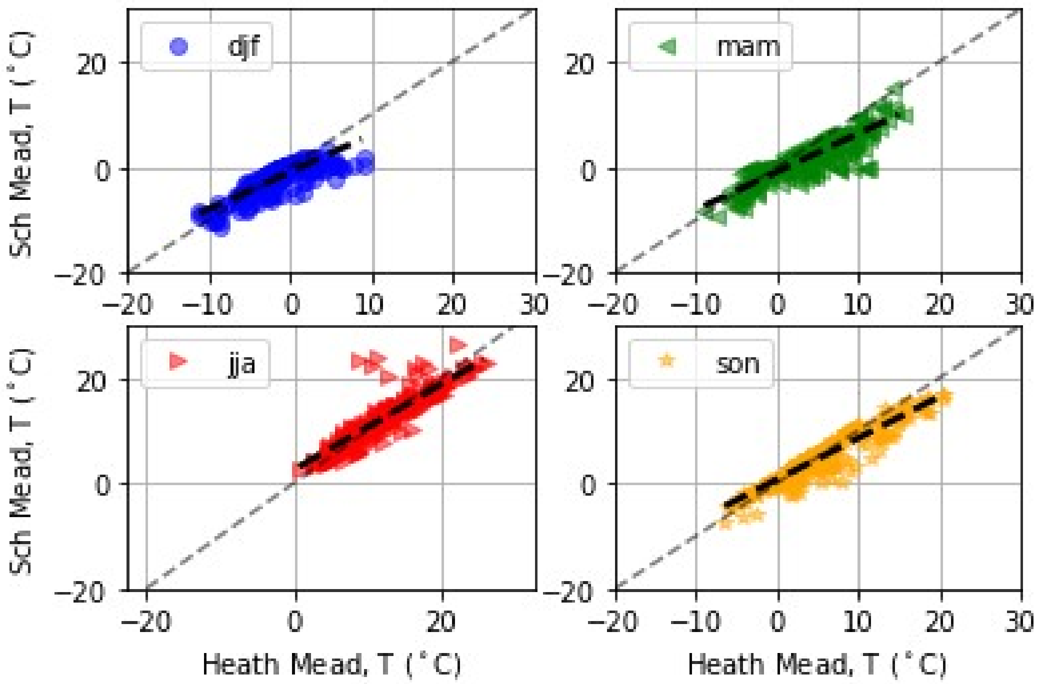

Figure 3. A linear regression of mean daily temperature from Heather Meadows, Mt. Baker, WA USA against mean daily temperature from Schrieber's Meadow, Mt. Baker WA, USA. These data are composited by season. The linear regression for each individual season is shown. A single event caused scatter at high values. The many values spread around zero are likely when snow is covering the sensor at Schrieber's Meadow.

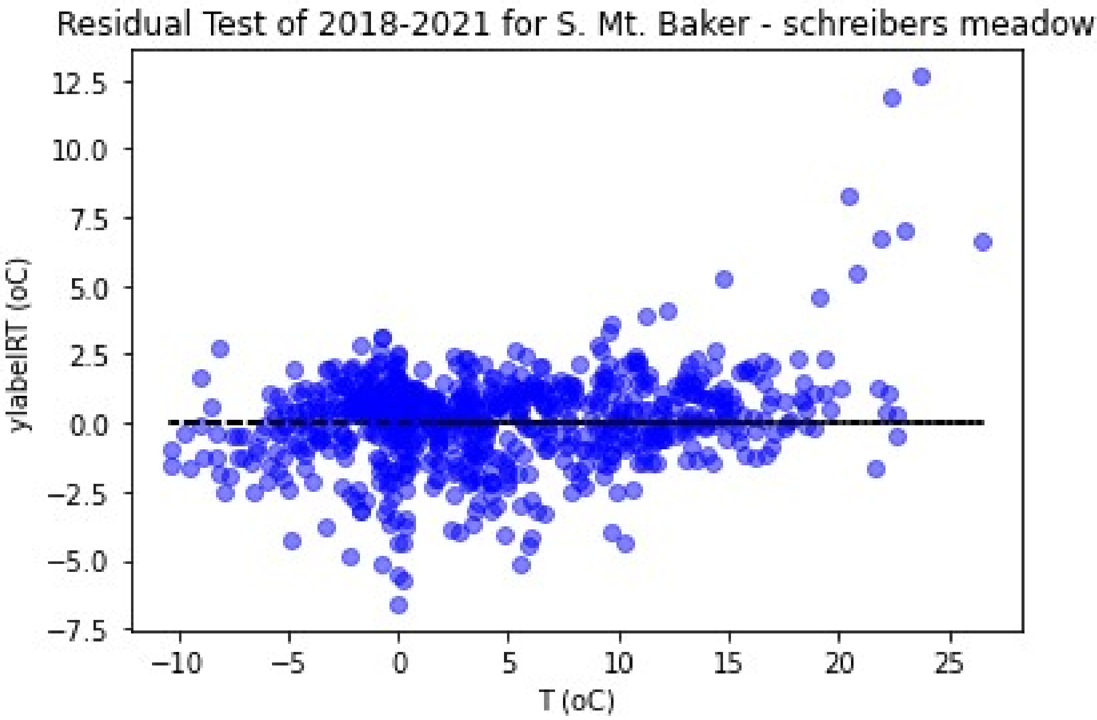

Figure 4. Mean daily predicted temperature residual against Schrieber's Meadow temperature for Jul 2018 – Jul 2021.

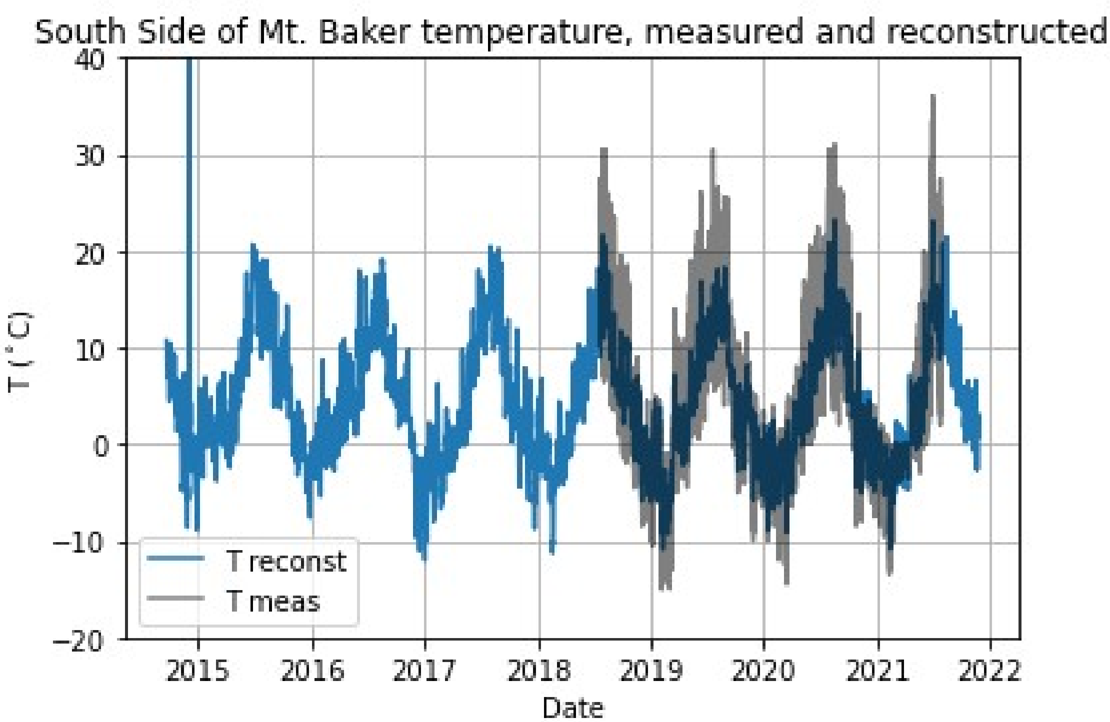

Figure 5. Observed mean daily 2-m air temperature (gray) and reconstructed mean daily 2-m air temperature for Schrieber's Meadow, Mt. Baker, WA USA. The spike in air temperature at the end of

2014 is due to a spike in air measured 2-m air temperature at Heather Meadows (a primary component of the air temperature model).

Model Evolution:

The model evolution is documented in the table below as features were included/excluded. More details can be found in Appendix 1. You can see that the model improved importantly when the seasonal (excluding MAM data) and precipitation (excluding light and moderate rain) information was included, but not when the wind speed and direction information was included.

Table 1. Model performance demonstrated by R2, and feature significance as demonstrated p-value. Reminder, the model target was mean daily 2-m air temperature for Schrieber's Meadow. More evolution and testing evidence is provided in Appendix 1.

Model Performance (R^2)	Model Features	Feature Performance (p-value)	Notes
0.758	Pan Dome 2-m T	0.000	One errant event in Summer visible.
0.758	Pan Dome 2-m T wind speed	0.000 0.135	
0.842	Pan Dome 2-m T (djf) wind speed jja mam son	0.000 0.047 0.000 0.208 0.000	Base case is djf.
0.763	Pan Dome 2-m T (N-NE) wind speed N-NW S W	0.000 0.194 0.075 0.570 0.465	Wind direction categories chosen based on EDA of monthly wind direction histograms. Base case is N-NE wind direction.
0.843	Pan Dome 2-m T (N-NE, djf) wind speed N-NW S W jja mam son	0.000 0.058 0.216 0.941 0.442 0.000 0.301 0.000	Base case is N-NE wind direction and djf.
0.935	Pan Dome 2-m T (N-NE, djf, drizzle) Heather Meadows 2-m T wind speed N-NW S W	0.014 0.000 0.147 0.843 0.148 0.688	Base case is N-NE wind direction, djf, and drizzle.

	jja	0.000	
	mam	0.836	
	son	0.000	
	heavy rain	0.014	
	light rain	0.069	
	moderate rain	0.383	
	no rain	0.000	
0.933	Heather Meadows 2-m T (djf, drizzle)	0.000	Base case is djf and drizzle.
	jja	0.000	
	mam	0.875	
	son	0.000	
	heavy rain	0.009	
	light rain	0.121	
	moderate rain	0.719	
	no rain	0.000	
0.758	Z-transformed Pan Dome Temp	0.000	

Z transform!:

Late in the game I performed a z-transform on the data and ran an OLS on only the Pan Dome data. It did not improve the regression performance in this case. Also, it did improve the QQ plot. See below.

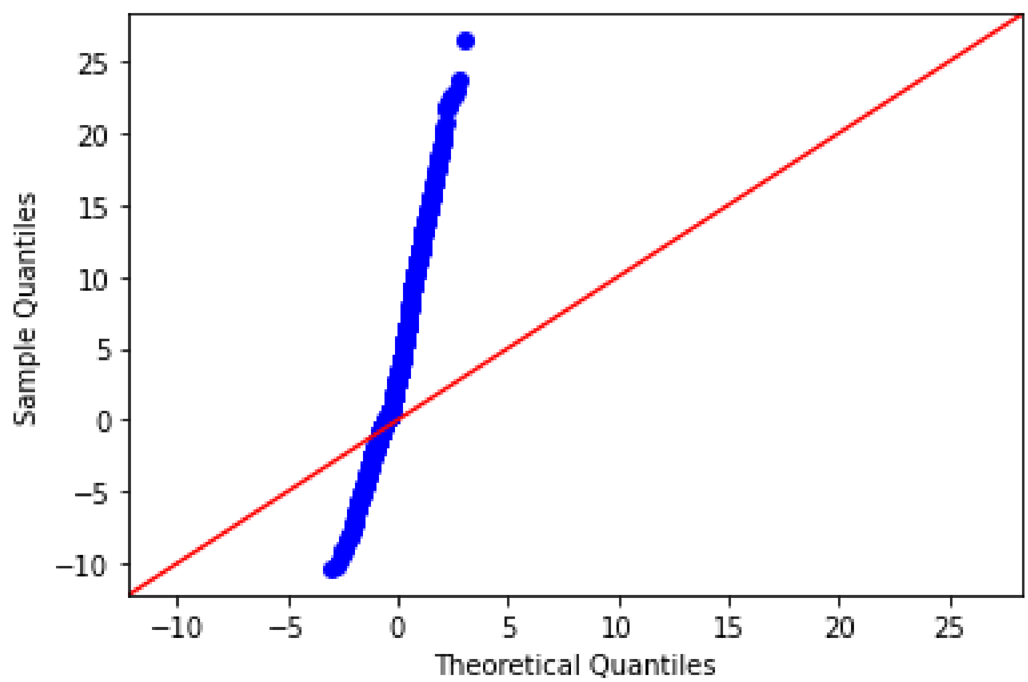

Figure 6. QQ plot of mean daily 2-m air temperature from Schrieber’s Meadow (training data set).

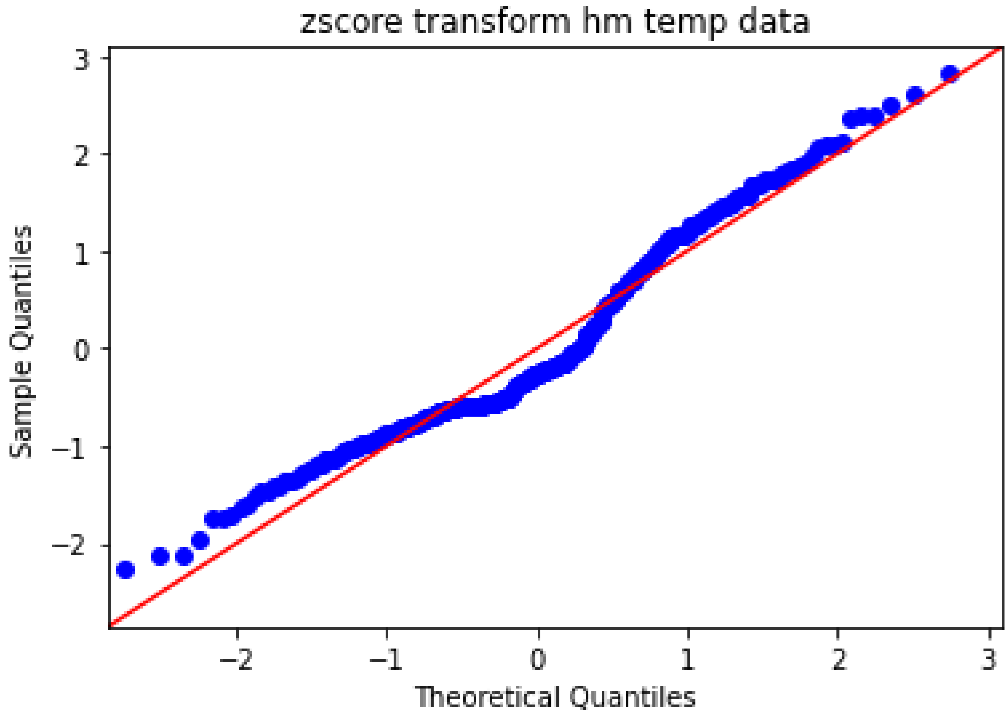

Figure 7. QQ plot of z-transformed mean daily 2-m air temperature training Schrieber's Meadow (training data set).