Derivation of meteorological reference year with hourly interval for Italy

Gianluca Antonacci – CISMA Srl, Bolzano, Italy
Ilaria Todeschini – CISMA Srl, Bolzano, Italy

Abstract

The present paper describes the work which has led to the creation of the meteorological reference year for the whole of Italy, according to the algorithm contained in the UNI EN ISO 15927-4 technical standard. The work was done on behalf of CTI (Comitato Termotecnico Italiano). The above standard specifies a method to derive the hourly values of weather data needed to assess the annual energy needs for building heating and cooling. The obtained time series is built up by 8760 hourly records extracted from longer time series (theoretically at least 10 years).

In meteorological data processing particular attention was paid to the quality check of collected data. The reference year thus contains records for the following parameters: temperature, solar radiation, relative humidity, wind speed. The first three are the key values for calculation of the energy budget needed for heating and cooling.

This approach enables the development of calculation methodologies towards the direction of a dynamic thermal design in which the external forcing is given with hourly interval, instead of through monthly averages as currently contemplated in the current version of UNI 10349 standard.

The reconstruction procedure of the meteorological reference year is based on the extraction of hourly data from multi-year records and subsequent statistical analysis; the algorithm is discussed in detail in the paper.

For each Italian administrative Province, weather stations operated by local agencies were taken into account where at least ten years of data were available. Besides the implementation of the algorithm, we discuss the practical problems encountered in managing these meteorological data, caused by strong heterogeneity resulting from different types of networks, sensors, data acquisition methods and standards. The uncertainty about the result caused by these issues is also qualitatively assessed and the advantages and disadvantages of the adoption of hourly varying time series with respect to averaged values for the thermal design of buildings is discussed.

1. Introduction

The aim of this work is to construct a reference year of hourly values of appropriate meteorological data suitable for assessing the average annual energy consumption for heating and cooling, according to the UNI EN ISO 15927-4 standard. In fact, to correctly simulate the building behaviour, not only the mean values of meteorological parameters are important, but also the frequency distribution of the individual parameters and the cross correlation between them. In this way, a dynamic project of building plants can be done.

Therefore, the aim of the present work is to provide the tools and the data to update the UNI 10349 with a new and more detailed set of values.

The reason for taking into account almost ten years (but indeed preferably more) can be explained by considering that unusual periods of warm or cold weather can in this way be excluded from the computation of the reference meteorological year.

The present paper basically copes with the methodological aspect of the work, description of the algorithm and encountered difficulties. On the other hand, detailed numerical results will be distributed to end users from CTI.

Already in the early phase of data collection some issues about the data quality became quite evident, making it clear that it would not be possible to obtain completely homogeneous results. Therefore, as explained and discussed further in the article, some inconsistencies have remained in the downstream process. While the initial idea was to "manipulate" the data as little as possible, during

the work we were forced to introduce corrections and compensations in order to achieve results which could be treated as valid for the whole of Italy.

2. Description of the algorithm

The procedure allows to identify a series of hourly data in which the mean value of individual variables, their frequency distribution and correlations between the different variables within each month are as close as possible to the corresponding calendar month of the long term data set.

The procedure is characterized by two main steps: firstly the selection from the multi-year record of the "best" (i.e. statistically most representative) month is made for each calendar month. Secondly the hourly values in the selected month are adjusted so as to provide a smooth transition between months deriving from different years.

The meteorological reference year must contain 8760 values of hourly dry-bulb temperature, solar radiation, relative humidity and wind speed. Whilst the first three parameters are crucial for the energy requirements computation for heating and cooling, the wind speed can be considered as a secondary parameter.

For each meteorological parameter, p, the procedure consists in:

calculation of the daily means, p, from at least ten years (but preferably more) of hourly data of p;

calculation, for each calendar month, of the cumulative distribution function of the daily means over all years in the data set, Φ(p,m,i), by sorting all the values in increasing order and then using Equation (1):

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

where K(i) is the rank order of the i-th value of the daily means within that calendar month in the whole data set;

calculation, for each year of the data set, of 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 Equation (2):

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

where J(i) is the rank order of the i-th value of the daily means within that month and that year;

calculation, for each calendar month, of the Finkelstein-Schafer statistic, FS(p,y,m), for each year of the data set using Equation (3):

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

ranking, for each calendar month, of the individual months from the multiyear record in order of increasing size of FS(p,y,m);

for each calendar month and each year, add the separate ranks for the three climate primary parameters (temperature, solar radiation and relative humidity);

for each calendar month, for the three months with the lowest total ranking, calculate 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 reference year.

When each calendar month is selected, the last eight values of the month and the first eight values of the next month for each meteorological parameter should be adjusted in order to avoid a sharp transition between months that derive from different years. This smoothing procedure should be done by means of interpolation. In this work a smoothing procedure with running medians of five hours windows has been performed. This adjustment also includes the last eight hours of December and the first eight hours of January so that the reference year can be used repeatedly in simulations lasting more than 12 months.

This statistical computation has been carried out by means of the statistical software R, whereby a semi-automatic implementation has been made possible.

Once the reference meteorological year has been selected according to this procedure, the hourly values of diffuse and direct radiation, along with the vapour pressure, were derived.

Most weather meteorological stations measure only the global solar radiation. For many applications, however, the knowledge of the amount of diffuse and direct radiation is

important. In this work the diffuse fraction of the solar radiation was obtained by adopting the function proposed by Boland et al. (2008). The major advantage of this kind of formulation is that this function is logistic and not piecewise-defined as other commonly used formulations (e.g. Erbst). The diffuse fraction of the solar radiation on the horizontal, k, can be evaluated by means of Equation (4):

$$k=\frac{1}{1+e^{-5.0+8.6K_T}} \quad (1)$$

where KT is the hourly clearness index is defined as the fraction between the global solar radiation Iglobal and the extraterrestrial radiation H0:

$$K_T=\frac{I_{global}}{H_0}. \quad (2)$$

The orbit of the Earth is not a circle, but an ellipse, hence the distance between Earth and Sun varies over the year, so an eccentricity correction factor can be defined as:

$$E_0=(a_0+a_1\cos(\Gamma)+a_2\sin(\Gamma)+a_3\cos(2\Gamma)+a_4\sin(2\Gamma)) \quad (3)$$

where Γ, the angle of the Sun for the Nth day of the year, can be evaluated as:

$$\Gamma=2\pi\frac{N-1}{365} \quad (4)$$

and the coefficient proposed by Spencer (1971) are summarized in Table 1.

a_0=1.000110	a_1=0.034221	a_2=0.001280
a_3=0.000719	a_4=0.000077	

Table 1 – Coefficients for the evaluation of the eccentricity correction factor, as proposed by Spencer (1971)

As a consequence, the direct normal radiation is defined as:

$$I_0=1367\cdot I_{global}$$

The solar declination δ can be calculated using equation proposed by Spencer (1971):

$$\delta=b_0+b_1\cos(\Gamma)+b_2\sin(\Gamma)+b_3\cos(2\Gamma)+b_4\sin(2\Gamma)+b_5\cos(3\Gamma)+b_6\sin(3\Gamma) \quad (5)$$

where the coefficients are summarized in Table 2.

b_0=0.006918	b_1=-0.399912	b_2=0.070257
b_3=-0.006758	b_4=0.000907	b_5=-0.002697
b_6=0.00148		

Table 2 – Coefficients for the evaluation of the solar declination, as proposed by Spencer (1971)

Thus, the equation of time adjustment (in hours) proposed by Spencer is:

$$E_0=\frac{24}{2\pi}(c_0+c_1\cos(\Gamma)+c_2\sin(\Gamma)+c_3\cos(2\Gamma)+c_4\sin(2\Gamma)) \quad (6)$$

where the coefficients are summarized in Table 3.

c_0=0.000075	c_1=0.001868	c_2=-0.032077
c_3=-0.014615	c_4=-0.040849	

Table 3 – Coefficients for the equation of time, as proposed by Spencer (1971)

The local solar time LsoT can be then estimated as:

$$LSoT=GMT+E_T+\frac{1}{15}(\lambda-LSTM) \quad (7)$$

where λ is the local longitude (positive for East) and LSTM the local standard time meridian.
Finally the sun height can be defined as:

$$H_S=\frac{\pi}{2}-\arccos[\cos(\varphi)\cos(\delta)\cos(\omega)+\sin(\varphi)\sin(\omega)] \quad (8)$$

with ϕ is the latitude and ω the hour angle can be defined as:

$$\omega=15(LSoT-12)\frac{\pi}{180}. \quad (9)$$

Now the extraterrestrial radiation can be calculated by means of equation (13):

$$H_0=E_0[\cos(lat)\cos(\delta)\cos(\omega)+\sin(lat)\sin(\omega)] \quad (10)$$

Once the diffuse radiation is evaluated by means of equation (4), the direct radiation is automatically defined as:

$$I_{direct}=I_{global}-I_{diffuse} \quad (11)$$

Besides, the saturation pressure can be estimated by means of the formulation contained in the ASHRAE handbook, depending on the temperature range. For the temperature range of -100°C to 0°C is given by:

$$\ln(p_{ws})=d_1T^{-1}+d_2+d_3T+d_4T^2+d_5T^3+d_6T^4+d_7\ln(T) \quad (12)$$

while for the temperature range of 0°C to 200°C is given by:

$$\ln(p_{ws})=d_8T^{-1}+d_9+d_{10}T+d_{11}T^2+d_{12}T^3+d_{13}\ln(T) \quad (13)$$

where the temperature T is the absolute temperature and the coefficients are summarized in Table 4.

d_1=-5.6745359E+03	d_2=6.3925247E+00
d_3=-9.6778430E-03	d_4=6.2215701E-07
d_5=2.0747825E-09	d_6=-9.4840240E-13
d_7=4.1635019E+00	d_8=-5.8002206E+03
d_9=1.3914993E+00	d_{10}=-4.8640239E-02
d_{11}=4.1764768E-05	d_{12}=-1.4452093E-08
d_{13}=6.5459673E+00	

Table 4 – Coefficients for the evaluation of the vapour saturation pressure (ASHRAE handbook)

Thus, the vapour pressure p_w can be calculated starting from the hourly values of relative humidity RH:

$$p_w = \frac{RH}{100} p_{ws} \qquad (17)$$

3. Choice of the meteorological station

For each Italian province, weather stations operated by local agencies were taken into account where at least ten years of data were available, even if in some cases the effectively usable record length turned out to be shorter. Only the stations where all the four meteorological parameters were measured were considered.

For each reference station, one of the most important criteria for the choice of the representative meteorological station was the length of the record. Also the amount of non-valid data has played a significant role in the selection process. If more than one weather station was available, the more suitable station for the entire territory of the province was selected.

In the lower layers of the atmosphere, average temperature approximately decreases with the altitude at a fairly uniform rate. For this reason, measured temperature values depend on the weather station altitude. It is therefore necessary to correct the hourly values of temperature and consequently vapour pressure when referring to another location characterized by a different altitude value with respect to the monitoring station.

The vertical temperature gradient depends on the geographical area. In Italy five different areas have been identified in the present UNI 10349 standard, whose corresponding values are reported in Figure 1. We adopted the same classification, and although it is quite rough, revisiting this classification is out of the scope of present work.

The corrected temperature T_{corr} at the altitude z can be calculated:

$$T_{corr} = T_{station} - (z - z_{station})\delta_t \qquad (18)$$

where zstation and Tstation are respectively altitude and measured temperature at the selected weather station.

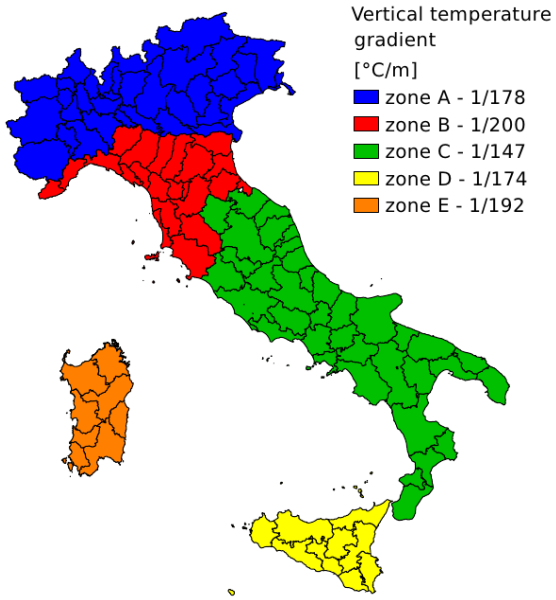

Fig. 1 – Geographical zone with different values of the vertical temperature gradient δt, according to UNI 10349 technical standard

4. Example of reference meteorological year

Following is an example of the result of the meteorological reference year building procedure (Fig. 2). For each month, it is possible to compare the hourly temperature values with the corresponding monthly average (black line) provided by the present UNI 10349. Please note that this average value does not correspond to the monthly average of the selected reference month but it refers to average values of three decades of meteorological data. A further discussion regarding the differences between the present method (UNI 10349) and the reference meteorological year as described in this work is presented in the following paragraph.

5. A synthesis of results

After the described elaboration, which clearly adopts a local approach, the obtained hourly meteorological data were also extrapolated, although in quite a approximate way, in order to obtain a synthetic result for the whole of Italy and to compare the meteorological reference years obtained in the different provinces.

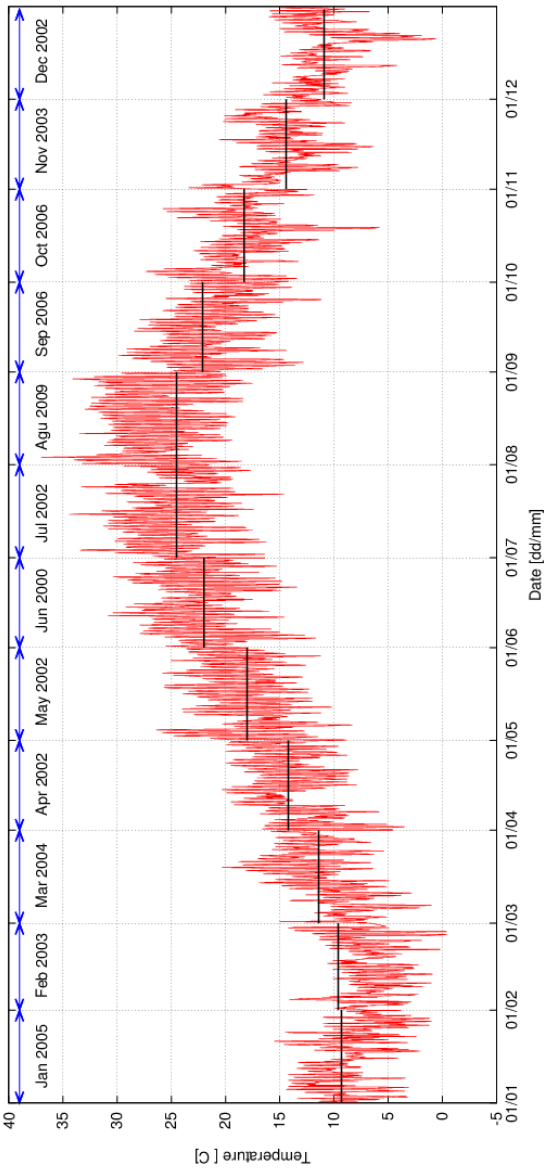

Fig. 2 – Example of the hourly temperature values identified by means (the figure is referred to the city of Brindisi) of the algorithm proposed by UNI EN ISO 15927-4 and monthly averages according to current version of UNI 10349 standard

Annual average values of temperature, global radiation, relative humidity and vapour pressure were calculated for each Italian municipality; the correction for different values of altitude were applied to temperature and consequently vapour pressure. For this reason their values vary within the same province, while relative humidity and global radiation, which were not corrected with the altitude, show a unique value within each province. This is obviously a rough approximation, and should therefore be used only as a first indication. The approach was used only in order to maintain consistency with the current version of the UNI 10349 standard, but simply extrapolating meteorological values to a whole territory would not make much sense, as the statistical approach contained in the UNI EN ISO 15927-4 cannot take into account the spatial variability of orography and land use, both strongly influencing wind speed, relative humidity and solar radiation. Therefore the only extrapolation on temperature, along the vertical coordinate, could be considered as acceptable from the climatological point of view.

6. Description of the encountered problems

Beside the implementation of the algorithm, a strong effort has been directed towards the solution of several practical problems encountered in managing meteorological data. The problems were principally caused by strong heterogeneity resulting from different types of networks, sensors, data acquisition methods and standards.

In fact, nowadays a major problem in managing Italian meteorological data is the heterogeneity of the observational network. It should also be considered that the current version of UNI 10349 is based on an analysis that might have been less accurate, but certainly more homogeneous, as at the time it was issued the observational network was managed by the National Weather Service before it was dismantled. Currently we have 20 different regional services with different stations, different sensor types, different criteria for installation, different placement (ranging from urban to rural), different procedures for data management and different ranges of available periods. As consequences, the mean monthly values of some stations as derived with the present

work deviate from those of the current standard (UNI 10349). With reference to this, it should also be considered that the adopted method is in any case different from the older one which led to the current version of the UNI 10349 standard. The latter is in fact built up of the monthly values of parameters, simply averaged over three decades, while the depicted algorithm gives a "characteristic" monthly value, namely the value less deviating from the average. The two values can be practically considered as coinciding only for very long time series (several decades), which indeed is not the case for the present analysis.

These difficulties are so relevant because it is not possible to define a unique procedure to overcome them, given the heterogeneity of the data and of

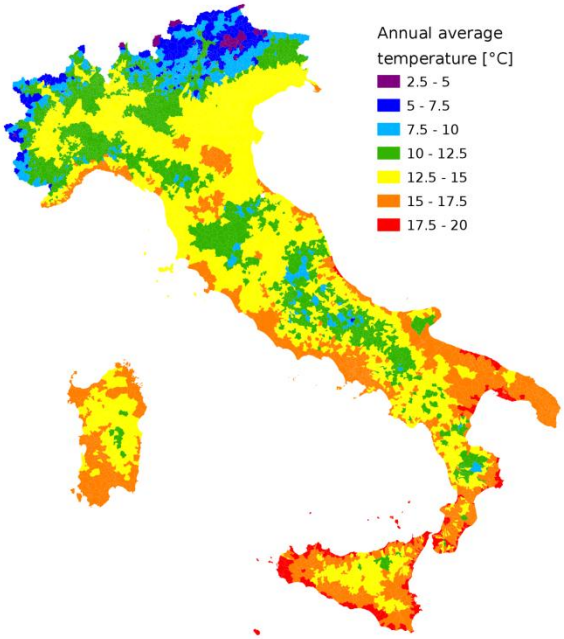

Fig. 3 – Map of the annual average temperature

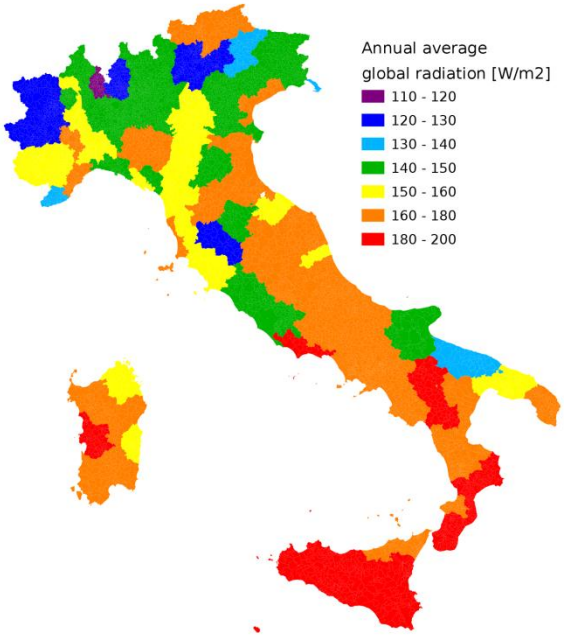

Fig. 4 – Map of the annual average global solar radiation

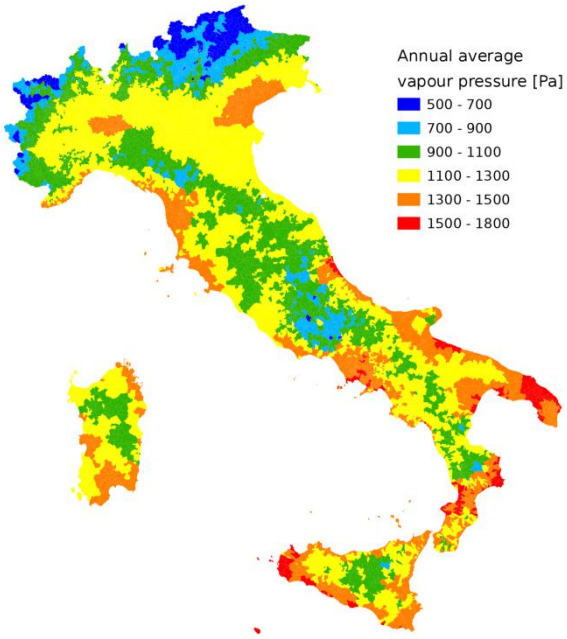

Fig. 5 – Map of the annual average vapour pressure

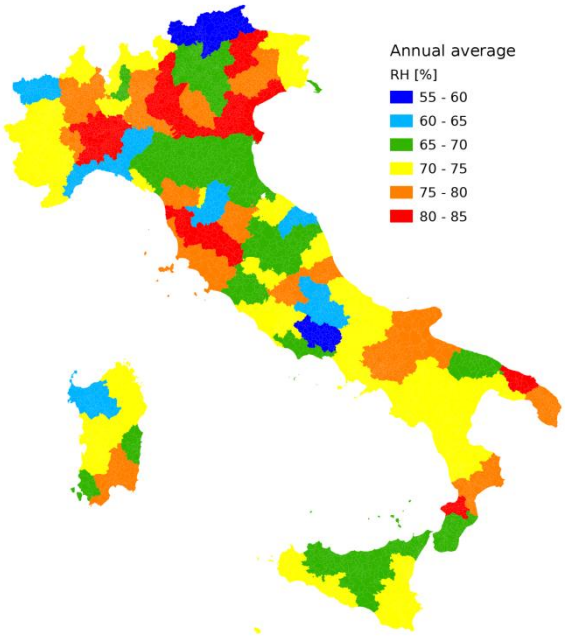

Fig. 6 – Map of the annual average relative humidity

their corresponding problems. Data quality is in fact strongly inhomogeneous and most of the times it is not evident.
A short overview of the main problems is presented in the following paragraphs, along with a brief description of the adopted solutions.

6.1 Lack of data

The problem of a lack of data could not be ignored, since periods without any valid data were present in almost every time series that has been analyzed.
The UNI EN ISO 15927-4 suggests much care should be exercised in the quality control of the raw data. In accordance with the regulation, missing values shall be generated by linear interpolation or by an estimate, and sudden, unnatural jumps or singular values shall be examined and corrected. This approach is cannot be adopted for long periods of time without any valid data, since the basic principle of the regulation is to use real measured data as much as possible instead of theoretical and interpolated values.
For this reason, to avoid non-valid data in the meteorological reference year, all the months with significant amounts of non-valid data, even of just one parameter, were not considered and not used by the algorithm.
The selection process of the "best" (i.e. statistically most significant) month to be included in the reference meteorological year is certainly influenced and modified by this choice. Nevertheless, this approach was preferred in order to preserve the original data.

6.2 Homogenization

Homogenization was necessary in order to make comparable data that derive from different sources. In fact, the analysed data were collected by the different local agencies, whose standards are not unique. Also the criterion adopted in the choice of the localization is strictly influenced by the aim of the station and it is different for example for agro-meteorological, air quality or hydrographic stations.
One example of this homogenization process is the adjustment of the time axis to the standard local time (UTC+1 for Italy).
Moreover, the theoretical hour of sunrise and sunset was calculated and compared with the corresponding measured ones in each weather station. In some cases, the length of the day was systematically shorter than the theoretical one, probably due to a bad positioning of the radiometer. Global radiation values at sunrise and sunset were therefore obtained by means of interpolation.
Relative humidity data larger than 100%, measured in case of oversaturation in some weather stations, were fixed to 100%.
Usually the lowest value of the measured wind speed coincides with the cut-in threshold value typical of the installed anemometer. A common value of 0.1 m/s was chosen as arbitrary threshold value for wind speed.
Although the homogenization process, the heterogeneity was not completely eliminated, as it can be seen for example in the annual average solar radiation map in Figure 4. Some provinces in the south of Italy are characterized by smaller values of solar radiation if compared with the contiguous ones. This behaviour is due to the fact that the considered meteorological stations are very different. The only option was to modify collected data before adopting the described procedure, but that was not the purpose of the UNI EN ISO 15927-4 technical standard.

6.3 Validation

The most part of the analysed meteorological data were supposed to be validated by the local agencies that manage the weather stations. Validation is a crucial process that consists in examination of data in order to detect possible errors, but also in the standard procedures that include periodical calibration of the instruments, data screening and also preventive maintenance.
Nevertheless, although theoretically validated, a large amount of data showed clearly unnatural values that had to be compensated or eliminated where corrections where not feasible.

The main corrections that were made are reported below:

- values exceeding 50% of the 99th percentile (both maximum and minimum values) were eliminated;
- values repeated for more than 5 hours were eliminated;
- global radiation values higher than the solar constant were eliminated;
- offset of the entire time series due, for example, to the existence of an urban heat island or to a change in the position of the sensor was eliminated;
- linear detrending was applied where the drift of temperature and solar radiation sensors was clear and identifiable.

6.4 Length of the temporal series

The UNI EN ISO 15927-4 standard provides the correct procedure to evaluate the reference meteorological year. Before applying the suggested algorithm it is therefore necessary to identify for each province the most representative weather station. As already explained, the most important criterion for the choice of the weather station was the length of the available temporal series. Significant differences exist between different regions. In fact, some regions have collected more than 20 years of meteorological data, in other regions less than 10 years were available (notice that a minimum 10 years length is the one recommended by the UNI EN ISO 15927-4). In some cases, the meteorological reference year was forcibly evaluated on the basis of shorter temporal series due to lack of data covering continuously a longer period.

7. Conclusions

A reference meteorological year is a data set of hourly values of temperature, solar radiation, relative humidity and wind speed throughout a year. The months have been selected from individual years and concatenated to form a complete year. This work describes the procedure adopted to build it for each Italian province, according to the UNI EN ISO 15927-4 standard.

The most important characteristic of this approach is that it enables the development of calculation methodologies towards the direction of a dynamic thermal design in which the external forcing is given with hourly interval, instead of through monthly averages as currently contemplated in the UNI 10349 standard. In fact, a more accurate evaluation of heating and cooling loads in buildings and the performance of solar thermal and photovoltaic systems requires accurate and reliable meteorological data series as input. For this reason a reference meteorological year is a more appropriate data source when evaluating the thermal performance of buildings through numerical simulations.

Unfortunately a great heterogeneity characterizes meteorological data provided by the local agencies all over Italy. A stronger effort should therefore be directed to enhance a homogenization procedure in order to obtain more comparable data sets. In fact, a major issue with the data management was that we were forced in some cases to do (or redo) the data validation process. Strictly speaking, this action would need to gather also all the information about the sensors, their calibration and data filtering algorithms. This was out of the scope of the present work, but according to the authors, it should be reconsidered for the future, at least for the most critical cases. As a consequence, the initial intention of barely modify original data, has been clearly and forcibly reconsidered.

As a conclusion, we believe that the obtained results, which will be distributed by CTI in the future, should be intended as a first valuable step that overcomes the limitation of the monthly mean climatological values, allowing dynamic simulations with higher time resolution. Results should, however, certainly be improved in the future, at least through the timely extension of the time series as they become available, but also through finer discrimination of reliable data. It should, in fact, be underlined once more that the whole procedure as foreseen by the UNI EN ISO 15927-4 standard is implicitly based on the assumption that quality checked, long enough and continuous data series are available.

A further reflection should also be reserved for the purpose for which these climatological series will be used. When using them to analyse meteorological parameters which are strictly referred to the measurement location (or the surrounding area for which the station can be considered as representative), the adoption of the presented time series can be considered as suitable. In fact, extending the results over a larger area is conceptually more questionable, as it introduces a strong hypothesis on spatial variation. In this case a more comprehensive (in space) approach should be taken into consideration, thus overcoming the limitation of punctual (always intended as spatially) statistics.

8. Acknowledgement

This work was carried out on behalf of the CTI (Comitato Termotecnico Italiano) as part of the research coordinated and carried out by ENEA. The authors would like to thank Prof. Baggio (University of Trento) for the valuable suggestions and Dr. Murano (Comitato Termotecnico Italiano) for the accurate checking of the results.

9. Nomenclature

Symbols

E_0	Eccentricity factor [-]
F	Cumulative distribution function of the daily means within each month
FS	Finkelstein-Schafer statistic
GMT	Greenwich Mean Time
H_0	Extraterrestrial radiation [W/m2]
H	Sun height [rad]
I_{global}	Global solar radiation [W/m2]
$I_{diffuse}$	Diffuse solar radiation [W/m2]
I_{direct}	Direct solar radiation [W/m2]
k	Diffuse fraction of the global solar radiation in the horizontal [-]
KT	Hourly clearness index [-]
LSTM	Local Standard Time Meridian
LsoT	Local Solar Time
p_{ws}	Saturated vapour pressure [kPa]
p_w	Vapour pressure [kPa]
RH	Relative humidity [%]
T	Temperature [°C]
z	Altitude [m]
Γ	Angle of the sun for the Nth day of the year [rad]
λ	Longitude [°]
δ	Solar declination [rad]
δ_t	Vertical temperature gradient [°C/m]
Φ	Cumulative distribution function of the daily means over all years
ω	Hour angle [rad]
ϕ	Latitude [°]

References

ASHRAE Fundamentals Handbook, 2005. Psychrometrics, Chapter 6, 6.2.

Crawley M. J., The R Book, Wiley 2007, ISBN 978-0-470-51024-7.

Boland J., Ridley B. and Brown B. 2008. Models of diffuse solar radiation, Renewable Energy, Vol. 33, 575-584.

Spencer J.W., 1971. Fourier series representation of the position of the Sun, Search 2 (5), 172.

UNI EN ISO 15927-4 2005. Hygrothermal performance of buildings, calculation and presentation of climatic data. Part 4: Hourly data for assessing the annual energy use for heating and cooling.

UNI 10349 1994. Riscaldamento e raffrescamento degli edifici. Dati climatici, CTI.