**Software Specification for the calculation of National Climate Monitoring Products - Annex to Guidance for the production of National Climate Monitoring Products**

The following specification details the mathematical steps necessary to calculate the six National Climate Monitoring Product (NCMPs). The context for these calculations is given in the Guide to National Climate Monitoring Products to which this document is an annex.

The software specification is split into two distinct parts. The first is the calculation of indices for individual stations. The second concerns the use of those indices to calculate area-averages and other outputs for the NCMPs.

The R-NCMPs User Manual provides step-by-step instructions on the installation of R, preparation of input data, quality control, computation of the indices at station level, and computation of the variograms and region-averages. It also includes the mathematical description of each NCMPs.

**1. Instructions for programming the 6 NCMPs at station level**

Let *Txij*and *Tnij* be the daily maximum and minimum temperature for day i and period j, *Tmij* is the mean of *Txij* and *Tnij* when both values are available. *Prij* is the daily precipitation amount on day i and period j. The temperature is given in °C and the precipitation in mm. Monthly NCMPs are calculated if no more than 6 days are missing in a month; annual NCMPs are calculated if no more than 18 days are missing in a year; the climatology is calculated if no more than 7 years are missing in the 30-year period; otherwise it is missing. The climatology is the mean of all monthly values of the base period. The percentiles are calculated for each day using a 5-day running mean for the base period. NCMPs are calculated for each month and for the year.

**Input data**

. ASCII text file, one file by station

. columns are Year, Month, Day, *Prij, Txij, Tnij*

. example: 1950 07 15 2.5 25.0 10.2

. missing value is -99.9

. records must follow the calendar date order

**NCMP1: mean temperature anomalies (°C)**

. monthly and annual mean temperature (*TMj*): average of *Tmij* over the month (year)

. climatology (*CTj)*: average of *TMj* over 1981-2010

. monthly and annual mean temperature anomaly (*TMAj*): *TMj* - *Cj*

**NCMP2: percentage rainfall (%)**

. monthly and annual total precipitation (*PRj*): sum of *Prij* over the month (year)

. climatology (*CPj*): average of *PRj* over 1981-2010

. monthly and annual precipitation ratio (*Rj*): *PRj* / *CPj* x 100

. monthly and annual precipitation normalized anomalies (*PRAj*): (*PRj* - *CPj*) / *CPj* x 100

. monthly and annual precipitation anomalies: (*PRj* - *CPj*)

**NCMP3: standardized precipitation index**

. 13 precipitation series (1 for each month and 1 for annual)

. calculate fraction of months with zero rain and fraction of months with non-zero rain.

. fit Gamma distribution to non-zero values in each series separately

. calculate cumulative-probability of each non-zero monthly (annual) value computed from observations

. adjusted cumulative probability is cumulative probability multiplied by fraction of months with non-zero rain plus fraction of months with zero rain

. convert adjusted cumulative-probability from monthly (annual) values to obtain precipitation deviation from normal distribution with mean zero and standard deviation one

. Reference: Standardized Precipitation Index User Guide, WMO No. 1090, 2012.

**NCMP4: warm days and warm nights**

. calendar day 90th percentile centered on a 5-day window for *Txij*andn years *(Txin90)*

. monthly and annual percentage of days when *Txij* > *Txin90 (Tx90p)*

. calendar day 90th percentile centered on a 5-day window for *Tnij*andn years *(Tnin90)*

. monthly and annual percentage of days when *Tnij* > *Tnin90 (Tn90p)*

. the routines were taken from the Pacific Climate Impacts Consortium library available at <https://www.pacificclimate.org/resources/software-library>

**NCMP5: cold days and cold nights**

. calendar day 10th percentile centered on a 5-day window for *Txij* and n years *(Txin10)*

. monthly and annual percentage of days when *Txij* < *Txin10 (Tx10p)*

. calendar day 10th percentile centered on a 5-day window for *Tnij* and n years *(Tnin10)*

. monthly and annual percentage of days when *Tnij* < *Tnin10 (Tn10p)*

. the routines were taken from the Pacific Climate Impacts Consortium library available at <https://www.pacificclimate.org/resources/software-library>

**NCMP6: extremes of temperatures and precipitation**

. monthly and annual highest and lowest *Txij* *(TXx* and *TXn)*

. monthly and annual highest and lowest *Tnij (TNx* and *TNn)*

. monthly and annual highest *Prij (RX1day)*

. date and value of highest and lowest *TXx* for each month and entire period

. date and value of highest and lowest *TNn* for each month and entire period

. date and value of highest *RX1day* for each month and entire period

**Output data**

. csv file, one file by station, one for each NCMPs

. columns are Year, Jan, Feb … Dec, Ann

. example: filename is Toronto\_TMA.csv

 1950 -10.2 -5.6 0.3 … 8.4

. missing value is -99.9

**2. Instructions for programming the calculation of country-wide averages**

NCMPs 1 through 5 are defined as country-wide averages of various indices. This section describes the calculation of the country-wide averages from the previously calculated station indices.

Let *N* be the number of stations in the country for which we have station indices in year *y* and month *m* where 1≤*m*≤12 where *m*=1 corresponds to January, *m*=2 corresponds to February etc.

The value of the index for NCMP *k*, at station *i,* in year *y* and month *m* is *Ikiym*.

• For NCMP1, I1 is the mean temperature anomaly

• For NCMP2, I2 is the precipitation anomaly normalized in percent

• For NCMP3, I3 is the Standardised Precipitation Index

• For NCMP4, I4 is the percentage of warm days (or warm nights)

• For NCMP5, I5 is the percentage of cold days (or cold nights).

The difference in the index between two stations *i* and *j* is *Δijym* and the separation between the stations is *Dij*. The separation is assumed to be constant in time. The climatology period is 1981 to 2010.

**Input data**

The input data are those output by the station index calculation previously described:

• csv file, one file per station, one file for each NCMPs

• columns are Year, Jan, Feb … Dec, Ann

• example: filename Toronto\_TMA.csv

• example format: 1950 -10.2 -5.6 0.3 … 8.4

• missing value is -99.9

**Calculating the variogram**

The first step is to calculate the variogram, which relates the separation between two stations to the expected difference. One variogram is calculated for each NCMP and for each calendar month *m*. The variograms only need to be calculated once for each NCMP. They can be saved and reused in all future calculations of the NCMP. However, if an important change is made to the available station data due to improved quality control, homogenization or a large change in the number of stations, then the variograms should be recalculated.

The variogram for month *m* is calculated as follows. For every pair of stations *i* and *j* where *j>i* and for every year *y* in the climatology period, calculate

*Δijym=Ikiym – Ikjym.*

Choose a maximum separation, *Dmax*, which is the smaller of the following two distance: maximum separation between stations (max(Dij)) and 2000 km for precipitation indices and 3000 km for temperature indices.

Separate the *Δijym* into bins of width *w* based on their separations *Dij*. Typically, the bin width is set to 20 km. Bin *l* contains the *Δijym* where

*lw ≤ Dij < (l+1)w* where *(l+1)w < Dmax.*

In each bin *l*, calculate the bin average *Bl* by taking the arithmetic mean of the *(Δijym)2* in the bin:

*Bl = (Σ(Δijym)2)/nl*

where *nl* is the number of *Δijym* in bin *l*. The bin separation *Dl*is

*Dl = lw - w/2*.

Plot *Bl* versus *Dl* for all *l*. The plot shows how the difference in the index varies as a function of separation between two stations. The aim is next to find a function which approximates this relationship. This function is known as the functional variogram *Vm(D,n,r,s).*

The functional variogram is found by finding the function *V* and parameters *n*, *r* and *s* that minimize the mean squared error *E*:

*E = Σl [Bl – Vm(Dl, n, r, s)]2*

A list of typical functions that are used for *V* is provided in Appendix A. It is possible, but not generally advisable, to perform this fit by eye. It is always advisable to check that the fit is reasonable. If there are very few stations, it might work better to minimize the mean absolute error MAE instead:

*MAE = Σl |Bl – Vm(Dl, n, r, s)*.

This process should be repeated to find a functional variogram *Vm* and parameters *n*, *r* and *s* for each of the twelve calendar months and for each NCMP.

**Interpolation onto a regular grid**

The next step is to estimate the value of the index for all points on a regular grid. This will be done using ordinary kriging which is a standard method in geostatistics.

Define a regular grid across the country. The grid should have regular spacing in latitude and longitude such that there are at least 100 grid boxes within the country’s borders. A particular grid box will be referred to using the letter *o* for a total of *M* grid boxes. The area of grid box *o* which falls within the country’s borders is *Ao.*

The interpolated value in grid box *o* for year *y* and month *m* proceeds as follows.

Create a data vector *d* which contains the station indices:

*di = Ikiym* for *1 ≤ i ≤ N*  (where *N* is the total number of stations)

*dN+1 = 1.*

Next create a matrix *C* with *N+1* by *N+1* elements:

*Cij = Vm(Dij, n, r, s) for 1 ≤ i ≤ N*

*Ci, N+1 = 1 for 1 ≤ i ≤ N*

*CN+1, j = 1 for 1 ≤ j ≤ N*

*CN+1, N+1 = 0*

and a matrix *F* with *N+1* by *1* elements:

*Fio = Vm(Dio, n, r, s) for 1 ≤ i ≤ N*

*FN+1,o = 1*

Where *Dio* is the separation between station *i* and the centre of grid box *o*. The interpolated value of the index for grid box *o* is then given by:

*Ikoym = dTC-1F.*

This process is repeated for each grid box.

**Calculation of the country average**

The countrywide average for month *m* and year *y* is calculated by taking the area-weighted average of all grid boxes within the country’s borders. The countrywide average is the NCMP for month *m* and year *y*. That is:

*NCMPkym = (Σo=1,M AoIkoym) / (Σo=1,M Ao).*

**Output data**

There are four outputs for NCMP 1 to 5:

1. NCMP*k*\_variogram.csv:
	* one file by NCMPs
	* columns are month, function, *n* ,*r*, *s*, mean squared error
	* example: January, spherical, 32.21, 3327.68,196.81, 2666.81
2. NCMP*k*\_Graphs.pdf:
	* variogram for each month and the year
3. NCMP*k*\_Region\_Avg.csv:
	* monthly average for the region
	* columns are year, month, index, number of stations
	* example: 1950, 1, 7.0, 51
4. N*k*\_*y*\_*m*.csv:
	* value for each grid box participating in the average for year *y* and month *m*
	* columns are grid number, latitude, longitude, area, index
	* example: 9, 58, -136, 26203, 2.95.

NCMP 6 consists of counts of stations that have broken their daily temperature and precipitation records in a particular period.

**Input data**

The input data are those output by the station index calculation previously described:

• csv file, one file per station, one file per NCMP

• columns are Year, Jan, Feb … Dec, Ann

• example filename: Toronto\_TMA.csv

• example format: 1950 -10.2 -5.6 0.3 … 8.4

• missing value is -99.9

**Counts of record stations**

For each variable *k*, which includes TXx, TXn, TNx, TNn, Pr, set the count *Ck* to zero and eligible station count *Ek* to zero.

For each station *i* and variable *k*, determine the length of record *Lik* for that station.

If the length of the record is greater than 30 years the add one to *Ek*.

If the length of the record is greater than 30 years and the value of the variable *V*, for the year *y* and month *m* Vikym is the highest (Txx, Tnx, Precip) in the record i.e.

*Vikym >= max(Vikm)*

or lowest in the record (TXn, TNn) i.e.:

*Vikym <= min(Vikm)*

then add one to *Ck*.

**Output data**

The outputs for NCMP6 are:

• csv file

• columns are Year, Month, Ck, Ek for each k in (TXx, TXn, TNx, TNn, Pr)

• example: filename is Canada\_NCMP6.txt

• example output: 1950 3 3 120 0 120 0 120 5 120 10 270

• missing value is -99.9.

**Appendix A: example functions for variograms**

In each case, *n*, *r* and *s* are the parameters of the function: *n* corresponds to the variance at zero separation, *r* is a range parameter that controls how quickly or slowly the variance changes with separation, and *s* corresponds to the variance at large separations.

**Exponential**

*V(D,n,r,s) = (s-n)(1 – exp(-D/r)) + n*

**Spherical**

*V(D,n,r,s) = (s-n)( 3D/2r – D3/2r3) + n, for D<r*

*V(D,n,r,s) = s, for D>r*

**Gaussian**

*V(D,n,r,s) = (s-n)(1 – exp(-D2/r2a)) + n*