Geophysica (2002), 38(1-2), 15-41

Homogeneity Testing and Adjustment of Climatic Time Series in Finland

Heikki Tuomenvirta

Finnish Meteorological Institute, P.O. Box 503, FIN-00101 Helsinki, Finland

(Received: June 2001; Accepted: December 2001)

Abstract

The statistical tests used for the detection and adjustment of homogeneity breaks in climatic time series are mostly objective. However, subjective decisions on the design of the testing procedure are usually required. These include the choice of critical significance levels to adopt, decisions on whether to test annual or seasonal values, and selection of methods for using station histories and other information on data (metadata). This leads into different approaches in testing. In Finland, the Standard Normal Homogeneity Test has been used together with metadata in the building of long-term climate data sets from the original weather observations. This paper describes the testing and adjustment process applied for temperature, precipitation and air pressure time series and provides examples. Homogenisation procedures such as these are essential for ensuring the reliability and suitability of long-term time series for studies of climatic changes and variations.

Key words: climatological time series, homogeneity testing, metadata, temperature, precipitation, Finland

1. Introduction

1.1 Background

Ideally, only reliable observational data should be used in the analysis of climate. Many types of disturbances can cause apparent changes in long-term climatological time series, which may distort the true climatic signal. Breaks in the homogeneity of time series may be caused by factors such as changes in instruments, in observation practices, in station location and in the environment. Homogenisation of time series is widely recognised to be one of steps that has to be taken in the process of construction reliable long-term data sets from original climate observations. *Peterson et al.* (1998) review many techniques to detect and adjust non-homogeneities in time series and they describe some approaches used in the building of homogenised data sets. Previous work in the Nordic region demonstrates that original temperature and precipitation series would have been systematically biased without adjustments, and inhomogneities of individual time series can be as large as decadal variations of climate (*Tuomenvirta,* 2001).

There are several ways to apply statistical homogeneity tests to climatological time series. Although the tests themselves are objective, subjective choices are still required in the application of tests and in the use of available "metadata" (used hereafter to refer to information on observations, including instruments, observation methods, observing practices and data processing). This results in different approaches being employed by different researchers. The objective of this paper is to describe the methodology used in Finland to detect and adjust temporal inhomogeneities of climatological time series with the help of metadata.

The Standard Normal Homogeneity Test (SNHT) developed by *Alexandersson* (1984, 1986) is described in detail, because it has been the main statistical homogeneity testing tool used in Finland. The procedures used in applying the test to a large number of time series is explained, and the advantages and disadvantages of SNHT are presented with reference to other published work.

1.2 Climatological data

The methods presented in this paper have been applied to observations from the Finnish meteorological station network compiled at a monthly resolution in the Finnish Meteorological Institute (FMI). Table 1 lists the elements that have undergone at least some testing. Data sets where some of the time series have been published are also given as well as papers where the author has analysed data. Figure 1 shows the locations of stations that have been included in Nordic data sets. Although fewer than twenty stations are typically distributed internationally, the testing involves and produces information on the homogeneity of around 50-250 stations in Finland (*Tuomenvirta,* 2001).

Table 1. Climatic elements tested with SNHT in Finland, data sets where some of the tested time series have been published, and articles where the author has used data. DATA SETS: North Atlantic Climatological Dataset (NACD), *Frich et al.* (1996); North Atlantic-European pressure observations (WASA dataset), *Schmith et al.* (1997); REWARD (Relating Extreme Weather to Atmospheric circulation using a Regionalised Dataset), *Førland et al.* (1998). STUDIES: TH96 = *Tuomenvirta* and *Heino* (1996), T2000 = *Tuomenvirta et al.* (2000), T2001 = *Tuomenvirta* (2001).

	Mean Temperature	Maximum Temperature	Minimum Temperature	Precipitation Total	Air Pressure at Sea Level	Cloud Cover*
SYMBOL	**T**	**Tx**	**Tn**	**R**	**P**	**C**
DATA SETS	**NACD**	**REWARD**	**REWARD**	**NACD**	**NACD, WASA**	**NACD**
STUDIES	**TH96, T2001**	**T2000, T2001**	**T2000, T2001**	**TH96, T2001**	**TH96, T2000**	**TH96, T2000**

* Only some experimental homogeneity testing has been performed with cloud cover.

In the process of constructing reliable data sets from long-term climatological observations, homogenisation of time series is a step that has to be taken after identifying the sources of observations, and digitisation and preliminary quality control of the data. In Finland, known homogeneity breaks were also adjusted before the testing of individual stations with SNHT (*Tuomenvirta and Heino* 1996). These include the nationwide methodological and instrumental changes identified by *Heino* (1994), that cause systematic biases in the original data. However, only a few of the documented station relocations could be adjusted based on side-by-side comparison measurements (*Tuomenvirta and Heino,* 1996).

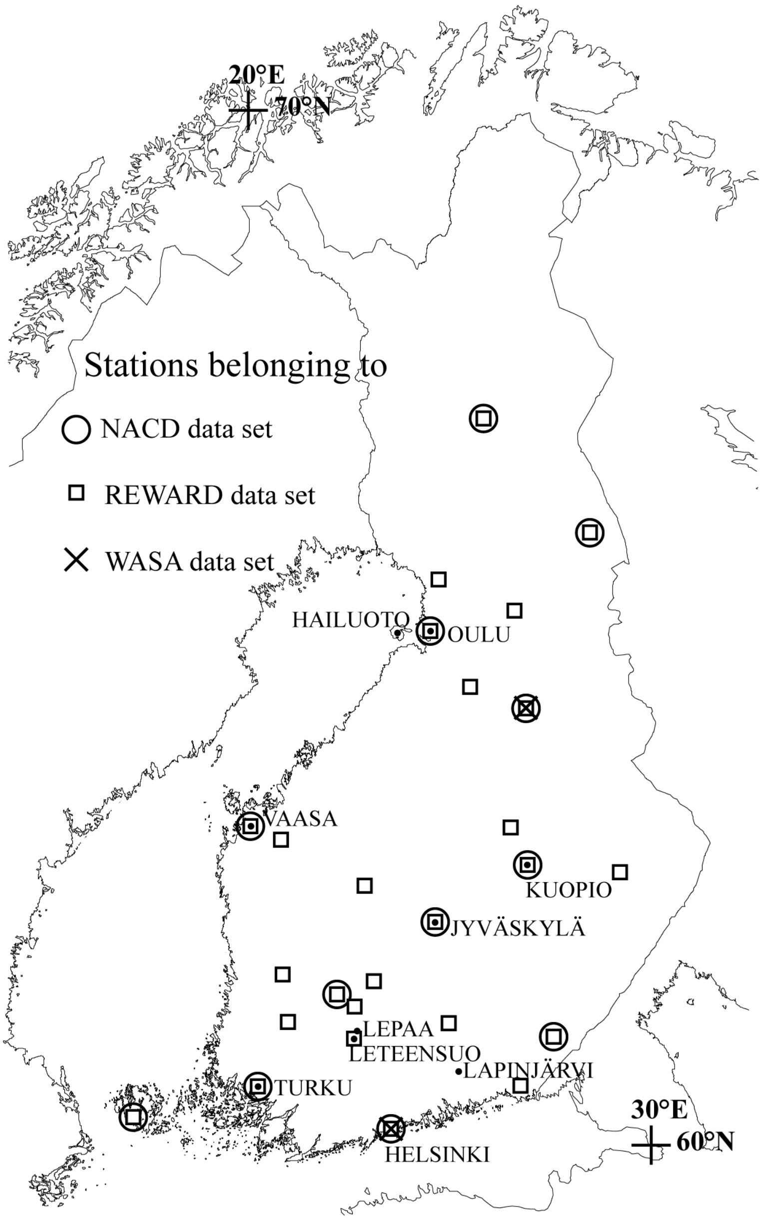

Fig. 1. Finnish stations included in the NACD, REWARD and WASA data sets (see Table 1), and location of stations mentioned in this paper (marked with dots and name given).

2. *Homogeneity testing and adjustment using the Standard Normal Homogeneity Test*

In order for a meteorological or climatological observational time series to be regarded as perfectly homogeneous, it should record variations that are attributable to weather and climate fluctuations alone (*Conrad and Pollack* 1950). This would require that observations be performed at the same site within an unchanged environment using the same calibrated instrument according to the same method. In reality, these requirements are rarely fulfilled in long time series, and their "absolute homogeneity" is always questionable. Instead, climatologists must make do with series that are "relatively homogeneous", where the differences or ratios between candidate station series and synchronous series at neighbouring (relatively) homogeneous stations are statistically random series.

Because metadata are incomplete, a technique is required that can both identify an inhomogeneity without knowing *a priori* the time of a break point in the time series, and can also estimate the magnitude of the identified break. For these purposes, the Standard Normal Homogeneity Test (SNHT) has been used in Finland.

2.1 Basic assumptions of the Standard Normal Homogeneity Test

SNHT is a parametric test using neighbouring station(s) as a reference to identify non-homogeneities in the time series of the station being tested (candidate station). It is used to detect abrupt or linearly developing differences between the candidate and the reference station(s). *Hawkins* (1977) presented a formulation of a testing method that was subsequently developed into SNHT and applied to climatological series by *Alexandersson* (1984, 1986). The SNHT is related to a curve fitting technique using the least squares principle (*Alexandersson and Moberg, 1997*, hereafter *AM,* 1997).

The basic assumption behind SNHT is that the ratio/difference, Q, between e.g. precipitation/temperature at the candidate station and a neighbouring reference station remains fairly constant in time. This requires a sufficient correlation between the test and reference stations. An inhomogeneity will be revealed as a systematic change in this ratio/difference, Q, which is defined in section 2.3 in which techniques to construct reference series are also discussed.

SNHT uses normalised series of the ratios/differences, Z_i, defined as

$$Z_i = (Q_i - \overline{Q}) / \sigma_Q \tag{1}$$

where $\overline{Q}$ is the sample mean value and σ_Q the sample standard deviation of the ratio/difference Q_i at time step i (denoted in many climatological applications as one year). In the following discussion, "year" will be used instead of "time step" or "unit time", although the time step is by no means restricted to one year.

After making the assumption that Z_i is described by a Normal distribution, N, the null hypothesis for all variants of SNHT is:

$$H_0: \quad Z_i \in N(0,1) \quad i \in \{1,...,n\}$$

i.e. the whole series is homogeneous. All values in the normalised series of ratios/differences are normally distributed with a mean value equal to zero and standard deviation equal to one.

2.2 *Single shift of the mean level*

The alternative hypothesis is that the series is inhomogeneous. One reason for this might be that there is a single shift in the mean level of the candidate station. At some unknown time the mean value changes abruptly, while the standard deviation remains unchanged (*Alexandersson,* 1986, *AM,* 1997). For example, a precipitation gauge is moved to a more sheltered site thus reducing the wind error and increasing the amount of measured precipitation. In this case, the alternative hypothesis, H_1, is written as

$$H_1 : \begin{cases} Z_i \in N(\mu_1,\sigma) & i \in \{1,...,a\} \\ Z_i \in N(\mu_2,\sigma) & i \in \{a+1,...,n\} \end{cases}$$

where μ_1 is the mean value during the first a years, μ_2 is the mean value during the last (n-a) years, and σ is the sample standard deviation.

Alexandersson (1986) and *AM* (1997) show how the test quantity, T, which separates H_1 from H_0, is derived. The interpretation of T is that a high value at year a suggests that μ_1 and μ_2 depart significantly from zero, making H_1 likely. The maximum value of T, denoted T^s_{max}, is

$$T^s_{max} = \underset{1\le a\le n-1}{Max}\left\{T^s_a\right\} = \underset{1\le a\le n-1}{Max}\left\{a\overline{z_1}^2 + (n-a)\overline{z_2}^2\right\} \tag{2}$$

where $\overline{z_1}$ and $\overline{z_2}$ are the mean values before and after the shift. The corresponding value of a is the most probable break point, i.e. the last year at the old level. The null hypothesis can be rejected, if T^s_{max} is above the selected significance level, which depends on the length of the series. *AM* (1997) give critical T values for 10%, 5% and 2.5% levels, T_{90}, T_{95}, and $T_{97.5}$ respectively. According to *Hawkins* (1977) there is an increased probability for high T values near the ends of series where a few low or high values of Z_i make T^s_a large.

The single shift SNHT has been the main tool applied for homogeneity testing in Finland. There are also three other variants of SNHT which have been used but to a lesser extent. *Alexandersson* (1995) formulated a test for the double shift of the mean level in difference/ratio series. The double break test is useful for dividing long time series into shorter periods and has been used for that purpose in Finland. *AM* (1997) presented a version of SNHT to test for a linear trend caused by gradual changes as well as a version for detecting a single shift of both the mean level and the standard deviation in difference/ratio series. The alternative hypotheses and test statistics of these three versions of SNHT are given in the Appendix.

2.3 *Building of reference series*

In theory one homogeneous series having a high correlation with the candidate series is sufficient for the purpose of testing. However, it is usually advisable to build a reference series from more than one series, when available. The use of several series reduces the effects of spatial variation and possible inhomogeneities in the reference series.

The formulae presented in sections 2.1 and 2.2 and Appendix (A1-A8) are independent of the formulation of the reference series. In several applications of SNHT, the ratio/difference series, Q, are defined in the following way. Y denotes the candidate series and Q_i denotes a specific value at year (time step) i. X_j denotes a reference series at station j (out of a total number of k series). The ratio term is formed as

$$Q_i = \frac{Y_i}{\left[\sum_{j=1}^{k} V_j \, X_{ji} \, \overline{Y}/\overline{X_j}\right] / \sum_{j=1}^{k} V_j} \tag{3}$$

and the difference term as

$$Q_i = Y_i - \left\{ \sum_{j=1}^{k} V_j \left[X_{ji} - \overline{X_j} + \overline{Y}\right] / \sum_{j=1}^{k} V_j \right\} \tag{4}$$

where V_j is a weighting factor for reference station j. Factor V_j allows more weight to be put on the "best" reference stations. Often V_j is the square of the correlation coefficient between candidate and reference stations, though equal weighting has mostly been used on the Finnish data. An overbar denotes a time mean. Normalisation allows the use of non-complete series as reference series. The mean values used in this normalisation must be calculated for the same time period for the candidate and all reference series. Otherwise natural climatic fluctuations may mask homogeneity breaks or cause a false detection of break. Normalisation causes the ratios to fluctuate around one and differences around zero.

The denominator in (3) and the term within brackets in (4) are called the reference values. They give an estimate of the corresponding value at the candidate station. In addition to the standard testing procedures, the Q-values can also be used to point out suspect single values at the candidate station (outlier testing) and in the interpolation of missing values in the candidate series (*Moberg and Alexandersson*, 1997)

Peterson and Easterling (1994) have developed a routine for building the reference series, where they use differences between successive years to reduce the effect of step-like discontinuities on the correlation coefficient. A similar large homogeneity break in both the candidate and the reference stations may artificially increase the correlation between these stations. Likewise, large breaks may mask the correlation between stations. Calculation of correlation coefficients from series showing the change in data per unit time reduces the risk of neglecting a potentially good reference station.

Figure 2 shows an example from Hattula in southern Finland where two stations 8 km apart from each other (mostly observing the same sky) have long cloud cover series. There seem to be homogeneity breaks in the original cloud cover series. The correlation coefficient of the annual mean cloud cover series between the two stations is -0.16. However, the correlation between change-in-cloud-cover (first difference dC/dt = C_{i+1}-C_i, here C_i is the annual mean cloud cover at year i) series is 0.84. This suggests that the two series are closely related. As stated by *Heino* (1994) the homogeneity breaks are probably due to changes of observers.

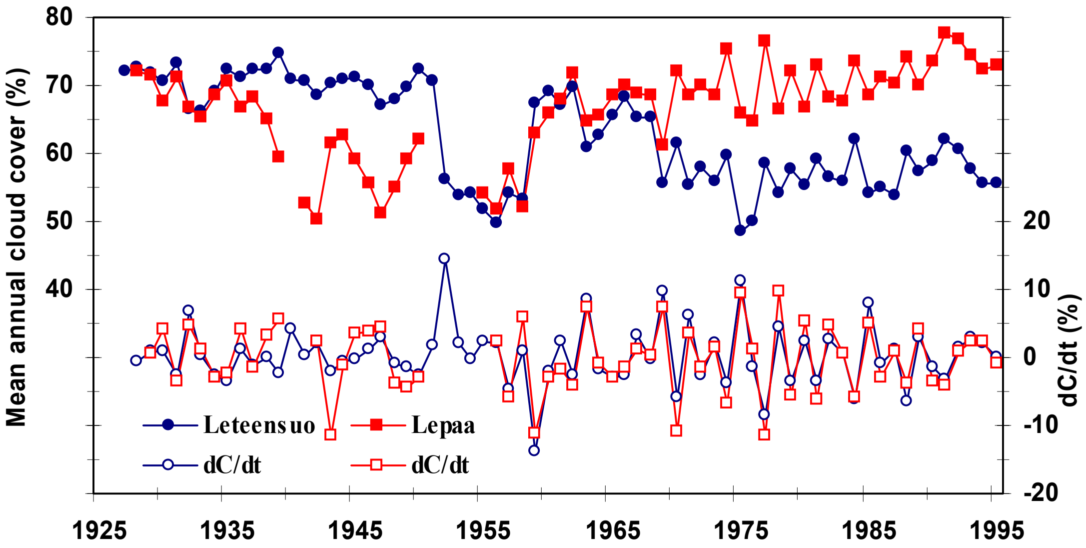

Fig. 2. Mean annual cloud cover (right axis) and change-in-cloud-cover (dC/dt) series from Hattula Leteensuo (61°04'N, 24°14'E) (circles) and Hattula Lepaa (61°08'N, 24°20'E) (squares), 1925-95. Some years of data are missing from Lepaa.

2.4 *Calculation of adjustments*

SNHT provides an estimate of the size of the detected discontinuity that can be used to adjust an inhomogeneous series. In the single shift test, the two levels of the ratios or differences derived from (1), $\overline{q_1}$ and $\overline{q_2}$, are

$$\overline{q_1} = \sigma_Q \overline{z_1} + \overline{Q} \tag{5}$$

$$\overline{q_2} = \sigma_Q \overline{z_2} + \overline{Q} \tag{6}$$

The adjustment for years from 1 to a is $\overline{q_2}/\overline{q_1}$ for ratios, and $\overline{q_2} - \overline{q_1}$ for differences. After the adjustments have been applied, the data are homogenised to the present measuring situation provided that the series contained only one homogeneity break. Double shift adjustments are calculated in a similar manner as for the single shift. Trend adjustments must be calculated with μ_1 and μ_2 from (A4) and (A5) and applied to (5) and (6), respectively.

One possible way to handle a change in both the mean level and standard deviation is described in *Tuomenvirta and Alexandersson* (1995). Firstly, linear regressions between the candidate and the reference series are formed, both before and after the homogeneity break. Secondly, the change of slope of the regression line is used as an amplification/damping factor in the adjustment procedure. If reference values are denoted with R, linear regression of the candidate value Y becomes

$$Y = \alpha R + \beta \tag{7}$$

where α is the slope and β is the y-axis intercept. The regressions are calculated before and after the break, denoted 1 and 2, respectively. The adjusted value at the candidate station, Y_i^a, can be calculated as

$$Y_i^a = \overline{Y_1} + \Delta Y + (a_2/a_1)(Y_i - \overline{Y_1}) \tag{8}$$

Here ΔY is the change of the mean level. The last term is the amplification/damping factor, which is the ratio of slopes times anomaly.

An example of the use of linear regressions in the calculation of adjustments is shown in Figure 3. Oulu climatological station (65°02'N, 25°29'E) was moved from the town centre by the sea to the university campus about 5 km from the shoreline in 1983. Figure 3a displays the linear regressions of the December mean temperature anomalies. The period 1972-82 values are above the period 1983-94 values, i.e. the old site is warmer than the new one due to the warming effects of the sea, the river and urban heating. The regression line at the university campus (1983-94) is steeper than in the town centre (1972-82) thus indicating larger variability. The December mean temperature adjustment is on average -0.73°C, but there is a strong temperature dependency (Fig. 3b). The old site at the town centre is more maritime and the cold temperature anomalies are damped. Therefore, the cold Decembers of the period 1972-82 require large negative adjustments to represent the new site.

Tuomenvirta and Alexandersson (1995) compared monthly mean temperature adjustments calculated with (8) and a "constant" adjustment $\overline{q_2} - \overline{q_1}$. They showed that (8) gave more reliable adjustments, especially during winter. This was due to the fact that, in certain weather types, the size of the adjustment is a function of the temperature anomaly. Equation (8) gives better adjustments in inversion situations, which are frequent at high latitudes in wintertime. However, calculation of the adjustment requires good reference series.

The use of reference series in the calculation of adjustments introduces a regionalization effect into the data. The adjustment of series to represent homogeneous conditions at a particular site can be difficult. It is possible that procedures using other climatic elements (e.g. temperature with wind, albedo, stability, etc.) might help, but this is not usually done. In any case, the original data must always be preserved because new and better approaches to homogeneity adjustments may be developed.

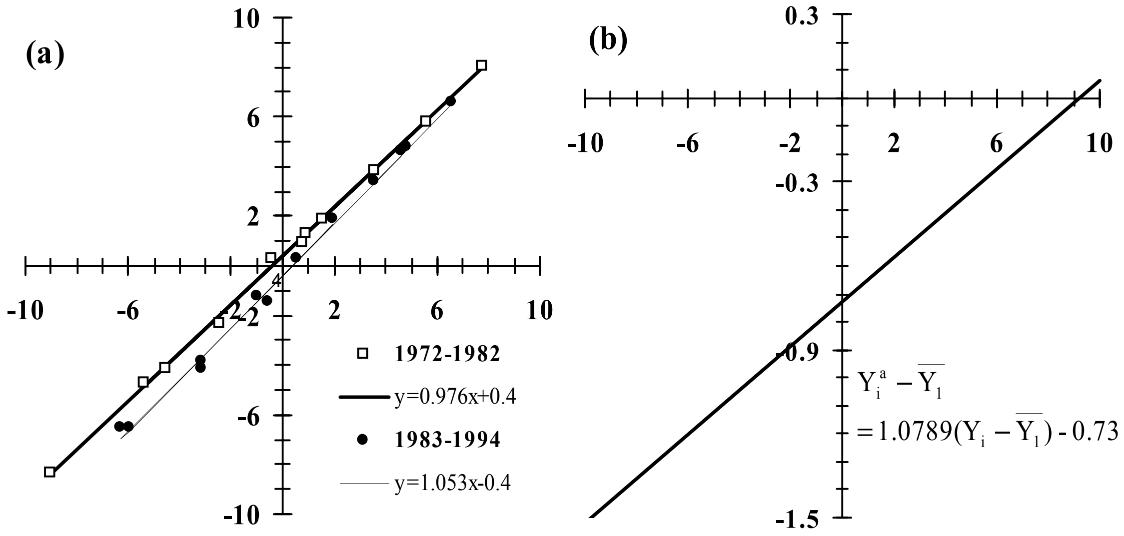

Fig. 3. (a) Linear regressions between the observed December mean temperature anomalies (y-axis) and the reference series (x-axis) at Oulu for the periods 1972-82 (town centre) and 1983-94 (university campus). (b) Adjustments of December mean temperature (y-axis) as a function of the temperature anomaly (x-axis) at Oulu. (All units: °C).

3. *Examples of inhomogeneities in the Finnish data*

Three examples of the use of different variants of SNHT are presented. In order to focus on the use of SNHT, there are homogeneous reference series available in the following examples. They have commonly been built from 6 to 9 relatively homogeneous series. Usually at the start of testing there are no homogeneity-tested series available. The problem of building a homogeneous reference series is discussed in section 4. The causes of homogeneity breaks are also described, according to the station history information. Finally, the seasonal variation of adjustments and their causes are discussed, with further examples.

3.1 *Abrupt change due to change of measurement site*

A weather station has been running continuously at Turku airport (60°31'N, 22°16'E) since 1955. However, the single shift SNHT for Turku airport annual precipitation series shows a T^s_{max} value (equation 2) of 18.1 in 1979, well above the 95% significance level (Fig. 4). So the null hypothesis, that the whole series is homogeneous, must be rejected. The Q-values before 1979 are commonly less than one and after 1979 mostly larger than one. The reason for the homogeneity break was a relocation of the observing site within the airport area. At the new site the precipitation gauge catches about 8% more precipitation than at the previous site. The adjustment ($\overline{q_2}/\overline{q_1}$ =1.08) is applied and SNHT run again. For the adjusted data the Q-value fluctuates around one and T^s_{max} is not statistically significant. The adjustment has made the series relatively homogeneous during the period 1960-99.

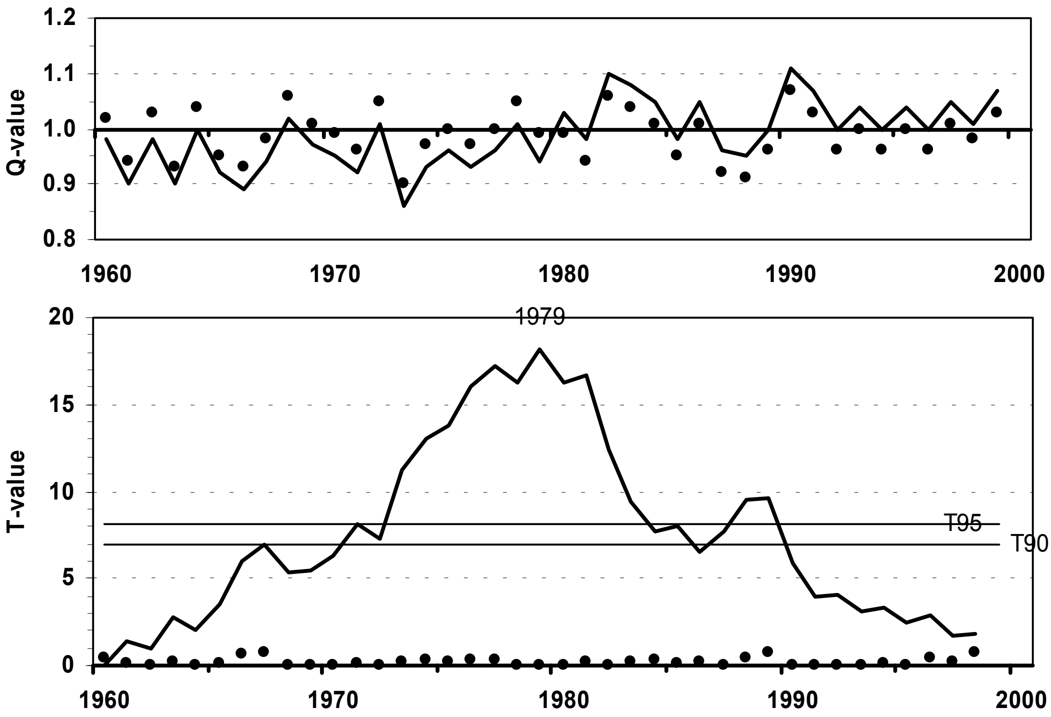

Fig. 4. Values of Q and T from performing the single shift SNHT on the annual precipitation series from Turku airport, 1960-99. Solid lines are results from the original data and dots from the adjusted data. Critical levels of 90% and 95% are also marked.

3.2 Gradual change in the environment of the measurement site

To illustrate the effect of a change in the environment of the observing site, the case of Lapinjärvi climatological station (60°37'N, 26°09'E) in southern Finland has been examined. The station was operating at the same farmhouse from 1955 to 1996. However, the surroundings of the temperature screen and precipitation gauge altered dramatically due to growth of a spruce fence, apple trees and bushes. There are buildings in the sector from west to north of the measurement site. No new buildings close to the observing site have been constructed. In 1955, the yard was open towards fields, but already in 1968 the fence around the yard is more than 1 metre high (Fig. 5) reaching the level of the precipitation gauge. In 1978 the fence is thick, about 3 metres high and there are high bushes and apple trees in the garden. The wind speed at gauge and screen height must have decreased and the yard has also become more shaded.

The environmental change that took place in Lapinjärvi is not common in the Finnish station network. It was chosen because it is well documented with photographs. Often environmental changes are difficult to notice from the station inspection reports.

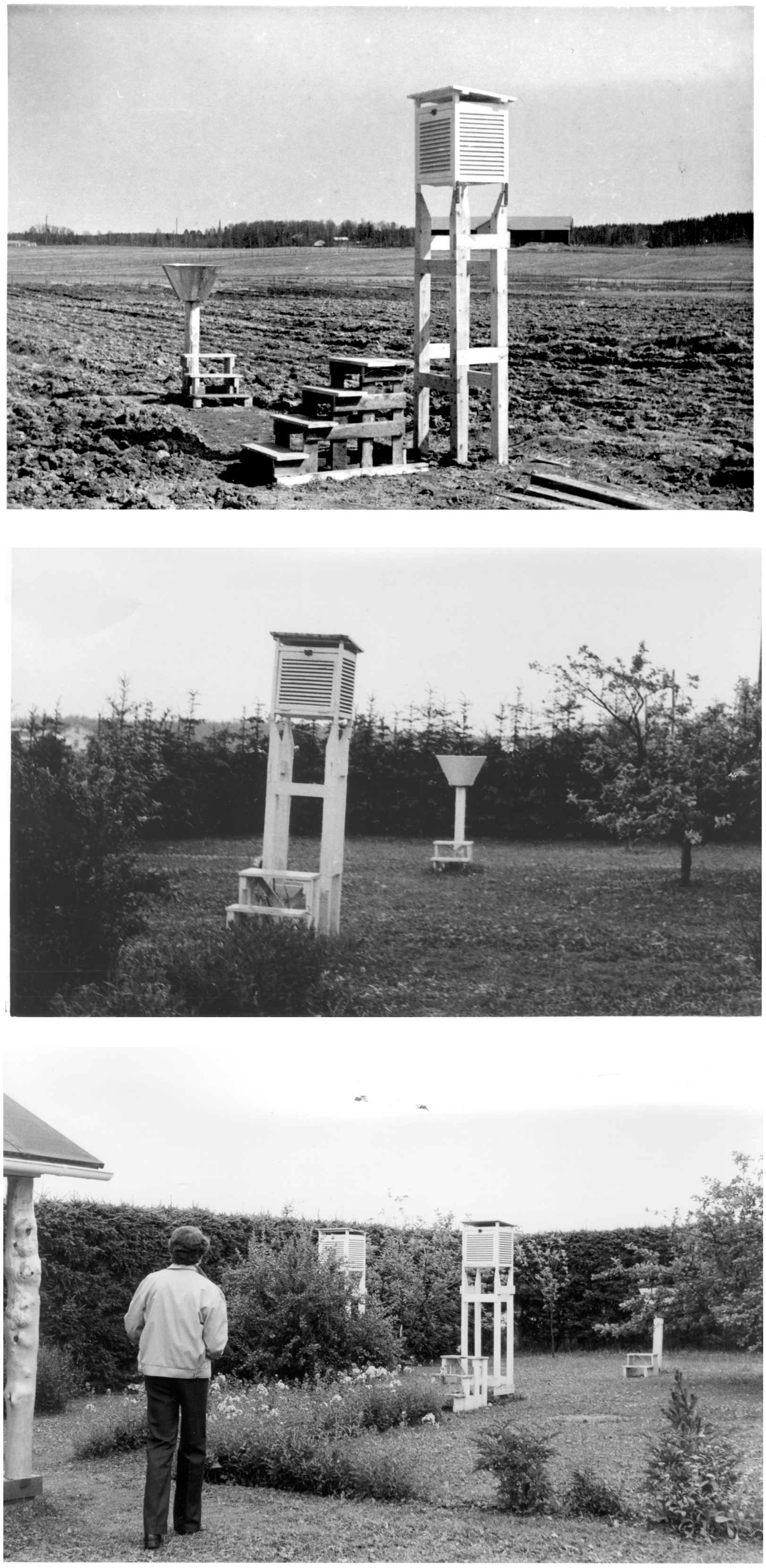

Fig. 5. Photographs from Lapinjärvi climatological station facing East – Southeast on 18.5.1955, 23.6.1968 and 7.8.1978 (starting from the top).

From the available evidence, it seems likely that environmental change has been gradual at Lapinjärvi. Therefore, the trend SNHT should suit the situation. The test value, T^{t}_{max} in (A2), for the annual precipitation series is 12.9 and for the annual mean temperature 16.6 while the 95% significance level is 7.1. From Figure 6 it can be seen that the precipitation Q-values are low in the early years but increase later on, while the opposite occurs for temperature. The trend SNHT for precipitation gives a=1963, b=1968 and $\overline{q_2} / \overline{q_1}$ = 1.15, and for temperature a=1969, b=1974 and $\overline{q_2} - \overline{q_1}$ = -0.22°C.

Vegetation growth has increased the gauge catch by reducing the wind speed near the orifice of the gauge (Fig. 6a). Precipitation measurements are thus more accurate, because the undercatch is reduced. However, at the same time the homogeneity of time series has been broken, with a measured increase in precipitation, especially in winter (by about 40%). The gauge was moved by about 8 meters in 1968, which may also have affected the measurements.

The cooling that also took place (Fig. 6b) occurred later than the precipitation increase. This may be related to the partial shadowing effect of vegetation. The cooling is strongest during the spring and autumn (from 0.3°C to 0.4°C) when the Sun's radiation comes at low angles, and it can be seen in the mean daily maximum temperatures, while the mean daily minimum temperatures are not affected. This might also explain why summer temperatures have remained practically unchanged.

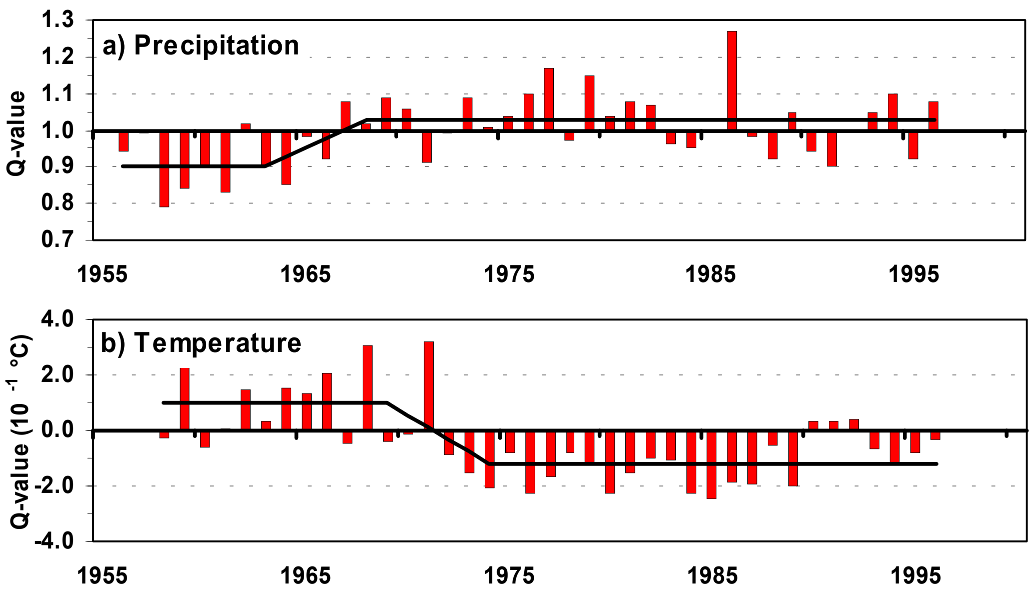

Fig. 6. Trend SNHT results: Q-series (bars) and linear trend fit (solid lines) for a) annual precipitation and b) annual mean temperature at Lapinjärvi, 1956-96.

Unmistakable changes in the vegetation have occurred at the Lapinjärvi station and the SNHT results are mostly physically consistent with them. There are large homogeneity breaks in the precipitation series and small, but detectable, breaks in the mean and maximum temperature series. Nonetheless, because the station history information is not complete, it cannot be ruled out that other factors could also have contributed to the detected inhomogeneities.

The single shift SNHT would have given almost as good a result at Lapinjärvi as the trend SNHT. *AM* (1997) suggest that short trend periods should be handled as abrupt breaks. The length of the trend period, b-a, should be at least five years. In practice, it can be difficult to distinguish between small successive shifts with the same sign and a trend (*Moberg and Alexandersson*, 1997).

3.3 Multiple breaks due to station relocations

The third SNHT example comes from Kuopio climatological station (62°54'N, 27°41'E). The single shift SNHT shows a break in 1922/23 in a series of annual mean daily maximum temperatures (Fig. 7). The shape of the T-series suggests another break around 1950. Because the single shift SNHT is designed to detect only one break, the test is run again for the period 1924-98. Now the break in 1950 becomes the largest ("Step 2", Table 2). The third step is to test the period 1951-98 which, although a high value of T^{s}_{max} is recorded, can still be classified as homogeneous. The high value occurs at the beginning of the test period and can be omitted (*Hanssen-Bauer and Førland*, 1994). In Step 4, the data before 1951 are adjusted (raised by 0.91°C) to make the period 1924-98 homogeneous. In the last two steps, the whole time series is tested. A break in 1923 is detected and it is adjusted by lowering data before 1924 by 1.26°C.

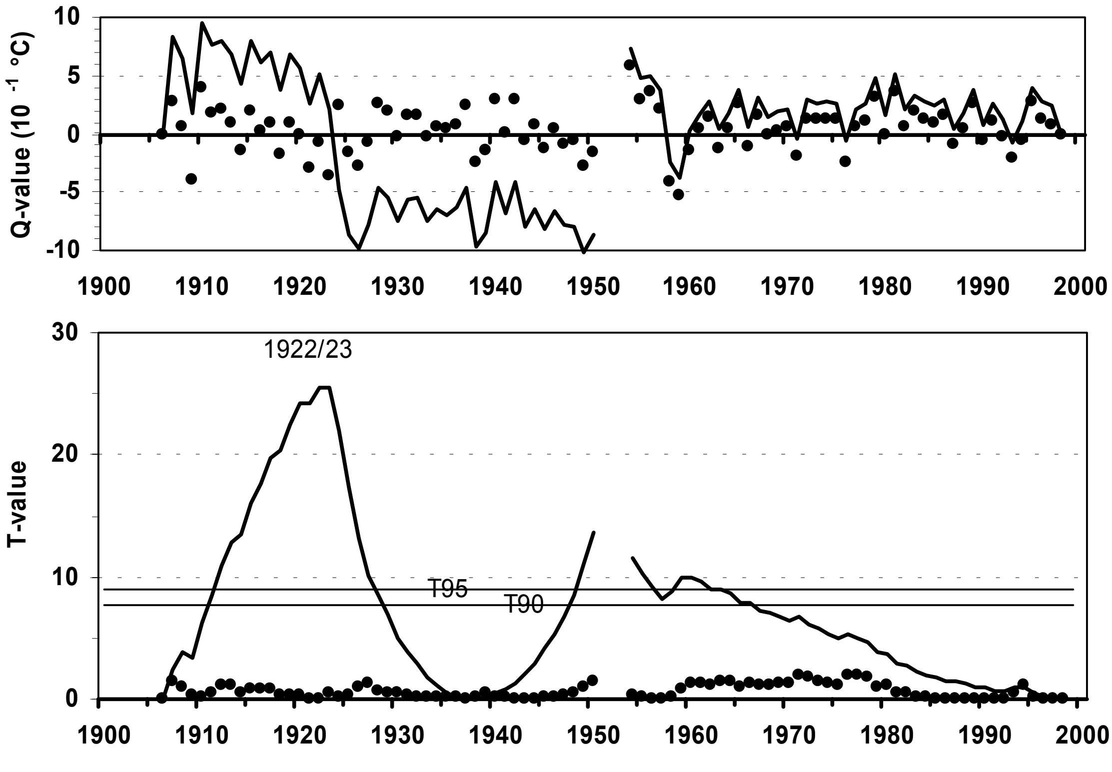

Fig. 7. Values of Q and T from performing the single shift SNHT on the annual mean daily maximum temperature series from Kuopio, 1907-98 (1951-53 Tx data not in digital format). Solid lines are results for the original data and dots for the adjusted data (Step 1 and Step 6 in Table 2, respectively). Critical levels of 90% and 95% are also marked.

If the break in 1922/23 were adjusted in Step 1 using the whole time series, the size of the adjustment would be badly defined. It would also lead to partial masking of the other homogeneity breaks. Even several test steps would not necessarily produce the same adjustments as in Table 2.

The double break SNHT (A1) is quite a powerful test to apply in the Kuopio case. It correctly detects the break years, but underestimates the size of the break in 1923 ("Step 1 DB" in Table 2). The series do not become homogeneous with these adjustments, but require further test steps. However, compared to the use of the single shift SNHT, the double break SNHT was capable of detecting both break years at Step 1, which could reduce the number of test steps.

The physical explanations for these two discontinuities are station relocations. In July 1924 the station was moved up to Puijo hill (about 100 meters above town level) and in June 1951 it was moved back down to the town area. These relocations have severely affected the homogeneity of climatological time series in Kuopio (*Heino,* 1994).

Table 2. Single shift SNHT (Step 1 – Step 6) and double break SNHT (Step 1 DB) results for annual mean daily maximum temperature at Kuopio.

	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**	**Step 1 DB**
Test period	1907-98	1924-98	1951-98	1924-98	1907-98	1907-98	1907-98
$T^{s}_{max}/\mathbf{T_{95}}$	25.4/9.0	59.5/8.8	9.0*/8.24	2.1/8.8	75.1/9.0	2.0/9.0	48.1/15.8
Break year(s), if significant	1922/23	1950			1923		1950 1923
Adjustment Year/size (°C)				1950/0.91	1950/0.91	1950/0.91 1923/-1.26	1950/0.68 1923/-0.75

* T^{s}_{max} at the beginning of series, not significant (see section 2.2)

3.4 Examples of monthly temperature and precipitation adjustments

To conclude this section, some examples of the seasonal distribution of adjustments are given. Some adjustments to data sets are applied only at the annual (e.g. *Young,* 1993; *Hanssen-Bauer and Førland,* 1994) or seasonal (e.g. *Moberg and Alexandersson,* 1997) level, but in Finland adjustments were calculated at a monthly level.

There is often a seasonal distribution in the size of required adjustments, as shown in Figure 8. Monthly temperature adjustments related to the relocation of the Kuopio station from Puijo hill (230 m above sea level) down to the town (several sites 120-100 m above sea level) are shown (Fig 8a-c). There are several factors involved that make the Tx, T and Tn adjustments different. In late spring and summer the average temperature lapse rate between the 1000 hPa to 950 hPa pressure levels is -0.6°C/100m according to radiosonde soundings at Jyväskylä (62°24'N, 25°39'E, about 130 km southwest of Kuopio) (*FMI,* 1994). This roughly explains the adjustments of Tx and T in summer, but Tn is influenced by nighttime inversions, which make lower sites cooler than the hill. In autumn the warming effect of Lake Kallavesi influences the town area, but not so much the top of the hill. In wintertime there are frequently surface inversions, which should make the low-lying town area cooler; on the other hand the town and lake exert their own warming effects.

Figure 8d is an example of opposite mean temperature adjustments between the periods March-June and July-February. The Hailuoto station was moved from the west coast of the island about 10 km to the east (65°02'N, 24°44'E), into the middle parts of the island. The temperature climate of the station became more continental. At an annual level the homogeneity break is almost undetectable -0.18°C. This is a good example of why it is preferable to test temperature series at seasonal level, if possible.

In contrast, air pressure and precipitation adjustments are rarely opposite from season to season in Finland. Precipitation series are much noisier than air pressure or temperature series. Testing of precipitation series can therefore be more effective at an annual level, although adjustments have been calculated for monthly data.

The monthly precipitation adjustments related to the homogeneity break in 1979 at Turku (Fig. 4) are shown in Figure 8e. The adjustments are largest in winter (about 20%), when most of the precipitation is in solid form and is influenced more by aerodynamic errors than liquid precipitation. In summer there are no large differences between the sites. Another example is Vaasa, which is typical of many long-term stations in Finland that have experienced relocation from town to airport (Fig. 8f). Airport sites are open, recording higher wind speeds, which exaggerate the problem of undercatch.

The adjustments shown in Figure 8 exhibit a quite smooth behaviour. In many other cases the adjustments contain more "noise", e.g. due to a low quality reference series or short time period for the determination of adjustments. Smoothing of adjustments with a 3-month running mean has sometimes been used in Finland. Alternatively, *Slonosky et al.* (1999) use Gaussian filtering. However, there can be noticeable differences between the adjustments for consecutive months due to real effects, at least for a small number of years (less than ten or so). For instance, the size of temperature adjustments can be a function of the temperature anomaly in the winter months. In such situations, the required adjustment for "mild Januarys" may be markedly smaller than for "cold Februarys" with surface inversions, and smoothing would bias the adjustments.

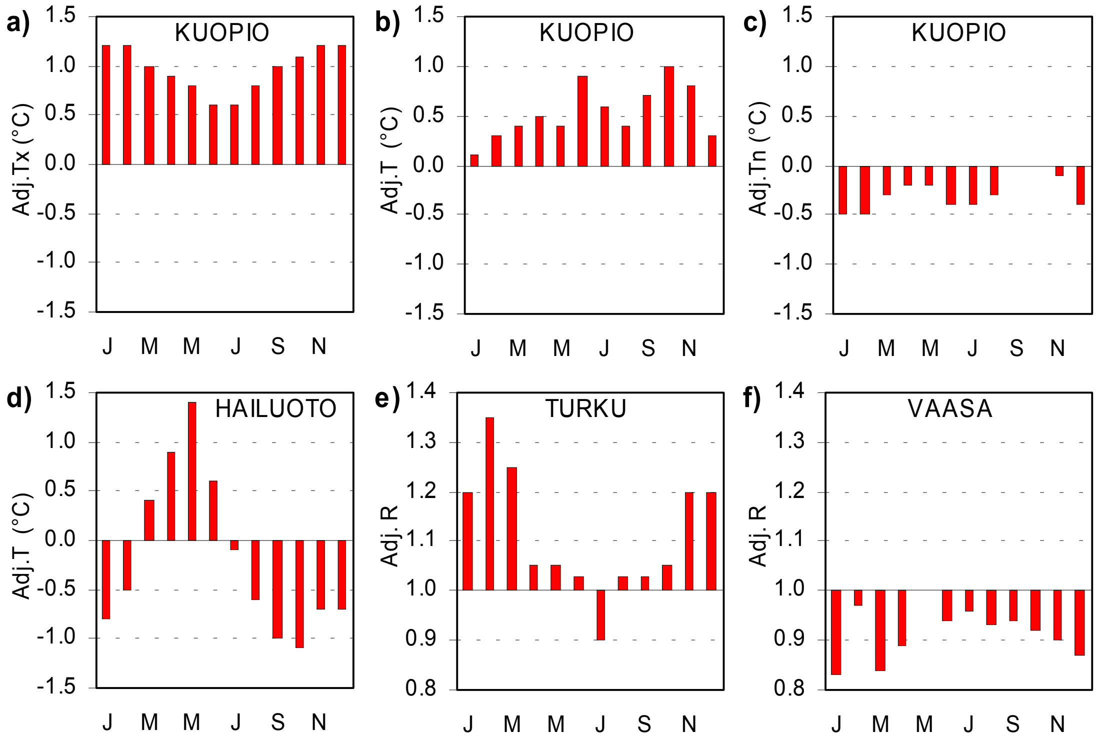

Fig. 8. Monthly adjustments applied to time series of a) mean daily maximum temperature at Kuopio (1951), b) mean temperature at Kuopio (1951), c) mean daily minimum temperature at Kuopio (1951), d) mean temperature at Hailuoto (1950), e) precipitation totals at Turku (1979) and f) precipitation totals at Vaasa (1946).

4. *Application of SNHT in the production of homogeneous data sets*

SNHT provides a means to test and adjust time series to produce relatively homogeneous data sets. However, there are two problems often encountered at the start of testing, which challenge the basic assumptions of SNHT. Firstly, *a priori*, there are no series that can be classified, unequivocally, as homogeneous, so it is impossible to build a perfectly homogeneous reference series. Secondly, SNHT is meant for detecting one (single shift and trend) or two (double shift) homogeneity breaks at a time, but the series may contain multiple non-homogeneities.

The lack of homogeneous reference series is commonly solved in an iterative way, where the homogeneity of all series are gradually improved at each stage of testing. Basically a test session consists of two test rounds. In the first round, all stations are used as candidate stations and they are tested against potentially homogeneous reference stations. For example, at the very first session all stations are reference stations, therefore, some of the detected breaks may not be real but artefacts from non-homogeneous reference series. As a result of the first round, the stations are classified as homogeneous or non-homogeneous with preliminary adjustments. In the second round, the non-homogeneous stations (original series) are tested using the homogeneous stations (or if they are too few, non-homogeneous stations with adjustments are also used) as reference series. As a result of the second round, non-homogeneous stations have more reliable adjustments. To test whether the homogeneity criteria are fulfilled, the first round is then run again. If homogeneity criteria are not met, the test session is repeated as many times as necessary. A lengthy and detailed description of this kind of procedure can be found in *Moberg and Alexandersson* (1997), while diagramatic summaries are presented by *Hanssen-Bauer and Førland* (1994) and *González-Rouco et al.* (2001).

The problem that SNHT is not designed to handle multiple non-homogeneities can be circumvented by investigating the series in periods containing no or only one non-homogeneity. In *Hanssen-Bauer and Førland* (1994), if multiple breaks were detected, a time series was classified as inhomogeneous. Although, this approach is in accordance with SNHT assumptions, it was not followed in Finland because only a few of the long time series would have been classified as homogeneous. Instead, the series were divided into periods containing no or only one non-homogeneity. If available, station history information was first used to determine potential break points; if not, SNHT results were used (see the Kuopio example in section 3.3). Break points close to each other (<5 years) were merged; 5 years was also the limit for omitting SNHT results from both ends of series.

The commonly used statistical significance level for detecting breaks is the 95% level (*Tuomenvirta and Drebs,* 1994; *Moberg and Alexandersson,* 1997; *Slonosky et al.,* 1999), and this has been used in most of this work, too. *Tuomenvirta and Drebs* (1994) imposed an additional requirement for adjusting a series – that each break detected by SNHT should be supported by some physical evidence from station history. However, this proved to be too strict a criterion because of incomplete metadata. Instead, the use of a lower, 90%-significance level in connection with supporting evidence from metadata was adopted, following *Hanssen-Bauer and Førland* (1994).

In their studies, *Peterson and Easterling* (1994) and *Moberg and Alexandersson* (1997) selected as reference stations those stations having the largest correlation coefficients (change-in-time-step series, cf. section 2.3) with the candidate station. In Finland, reference stations were selected manually. A high correlation coefficient was one of the criteria, but the geographical distribution of reference stations and metadata information were also used. For example, a station with several relocations or missing years of data was replaced with a more suitable one if available. If many (>10) good reference stations were available, the test was run twice with different sets of reference series. However, with the early data the problem is more often that there are too few highly correlated reference stations, especially in northern Finland where the station network is relatively sparse.

Only annual values of mean temperature, precipitation and air pressure have been tested in the Finnish data set. Maximum and minimum temperatures have been tested at a seasonal level; testing at a monthly level has too high a noise level. However, adjustments have been calculated on a monthly basis. The time periods to be used in the calculation of adjustment terms is not specified in the SNHT. The length of homogenous periods before and after a break imposes practical limits on the calculation period. The use of a long period (more than 20 years) would normally be recommended. However, sometimes it is wise to use only those periods of a reference series that are known to be of good quality. The exact dates of breaks were taken from metadata, when available. The series were adjusted relative to the most recent observations.

5. *Discussion*

5.1 *The practise and objectives of homogeneity testing*

Maximising the amount and quality of information is a good principle to guide the application of SNHT, or any other homogeneity testing method. Firstly, all relevant data must be in usable (digital) form. Time series from stations which are no longer operating, but have reasonably long series may prove to be useful in testing and adjusting in data sparse regions or time periods. It is advisable to initiate testing in regions and time periods with good data coverage, and for the analysis to "migrate" gradually into regions/periods of data sparsity.

The testing procedure must be designed to best meet the targets set for homogeneity testing within the resources available. For example, determining homogeneous series from a large number of stations, e.g. 10 years of precipitation totals (*Nordlund and Tuomenvirta,* 1998), is a quite different task than trying to adjust over a 200-year long temperature series (*Moberg and Bergström,* 1997). Moreover, the spatial scale of interest determines whether local variations caused by environmental changes are adjusted or not, e.g. urbanisation in Helsinki (*Heino,* 1994) and draining of peatlands (*Venäläinen et al.,* 1999) in Finland.

5.2 *Metadata*

Metadata can be an important source of extra information. Firstly, metadata provide information that cannot be substituted by statistical homogeneity testing. Nationwide methodological and instrumental changes are difficult to detect with relative testing methods and they may cause systematic biases in the data. These kinds of inhomogeneities were adjusted in the Finnish data prior to statistical homogeneity testing (*Heino,* 1994; *Tuomenvirta,* 2001). Moreover "side-by-side" comparison measurements sometimes exist to adjust homogeneity breaks at individual stations or there have been special campaigns to compare instruments (e.g. *Huovila et al.,* 1988). Results from "side-by-side" comparisons should be utilised when available.

Secondly, metadata offers a useful aid for guiding and shortening the process of statistical homogeneity testing. Before testing, station history information can be used to select reference stations that seem homogeneous. In piecewise testing, station histories can be used to define suitable time periods to be tested. Detected breaks can be confirmed from metadata, the reason and exact date of a discontinuity determined and consistency between the adjustment applied and the physical reason for the break established. Metadata can also help to avoid the danger of adjusting time series to be homogeneous relative to each other, but not necessarily correctly describing the climate. An example is a case where there are only two series available and one discontinuity is found. A statistical test would not be able to tell which of the series should be adjusted. In these kinds of situations metadata, if available, can provide invaluable guidance.

Unfortunately, the use of metadata can be problematic. Metadata are usually incomplete and seldom can be fully relied on. Therefore, statistical tests for unknown breaks are needed (though the *t*-test would be a sufficient tool if all breaks were known beforehand). The amount and quality of metadata varies a lot. Important pieces of information may not have been recorded, e.g. measurement site changes. There can be problems in interpretation of metadata, e.g. is a change in geographical co-ordinates due to a station move or to improved surveying. There can be erroneous information included, e.g. instructions for a rain gauge to be moved are recorded, but in reality no action is taken.

The use of metadata can be time consuming and tedious. For global data sets it can be unfeasible due to problems in collection and interpretation of various national documents. Some sources of metadata contain large amounts of irrelevant information, making extracting the important pieces laborious.

5.3 Lack of reference stations

Sometimes there are no highly correlated reference stations available, e.g. on isolated islands, and reference series cannot be built for a given element. In such cases, SNHT has been applied to Q-series that have been formed from different elements at the same station (internal testing). In this way, the testing can be done although neighbouring stations are not available. However, a disturbance affecting all observations, e.g. a site change, may result in the homogeneity break not being detected, or at least the size of it being under- or overestimated. The use of physically different elements (e.g. daytime cloud cover and maximum temperature) may violate the basic assumption of SNHT that the ratios/differences stay constant over a long period. For example, a change in the dominant circulation type may break the relationship between different elements. The best results have been achieved when station history information has been used in the interpretation of test results. The use of other elements has proved useful but it must be applied with care. For example, if Tx and Tn are homogenised relative to each other, by definition, diurnal temperature range (Tx-Tn) cannot change significantly.

Another solution to the problem of having no reference stations is to use a non-parametric test operating solely on the candidate series (absolute testing) to search for break points, trends and other non-random behaviour from the series (e.g. *Sneyers,* 1990; *Sneyers et al.,* 1998). However, some criteria to separate climatic break points from artificial ones must be developed. It is also worth noting that absolute tests can be applied to homogeneity testing of multiple series. The results from several stations are compared and climatic break points can be determined as those breaks that occur at the same time at several stations, leaving the remaining break points as potential inhomogeneities (*Ortiz et al.,* 1999).

5.4 SNHT and new methods

SNHT has performed well in intercomparisons of different homogeneity tests, being ranked the best or one of the best methods (*Easterling and Peterson,* 1992; *Easterling and Peterson,* 1995; *Bosshard,* 1997). However, new homogeneity testing methods employing a different approach and utilising more statistics have recently been developed. *Peterson et al.* (1998) contains a review of both the traditional and newly emerging approaches.

Easterling and Peterson (1995) have developed a method combining regression analysis and non-parametric statistics. It can be fully automated and it has been used to homogenise the Global Historical Climatology Network (GHCN; *Vose et al.,* 1992) temperature time series. For the analysis of temperature and precipitation trends in Canada (*Zhang et al.,* 2000), *Vincent* (1998) developed a technique where four linear regression models are used to determine whether the series are homogeneous or not. A model is finally accepted when the residuals (differences between the values of the candidate series and the fitted values given by the model) are considered to be random variables.

SNHT does not provide any estimate of confidence levels of adjustments. *Szentimrey* (1999) and *Bosshard* (1997) have both elaborated ways to evaluate confidence intervals for the size of adjustments. *Szentimrey* (1999) also calculated confidence intervals for the date of break.

The requirement of statistical tests for a homogeneous reference series can be avoided by assuming that series contain homogeneous sub-periods and that these sub-periods can be used as reference series. All series within the same climatic region are compared to each other. Both *Szentimrey* (1999) and *Mestre* (1999) use this technique. *Mestre* (1999) uses a Bayesian approach for the detection of break points and recommends the use of a double-step model in the detection (see also *Caussinus and Mestre,* 1996). The approach is elegant, but *Slonosky et al.* (1999) found that the method has problems when there were few homogeneous, neighbouring stations. One can use several statistical methods to complement each other, but then decision rules must be developed for the interpretation of test results (*Auer et al.,* 1999; *Begert et al.,* 1999).

5.5 Concluding remarks

SNHT is a practical tool for homogeneity testing. All versions of SNHT give the following information: date of the break, size of the break and significance of the break. SNHT has been in use for more than a decade (e.g. *Alexandersson,* 1986; *Hanssen-Bauer and Førland,* 1994; *Tuomenvirta and Drebs,* 1994; *Quintana-Gomez,* 1995; *Steffensen,* 1996; *Nordli et al.,* 1996; *Moberg and Alexandersson,* 1997; *Slonosky et al.,* 1999; *González-Rouco et al.,* 2001). The single shift test has been especially widely used, and a reasonable amount of experience on the application and capabilities of the technique has been gained and published.

In Finland, SNHT has been used to systematically test temperature, precipitation and pressure series. New test methods could equally well be applied, but it seems that different methods produce fairly similar adjusted series (*Peterson et al.,* 1998; *Slonosky et al.,* 1999) and SNHT has proved to be a practical alternative. One of the most important results of this work is that most of the Finnish long-term stations have been tested. The risk of misinterpreting the record of climate and climate changes due to bad data is now much reduced.

The use of non-homogeneous climatological time series (i.e. containing non-climatic induced variations) can lead to inconsistent conclusions. It can be recommended that, besides routine quality control, the homogeneity of data should be evaluated before performing studies of climatic changes.

Acknowledgements

I would like to thank Drs. Raino Heino and Hans Alexandersson for discussions related to homogeneity problems and Mr. Achim Drebs for his assistance with metadata. Dr. Tim Carter made the linguistic revision and provided many helpful comments. This work was supported by the Academy of Finland through the FIGARE programme and the Yrjö, Kalle and Vilho Väisälä Foundation.

References

Alexandersson, H., 1984. A Homogeneity Test Based on Ratios and Applied to Precipitation. Report 79, Department of Meteorology, Uppsala, Sweden, 55 pp. (PhD dissertation)

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. *J. Climatol*., **6**, 661-675.

Alexandersson, H., 1995. Homogeneity testing, multiple breaks and trends. *Proceedings of the Sixth International Meeting on Statistical Climatology*, Galway, Ireland, 439-441.

Alexandersson, H. and A. Moberg, 1997. Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol.,* **17**, 25-34.

Auer, I., R. Böhm, M. Hagen and W. Schöner, 1999. ALOCLIM Austrian-Central European Long-term Climate – creation of a multiple homogenized long term climate data set. In: Szalai et al. (eds). Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 9-13 November 1998*, WCDMP*-No. **41**, WMO, Geneva, 47-72.

Begert, M., M. Giroud, R. Kegel, G. Seiz, V. Koehli, O. Bochnicek, M. Fukasz, E. Nieplova and L. Sramo, 1999. Operational homogenization of long term climate data series at SMI and SHMI. In: Szalai et al. (eds). Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 9-13 November 1998, *WCDMP*-No. **41**, WMO, Geneva, 177-184.

Bosshard, W., 1997. Homogenisierung klimatologischer Zeitreihen, dargelegt am Beispiel der relativen Sonnenscheindauer. Philosophical Faculty II, University of Zurich, 114 pp. (PhD dissertation with summary in English)

Caussinus, H. and O. Mestre, 1996. New mathematical tools and methodologies for relative homogeneity testing. *Proceedings of the First Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, Hungarian Meteorological Service, 63-82.

Conrad, V. and L.W. Pollack, 1950. *Methods in Climatology*. 2nd edition. Harvard University Press, Cambridge, Massachusetts, 459 pp.

Easterling, D.R. and T.C. Peterson, 1992. Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review. *5th International Meeting on Statistical Climatology*, Toronto, Canada, J28-J32.

Easterling, D.R. and T.C. Peterson, 1995. A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatology*, **15**, 369-377.

FMI, 1994. Radiosonde observations in Finland 1982-1991. *Meteorological Yearbook of Finland*, Vol. **82-91** Part 3, Finnish Meteorological Institute, Helsinki, 182 pp.

Frich, P. (Co-ordinator), H. Alexandersson, J. Ashcroft, B. Dahlström, G. Demarée, A. Drebs, A. van Engelen, E.J. Førland, I. Hanssen-Bauer, R. Heino, T. Jónsson, K. Jonasson, L. Keegan, P.Ø. Nordli, T. Schmith, P. Steffensen, H. Tuomenvirta and O.E. Tveito, 1996. North Atlantic Climatological Dataset (NACD Version 1) - Final Report. Danish Meteorological Institute, *Scientific Report* **96-1**, 47 pp.

Førland, E.J (ed), H. Alexandersson, B. Dahlström, A. Drebs, P. Frich, I. Hanssen-Bauer, R. Heino, J. Helminen, J. Jónsson, Ø. Nordli, T. Palsdottir, T. Schmith, H. Tuomenvirta, O.E. Tveito and H. Vedin, 1998. REWARD: Relating extreme weather to atmospheric circulation using a regionalised dataset. The Final Report. *DNMI Report KLIMA* **17/98**, 23 pp.

González-Rouco, J.F., J.L. Jiménez, V. Quesada and F. Valero, 2001. Quality control and homogeneity of precipitation data in the southwest of Europe. *J. Climate*, **14**, 964-978.

Hansen-Bauer, I. and E.J. Førland, 1994. Homogenizing long Norwegian precipitation series. *J. Climate*, **7**, 1001-1013.

Hawkins, P.M., 1977. Testing a sequence of observations for a shift in random location. *J. Am. Statist. Assoc.,* **73**, 180-185.

Heino, R., 1994: Climate in Finland during the period of meteorological observations. *Finnish Meteorological Institute Contributions*, No. **12**, Finnish Meteorological Institute, 209 pp.

Huovila, S., E. Elomaa, K. Leminen, B. Tammelin and A. Tuominen, 1988. Comparison of snow gauges used in Nordic countries – Contribution of Finland to WMO solid precipitation intercomparison. Part I: System description. Finnish Meteorological Institute, *Meteorological publications,* No. **9**, Helsinki, 61 pp.

Mestre, O., 1999. Step-by-step procedures for choosing a model with change-points. In: Szalai et al. (eds). Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 9-13 November 1998, *WCDMP*-No. **41**, WMO, Geneva, 15-26.

Moberg, A. and H. Alexandersson, 1997. Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *Int. J. Climatol.,* **17**, 35-54.

Moberg, A. and H. Bergström, 1997. Homogenization of Swedish temperature data. Part III: The long temperature records from Uppsala and Stockholm. *Int. J. Climatol.,* **17**, 667-700.

Nordli, P.Ø., I. Hanssen-Bauer and E.J Førland, 1996. Homogeneity analyses of temperature and precipitation series from Svalbard and Jan Mayen. Report No. 16/96 Klima, 41 pp.

Nordlund, G. and H. Tuomenvirta, 1998. Spatial variation in wet deposition amounts of sulphate due to stochastic variations in precipitation amounts. *Atm. Env*., **32**, 2913-2921.

Ortiz, A.G., A.C. Bernabé and C.A. Mata, 1999. Homogeneity analysis of the annual averages of the air temperature of Huesca (Spain). In: Szalai et al. (eds). Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 9-13 November 1998, *WCDMP*-No. **41**, WMO, Geneva, 141-148.

Peterson, T.C. and D.R. Easterling, 1994. Creation of homogeneous composite climatological reference series. *Int. J. Climtol.,* **14**, 671-679.

Peterson T.C., D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullet, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E.J. Førland., I. Hanssen-Bauer, H. Alexandersson, P. Jones and D. Parker, 1998. Homogeneity adjustments of *in situ* atmospheric climate data: A review. *Int. J. Climatol.,* **18**, 1493-1517.

Quintana-Gomez, R.A., 1995. Trends of maximum and minimum temperatures in Northern South America (Venezuela and Columbia) during the period 1918-1990. Homogeneity test. *Proceedings of the Sixth International Meeting on Statistical Climatology*, Galway, Ireland, 459-462.

Schmith, T., H. Alexandersson, K. Iden and H. Tuomenvirta, 1997. North Atlantic-European pressure observations 1868-1995 (WASA dataset version 1.0). Danish Meteorological Institute, *Technical Report* **97-3**, 13 pp.

Slonosky, V.C., P.D. Jones and T.D. Davies, 1999. Homogenization techniques for European monthly mean surface pressure series. *J. Climate*, **12**, 2658-2672.

Sneyers, R., 1990. On statistical analysis of series of observation. Technical note No. 143, WMO No. 415. WMO, Geneva, Switzerland, 192 pp.

Sneyers, R., H. Tuomenvirta and R. Heino, 1998. Observational inhomogeneities and detection of climate change - The case of the Oulu (Finland) air temperature series. *Geophysica*, **34**(3), 159-178.

Steffensen, P., 1996. Standard Normal Homegeneity Test for Windows, User Guide. Danish Meteorological Institute, *Technical Report* **96-13**, 35 pp.

Szentimrey, T., 1999. Multiple analysis of series for homogenization (MASH). In: Szalai et al. (eds). Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 9-13 November 1998, *WCDMP*-No. **41**, WMO, Geneva, 27-46.

Tuomenvirta, H. and A. Drebs, 1994. Homogeneity testing and management of metadata in Finland. In: Heino, R. (ed*.),* Climate Variations in Europe. Proceedings of the European Workshop*,* Kirkkonummi, Finland*. Publications of the Academy of Finland* **3/94**, 321-329.

Tuomenvirta, H. and H. Alexandersson, 1995. Adjustment of apparent changes in variability of temperature time series. *Proceedings of the Sixth International Meeting on Statistical Climatology*, Galway, Ireland, 443-446.

Tuomenvirta, H. and R. Heino, 1996. Climatic changes in Finland - recent findings. *Geophysica*, **32**(1-2), 61-75.

Tuomenvirta, H., H. Alexandersson, A. Drebs, P. Frich and P.O. Nordli, 2000. Trends in Nordic and Arctic temperature extremes and ranges. *J. Climate*, **13**, 977-990.

Tuomenvirta, H., 2001. Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data*. Int. J. Climatol*., **21**, 495-506.

Venäläinen, A., L. Rontu and R. Solantie, 1999. On the influence of peatland draining on local climate, *Boreal Env. Res.*, **4**, 89-100.

Vincent, L.A., 1998. A technique for the identification of inhomegeneities in Canadian temperature series. *J. Climate*, **11**, 1094-1104.

Vose, R.S., R.L. Schmoyer, M. Steurer, T.C. Peterson, R. Heim, T.R. Karl and J. Eischeid, 1992. The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure and station pressure data. ORNL/CDIAC-53, NDP-041. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 315 pp.

Young, K.C., 1993. Detecting and removing inhomogeneities from long-term monthly sea level pressure time series. *J. Climate*, **6**, 1205-1220.

Zhang, X., L.A. Vincent, W.D. Hogg and A. Niitsoo, 2000. Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean*, **38**, 395-429.

APPENDIX

This section summarises the other variants of SNHT, namely, tests for: double shift of the mean level, trend in the mean level, and single shift of the mean level and variance.

Double shift of the mean level

Alexandersson (1995) formulated a test for the double shift of the mean level. The alternative hypothesis is written as

$$H_1 : \begin{cases} Z_i \in N(\mu_1,1) & i \in \{1,...,a\} \\ Z_i \in N(\mu_2,1) & i \in \{a+1,...,b\} \\ Z_i \in N(\mu_3,1) & i \in \{b+1,...,n\} \end{cases}$$

where μ_1, μ_2, and μ_3 refer to the three mean levels. The series are divided into three time periods by years a and b. The test statistic becomes

$$T^d_{max} = \underset{1 \le a < b \le n}{Max} \left\{ a\overline{z_1}^2 + (b-a)\overline{z_2}^2 + (n-b)\overline{z_3}^2 \right\} \tag{A1}$$

where $\overline{z_1}$, $\overline{z_2}$, and z_3 are the mean values at three levels. Analogously SNHT could be developed for a higher number of breaks, though this soon results in a rapid increase in the computing time.

Trend in the mean level

Homogeneity breaks are not always abrupt but may have a trend-like behaviour. For example, the effect of urbanisation on temperatures is often gradual. SNHT can be formulated to include a model where Q-series have a linear trend from year a to b. The alternative hypothesis is written as

$$H_1 : \begin{cases} Z_i \in N(\mu_1,1) & i \in \{1,...,a\} \\ Z_i \in N(\mu_1 + \dfrac{(i-a)(\mu_2-\mu_1)}{(b-a)},1) & i \in \{a+1,...,b\} \\ Z_i \in N(\mu_2,1) & i \in \{b+1,...,n\} \end{cases}$$

where μ_1 and μ_2 are the mean levels at the beginning and end of the series (*AM,* 1997). During years from *a* to *b* the mean level changes linearly from μ_1 to μ_2. The trend may extend throughout the whole series. The test statistic T^t_{max} can be calculated as

$$\begin{aligned} T^t_{max} = \underset{1 \le a < b \le n}{Max} \{ & -a\mu_1^2 + 2a\mu_1\overline{z_1} - \mu_1^2 SB - \mu_2^2 SA + 2\mu_1 SZB + 2\mu_2 SZA \\ & - 2\mu_1\mu_2 SAB - (n-b)\mu_2^2 + 2(n-b)\mu_2\overline{z_2} \} \end{aligned} \tag{A2}$$

where

$$SA = \sum_{i=a+1}^{b} (i-a)^2 / (b-a)^2 \tag{A3a}$$

$$SB = \sum_{i=a+1}^{b} (b-i)^2 / (b-a)^2 \tag{A3b}$$

$$SZA = \sum_{i=a+1}^{b} z_i (i-a) / (b-a) \tag{A3c}$$

$$SZB = \sum_{i=a+1}^{b} z_i (b-i) / (b-a) \tag{A3d}$$

$$SAB = \sum_{i=a+1}^{b} (b-i)(i-a) / (b-a)^2 \tag{A3e}$$

μ_1 and μ_2 can be obtained from

$$\mu_1 = \frac{a\overline{z_1} + SZB - SL \cdot SAB}{a + SB + SK \cdot SAB} \tag{A4}$$

$$\begin{aligned} \mu_2 &= \mu_1 SK + SL \\ &= \mu_1 \frac{(-SAB)}{SA + n - b} + \frac{(n-b)\overline{z_2} + SZA}{SA + n - b} \end{aligned} \tag{A5}$$

where mean levels before and after the trend section are $\overline{z_1}$ and $\overline{z_2}$, respectively. If b=a+1, then trend SNHT is equal to the single shift case.

Single shift of the mean level and variance

There are also situations when the variance of the series changes abruptly. Usually this coincides with a change in the mean level. For example, if a station is moved from the coastline to inland, the mean level as well as the variance of temperature may change. If two different standard deviations σ_1 and σ_2 are used, the alternative hypothesis becomes

$$H_1 : \begin{cases} Z_i \in N(\mu_1, \sigma_1) & i \in \{1, ..., a\} \\ Z_i \in N(\mu_2, \sigma_2) & i \in \{a+1, ..., n\} \end{cases}$$

allowing both the mean level and the variance to change (*AM,* 1997). The test parameter and the critical levels are different from the case where only the mean value changes. The test parameter, T^{s2}_{max}, is

$$T^{s2}_{max} = \underset{2 \le a \le n-2}{\mathrm{Max}} \{-2a \ln\sigma_1 - 2(n-a)\ln\sigma_2 - 1\} \qquad \text{(A6)}$$

where

$$\sigma_1 = \sqrt{\frac{\sum_{i=1}^{a} z_i^2 - (\sum_{i=1}^{a} z_i)^2 / a}{a}} \qquad \text{(A7)}$$

and

$$\sigma_2 = \sqrt{\frac{\sum_{i=a+1}^{n} z_i^2 - (\sum_{i=a+1}^{n} z_i)^2 / (n-a)}{(n-a)}} \qquad \text{(A8)}$$

Theoretically this version is more multipurpose than the single shift (2) because it looks for the change in variance, too. However, as stated in *AM* (1997), this version is oversensitive at the beginning and end of a series. A few values with a small or high variance can result in a high value of T^{s2}_{max} in (A6), making the test quite unreliable near both ends of the series.