Data summary procedures for NorWeST
Upper Columbia Yakima Processing Unit

NorWeST Stream Temp

Data for this project should be considered provisional. We have attempted to remove erroneous data but not all can be caught. Each user should run their own QA/QC on this data to fit the objectives of their study.

Data QA/QC procedures

Data from each source is imported into a Deployment table and a RawData table

Deployment table includes at minimum the following fields:

Field Name	Format	Description
Perma_FID	Numeric: Long Integer	Unique site value assigned by RMRS
Stream	Text	Stream name
Source	Text	Agency contributing the temperature data
Srce_Name	Next	Name of person that sent the temperature data
Deployment_Code	Text	Unique value for each survey. It is the concatenation of Perma_FID, StartDate and EndDate
StartDate	Date: mm/dd/yyyy	First full day of data
EndDate	Date: mm/dd/yyyy	Last full day of data
UOM	Text	Unit of measure of the contributed data: DegF = degrees Fahrenheit; DegC = degrees Celsius
UseData	Text	Y=deployment is good to use; N=deployment should not be used
Interval	Numeric: double	Interval in minutes between recording
Filename	Text	Filename of the contributed data
RMRS_Comments	Text	RMRS Comments added such as if some of the observations were considered erroneous or why the deployment was not used

Other data is imported into this table if it is available such as Lat/Long or comments from the contributor. Much of this table may be provided by the contributor. It is possible the start date and end date will differ from the data if that day was found to be a partial day.

Structure of the RawData table:

Field Name	Format	Description
Deployment_Code	Text	Link to Deployment table
SampleDate	Date: mm/dd/yyyy	Date of observation
SampleTime	Time: hh:mm	Time of observation
Temperature	Numeric: double	Temperature
UseData	Text	Y=observation can be used (default value); N=observation needs to be excluded

Initial cleaning of temperature data

Each source's data is imported into Access as a Deployment Table and RawData table:

- Once all raw data for a source has been combined into one file a check for duplicate Perma_FID*SampleDate*SampleTime is conducted using the Find Duplicates query in Access. If any duplicates are found, they are corrected or removed from the analysis if it is found that 2 different data streams were given for a single Perma_FID.
- The Find Unmatched query is used to determine if there are any Deployments that do not have rawdata and vice versa

SAS statistical software is used to make the following checks:

- If UOM is "DegF" then TemperatureC=(Temperature-32)*(5/9)
  - Else TemperatureC=Temperature
- If SampleDate is less than StartDate then UseData=N
- If SampleDate is EndDate then UseData=N
- If Temperature is less than -1C then UseData=N

Once the initial cleaning is done then graphs are created for each deployment where UseData=Y. These graphs are visually inspected to determine if other data needs to be removed from the analysis (set UseData=N).

Calculation of daily metrics:

Daily metric calculations are also carried out by source. Again, SAS statistical software is used to make the calculations.

The following Metrics are output:

- DailyMean: Average daily temperature
- DailyMaximum: Maximum daily temperature
- DailyMinimum: Minimum daily temperature
- DailyRange: Maximum daily minus minimum daily temperature
- N_Obs: Number of observations used for the SampleDate

The mode is calculated for each deployment, and this is used to compare to N_Obs. If N_Obs is less than 0.9*mode, then the day is considered a partial day and is not used in the analysis (UseDay=N). There is still a potential that partial days get through this procedure, but we feel this will catch the majority of the partial days.

The resulting data table is imported in Access and each source is appended. After each source is appended the find duplicate query is run to determine if multiple sources sampled the same site on the same date. If multiple contributors were found to have sampled the same site, then the source with the longest stream of data for that year is used and the other data is set to UseDay=N.

Once this dataset is created the data that was contributed as only daily metrics is appended. This dataset is exported, and SAS is used to calculate the 7-day running mean maximum and the 7-day running mean minimum.

- 7-day running mean maximum is the average of the maximum temperatures on the day with the 6 previous days

- 7-day running mean minimum is the average of the minimum temperature on the day with the 6 previous days
- If there is a day missing, then the calculations start over again

The Daily metrics table provided for download is:

Field Name	Format	Description
OBSPRED_ID	Numeric: long integer	Unique value assigned for the Perma_FID*Year
NorWeST_ID	Text	Represents the combination of the Processing Unit and OBSPRED_ID
SampleDate	Date: mm/dd/yyyy	Sample date
DailyMax	Numeric: double	Daily maximum temperature
DailyMin	Numeric: double	Daily minimum temperature
DailyMean	Numeric: double	Daily average temperature
DailyRange	Numeric: double	Daily range (max temp – min temp)
N_Obs	Numeric: long integer	Number of observations on the sample date Note: if data was submitted as daily metrics only then this value is 1
Mov_7mmx	Numeric: double	Moving 7 day mean maximum
Mov_7mmn	Numeric: double	Moving 7 day mean minimum

Monthly Metrics:

Daily metrics are used to summarized monthly metrics. Monthly metrics are output for months with at least 90% of days available:

- January, March, May, July, August, October, December – greater than 28 days
- April, June, September, November – greater than 27 days
- February – greater than 25 days

Field Name	Format	Description
OBSPRED_ID	Numeric: long integer	Unique value assigned for the Perma_FID*Year
NorWeST_ID	Text	Represents the combination of the Processing Unit and OBSPRED_ID
SampleYear	Numeric: long integer	Year of sample
MonthlyMax	Numeric: double	Maximum of the daily maximums for the month
MonthlyMin	Numeric: double	Minimum of the daily minimums for the month
MonthlyMean	Numeric: double	Average of the daily averages for the month
MaxDailyRange	Numeric: double	Maximum of the daily ranges for the month
NDays	Numeric: long integer	Number of days summarized for the month
SampleMonth	Numeric: long integer	Numeric value of the sample month (i.e. January=1)

For more information, see the NorWeST website