

GEOG309-23S2

Research for resilient environments and communities

ANALYSIS OF WEATHER DATA AT MT
HUTT AND THE WIDER CANTERBURY
ALPINE REGION



Nina Harris, Thomas Dawe, Ella Nicholls, Quinn Hanham, Jordan Sadeghi

Table of Contents

Executive summary.....	3
1. Introduction.....	4
2. Study Site.....	4
3. Stakeholders.....	5
4. Background Information.....	5
4.1. Scales.....	5
4.1.1. Energy Budget.....	6
4.1.2. Heat Fluxes.....	6
4.1.3. Lapse Rates.....	6
4.2. Weather station data collection.....	7
4.3. Validating Weather Station Data.....	7
5. Methodology.....	7
5.1. Weather Station Data Wrangling.....	7
5.2. Temperature.....	8
5.3. Rainfall.....	8
5.4. Wind.....	8
6. Results.....	9
6.1. Temperature.....	9
6.2. Rainfall.....	12
6.3. Wind.....	14
7. Discussion.....	15
7.1. Lapse Rates on Mt Hutt.....	15
7.2. Microclimates on Mt Hutt.....	15
7.3. Synoptic Scale Processes.....	17
8. Limitations.....	20
9. Conclusions.....	21
10. Acknowledgements.....	22
11. Appendix 1.....	22
12. References.....	31

Executive summary

- The research question for this project is: How can we better understand the synoptic and microscale weather processes at Mt Hutt?
- This report discusses the technical data manipulation techniques used to identify weather processes. Different datasets provided from our community partner were wrangled using R studio. Visualisation of the data was done using R studio & Excel. Analysis was also completed to validate our community partners weather stations.
- Mt Hutt has complex weather dynamics due to the topography of the southern alps, links between microscale and synoptic processes were made to explain the results we found.
- This project provides a solid foundation for future climatic research in Mt Hutt and the wider Canterbury alpine region. In the light of climate change, historic climate data is becoming more relevant and important to understand and combat this increasing threat.
- Limitations were found when undertaking this project such as group resources as well as inherent weather station errors/missing data.
- Code will be provided to our community partner to improve the data wrangling process.

1. Introduction

This report focuses on the analysis of weather station data in the Southern Alps, specifically the Mt Hutt region. A Cass Research Area Management Group member Dave Kelly has set up three weather stations at different elevations adjacent to the popular Mt Hutt Ski Field Access Road with the first station erected in 1995 and still running at present. These stations record temperature, rainfall, and sunlight in different time scales, this is good for our community partners research scope in biological sciences but lacks any windspeed and wind direction data

which is fundamental for climate analysis. To analyse weather station data in the Mt Hutt region, three additional weather stations were chosen from NIWA (CliFlo, 2023) to understand wind patterns from the head of the main divide catchments to the mid canterbury plains (Rakaia/Snowdon/Winchmore). This project builds an understanding of the microclimate and synoptic processes found within the Mt Hutt region while providing code to help with further data wrangling in the future.

2. Study Site

The region of interest in this project was Mt Hutt, found to the west of the Canterbury Plains and reaching a height of 2190m. The mountain houses a popular ski field approximately 100km from Christchurch as seen in Figure 1 within the Hakatere conservation park between the Rakaia and Rangitata rivers. This region of the southern alps is in the lee of the Main Divide which tends to receive less precipitation and warmer temperatures on average compared to the west coast (Sturman & Spronken-Smith, 2001).

Rakaia AWS is found deep in the alps at an elevation of 1752m above sea level. At this site, there is a strong influence of dynamic weather and funnelling due to the complex topography which can exacerbate weather station recordings. Snowdon AWS is found at the head of the braided river segment of the Rakaia River. This site is 420m above sea level which is significantly lower than Rakaia AWS. The final NIWA site is Winchmore, which is in the heart of the mid canterbury plains found 12km from Ashburton and 30km from the coast. These three independent sites were chosen to give us three different topographical areas within the region to understand the effects of weather.

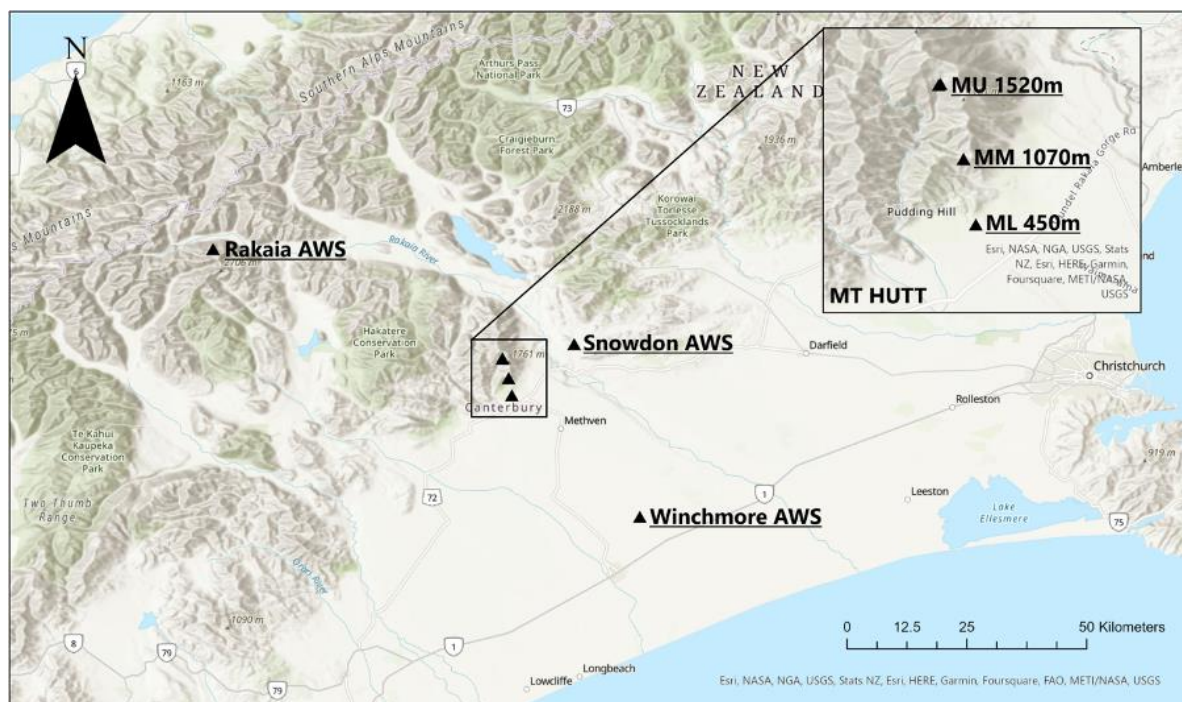


Figure 1: Map representing the study site. Mt Hutt found within the subset.

3. Stake Holders

Our primary stakeholder is Dave Kelly (member of the Cass Research Area Management Group). They provided valuable temperature and rainfall measurements to better understand weather patterns. A key focus was automation of data wrangling to validate the weather station readings because this process is currently labour intensive.

This project has significance for local Iwi in future research because climate data can be used to understand ecological processes that implicate food security and biodiversity. Climate data also informs research around potential health and wellbeing issues derived from climate change. The information validated during this study can be used to gain insights of how the local environment has been affected over time.

4. Background Information

4.1 Scales

It is important to understand microclimates and how they affect weather station readings. Microclimates are local atmospheric conditions which differ from the surrounding region. This can be due to a range of factors such as topography, vegetation, anthropogenic interference, or atmospheric processes such as energy budgets and lapse rates. Larger scale weather processes are found on the synoptic scale, this scale is where pressure systems have an influence on the weather and control the environment at a nationwide scale.

I. Energy Budget

Weather at a global scale is driven by the surface energy budget. The sun emits shortwave radiation, which is mostly absorbed by the earth's surface, longwave radiation is then released thermally. This outgoing longwave radiation is what drives the global circulation by heating parcels of air producing large-scale winds and cells. Incoming shortwave radiation can also be deflected back into the atmosphere from the surface, this outgoing shortwave reflects off surfaces with a high albedo such as snow and ice which absorbs extraordinarily little of this energy. Longwave outgoing can also be reflected into the earth's surface from things such as canopies and clouds. The four inputs for this system work together to circulate air parcels around the globe (Barry, R. G. et al 2009).

II. Lapse rates

Lapse Rates are an important thermal dynamic concept used extensively in this project. This is the rate of temperature change in the atmosphere with elevation. This is due to the decrease in air pressure at higher elevations where expansion of the air parcel occurs. This expansion does work on its surroundings without exchanging heat that in turn lowers the air temperature (Daidzic, N. 2019). There are numerous types of lapse rates, namely Environment Lapse Rate (ELR) and Dry Adiabatic Lapse Rate (DALR). ELR is the actual rate at which temperature

decreases with altitude whereas DALR is the rate at which temperature of dry air parcels decreases with altitude. DALR is used as a ‘ruler’ to identify if the atmosphere is stable or not. Stability is how likely an air parcel is to return to its position. An analogy to understand this process is pushing a ball up a valley wall, this is very likely to return to the bottom of the valley rendering it stable. The inverse of this is pushing a ball off the top of a hill, this is very unlikely to result in the ball returning to the top of the hill. If the ELR falls below the DALP then the air parcel is considered unstable and vice versa (Dawe, 2022).

4.2 Weather Station Data Collection

Literature on weather station data collection shows examples of difficulties and the usefulness of using weather stations to record data. Weather station data has been found to be accurate for the vicinity of the station, however interpolation is often required between stations and require maintenance along with constant periods of measurement to provide useful data (Mendelsohn et al., 2007). Research performed by Hahn et al. (2022) found that privately owned weather stations have the potential to provide a positive impact on meteorology and climate data. Although it is also mentioned that careful quality control is required for the data to be useful due to the non-uniform monitoring of the privately owned stations. They emphasise that unknown environmental conditions and errors in measurement can lead to possible impairments in analysis.

4.3 Validating Weather Station Data

The process of data wrangling for weather stations is complicated by the variability of valid results as weather conditions fluctuate often. When reviewing literature about this issue, two main steps were identified. These include statistical processing at the single station level, and then spatial autocorrelation of weather stations in proximity to each other (Beele, et al. 2022), (Liu & Zhang 2021). This can be done because climatic processes are continuous in space.

Beyond these two broad approaches, the methods used for data correction are very dependent on the number of available weather stations, the density of stations, and the uniformity of the data collection (Beele, et al. 2022). More stations and longer duration of data collection produces better results because statistical error is reduced, while dense point data improves the accuracy of spatial analyses (Skrynyk, et al., 2023), (Beele, et al. 2022).

5. Methods

5.1 Weather Station Data Wrangling

This project will be limited by the number of available stations, largely because the alpine region is less populated and therefore there is less interest in crowd sourced weather monitoring, and it is harder to maintain the hardware. However, there is a good density of weather stations, especially for temperature in the past three years because of ongoing research by RainfallNZ, who have placed a transect up Mt Hutt. These stations were invaluable to this project, confirming readings and trends identified on the mountain.

Figure 2 shows the first stages of the data wrangling for all three stations. This code takes two excel files containing weather station data and formats them into data frames that are ready for comparison. The first few steps include reading the data, and defining three functions, two of which relate to reformatting the data for further processing and one which applies a correction for the lapse rate between the two supplied data sets. The lapse rate established by RainfallNZ is -0.05 degrees/meter. This value has been used for these steps.

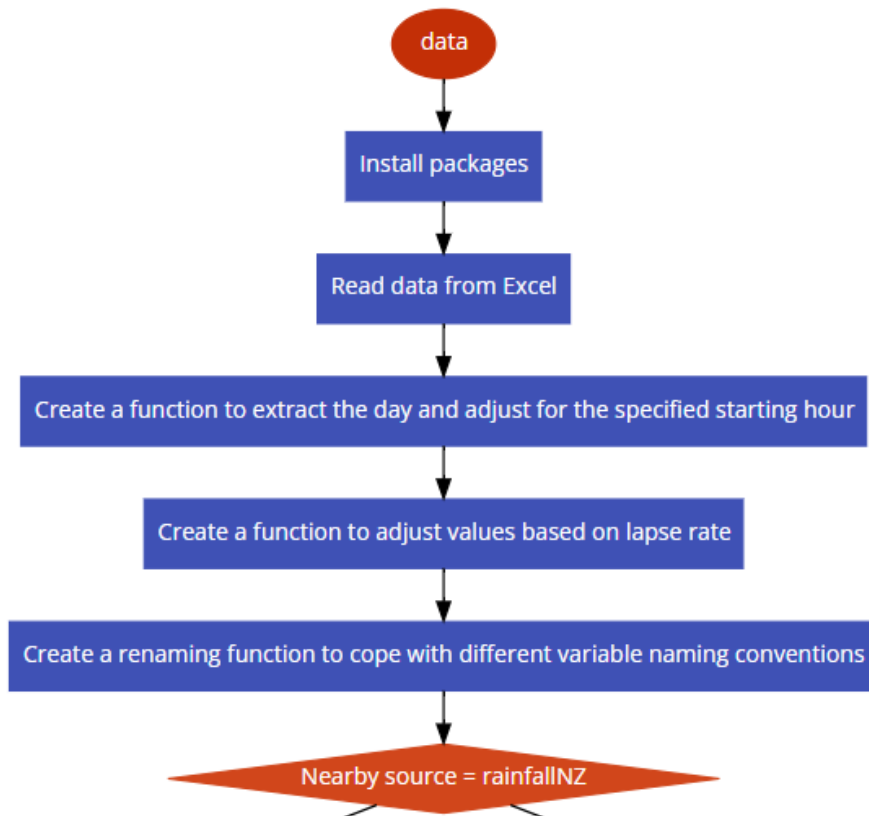


Figure 2: Flow diagram of steps used in data wrangling

The next step is an if statement which checks which secondary source is being used (Figure()). This is done because RainfallNZ and the three study stations have different processing requirements. The output of the if statement is always a data frame with four columns, for date, maximum, mean, and minimum daily averages. A second date data frame is created using specified start and end dates so that the user can control their period of interest. The data from the primary weather station is read into a data frame and merged onto the dates data frame. This will help users identify areas of missing data collection. Finally, the renaming function is applied to the primary data set.

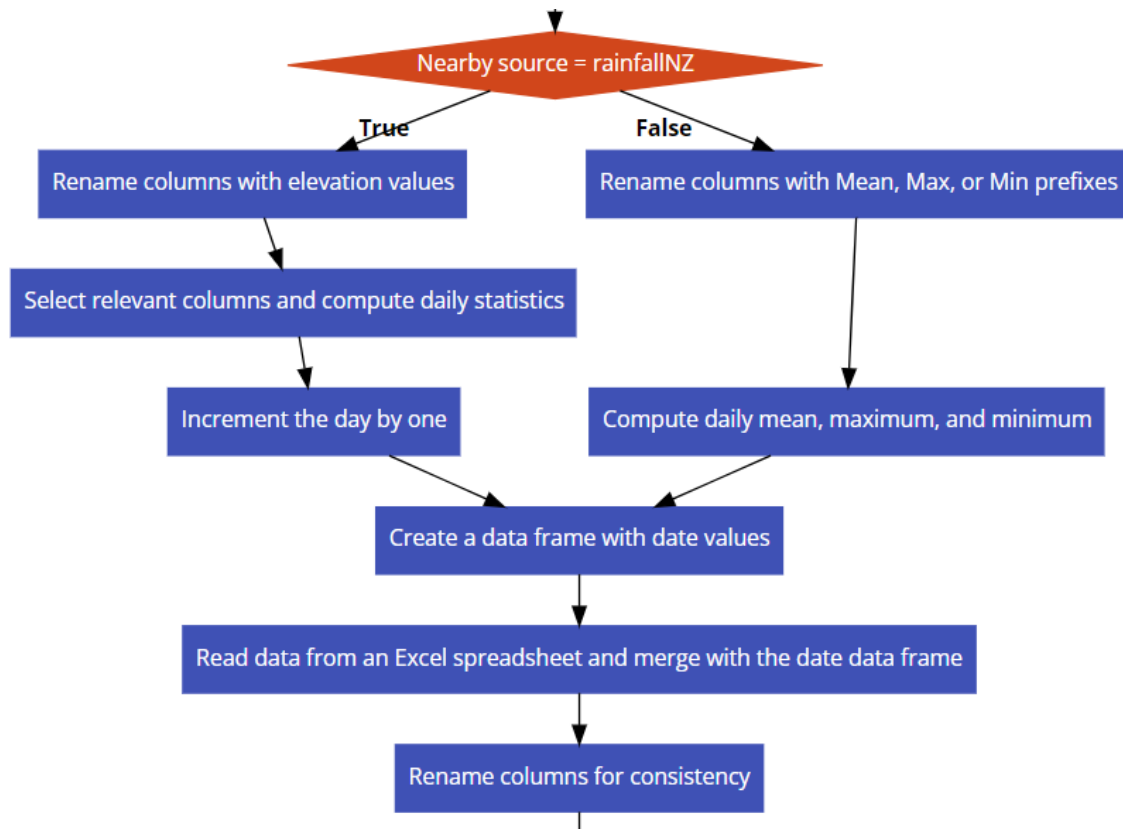


Figure 3. Flow diagram of steps used in data wrangling.

One gross error that has been identified is due to recurring battery failures. This sometimes produces a repetitive extreme reading for the panel sensors which are filtered out of the dataset (Figure 4). For this reason, it is helpful to record the status of the battery, because it can be an indicator of poor hardware performance. The last three steps are statistical comparisons of readings, first within the primary data set and then to the averaged values of the secondary dataset. Thresholds are set based on the maximum difference between stations after a correction for bias has been applied. Values are only modified if they will not intersect with their corresponding minimum, mean, and maximum daily readings. This was an issue which arose when applying spatial correlation analysis over larger distances. This is due to higher variability in the data, which lead to greater processing error.

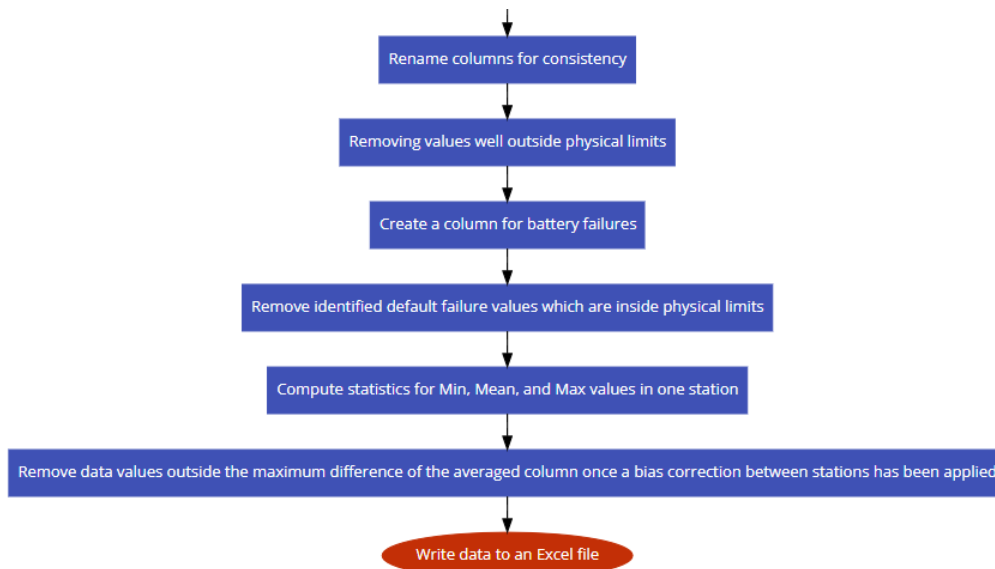


Figure 4. Flow diagram of steps used in data wrangling

For further detail please see Appendix 1.

5.2 Temperature

RainfallNZ has ten temperature sensors situated on Mount Hutt in an elevation transect. For this project we had a specific focus on their Site 6 which has an elevation of 1090m and is a short distance to our middle elevation site. Their data is recorded in half an hour intervals and measures temperature in degrees Celsius. Our community partner collects daily temperature averages, so a pivot table was used to generate daily temperature averages from the hourly values of the RainfallNZ data. From there a comparison of daily average temperatures was conducted between the two sites by creating a conditional formatting rule to display a heat map substitute on excel.

5.3 Rainfall

The low and mid-altitude weather stations employ tipping bucket rain gauges for rainfall measurement. Rainfall measurement was not performed at the highest altitude site as weather conditions at high altitudes require more upkeep and are more likely to be damaged by environmental hazards. The tipping bucket rain gauges measure the amount of rainfall by collecting precipitation until a specific volume of water is in the bucket, at which point the bucket tips and the data is recorded by a connected data logger. The rain gauges used in the weather stations were set to tip when the bucket collected 0.2mm of water, meaning that a reading was made every 0.2mm. Comparisons were made between the low and mid-altitude stations along with weather stations set up by Environment Canterbury (ECan) around the Mt Hutt area.

5.4 Wind

To analyse wind in the Mt Hutt area, 3 NIWA stations were gathered (Rakaia, Snowdon, Winchmore) to give three different topographical areas within the region to understand the effects of weather. These sites are located at over 1700m asl at the base of the main divide to 160m asl in the heart of the canterbury plains. To analyse the wind at these sites, windspeed vs direction graphs were created overlaying all sites in 3-month periods from summer 2015/16 and winter 2016. Windspeed vs temperature graphs were also created to understand the effects of atmospheric mixing and temperature structures.

6. Results

6.1 Temperature

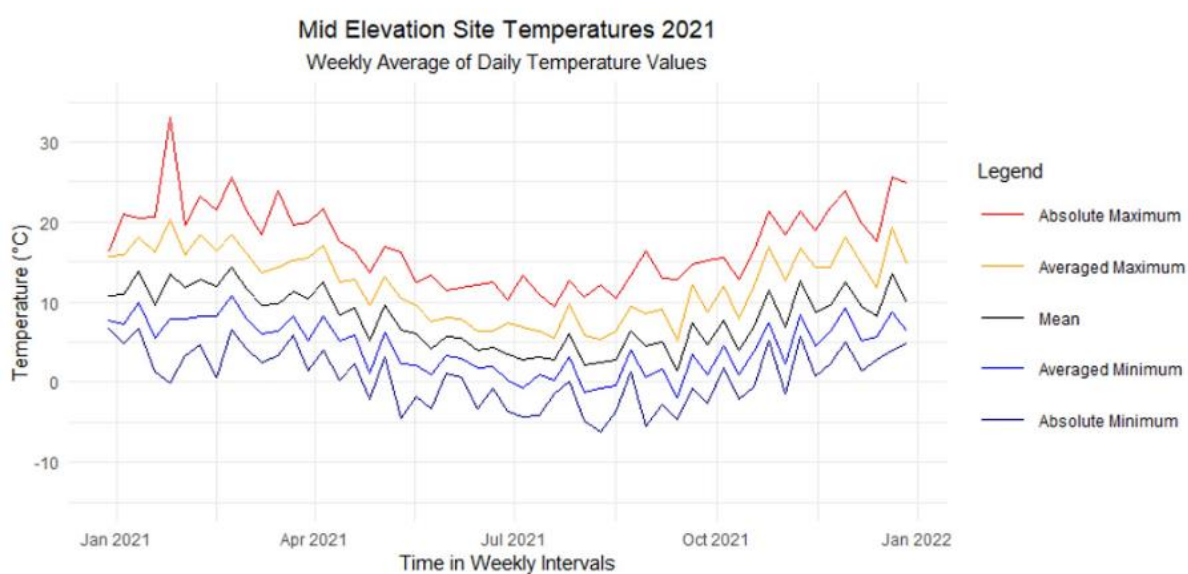


Figure 5. Line graph showing weekly average of daily temperatures for 2021 at the Mid elevation site on Mt Hutt.

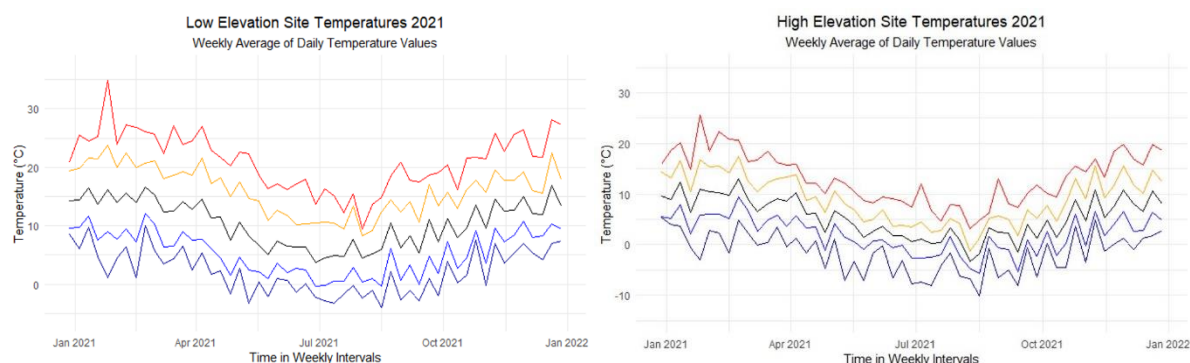


Figure 6. Line graphs showing weekly average of daily temperatures for 2021 at the Low elevation site (left) and high elevation site (right) on Mt Hutt

Figures 5 & 6 represent weekly averages of daily temperature for 2021 at all 3 sites on Mt Hutt. From the graphs, there seems to be a clear temperature lapse rate. For the higher elevation site, the variability between the max and min absolute values is less dispersed compared to the array in the lowest elevation site. Large temperature fluctuations are present during warmer summer months and tend to dampen in winter months.

Following Figures 5 & 6, Figure 7 shows a clear positive correlation between windspeed and temperature.

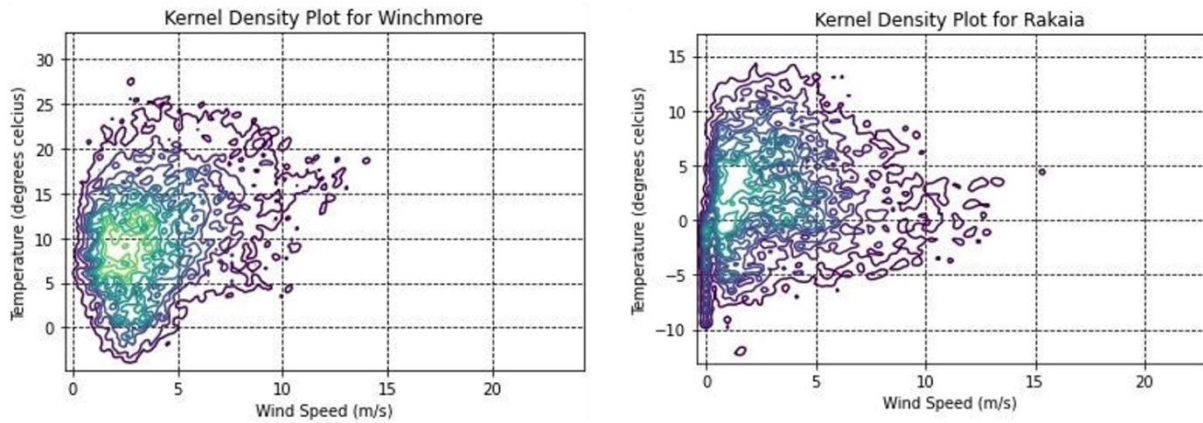


Figure 7. Kernel density plots showing the relationship between Wind Speed (m/s) and Temperature (degrees Celsius) for Winchmore and Rakaia. This is data from the 10/03/2022 - 10/10/2023. Areas with purple show less frequent occurrence, and areas with blue to green indicate a higher frequency of occurrence. *Note y-axis scale is not the same as technical difficulties were endured with the python coding process.

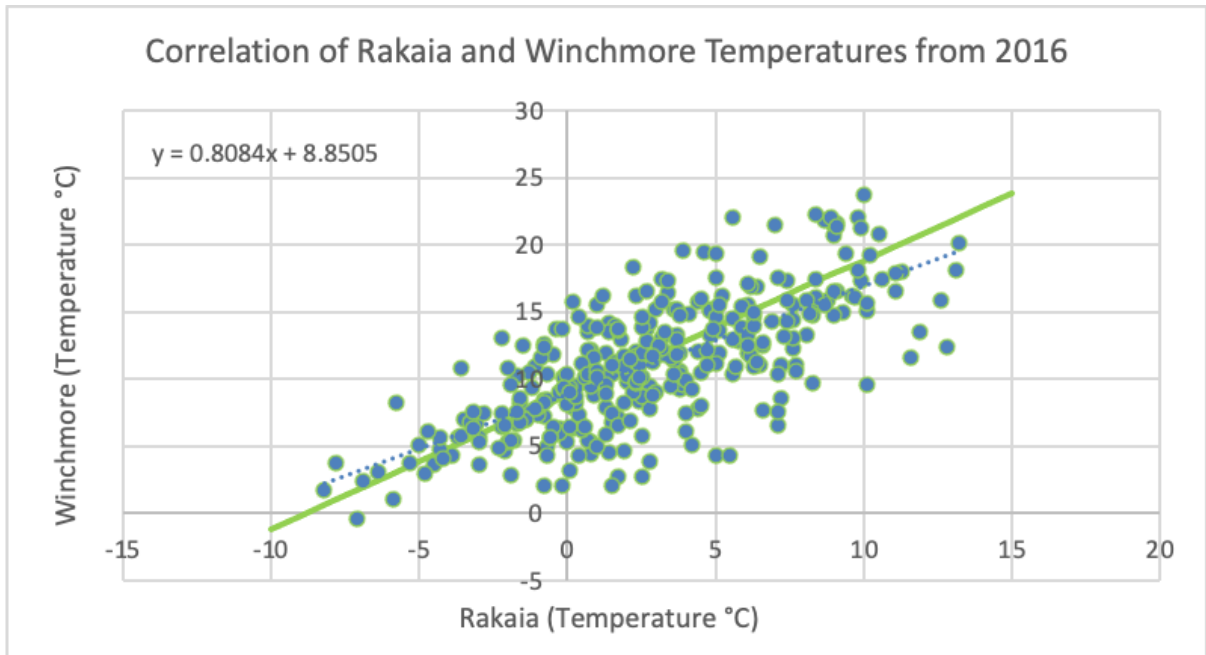


Figure 8. The gradient of 0.8084 means that for every increase in temperature in Rakaia of 1 degree, temperature increases by 0.8084 degrees in Winchmore. The y-intercept of 8.8505 means that when the temperature in Rakaia is 0 degrees, the temperature in Winchmore is 8.8505 degrees.

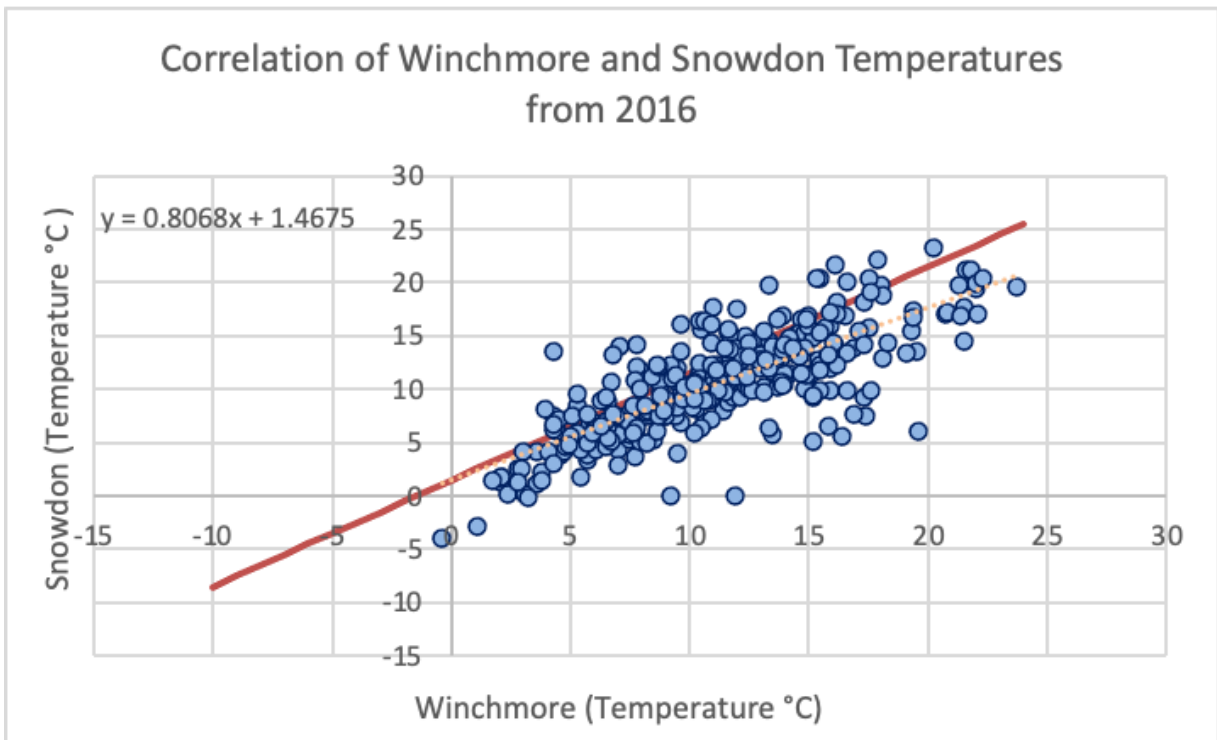


Figure 9. The gradient of 0.8068 means that for every increase in temperature in Winchmore of 1 degree, temperature increases by 0.8068 degrees in Snowdon. The y-intercept of 1.4675 means that when the temperature in Winchmore is 0 degrees, the temperature in Snowdon is 1.4675 degrees.

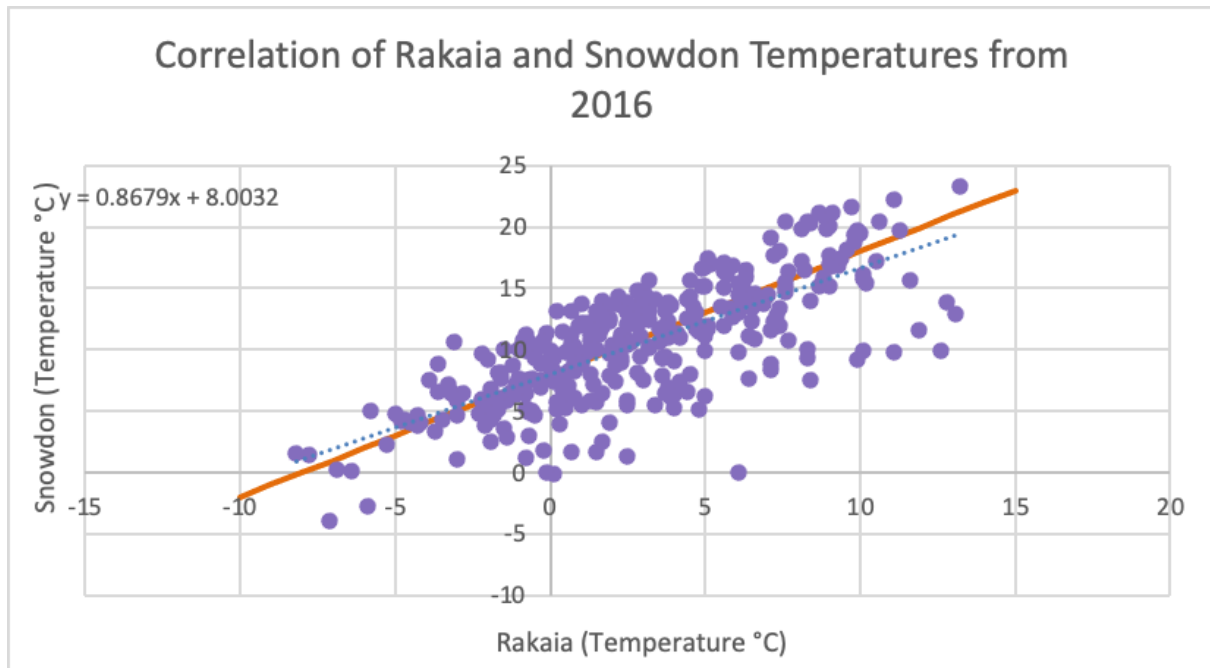


Figure 10. The gradient of 0.8679 means that for every increase in temperature in Rakaia of 1 degree, temperature in Snowdon increases by 0.8679 degrees. The y-intercept of 8.0032 means that when the temperature in Rakaia is 0 degrees, the temperature in Snowdon is 8.0032 degrees.

6.2 Rainfall

Figure 11 shows the accumulation of rainfall between the mid elevation site (MM) and ECan's station located at a similar altitude and within proximity. The findings show that ECan's station recorded higher amounts of rainfall when compared to the MM site in which the disparity becomes clearer as time progresses. Figure 12 shows the same stations being compared although focuses from 2016 onward in which an Equipment change to the mid altitude tipping bucket rain gauge was made. Similar findings were found between the two stations such as the reduction of rainfall gathered around the end of 2018 and the rise in the middle of 2021, although the mid altitude station is shown to be recording significantly lower measurements. Figures 13 and 14 show the accumulation of rainfall between the MM site and the low elevation site. Figure 13 shows that the mid- elevation site gathered more rainfall than the low-elevation site. However, Figure 14 shows that during 2016-2023, the low elevation site was gathering higher amounts of rainfall compared to the mid elevation site. These results suggest there may have been an error with the equipment collecting rainfall or difficulties with the environment at the mid-elevation site. This is supported by Figure 15 in which shows the low altitude site is collecting lower amounts of rainfall when compared to ECan's station near the mid elevation site.

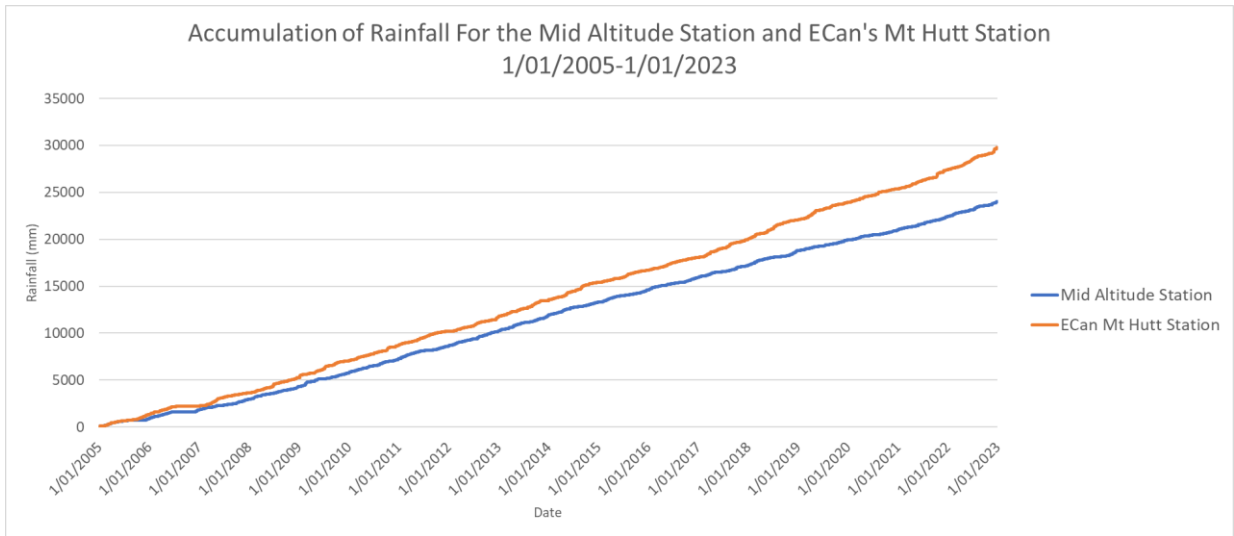


Figure 11. Line graph showing the accumulation of rainfall recorded at the mid elevation site alongside the recordings from the ECan Mt Hutt Station. The recordings shown are during 1/01/2005-1/01/2023.

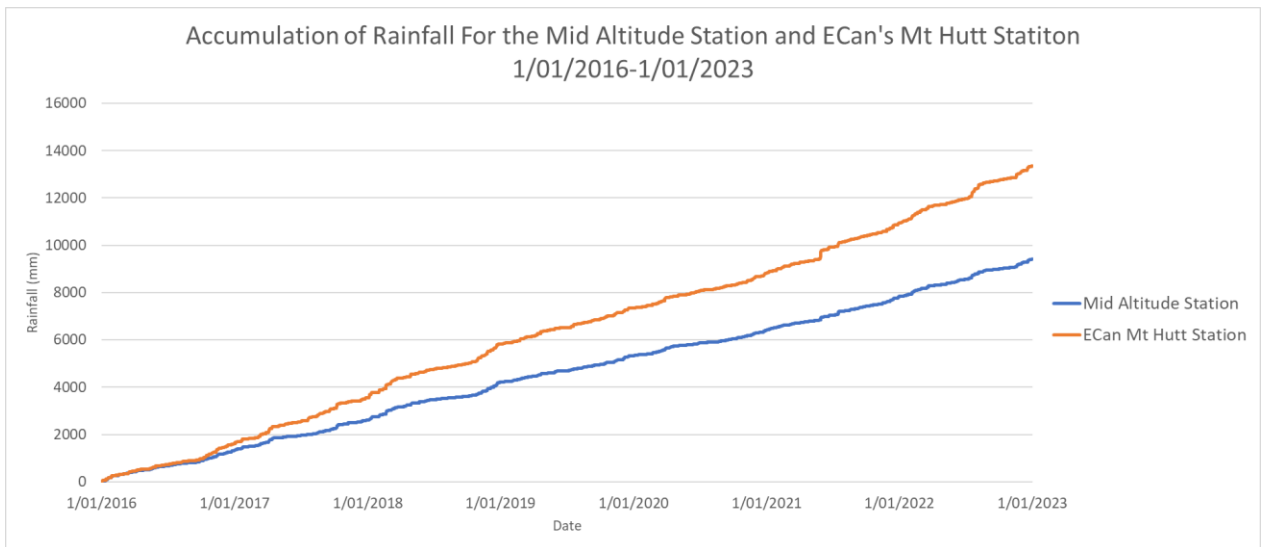


Figure 12. Line Graph showing the accumulation of rainfall recorded at the mid elevation site alongside the recordings from the ECan Mt Hutt Station. The recordings shown are during 1/01/2016-1/01/2023.

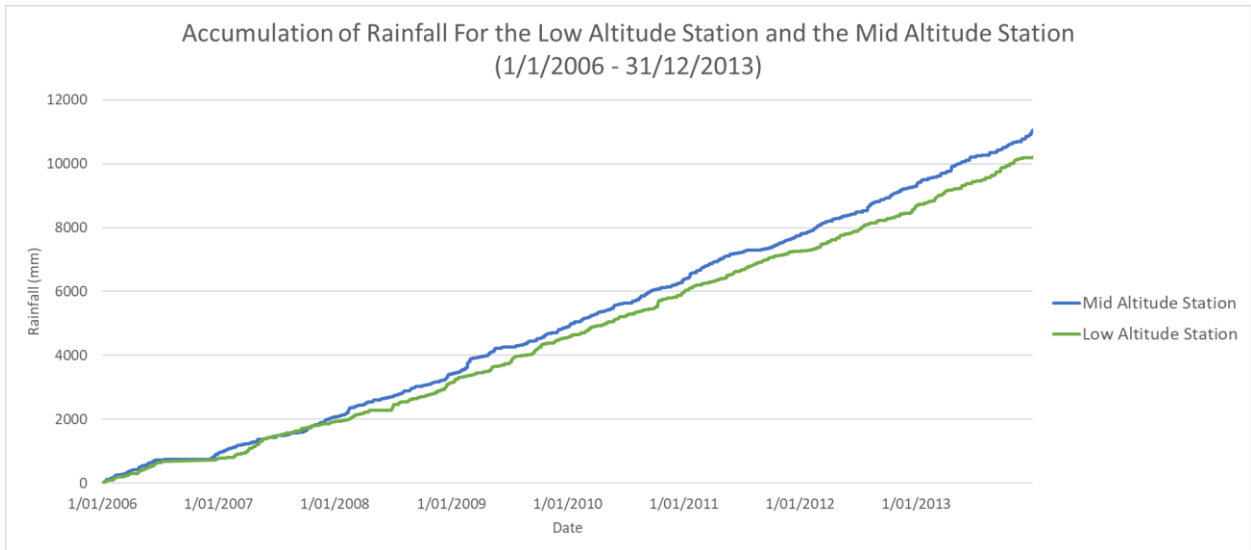


Figure 13 Line graph showing the accumulation of rainfall recorded at the low elevation site alongside the recordings from the mid elevation site. The recordings shown are from 1/01/2006-31/12/2013.

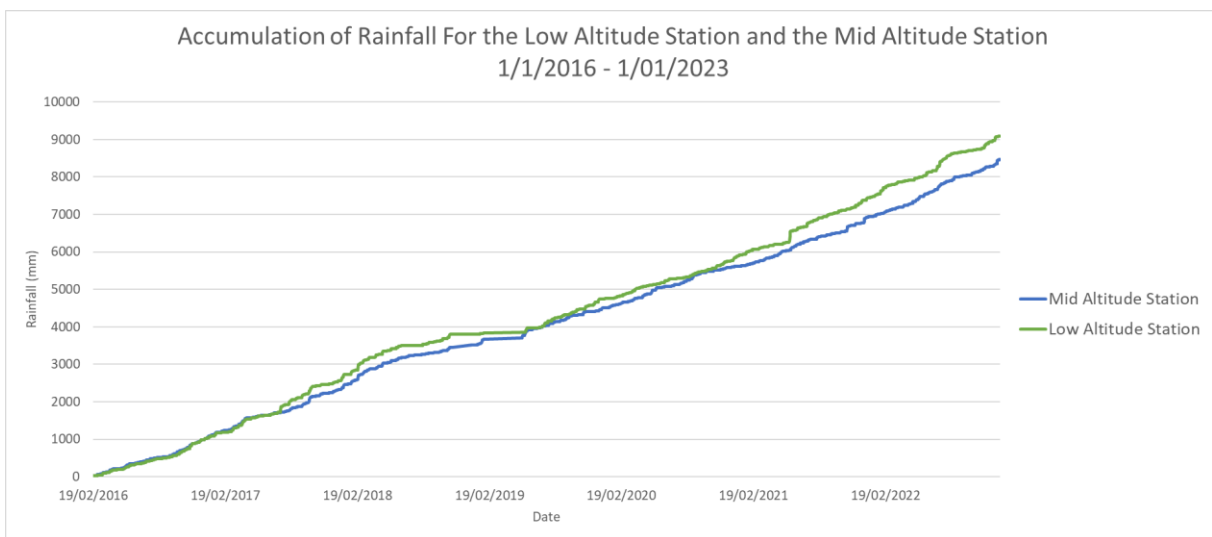


Figure 14. Line graph showing the accumulation of rainfall recorded at the low elevation site alongside the recordings from the mid elevation site. The recordings shown are from 1/01/2016-1/01/2023

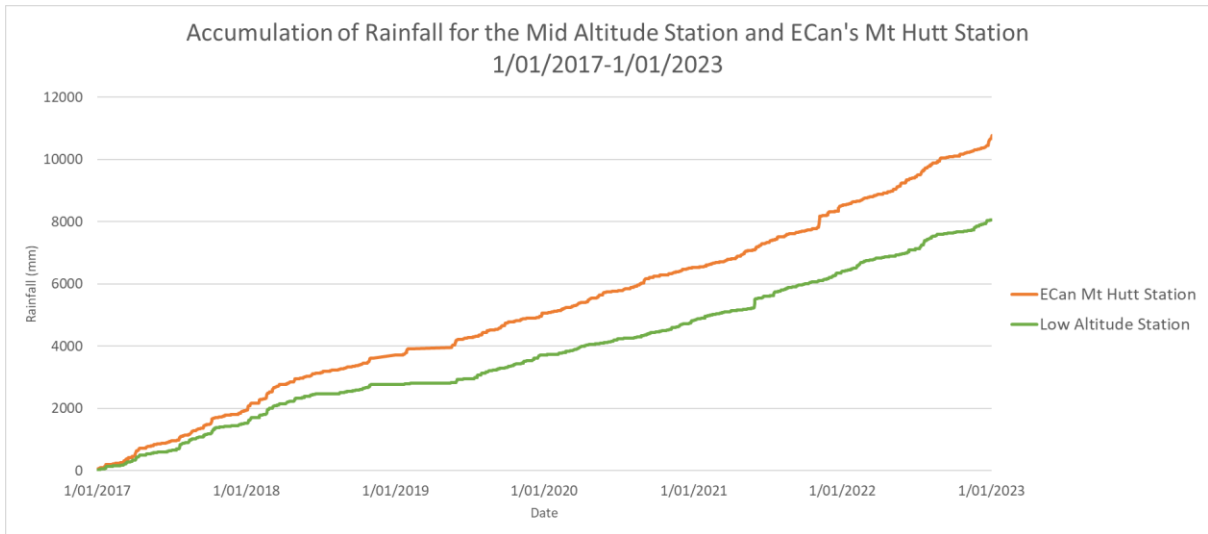


Figure 15. Line graph showing the accumulation of rainfall recorded at the low elevation site alongside the recordings from the ECan Mt Hutt station. The recordings shown are from 1/01/2017-1/01/2023

6.3 Wind

Figures 16 & 17 show how topography can alter the local wind patterns in an area. Figure 13 (Rakaia) shows that NW winds tend to produce the highest windspeeds, there seems to be much lower wind speeds at all the other wind directions. Comparing summer and winter plots in Figure 16, there tends to be lower winds speeds during the cooler winter months. This is supported by Figure 7's kernel density plot. N/NE winds tend to have the highest windspeeds in summer whereas the NW winds dominate during the winter months. Winchmore's site has little to no topographical relief which can interfere with wind patterns thus becoming more susceptible to large scale wind patterns rather than dynamic small-scale patterns.

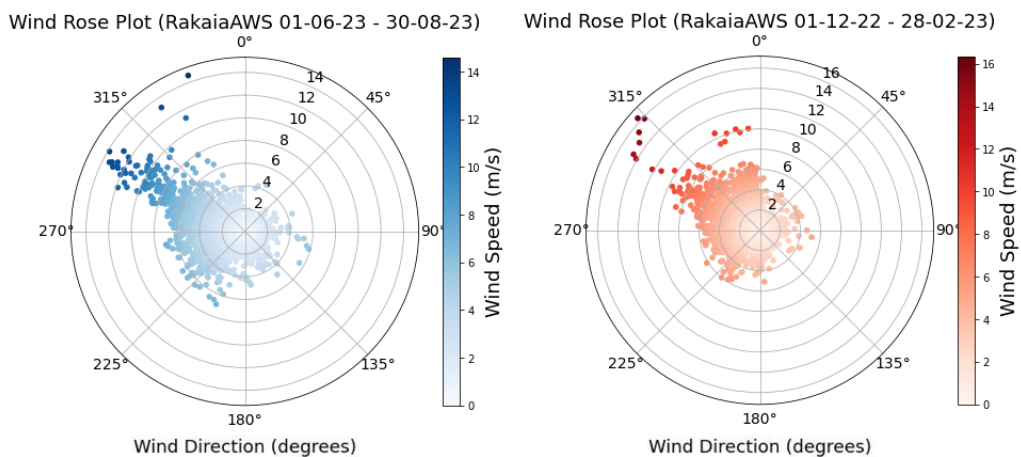


Figure 16. Wind rose plots for Rakaia WS representing Windspeed vs Direction (Left, 2023 Winter / Right, 2023 Summer)

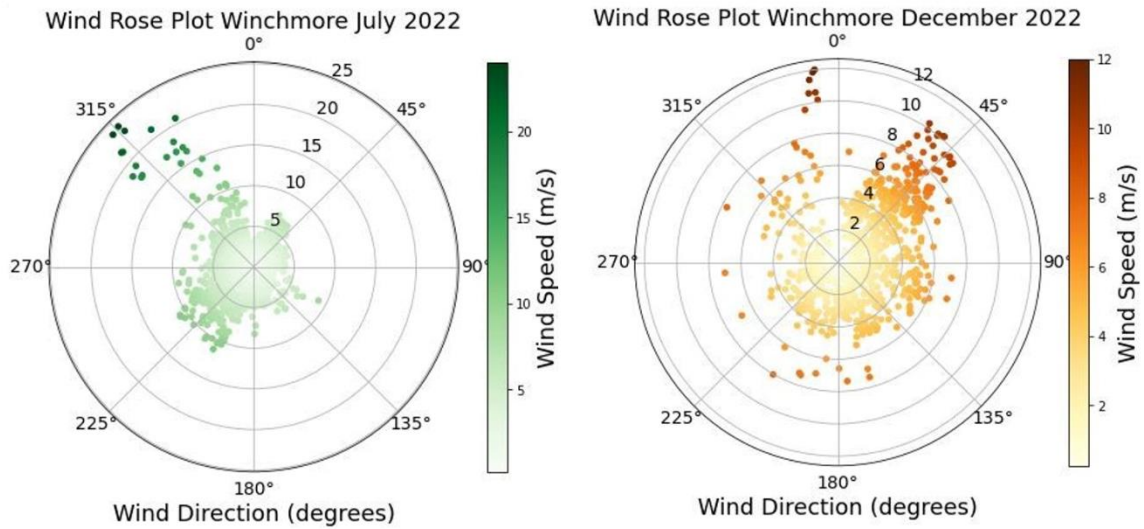


Figure 17. Wind rose plots for Winchmore WS representing Windspeed vs Direction. (Left, July 2022 / Right, December 2022)

7. Discussion

7.1 Lapse rates on Mt Hutt

13.1531042	14.2666667
19.9307083	21.1633333
18.2853958	17.49
15.59025	15.03
13.8756875	13.1666667
5.71145833	5.41466667
7.554375	10.3
14.7659792	17.2533333
16.6451042	15.4366667
15.28325	16.18
13.7654792	12.4166667
8.97545833	7.71
5.80814583	7.40666667
10.4606875	12.1166667
13.4421042	15.0866667
15.38825	17.3366667

Interlinking with the data provided by Tim Kerr and Heather Purdie has created a guided understating of how air temperature lapse rate occurs on Mount Hutt. Shown Figure 18, data in the left column represents 1090m (Kerr and Purdie) and 1070m data in the right column (Kelly). Although there are some days where the 1090m site experiences higher temperature recordings, the difference between the sites is insignificant. When the 1090m sight recorded cooler temperatures, it reflects active air temperature lapse rates that occurred

Figure 18 (to the left). Snapshot of the conditionally formatted data in excel showing the differences in temperature for the 1090m and 1070m sites. Specific to the first few days of January 2022.

7.2 Microclimates at Mt Hutt

Understanding microclimates at our community partners sites can build trust in the data to be representative. The first site at 450m is raised on a mound above a paddock with a line of trees towards the south that shelter it from strong wind. At 450m, snow is rare which keeps annual temperatures consistent with the wider region. From installation in 2005 to Feb 2020, a handful of eucalyptus trees shaded the weather station during winter months. At Noon sun was halfway up the tree when seen from solar panel. This is an environmental factor which may affect the representation of data for the wider region. These trees were cut down in February 2020.

The second 1070m site is located at Scott's Saddle (approx. ½ up access rd.) (figure 16). This site has had two weather station locations. The first location (1995-2005) was towards the road end of the ridge, slightly higher than the present site. This can be subject to higher winds and sun from all directions. The second location (2004-present) is located slightly south below the ridge, which is more protected from prevailing NW winds, this may also see shade earlier in the day compared to old site which will yield cooler air temperatures. Both sites are found in high density shrubbery which can obstruct sensors if growth is uncontrolled. Snowfall can occur at this elevation which lowers air temperature due to a high albedo. When snowfall is found, this lowers surface temperatures to 0 degrees Celsius regardless of air temperatures which can skew data. Snow at this elevation is intermittent and tends to melt within a few days/weeks.

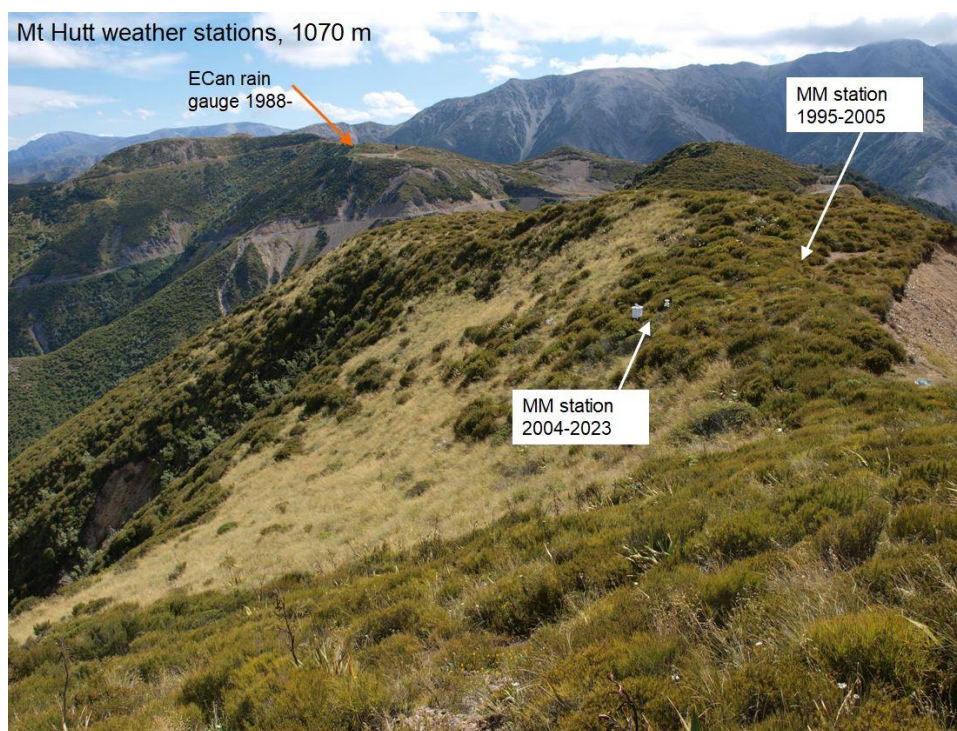


Figure 19: Mount Hutt MM Site 1070m (Kelly 2001)

The third site is located below the top car park of the Mt Hutt Ski Field. This site has had two weather station locations. The first location (2004-2009) was at 1560m near steep bluffs. The second location (2009-present) is located at 1520m in flatter basin, 40m below previous site with no major features in proximity other than a stream to the northeast and the main carpark to north. These sites are subject to constant snowpack throughout winter and into spring (June-October) which can at times bury the station. During summer, the site is subject to scree (loose rock) slopes with little vegetation.

7.3 Synoptic Scale Processes

Synoptic scale circulation represents a small cog in the larger global circulation system. The data analysed in this report has emphasised how synoptic conditions interlink with local

weather patterns. As mentioned, solar radiation plays a pivotal role in all-weather activity. Unequal distribution of heat leads to horizontal motion known as wind and is subsequently associated with vertical motion which causes weather.

Over seasonal to decadal time scales, the advection of air masses over the Southern Alps is heavily influenced by the position of the Westerly wind belt and meridional circulation anomalies of the New Zealand region (Clare et al., 2002). These westerly winds tend to be stronger in the Southern Hemisphere due to the scarcity of land mass. Surface roughness such as the Southern Alps can however reduce surface wind speed due to the dissipation of kinetic energy to molecular heat and friction. In a stable atmosphere, stronger airflow has sufficient energy and can be lifted accelerated when passing over hills or mountain tops. The interaction of airflow with the Southern Alps can cause the development of features such as the Foehn effect, Lee waves and severe downslope winds (Sturman & Spronken-Smith, 2001). According to Sturman and Spronken-Smith (2001), the interaction of synoptic systems with the mountains causes subtle changes in the pressure field perturbation to which local airflow responds, making mesoscale forecasting difficult. An example of this includes the lee-trough north-easterlies over Canterbury, which may combine with the local sea breeze system to produce enhanced on-shore flow – as shown in Figure 20.

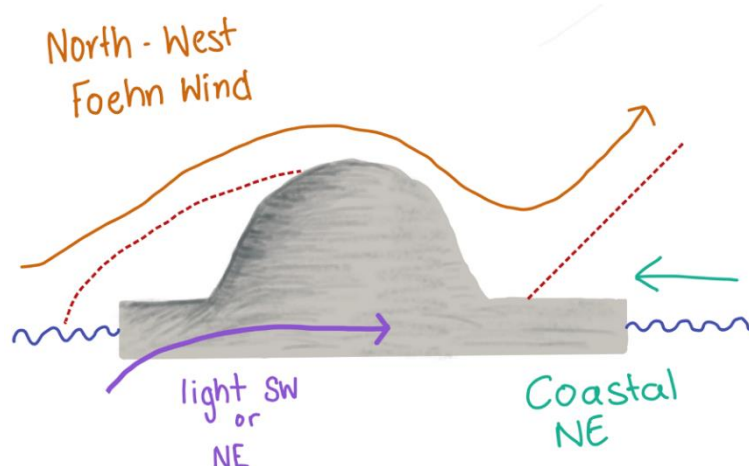


Figure 20. Schematic showing how the Southern Alps can generate Foehn events and how other directional systems are affected.

Synoptic conditions can also be the driver behind temperatures recorded and anomalies. There appears to be a large peak in temperature on February 22, 2021 seen in Figures 5 and 6. Further investigation was conducted to conclude whether synoptic circulation may be the cause for the high temperature recorded. Figure 21 displays mean sea level pressure (MSLP) maps acquired from the Australian Bureau of Meteorology (Bureau of Meteorology, 2019), for the day prior and the day of the peak temperature event. The maps indicate a large high-pressure system over the Canterbury region. The proximity of the isobars suggests there also was little to no wind on these days. As noted previously, high-pressure systems represent clear skies and the atmosphere being statically stable near the surface. Inferring from this, the peak in temperature is relatively normal given the synoptic conditions on that day.

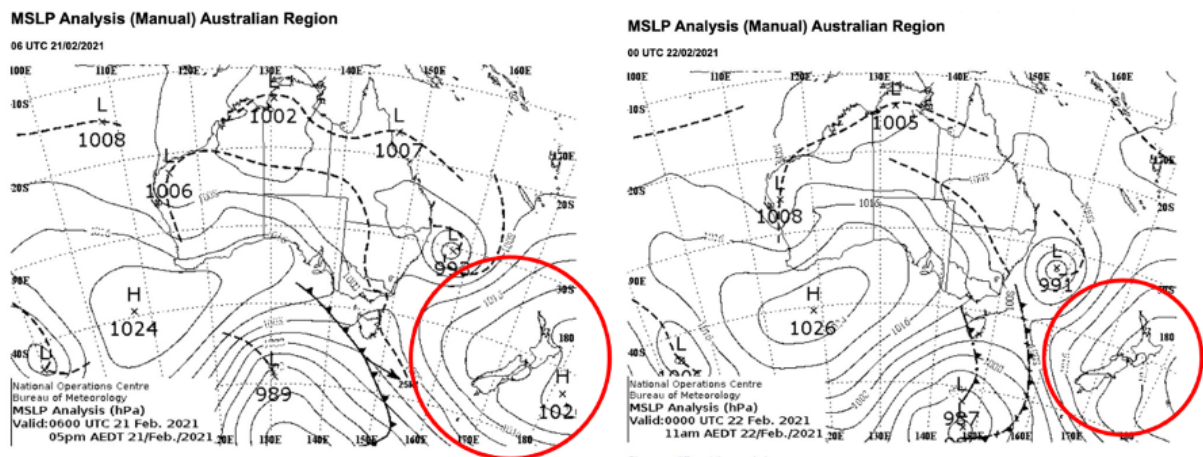


Figure 21. MSLP charts show the large high-pressure system migrating eastward over the South Island

8. Limitations

During this research project, constraints and challenges that impacted the scope and outcome of the study were observed. The following section provides the limitations faced during the research period.

The first limitation to mention is the time constraints of the project. The period for the project was twelve weeks and with many of the members having external influences reducing the time able to spend on the project. The time constraints led to the scale of the project being reduced.

Additionally, the quality of the weather station data provided from our community partner provided a limitation as our research heavily relied on it. Inconsistencies in data collection, recording errors and a lack of wind speed or direction measures raised concerns regarding the accuracy and reliability of the findings to be made. Additional station data from third party sources helped to mitigate this limitation.

The weather station data provided also showed challenges regarding homogeneity regarding the changes in instrumentation, location, and data collection practices over time. These factors can potentially add biases onto the data, leading to possible inaccuracies and misinterpretation and should be considered when interpreting the results.

9. Conclusion

Throughout the course of the project, many secondary data sources were collected to be used along with our study stations. The three NIWA stations gathering wind speed and direction, the temperature lapse rate project by RainfallNZ, and ECan's weather stations, were all used to observe the variability of the microclimate of the Southern Alps.

Observations on wind speed and direction, temperature, and rainfall were analysed for their effects in the Mt Hutt area, showing relationships between microclimates and synoptic weather patterns. The effect of the lapse rate is particularly prominent in this alpine region.

Alongside analysing the findings from the stations. The code produced will reduce labour and human error in future collection of this data, which occurs frequently though out the year.

10. Acknowledgements

We would like to acknowledge Professor Peyman Zawar-Reza for guiding us through the research and analysis process. We would also like to acknowledge Dave Kelly and his role in the Cass Research Area Management group for providing us with the data to perform an in-depth weather analysis. Both individuals have dedicated their time to helping us achieve objectives and gain an understanding of research processes. We would further like to acknowledge Tim Kerr and Associate Professor Heather Purdie for the use of their temperature lapse rate project data in our analysis. We would like to acknowledge Tipene Merritt for allowing us to engage with Mana Whenua. Lastly, we would like to acknowledge the rest of the GEOG309 team for making this project possible.

Appendix 1

```
#####  
#####  
# INPUT DATA: Adjust as needed  
# Excel sheet information for the station of interest (primary data) (file  
types .xls or .xlsx)  
file_Path <- 'GEOG309/MtHuttRawWeather_MM_2015-2023.xls'  
#directory of file  
sheet_name <- 'Raw' # name of the workbook sheet which contains the data.  
empty_rows = 1 # number of rows above the header, not including column names.  
site_name = 'MM' #choose from 'ML', 'MM', or 'MU'  
elevation = 1070 # elevation of the primary data site  
  
# Secondary data set information from a nearby station(s) (file type is .xls  
or .xlsx)  
nearby_file_Path<-"GEOG309/MtHuttRawWeather_ML450m.xls" #directory of file  
nearby_sheet_name <- 'Raw2017_2023'# name of the workbook sheet which  
contains the data.  
nearby_empty_rows = 2 # number of rows above the header, not including column  
names.  
nearby_elevation = 450 # elevation of the secondary data site  
nearby_source = 'ML' #choose from 'rainfallNZ', 'ML', 'MM', or 'MU'  
  
# dates of interest  
start_hour <- 8 # time data is collected each day for the primary station  
data set.  
start_date<- ymd(20220101) # in format year month day  
end_date <- ymd(20221230)  
  
# File name for the result to write to:
```

```

output_path = "GEOG309/ReliableData_Example.xlsx"
#####
#####
lapse_rate = -0.05 # Value obtained from RainfallNZ transect: rate is -0.05
degrees/metre.
# setup libraries and package installs-----
-----

install.packages('tidyverse')
install.packages("readxl")
install.packages("writexl")
install.packages("openxlsx")
library(writexl)
library(tidyverse)
library(lubridate)
library(readxl)
library(data.table)
library(openxlsx)

#####
#####
# Define Functions
extract_day = function(datetime_obj) {
  #floor the date according to the time of day specified
  # Extract the year, month, and day from the date-time object
  date_part <- as.Date(datetime_obj)

  # Check if the hour for each date-time is earlier than the specified
  starting hour
  for (i in 1:length(datetime_obj)) {
    if (hour(datetime_obj[i]) < start_hour) {
      # Subtract one day from the corresponding date
      date_part[i] <- date_part[i] - days(1)
    }
  }

  return(date_part)
}

lapse_rate_adjust = function(value){
  # apply a correction to secondary data accounting for the lapse rate
  adjusted_value = value + (nearby_elevation-elevation)*lapse_rate
  return(adjusted_value)
}

rename_columns <- function(data_set){
  #generalise column names for further processing.
  data_set <- data_set %>%
    rename_with(~gsub("Me", "Mean", .x), .cols = contains("Me"))
  data_set <- data_set %>%
    rename_with(~gsub("av", "Mean", .x), .cols = contains("av"))
  data_set <- data_set %>%
    rename_with(~gsub("avg", "Mean", .x), .cols = contains("avg"))
  data_set <- data_set %>%
    rename_with(~gsub("mx", "Max", .x), .cols = contains("mx"))
  data_set <- data_set %>%

```



```

    rename_with(~gsub("Mx", "Max", .x), .cols = contains("Mx"))
data_set <- data_set %>%
  rename_with(~gsub("av", "Mean", .x), .cols = contains("av"))
data_set <- data_set %>%
  rename_with(~gsub("Mn", "Min", .x), .cols = contains("Mn"))
data_set <- data_set %>%
  rename_with(~gsub("mn", "Min", .x), .cols = contains("mn"))
data_set <- data_set %>%
  rename_with(~gsub("mean", "Mean", .x), .cols = contains("mean"))
data_set <- data_set %>%
  rename_with(~gsub("min", "Min", .x), .cols = contains("min"))
data_set <- data_set %>%
  rename_with(~gsub("max", "Max", .x), .cols = contains("max"))
data_set <- data_set %>%
  rename_with(~gsub("Min.batt", "Battery", .x), .cols = contains("batt"))
data_set <- data_set %>%
  rename_with(~gsub("Meanan", "Mean", .x), .cols = contains("Meanan"))
return(data_set)
}

#####
#####
# Read data from Excel
temps2 <- read_excel(nearby_file_Path,
                     sheet = nearby_sheet_name,
                     skip = nearby_empty_rows,
                     trim_ws = TRUE)

if (nearby_source == 'rainfallNZ') {
  # Rename columns with elevation values
  temps2 <- temps2 %>%
    rename('1755' = Site1,
           '1650' = Site2,
           '1490' = Site3,
           '1340' = Site4,
           '1205' = Site5,
           '1090' = Site6,
           '1000' = Site7,
           '880' = Site8,
           '175' = Site9,
           '560' = Site10
    )
  number_strings <- colnames(temps2)
  # Target value
  target_value <- elevation
  # Convert string values to numeric
  numbers <- as.numeric(number_strings)
  # Calculate the absolute differences
  differences <- abs(numbers - target_value)
  # Find the index of the closest number
  closest_index <- which.min(differences)
  # Get the closest number
  nearestSite <- as.character(numbers[closest_index])

  # Select relevant columns and compute daily statistics
  day_values <- temps2 %>%
    group_by(day = extract_day(Date)) %>%

```

```

    summarise(
      site_avg = mean(!sym(nearestSite), na.rm = TRUE),
      site_max = max(!sym(nearestSite), na.rm = TRUE),
      site_min = min(!sym(nearestSite), na.rm = TRUE)
    )
  day_values$day <- day_values$day + days(1)
  # Apply lapse rate correction
  day_values[, -which(names(day_values) == 'day')] <- lapply(
    day_values[, -which(names(day_values) == 'day')],
    lapse_rate_adjust
  )
} else {
  temps2 <- rename_columns(temps2)
  # Clean data by removing values outside physical limits
  day_values <- temps2 %>%
    select(date)
  day_values <- day_values %>%
    rename(day = date)
  temps2 <- temps2 %>% select(contains(c("Mean", "Min", "Max")))
  temps2 <- temps2 %>%
    mutate(
      across(everything(), ~ case_when(.x > 50 ~ NA_real_,
                                       .x < -30 ~ NA_real_,
                                       TRUE ~ .x))
    )
  min_columns <- temps2 %>%
    select(contains("Min"))
  min_columns <- min_columns %>%
    rowwise() %>%
    mutate(min = mean(c_across(everything()), na.rm = TRUE))

  mean_columns <- temps2 %>%
    select(contains("Mean"))
  mean_columns <- mean_columns %>%
    rowwise() %>%
    mutate(mean = mean(c_across(everything()), na.rm = TRUE))
  max_columns <- temps2 %>%
    select(contains("Max"))
  max_columns <- max_columns %>%
    rowwise() %>%
    mutate(max = mean(c_across(everything()), na.rm = TRUE))
  day_values$site_min <- min_columns$min
  day_values$site_avg <- mean_columns$mean
  day_values$site_max <- max_columns$max
}

#####
#####
# create datetime dataframe -----
-----

unit_period = period(1, 'days')

date <- seq(from = start_date, to = end_date, by = "days")

dates_df <- data.frame(date)

```

```

dates_df$JulianDayOfYear <- yday(dates_df$date)
dates_df$Year <- year(dates_df$date)

dates_df$DayOfYear <- paste(dates_df$JulianDayOfYear, dates_df$Year)

#####
#####
# create data frame from excel spreadsheet and merge with date data frame--
-----

Excel_data <- read_excel(file_Path, sheet = sheet_name, skip=empty_rows, trim_ws = TRUE)

Excel_data$DayOfYear <- paste(Excel_data$day, Excel_data$year)

data_set = left_join(dates_df, Excel_data, by = join_by(x$DayOfYear == y$DayOfYear))
data_set <- setorder(data_set, date)
data_set <- data_set %>% select(-year, -JulianDayOfYear, -day)
#####
#####
# Rename columns for consistency

data_set <- rename_columns(data_set)
# Filter data -----
-----

Reliable_data <- data_set %>%
  select(-DayOfYear, -date, -Year, -time, -Battery)

# remove values outside physical limits (generous estimates [max 50 min -30])
Reliable_data <- Reliable_data %>%
  mutate(
    across(everything(), ~ case_when(.x > 50 ~ NA_real_,
                                     .x < -30 ~ NA_real_,
                                     TRUE ~ .x))
  )

#####
#####
data_set$BatteryFailure <- 0

# data_set
data_set <- data_set %>%
  mutate(BatteryFailure = ifelse(Battery < 10, BatteryFailure + 1, BatteryFailure))
data_set <- data_set %>%
  mutate(BatteryFailure = ifelse(Battery > 13.4, BatteryFailure + 1, BatteryFailure))
Reliable_data$BattFailed <- data_set$BatteryFailure
Reliable_data <- Reliable_data %>%
  mutate(BattFailed = ifelse(BattFailed > 0, TRUE, FALSE))

```

```

# remove max panel default failure values
Reliable_data <- Reliable_data %>%
  mutate(Max.panel= ifelse(Max.panel==35.44 & BattFailed==TRUE,
NA,Max.panel) )

#####
#####
# Average across different sensor readings
# MIN
min_columns <- Reliable_data %>%
  select(contains("Min"))

row_averages1 <- min_columns %>%
  rowwise() %>%
  mutate(Min = mean(c_across()),na.rm = TRUE))

Reliable_data$Min <- row_averages1$Min
# MEAN
mean_columns <- Reliable_data %>%
  select(contains("Mean"))

row_averages2 <- mean_columns %>%
  rowwise() %>%
  mutate(Mean = mean(c_across()),na.rm = TRUE))

Reliable_data$Mean <- row_averages2$Mean

# MAX
max_columns <- Reliable_data %>%
  select(contains("Max"))

row_averages3 <- max_columns %>%
  rowwise() %>%
  mutate(Max = mean(c_across()),na.rm = TRUE))

Reliable_data$Max <- row_averages3$Max
Reliable_data$date <- data_set$date
day_values <- day_values %>%
  mutate(day = format(day, format = "%Y-%m-%d"))
Reliable_data <- Reliable_data %>%
  mutate(date = format(date, format = "%Y-%m-%d"))
day_values <- day_values %>%
  distinct(day, .keep_all = TRUE)

#####
#####
# compare primary and secondary station data
stations_data_set = left_join(Reliable_data,day_values,by = c('date'='day'))
cols = c('date', 'Min', 'site_min', 'Mean', 'site_avg', 'Max', 'site_max')
stations_data_set <- stations_data_set %>%
  select(all_of(cols))

#####
#####
# MEAN

stations_data_set <- stations_data_set %>%

```

```

mutate(Mean_diff = stations_data_set$site_avg - stations_data_set$Mean)
av_bias      =      sum(stations_data_set$Mean_diff,na.rm      =
TRUE)/length(stations_data_set$Mean)

stations_data_set <- stations_data_set %>%
mutate(MeanBiasCorrection = stations_data_set$site_avg - av_bias)

stations_data_set <- stations_data_set %>%
mutate(Mean_diff      =      stations_data_set$MeanBiasCorrection      -
stations_data_set$Mean)
av_bias      =      sum(stations_data_set$Mean_diff,na.rm      =
TRUE)/length(stations_data_set$Mean)

threshold_mean = max(stations_data_set$Mean_diff,na.rm=TRUE)

mean_columns <- Reliable_data %>%
select(contains("Mean"),-Mean)

mean_columns <- mean_columns %>%
mutate(across(everything(), ~ ifelse((.x > .x+threshold_mean | .x < .x-
threshold_mean), NA,.x)))

row_averages4 <- mean_columns %>%
rowwise() %>%
mutate(Mean = mean(c_across()),na.rm = TRUE))

# MIN
stations_data_set <- stations_data_set %>%
mutate(Min_diff = stations_data_set$site_min - stations_data_set$Min)
min_bias      =      sum(stations_data_set$Min_diff,na.rm      =
TRUE)/length(stations_data_set$Min)

stations_data_set <- stations_data_set %>%
mutate(MinBiasCorrection = stations_data_set$site_min - min_bias)

stations_data_set <- stations_data_set %>%
mutate(Min_diff      =      stations_data_set$MinBiasCorrection      -
stations_data_set$Min)
min_bias      =      sum(stations_data_set$Min_diff,na.rm      =
TRUE)/length(stations_data_set$Min)

threshold_min = max(stations_data_set$Min_diff,na.rm=TRUE)

min_columns <- Reliable_data %>%
select(contains("Min"),-Min)

min_columns <- min_columns %>%
mutate(across(everything(), ~ ifelse((.x > .x+threshold_min | .x < .x-
threshold_min), NA,.x)))

row_averages5 <- min_columns %>%
rowwise() %>%
mutate(Min = mean(c_across()),na.rm = TRUE))

# Max
stations_data_set <- stations_data_set %>%
mutate(Max_diff = stations_data_set$site_max - stations_data_set$Max)

```

```

mx_bias          =          sum(stations_data_set$Max_diff,na.rm          =
TRUE)/length(stations_data_set$Max)

stations_data_set <- stations_data_set %>%
  mutate(MaxBiasCorrection = stations_data_set$site_max - mx_bias)

stations_data_set <- stations_data_set %>%
  mutate(Max_diff          =          stations_data_set$MaxBiasCorrection          -
stations_data_set$Max)
mx_bias          =          sum(stations_data_set$Max_diff,na.rm          =
TRUE)/length(stations_data_set$Max)

threshold_max = max(stations_data_set$Max_diff,na.rm=TRUE)

max_columns <- Reliable_data %>%
  select(contains("Max"),-Max)

max_columns <- max_columns %>%
  mutate(across(everything(), ~ ifelse((.x > .x+threshold_max| .x < .x-
threshold_max), NA,.x)))
# max_columns
row_averages6 <- max_columns %>%
  rowwise() %>%
  mutate(Max = mean(c_across()),na.rm = TRUE))

Reliable_data$Max_adjusted <- row_averages6$Max
Reliable_data <- Reliable_data %>%
  mutate(
    across(Max_adjusted, ~ case_when(.x < Mean ~ Max,
TRUE ~ .x))
  )
Reliable_data$Min_adjusted <- row_averages5$Min
Reliable_data <- Reliable_data %>%
  mutate(
    across(Min_adjusted, ~ case_when(.x > Mean ~ Min,
TRUE ~ .x))
  )
Reliable_data$Mean_adjusted <- row_averages4$Mean

Reliable_data <- Reliable_data %>%
  mutate(
    across(Mean_adjusted, ~ case_when(.x > Max ~ Mean,
.x < Min ~ Mean,
TRUE ~ .x))
  )

#####
#####
# Select data for outputs
write_out1 <- Reliable_data %>%
  select(date,Max_adjusted,Mean_adjusted,Min_adjusted,BattFailed)
write_out1 <- write_out1 %>%
  rename('Max' = Max_adjusted,
'Min' = Min_adjusted,

```

```

      'Mean' = Mean_adjusted)
write_out2 <- write_out1
write_out2$Max <- ifelse(is.na(Reliable_data$Max),
                        stations_data_set$MaxBiasCorrection,
                        Reliable_data$Max)
write_out2$Mean <- ifelse(is.na(Reliable_data$Mean),
                          stations_data_set$MeanBiasCorrection,
                          Reliable_data$Mean)
write_out2$Min <- ifelse(is.na(Reliable_data$Min),
                          stations_data_set$MinBiasCorrection,
                          Reliable_data$Min)

# Write data to an Excel file with multiple sheets
write_xlsx(list(SingleStationValues = write_out1,
               SubstitutedValues = write_out2,
               RawData = data_set),
           output_path)

```

```

#####
# INPUT DATA: Adjust as needed
# Excel sheet information for the station of interest (primary data) (file types .xls or .xlsx)
file_Path <- 'GEOG309/MtHuttRawWeather_MM_2015-2023.xls' #directory of file
sheet_name <- 'Raw' # name of the workbook sheet which contains the data.
empty_rows = 1 # number of rows above the header, not including column names.
site_name = 'MM' #choose from 'ML', 'MM', or 'MU'
elevation = 1070 # elevation of the primary data site

# Secondary data set information from a nearby station(s) (file type is .xls or .xlsx)
nearby_file_Path <- "GEOG309/MtHuttRawWeather_ML450m.xls" #directory of file
nearby_sheet_name <- 'Raw2017_2023' # name of the workbook sheet which contains the data.
nearby_empty_rows = 2 # number of rows above the header, not including column names.
nearby_elevation = 450 # elevation of the secondary data site
nearby_source = 'ML' #choose from 'rainfallNG', 'ML', 'MM', or 'MU'

# dates of interest
start_hour <- 8 # time data is collected each day for the primary station data set.
start_date<- ymd(20220101) # in format year month day
end_date <- ymd(20221230)

# File name for the result to write to:
output_path = "GEOG309/ReliableData_Example.xlsx"
#####

```

References

- Barry, R. G., & Chorley, R. J. (2009). Atmosphere, weather and climate. *Routledge*.
- Beele, E., Reyniers, M., Aerts, R., & Somers, B. (2022). Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium. *Earth System Science Data*, 4681-4717.
- Bureau of Meteorology. (2019). *Latest Colour Mean Sea-Level Pressure Analysis*. Bom.gov.au; Bureau of Meteorology. http://www.bom.gov.au/australia/charts/synoptic_col.shtml

- CliFlo, 2023. The National Climate Database NIWA <https://cliflo.niwa.co.nz/>
- Clare, G. R., Fitzharris, B. B., Chinn, T. J. H., & Salinger, M. J. (2002). Interannual variation in end-of-summer snowlines of the Southern Alps of New Zealand, and relationships with Southern Hemisphere atmospheric circulation and sea surface temperature patterns. *International Journal of Climatology*, 22(1), 107–120. <https://doi.org/10.1002/joc.722>
- Daidzic, N. E. (2019). On atmospheric lapse rates. *International Journal of Aviation Aeronautics and Aerospace*, 6(4). <https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=1062&context=avia-fac-pubs>
- Dawe, T., 2022. GEOG211 Mountain Weather and Climates
- De Frenne, P., Lenoir, J., Luoto, M., Scheffers, B. R., Zellweger, F., Aalto, J., Ashcroft, M. B., Christiansen, D. M., Decocq, G., De Pauw, K., Govaert, S., Greiser, C., Gril, E., Hampe, A., Jucker, T., Klings, D. H., Koelemeijer, I. A., Lembrechts, J. J., Marrec, R., & Meeussen, C. (2021). Forest microclimates and climate change: Importance, drivers and future research agenda. *Global Change Biology*, 27(11), 2279–2297. <https://doi.org/10.1111/gcb.15569>
- Fridley, J. D. (2009). Downscaling Climate over Complex Terrain: High Finescale (<1000 m) Spatial Variation of Near-Ground Temperatures in a Montane Forested Landscape (Great Smoky Mountains)*. *Journal of Applied Meteorology and Climatology*, 48(5), 1033–1049. <https://doi.org/10.1175/2008jamc2084.1>
- Gultepe, I. (2015). Mountain weather: Observation and Modelling. *Advances in Geophysics*, 56, 229–312. https://www.researchgate.net/profile/IsmailGultepe/publication/281717358_Mountain_Weather_Observation_and_Modeling/links/59f1f4c4a6fdcc1dc7bb0a7b/Mountain-Weather-Observation-and-Modeling.pdf
- Hahn, C., Garcia-Marti, I., Sugier, J., Emsley, F., Anne-Lise Beaulant, Oram, L., Strandberg, E., Lindgren, E., Sunter, M., & Ziska, F. (2022). *Observations from Personal Weather Stations—EUMETNET Interests and Experience*. 10(12), 192–192. <https://doi.org/10.3390/cli10120192>
- Kamal, A., Abidi, S. M. H., Mahfouz, A., Kadam, S., Rahman, A., Hassan, I. G., & Wang, L. L. (2021). Impact of urban morphology on urban microclimate and building energy loads. *Energy and Buildings*, 253, 111499. <https://doi.org/10.1016/j.enbuild.2021.111499>
- Liu, T., & Zhang, D. (2021). Advances in the Quality Control Methods of Air Temperature Data at Surface Automatic Weather Stations. *IOP Conference Series: Earth and Environmental Science*.
- Maclean, I. M. D., Duffy, J. P., Haesen, S., Govaert, S., De Frenne, P., Vanneste, T., Lenoir, J., Lembrechts, J. J., Rhodes, M. W., & Van Meerbeek, K. (2021). On the measurement of microclimate. *Methods in Ecology and Evolution*, 12(8), 1397–1410. <https://doi.org/10.1111/2041-210x.13627>
- Mendelsohn, R., Kurukulasuriya, P., Basist, A., Kogan, F., & Williams, C. (2007). Climate analysis with satellite versus weather station data. *Climatic Change*, 81(1), 71–83. <https://doi.org/10.1007/s10584-006-9139-x>
- Mount Hutt Skifield - Methven / Ashburton's treasures / Schools' Stories / Home - Living Heritage. (n.d.). www.livingheritage.org.nz. Retrieved October 19, 2023, from <https://www.livingheritage.org.nz/Schools-Stories/Ashburton-s-treasures/Mount-Hutt-Skifield-Methven#:~:text=In%201972%2C%20Mount%20Hutt%20began>
- Nature Climate Change. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, 5(5), 424–430. <https://doi.org/10.1038/nclimate2563>
- Skrynyk, O., Sidenko, V., Aguilar, E., Guijarro, J., Skrynyk, O., Palamarchuk, L., . . . Osadchy, V. (2023). Data quality control and homogenization of daily precipitation and air temperature (mean, max and min) time series of Ukraine. *International Journal of Climatology*.

Sturman, A., & Spronken-Smith, R. (Eds.). (2001). *The Physical Environment: A New Zealand Perspective*. Oxford University Press.

Sturman, A., & Wanner, H. (2001). The Alps in Europe and the Southern Alps in New Zealand. *Source: Mountain Research and Development*, 21(4), 359–369.

Wratt, D. S., Ridley, R. N., Sinclair, M. R., Larsen, H., Thompson, S. M., Henderson, R., Austin, G. L., Bradley, S. G., Auer, A., Sturman, A. P., Owens, I., Fitzharris, B., Ryan, B. F., & Gayet, J.-F. . (1996). The New Zealand Southern Alps Experiment. *Bulletin of the American Meteorological Society*, 77(4), 683–692. [https://doi.org/10.1175/1520-0477\(1996\)077%3C0683:TNZSAE%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3C0683:TNZSAE%3E2.0.CO;2)