** **

**ICPAC QUALITY CONTROL ASSURANCE AND PROCEDURES**

It is widely known that raw digital climate data are subject to a wide range of errors, which can be introduced along the chain of collecting, digitizing, processing, transferring, storing and transmitting the data (Anonymous 2010). A wide set of procedures for quality controlling (QC) raw data have been developed, which are recommended by the World Meteorological Organization (WMO) and implemented or used by investigators to detect and flag potentially erroneous values in any dataset (e.g. Abbott 2sgma check1986).

The 2sgma check and neighboring station check has been done to the data. The Consistence check to compare the available records from different sources has also been made. ICPAC, as an RCC is obliged to support NMHS build capacity in climate data management which includes climate data quality control.

As mentioned before ICPAC has no mandate to observe climate data, rather receives digital data from 137 IGAD NMHS station data on bilateral agreements on a 10day basis. This is augmented with the rainfall estimates and other datasets from the public domain like http://climexp.knmi.nl/selectdailyseries.cgi which serves in climate monitoring. There are two levels of quality control at ICPAC and they are Primary and Secondary level of quality control.

**Primary level QC**

Observational data error detection and possible error corrections in order to ensure the highest possible reasonable standard accuracy for real-time can follow the following steps;

1. Quick inspection of the data to isolate anomalous deviations and suspicions zeros in a space of missing data.
2. Action taken in the real time situation is to contact the NMHS where data originated by phone or email such that the person in charge quickly confirms the validity of the value.
3. If the value is too suspicious beyond 4sgma is put aside for further investigation.

**Secondary level QC**

When data is received at ICPAC it is subjected to further rigorous checks to come up with quality controlled climate data series for long range monitoring, detection and attribution. Following are some of the checks employed using both ***Macros*** and ***Climsoft***:

* Repeated data check
* Consistent checks with other data existing at ICPAC
* 2sgma Check for Dekadal monthly data sets
* 10th Percentile check for the Rainfall Lower limits
* Daily rainfall checked with upper limit of 100mm but this can change according to location
* Temperature rates of change from the previous day

**The 2sgm Test:**

# The distribution of rainfall and temperature follows 2sgma away from the mean as recommended by Abbot in 1985 and can be used to set lower and upper limits checks. Both errors and extreme values sit in the tails beyond the lower and upper limit but this can only be true if the data is having a normal distribution. In the case of the skewed data like rainfall, the upper limit is used and the lower limit is set at the 10th percentile .However in the case of temperature which is normally distributed, this formulae works for the lower and upper limit

Organize the data you need to quality control whether Monthly or dekadal data from which the statistics can be generated.

Getting the upper limit to be used to quality controlling the rainfall data using the 2sgma and this is before you use the macro.





Upper limit

Lower limit

# Performing Quality Control using Climsoft

Data brought in to ICPAC from the regional NMHS is already digitized and can contain inherent errors due observation or transmission which may have not been detected from the source. The data is imported into Climsoft database and subjected to further quality control checks.

Outliers causing noise in the data and Inconsistencies between data items for example, the minimum temperature <= maximum temperature for the same period at the same location can be detected by Climsoft which has inbuilt modules; these checks can be performed as instructed below:

Before performing any checks, it is important to establish the plausible upper and lower limit of the elements you are to check; some people prefer to put the absolute or highest value ever recorded. This is not advisable because the return period of such values can be more than 30years. The lower and upper limits for precipitation and temperature generated at ICPAC for each stations and then grouped as par climatic zones for easy of operations, The table below shows these grouped limits.

|  |  |  |  |
| --- | --- | --- | --- |
| Areas of interest | Rainfall | Tmax | Tmin |
| Limits | Lower | Upper | Lower | Upper | Lower | Upper |
| Cold places | 0 | 100mm | 7°C | 24°C | -2°C | 16.5°C |
| Places with winter | - | - | 0°C | 20°C | -4°C | 10°C |
| Coastal regions | 0 | 150mm | 15°C | 34°C | 14C | 24°C |
| Mountainous regions | 0 | 130mm | 8°C | 26°C | -2.1° | 16° |
| Lift valleys | 0 | 70mm | 16°C | 35°C | 14.5° | 23° |
| Deserts& marginal areas | 0 | 60mm | 20°C | 38°C | 13.6°C | 25°C |
| Warm Places | 0 | 100 | 13.5° | 33.5°C | 10°C | 22°C |

**Setting the limits**

When logged on Climsoft the welcome data management form will appear as shown in the Fig1a then click on the Metadata icon and in then in check on the element Fig1b. This will give you an interface to set the element upper and lower limits as shown in Fig2a, then click on update to save the changes.
 

**Fig1a Fig1b**

After setting the limits and still logged on Climsoft on the welcome data management form Fig1a click on the QC icon to begin the quality control process, then check in the absolute limit as shown in Fig2b below and click ok

 **Fig2a Fig2b**

The output excel table containing the values to be checked for their correctness will appear as below. NB the output values have to be scaled down because they were factored by 10.



Check the output data against the original to see if it is not a mismatch. From above table, the station within this region falls under this set limit of 700mm except Station ID No 95320520 exceeding the limit with 85.5 mm is to be checked further with neighboring stations to confirm its validity. NB It require a qualified and experienced climatologist to do this kind of work not to discard true anomalous values. It is recommended to flag these values and keep the database as original as possible

Steps to run the quality Control Macro

These ICPAC Macros were developed to do further checks like repeated data, temperature step checks, temporal and spatial consistence of a station, They are used at post data ingestion level away from the DBMS and they are very easy to use. They are intended to manifest the errors or suspicious values which were undetected at the lower level base level before the data can be used for analysis to generate dependable results. The steps used to run the macros are listed below.

1. Open the excel file containing the Macro;
2. Enable the macro by clicking option and enable Macro before it can run;
3. Select and Copy the original dataset and paste it in sheet1 of the macro;
4. Ensure that the data starts on column "I" for the Monthly data;
5. Remove all the -999 from the dataset and replace them with blank;
6. Pre-calculations of Upper and Lower limits have been embedded within sheet macro as it can be seen in the table below;
7. Press Cntr\_Shift\_U for the Upper-limit to run which will highlight or flagged values exceeding the limit;
8. Press Cntr\_Shift\_D this runs temperature rates of change
9. Press Cntr\_Shift\_L to run the Lower-limit separately in case ;
10. Press Cntr\_Shift\_R to run the Repeat check macro;
11. Press Cntr\_Shift\_C to clear the flags ;

After running the macro the anomalous values are flagged in blue or red depending on the magnitude of the error.

**Step2**

This is investigation of the validity of the suspicious values; it should be noted that not all flagged values are wrong. For example the blue anomalous value in Jan 1998 was due to El-Nino and then the red value of 677 in Sept 1986 was missing a decimal point. This kind of work requires patience and a critical mind with lots of experience in climatology.



**Comparison with neighbors test**

Cntr\_Shift\_ N; will sort data by year latitude longitude to compare neighbors.

Check if these high values are appearing in neighboring stations to confirm the consistence of the values identified and how they compare with neighbor in the similar years.

 However this may not work in mountainous regions so other criteria checks can be employed such ENSO years or the temporal checks as par that very station time series data.

**The Temperature quality control**

The Temperature checks outside the database using the Macro and following the same steps highlighted above. The table below uses both the 2sgma upper limit and Step change from the previous day checks. The blue flags have been accepted but the one in red are still suspicious values set for further investigation

**MINIMUM TEMPERATURE FLAGGED VALUES**



**NMHS** **support**

The NMHSs Data managers have been provided with these Macros and trained on error detection procedures. Tanzania Meteorological Authority (TMA) requested technical support to quality control their climate data sets (Rainfall and Temperature). A diagnostic approach to establish the error sources and causes is being done at ICPAC