SOFTWARE FOR WEATHER DATABASES MANAGEMENT AND CONSTRUCTION OF REFERENCE YEARS

Marco Beccali[1], Ilaria Bertini[2], Giuseppina Ciulla[1], Biagio Di Pietra[2], and Valerio Lo Brano[1]
[1]Department of Energy, University of Palermo, Palermo, Italy
[2]ENEA, UTEE-GED, Rome, Italy

ABSTRACT

The purpose of this paper is to illustrate a procedure that permits, starting from a sufficiently long database of time series, the construction of a Reference Year (RY) of hourly weather data according to the rules of ISO 15927-4 standard. In order to facilitate the management of the weather database and to allow the users to easily generate the file, an algorithm has been implemented in Microsoft Visual Basic for Application (VBA). By this way, the application of the ISO 15927-4 is possible even using a popular software tool such as Microsoft Excel or Access without any other expensive specialized software. Such tool allows to fulfil all the procedures mentioned in ISO 15927-4 giving as result a time series of 8760 values of several weather variables, ready to be used in any software for energy simulation of buildings.

INTRODUCTION

A sustainable building design approach involves architects and engineers to perform detailed analyses of all the phenomena that influence the thermal behaviour of buildings in order to correctly assess the energy needs. Being the building behaviour strictly linked to the latitude of the site, the knowledge of correct and well organised meteorological data is a crucial issue (Vertačnik, 2008). For this scope, the use of long periods (at least ten years but preferably more) of hourly meteorological data is preferred where possible..

The RY is usually made of up of 12 representative months created from data of a long period of observation. The international standard ISO 15927-4 (UNI EN ISO 15927-4) indicates the procedure for the construction a RY of weather parameters consisting of 8760 hourly values, starting from a consistent database of monitored data.

Although the international standard ISO 15927-4 is clear in the definition of the operations to perform for the manipulation of the data, of course does not indicates the exact algorithms or procedures to be adopted in the calculation and sometimes simply indicates the targets to be achieved without any tips on how to reach them.

The paper illustrates a procedure that permits, starting from a data set of time series, the construction of a reference year of hourly weather data according to the recomendations of ISO 15927-4 standard. The algorithms and the subroutines presented in this work clearly indicate the procedures and the mathematical methods to be used for the application of the technical standard ISO 15927-4. The simplicity and the immediacy of the programming language also permit an easy implementation of the code into other more complex software and allow a fast update or improvement.

THE ISO 15927-4 METHODOLOGY

The International Standard ISO 15927 "Hygrothermal performance of buildings-Calculation and presentation of climatic data", Part 4 "Hourly data for assessing the annual energy use for heating and cooling" specifies a method for constructing a reference year (RY) of hourly values of appropriate meteorological data suitable for assessing the average annual energy for heating and cooling. Other RYs representing average conditions can be constructed for special purposes. The procedures in ISO 15927-4 are not suitable for extreme or semi-extreme years for simulation and for i.e. moisture damage or energy demand in cold years.

The ISO 15927-4 clarifies that a correct simulation of the humidity of a building does not only depend by an appropriate average value of weather parameters, but also by their frequency distributions and by the correlations between them. The purpose of the methodology is to obtain a series of 8760 hourly values of meteorological characteristics derived from a long data set (at least 10 years).

The goal is pursued through a series of statistical analyses, starting from a "long period" (LP) year, obtained as arithmetic mean of daily values of all parameters for all the available years.

Such LP year is the benchmark for the selection of the months that will be part of the RY. The selection of the most representative months is made by using the Finkelstein-Schafer (FS) statistic. After identifying the three months characterised by the lowest value of FS (for each months of the

calendar), the difference between the wind speed monthly average and the wind speed monthly average for the Long Period year is calculated. The month the will be selected in order to be included in the RY as the one with the lowest difference between these two values.

In the following paragraphs, the procedures that have been used to build a RY through a simple tool are described.

Required parameters

Hourly weather data from which will be extrapolated the RY must contain at least the following meteorological parameters:

- dry bulb air temperature;
- global radiation on a horizontal surface;
- relative humidity;
- dew point temperature;
- atmospheric pressure;
- wind speed.

Since the temperature, radiation and relative humidity are key factors for the calculation of heating and cooling demand of buildings, these will be taken as fundamental parameters for the construction of the RY.

Outline of the ISO 15927-4 methodology

For each chosen climatic parameters p it is necessary:

- to calculate the daily means $\overline{p}$ from at least 10 years of hourly values,
- to calculate the cumulative distribution function of the daily means over all years in the data set $\Phi(p,m,i)$ for each calendar month, by sorting all the values in increasing order and then using Eqn (1)

$$\Phi(p,m,i) = \frac{K(i)}{N+1} \qquad (1)$$

where $K(i)$ is the rank order of the ith value of the daily means within that calendar month in the whole dataset, N is the number of days in any calendar month in the whole data set, p is the climate parameter, m is the month of the year;

- to calculate, for each year y of the data set, the cumulative distribution function of the daily means within each calendar month, $F(p,y,m,i)$ by sorting all the values for that month and that year in increasing order and then using Eqn. (2)

$$F(p,y,m,i) = \frac{J(i)}{n+1} \qquad (2)$$

- where $J(i)$ is the rank order of the ith value of the daily means within that months and that year and n is the number of days in an individual month;
- to calculate, for each calendar month, the Finkelstein-Schafer statistic $F_S(p,y,m)$ for each year of the data set using Eqn. (3)

$$F_S(p,y,m) = \sum_{i=1}^{n} \left| F(p,y,m,i) - \Phi(p,m,i) \right| \ (3)$$

- to rank, for each calendar month, the individual months from the multiyear record in order of increasing size of $F_S(p,y,m)$
- to add, for each calendar month and for each year the separate ranks for the three climate parameters.
- to calculate, for each calendar month, for the three months with the lowest total ranking, the deviation of the monthly mean wind speed from the corresponding multi-year calendar-month mean. The month with the lowest deviation in wind speed is selected as the "best" month to be included in the RY.
- finally, to make some corrections in order to fit and adjust the last eight hours each month and the first eight hours of the next month. These changes are made through interpolations and adjustments to ensure that there are no abrupt transitions between the "best" selected months that are joined together. These adjustments include the last eight hours of December and the first eight hours of January so that the RY can be repeatedly used in simulations.

Following this methodology, the RY is provided in the same format of the used data set and consists of 8760 hourly values of climatic parameters selected as primary (air temperature, relative humidity, solar radiation) and secondary (wind speed).

APPLICATION OF THE PROCEDURE

The original data must be pasted in the worksheet in chronological order, as shown in the following Figure 1.
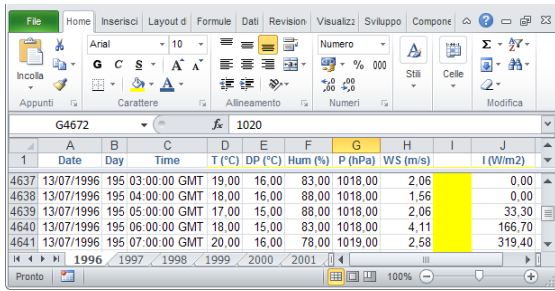
Figure 1: Monitored data in a worksheet

Different years of monitored data can be inserted in different tabs. Before starting any type of data processing, the standard requires a review of the "quality" of data in order to have a clear understanding of the possible anomalies that may be present. To this end, a manual control of the data should be performed in this way:

- Count the number of available data for the calculation for every single year;
- Eliminate possible outliers.

Control Panel

To perform all the previously described data handling operations authors have developed a software tool in a Visual Basic for Application (VBA) macro (pieces of code that can automatically run any Excel command) nested into an Excel worksheet. No special math library is required. A very simple control panel was designed to permit the user to manage the generation of the RY.
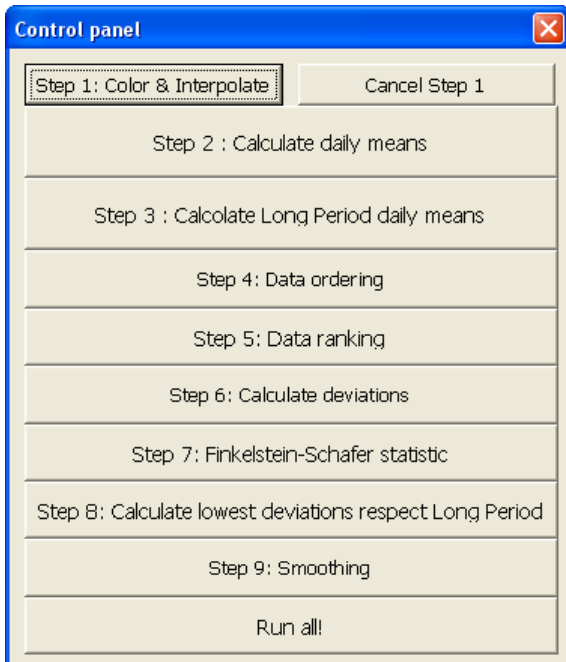

Figure 2: Control panel of the software

Step 1: Color & Interpolate

The first command button performs the following operations:

1. find and fill all the empty cells with yellow;

2. find all the empty cells ranges with an extension ≤ eleven cells;

3. make a linear interpolation that permits to fill the missing data in all void cells ranges that are no more than eleven cells

4. mark with blue color interpolated values for easy recognition.

The data intervals larger than eleven cells will remain void. By these operations user can fix almost all of the deficiencies present in the dataset. Moreover, all the missing or previously removed values (i.e. because unrealistic) are replaced by, interpolated values.
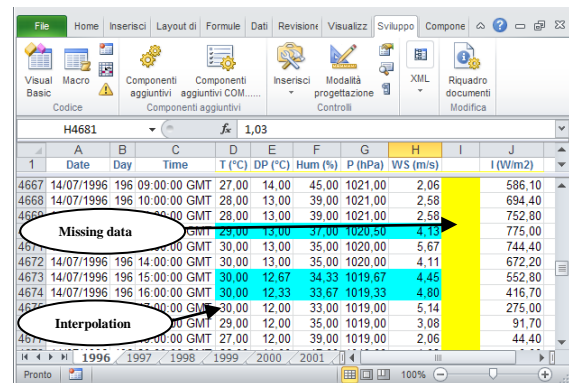

Figure 3: The worksheet after the step 1.

Step 2: Calculate daily averages

The second step of calculation consists of the calculation of the average daily values of all climatic parameters for all the years. This will permit the subsequent calculation of the LP averages. The LP average is the arithmetic average of all daily averages of the climatic parameters and represents the normal figure of temperature, relative humidity, solar radiation and wind speed seen over the years under review.

For each year of data, previously corrected and interpolated, the daily average of all climatic parameters is calculated as the mean of 24 hourly values; the VBA function used to perform the mathematical operation is "*WorksheetFunction.Average*"

Step 3: Calculate Long Period daily averages

In this step, the algorithm automatically calculates the LP daily average: one year consisting in 365 figures representing their daily average over the whole period of calculation. At the end of the calculation, the VBA procedure automatically creates a file called "MedieGiornaliere.xls" and save it in the current folder. The macro is based upon a little routine that uses the combination of "*WorksheetFunction.Average*" function and "*WorksheetFunction.Count*" function.

Step 4 and Step 5: Data ordering and Data ranking

Following the instructions of the standard ISO 15927-4, the next step consists in the calculation of the Cumulative Distribution Function (CDF) which describes the probability that a real-valued random variable *X* with a given probability distribution will have a value less than or equal to *x*. The CDFs are used to select the best "month" to be included in the RY. To this aim, these steps perform a sorting and a ranking operation. In the sorting process, all climatic variables for each month of every year are arranged in ascending order; the operation is performed also on LP data. The ranking operation returns the rank of each recorded value (for every climatic variable) for each month of every year and for LP data.

Ranking and sorting operations are performed using Excel built-in functions such as the property "*Selection.Sort*" and the method "*WorksheetFunction.Rank*".

Step 6: Calculate deviations

Previous operations are used in this step to numerically build the CDFs for each month, for all the years and for the LP data in accordance with the Eqn. (2). In other words, for each climate variable, the occurrence frequency of each value within each month for all years is numerically determined.

At the end of the calculation, the VBA macro automatically creates two files respectively named "datiordinati.xls" and "RankingDeiDati.xls" and saves them in the current folder.

Step 7: Finkelstein-Schafer statistic

As can be seen from the Eqn. (3), the value of the F-S statistic is given by the sum of the differences in absolute values between the monthly CDFs and LP CDFs.

In detail, the implemented automatic procedure performs the following operations:

- Calculates deviations between monthly CDFs and LP CDFs for each calendar month and for each year. This calculation is done for the primary parameters defined in the standard: dry bulb temperature, relative humidity and solar radiation;
- Once calculated the F-S statistics for each month of each available year and for all the three primary climate parameters, it sorts F-S output values in order of increasing size.

Step 8: Calculate lowest deviations respect Long Period

For each calendar month, for the three months with the lowest total F-S, the macro calculates the deviation of the monthly mean wind speed from the corresponding multi-year calendar-month mean. The month with the lowest deviation in wind speed

is selected as the "best" month to be included in the RY.

Figure 4 shows an example of the file "StatisticFS" where are saved the matrix that permits to identify the best month to be selected for the RY. The procedure automatically marks with blue color the cell with the lowest value.

Figure 4: Selection of the best month for the RY

The sheet showed in Figure 4 permits to better understand the quality of available data. The relatively high value of deviations in the first months of 1996 are due to several missing values in the data set that cannot be fixed by interpolation because the range of the missing data is more than 11 hours.

Step 9: Smoothing

The last step of the procedure is the adjustment of the 8760 hourly values selected to build the RY. The standard ISO 15927-4 recommends to adjust the parameters in the last eight hours of each month and in the first eight hours of the next month by interpolation to ensure a smooth transition when the best months are nested in order to build the RY. Furthermore, this procedure shall include the last eight hours of December and the first eight hours of January so that the RY can be used repeatedly in simulations. The smoothing procedure permits to avoid abrupt changes in the time series that would be unfair in simulations. The standard ISO 15927-4 does not suggest any smoothing procedure leaving the user free to choose the most suitable algorithm.

In the application it was decided to use the mathematical procedure known as LOWESS (Locally Weighted Scatterplot Smoothing) (Cleveland, 1979), (Cleveland et al., 1988).

LOWESS is a data analysis technique for producing a "smooth" set of values from a time series that has been contaminated with noise, or from a scatter plot with a "noisy" relationship between the 2 variables. In a time series context, the technique is an improvement of the least squares smoothing technique that works even if the data is not equally spaced (as least squares smoothing assumes).

The algorithm implemented in the VBA macro is adapted from the one published by the National Institute of Standards and Technology, NIST. The effect of the LOWESS smoothing can be observed in the Figure 5 where last eight hourly values are joined with the first eight values of another time series.
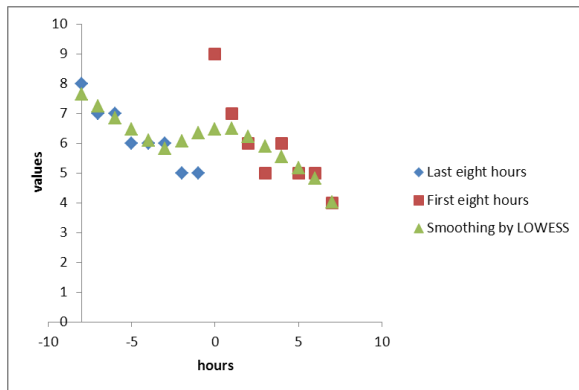

Figure 5: Application of LOWESS smoothing

At the end of the smoothing procedure a file containing the 8760 records is generated and saved in the current folder. The smoothed values are marked in green colour to better identify them, as shown in Figure 6.

Figure 6: The smoothed values in the RY marked in green.

CONCLUSION

The aim of this work was to develop a software tool based on the standard UNI EN ISO 15927-4, which permits to obtain a set of 8760 hourly values handling the original figures of the key climatic parameters: the air temperature, the solar radiation and the relative humidity.

In order to demonstrate how this algorithm can be implemented through simple and popular software tools, the authors used macros in Microsoft Excel. All of the designed routines were developed using standard mathematical functions available in VBA environment. The collection of these routines makes possible to create automatically a Test Reference Year.

The development of an algorithm for generating a RY allows the users to overcome some of the limitations encountered in the application of the most commonly used methods, i.e. the Italian standard UNI 10349 (UNI 10349) which refers to monthly climate parameters only for the most important cities and that, in case of minor cities or for areas not present in the tables, needs interpolations and other approximations.

The possibility to easily implement the UNI EN ISO 15927-4 standard procedure allows to users who have enough weather time series for any location, to autonomously generate a RY.

With the creation of a personalised RY is therefore possible to obtain a set of hourly climatic variables characteristic and exclusive of the area under examination. Furthermore, the procedure can be applied in different ways, changing the selection of key parameters in order to calculate different possible RY types. RY dedicated to annual solar radiation that can be used in solar and photovoltaic simulation or RY dedicated to wind speed and direction useful in case of preliminary evaluation of exploitation of wind energy can be easily generated. A comparison with the standard RY for the city of Palermo has produced good results. The software will be soon available for download at the website www.dream.unipa.it.

REFERENCES

Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". Journal of the American Statistical Association.

Cleveland, W.S.; Devlin, S.J. (1988). "Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting". Journal of the American Statistical Association.

UNI 10349 Riscaldamento e raffrescamento degli edifici. Dati climatici.

UNI EN ISO 15927-4: Hygrothermal performance of buildings -- Calculation and presentation of climatic data -- Part 4: Hourly data for assessing the annual energy use for heating and cooling

Vertačnik G. (2008); Test Reference Year in Environmental Agency of the Republic of Slovenia, Ljubljana.