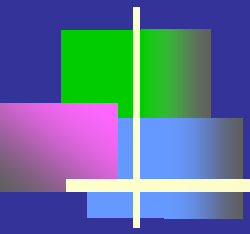
Introduction to Climate Data Homogenization techniques



By Thomas Peterson

Using material stolen from Enric Aguilar*

CCRG
Geography Unit
Universitat Rovira i Virgili de Tarragona
Spain

* Who in turn stole material prepared by Lucie Vincent, Climate Research Branch, Meteorological Service of Canada Environment Canada

Objective

Detecting steps in climatological time series, even without the prior knowledge of the position in time and magnitude of the inhomogeneity

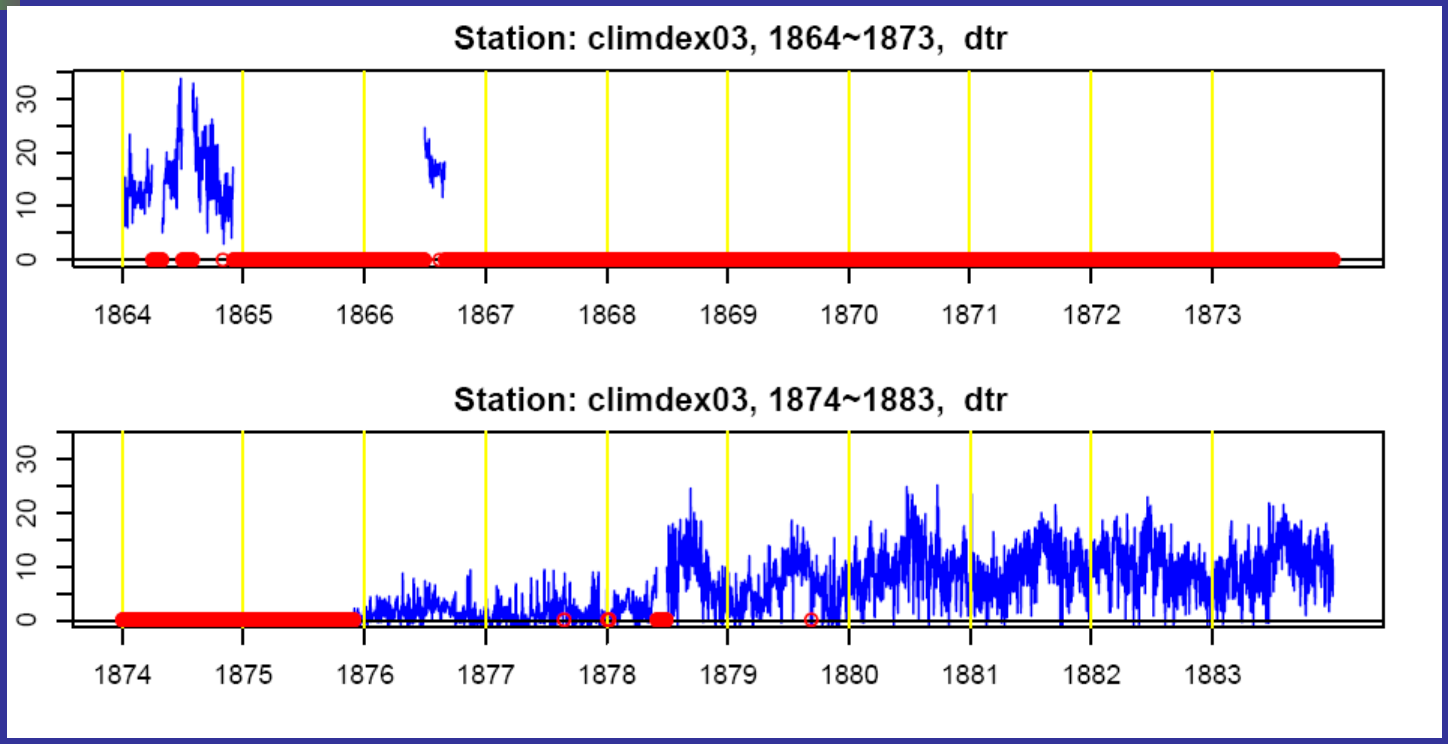
USING WORKSHOP SOFTWARE TO TEST HOMOGENEITY OF TIME SERIES

1) CHECKING THE DATA

2) CHECKING INDICES

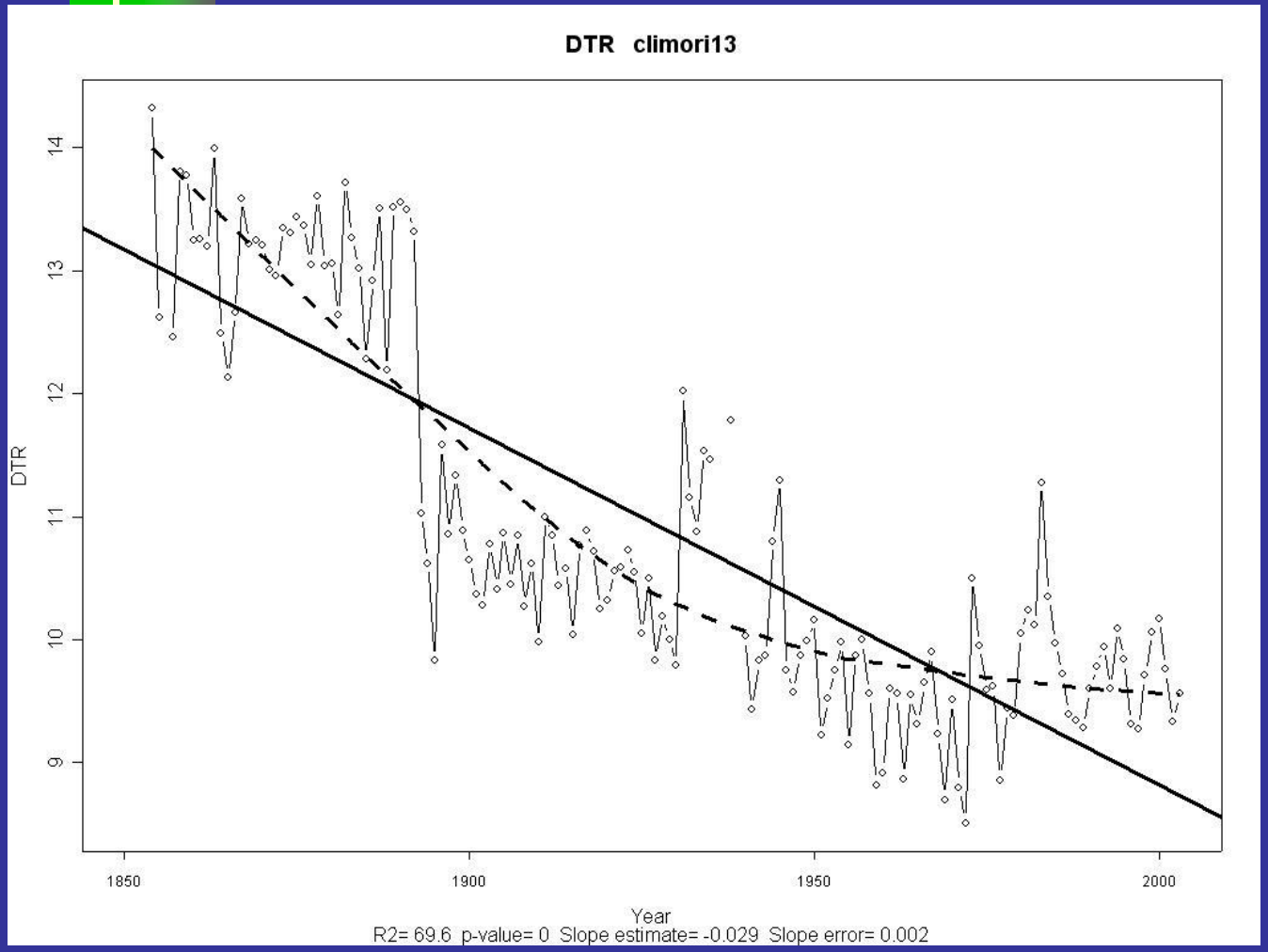
3) HOMOGENEITY TESTING BASED ON REGRESSION MODELS: F-test for the comparison of regression models and for detecting the position in time and magnitude of significant steps (VERY RECENTLY implemented in Rclimdex; Fortran program available)

CHECKING THE DATA



DAILY VALUES OF DTR FOR BADAJOZ, SPAIN, 1864-1884.
Notice obvious change in DTR between 1878 & 1879. File was produced with Rclimdex's QC utility

CHECKING THE INDICES (I)

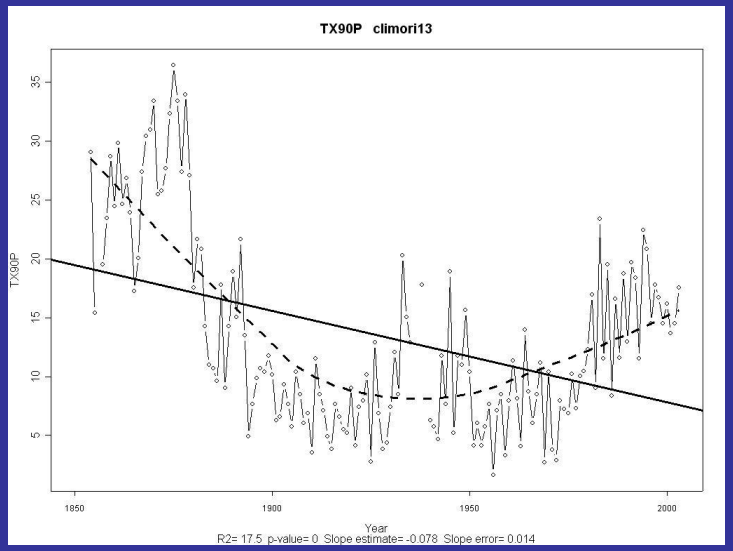


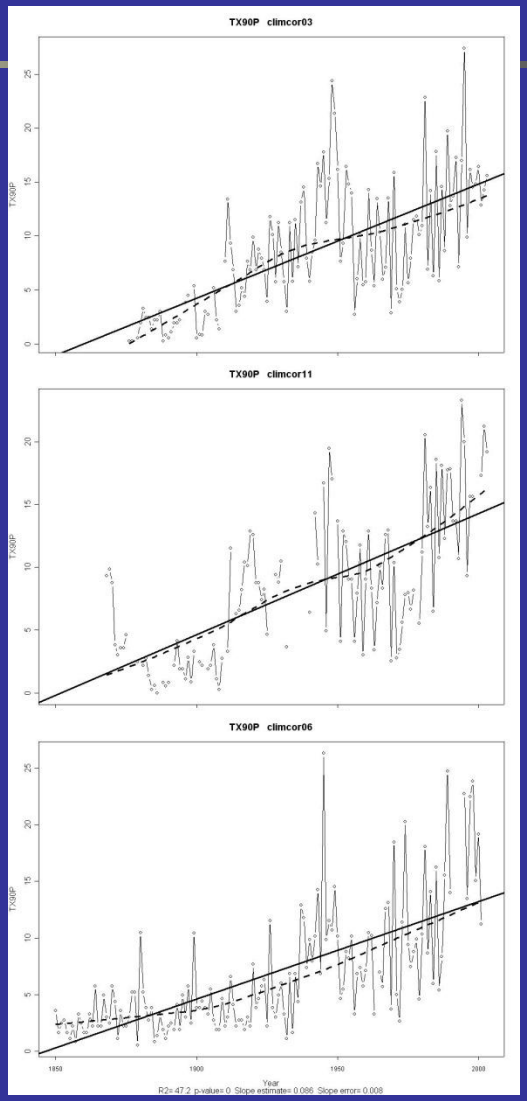
Data for Madrid, Spain (non-homogenized)

Obvious Change in DTR index values IN 1893.

Metadata reports a change in shelter

CHECKING THE INDICES. CONTRASTING STATIONS (II)





DTR INDEX: RESULTS FOR MADRID (LEFT) LOOK VERY DIFERENT TO RESULTS FOR BADAJOZ, HUESCA & CÁDIZ (RIGHT, TOP TO BOTTOM)

SOME NATION-WIDE PROBLEMS MAY NEED CONTRAST TO FOREIGN STATIONS

Techniques for the detection of discontinuities in climatological series

- Many techniques developed and applied on annual and monthly temperature and precipitation
- Techniques based on statistical approach
- Detailed review in Peterson et al. 1998
- New updated homogeneity evaluation from European COST-HOME project

- Bayesian Approach
  Perreault et al. 2000 – INRS, Canada
- Caussinus – Mestre Technique
  Caussinus & Mestre 1996 – Meteo-France
- Multiple Analysis of Series for Homogenization (MASH)
  Szentimrey 1996 – Hungarian Meteorological Service
- Multiple Linear Regression
  Vincent 1998 – Climate Research Branch, Canada
- Potter's Method
  Potter 1981 - ?
- Standard Normal Homogeneity Test
  Alexandersson 1986 – Swedish Meteorological Institute
- Two-Phase Regression
  Easterling & Peterson 1995 – NCDC, USA

and many more …

Technique based on regression models

(Easterling & Peterson 1995; Vincent 1998; Lund & Reeves 2002)

Model 1 :

$$y_i = a_1 + b_1 t_i + e_i$$

where y_i : candidate series

t_i : time

Model 2 :

$$y_i = a_2 + b_2 t_i + c_2 I_i + e_i$$

where $I_i = 0$ for $I = 3,\ldots,p-1$

$I_i = 1$ for $I = p,\ldots,n-3$

$$F^* = [(SSE1-SSE2)/1] / [SSE2/(n-3)]$$

if $F^* > F_{max}$ accept Model 2

(F_{max} : Wang 2003)

Example

1917-2000

Model 1 :

$y_i = 1.1 - 0.016t_i + e_i$

SSE1 = 20.7

Model 2 :

$y_i = 0.9 + 0.007t_i - 1.3I_i + e_i$

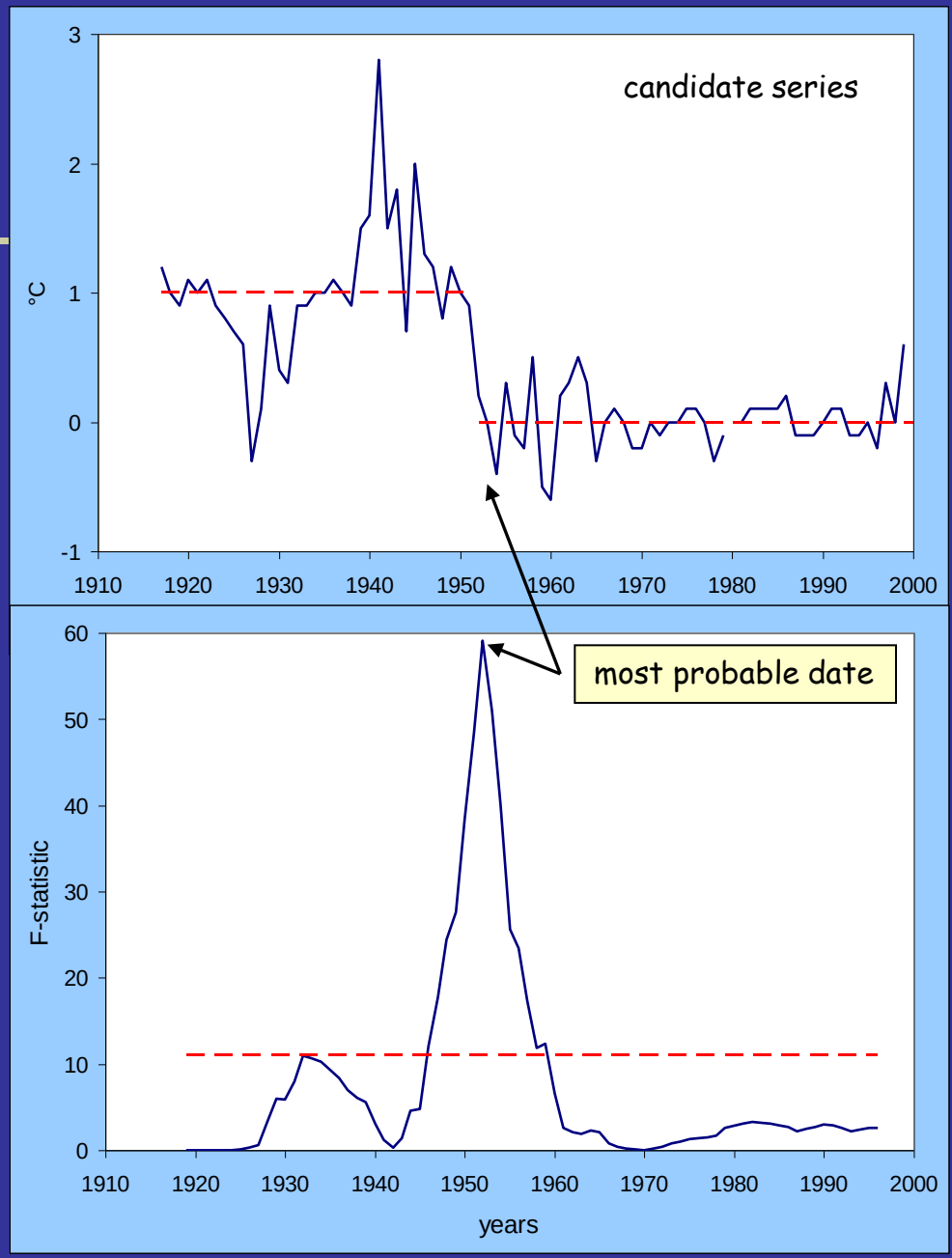
date of the step = 1952

magnitude = -1.3°C

SSE2 = 11.9

F* = 59.1 > 11.1

therefore there is a step in 1952



Example

1917-1951

Model 1 :

$y_i = 0.7 + 0.019t_i + e_i$

SSE1 = 8.6

Model 2 :

$y_i = 1.0 - 0.020t_i + 0.9I_i + e_i$

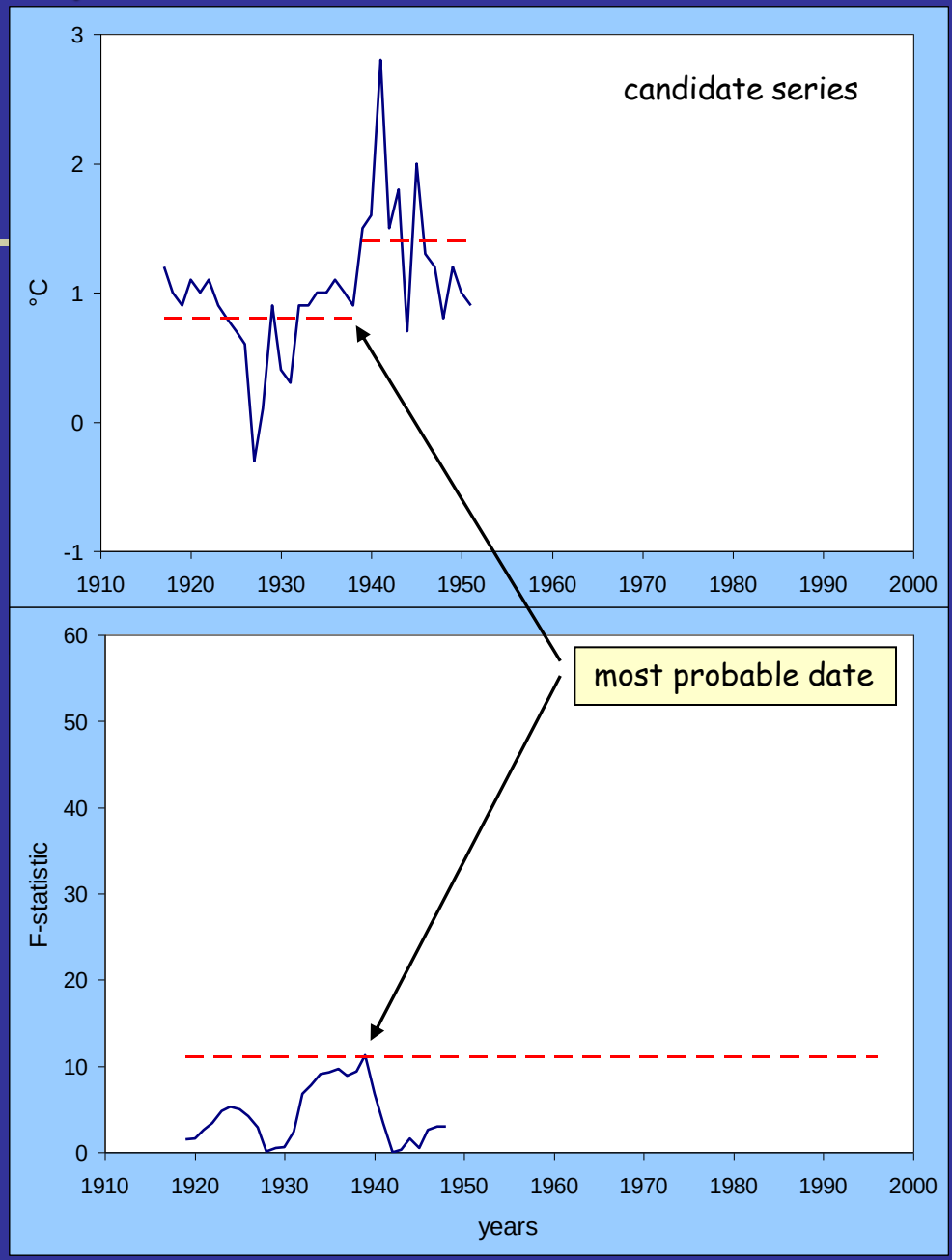
date of the step = 1939

magnitude = 0.9°C

SSE2 = 6.3

F* = 11.3 > 11.1

therefore there is a step in 1939



Example

1952-2000

Model 1 :

$y_i = -0.1 + 0.002t_i + e_i$

SSE1 = 2.59

Model 2 :

$y_i = -0.1 + 0.007t_i - 0.2I_i + e_i$

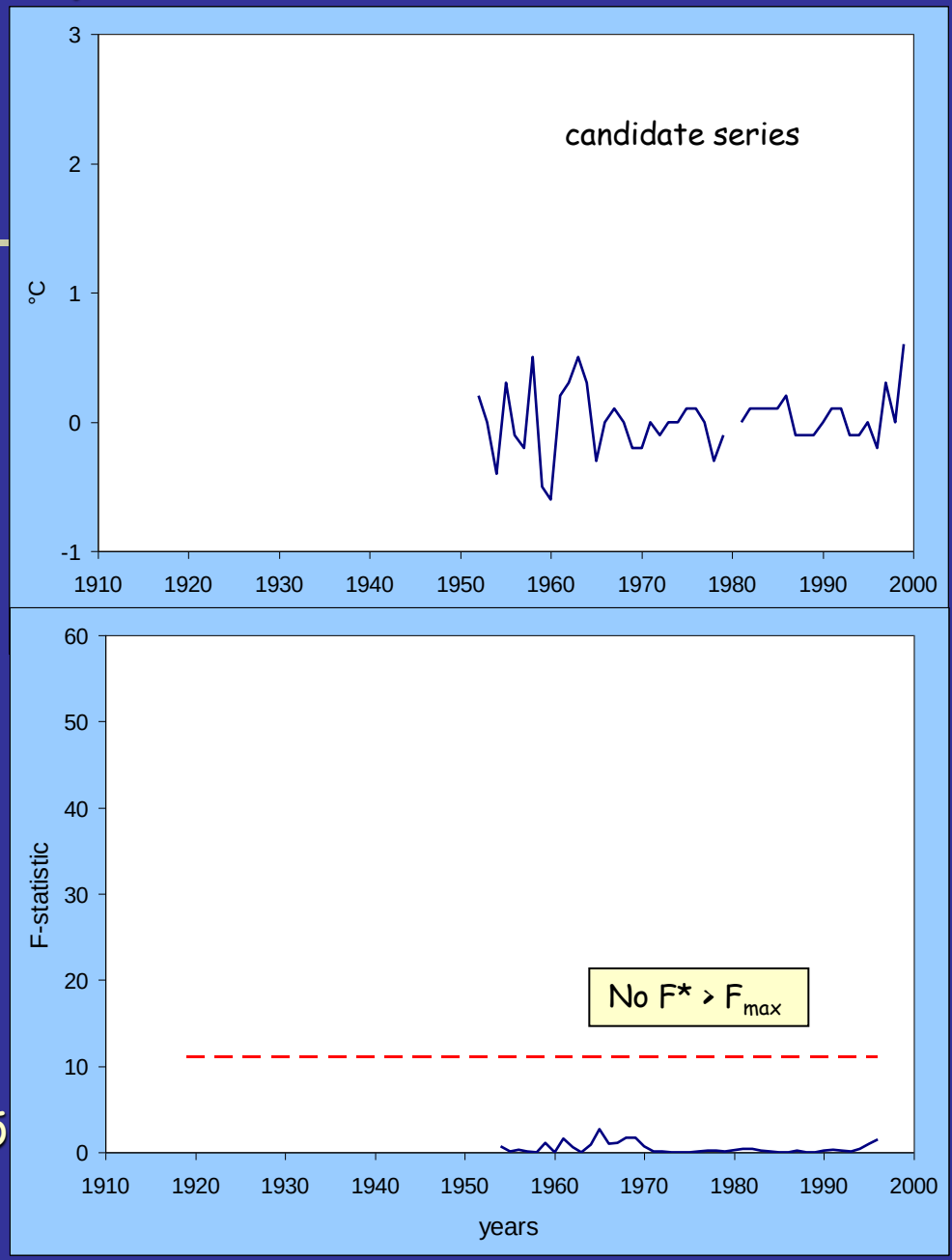
date of the step = 1965

magnitude = -0.2°C

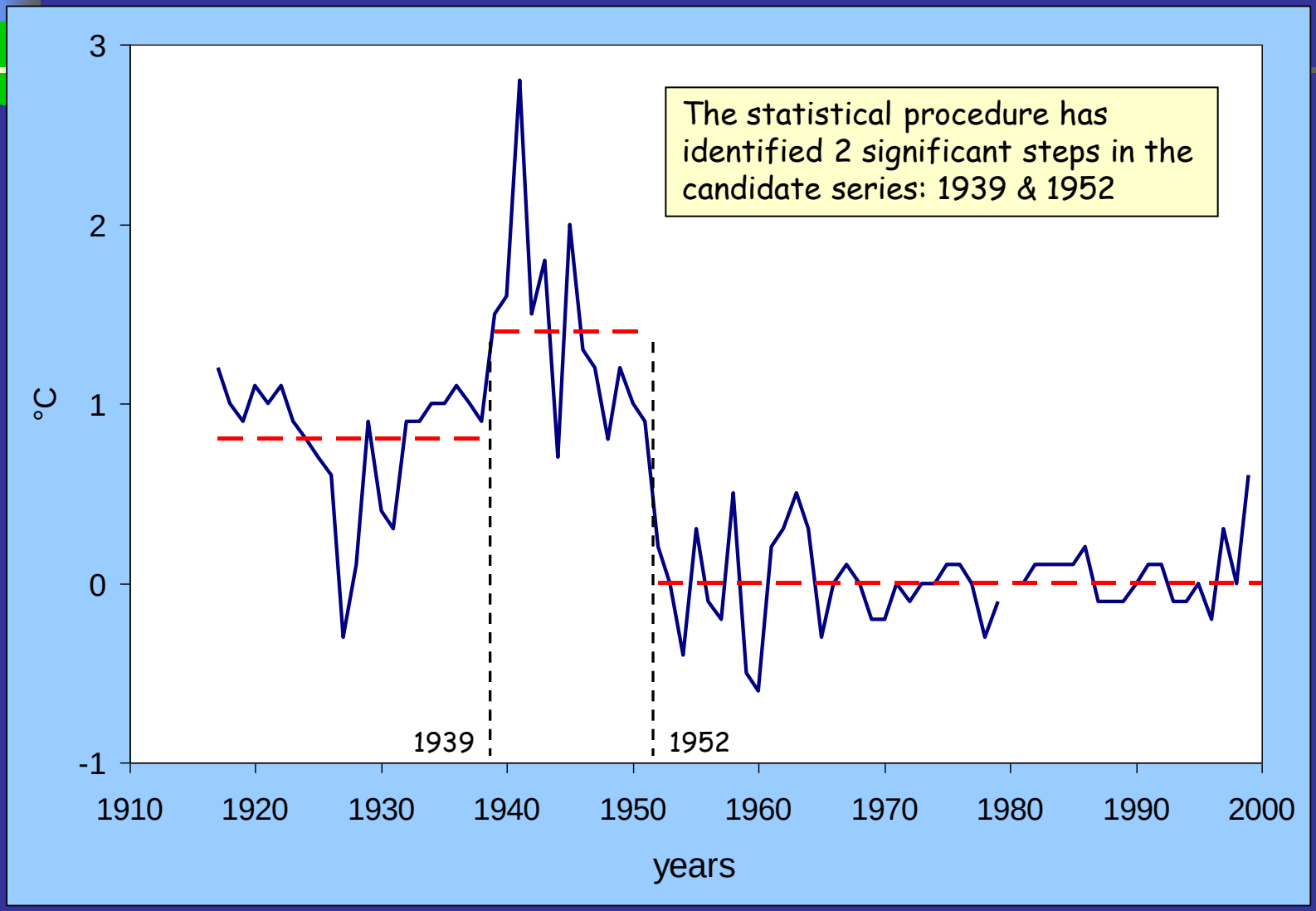
SSE2 = 2.58

F* = 0.01 < 11.1

therefore there is no step in 1965



Example


The statistical procedure has identified 2 significant steps in the candidate series: 1939 & 1952
3
2
1
0
-1
°C
1939
1952
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
years

METADATA & STATISTICS

INM code	Name	Address	Location	Lon	Lat	Altitude	Starts	Ends
2030	SORIA	Instituto 2ª Enseñanza	¿?	02º 28' W	41º 49' 10''	1058.5 m	01/09/1871	31/12/1879
2030	SORIA	Instituto 2ª Enseñanza	¿?	02º 28' W	41º 49' 10''	1058.5 m	01/01/1880	28/02/1893
2030	SORIA	Instituto 2ª Enseñanza	¿?	02º 28' W	41º 49' 10''	1058.5 m	01/03/1893	30/10/1893
2030	SORIA	Instituto 2ª Enseñanza	¿?	02º 28' W	41º 49' 10''	1058.5 m	01/11/1893	31/12/1900
2030	SORIA	Instituto 2ª Enseñanza	¿?	02º 28' W	41º 49' 10''	1058.5 m	01/01/1901	1910
2030	SORIA	Instituto 2ª Enseñanza	Jardín	02º 28' W	41º 49' 10''	1058.5 m	1911	31/12/1942
2030	SORIA	Piso c/ Navas de Tolosa	Terraza	02º 28' W	41º 46'	1083 m	**01/11/1943**	31/12/2002

ANNUAL Homogenouse test statistics report:

Year=1919,Fstat=3.3462,pval=0.9254,Fm90=9.26,Fm95=11.15,Fm99=15.75,StepSize=0.4,SegmLen=44

Year=**1946**,Fstat=7.1762,***pval=0.9902***,Fm90=9.3,Fm95=11.09,Fm99=15.23,StepSize=0.41,SegmLen=57

Year=1981,Fstat=7.6858,pval=0.9924,Fm90=9.3,Fm95=11.09,Fm99=15.23,StepSize=0.76,SegmLen=57

CONCLUSIONS

- HOMOGENIZATION ASSESSMENT ON AN ANNUAL/MONTHLY BASIS WILL PREVENT MAJOR INHOMOGENEITIES TO CORRUPT THE TRENDS ANALYSIS, DISCARDING SERIES OR INHOMOGENEOUS SEGMENTS
- EVEN WHEN A CANDIDATE STATION IS/LOOKS HOMOGENEOUS AT MONTHLY & ANNUAL SCALE, INHOMOGENEITIES MAY REMAIN ON A DAILY BASIS


Thank You