Geophysical Research Abstracts,
Vol. 11, EGU2009-13763, 2009
EGU General Assembly 2009


# Linear regression significance: from statistics to climatology

A. Douguédroit and S. Bridier

Institut de Géographie, UMR 6012 ESPACE, 29 Av. R. Schuman, 13621 Aix en Provence, France
(Annick.Douguedroit@univ-provence.fr)

The question is to know which interest possesses for the climatology the significance of a linear regression defined from the statistical point of view. The study of the evolution, between 1951 and 2007, of annual minimal, maximal and average temperatures of French Mediterranean stations gives a first answer.

The linear regressions of all the series considered as homogeneous have been calculated. The significance of the linear regressions at the 5% level have been checked by the application of the test of Student (u-Test). The threshold for each number of pairs retained by the test is based on a low coefficient of correlation ($r$), giving a weak coefficient of determination ($r^2$) that represents the explained variance of the concerned temperatures.

Here, 57 (from 1951 to 2007) $-2 = 55$ pairs have a significant regression at the 5% level for $r = 0.26$ which corresponds to $r^2 = 0.07$. This very weak, representing less than 10% of the total variance. More than 90% is not represented by the regression. Such regressions, significant from the statistical point of view, are from a climatic viewpoint one only weakly representative.

Among the 111 series of the 37 French stations retained, only three belonging to the same station have non-significant linear regressions representing their evolution from 1951 to 2007. The $r$ of the other series range from 0.27 to 0.84 and $r^2$ from 0.07 to 0.71. From the climatic point of view they represent very different evolutions which have been investigated by studying the variation of the annual explained variances throughout the series. In each series the length of successive annual values beginning in 1950 or ending in 2007 (or before for a few series ending before 2007) represented by the highest $r$ and $r^2$ have been determined. It corresponds to the total period for a very few series only. Most of the series consisted, as a general rule, of two successive periods; the first one, often characterized by a not-significant, decreasing regression, is followed by a second one with an increasing regression with high explained variance representing the current warming. The reserved criterion for the climatic point of view is then the one of the level of the explained variance. It is, within regressions significant from the statistical point of view, to research the part of the series with highest represented variance (that is, the highest $r^2$) and to appreciate the part of the variance represented in the $r^2$. It allows, in that case, to determine the date of the beginning of the current warming.

Results obtained for annual minimal, maximal and average temperatures from 1951 to 2007 have been collected for each station, and groups have been formed with stations of similar evolutions. Groups are strongly related to the evolution of the environment of the stations: (1) the country, (2) the suburbs and (3) the city centre.