Research Papers
Issue RP0203
December 2013

CIP - Climate Impacts and Policy Division

Database of observed time series of climate variables and scenarios of future climate change

By **Paola Marson**
Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)

Emanuele Massetti
Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Fondazione Eni Enrico Mattei (FEEM)
emanuele.massetti@cmcc.it

SUMMARY This document summarizes a dataset with historical, observed, time series of climatic variables and scenarios of future climate, at global level created as part of the activities of the GEMINA, WP 7.1.3. We briefly present the main data sources and methods using Italy as a case study. For questions about the database please write to Emanuele Massetti, emanuele.massetti@cmcc.it.

This report represents the Deliverable P149 developed within the framework of Work Package 7.1.3 of the GEMINA project, funded by the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea.

1 Observations of past climate trends

• The ERA-interim reanalysis

ERA-Interim is the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It integrates a variety of observing systems with numerical forecasting models to produce a record of the global atmospheric circulation. The reanalysis is produced in a 12-hourly cycle of forecast and analysis steps: the observed fields are considered as the true state of the atmosphere and taken as input to initialize the parameters of a short-range numerical forecast model (the ECMWF IFS). The prediction obtained is combined to new available observations in the analysis step and the results become the initial state for a new cycle of forecast. The physical parameters estimated by the forecast model are constrained in the analysis step by the observations with physical and temporal laws. The consistency with the laws of physics and the valuable contribution of observations differentiates the reanalysis method from others in estimating geophysical parameters from observations.

A detailed description of the sources of observed data, the methods and procedures involved in the data assimilation and reanalysis process can be found in D.P. Dee et al. (2011a).

• Data characteristics

The data are provided without any restriction of use for non-commercial purposes by the European Centre for Medium-Range Weather Forecasts (ECMWF), UK. We focus here on the variable Air Temperature at 2 meters from surface, the standard air temperature variable.

The ERA-Interim archive currently contains 6-hourly gridded estimates of fields from 1979 to the present following a monthly update schedule.

The dataset covers the whole globe using a N128 Gaussian grid. Data is also available with other spatial resolution (0.75°x0.75°; 1°x1°; 1.125°x1.125°; 1.5°x1.5°; 2°x2°; 2.5°x2.5°; 3°x3°). For the studies conducted within GEMINA project we chose to use the highest available resolution, namely 0.75°x0.75°. This means that we have climate data for the whole world in a 480x241 grid, for a total of 115,680 observations per unit of time. For each observation unit we have data at 6-hour time-intervals (00:00 UTM, 06:00 UTM, 12:00 UTM, 18:00 UTM). Our dataset extends from 1979 to present day, for a total of over 5 billion observations over time and space.

This highly disaggregated (both spatially and temporally) data will be used for a new generation of studies that will assess the impact of climate change on agriculture. Thanks to the very high temporal resolution it will be possible to study how different time sequences of temperatures affect land values and crop yields. The data reveals the existence of climate extremes and the peculiar characteristics of these extremes, over time and space.

Here we provide a brief description of the data by projecting the dataset on the Italian municipalities.

• The ERA-Interim projected on Italian municipalities

Data from ERA-interim reanalysis were used to obtain a description of observed climate in Italy, at the level of the 8,100 Italian municipalities, from 1981 to 2010.

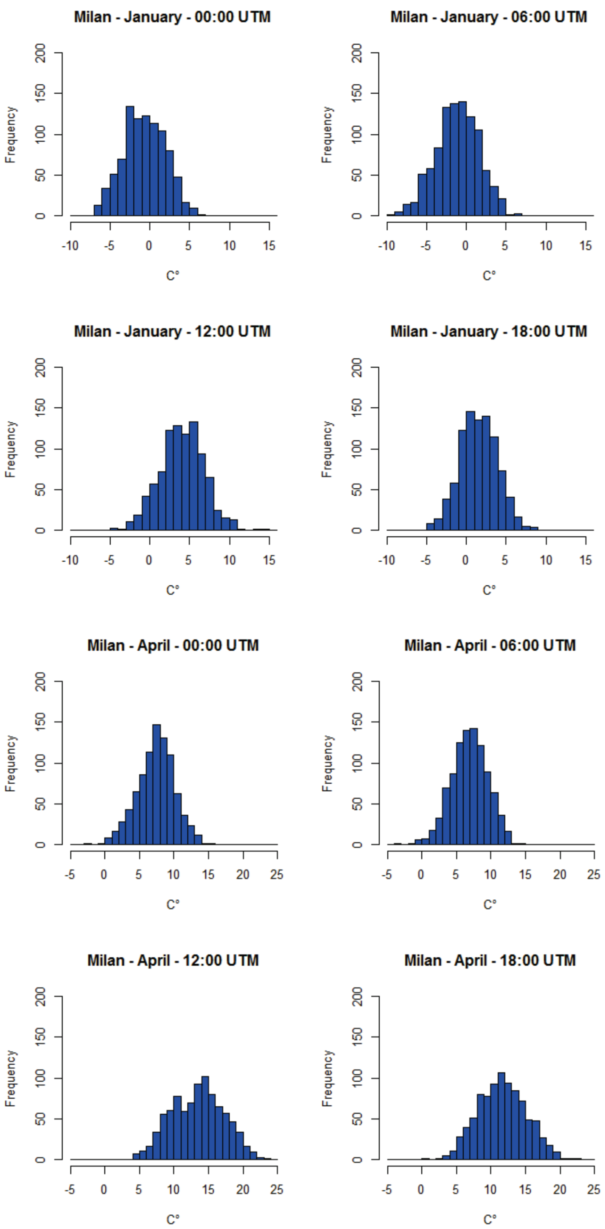
We started by checking for the presence of outliers (temperatures too low or too high with respect to the season considered). Outliers were removed and substituted by interpolating adjacent values..

The climate data was attributed to each of the 8,100 Italian municipalities by means of a simple interpolation between the centroid of each municipality and the four closest grid-points of the ERA-Interim dataset, with weights inversely proportional to distance. The geo-referenced data on Italian municipalities is from the Global Administrative Areas Database (GADM) version 2.0 (January 2012) (www.gadm.org).

The interval of variation of the recorded temperature per month has been divided in bins: each bin corresponds to an interval of one degree Celsius of variation. From data projected on each municipality we computed the count of observations falling inside each bin. The high frequency of data recorded per day permits to take into account in the analysis the variations of temperature inside the daily cycle, and for this reason the extreme events (both in high and low temperatures) occurred.

As an example of the possible use of the data, we display the distribution of 6-hour temperature intervals in the areas of Milan, Rome and Palermo, during the central months of the four seasons in Figure 2, Figure 3 and Figure 4.

Figure 1. GADM database of Global Administrative Areas (version 2.0 January 2012) - Italian Municipalities and ERA-interim reanalysis grid points in the selected box.


Milan - January - 00:00 UTM
Milan - January - 06:00 UTM
Milan - January - 12:00 UTM
Milan - January - 18:00 UTM
Milan - April - 00:00 UTM
Milan - April - 06:00 UTM
Milan - April - 12:00 UTM
Milan - April - 18:00 UTM
Frequency
C°

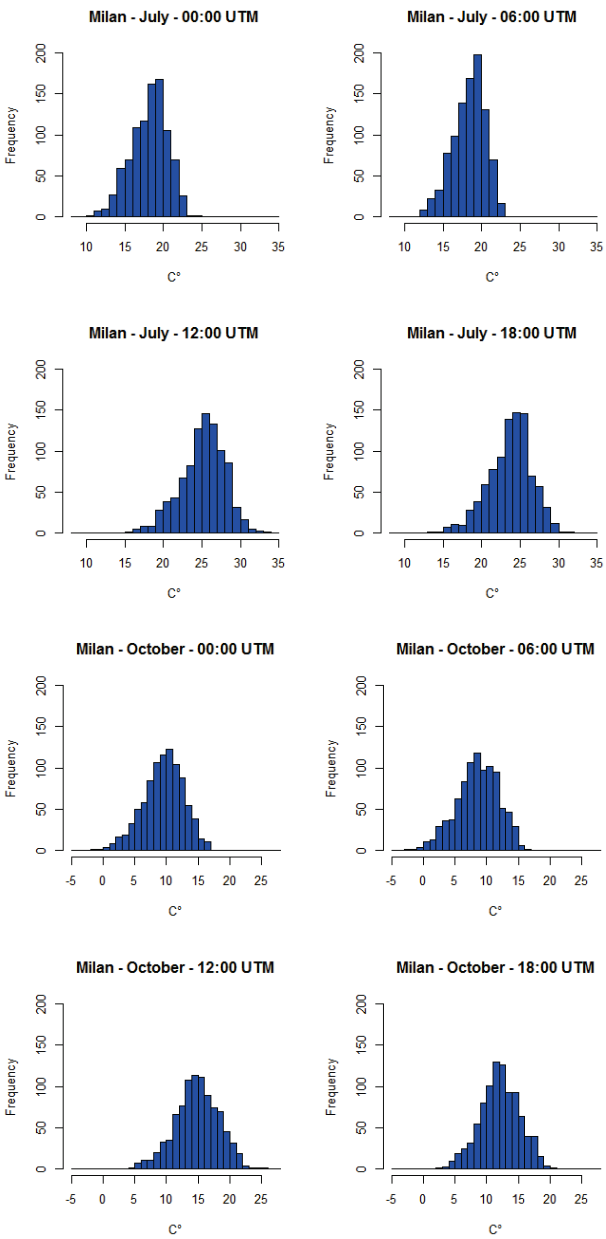
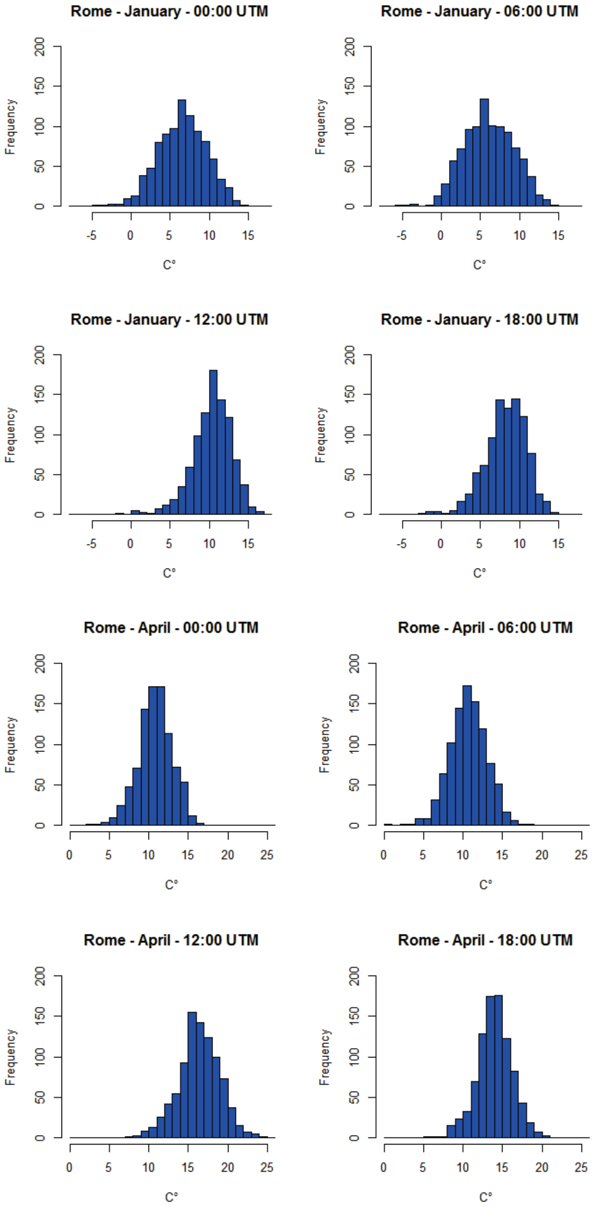


Figure 2. Milan: Occurrence of observations per bin considering January, April, July and October in the period from 1979 to 2010. Data: ERA-interim ECMWF reanalysis interpolated to the area of Milan.


Rome - January - 00:00 UTM
Rome - January - 06:00 UTM
Rome - January - 12:00 UTM
Rome - January - 18:00 UTM
Rome - April - 00:00 UTM
Rome - April - 06:00 UTM
Rome - April - 12:00 UTM
Rome - April - 18:00 UTM
Frequency
C°

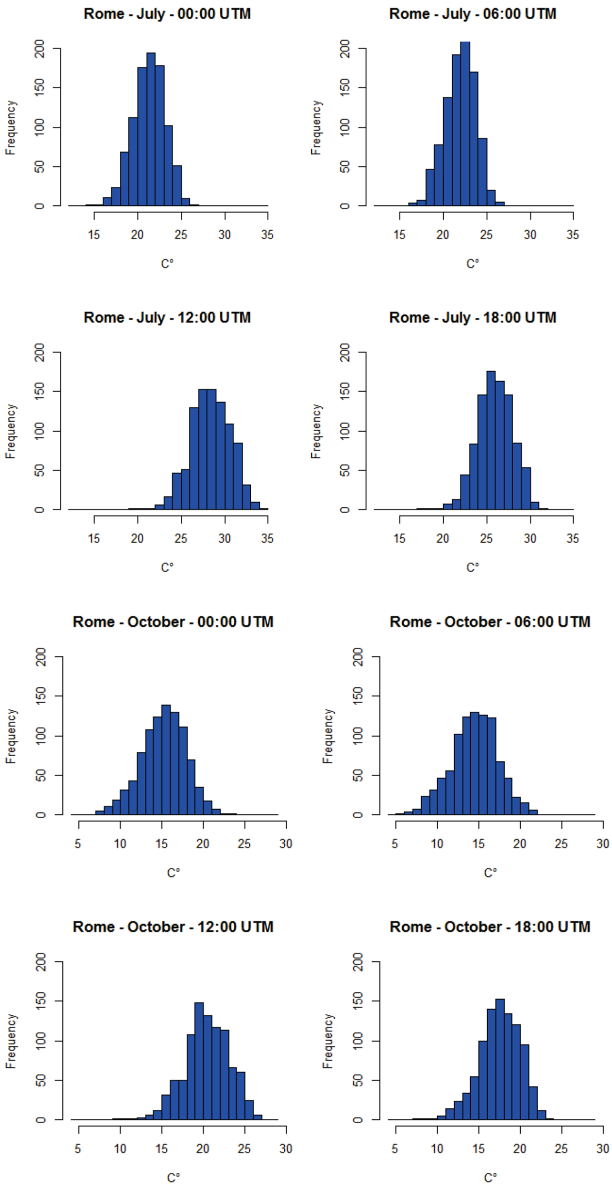
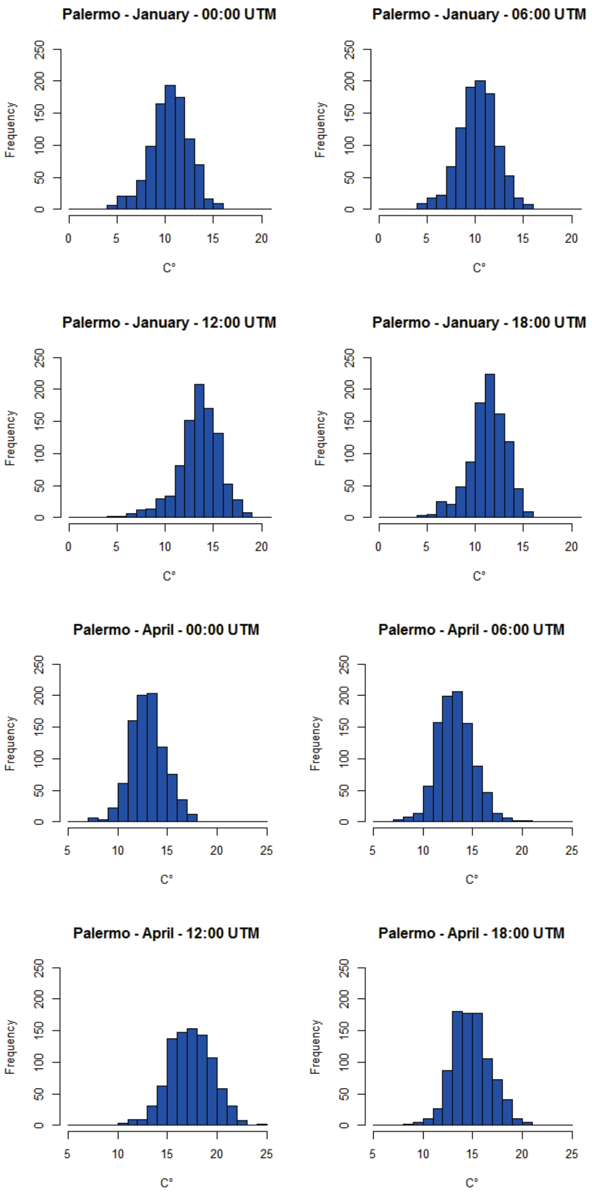


Figure 3. Rome: Occurrence of observations per bin considering January, April, July and October in the period from 1981 to 2010. Data: ERA-interim ECMWF reanalysis interpolated to the area of Rome.


Palermo - January - 00:00 UTM
Palermo - January - 06:00 UTM
Palermo - January - 12:00 UTM
Palermo - January - 18:00 UTM
Palermo - April - 00:00 UTM
Palermo - April - 06:00 UTM
Palermo - April - 12:00 UTM
Palermo - April - 18:00 UTM
Frequency
C°

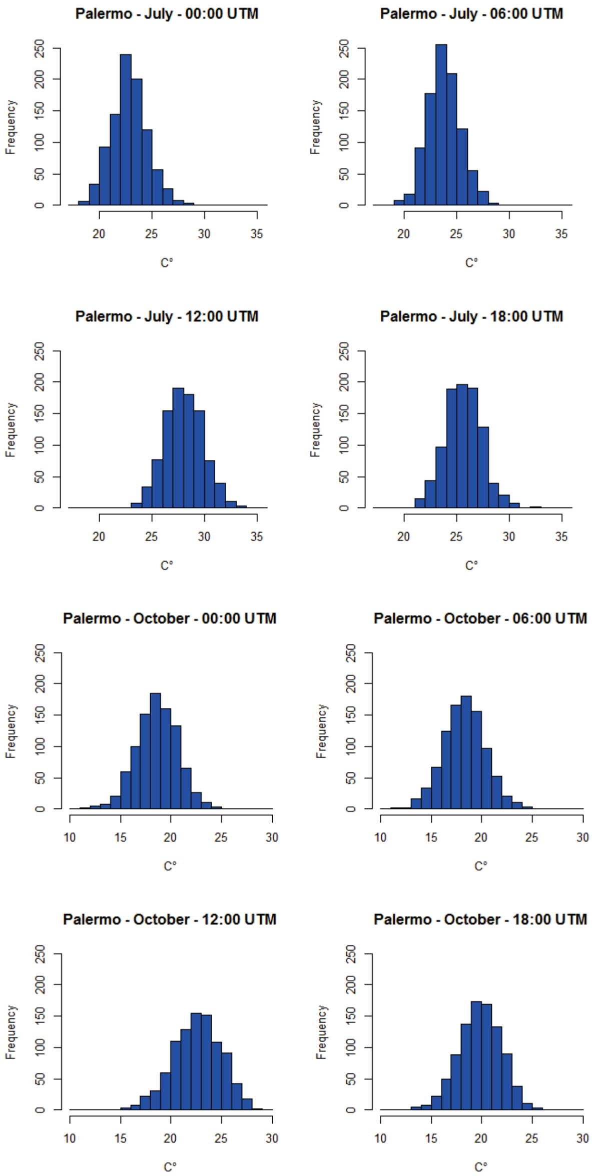


Figure 4. Palermo: Occurrence of observations per bin considering January, April, July and October in the period from 1979 to 2010. Data: ERA-interim ECMWF reanalysis interpolated to the area of Palermo future climate.

2 Climate simulation from the CCSM4.0 Global Circulation Model

A General Circulation Model simulates atmospherical fields by numerically solving equations which include physical, biological and chemical knowledge of the climate system. In addition to these information a GCM requires as input for the initialization boundary conditions (values from observations or other studies kept fixed during the simulation) and external forcings which drive the changes in climate. The generated prediction is discrete in time and space and can describe the changes in climate as a result of slow changes in input constrains (I.e the rensponse of climate to different emissions of greenhouse gases and aerosol).

2.1 Data characteristics

Data are obtained from the Earth System Grid Federation (ESGF) Portal, which comprehends all the simulations computed by the research institutions taking part to the Coupled Model Intercomparison Project Phase 5 (CMIP5), within WCRP's Working Group on Coupled Modelling.

The simulation here considered is produced by the National Center of Atmospheric Research (NCAR) in the USA, using the Community Climate System Model (CCSM) version 4.0. Experiments conducted running this model both in the past and in the future, are conducted to replicate past climate observations and to provide scenarios of future climate under alternative radiative forcing conditions. With the exception of exogenous radiative forcing, all other homogeneous conditions remain unchanged in the experiments (in terms of boundary conditions and time – space resolution) so that the results obtained can be used to compare climate conditions in the two periods.

For GEMINA project aims we only developed algorithms to download, store and process data from the CMIP5 exercise in an efficient way. Here we provide examples using Air Temperature at 2 meters from surface : the following description of data refers to this field.

The NCAR CCSM model provides data in a grid of about 1.25 x 1.00 degrees (288 X 192 points) of Longitude and Latitude respectively.

The experiment that reproduces past climate covers the period from 1850 to 2005, while the future run scenario of future climate change starts from 2006 and ends in 2100. The data can provide a description of changes in temperatures and can be combine to past descriptions already obtained from other investigations (in particular the one from ERA-interim reanalysis here portrayed). We use temperature data with daily temporal resolution (for both the experiments) for a total of more than 20 billion observations over time and space.

The new set of emission scenarios adopted by climatologists spans four representative concentration pathways (RCPs). Here we considered the CCSM4 output obtained under the Representative Concentration Pathway 6.0. The RCP 6.0 is developed by the AIM modeling team at the National Institute for Environmental Studies (NIES) in Japan. It is a stabilization scenario where the total radiative forcing is supposed to be constant at 6.0 W/m^2 after 2150. The details of the scenario are described in Fujino et al. (2006) and Hijioka et al. (2008).

2.2 Future change of the distribution of daily temperatures in Italy

We compare the 2031-2060 climatology of the distribution of daily mean temperature in the RCP 6.0 scenario (i.e. the observed frequency of different levels of daily mean temperature) with the 1975-2005 climatology of the distribution of mean daily temperature in the historical emissions scenario. We then apply the emerged change of the distribution to the observed present climatologies of mean daily temperatures from ERA-interim reanalysis data.

From the retrieved global gridded data, we selected the same box as in the ERA-interim exercise. It contains the Italian peninsula and part of the Tyrrhenian, Adriatic and Ionian sea, as shown in the following figure:

Figure 5. GADM database of Global Administrative Areas (version 2.0 January 2012) - Italian Municipalities and CCSM4 grid points in the selected box.

Also in this case, we performed a series of data quality tests. If present, we replaced outliers by interpolating adjacent values, in space and time. The values have been then interpolated to each centroid of Italian municipalities using the same interpolation method as before.

We computed the same counts of observation as in the ERA-interim exercise. The following graphs show the counts in the past (black) and in the future (green) according to the simulation from CCSM4. As before, the results presented refer to the area of Milan, Rome and Palermo in January, April, July and October.

To build the future climate in the decades from 2031 to 2060, we added to the daily mean from ERA-interim the expected change in temperature derived from the comparison between past and future CCSM4 simulations (Figures 9-11).

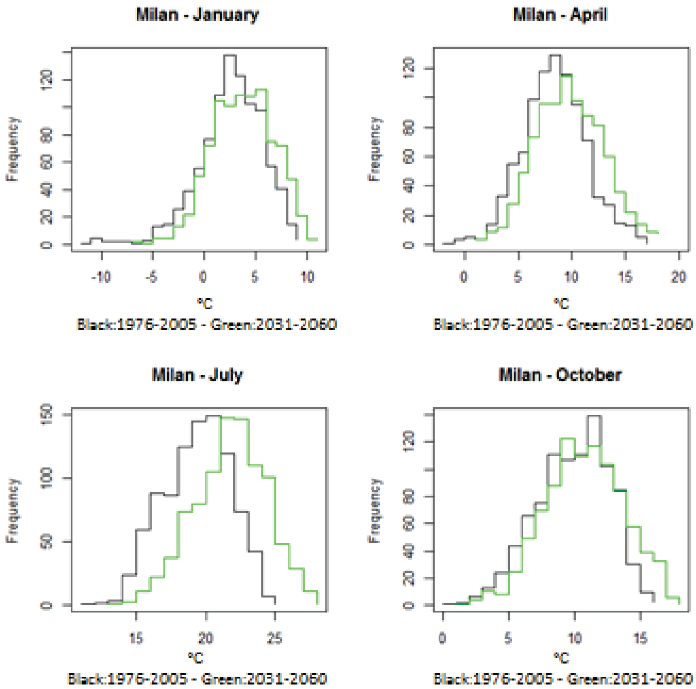


Figure 6. Milan – Number of days per degree Celsius, CCSM4 model, RCP 6.0. Black: 1976 – 2005, Green: 2031 – 2060.

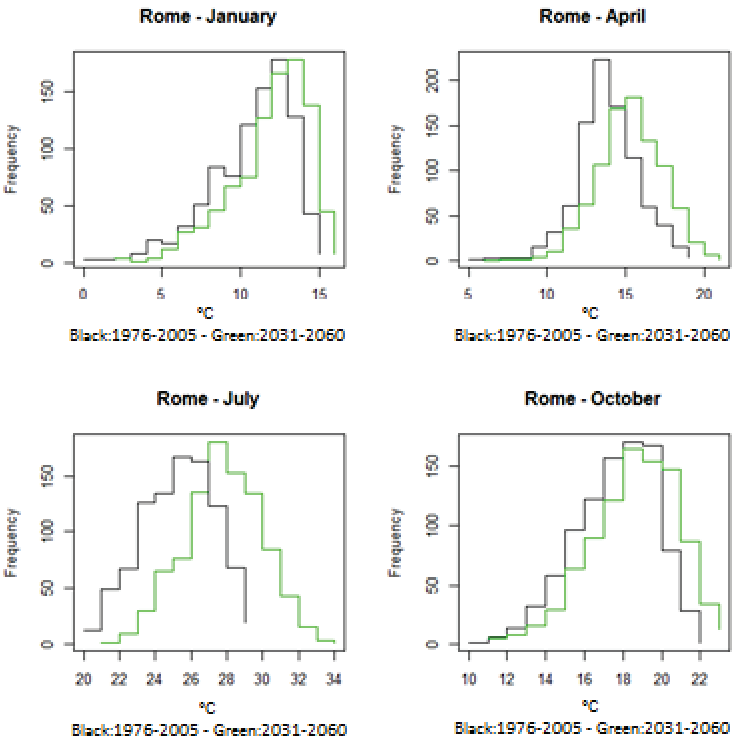


Figure 7. Rome – Number of days per degree Celsius, CCSM4 model, RCP 6.0. Black: 1976 – 2005, Green: 2031 – 2060.

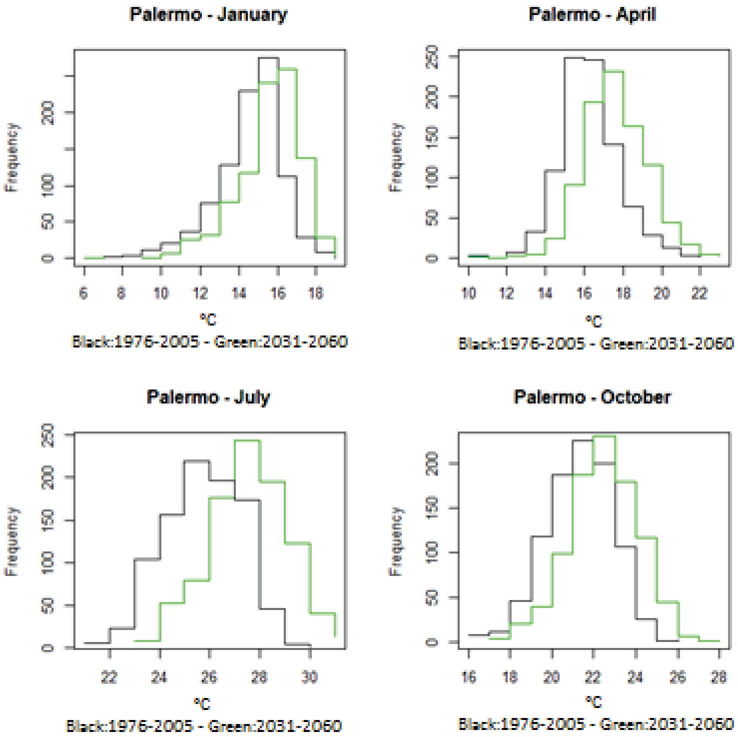


Figure 8. Palermo - Number of days per degree Celsius, CCSM4 model, RCP 6.0. Black: 1976 – 2005, Green: 2031 – 2060.

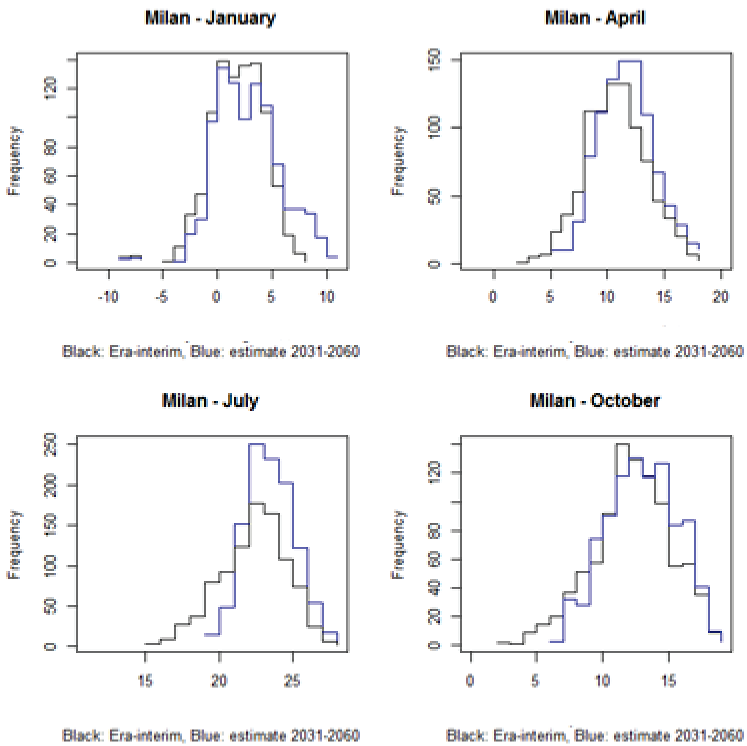


Figure 9. Milan - Black: present daily temperature (ERA-interim); Blue: estimate future count of days per degree Celsius built by adding to the daily mean from ERA-interim the expected change derived from the comparison of past and future CCSM4 simulations.

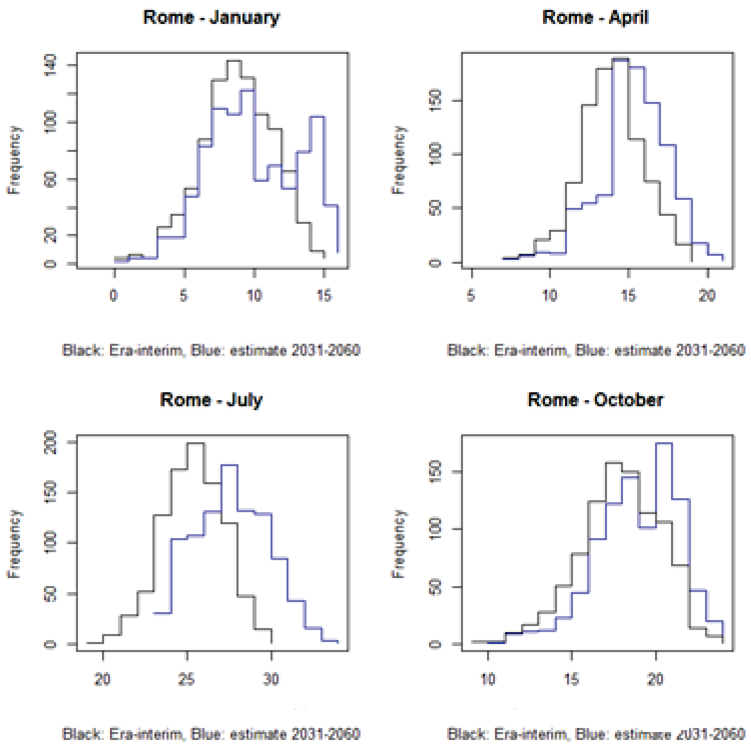


Figure 10. Rome - Black: present daily temperature (ERA-interim); Blue: estimate future count of days per degree Celsius built by adding to the daily mean from ERA-interim the expected change derived from the comparison of past and future CCSM4 simulations

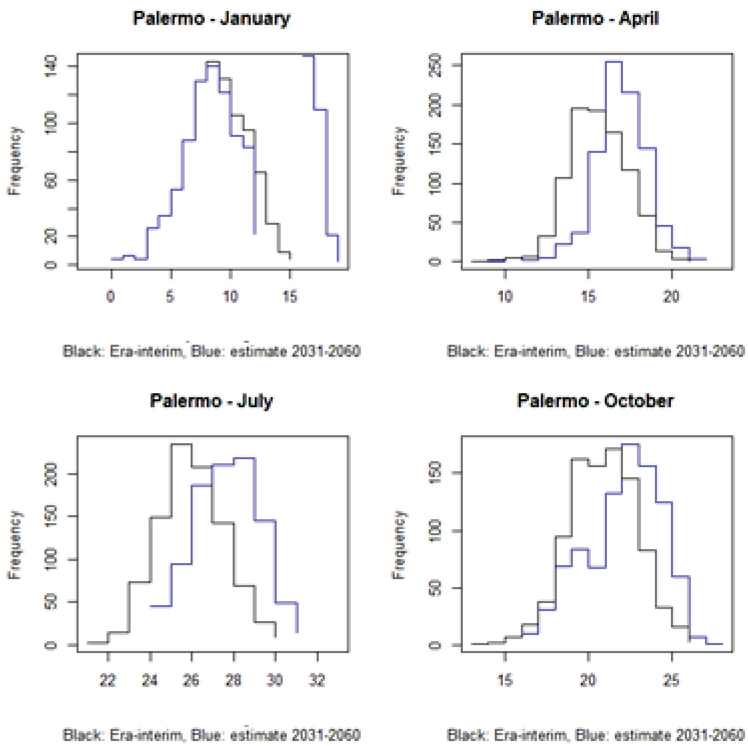


Figure 11. Palermo - Black: present daily temperature (ERA-interim); Blue: estimate future count of days per degree Celsius built by adding to the daily mean from ERA-interim the expected change derived from the comparison of past and future CCSM4 simulations

3 Excess Heat Factor

As a second exercise with our data we computed an index to measure events characterized by extreme heat. The index used is the *Excess Heat Factor* (*EHF*) defined by Nairn et al. (2009).The index is a combination of two measures based on daily mean temperature:

a) The ***Excess Heat Index*** *measures an excess heat condition, defined as an unusually high heat arising from a high daytime temperature not sufficiently discharged overnight due to unusual overnight temperature. A* short-term averaged temperature is compared to a long-term climate reference value to establish the existence of significant excess heat. The authors set the reference value as the 95th percentile of observed mean daily temperature for all days spanning (about) 30 years (hereafter T_{95}). When a three-days mean daily temperature anomaly with respect to the reference is positive we observe a significant excess heat event. On day *i the Excess Heat Index* is defined as*:*

$$EHI_{sig}(i) = (T_i + T_{i-1} + T_{i-2})/3 - T_{95}$$

b) The ***Heat Stress*** compares the three-day daily mean temperature to the 30 days previous state in order to get an evaluation of the acclimatization and the stress condition of the system on day *i.*

$$EHI_{acc}(i) = (T_i + T_{i-1} + T_{i-2})/3 - (T_{i-3} + \ldots + T_{i-32})/30$$

The units of both indices are °C.

The combined effect of *Excess Heat* and *Heat Stress* provides a measure of intensity and duration of a heatwave event, defining the **Excess Heat Factor (EHF)** as:

$$EHF(i) = EHF_{sig}(i) \times max[1, EHI_{acc}(i)]$$

The product of the two quantities increases as a quadratic response: the second term enhances the excess heat when the observed day *i* is preceded by a relative medium-range rise in temperature, preserving the sign given by $EHI_{sig}(i)$.

The 3-days average implicitly ensures a minimum period of persistence to make an anomalous condition to be meaningful. Furthermore, the authors defined the *heat load* of an event as the sum of consecutive positive EHFvalues.

3.1 A modified version of the *Excess Heat Factor*

The EHF can be only applied to data referring to the summer season. This limitation lies in the threshold defining an *excess heat state* set as the 95th percentile of the empirical distribution of daily means across the whole annual cycle in the reference period. This approach does not allow for the identification of an unusual condition in colder seasons, which may be instead desirable in some studies on impacts.

To overcome this limitation, we account for a new threshold defined as the 95th percentile of the empirical probability distribution of each of the four seasons across a 30 years reference period (hereafter sT_{95}).

The $EHI_{acc}(i)$ factor remains unchanged in the new version, while for day *i* the $EHI_{sig}(i)$ becomes:

$$sEHI_{sig}(i) = (T_i + T_{i-1} + T_{i-2})/3 - sT_{95}$$

where *s* states for the season considered. As a consequence:

$$sEHF(i) = sEHF_{sig}(i) \times max[1, EHI_{acc}(i)]$$

We performed a preliminary test of the modified index on the ERA interim dataset covering Italy. We compared the results produced by the two approaches for the months from June to August, when the comparison is more demanding. The reference period spans from 1981 to 2010. We portray the results obtained from some grid points covering the Padan Plain and Sicily for summer 2003 (Figure 12 and Figure 13)and for summer 2000(Figure14 and Figure 15). Summer 2003 was characterized by known extreme heat conditions while summer 2000 is not considered an anomalous season in the literature. The figures also display the 3-days average temperature as an added reference to compare the behaviour of the indices. The average is computed homogenously to the index: $T_a(i) = (T_i + T_{i-1} + T_{i-2})/3$.

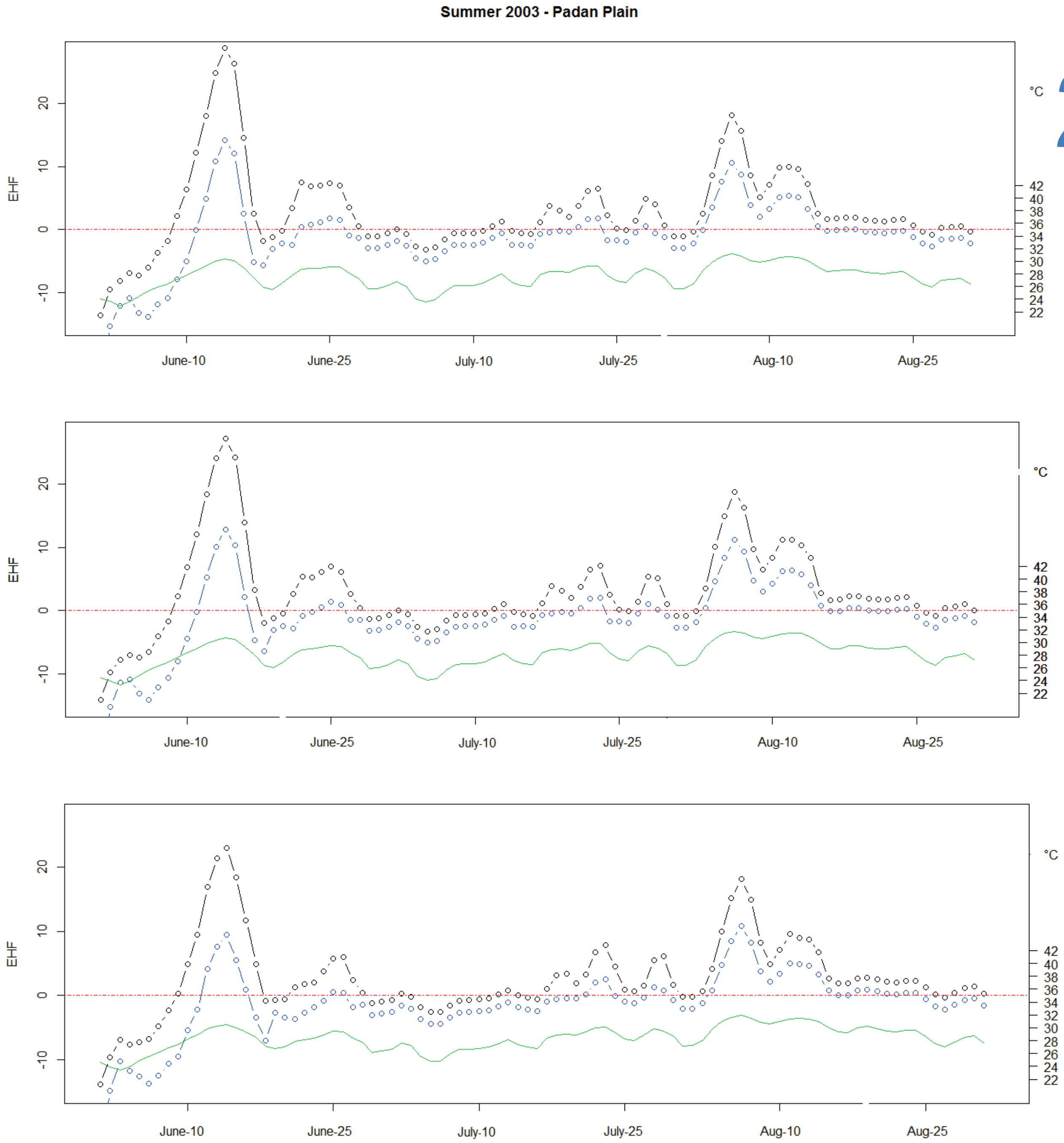


Figure 12 – Each panel represents the value of the two indices and the 3-day average temperature relative to 3 grid points over the Padan Plain.. Black: EHF, Blue: sEHF, Green: 3-days average temperature .Period: JJA 2003

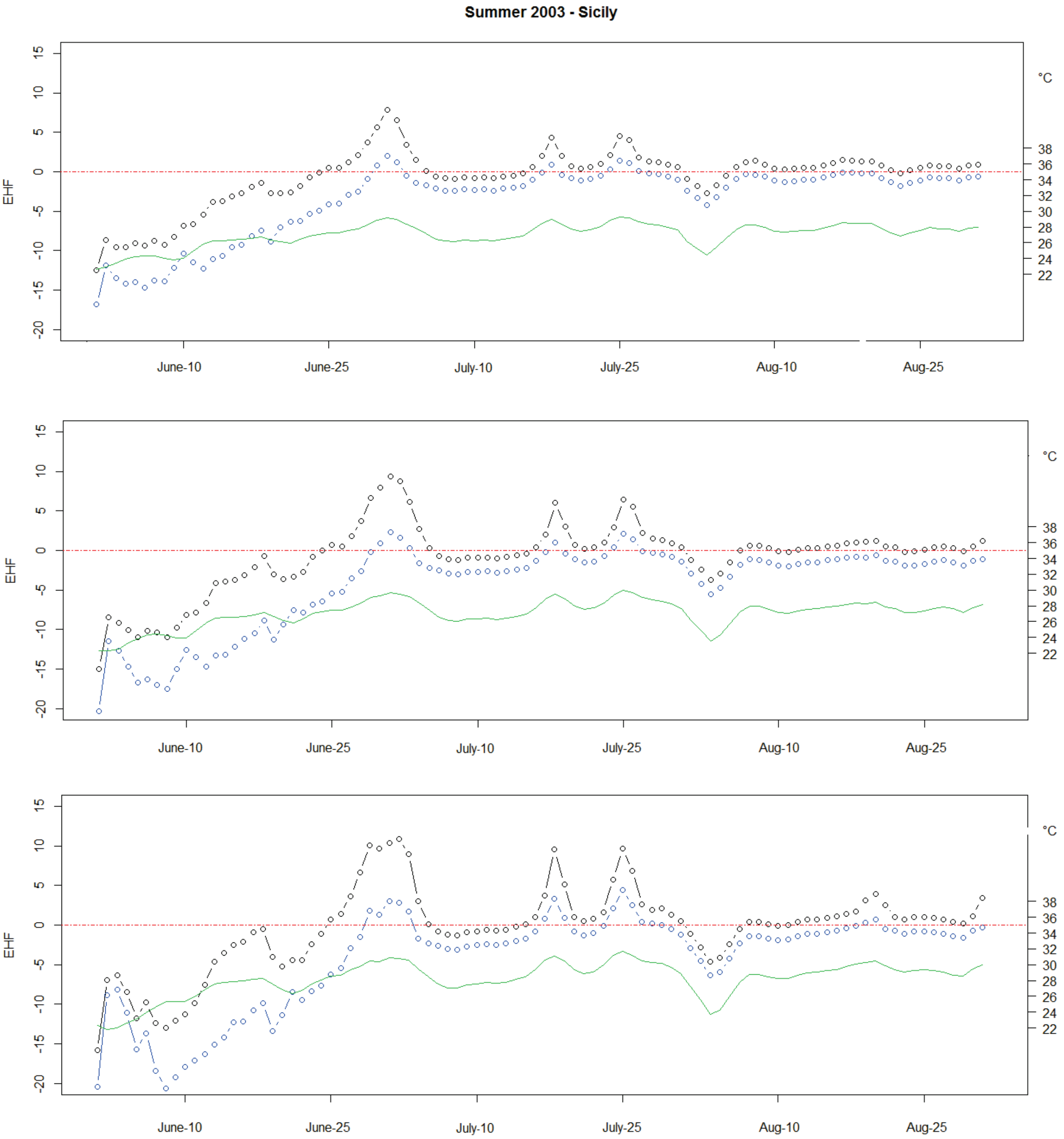


Figure 13 - Each panel represents the value of the two indices and the 3-day average temperature relative to 3 grid points over Sicily.. Black: EHF, Blue: sEHF, Green: 3-days average temperature .Period: JJA 2003

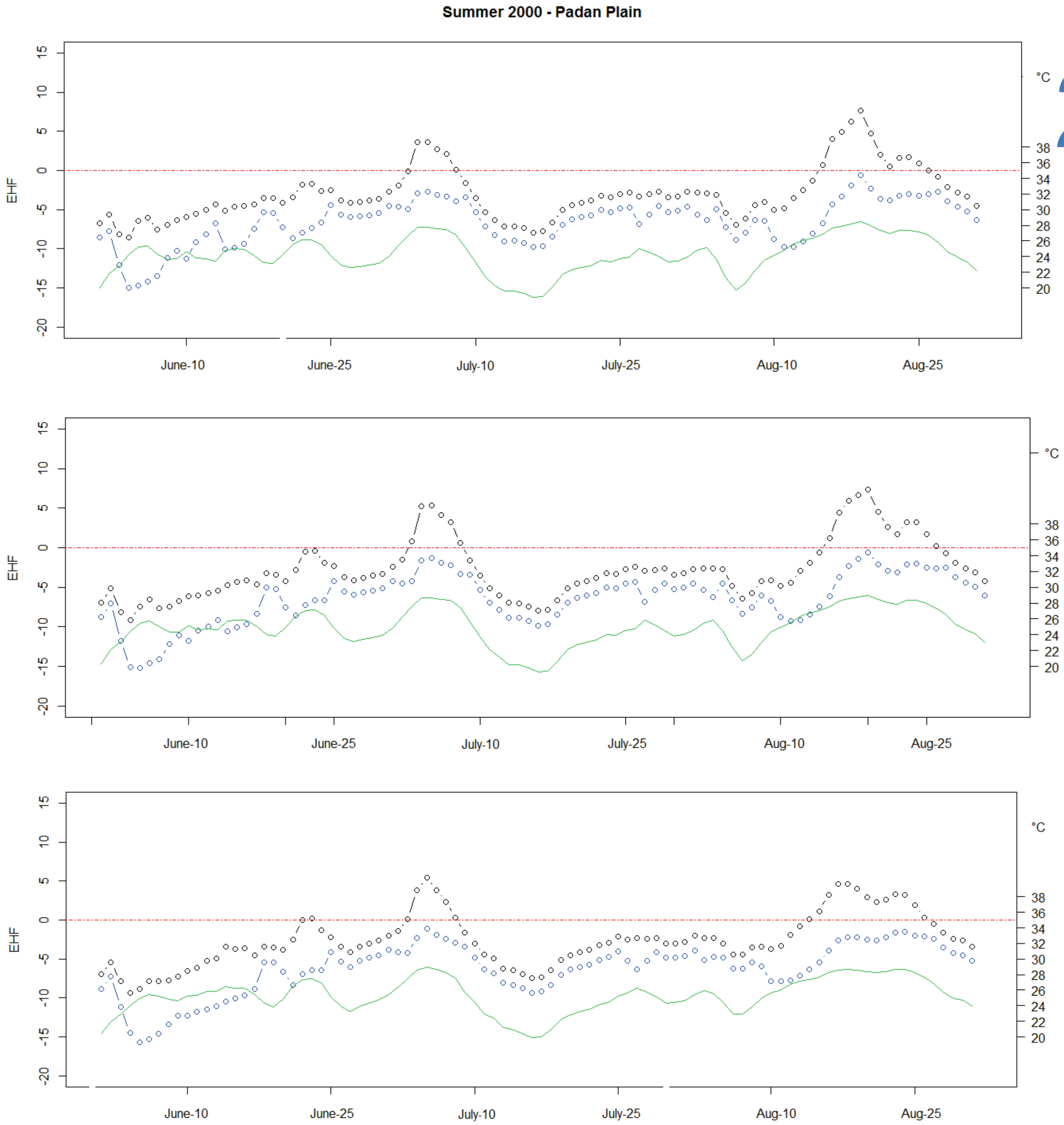


Figure 14 - Each panel represents the value of the two indices and the 3-day average temperature relative to 3 grid points over the Padan Plain.. Black: EHF, Blue: sEHF, Green: 3-days average temperature .Period: JJA 2000.

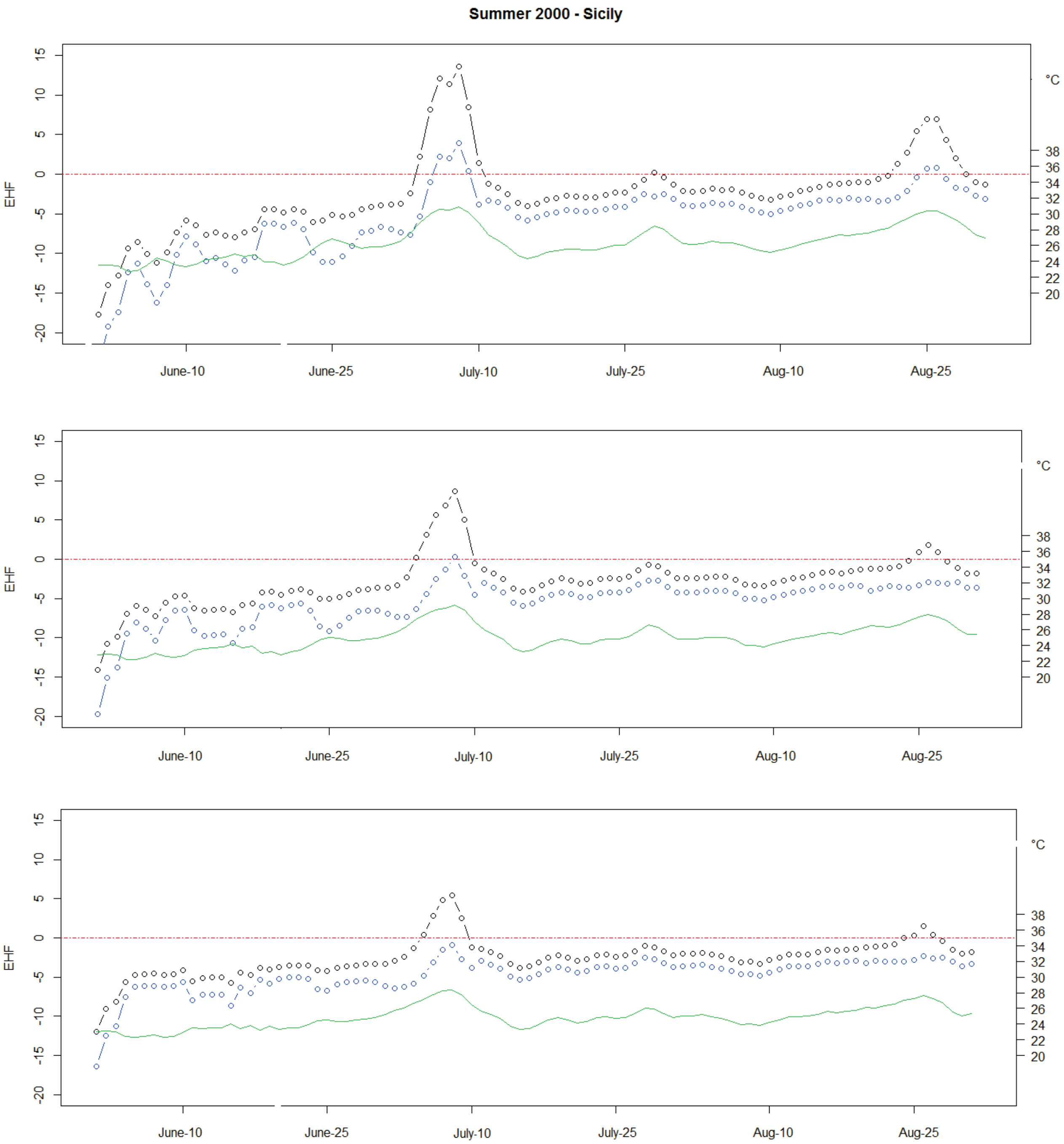


Figure 15 - Each panel represents the value of the two indices and the 3-day average temperature relative to 3 grid points over Sicily.. Black: EHF, Blue: sEHF, Green: 3-days average temperature .Period: JJA 2000

The indices track a similar trend in all the panels. The sEHF is more cautious in identifying excess heat events than the original version in both years here investigated. The extreme conditions revealed by the sEHF last fewer days and are detected some days after with respect to the EHF. It is an expected result, consistent with the choice of the reference distribution: the factor $sEHI_{sig}(i)$ is expected to be positive (in summer) fewer times than $EHI_{sig}(i)$. By contrast, the first index will reveal a higer number of extreme heat events during colder seasons.

Since $EHI_{sig}(i)$> $sEHI_{sig}(i)$ in summer ,the modified version $sEHF$ is less sensitive to the acclimatization factor and its value has more moderate adjacent changes.

Besides the above mentioned brief considerations on the behavior of the indices on the analyzed data in summer, the modification has a general property. The 95th percentile in the *significance* factor provides a quantitative definition of an extreme heat. The relative threshold based on the seasonal distribution sT_{95} reveals what is an unusual state in the period of interest within the year. This property can bring some advantages: as a trivial example, treating a typical summer crop, one can be interested in identifying unusual heat conditions relative to the life cycle of those specific plants.

Both versions of the index will be computed on the dataset from CCSM4.0. The historical experiment will constitute the baseline on which to estimate T_{95} and sT_{95}. The thresholds will be used to calculate the values of the indices both on the historical and the future simulations and compare the time series of the results and the *heat loads*.

Visit **www.cmcc.it** for information on our activities and publications.

The Euro-Mediteranean Centre on Climate Change is a Ltd Company with its registered office and administration in Lecce and local units in Bologna, Venice, Capua, Sassari, Viterbo, Benevento and Milan. The society doesn't pursue profitable ends and aims to realize and manage the Centre, its promotion, and research coordination and different scientific and applied activities in the field of climate change study.

cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici