Introduction to MarClimEX

Xiaolan Wang, Yang Feng and Rodney Chan

Climate Research Division
Environment and Climate Change Canada
Downsview, Ontario
Canada

May 26, 2016

Maintained by Yang Feng
Please contact yang.feng@canada.ca for any question or comment.

Contents

1 Introduction

MarClimEX is designed to provide a user friendly interface to compute indices of marine climate extremes. It computes all 37 core marine climate extrme indices recommended by the CCl/CLIVAR Expert Team for Climate Change Detection Monitoring and Indices (ETCCDI) including extreme wave height indices, extreme wind indices and extreme water level indices. This version of MarClimEX has been developed under **R** 2.15.2. It also depends on the **R** library of PCICt (Version 0.5-4) for computing the 37 core indices as well as the **R** library of Tcl/Tk (Version 2.15.2) for the graphical user interface. This MarClimEX package should run with **R** 2.15.2 or a later version. The depended **R** libraries will be downloaded and installed during the installation of this MarClimEX package. This manual provides step-by-step instructions of the following.

1. The installation of **R** and setting up the user environment
2. The installation of MarClimEX
3. Using MarClimEX

2 Installation of R

2.1 Brief introduction to R

R is a language and environment for statistical computing and graphics. It is a GNU implementation of the S language developed by John Chambers and colleagues at Bell Laboratories (formerly AT&T, now Lucent Technologies). S-plus provides a commercial implementation of the S language.

2.2 Installing R

MarClimEX requires the base package of **R** (Version 2.15.2 or later). The installation of **R** involves a very simple procedure. First, connect to the **R** project website at www.r-project.org, then follow the links to download the most recent version of **R** for your computer operating system from any mirror site of CRAN.

For Microsoft Windows XP and later, download the base **R** Windows installer. Run that installer and **R** will be automatically installed on your computer, with a shortcut to **R** on your desktop. The Tcl/Tk library is included in the default installation of **R**.

For Linux, download the proper precompiled binaries and follow the instruction to install **R**. For other UNIX systems, you may need to download the source code and compile it yourself.

For Mac OS X 10.9 (Mavericks) and above, download the latest version of the **R** signed package. Validate the signature using pkgutil for example *–check-signature R-3.2.1.pkg* in the *Terminal*. Run the **R** signed package to install. Select custom install during installtion to enable Tcl/Tk library. Connect to XQuartz website at xquartz.macosforge.org to use X11 which is required for the graphical user interface of MarClimEX. Download the latest XQuartz image and install. There is now an R app in Lauchpad.

For Mac OS X 10.5 (Leopard) to 10.8 (Mountain Lion), download the last supported legacy version of the **R** signed package of the corresponding Mac OS X version. Validate the signature by checking the MD5 checksum with the website. For example to check the MD5 checksum type in *md5 R-3.2.1.pkg* in the *Terminal* application. Run the **R** signed package to install and during installation select custom install to enable Tcl/Tk library. **R** is now installed and in the Application folder. For Mac OS X 10.8 (Mountain Lion), connect to XQuartz website at xquartz.macosforge.org to use X11 which is required for the graphical user interface of MarClimEX. Download the latest XQuartz image and install.
x

2.3 Running R

For Windows, double click the **R** icon on your desktop, or launch it through Windows *Start* Menu. This usually gets you into the **R** user interface. **R** 64bit will also be installed if your system supports 64bit. It is recommended to use

R 64bit if your system is 64bit. You may quit the program by clicking on the top menu under *File* then *Exit*.

Under Linux, just run the command *R* to give you the **R** console. You may quit by typing in the command *q()*.

Under Max OS X, click on the **R** app in Lauchpad, or double click the **R** icon in your Application folder. This usually gets you into the **R** user interface. You may quit the program by clicking on the top menu under *R* then *Quit R*.

3 Installation of MarClimEX

MarClimEX is an **R** package with the most recent version avaliable from the ETCCDI website at http://etccdi.pacificclimate.org/software.shtml, where registration is required. Please install MarClimEX as a local package in **R**. First launch **R** in the same directory as the MarClimEX package. Then run the following line:

```
install.packages("MarClimEX_0.1.tar.gz", repos=NULL, type="source")
```

If you had used an older version of the **R** package PCICt or an **R** package that uses PCICt like MarClimEX then you may need to update the PCICt pacakge. With an internet connection, please use the following command and select a minor to update the **R** package of PCICt.

```
install.packages("PCICt")
```

MarClimEX has been developed under **R** 2.15.2. This version of MarClimEX depends on the **R** library of PCICt (Version 0.5-4) for computing the 37 core indices as well as the **R** library of Tcl/Tk (Version 2.15.2) for the graphical user interface. The depended **R** libraries will be downloaded and installed automatically during the installation of the MarClimEX package.

For Windows, you may change the current **R** working directory by clicking on the top menu under *File* then *Change dir...*, afterwards select the directory where you store the MarClimEX package before installing.

Under Mac OS X, please click on the top menu under *Misc* then *Change Working Directory...* to change the current **R** working directory. Please select the directory where MarClimEX package is stored before installing.

4 MarClimEX

4.1 Getting started on MarClimEX

MarClimEX can be loaded like any other **R** packages. All dependent libraries will also be loaded. Please type in the following into the **R** console:

```
library("MarClimEX")
```

In order to launch MarClimEX user interface and to begin using RClimdex, simply type in the following into the **R** console.

```
marclimex.start()
```

You may type in the command into the **R** console again to relaunch the user interface. MarClimEX is not programmed to support concurrent sessions, therefore please launch one user interface at a time.

Figure 1: MarClimEX main menu.

4.2 Quality control procedures

There is a simple data quality control consists of the following procedures.

1. Replace all user-defined missing values, with default as -99.9, into NA (internal R missing value).
2. Replace all unreasonable values into NA. Those values include:
   (a) All values corresponding to an impossible date. (i.e. 32nd March 2013, 12th June 20AA, etc.)
   (b) Any non-numeric values.

4.3 Loading datasets

In the *File Open Window*, you may select the dataset for MarClimEX. You may need to change directory to where the dataset is stored. MarClimEX expects file extension of either *.csv* or *.txt* however you may display files of all types by selecting *ALL files* in the *Files of type* dropdown menu (See Figure 2). Please note that the dataset must be formatted as described in Appendix B.

Windows and Linux users may press and hold *shift* key or *crtl* key (Mac OS X users may use *cmd* key) to select multiple datasets. When multiple datasets are selected, MarClimEX would calculate indices based on the same parameters selected on the menu to all datasets in sequence. The output filenames of multiple datasets correspond with their filenames respectively and cannot be changed.

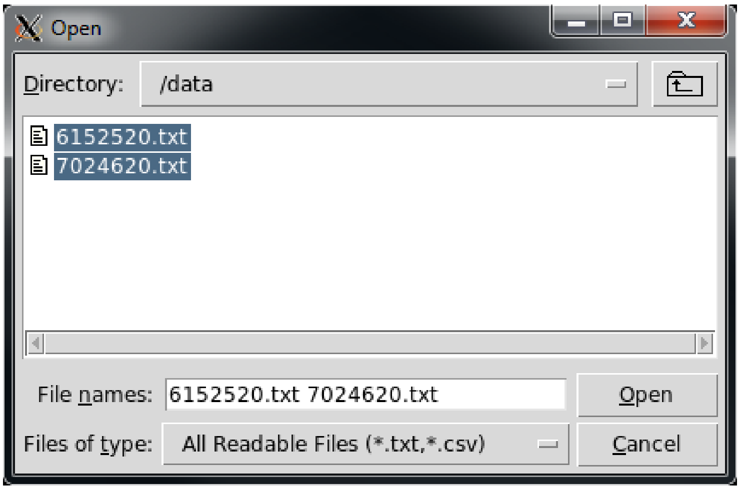

Figure 2: File selection for MarClimEX.

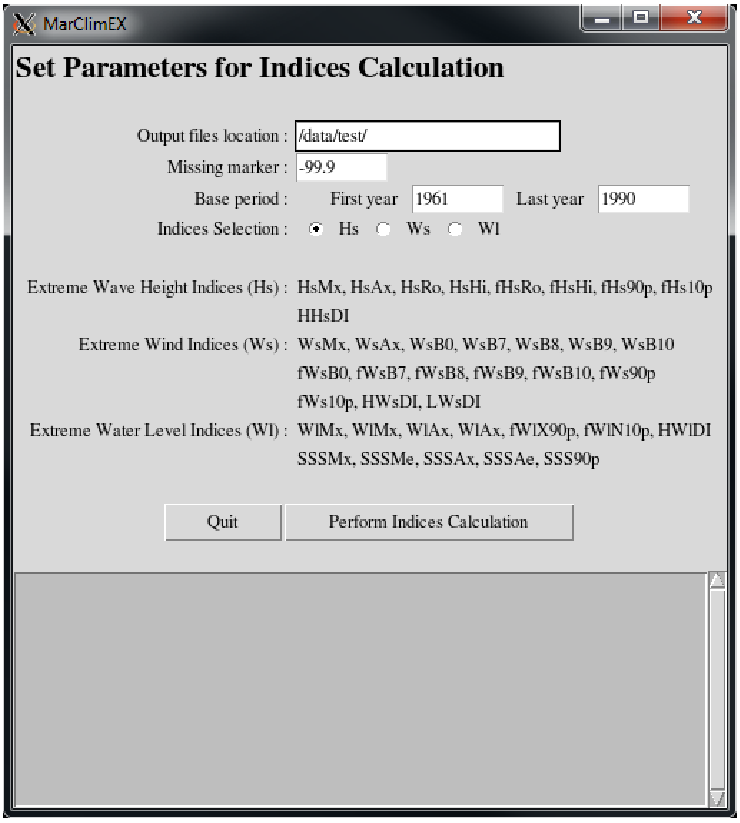

Figure 3: Parameters selection for MarClimEX.

4.4 Parameters selection and indices calculations

The following is a brief descripition to each parameter as shown in Figure 3.

1. Output file location: The local directory where MarClimEX will store the output files from the indices calculation.
2. Station name or code: Output filename prefix. (Only available when processing one dataset)
3. Missing marker: Character string as the indicator for missing values in input dataset.
4. Base period: The base period usually is a long enough period preferably more than 10 years in which climatology do not change much for a given station.
   (a) First year: Starting year of base period. (Always begin on 1st of January)
   (b) Last year: Ending year of base period. (Always end on 31st of December)
5. Indices Selection: Select the type of climate change/extreme indices
   (a) Hs: Extreme wave height indices
   (b) Ws: Extreme wind speed indices
   (c) Wl: Extreme water level indices

The output file location is where all outputs from MarClimEX will be stored. In addition, there is a log file in *txt* format with details about the MarClimEX session. There are also the *indices* subdirectory which would be where the indices calculation outputs be stored. You may only rename the output filename if you are not processing multiple dataset. The current version of MarClimEX does not support different indices calculation selection for each individual dataset when processing multiple dataset. For example, extreme wave heigh indices will be proccessed to all dataset regardless if all datasets are Hs when it is selected, therefore a mix of Hs, Ws and Wl datasets cannot be proccessed in the same batch.

Please note that the default missing marker of -99.9 will always be used in addition to your user-defined missing marker and the base period always begin from the 1st of January and ends on the 31st of December of the user defined period.

There is a brief summary of the indicators for each type of indices. More information about the indices can be found in Appendix A and full definition can be found in Appendix C. Press the *Perform Indices Calculation* button to begin and press the *Quit* button to exit to main menu. Also note that indices calcualtion will not automatically quit after completion.

MarClimEX communicates with the user in the log message box below the buttons. This is where details about the current status, such as which dataset is loaded, what process is being performed, number of datasets remaining, etc. are displayed.

A List of ETCCDI core marine climate change/extreme indices

A.1 Extreme wave height indices

ID	Indicator name	Definitions	Units
HsMx	Monthly max significant wave height	Monthly maximum value of significant wave height (Hs)	m
HsAx	Annual max significant wave height	Annual maximum value of significant wave height (Hs)	m
HsRo	Rough wave days	Annual count of days when daily max Hs > 2.5 m	days
HsHi	High wave days	Annual count of days when daily max Hs > 6 m	days
fHsRo	Frequency of rough wave days	Annual percentage of days when daily max Hs > 2.5 m	%
fHsHi	Frequency of high wave days	Annual percentage of days when daily max Hs > 6 m	%
fHs90p	Frequency of top decile wave days	Annual percentage of days when daily max Hs $>$ 90th percentile	%
fHs10p	Frequency of low decile wave days	Annual percentage of days when daily max Hs $<$ 10th percentile	%
HHsDI	Top decile wave spell duration indicator	Annual count of days with at least 2? consecutive days when daily max Hs $>$ 90th percentile	days

A.2 Extreme wind indices

ID	Indicator name	Definitions	Units
WsMx	Monthly max wind speed	Monthly maximum value of wind speed (Ws)	m/s
WsAx	Annual max wind speed	Annual maximum value of wind speed (Ws)	m/s

ID	Indicator name	Definitions	Units
WsB0	Calm wind days	Annual count of days when daily max Ws < 0.514 m/s (Beaufort Scale 0)	days
WsB7	Near gale-force wind days	Annual count of days when daily max Ws > 14.403 m/s (Beaufort Scale 7)	days
WsB8	Gale-force wind days	Annual count of days when daily max Ws > 17.222 m/s (Beaufort Scale 8)	days
WsB9	Strong gale-force wind days	Annual count of days when daily max Ws > 20.833 m/s (Beaufort Scale 9)	days
WsB10	Storm-force wind days	Annual count of days when daily max Ws > 24.722 m/s (Beaufort Scale 10)	days
fWsB0	Frequency of calm wind days	Annual percentage of days when daily max Ws < 0.514 m/s (Beaufort Scale 0)	%
fWsB7	Frequency of near gale-force wind days	Annual percentage of days when daily max Ws > 14.403 m/s (Beaufort Scale 7)	%
fWsB8	Frequency of gale-force wind days	Annual percentage of days when daily max Ws > 17.222 m/s (Beaufort Scale 8)	%
fWsB9	Frequency of strong gale-force wind days	Annual percentage of days when daily max Ws > 20.833 m/s (Beaufort Scale 9)	%
fWsB10	Frequency of storm-force wind days	Annual percentage of days when daily max Ws > 24.722 m/s (Beaufort Scale 10)	%
fWs90p	Frequency of top decile wind days	Annual percentage of days when daily max Ws > 90th percentile	%
fWs90p	Frequency of low decile wind days	Annual percentage of days when daily max Ws < 10th percentile	%
HWsDI	Top decile wind spell duration indicator	Annual count of days with at least 2? consecutive days when daily max Ws > 90th percentile	days
LWsDI	Low decile wind spell duration indicator	Annual count of days with at least 2? consecutive days when daily max Ws < 10th percentile	days

A.3 Extreme water level indices

ID	Indicator name	Definitions	Units
WlMx	Monthly max water level	Monthly maximum value of water level (Wl)	m
WlMn	Monthly min water level	Monthly minimum value of water level (Wl)	m
WlAx	Annual max water level	Annual maximum value of water level (Wl)	m
WlAn	Annual min water level	Annual minimum value of water level (Wl)	m
fWlX90p	Frequency of extreme high water level days	Annual percentage of days when daily max Wl > 90th percentile	%
fWlN10p	Frequency of extreme low water level days	Annual percentage of days when daily max Wl < 10th percentile	%
HWlDI	High water level spell duration indicator	Annual count of days with at least 2? consecutive days when daily max Wl > 90th percentile	days
SSSMx	Monthly max skew storm surge	Monthly maximum skew storm surge	m
SSSMe	Monthly mean skew storm surge	Monthly mean skew storm surge	m
SSSAx	Annual max skew storm surge	Annual maximum skew storm surge	m
SSSAe	Annual mean skew storm surge	Annual mean skew storm surge	m
SSS90p	All-time 90th percentile skew storm surge	???	m

B Data format

All indices calculation are outputted as comma-separated values (CSV) files. MarClimEX likewise accepts CSV files as input data. MarClimEX also accepts space-delimited ASCII text file as input data. Space-delimited format has each element separated by one or more spaces. The input dataset must satisfy the following requirements.

1. The input dataset **must** have the file extension *.csv* or *.txt*
2. Columns **must** be *YEAR MONTH DAY VAR* in that order where *VAR* has to be either *Hs*, *Ws* or *Wl*
3. The records **must** be in calendar date order. Missing dates are allowed.

This is an example of an input dataset as a space-delimited ASCII file.

```
1950  2  3  6.32
1950  2  4  4.79
1950  2  5  8.86
1950  2  6  15.67
1950  2  9  10.41
```

This is an example of an input dataset as a CSV file.

```
1950,2,3,6.32
1950,2,4,4.79
1950,2,5,8.86
1950,2,6,15.67
1950,2,9,10.41
```

C Indices definition

Definitions for indicators are listed in Appendix A. For practical reasons, in this version of the software, not all indices are calculated on a monthly basis. Monthly indices are calculated if no more than 3 days are missing in a month, while annual values are calculated if no more than 15 days are missing in a year. No annual values will be calculated if any one month's data are missing. For threshold indices, a threshold is calculated if at least 75% of data are present. For spell duration indicators (marked with a *), a spell can continue into the next year and is counted against the year in which the spell ends.

Top decile wave spell (HHsDI) for example beginning on 1st December 2010 and ending on 6th January 2011 is counted towards the total number of top decile wave spells in 2011.

1. **HsMx**

   Let Hs_{ij} be the wave height in month i, period j. The maximum wave height of each month is then:

   $$HS_{ij} = max(Hs_{ij})$$

2. **HsAx**

   Let Hs_{ij} be the wave height in year i, period j. The maximum wave height of each year is then:

   $$HS_{ij} = max(Hs_{ij})$$

3. **HsRo**

   Let Hs_{ij} be the wave height on day i in period j. Count the number of days where:

   $$Hs_{ij} > 2.5 \text{ m}$$

4. **HsHi**

   Let Hs_{ij} be the wave height on day i in period j. Count the number of days where:

   $$Hs_{ij} > 6 \text{ m}$$

5. **fHsRo**

   Let Hs_{ij} be the wave height on day i in period j. The percentage of time is determined where:

   $$Hs_{ij} > 2.5 \text{ m}$$

6. **fHsHi**

   Let Hs_{ij} be the wave height on day i in period j. The percentage of time is determined where:

   $$Hs_{ij} > 6 \text{ m}$$

7. **fHs90p**

   Let Hs_{ij} be the wave height on day i in period j and let $Hs_{in}90$ be the calendar day 90th percentile. The percentage of time is determined where:

   $$Hs_{ij} > Hs_{in}90$$

8. **fHs10p**

   Let Hs_{ij} be the wave height on day i in period j and let $Hs_{in}10$ be the calendar day 10th percentile. The percentage of time is determined where:

   $$Hs_{ij} < Hs_{in}10$$

9. **HHsDI***

   Let Hs_{ij} be the wave height on day i in period j and let $Hs_{in}90$ be the calendar day 90th percentile. Then the number of days per period is summed where, in interval, of atleast 2? consecutive days:

   $$Hs_{ij} > Hs_{in}90$$

10. **WsMx**

    Let Ws_{ij} be the wind speed in month i, period j. The maximum wind speed of each month is then:

    $$WS_{ij} = max(Ws_{ij})$$

11. **WsAx**

    Let Ws_{ij} be the wind speed in year i, period j. The maximum wind speed of each year is then:

    $$WS_{ij} = max(Ws_{ij})$$

12. **WsB0**

    Let Ws_{ij} be the wind speed on day i in period j. Count the number of days where:

    $$Ws_{ij} < 0.514 \text{ m/s}$$

13. **WsB7**

    Let Ws_{ij} be the wind speed on day i in period j. Count the number of days where:

    $$Ws_{ij} > 14.403 \text{ m/s}$$

14. **WsB8**

    Let Ws_{ij} be the wind speed on day i in period j. Count the number of days where:

    $$Ws_{ij} > 17.222 \text{ m/s}$$

15. **WsB9**

    Let Ws_{ij} be the wind speed on day i in period j. Count the number of days where:

    $$Ws_{ij} > 20.833 \text{ m/s}$$

16. **WsB10**

    Let Ws_{ij} be the wind speed on day i in period j. Count the number of days where:

    $$Ws_{ij} > 24.722 \text{ m/s}$$

17. **fWsB0**

    Let Ws_{ij} be the wind speed on day i in period j. The percentage of time is determined where:

    $$Ws_{ij} < 0.514 \text{ m/s}$$

18. **fWsB7**

    Let Ws_{ij} be the wind speed on day i in period j. The percentage of time is determined where:

    $$Ws_{ij} > 14.403 \text{ m/s}$$

19. **fWsB8**

    Let Ws_{ij} be the wind speed on day i in period j. The percentage of time is determined where:

    $$Ws_{ij} > 17.222 \text{ m/s}$$

20. **fWsB9**

    Let Ws_{ij} be the wind speed on day i in period j. The percentage of time is determined where:

    $$Ws_{ij} > 20.833 \text{ m/s}$$

21. **fWsB10**

    Let Ws_{ij} be the wind speed on day i in period j. The percentage of time is determined where:

    $$Ws_{ij} > 24.722 \text{ m/s}$$

22. **fWs90p**

    Let Ws_{ij} be the wind speed on day i in period j and let $Ws_{in}90$ be the calendar day 90th percentile. The percentage of time is determined where:

    $$Ws_{ij} > Ws_{in}90$$

23. **fWs10p**

    Let Ws_{ij} be the wind speed on day i in period j and let $Ws_{in}10$ be the calendar day 10th percentile. The percentage of time is determined where:

    $$Ws_{ij} < Ws_{in}10$$

24. **HWsDI***

    Let Ws_{ij} be the wind speed on day i in period j and let $Ws_{in}90$ be the calendar day 90th percentile. Then the number of days per period is summed where, in interval, of atleast 2? consecutive days:

    $$Ws_{ij} > Ws_{in}90$$

25. **LWsDI***

    Let Ws_{ij} be the wind speed on day i in period j and let $Ws_{in}10$ be the calendar day 10th percentile. Then the number of days per period is summed where, in interval, of atleast 2? consecutive days:

    $$Ws_{ij} < Ws_{in}10$$

26. **WlMx**

    Let Wl_{ij} be the water level in month i, period j. The maximum water level of each month is then:

    $$WL_{ij} = max(Wl_{ij})$$

27. **WlMn**

    Let Wl_{ij} be the water level in month i, period j. The minimum water level of each month is then:

    $$WL_{ij} = min(Wl_{ij})$$

28. **WlAx**

    Let Wl_{ij} be the water level in year i, period j. The maximum water level of each year is then:

    $$WL_{ij} = max(Wl_{ij})$$

29. **WlAn**

    Let Wl_{ij} be the water level in year i, period j. The minimum water level of each year is then:

    $$WL_{ij} = min(Wl_{ij})$$

30. **fWlX90p**

    Let Wl_{ij} be the water level on day i in period j and let $Wl_{in}90$ be the calendar day 90th percentile. The percentage of time is determined where:

    $$Wl_{ij} > Wl_{in}90$$

31. **fWlN10p**

    Let Wl_{ij} be the water level on day i in period j and let $Wl_{in}10$ be the calendar day 10th percentile. The percentage of time is determined where:

    $$Wl_{ij} < Wl_{in}10$$

32. **HWlDI***

    Let Wl_{ij} be the water level on day i in period j and let $Wl_{in}90$ be the calendar day 90th percentile. Then the number of days per period is summed where, in interval, of atleast 2? consecutive days:

    $$Wl_{ij} > Wl_{in}90$$

33. **SSSMx**

    Let Sss_{ij} be the skew storm surge in month i, period j, where the skew storm surge is the simple difference between observed high water level and tidal prediction in that same tidal cycle. The maximum skew storm surge of each month is then:

$$SSS_{ij} = max(Sss_{ij})$$

34. **SSSMe**

    Let Sss_{ij} be the skew storm surge in month i, period j, where the skew storm surge is the simple difference between observed high water level and tidal prediction in that same tidal cycle. The mean skew storm surge of each month is then:

    $$SSS_{ij} = mean(Sss_{ij})$$

35. **SSSAx**

    Let Sss_{ij} be the skew storm surge in year i, period j, where the skew storm surge is the simple difference between observed high water level and tidal prediction in that same tidal cycle. The maximum skew storm surge of each year is then:

    $$SSS_{ij} = max(Sss_{ij})$$

36. **SSSAe**

    Let Sss_{ij} be the skew storm surge in year i, period j, where the skew storm surge is the simple difference between observed high water level and tidal prediction in that same tidal cycle. The mean skew storm surge of each year is then:

    $$SSS_{ij} = mean(Sss_{ij})$$

37. **SSS90p**

    ???

    ???

D Empirical quantile estimation

The quantile of a distribution is defined as

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \geq p\}, 0 < p < 1$$

where $F(x)$ is the distribution function. Let $\{X_{(a)}, ..., X_{(n)}\}$ denote the order statistics of $\{X_1, ..., X_n\}$ (i.e. sorted values of $\{X\}$), and let $\hat{Q}_i(p)$ denote the ith sample quantile definition. The sample quantiles can be generally written as

$$\hat{Q}_i(p) = (1 - \gamma)X_{(j)} + \gamma X_{(j+1)}$$

Hyndman and Fan (1996) suggest a formula to obtain medium un-biased estimate of the quantile by letting j =int$(p \times n + (1 + p)/3)$ and letting $\gamma = p \times n + (1 + p)/3 - j$, where $\mathrm{int}(u)$ is the largest integer not greater than u. The empirical quantile is set to the smallest or largest value in the sample when $j < 1$ or $j > n$ respectively. That is, quantile estimates corresponding to $p < 1/(n{+}1)$ are set to the smallest value in the sample, and those corresponding to $p > n/(n + 1)$ are set to the largest value in the sample.

References

[1] Hyndman, R.J., and Y. Fan, 1996: Sample quantiles in statistical packages. *The American Statistician*, **50**, 361-367.