**R-NCMPs**

**User Manual**

**Developed by the**

**Expert Team on National Climate Monitoring Products**

**Version 1.0**

**June 14, 2017**

**Table of Contents**

1. Introduction ……………………………………………………………………………………………………………… 3

2. Installation and running R …………………………………………………………………………………………. 4

2.1 Download R …………………………………………………………………………………………………… 4

2.2 Install packages …………………………………………………………………………………………….. 5

2.3 Working directory …………………………………………………………………………………………. 6

2.4 Running a script ……………………………………………………………………………………………. 6

2.5 Quitting R ……………………………………………………………………………………………………... 7

3. Files to start …………………………………………..……………………………………………………………..….. 7

4. Input data …………………………………………………………………..……………………………………………. 8

5. Quality Control …………………………………………………………………………………………………………. 8

6. Indices at station level ……………………………………………………………………………………………… 9

7. Variogram ……………………………………………………….………………………………………………………. 10

8. Region-average NCMPs ……………………………………………………………………………………………. 10

9. Trends and graphs ……………………………………………………………………………………………………. 11

10. Counts of records ...…………………………………………………………………………………….………..… 11

11. Error messages ………………………………………………………………………………………………………. 12

12. Bug Report ………………………………………………………………………………………………….………… 12

13. References ……………………………………………………………………………………………………………. 12

Appendix A: List of 21 indices and 6 NCMPs.………………………………………….……………………… 13

Appendix B: Definition of the indices ……………………………………………………………………………. 14

Appendix C: List of routines with input/output directories ………………………………………….… 17

Appendix D: Fitting the variogram …………………………………………………………………………….…… 18

1. **Introduction**

R-NCMPs is a user-friendly software package for the calculation of 6 products (hereafter called National Climate Monitoring Products or NCMPs) which can be used for monitoring the temperature and precipitation conditions of a region. For the production of NCMPs in a consistent manner to allow comparison between regions, the Expert Team on National Climate Monitoring Products (ET-NCMP) has defined the 6 following products which can be calculated on a monthly basis:

* mean temperature anomaly averaged across the country
* percentage rainfall and rainfall anomaly averaged across the country
* standardized precipitation index (SPI) averaged across the country
* warm days averaged across the country
* cold nights averaged across the country
* counts of temperature and precipitation extremes.

The NCMPs are computed from 21 indices calculated for each station. A complete list of the 21 indices is provided in Appendix A along with their description in Appendix B.

The input datasets contains the daily temperature and precipitation for a number of observing stations. R-NCMPs includes a simple quality control procedure and does not homogenize climate data. A recommended climate data homogenization software package is RHtestsV4 which is available at

<http://etccdi.pacificclimate.org/software.shtml>. Six computer routines have been developed in R to produce the indices and NCMPs: the name of the routines along with the input and output directories are provided in Appendix C. This user’s manual provides step-by-step instructions on the installation of R, preparation of input data, quality control of the climate data, computation of indices and NCMPs.

**Acknowledgment**

The R-NCMPs program and user manual were originally developed by Lucie Vincent and Megan Hartwell from the Climate Research Division of Environment and Climate Change Canada. The ET-NCMP would like to thank Simon Grainger and James Adams for their important contributions to the development and testing of the computer routines.

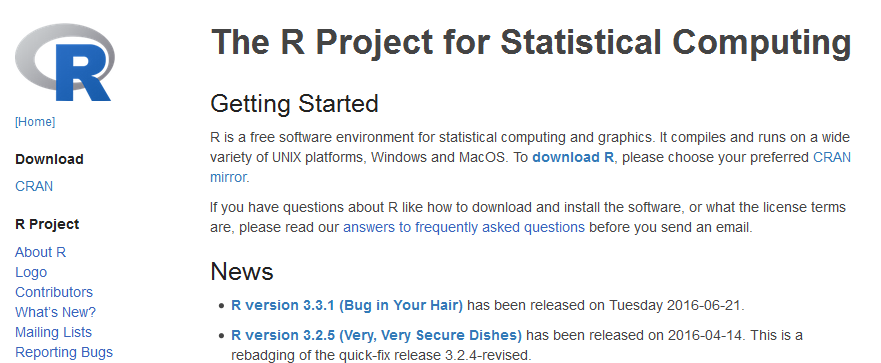
**Expert Team on National Climate Monitoring Products (ET-NCMP)**

As of June 2017, the members of the ET-NCMP are John Kennedy (lead, UK), Lucie Vincent (co-lead, Canada), Jessica Blunden (USA), Karl Braganza (Australia), Ladislaus Chang’a (Tanzania), Fatima Driouech (Morocco), Peer Hechler (Germany), Kenji Kamiguchi and Akihiko Shimpo (Japan), and Andreas Ramos (Brazil).

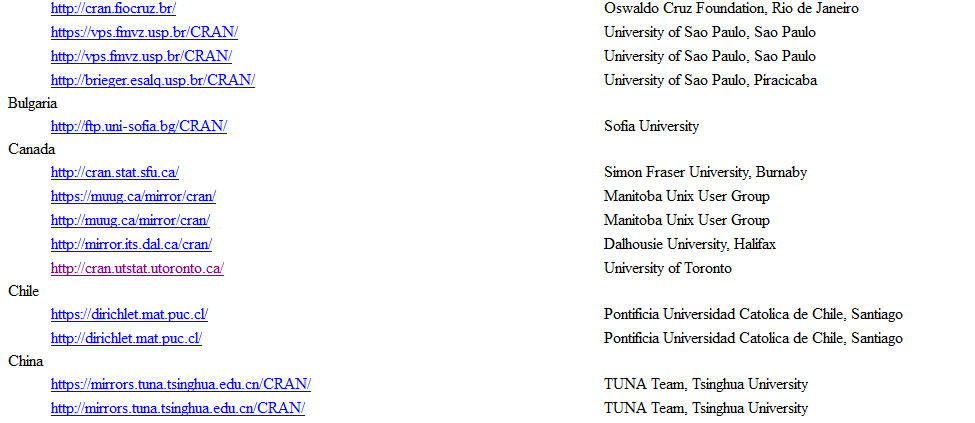
**2. Installation and running R**

***2.1 Download R***

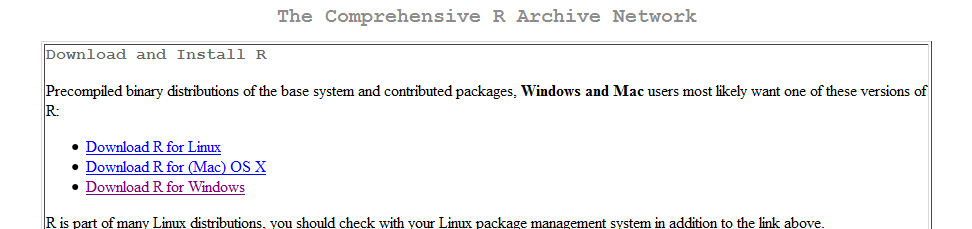
a. Go to <https://www.r-project.org/> and select “Download R” in the red box below:



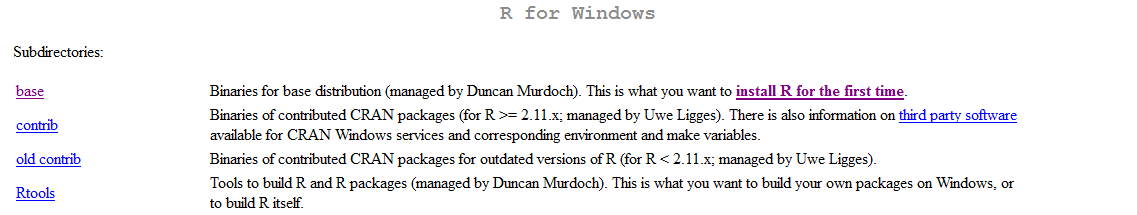
b. Select a location nearest to where you are:



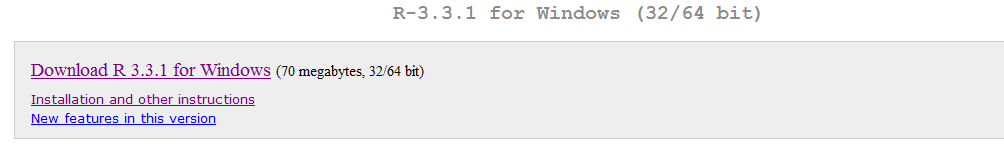
c. Select your operating system to download the appropriate installer:



d. Select “base”:



e. Select the Download R link, listed with the version number:



This begins downloading R, R-3.3.1-win.exe

Select this file to run (or save and then run it), and an installer will open. Select your language for installation, and select next in the set-up wizard. Continue selecting next until wizard finishes. You have now installed R!



***2.2 Install Packages***

a. To begin, open R, there will be a link on your desktop after installing:



b. Install packages required for the program you wish to run. For the NCMPs, enter:

**install.packages(c("data.table","stats","zoo","reshape2"))**

**install.packages(c("PCICt","climdex.pcic"))**

**install.packages(c("maps","mapdata","zyp","maptools"))**

You must enter “yes” for a personal library” and select your location to download these packages. R will install all the packages on its own and tell you when it’s done (and where to find them). This can take a few minutes.

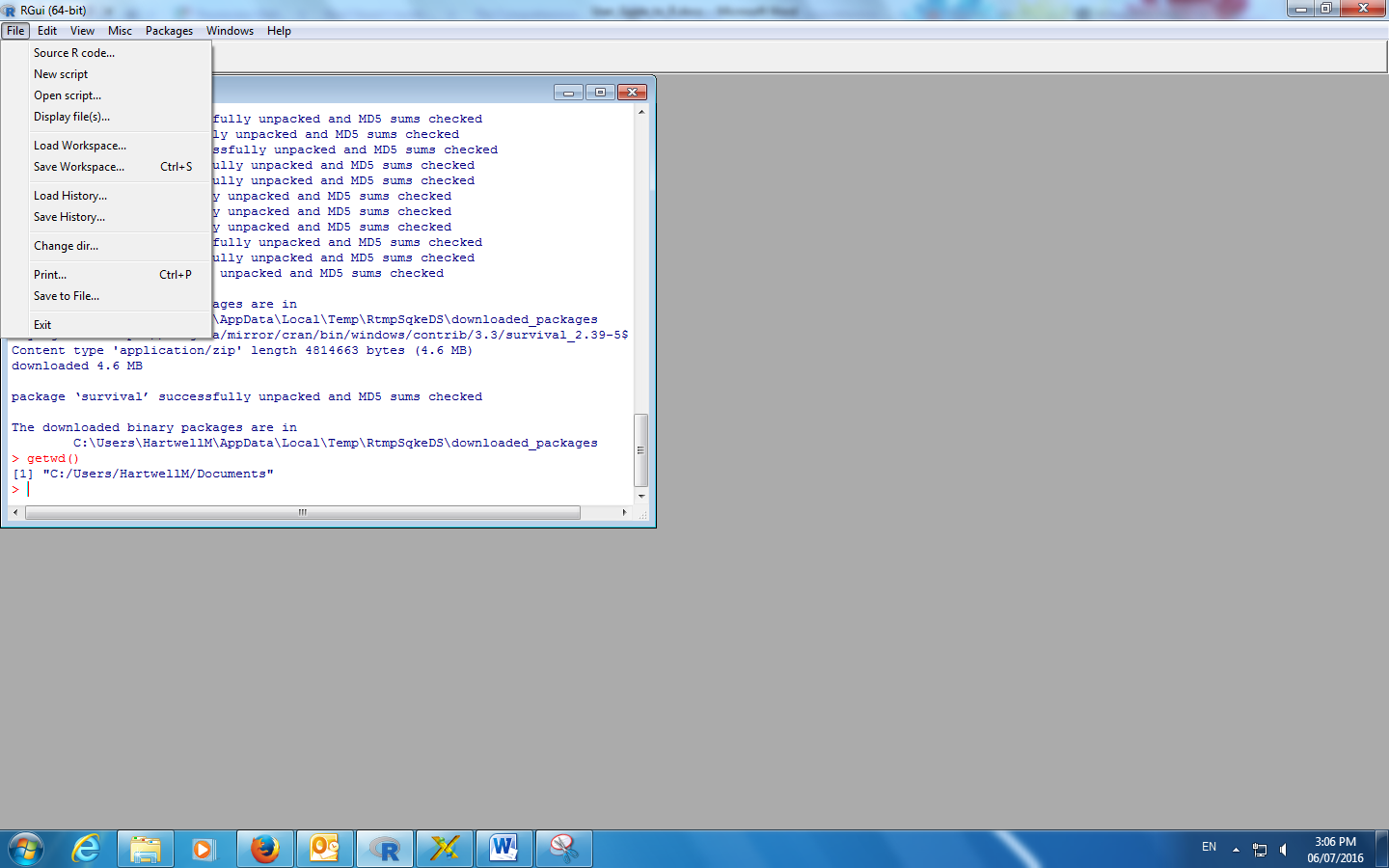
***2.3 Working Directory***

In R you are working within a specific directory, called your “working directory”. If you are accessing or printing any files they must be in the same directory you are in.

To determine where you are working, enter:

**getwd()**

If you wish to work in a different directory, there are two options you can use. The first option is to use your mouse to select File, and then select “Change dir…”



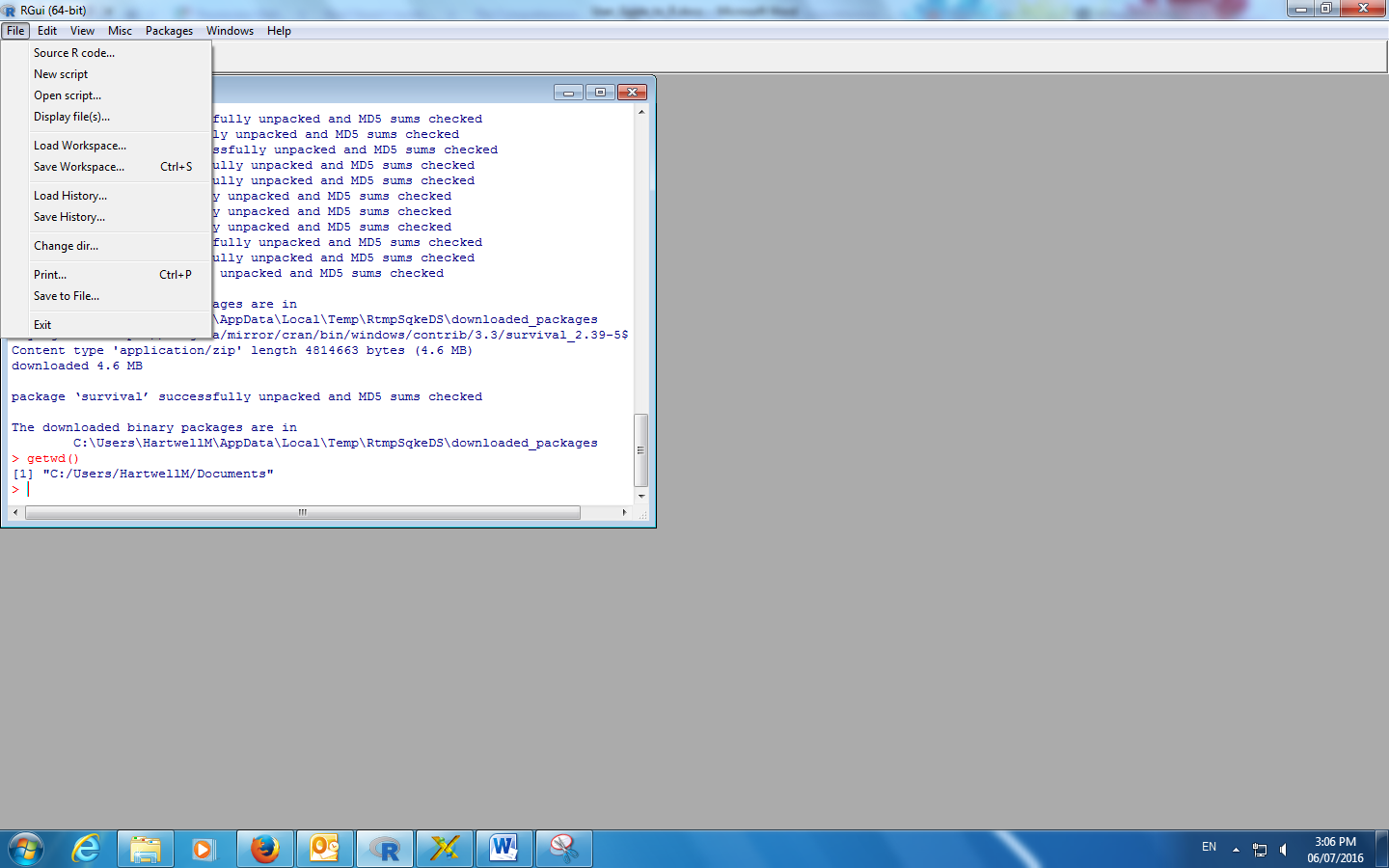
The second option is to type in the path of the directory you wish to access. You can type the path as continued from your current working directory or by typing it out from the C drive. The command is setwd(dir) where dir is a character string (so it must be in quotations) of the path of your target directory. Directories are separated by “/” or “\\”. To set a new directory, enter:

**setwd("C:/Users/NCMPs/")**

***2.4 Running a script***

Before you begin, ensure you have the proper input files ready for the script to read. You should also check to see if you have anything else in that directory under the name of any of the output files, as that information ***will be overwritten*** when the script runs.

If you wish to run a script, there are two options you can use. The first option is to use your mouse to select File, and then select “Source R code…”



The second option is to enter directly ‘source(“ScriptName.R”)’. For example:

**source("P1\_Quality\_Control\_Aug2016.R")**

The scripts will print updates on their progress as they run. You will have to scroll down to see the most recent line printed. The script will tell you when it is done running, or you can look for the “>” pointer and flashing cursor on the next line, where you will be able to type again. While the script is running the cursor may appear as a circling “loading” symbol while hovering over the R window.

***2.5 Quitting R***

To quit R, enter:

**q()**

R will ask if you want to save your workspace, answer accordingly, and then R will close. You may resume any saved workspace upon opening R.

**3. Files to start**

It is recommended that you download the Programmer’s File Editor available at <http://en.freedownloadmanager.org/Windows-PC/Programmer-s-File-Editor-FREE.html>.

You will also need to download the following 6 computer routines available at: <https://github.com/ET-NCMP/NCMP>

P1\_Quality\_Control.R

P2\_Indices.R

P3\_Variogram.R

P4\_Region\_Average.R

P5\_Trends\_Graphs.R

P6\_Counts\_Records.R

You will need to create an input file called “P0\_Station\_List.txt” and a directory called “A0\_Input\_Data” as described in section 4: this file and directory need to be located in the same directory as the 6 computer routines. Output directories will be created (see Appendix C).

**4. Input data**

You will need an input file called “P0\_Station\_List.txt” containing the filenames of each station with corresponding latitude and longitude, as for example:

Toronto\_\_\_\_\_\_\_\_\_\_\_\_\_\_ON.txt 45.02 -74.75

Fort\_Frances\_\_\_\_\_\_\_\_\_ON.txt 48.65 -93.43

Haliburton\_\_\_\_\_\_\_\_\_\_\_ON.txt 45.03 -78.53

Harrow\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ON.txt 42.03 -82.90

Kapuskasing\_\_\_\_\_\_\_\_\_\_ON.txt 49.42 -82.47

In addition, each station has a data file that lists the date (year, month, and day), daily precipitation, maximum and minimum temperatures in calendar order. Missing values are recorded as -99.9 (blank and N/A are not accepted). Each calendar day needs a record and each record must follow the calendar date order. It is important that all input files are ASCII text file and are available in the directory “A0\_Input\_data” (Appendix C). The length of the filenames must be 27 characters. For example, the input file “Toronto\_\_\_\_\_\_\_\_\_\_\_\_\_\_ON.txt” contains the following:

2010 4 30 0.0 21.0 -99.9

2010 5 1 15.4 20.5 6.5

2010 5 2 -99.9 -99.9 -99.9

2010 5 3 2.0 21.5 12.5

2010 5 4 1.4 17.0 7.2

The first column is the year, the second column is the month (January is 1, February is 2 etc), the third column is the day. The fourth column is the daily precipitation total in millimeters, the fifth column is the daily maximum temperature in degrees Celsius and the sixth column is the daily minimum temperature in degrees Celsius. Columns must be separated by at least one space.

**5. Quality Control**

Make sure that you are in the appropriate directory. To run the program for quality control, in R enter:

**source("P1\_Quality\_Control.R")**

and answer the following questions:

1. Enter the number of stations (1 to 200, or 0 for all)
2. Enter value (mm) to identify daily precipitation outlier (between 100 and 500, ex. 300)
3. Enter number of standard deviation to identify daily temperature outlier (between 3 and 7, ex. 5).

“ex.” indicates the recommended value. Observations that might be incorrect (outliers) are identified when:

* tmax ≤ tmin
* abs (tmax) > tx\_mean + *n* standard deviations
* abs (tmin) > tn\_mean + *n* standard deviations
* prec < 0 and prec > *xxx* mm

where *n* and *xxx* are user defined. The mean and standard deviation are computed from the daily temperatures of the entire period.

The directory “A1\_Quality\_Control” is created. In this directory, four output files are produced, one for each type of outlier. For example, the output file “AA\_Pr\_Outliers.csv” contains information in the following format:

Station Year Mo Dy Prec MnVal MxVal Error

Z\_SillyStation\_\_\_\_\_\_\_XX.txt 1980 12 11 -0.2 0.0 300.0 Negative prec

Z\_SillyStation\_\_\_\_\_\_\_XX.txt 1980 12 14 461.5 0.0 300.0 Prec above 300 mm

Look at each observation identified by the program and apply corrections to the input data files if needed. It is recommended to keep the original data in a separate directory. It takes about a minute to quality control 100 stations.

**6. Indices at station level**

Make sure that you are in the appropriate directory. To run the program for the indices, in R enter:

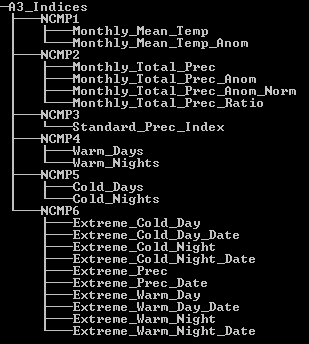
**source("P2\_Indices.R")**

and answer the following questions:

1. Enter the number of stations (1 to 200, or 0 for all)
2. Enter beginning year for base period (between 1901 and 2001, ex. 1981)
3. Enter ending year for base period (between 1910 and 2010, ex. 2010).

While the script is running the cursor may appear as a circling “loading” symbol while hovering over the R window. It can take 15 minutes to produce the indices for 100 stations.

The directory “A2\_Indices” is created. In this directory, there are 21 directories created and distributed between 6 parent directories, as shown in the tree diagram below:



In each directory, there is one output file created for each station. The list of the NCMPs is provided in Appendix A along with their description in Appendix B. If the warming message “*base thresholds longer object length is not a multiple of shorter object length*” appears, it means that some stations have a shorter reference period that the one defined by the user and the calculation will be done using the shorter period.

**7. Variogram**

The variogram is created by analyzing the distance (separation) between pairs of stations and the difference in the monthly index (Appendix D). Since it is computing intensive, the program can produce the variogram for one NCMP at the time or all (as defined by the user). It can take 30 minutes to run the routine for one NCMP using 100 stations.

Make sure that you are in the appropriate directory. To run the program for the variogram, in R enter:

**source("P3\_Variogram.R")**

and answer the following questions:

1. Enter beginning year to calculate variogram (between 1950 and 2015, ex. 1981)
2. Enter ending year to calculate variogram (between 1982 and 2016, ex. 1982)
3. Enter the desired NCMP number (between 1 and 7, or 0 for all).

The directory “A3\_Variogram” is created. In this directory, two files are produced for each NCMP. For example, the file “NCMP\_TMA\_Variogram.csv” contains the function name, the estimated parameter for *n*, *r* and *s*, and the mean squared error for each month, and the file “NCMP\_TMA\_Graphs.pdf” contains the graph (variogram) for each month and the year.

It is a good idea to look at the graphs for each variogram to ensure that the red line is a good match to the blue points.

**8. Region-average NCMPs**

The average is computed for a country or a region. For a country, the user has to provide the 3-digits numeric code available at <http://www.nationsonline.org/oneworld/countrycodes.htm>; for example, the 3-digits code for Canada is 124. The average is then calculated using the grid points within the country boundaries.

For a region, a window is defined by the latitude and longitude of four stations which are located the furthest from the window center, and the average is calculated from the grid points which fall on land. This option accommodates the calculation of an average over several countries (for example, the Caribbean region) or a small section of a country (for example, western Canada).

Make sure that you are in the appropriate directory. To run the program for the regional average, in R enter:

**source("P4\_Region\_Average.R")**

and answer the following questions:

1. Enter 0 for a country average or 1 for regional average
2. Enter the numeric 3 digit UN code for your country
3. Enter beginning year for region average (between 1950 and 2010, ex. 1950)
4. Enter ending year for region average (between 2010 and 2020, ex. 2015).
5. Enter the desired NCMP number (between 1 and 7, or 0 for all)
6. Enter the desired grid resolution 0.1, 0.25, 0.5, 1.0 or 2.0

The country average is computed for first five NCMPs (please make sure that the indices and variogram are produced for all NCMPs). It can take 10 minutes to produce the region-average using 100 stations. It will take longer for smaller grid resolutions, but the result will typically be more accurate.

The directory “A4\_Region\_Average” is created along with the subdirectory “A1\_Grid\_Squares”. In the parent directory, a file is created for each NCMP containing the region-average. For example, the file “NCMP\_TMA\_Region\_Avg.csv” contains the year, month, region-average and number of stations participating in the average. In the directory “A1\_Grid\_Squares”, a file is produced for each year and each month with the grid points participating in the average. For example, the file “TMA\_1950\_01.cvs” contains the grid number, latitude and longitude of the center of the grid, the area (km2), and the area-average for the temperature mean anomalies for January 1950.

**9. Trends and graphs**

The trends are computed for each index, each station and the region-average for a user’s defined period of time (if fewer than 10 years are missing) using the Sen's slope method (Sen, 1968) and removing lag-1 autocorrelation procedure presented in Zhang (1999). The computer routines are provided at <https://www.pacificclimate.org/resources/software-library>.

Make sure that you are in the appropriate directory. To run the program for the trends, in R enter:

**source("P5\_Trends\_Graphs.R")**

and answer the following questions:

1. Enter number of stations (1 to 200)
2. Enter beginning year to calculate trends (between 1950 and 2000, ex. 1950)
3. Enter ending year to calculate trends (between 1950 and 2050, ex. 2015).

The directory “A5\_Trends\_Graphs” is created which contains the trends and graphs for each region and each station, the rank for each region, and the maps of the trends for each index (trends statistically significant at the 0.05 level are represented by solid triangles).

**10. Counts of records**

The counts of records are provided for five indices: TXx, TXn, TNx, TNn and Rx1day. These represent the number of stations that break their record from station’s first 30-year period.

Make sure that you are in the appropriate directory. To run the program for the trends, in R enter:

**source("P6\_Count\_Records\_Aug2016.R")**

and answer the following questions:

1. Enter number of stations (1 to 200)
2. Enter beginning year to start the counts of records (between 1950 and 2000, ex. 1950)
3. Enter ending year to end the counts of records (between 1950 and 2015, ex. 2015).

The directory “A6\_Count\_Records” is created which contains the number of stations that break their record and the total number of stations participating in the count for TXx, TXn, TNx, TNn and Rx1day.

**11. Bug Report**

Please report any bugs/errors to [john.kennedy@metoffice.gov.uk](mailto:john.kennedy@metoffice.gov.uk) or [Lucie.Vincent@canada.ca](mailto:Lucie.Vincent@canada.ca). Please include the error message and data being used for the calculation of the NCMPs.

**12. Error messages**

Error “cannot open connection: no such a file…” Check if the file exists and named correctly.

Error “cannot open connection: permission denied…” Check if the file is opened and close it.

Error “character string is not in a standard format…” Check if there is data for a nonexistent date (example June 31).

Error “dims[product1] do not match the length of the object…” Caused when there is no temperature or precipitation data for the entire period. Ignore: the computations are done for the data available.

Warning “nls (Bl Spherical (Dl,a,b,s) …” Caused when it was not possible to fit the spherical function when producing the variogram. Ignore: the computations are done for the other fitted functions.

**13. References**

Sen, P.K., 1968: Estimates of the regression coefficient based on Kendall’s tau. J. Amer. Statist. Assoc., 63, 1379-1389, doi:10.1080/01621459.1968.10480934.

Zhang, X., L.A. Vincent, W.D. Hogg and A. Niitsoo, 2000: Temperature and precipitation trends in Canada during the 20th Century. Atmos.-Ocean, 38, 395-429, doi:10.1080/07055900.2000.9649654.

World Meteorological Organization, 2012. Standardized Precipitation Index User Guide. WMO No. 1090.

**Appendix A**

List of the 21 indices used to compute the 6 NCMPs. Monthly and annual values are computed for each index. There are 10 indices which are the same as those produced by RClimdex (denoted by “\*”). The “no” indicates no units for the corresponding index. The SPI values are computed following the procedure described in the Standardized Precipitation Index User Guide (WMO No. 1090, 2012). The indices TX90p, TN90p, TX10p and TN10p are computed using routines developed by the Pacific Climate Impacts Consortium available at <https://www.pacificclimate.org/resources/software-library>.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | NCMP | ID | Names | Units |
| 1 |  | TM | Monthly mean temperature | °C |
| 2 | 1 | TMA | Monthly mean temperature anomaly | °C |
| 3 |  | Pr\* | Monthly total precipitation | mm |
| 4 |  | PrA | Monthly total precipitation anomaly | mm |
| 5 | 2 | PrAn | Monthly total precipitation anomaly normalized | no |
| 6 |  | PrR | Monthly total precipitation ratio | no |
| 7 | 3 | SPI | Standardized Precipitation Index | no |
| 8 | 4a | TX90p\* | Warm Days | % of days |
| 9 | 4b | TN90p\* | Warm Nights | % of days |
| 10 | 5a | TX10p\* | Cold Days | % of days |
| 11 | 5b | TN10p\* | Cold Nights | % of days |
| 12 |  | TXx\* | Extreme Warm Day | °C |
| 13 |  | TNx\* | Extreme Warm Night | °C |
| 14 |  | TXn\* | Extreme Cold Day | °C |
| 15 |  | TNn\* | Extreme Cold Night | °C |
| 16 |  | RXday1\* | Extreme Precipitation | mm |
| 17 |  | TXx\_date | Extreme Warm Day date | date |
| 18 |  | TNx\_date | Extreme Warm Night date | date |
| 19 |  | TXn\_date | Extreme Cold Day date | date |
| 20 |  | TNn\_date | Extreme Cold Night date | date |
| 21 |  | RXday1\_date | Extreme Precipitation date | date |
| xx | 6 | CountsRec | Counts of records | # of stations |

**Appendix B**

Definitions of the 21 indices.

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 reference period. The percentiles are calculated for each day using a 5 day running window for each calendar day of the reference period. NCMPs are calculated for each month and for the year.

1. **TM**

Let *Tmikj* be the daily mean temperature for day i, month k, period j. The monthly mean temperature is:

1. **TMA**

Let *TMkj* be the monthly mean temperature for month k, period j and let *Ctk* be the temperature climatology for month k. The monthly mean temperature anomaly is:

1. **Pr**

Let *Prikj* be the daily precipitation for day i, month k, period j. The monthly total precipitation is:

1. **PrA**

Let *Prkj* be the monthly total precipitation for month k, period j and let *Cpk* be the precipitation climatology for month k. The monthly total precipitation anomaly normalized is:

1. **PrAn**

Let *Prkj* be the monthly total precipitation for month k, period j and let *Cpk* be the precipitation climatology for month k. The monthly total precipitation anomaly is:

1. **PrR**

Let *Prkj* be the monthly total precipitation for month k, period j and let *Cpk* be the precipitation climatology for month k. The monthly total precipitation ratio is:

1. **SPI**

Let *Prkj* be the total monthly precipitation for month k, period j. We calculate two parameters which are used to fit the gamma distribution, and we find the probability. The probability is the converted to SPI using the quantile function of the standard normal distribution with mean 0 and SD 1.

1. **TX90p**

Let *Txikj* be the daily max temperature for day i, month k, period j and *Tx90ij* be the value of the 90th percentile of maximum temperature for day i, month k. Then the warm days is:

1. **TN90p**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j and *Tn90ij* be the value of the 90th percentile of minimum temperature for day i, month k. Then the warm nights is:

1. **TX10p**

Let *Txikj* be the daily max temperature for day i, month k, period j and *Tx10ij* be the value of the 10th percentile of maximum temperature for day i, month k. Then the cold days is:

1. **TN10p**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j and *Tn10ij* be the value of the 10th percentile of minimum temperature for day i, month k. Then the warm nights is:

1. **TXx**

Let *Txikj* be the daily max temperature for day i, month k, period j. Then the extreme warm day is:

1. **TNx**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j. Then the extreme warm night is:

1. **TXn**

Let *Txikj* be the daily max temperature for day i, month k, period j. Then the extreme cold day is:

1. **TNn**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j. Then the extreme warm night is:

1. **RXday1**

Let *Prikj* be the daily precipitation for day i, month k, period j. Then the extreme precipitation is:

1. **TXx date**

Let *Txikj* be the daily max temperature for day i, month k, period j and *TXxkj­* be the extreme warm day for month k, period j. The extreme warm day date is:

1. **TNx date**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j and *TNxkj­* be the extreme warm night for month k, period j. The extreme warm night date is:

1. **TXn date**

Let *Txikj* be the daily max temperature for day i, month k, period j and *TXnkj­* be the extreme cold day for month k, period j. The extreme cold day date is:

1. **TNn date**

Let *Tnikj* be the daily minimum temperature for day i, month k, period j and *TNnkj­* be the extreme cold night for month k, period j. The extreme cold night date is:

1. **RXday1 date**

Let *Prikj* be the daily precipitation for day i, month k, period j and *RXday1kj* be the extreme precipitation for month k, period j. Then the extreme precipitation date is:

**Appendix C**

List of the 6 computer programs needed to produce the NCMPs along with the input and output directories.

|  |  |  |
| --- | --- | --- |
| **Programs** | **Input** | **Output** |
| P1\_Quality\_Control.R | A0\_Input\_Data | A1\_Quality\_Control |
| P2\_Indices.R | A0\_Input\_Data | A2\_Indices |
| P3\_Variogram.R | A2\_Indices | A3\_Variogram |
| P4\_Region\_Average.R | A2\_Indices A3\_Variogram | A4\_Region\_Average |
| P5\_Trends\_Graphs.R | A2\_Indices A4\_Region\_Average | A5\_Trends |
| P6\_Counts\_Records.R | A2\_Indices | A6\_Counts\_Records |

**Appendix D**

Process of fitting the functional variogram

The variogram is created by looking at each pair of stations and analyzing the distance (separation) between them and the difference in the monthly index. These points are divided into bins of 20 km. The maximum distance for temperature is set to 3000 km and in this case there are 150 bins; it is set to 2000 km for precipitation and there are 100 bins. The average of each bin is plotted. The functional variogram is then fit to minimize the mean squared error.

The functional variogram is one of three functions, all using parameters *n, r*, and *s* where *n* corresponds to the variance at zero separation, *r* is a range parameter that controls how quickly or slowly the variance changes with separation, *s* corresponds to the variance at large separations.

The *s* parameter is calculated by the program to be the average index value of the 20% largest bin separations. The variogram is then fit automatically to the other two parameters, using one of the three functions below:

**Exponential**

**Spherical**

**Gaussian**

.

The user should check the variogram to make sure that it is proper to its data. Please note that different functions can be used for different months.