VALIDITY OF HOMOGENEITY TESTS FOR METEOROLOGICAL TIME SERIES DATA: A SIMULATION STUDY* (LA VALIDITÉ DES TESTS HOMOGÉNÉITÉ MÉTÉOROLOGIQUE POUR LES SÉRIES CHRONOLOGIQUES: UNE ÉTUDE DE SIMULATION)

Yozgatligil, Ceylan*
E-mail: *ceylan@metu.edu.tr*

Yazici, Ceyda*
E-mail: *cyazici@metu.edu.tr*

Purutcuoglu, Vilda*
E-mail: *vpurutcu@metu.edu.tr*

Batmaz, Inci*
E-mail: *ibatmaz@metu.edu.tr*

**Middle East Technical University*
Department of Statistics
Ankara (06800) TURKEY

Climate change and related problems have a serious impact on the nature. Analyzing, modeling and forecasting meteorological variables are crucial issues to prevent disasters. To be able to obtain reliable results, quality of meteorological data is very important. One of the important quality control methods in meteorological data is the test of homogeneity. If there are changes in the series due to non-climate reasons, they are known as inhomogeneous. Non-climatic factors may hide the true climatic pattern, and then, analysis of climate and hydrology may give bias results (Costa & Soares, 2009). Therefore, temporal homogeneity of series is necessary in climatological research.

There are several causes of non-homogeneity such as abrupt discontinuities, gradual or instant changes or changes in the variability. Changes in the location of the station, in the instrumentation or in the calculations of averages are known as abrupt discontinuities (WMO, 1996). Gradual change can be the result of change in the surroundings of the station, urbanization, or changes in the instrumental characteristics. In anyway, inhomogeneity of the series must be detected and corrected before analyzing the data.

In the literature, there are two groups of homogeneity tests: the ones considering within homogeneity of the series (Gokturk et al., 2008; Tayanc et al., 1998 and Turkes, 2010), and considering the homogeneity of the series using the relationship between neighbor stations (Alexandersson, 1986; Buishand, 1982 and Yozgatligil et al., 2011). In this study, we only focus on the former one. Many of the previous studies on homogeneity uses the tests developed for independent data. Actually, meteorological series are

*This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.

time series data, and thus, have autocorrelation. Therefore, the strength of the homogeneity tests used in the literature is doubtful. In this study, we aim to evaluate the validity of the homogeneity tests used in the literature on time series data such as Kruskal-Wallis (KW), and compare them with the ones that consider data dependency such as Friedman test and KPSS stationarity test using Monte Carlo simulation technique.

After the introduction to the tests that have been used in this study, the conducted simulation study will be explained in detail. Then, concluding remarks and future work will be presented.

Homogeneity Tests within Series

In the literature, KW Test is one of the mostly applied tests for checking the within homogeneity of meteorological data. However this test assumes independent data. So, it is not suitable for testing the mean differences within time series.

The KW test (Kruskal, 1952; Kruskal & Wallis, 1952) is a well known nonparametric test used to compare two or more independent groups of sampled data. This test is an alternative to the one-way Analysis of Variances (ANOVA) test for comparing the means of independent groups, when the assumptions of the test are not met. One of the assumptions of the KW test is that the observations are drawn randomly and independently from their respective populations. In the computations, we initially combine all series as a single dataset, and then, find the rank score for each of the small series separately from the rank order of the pooled data. Finally, the test statistics H is found by

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \tag{1}$$

where N is the total sample size in the pooled data, n_i and R_i are the sample size and the rank of the i th series, respectively. In the application of this test, a series is divided into several groups covering seven or ten years in each one. Therefore, there is dependence within and between groups. The KW test gives little information about the probable date for a shift in the median, and no information about the magnitude of the break.

KW test does not take into account the autocorrelation in the series so that we want to look at the performance of the Friedman Test which is a nonparametric version of the repeated ANOVA (RANOVA). Here, we created two different data sets: first is the original one as a vector, and the other is the transpose of the original one. By taking the transpose, we reduce the dependency within each group. Then, we compute the sum of ranks within each column by the following expression

$$F = \frac{12}{bt(t+1)} \sum_{i=1}^{t} R_i^2 - 3b(t+1) \tag{2}$$

where R_i denotes the underlying rank for the i th series ($i = 1,\ldots,t$), b and t indicate the number of columns and rows of the data matrix, respectively. For the climatological dataset, b represents the number of stations where the data are collected, and t denotes as the number of time points. Critical values of the test can be found in Sprent & Smeeton (2007).

*This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.

Above tests are not taken the autocorrelation structure of the meteorological data into account to test the homogeneity within the series (Yazici et al., 2011). Stationarity tests which are developed specifically for time series data can also be applied to detect the existence of inhomogeneity within the series. One of the well-known stationarity tests, KPSS test (Kwiatkowski et al., 1992), is used for testing stationarity. One may want to combine this with the structural break tests like Chow test (Chow, 1960) to define the exact year at which the inhomogeneity occurs. However, we need to build a linear model for these kinds of tests. KPSS test is conducted by regressing Y_t on a constant, and a trend, and constructing the ordinary least squares (OLS) residuals $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$, obtaining the partial sum of the residuals as $S_t = \sum \varepsilon_i$. Then, the KPSS test statistic is calculated by

$$KPSS = n^{-2} \sum_{t=1}^{n} \frac{S_t}{\hat{\sigma}^2} \quad (3)$$

where $\hat{\sigma}^2$ is the estimate of the long-run variance of residuals. If the series is not stationary, then we have to reject the null hypothesis and conclude that the series is not homogeneous. Asymptotic distribution of the test statistic uses the standard Brownian bridge. As far as the authors' knowledge, in the literature, this is the first study, where Friedman and KPSS tests are considered to detect the inhomogeneity in the series.

Simulation Study

In the simulation study, we have generated data to represent temperature series which is similar to the series in Turkey. A time series model is developed for Turkish monthly average temperature data from 1950-2006; then, monthly temperature data are generated using the following model with seasonal dummies

$$Y_t = 14 + S_{it} * I_t + \varepsilon_t ; t = 1,2,\cdots,n; i = 1,2,\cdots,12 \quad (4)$$

where

$$S_{1t} = -7.3, S_{2t} = -7.55, S_{3t} = -6.3, S_{4t} = -3.0, S_{5t} = 1.2, S_{6t} = 5.8$$
$$S_{7t} = 8.5, S_{8t} = 8.4, S_{9t} = 5.2, S_{10t} = 1.5, S_{11t} = -1.8 \; and \; S_{12t} = -5.1$$

and I_t is a dummy variable for months. Also, ε_t 's are generated from the Normal distribution with mean zero and three different variance values 1, 9 and 25. Hence, we could be able to compare the performances of the tests when we have low and high variability in the temperature series. We have generated 720 time points (60 years) in the series. Then, yearly series are obtained by aggregating the monthly values.

Several scenarios have been considered as the occurrence of the inhomogeneity: mean shift at the beginning, at the middle and at the end of the series for the change in the location of the station, gradual change for the instrument change case, sharp decrease (–20 C^o from the yearly aggregates) at one time point for the sudden change and trend for the change in the environment like urbanization. Gradual change has been created by adding $5 + 1/(1+i), i = 1,2,\cdots,6$ to the observations at the beginning, middle and end of the series considering the effect as six years, and trend has been created as

*This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.

$Y_t + \alpha t; \alpha = 0.01, 0.04; t = 1, 2, \cdots, 60.$ Each series simulated 10,000 times, and in each simulation, KW, Friedman, transposed Friedman (Friedman^T), and KPSS tests have been applied to the series. Then, we counted how many times these tests catch the artificially created inhomogeneity correctly, and how many times the tests find inhomogeneity although the series is homogeneous. In the literature, KW test is applied to series after it is divided to 7-year or 10-year periods. Then, the averages of these subseries are compared for the possible mean shift in the series. Because of this, we have applied the KW test using the 10 year periods.

Findings and Conclusion

Simulation results given in the Tables 1 to 8 indicate that all applied tests reduce their performances with an increase in the variance. Column with Y_t gives the results for the original data, and the next column gives the results after the applications of the scenarios. The results for the original series are the close values to the Type I error probability of 5% but KPSS test shows superior performance. Since the conclusions for the inhomogeneity applied from the beginning, middle and end of the series are similar, we have presented only one of them. Only the results for mean shift are different. When the shift is at the beginning of the series, the chance of catching it is low. KW test gives better results than the Friedman test although independence assumption of data for the KW test is not validated. Transposed Friedman demonstrates the worst results. This test should not be considered as a homogeneity test. KPSS test gives slightly better results than the others to detect the mean shift. KW test performs vaguely better to detect gradual changes for low variance. Tests catch the mean shift especially high mean shifts well but they fail to identify other inhomogeneity scenarios. Also, although results are not given here, we can say that KW and Friedman tests detect wrong points as a break point sometimes. Hence, KW may not be a reliable test to detect the homogeneity, especially if there are no metadata. KPSS test gives information whether the series homogeneous or not but it does not show the starting point of the inhomogeneity. Hence, these tests are not enough to say the series is homogeneous. One should consider other homogeneity tests but they also suffer from the dependency of meteorological data. As a future study, the performance of other homogeneity tests using the neighbor stations and control charts discussed in Yazici at al. (2011) will be considered by a Monte Carlo simulation study.

Table 1 Percentage of non-homogeneous results when the mean shift of 6 unit increase is used at the beginning of the series

	$\sigma^2 = 1$		$\sigma^2 = 9$		$\sigma^2 = 25$	
	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$
Friedman	0.0437	0.2731	0.0371	0.2076	0.0444	0.0542
Friedman^T	0.0345	0.0350	0.0374	0.0365	0.0349	0.0344
KW	0.0459	0.3212	0.0398	0.2426	0.0442	0.0554
KPSS	0.0501	0.3860	0.0438	0.4028	0.0470	0.0598

∗This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.

Table 2 Percentage of non-homogeneous results when the mean shift of 6 unit increase is used at the middle of the series

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$
Friedman	0.0414	0.0529	0.0447	0.2117	0.0425	0.0915
Friedman^T	0.0310	0.0318	0.0360	0.0375	0.0363	0.0364
KW	0.0412	0.0534	0.0462	0.2393	0.0427	0.0960
KPSS	0.0442	0.0590	0.0468	0.3949	0.0442	0.1697

Table 3 Percentage of non-homogeneous results when the mean shift of 24 unit increase is used at the beginning of the series

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$
Friedman	0.0420	0.4941	0.0425	0.3841	0.0416	0.2037
Friedman^T	0.0351	0.0324	0.0346	0.0334	0.0350	0.0348
KW	0.0416	0.5881	0.0411	0.4583	0.0404	0.2409
KPSS	0.0458	0.9999	0.0489	0.5972	0.0444	0.2727

Table 4 Percentage of non-homogeneous results when the mean shift of 24 unit increase is used at the middle of the series

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$	Y_t	$Y_t, shift$
Friedman	0.0406	1.0000	0.0461	0.9999	0.0413	0.9043
Friedman^T	0.0336	0.0343	0.0373	0.0371	0.0368	0.0370
KW	0.0446	1.0000	0.0454	1.0000	0.0388	0.9528
KPSS	0.0468	1.0000	0.0496	1.0000	0.0445	0.9852

Table 5 Percentage of non-homogeneous results when the gradual increase is used at the middle of the series

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	$Y_{t,gradual}$	Y_t	$Y_{t,gradual}$	Y_t	$Y_{t,gradual}$
Friedman	0.0401	0.1276	0.0399	0.0524	0.0419	0.0454
Friedman^T	0.0344	0.2092	0.0368	0.0355	0.0377	0.0368
KW	0.0399	0.1514	0.0424	0.0542	0.0391	0.0440
KPSS	0.0471	0.0964	0.0442	0.0518	0.0442	0.0494

Table 6 Percentage of non-homogeneous results when the sudden decrease is used at the end of the series

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	$Y_{25}-20$	Y_t	$Y_{25}-20$	Y_t	$Y_{25}-20$
Friedman	0.0425	0.0420	0.0433	0.0452	0.0427	0.0428
Friedman^T	0.0348	0.0329	0.0340	0.0365	0.0371	0.0384
KW	0.0424	0.0439	0.0447	0.0478	0.0406	0.0405
KPSS	0.0446	0.0152	0.0462	0.0415	0.0429	0.0418

*This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.

Table 7 Percentage of non-homogeneous results when the increasing trend with slope 0.01 is applied

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	Slope=0.01	Y_t	Slope=0.01	Y_t	Slope=0.01
Friedman	0.0451	0.0487	0.0460	0.0460	0.0403	0.0406
Friedman^T	0.0342	0.0334	0.0369	0.0370	0.0349	0.0347
KW	0.0447	0.0474	0.0433	0.0440	0.0405	0.0405
KPSS	0.0453	0.0600	0.0467	0.0479	0.0435	0.0429

Table 8 Percentage of non-homogeneous results when the increasing trend with slope 0.04 is applied

	$\sigma^2=1$		$\sigma^2=9$		$\sigma^2=25$	
	Y_t	Slope=0.04	Y_t	Slope=0.04	Y_t	Slope=0.04
Friedman	0.0424	0.1218	0.0407	0.0499	0.0448	0.0474
Friedman^T	0.0353	0.0364	0.0325	0.0329	0.0355	0.0349
KW	0.0419	0.1390	0.0395	0.0490	0.0420	0.0451
KPSS	0.0491	0.3121	0.0461	0.0725	0.0480	0.0571

REFERENCES (RÉFERENCES)

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Int. J. Climatology,* **6**, 661–675.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** (3): 591–605.

Buishand T. A. (1982). Some methods for testing the homogeneity of rainfall records. *J. Hydrology,* **58**, 11–27.

Costa, A.C. and Soares, A. (2009). Homogenization of climate data: Review and new perspectives using geostatistics. *Math. Geosci.,* **41**, 291–305.

Gokturk, O. M., Bozkurt, D., Lutfi, O. & Karaca, M. (2008). Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes,* **22**, 3210 - 3218.

Kwiatkowski, D., Phillips P.C.B., Schmidt, P. & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics,* **4**(1-3), 159-178.

Kruskal, W.H. (1952). A nonparametric test for the several sample problem. *Ann. Math. Stat.,* **23**, 525–540.

Kruskal, W.H. & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.,* **47**, 583–621.

Sprent, P. & Smeeton, N.C. (2007). *Applied nonparametric statistical methods*, 4th ed. Chapman and Hall/CRC.

Tayanc, M., Dalfes, N., Karaca, M. & Yenigun, O. (1998). A comparative assessment of different methods for detecting inhomogeneties in Turkish temperature data set. *Int. J. Climatology,* **18**: 561–578.

Türkeş, M. (2010). *Klimatoloji ve meteoroloji*, Istanbul: Kriter Publishing.

Meteorological Organization World (WMO) (1996). *Guide to meteorological instruments and methods of observation*, Part III – Chapter 3 Quality Management WMO-No. 8, Geneva.

Yazici, C., Purutcuoglu, V., Yozgatligil, C., Batmaz, I., Bayramoglu, K. & Iyigun, C. (2011). Homogeneity analysis for dependent climate data. Presented at *7th International Statistical Conference ISTKON*, April 28-May1, Antalya, Turkey.

Yozgatlıgil, C., Purutçuoğlu, V., Yazıcı, C. & Batmaz, I. (2010). Plausibility of the SNHT on Turkish precipitation data. *7th National Symposium on Statistical Days*, 32-34, June 28 - 30, Ankara, Turkey.

∗This study is supported by Middle East Technical University under the contract number BAP-2008-01-09-02.