COST ES0601 STSM
Short Scientific Report

Beneficiary's Name and Institution: Mr Julien VIARRE, ENSEEIHT
Host's Name and Institution: Enric AGUILAR, University Rovira i Virgili de Tarragona
Period: from 6/07/2009 to 20/07/2009
Place: Tarragona (ES)

Purpose of the visit

This Short-Term Scientific Mission aimed to homogenize monthly rainfall and temperature series of the Benchmark dataset using a Penalised Maximal F-Test (Wang, 2008). Both « absolute » (without reference) and « relative » (with reference) approaches have been implemented. Only surrogate series have been corrected.

Description of the scientific content

Programs developed here have been computed with R software.

We used here the second version of the RHtest software package to detect and adjust multiple change-points (mean-shifts) which could exist in data series which may have first order autoregressive errors. Briefly, this version has significant improvements over its previous version, which did not account for any autocorrelation and did not resolve the problem of uneven distribution of false alarm rate and detection power (Wang et al. 2007, Wang 2008a).

First Step. Processing data.

All the rainfall and temperature series from the twenty networks of the Benchmark Dataset have been converted in the suitable format. Then, two distinct procedures have been completed. The first is obtained when a reference series is not available (absolute approach). In this case, the software ensures the detection of change-points. The second is implemented when a reference is computed (relative approach).

Second Step. Automating RhtestV2. Absolute approach (without reference)

A program was computed to automatically process all the series from the twenty networks of the Benchmark Dataset. For each case, a “break file” is generated providing the number of change points identified, their type (type 1 if there is not metadata available), dates, p-values, Pfmax statistics, the 95th percentiles of Pfmax and its 95% uncertainty range (see figure1).

```
5 changepoints in Series ~/Bureau/Tarragona/RHsinref/temp/sur1/000001/ratnm00000150d.txt
     1  19060400    1.0000(    1.0000-    1.0000)  140.1805   20.7459(   18.6211-   23.1128)
     1  19290600    1.0000(    1.0000-    1.0000)  278.2148   20.7572(   18.6308-   23.1260)
     1  19360300    1.0000(    1.0000-    1.0000)   19.0986   19.7403(   17.7434-   21.9672)
     1  19410600    1.0000(    1.0000-    1.0000)   24.7879   21.0622(   18.8913-   23.4854)
     1  19820400    1.0000(    1.0000-    1.0000)   65.8126   21.1951(   19.0048-   23.6431)
```

figure1. Example of break file generated for network 1, station code 00000150

The program computed here lets the choice of considering a break significant if :

- the Pfmax statistics is smaller than its 95th percentiles (choice 1)
- the Pfmax statistics is smaller then the upper branch of the 95% uncertainty range of the 95th percentiles (choice 2)

Results sent have been computed with the first option.

Third Step. Automating RhtestV2. Relative approach (with reference)

The burning issue of considering a series as a reference lies in the method employed to calculate this series. In our study, we estimated the correlation (Pearson's method) between each constituent series of a network. For each series and each date, the reference value is obtained by weighting the best correlated series by their correlations. If only one value is available for a chosen date, the reference value is defined as unavailable. We let the user decide the minimal correlation needed to consider the series in the constitution of the reference. If no series correlations overpass this limit, we ask for a minimal number of series to be used in this case.

Fourth Step. Processing results.

Data adjusted are finally converted in the suitable format (Benchmark Dataset). Comparisons between “absolute” and “relative” approach is implemented to estimate the effects of considering a reference in the determination of multiple change-points.

Example of Network 1. Temperature series. Station Code 00000150 “Caldes de Montbui”

Absolute Approach. *See Graphs in Annex1*

4 changepoints in Series ~/Bureau/Tarragona/RHsinref/temp/sur1/000001/ratnm00000150d.txt

1	19060400	1.0000(	1.0000-	1.0000)	137.1191	21.1339(	18.9707-	23.5477)	0.3195(	0.2677-	0.3694)
1	19290600	1.0000(	1.0000-	1.0000)	266.7321	21.2701(	19.0882-	23.7103)	0.3195(	0.2677-	0.3694)
1	19410600	1.0000(	1.0000-	1.0000)	70.2042	21.5476(	19.3255-	24.0364)	0.3195(	0.2677-	0.3694)
1	19820400	1.0000(	1.0000-	1.0000)	65.5266	21.5942(	19.3653-	24.0911)	0.3195(	0.2677-	0.3694)

Relative Approach. *See Graphs in Annex2*

8 changepoints in Series ~/Bureau/Tarragona/RHconref/temp/sur1/000001/rutnm00000150d.txt

1	19160300	1.0000(	1.0000-	1.0000)	8.2803	5.2020(	4.9021-	5.5117)			
1	19290600	1.0000(	1.0000-	1.0000)	35.2548	5.2336(	4.9330-	5.5433)			
1	19370100	1.0000(	1.0000-	1.0000)	5.7239	5.1484(	4.8567-	5.4582)			
1	19401100	1.0000(	1.0000-	1.0000)	11.7141	5.1868(	4.8933-	5.4965)			
1	19511200	1.0000(	1.0000-	1.0000)	7.2082	5.2336(	4.9330-	5.5433)			
1	19610900	1.0000(	1.0000-	1.0000)	9.7099	5.3034(	5.0020-	5.6237)			
1	19871000	1.0000(	1.0000-	1.0000)	11.9277	5.2956(	4.9942-	5.6147)			
1	19950500	1.0000(	1.0000-	1.0000)	10.2686	5.1589(	4.8669-	5.4687)			

Only the break in 1929 remains significant for the two procedures. The increase of break points number when we use a reference may come from the method used to compute reference values. A comprehensive analysis is needed to compare both of the methods.

This theme could be the target of a future collaboration with host institution. Moreover, we begun the implementation of the daily adjustment presented by Vincent et al. (2001). The program computed has to be improved in order to take benefits of daily climate data.

All the series have thus been adjusted using RhtestV2 software package. Programs computed to generalize Rhtest in order to process all the series successively are available (julien.viarre@gmail.com).