ClimateCHIP = Climate Change Health Impact & Prevention: http://www.climatechip.org/

23 October 2014

Technical report 2014: 3

Hothaps-Soft: A tool for the estimation and analysis of local climate and population heat exposure

Matthias Otto [1*], Bruno Lemke [1], Tord Kjellstrom [2,3,4,5]

[1]Nelson Marlborough Institute of Technology, Nelson, New Zealand
[2]National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia
[3]Centre for Global Health Research, Umea University, Umea, Sweden
[4] International Health, University of Tromso, Norway
[5] Health and Environment International Trust, 168 Stafford Drive, Mapua, New Zealand

* Corresponding author: Matthias Otto, matthias.otto@nmit.ac.nz

ABSTRACT

Researchers often access weather station data from multiple sources to identify and estimate current and future heat exposures. Many resources are freely available; however, numerous steps are often involved to convert raw data to meaningful values for evaluation, research and publication.

Global weather data collected and re-distributed by NOAA, "Surface data, Global Summary of the Day" (GSOD 2012) was chosen as the primary source for the analysis and presentation software described here. No attempts were made to correct errors or interpolate missing data: missing data is shown as gaps in the charts and not included in the statistics. Data by the Climatic Research Unit (CRU-3) is included for quality control, which has been rigorously tested and is commonly used by climate modellers.

Assuming 18,500 weather stations collected by NOAA containing at least one month of useful daily data over the 30 years needed to detect meaningful trends, and the parameters contributing to human heat exposure, more than 800 million data points are eligible for analysis.

The developed software tool, named Hothaps-Soft, allows for rapidly locating weather stations by name or longitude/latitude, analysing data density, and importing into a database. Once imported, time trends, statistics and analyses of the readings and calculated heat exposure indices (WBGT, UTCI) can be exported numerically or displayed/printed in graphical, publication-ready, and customisable formats.

Hothaps-Soft, including the source dataset, will run off a USB memory device on a Windows-compatible computer[*] without any installation or Internet connection, an important aspect for researchers in developing countries where heat exposure is of particular concern.

[*] The software has been tested on Windows XP, Vista, Windows 7 and 8

INTRODUCTION

The growing awareness of the impact of climate conditions on human society and the projections of future climate change (IPCC 2007) necessitates user-friendly tools for analysis and interpretation of local climate and weather data. A key issue for human health and work productivity is daily and hourly heat exposure in workplaces (Kjellstrom et al. 2009a). Physiological limits to human body heat balance and the narrow range of core body temperature that needs to be maintained to protect health and performance (Parsons 2003), means that accurate estimates of heat stress are needed so that preventive actions can be taken when local climate conditions exceed certain limits (ISO 1989, ACGIH 2009).

Most of the global population live and work in hot tropical or sub-tropical areas and climate change projections of standard climate variables (IPCC 2007) indicate that the ambient heat exposure situation is going to get worse. A tool that uses such variables from daily routine weather station data and grid cell based climate estimates was therefore developed by our team.

We carried out preliminary work to study weather station data and global grid cell data, and to calculate heat stress indices and time trends. This work used spreadsheets and macros to extract the relevant information from various data sources involving a number of stages, many of which required manual input. In addition to being time-consuming, these processes generated large numbers of individual files.

We decided to prepare a user-friendly software tool because no "off-the-shelf" solutions provided the functions necessary for extracting data from raw text-files and producing statistical analyses and publication-ready charts. The development and initial use of the software has been a component of the global HOTHAPS (High Occupational Temperature Health and Productivity Suppression) research program (Kjellstrom et al. 2009b), and testing in different locations has created important feedback for improvements to the software.

The purpose of Hothaps-Soft is to provide researchers, government officers, community organizations, enterprise managers, occupational health staff and environmental professionals with a tool to locate weather stations of interest, based on geographic or other criteria, from a large global catalogue, study the characteristics of the climate variable measurements and calculate selected heat stress indices.

DATA AND METHODS

Source data

Datasets freely available (to scientists) that have both humidity and temperature data required for calculating heat stress indices include:

- Daily weather station data readings published by the National Oceanic and Atmospheric Administration as the Global Summary of the Day (GSOD 2012)
- Monthly grid-cell (0.5° x 0.5°) estimates, derived from weather station readings produced by the Climate Research Unit at the University of East Anglia, UK (CRU-3)

The GSOD set consists of a series of space-delimited text files, which can be downloaded as annual archives, each containing the weather data recordings of all stations collected by NOAA. The compressed archives for a 30 year period amount to about 2GB of data.

GSOD is derived from the Integrated Surface Hourly (*ISH*) database by calculating daily averages. The readings cover a period from early in the 20th century up to the present day for approx. 30,000 weather stations to which only minimal eliminations of obviously erroneous data were applied (Lott 2004). No attempts were made to add or interpolate missing data. Recorded parameters include dew point, rainfall, maximum, minimum and mean temperature, maximum and mean wind speed. Missing or grossly implausible readings are marked as invalid. As weather data prior to 1980 was often recorded manually rather than electronically, we have chosen daily records since 1980 until the present. Annual updates can be applied as soon as new data becomes available, normally in the first few days in January.

The CRU dataset consists of a global data-grid where all (land-based) cells have been calculated by processing data from relevant weather stations with the aim of excluding local "urban heat island" effect (Oke 1973). Thus, rural weather station data inside a grid cell are given more influence on the grid cell results than urban stations (Mitchell & Jones 2005). The version of the CRU data set used, CRU-TS3.1, covers monthly averages of Tmin, Tmax, Tmean, humidity, cloud cover, precipitation, frost and wet day frequency data, and is available from 1901 to 2009 (CRU-3).

While GSOD is the primary data source, the data from the CRU-TS3.1 set has been imported into the database to provide quality control *comparison values* for dew point, mean and maximum temperatures. The CRU data is only available in the form of monthly averages that disallow daily threshold analyses (see below). There are credibility risks in manipulating and homogenising raw data (Hickman & Randerson 2009), so no data manipulation is attempted by Hothaps-Soft. However, where readings are covering less than 90% of the relevant period (month or year) they are excluded from monthly and annual statistics, and are not presented in the software's output.

Lazy loading

After removing obviously unsuitable stations (see Data Reliability and Filtering below) there are still 18,000+ land-based GSOD stations available to the user. Importing all these stations into the database did not seem sensible. The project's strategy employs an on-demand (*lazy*) loading strategy for these reasons:

- The majority of these stations contain insufficient data for many studies.
- Filling the database with all available GSOD data would impair performance and unnecessarily increase memory demand, especially on portable computing devices.
- The import, filtering and pre-processing algorithms have been modified several times throughout the development process. Importing all available raw data for each iteration was not feasible, given the time required for a complete import.

The user can locate a weather station by code, name, country or geographic coordinates from the raw dataset, ascertain how much data is available for that station by running an analysis step (see below), and then import the station's data if it meets the requirements. Importing and pre-processing data for one station takes about 15 seconds on a present day PC/laptop.

Data importing

Hothaps-Soft utilises a database managed by a ubiquitous open-source database management system (MySQL); common Microsoft software products (Microsoft .NET, Mitchell 2009) were chosen as the programming platform for their well supported rapid development systems. The programming language is C#.

For analysing the raw GSOD data and importing them into the database the compressed files are expanded into plain text and then parsed. Heat stress indices WBGT and UTCI are calculated using published formulas (see below) and stored during importing. Unavailable readings, represented as 999 or 9999 values in the raw data, are recorded as *nulls* in the database.

In order to improve speed and flexibility of use, and to be able to rapidly plot any combination of graphs and trends in the same chart, some database *denormalisation*[†] had to be applied. For example, to produce one 30 year graph of a single measurement (e.g. maximum temperature), 10,950 records (30 years * 365 days) would have to be read from the database, counted, then filtered according to the imposed 90% *reliability criterion* (see below) and subsequently averaged for each year or month. A composite chart may contain 10 graphs requiring reading in excess of 100,000 data points. While these numbers are not exorbitant for a modern computer, it would make the user experience sluggish and less useful when exploring and analysing many combinations of readings. For this reason, as part of the *import* step the daily data is processed and the results stored in a second table. The *Monthly Averages* table contains *monthly* as well as *annual averages* for each measurement and station which passes the 90% criterion. This reduces the number of database reads per graph to one per year, with no filtering, averaging or other processing required. Once imported, viewing any combination of data and trends in graphical form, or exporting statistical data in numeric form, is virtually instant.

Data reliability, analyzing and filtering

It was anticipated that data availability and reliability was going to be a major concern. NOAA aggregates data from thousands of stations. Variability of local conditions (e.g. stations being moved, renamed, equipment replaced, upheavals like natural disasters, wars, strikes, and more mundane electricity interruptions) over the 30 year period renders some weather data as patchy.

During the early development stages of Hothaps-Soft a number of special methods and functions have been implemented to deal with the completeness (or "data density") of the GSOD data set. One of the metadata[‡] files provided by NOAA is the *Historic Integrated Surface Hourly Database Station List* (ISH-History). It contains information pertaining to all past and present stations. In 2010 the number of stations in this list was in the region of 30,000.

[†] Denormalisation is a concept in database design, where certain best-practice rules are broken, for example by storing calculated data, thus causing logical duplication and compromising data integrity. Denormalisation is often used to improve performance.

[‡] Metadata, sometimes called "data about data" is not the actual data of the problem domain, but *describes* the relevant data. Metadata may entail, for example, column headings, data types (e.g. numbers or text), the meaning or context of the data (e.g. a legend) or the number of records.

To make Hothaps-Soft more useful, the likelihood of the user locating stations that don't contain any useful data for the desired period had to be reduced. In the first instance any so-called *Bogus* stations where removed, as well as data from buoys, and any that could clearly be identified as invalid or test entries. Further, a function was built into the software to identify and purge stations that contain no data for the specified year range (1980 to the present). The remaining list of approx. 18,500 stations forms the table currently used for the user to choose stations from.

Additionally, a function has been implemented that produces a breakdown of the number of stations and their complete years (containing 90% or more complete readings). This shows that just 1,733 stations having near complete readings for all of the years between 1980 and 2010. Table 1 shows details of the station attrition.

Table 1: Historic integrated surface (ISH) hourly station breakdown

		Stations	
All stations (ISH-History)		**29,704**	
Bogus, buoys etc.	-2,090	27,614	
1980-2010:			
No data	-9,035	18,579	=> Number of stations offered in the software
Less than 1 yr of data	-4,488	14,091	
Between 1 and 10yrs of data	-7,351	6,740	
Between 11 and 20yrs of data	-2,934	3,806	
Between 21 and 30yrs of data	-2,073	**1,733**	=> Number of stations containing 90% or more data in each year

Hothaps-Soft provides a number of data analysis functions with regards to data completeness. Before any data is imported into the database, the raw NOAA/GSOD set can be scanned for availability of data. Table 2 shows the analyses for 2 weather stations: Dallas Fort Worth airport shows a near complete set of readings for the specified period (1980 to 2012). The numbers in brackets show the "data density" (percentage of days containing any recordings). Note the completeness value of 99.7% in 2007, caused by one day without data for that year. This is tolerated by the software as it is well above the imposed acceptability threshold of 90%, and the data may be used for statistics, such as calculating averages and for producing complete graphics.

Table 2: Sample output analysis of raw daily NOAA/GSOD data from 2 weather stations: Dallas Airport (near complete) and Mexico City (incomplete)

Analysing NOAA for DALLAS-FORT WORTH/F (722590-03927): 33 year(s), 12053 records (99.9%) 1980(100.0%), 1981(100.0%), 1982(100.0%), 1983(100.0%), 1984(100.0%), 1985(100.0%), 1986(100.0%), 1987(100.0%), 1988(100.0%), 1989(100.0%), 1990(100.0%), 1991(100.0%), 1992(100.0%), 1993(100.0%), 1994(100.0%), 1995(100.0%), 1996(100.0%), 1997(100.0%), 1998(100.0%), 1999(100.0%), 2000(100.0%), 2001(100.0%), 2002(100.0%), 2003(100.0%), 2004(100.0%), 2005(100.0%), 2006(100.0%), 2007(99.7%), 2008(100.0%), 2009(100.0%), 2010(100.0%), 2011(100.0%), 2012(100.0%)
Analysing NOAA for MEXICO CITY (766800-99999): 33 year(s), 7006 records (58.1%) 1980(68.9%), 1981(71.0%), 1982(53.2%), 1983(66.0%), 1984(68.6%), 1985(60.3%), 1986(51.2%), 1987(38.9%), 1988(29.5%), 1989(76.4%), 1990(84.1%), 1991(83.8%), 1992(65.3%), 1993(59.5%), 1994(68.5%), 1995(83.3%), 1996(96.7%), 1997(95.3%), 1998(88.8%), 1999(59.5%), 2000(20.5%), 2001(3.3%), 2002(4.4%), 2003(2.5%), 2004(4.1%), 2005(40.3%), 2006(52.3%), 2007(59.5%), 2008(60.9%), 2009(66.6%), 2010(66.0%), 2011(89.6%), 2012(79.5%)

While many of the major city weather stations, especially those allied to international airports, provide near complete data, some locations present interesting challenges: Downtown Mexico City (Table 2) is an example of a station with highly patchy data. It appears that this station hasn't got a single complete year of data, and only 2 years above the 90% completeness threshold.

If the selected stations have already been imported into the database, the *imported* data is also analysed and the results displayed in the same way as in Table 2. By comparing the NOAA/GSOD and database analyses the user can ascertain how many records (years) have been imported already. This is useful for longitudinal studies: as NOAA is regularly adding new data to their collection, this step helps the user decide whether it is necessary to import newly available data.

Once imported, the user can execute a *Detailed Analysis* to find how many variables above the 90% threshold are available for each month of each year (Table 3). A maximum of 7 variables (dew point, rainfall, maximum, minimum and mean temperature, maximum and mean wind speed) are extracted from GSOD. The numbers shown indicate how many of these 7 variables reach the 90% threshold in each given month (1-12), and for the whole year (first column, labeled *Yr*).

Table 3: Sample output *detailed analysis of data density*:

MEXICO CITY	Month:												
	Yr	1	2	3	4	5	6	7	8	9	10	11	12
1980:	0	0	0	0	0	0	0	0	0	0	0	0	0
1981:	0	0	0	0	0	0	0	0	0	0	0	0	0
…													
1994:	0	0	0	0	0	0	0	0	6	0	0	0	0
1995:	0	0	0	0	0	7	0	7	7	6	7	5	0
1996:	7	7	7	7	7	7	7	7	7	7	7	7	7
1997:	7	7	7	7	7	7	7	7	7	7	7	7	0
1998:	0	0	7	7	7	7	7	7	0	7	7	7	0
1999:	0	7	7	0	0	0	0	0	0	0	0	0	0
…													
2007:	0	0	0	0	0	0	0	0	5	0	0	0	0
2008:	0	0	0	0	0	0	0	0	0	0	0	0	0
2009:	0	0	0	0	0	0	0	0	0	0	0	0	0
2010:	0	0	0	0	0	0	0	0	0	0	7	0	0
2011:	0	0	7	7	0	7	7	7	7	7	7	7	0

No variable with at least 90% data density in any month in these years

A year with a complete data set

Even though the majority of months have sufficient data, there are not enough days with readings to calculate an annual average.

5 of possible 7 values (temp, dew point etc.) have sufficient data. User must view chart to see which.

90% Threshold

A tolerance threshold is required since only a limited number of stations have 100% records. If too much data is missing the annual and monthly data becomes too distorted. After working with datasets of varying degrees of completeness, it was found that, depending on location and clustering of the data loss, up to 10% missing daily data (three days missing in a month) has no substantive effect on averages and trends in most cases.

Monthly averages are only calculated and included in the "Monthly Averages" database table if there are no more than two missing days each month. As a result a graph showing the average November temperatures for a station would have a gap in 1999 for example, if in November of that year less than 27 readings are available (Figure 1).

Annual averages are calculated by averaging a year's monthly averages - irrespective of their level of completeness - after weighting them with the numbers of days in each month. The 90% threshold is applied to annual averages so that annual average data points are stored for plotting or trend analysis only if at least 329 daily recordings are available for that year. Figure 1 shows an annual graph, a graph for the coolest month (November) and the hottest month (May). Gaps are apparent for years and months where available readings fall below the 90% threshold. The monthly gaps in 1982 create an annual gap in the same year, while the November gap in 1999 does not create an annual gap[§].

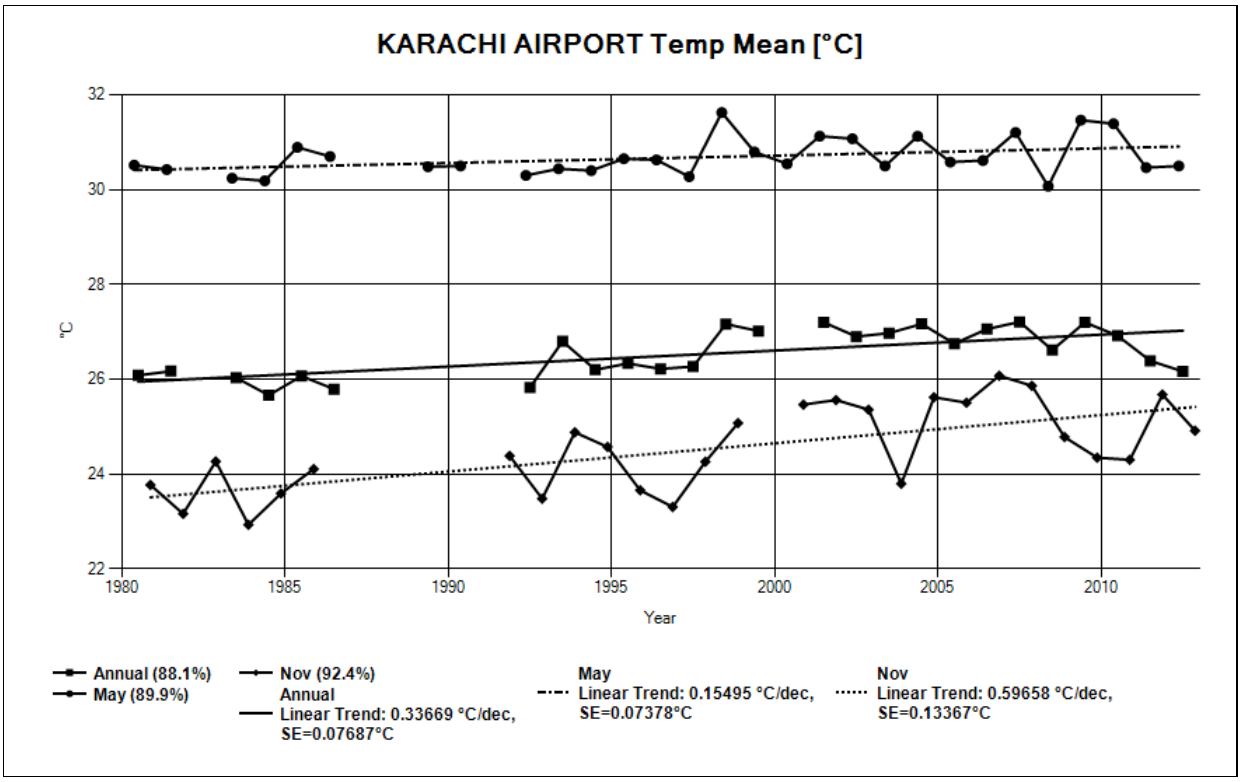

Fig 1: Sample chart with gaps in monthly and annual graphs

WBGT, heat stress index

In addition to filtering and importing raw weather station measurements Hothaps-Soft calculates two heat stress indicators. They are added to the database and can be plotted like the actual weather data readings and statistical calculations performed on them. The widely used Wet Bulb Globe Temperature (WBGT) is calculated upon importing GSOD weather station data. Calculating *outdoor* WBGT requires solar radiation data. While direct solar radiation hitting the planet is well known, actual local radiation is affected by cloud cover, which varies greatly by location and time. Actual solar radiation data has not routinely been recorded by weather stations over the desired 30-year time span, and calculation of outdoor WBGT has therefore been excluded from the scope of this software application in favour of *indoor* WBGT, which corresponds to the human body's experience of heat stress indoors or in full shade.

[§] Note that the X-axis gridlines mark the ***beginning*** of the labelled year.

WBGT normally also requires knowledge of the prevailing *wind speed*. While wind speed is part of the GSOD data set, its availability is often patchy. If the same rigorous criteria (i.e. 90% data density) used for the other readings were applied to wind speed, available WBGT values would be very sparse. For WBGT in the shade a constant wind speed of 1 m/s was used, as this corresponds to the air movement over skin generated by a person moving at walking speed, an assumption that would apply to *working people* involved in continuous physical labour.

If it is assumed that indoors, or full shade, heat exposure does not include radiant heat from the sun (or other source), the variables needed for WBGT calculations are Ta (air temperature) and Td (dew point temperature) (Lemke & Kjellstrom 2012).

Accurate indoor WBGT calculations commonly rely on *hourly* temperature and humidity profiles. Because readily available *daily* station data is used as the source Lemke and Kjellstrom's formula was used to calculate three different WBGT values, based on Tavg (average daily temperature), Tmax (maximum temperature) and the *midway point* between these two. Typical hourly WBGT modelling (Kjellstrom et al. 2013) has shown that WBGTmax corresponds to the WBGT during the four hottest hours (e.g. 12noon to 4pm), WBGTmid to the next hottest four hours (10am-12noon and 4-6pm) and WBGTavg to the third hottest four hours (8-10am and 6-8pm). Our analysis of indoor and outdoor WBGT in selected hot locations (Kjellstrom et al. 2013 and private communication) has shown that in the full sun during hot afternoons WBGT will be approximately 2.5 C higher than the indoor WBGT.

UTCI, heat stress index

The Universal Thermal Climate Index (UTCI) is a heat stress index based on the physiological response of humans to heat in a specific physical activity setting (walking on flat ground at 4 km/hr = approx. 1.1 m/s) (Bröde et al. 2011). The actual index requires sophisticated physiological modelling (Fiala et al. 2001), however, this has been converted to a 200+ term regression formula that is available from the UTCI website (Bröde 2009) which is used by Hothaps-Soft to calculate UTCImax, UTCImid and UTCIavg as outlined in the WBGT section above.

While UTCI is a rational heat index (based on human physiology), WBGT has been tested in real situations and has well determined limits to permissible work under different heat stress conditions (ISO 1989, ACGIH 2009).

RESULTS, GRAPHICAL OUTPUTS

After the user selects a range of stations, values (e.g. temperatures, dew point) and time modes (e.g. monthly, annual or specific months, and/or year ranges respectively) the following outputs can be generated by Hothaps-Soft:

Distribution charts

Box and whisker plots can be produced to show the statistical distribution of the chosen reading(s). The box-plot percentiles can be customised, as can be the display of mean and median. The example chart in Figure 2 gives the data distributions over 2 different year periods, and clearly shows higher mean temperatures for each month and higher variability for certain months (Jan, Jun, Dec) during the recent 5 years compared to 1981-1985.

Text sizes, grid lines, legends and colours can be customised. Chart data can be exported numerically; charts themselves can be saved as PNG or JPG files to be used in report manuscripts.

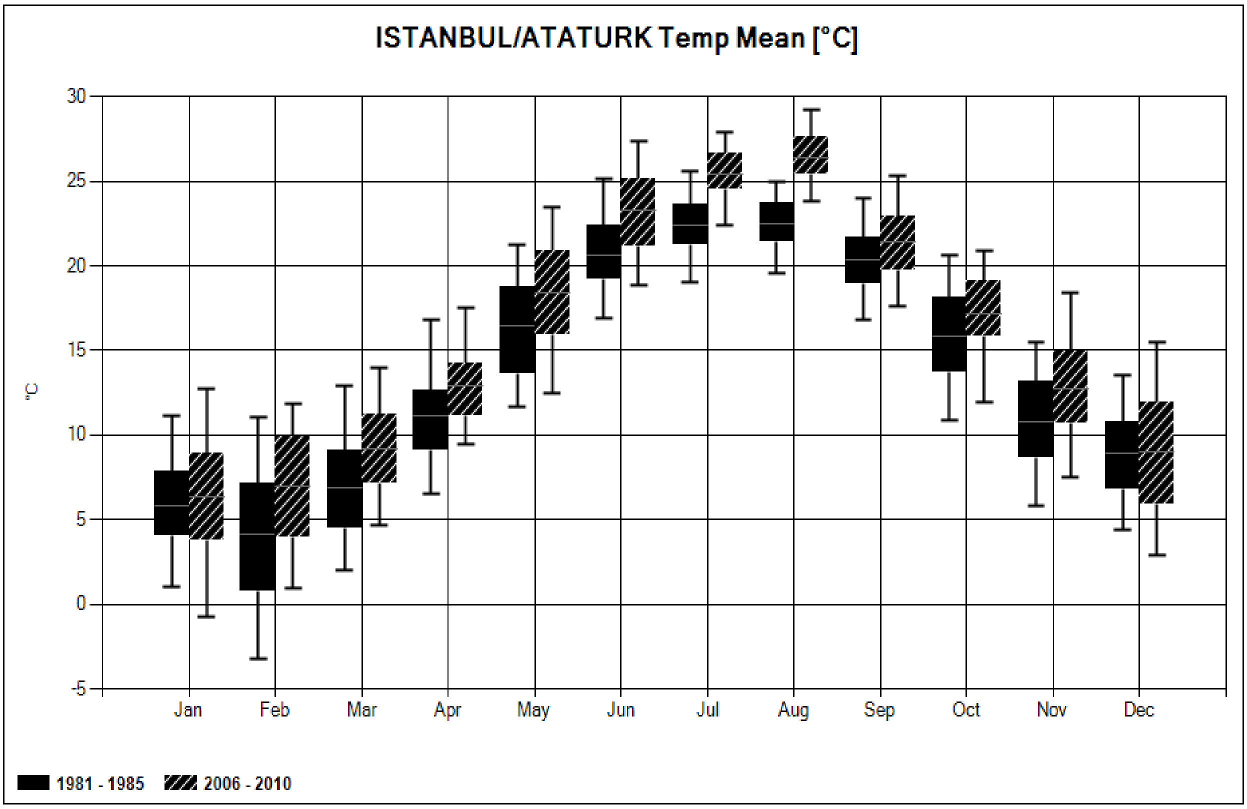

Fig 2 Sample data distribution chart

Time value charts

Figure 1 shows an example of a typical XY-graph of one weather station's time trend for one selected measurement and three time modes (*annual, May* and *November averages*). Any number of stations, measurements and time modes can be combined in a chart.

Fitted trend lines are optional. If selected, linear trend lines are accompanied by slope coefficient and standard error in the legend (Figure 1). Also available are logarithmic and polynomial trends. Trends (regression) are calculated even with incomplete data.

Additionally, CRU-3 graphs (data taken from a grid cell that contains the selected weather station) can be overlaid to serve as a quality reference, and to cover sections where no reliable GSOD data is available (Figure 3). As the CRU reference represents data homogenised over the 0.5x0.5 degree grid cell by including a range of applicable weather stations (Mitchell & Jones 2005), urban heat island effects can be studied when overlaying an urban weather station with gridded CRU data.

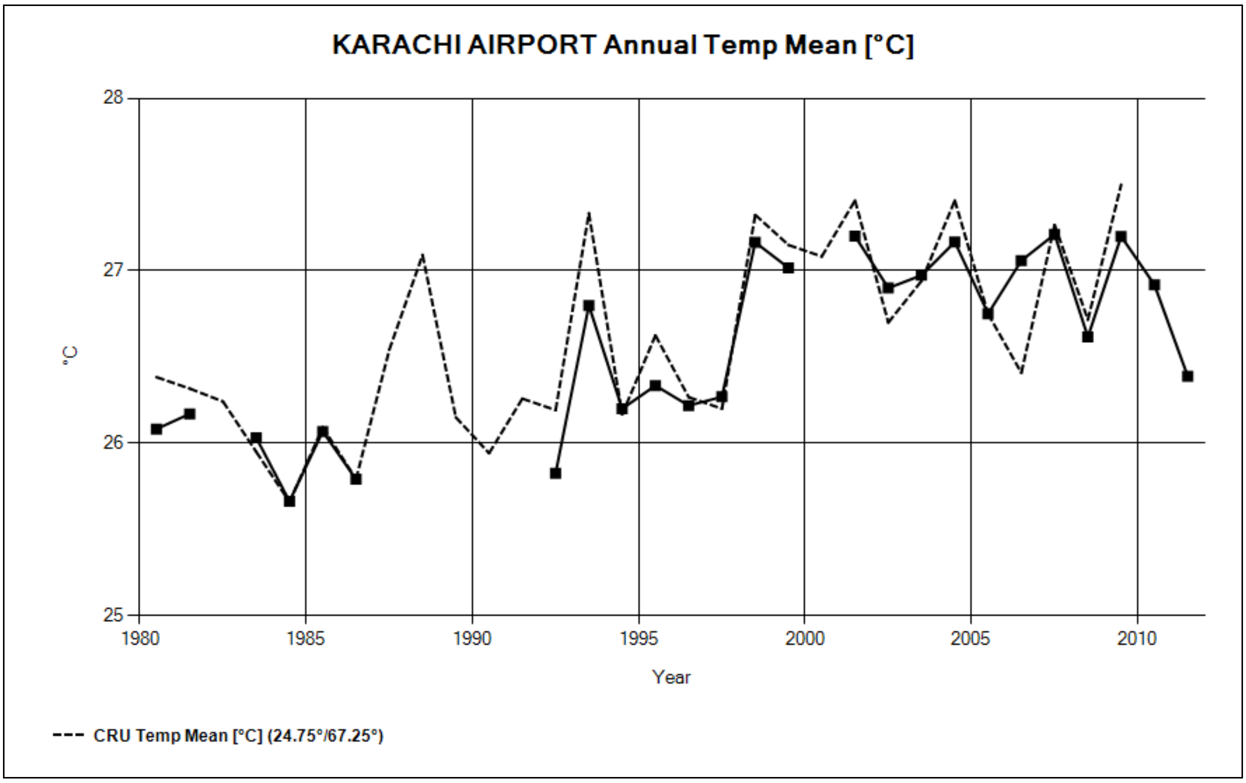

Fig 3 Sample chart including CRU-reference

While airports are usually located away from built-up areas, weather station and CRU grid cell reference will not be too dissimilar (Figure 3). However, Figure 4 contrasts annually averaged maximum temperatures recorded by Melbourne's (Australia) central city weather station with the relevant homogenised gridded data. The urban trend is twice that of the reference grid, and the actual temperatures recorded in the city over the last 10 years are about 1 degree higher than the average for the grid.

Text sizes, grid divisions, legend positions and colours can be customised. Again, charts and their underlying data can be exported graphically and numerically.

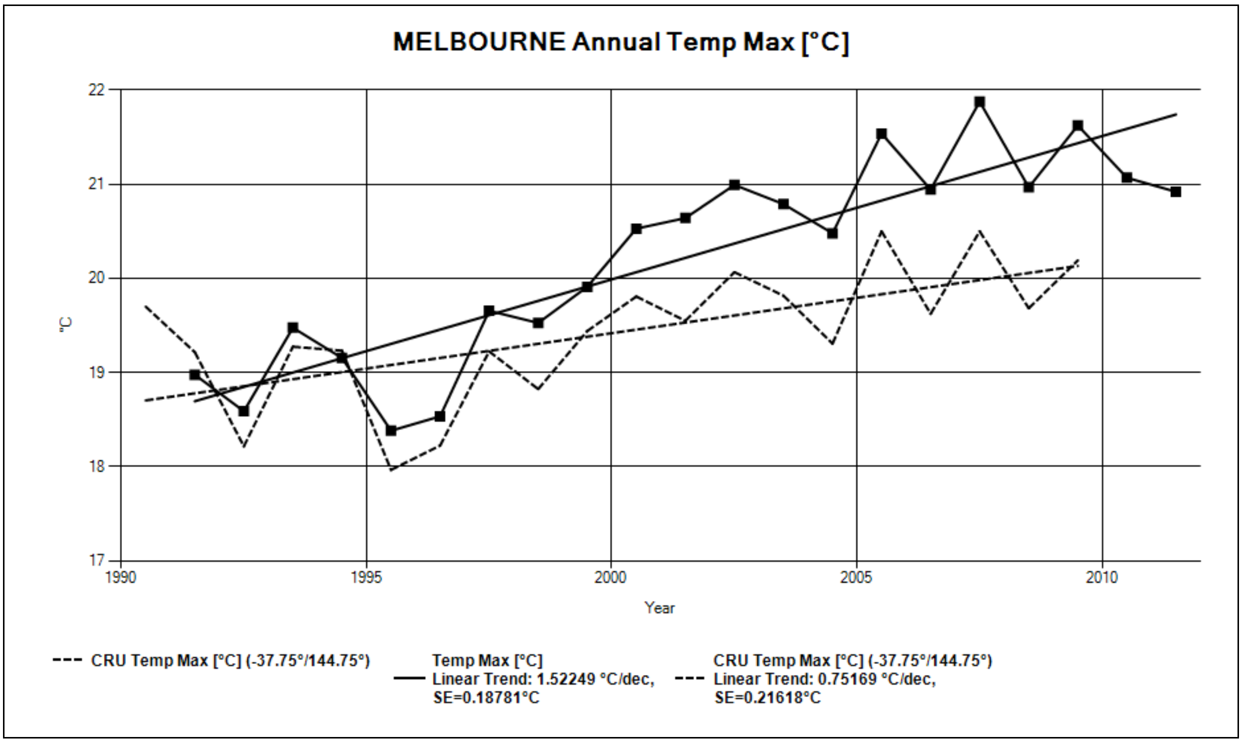

Fig 4 Comparison between an urban station and homogenised grid cell data

Threshold charts

While common climate analysis focuses on average trends, the trends in the tails of a distribution are often much more pronounced. Hence Hothaps-Soft incorporates a *threshold analysis*. Similar to time value charts, customisable threshold charts can be produced, where the Y-axis displays the number of days per year that a station's readings exceed a threshold specified by the user. Figure 5 shows the number of days exceeding 35 °C annually in Athens. The trend line indicates that the number of days with Tmax above 35 °C has increased by nearly 6.5 days per decade (standard error 1.25 days). Unlike the time value charts described above, where annual or monthly *averages* are plotted, threshold charts use *daily data* and allow the identification of weather extremes like heat waves, which are usually masked when observing average values.

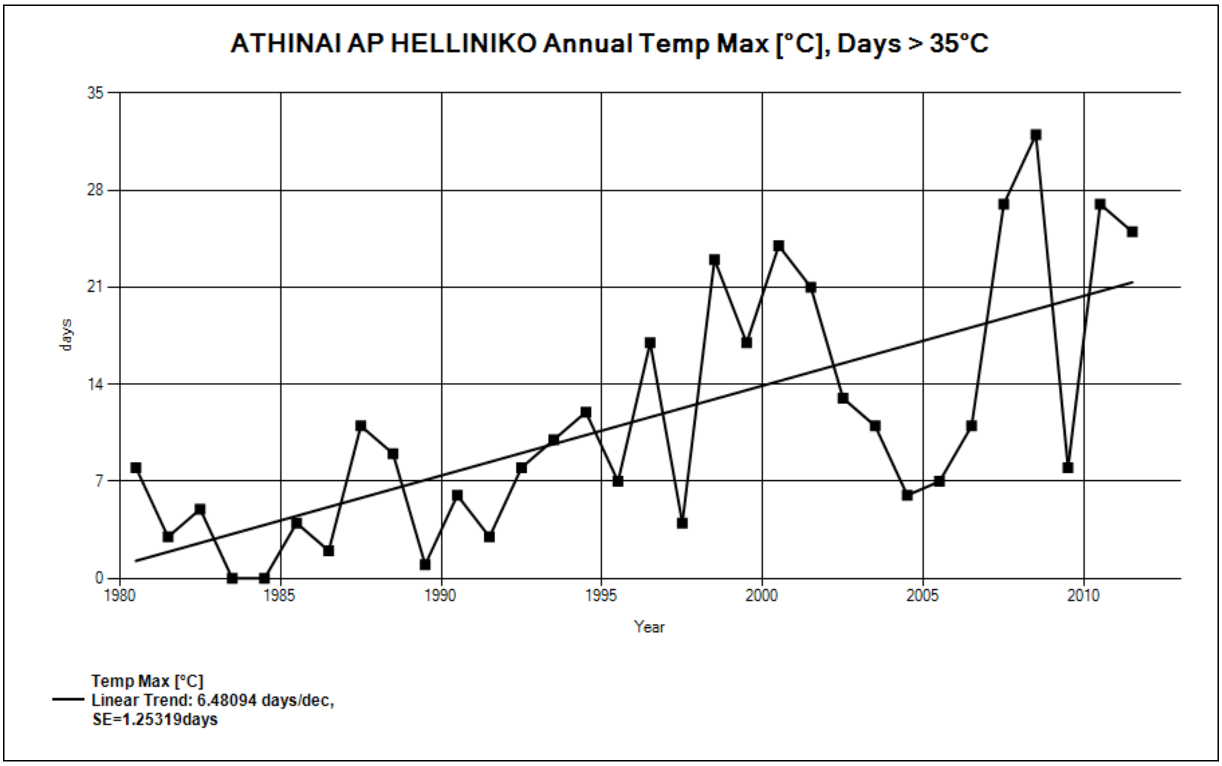

Fig 5 Sample threshold chart

RESULTS, NUMERICAL OUTPUTS

Averages, regressions, residuals

After the selection of stations, values (e.g. temperatures, dew point) and time modes (e.g. annual or specific months), averages, regressions and residuals can be calculated and exported, for example, into a spreadsheet.

Numerical representation of plots and trend lines in graphic outputs

The numerical data underlying any of the charts, including optionally enabled trend lines or CRU reference data, can be exported in the form of comma-delimited text files for further processing using other software tools.

Daily data

The daily data used as the source, i.e. raw measurements and calculated heat indices, can be exported in the form of comma-delimited text files for further processing using other software tools.

DISCUSSION

The relationship between GSOD data and the quality control processed *Global Historical Climatology Network* dataset (GHCN) also available from NOAA, has been summarised in the YALE forum on Climate Change & the Media (Hausfather 2010). GSOD covers more stations than GHCN since 1973, and *significantly* more stations from 1992. When temperatures are compared, GSOD tends to run colder (i.e. more conservative in climate-change terms) than equivalent GHCN land temperatures. While Hausfather recommends caution when interpreting GSOD data, he concludes that raw GSOD data are closely in line with the major GHCN-based land temperature series.

This discussion is significant for this software project because the GHCN dataset doesn't include humidity data, fundamental to any valid heat stress index calculations. It has also been pointed out that GHCN suffers from changes in the distribution characteristics that result from site location and urbanisation adjustments (Guttman 1996). Also, GHCN stations have been deliberately chosen to minimise any urban heat island effects and these are precisely the values researchers are trying to capture to make meaningful predictions about places where most people live and work.

This software was developed as a contribution to the tools available for the HOTHAPS global research programme on how heat exposure affects working people (Kjellstrom et al. 2009b). It has been tested in different locations, and data and charts resulting from Hothaps-Soft have already been published (Kjellstrom et al. 2011 2012a 2012b 2013).

The rapid access to these large historic climate data sets is considered a major advantage for the various teams involved in HOTHAPS (Kjellstrom et al. 2009b). The ability to compare actual weather station data with grid cell based estimates from the CRU is a valuable feature, as it provides a rapid method to assess the likely validity of the climatic conditions indicated by the weather station in question.

There are a number of tools and software products that allow the extraction and graphic representation of weather station data. The National Climatic Data Center (NCDC) at NOAA permits users to view selected variables and stations contained in the GSOD dataset free of charge. The web-based interface produces line graphs based on otherwise unprocessed daily records. The user can adjust the X-axis resolution (entire period, one year or one month) and scroll through the time period. This tool is useful to ascertain which data points are available, and possibly to locate extreme weather events. It gives immediate access to all weather stations catalogued by NOAA, and charts can be printed. However, it offers only very limited choices to customise and combine graphs, has no statistical features like trend lines, averages, distributions etc., and no heat stress indices.

The software project MeteoInfo (Wang 2012) accepts spatial (gridded) meteorological input in a variety of formats. It allows viewing and analysing interactively. The outputs are generally maps, offering GIS functionality specialised for weather/climate data use. In addition to the free software package, the author freely distributes class libraries[**] for use by other software developers. This software tool allows some excellent geographical view of meteorological data. It doesn't seem to include statistical or heat stress related functionality. It is conceivable that with appropriate pre-processing of gridded data MeteoInfo might be able to produce spatial views of heat stress indices.

A Windows toolkit was has been developed for providing a wide variety of mainly statistical functions that can be performed on palaeo-environmental data, published under C2 (Juggins 2010). Data can be imported and exported in a number of different formats. The package includes modelling tools that derive transfer functions from a combination of datasets. Data can be analysed and viewed in a spreadsheet-like interface as well as in a number of chart types. C2 appears to be a specialised set of tools suitable for processing, analysing and viewing climate data sets and models. With a capacity of 20,000 observations (rows) it may well be suitable to analyse for example GSOD station data. The mathematical and statistical features of this program exceed that of the software described here, which has been primarily designed for ease of use and for initial analysis of climate conditions and trends at a specific location.

The popular Gapminder software (Rosling 2005) is significant and different from any other tool as it allows advanced dynamic presentation of various health, socio-economic and environmental data for countries. Sub-country data are also available for some countries. It presents time trends in the form of animated coloured bubbles that move spatially and change size. The graphics are a very effective medium to demonstrate statistical variables as they develop over time. Rosling has given many stimulating lectures in his field of world ecology, population, wealth and health using this tool. While climate and heat stress data can potentially be used as input, it has to be noted that the data needs to be prepared and processed before it is suitable for presentation in Gapminder. This software is not a

[**] A class library is a set of software components giving the developer access to features and functions.

replacement for graphic outputs required in printed publications and will not quantify time trends, provide data reliability information or other common statistics.

There are software tools for both, online and offline use, all suited to different specific areas in the wider field of analysing and presenting climate data. The software presented in this paper covers specific aspects of the needs in the climate and health research arena, which are not easily available with other software. This includes analysing raw data numerically in several ways, presenting data in customisable, publishable graphical formats, calculating heat stress indices, making all visual and calculated outputs available in numeric form for further processing, and executing Hothaps-Soft on a minimal hardware platform without installation or the need for an Internet connection (as dictated by some locations of the developing world).

Hothaps-Soft has been documented by way of a 3-level manual set (introduction/features, manual and technical reference). It has been tested among a group of researchers and is currently undergoing beta testing. Thus far, it has been well received by users. Access to Hothaps-Soft may be gained via the development team's website www.climatechip.org

CONCLUSIONS

The work described here is a snapshot in the process of adapting a software tool designed to meet the needs of researchers in the field of climate data observations mainly focussed on human health impacts. It allows a rapid analysis and importation of weather station and global grid cell data, and produces versatile graphic and numeric outputs. Small memory devices make the software highly portable, able to be fitted and run on a single 8GB flash-drive, including the database as well as the raw input data.

While the software meets the current needs of the HOTHAPS research group (Kjellstrom et al. 2009b), further improvements and additions are envisaged. Some usability issues have already been identified and it is likely that more will be found by the beta-testing community. In addition, there are suggestions for new features and outputs as research methods progress. The internal architecture of Hothaps-Soft is designed to accommodate changes and ongoing development.

REFERENCES

ACGIH (2009) TLVs for chemical substances and physical agents & Biological exposure indicators. Cincinatti, American Conference of Government Industrial Hygienists.

Bröde P, Fiala D, Blazejczyk K, Holmér I, Jendritzky G, Kampmann B, Tinz B, Havenith G (2011) Deriving the operational procedure for the Universal Thermal Climate Index UTCI. International Journal of Biometeorology: Special issue (UTCI)

Bröde P (2009) FORTRAN code of the UTCI formula www.utci.org/public/UTCI Program Code/UTCI_a002.f90 (accessed May 2013)

CRU-3, Climatic Research Unit at the University of East Anglia, CRU-3 download site: badc.nerc.ac.uk/browse/badc/cru/data/cru_ts_3.10 (accessed October 2012)

Fiala D, Lomas K, Stohrer M (2001) Computer predictions of human thermoregulatory and temperature responses to a wide range of environmental conditions. International Journal of Biometeorology 45, 143-59

GHCN (2012) Global Historical Climatology Network. National Atmospheric and Oceanographic Administration, USA: http://www.ncdc.noaa.gov/oa/climate/ghcn-daily (accessed May 2013)

GSOD (2012) Global Summary of the Day. National Atmospheric and Oceanographic Administration, USA: www7.ncdc.noaa.gov/CDO/dataproduct (accessed May 2013)

Guttman NB (1996) Statistical Characteristics of U.S. Historical Climatology Network Temperature Distributions, Climate Research: www.int-res.com/articles/cr/6/c006p033.pdf (accessed May 2013)

Hausfather Z (2003) YALE forum on Climate Data and the Media. Yale University Forum. www.yaleclimatemediaforum.org/2010/08/an-alternative-land-temperature-record-may-help-allay-critics-data-concerns (accessed May 2013)

Hickman L, Randerson J (2009) Climate sceptics claim leaked emails are evidence of collusion among scientists. The Guardian. www.guardian.co.uk/environment/2009/nov/20/climate-sceptics-hackers-leaked-emails (accessed May 2013)

IPCC (2007) Climate Change 2007: The physical science basis. Summary for policymakers. Contribution of working group I to the fourth assessment report. Geneva: The Intergovernmental Panel on Climate Change.

ISO (1989) Hot environments - Estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature). ISO Standard 7243. Geneva: International Standards Organization.

Juggins S (2010) C2 Windows program, Newcastle University, www.staff.ncl.ac.uk/staff/stephen.juggins/software/C2Home.htm (accessed May 2013)

Kjellstrom T, Holmer I, Lemke B (2009a) Workplace heat stress, health and productivity - an increasing challenge for low and middle-income countries during climate change. Global Health Action 2: Special Volume, pp. 46-51. (DOI: 10.3402/gha.v2i0.2047)

Kjellstrom T, Gabrysch S, Lemke B, Dear K (2009b) The "HOTHAPS" program for assessment of climate change impacts on occupational health and productivity: An invitation to carry out field studies. Global Health Action: www.globalhealthaction.net. DOI 10.3402/gha.v2i0.2082.

Kjellstrom T, Lemke B, Otto M (2011) Climate change, occupational heat stress and impacts on health and productivity in Africa. African Newsletter on Occupational Health and Safety, Dec 2011, 21, 44-47.

Kjellstrom T, Lemke B, Otto M, Dear K (2012a, in press) Global assessment of the health impacts of climate change: Occupational heat stress. Technical report for the WHO.

Kjellstrom T, Lemke B, Otto M (2013) Mapping occupational heat exposure and effects in South-East Asia: Ongoing time trends 1980-2009 and future estimates to 2050. Industrial Health 51: 56-67.

Kjellstrom T, Otto M, Lemke B (2012b, in press) Manual for the Hothaps-Soft software for analysis of local climate conditions based on existing weather station data. Technical Report, HEIT, New Zealand.

Lemke B, Kjellstrom T (2012) Calculating workplace WBGT from meteorological data. Industrial Health, 50, 267-278.

Lott JN (2004) The quality control of the integrated surface hourly database. 84th American Meteorological Society Annual Meeting, 2004, Seattle, WA, American Meteorological Society, Boston, MA, 7.8 (7p.)

Microsoft .NET. A comprehensive and consistent programming model for building applications www.microsoft.com/net (accessed May 2013)

Mitchell S (2009) Microsoft Chart Controls, Visual Studio Automatic Code Snippets and more msdn.microsoft.com/en-us/magazine/dd569763.aspx (accessed May 2013)

Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology, 25, pg 693-712

MySQL. The world's most popular open source database. mysql.com

NCDC. National Climatic Data Center, US Department of Commerce, National Oceanic and Atmospheric Administration, datasets in numerical and graphical format: www7.ncdc.noaa.gov/CDO/cdoselect.cmd (accessed May 2013)

NOAA. National Oceanic and Atmospheric Administration: http://www.noaa.gov/

Oke TR (1973) City size and the urban heat island. Atmos Environ 7: 769-779

Parsons K (2003) Human thermal environments. The effects of hot, moderate and cold temperatures on human health, comfort and performance. 2nd edition. Taylor & Francis London. ISBN 0-415-23793-9

Rosling H (2005) Gapminder. www.gapminder.org

UTCI. Universal Thermal Climate Index. www.utci.org

Wang YQ (2012) MeteoInfo. GIS software for meteorological data visualization and analysis, Wiley Online Library DOI: 10.1002/met.1345 onlinelibrary.wiley.com/doi/10.1002/met.1345/abstract, www.meteothinker.com (accessed May 2013)

WBGT. Wet bulb globe temperature index: www.hse.gov.uk/temperature/heatstress/measuring/wetbulb.htm (accessed May 2013)