Creating surface temperature datasets to meet 21st Century challenges
Met Office Hadley Centre, Exeter, UK
7th-9th September 2010

White papers background

Each white paper has been prepared in a matter of a few weeks by a small set of experts who were pre-defined by the International Organising Committee to represent a broad range of expert backgrounds and perspectives. We are very grateful to these authors for giving their time so willingly to this task at such short notice. They are not intended to constitute publication quality pieces – a process that would naturally take somewhat longer to achieve.

The white papers have been written to raise the big ticket items that require further consideration for the successful implementation of a holistic project that encompasses all aspects from data recovery through analysis and delivery to end users. They provide a framework for undertaking the breakout and plenary discussions at the workshop. The IOC felt strongly that starting from a blank sheet of paper would not be conducive to agreement in a relatively short meeting.

It is important to stress that the white papers are very definitely not meant to be interpreted as providing a definitive plan. There are two stages of review that will inform the finally agreed meeting outcome:

1. The white papers have been made publicly available for a comment period through a moderated blog.
2. At the meeting the approx. 75 experts in attendance will discuss and finesse plans both in breakout groups and in plenary. Stringent efforts will be made to ensure that public comments are taken into account to the extent possible.

Climate Data Policy

Albert M.G. Klein Tank (KNMI), Philip D. Jones (UEA), Thomas C. Peterson (NCDC/NOAA).

Draft white paper for discussion at the international workshop:
"Creating surface temperature datasets to meet 21st Century challenges",
Met Office Hadley Centre, Exeter, UK, 7th-9th September 2010

Remit

This white paper discusses data policy issues pertaining to:

1) the collection of high resolution observations (both the raw daily and sub-daily data and the metadata) into a single databank for the GCOS-defined Essential Climate Variable surface air temperature over land;
2) the development and dissemination of derived, value-added datasets from this temperature databank, including gridded data products and so-called Climate Data Records in which the time dependent biases have been removed.

Current state-of-the-art

Existing worldwide surface air temperature datasets used for regular monitoring of long-term climate change, such as those from CRU and NOAA-NCDC[1], are currently restricted to monthly averages for a subset of the possible stations only. Much of the initial data collection for these datasets stems from World Weather Records, a major climate data project initiated in 1923 and still updated every decade under the auspices of World Meteorological Organization (WMO; Le Truet et al., 2007). World Weather Record data are supplemented with data from dozens of other sources (see Peterson and Vose (1997) for a sample list). For near real-time data, these monthly datasets rely on CLIMAT messages, which are exchanged over the WMO's Global Telecommunications System (GTS) amongst National Meteorological and Hydrological Services (NMHSs). As a result, both datasets share much but certainly not all of the underlying archive. Indeed, sometimes their data for a given station are different. For example, the source for one dataset may be monthly means calculated as the average of 3-hourly observations while the other dataset used the mean of average daily maximum and minimum temperatures. The difficulties caused by different ways to calculate mean monthly temperatures has long been known (e.g., Jones et al., 1985). It is important to notice that neither the CRU archive nor the NOAA-NCDC archive has the original observations that were made, but rather just possess the calculated monthly mean temperature values.

A single, comprehensive databank holding for all available observations of surface air temperature, as originally recorded at weather stations around the world, does not, at present, exist. Given the changes in observing and recording practices over the years (particularly over the last 20 years), considerable intellectual, financial and diplomatic resources would be required to collect all available historical information, including the original source data. Even for modern data, the prompt and regular flow to the international data centres is currently inadequate (GCOS, 2010; see also Figure 1). Lack of engagement, prevalence of short-term

[1] the third, well known global temperature dataset from NASA-GISS makes use of the NOAA-NCDC data for the land surface part

research funding or overall lack of resources, and inadequately integrated data-system infrastructures are among the causes. But the main reason for the lack of a comprehensive temperature databank is most likely related to the difficulty caused by data policy issues of many NMHSs. Such data policy issues are pertinent for both the historical data and modern data.

International datasets for daily temperature data do exist but have less coverage in both time and space than monthly data. The GHCN-Daily archive at NOAA-NCDC currently has data from 44,010 stations but only 23,000 have temperature data and many station records cover only a few decades. Furthermore, this high number is dominated by contributions of essentially entire national archives from a few countries (see Figure 2). Yet, there is a potential to dramatically increase these data. Daily data generally reside in national archives which are under the authority of individual NMHSs. Besides the data policy issues that restrict data exchange, with daily data facing more restrictions than monthly data, in many cases much of the daily data have not yet been digitized and reside only on paper archives (see the White paper on recovery of historical data).

Higher resolution sub-daily data are also exchanged over the GTS in the form of SYNOP and other messages, but the quality and completeness of these messages, which are exchanged for weather forecasting purposes, limit their applications for climate research. Yet, some groups take advantage of the SYNOP messages and temporarily use them as near real-time supplements to the climatic time series under the condition that they will be replaced at a later stage. International datasets for higher quality sub-daily temperature data collected from the national archives of individual NMHSs have yet to be developed.

Theory and practice

In theory, WMO Resolution 40[2] on the free exchange of data produced by the NMHSs states that: “As a fundamental principle …, WMO commits itself to broadening and enhancing the free and unrestricted international exchange of meteorological and related data and products”. In recognition of the increased demand, the WMO has issued many requests to member states since the resolution passed in 1995, asking them to send their daily data to the international data centres so that the information may be made freely available for research and operational use.

However, in practice, there are still large obstacles to data being accessible to scientists. Even with the formal arrangements for international data exchange in place, there is still a lack of data in the international repositories and, moreover, for some of these data restrictions are imposed by the data providers which may limit accessibility. Often, the data acquired have come with the restrictions that the data are used only for academic purposes and are not passed onto third parties. The development of a subset of available station data in the public domain, such as the GCOS Surface Network (GSN), was an important step since the GSN stations were identified as the world’s best stations for long-term climate monitoring (Peterson et al., 1997). However, the network is limited to approximately 1000 stations worldwide (out of an estimated total of two orders of magnitude more stations), and the GSN archive of daily records is incomplete despite years of concerted efforts to obtain

[2] Res40Cg-XII WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities and its annex 4: definitions of terms in the practice and guidelines; available from http://www.wmo.int/pages/about/Resolution40_en.html

contributions to it (Figure 3). Even the historic monthly records for many GSN sites are not as complete as they should be and could be.

Many NMHSs charge for data

The station network for near-surface climate observations is managed by a large number of (predominantly) NMHSs, each of which has its own data archive and distribution policy. Many NMHSs impose conditions and charge a fee for access. The problem arises from the need to be, or aim to be, cost neutral i.e., the need to sell the data in order to recoup the costs of making observations and preparing the data. In many countries, the NMHSs are made to cover part of their costs by their respective national governments. Some NMHSs are on exceptionally tight budgets; so tight, in fact, that should they not receive revenues from the sale of their data, they might not have the resources to take and process the observations in the first place. Even if scientists have the funds to buy the data for a particular project, regulations usually prevent copies from being included in the international databases for future use. Often, the data has to be repurchased for a subsequent project. So a transparent comprehensive international database of surface temperature observations has complex financial implications in many parts of the world.

Many NMHSs are unaware of the importance of their data for climate change research and policy. At present, a clear disconnect exists between NMHSs and international institutions such as the IPCC, SUBSTA, UNFCCC and related bodies such as GCOS and the Group on Earth Observations (GEO). UNFCCC advocates open access to data and the GEO has set principles for promoting the free and open access to existing databanks in accordance with set principles. However, these institutions do not enforce anything and leave open the possibility to charge for data by stating that national politics and legislation should be recognized. GCOS does monitor the progress in contributing to international data archives, but GCOS is not in a position to impose sanctions if countries fail to co-operate. The disconnect between NMHSs and international climate change research/policy is illustrated by the so-called Oslo Declaration which has been issued by the directors of the NMHSs in Europe in 2009[3]. They state: “Recognizing the different funding policies associated with different economic models for NMHSs and associated different official mandates, the directors of the NMHSs in Europe have reached consensus recently on progressive expansions of their set of *Essential* data made available on a free and unrestricted basis”. However, there is no mention of the international requirements for climate, the directors don’t specify whether (sub-)daily data are part of the free set, and they leave open the possibility for continuing to license as appropriate.

Possible alternatives

Limiting an international databank to include only the data that comes without restriction would, in turn, exclude vast amounts of data. For example, at the moment roughly 40% of the daily data in Europe and estimated larger percentages in Asia, Africa and South America would be unavailable for use. In many value-added datasets, such as gridded temperature datasets, it is important to use as much station data as possible to fully characterise global- and regional-scale changes. Hence, restricting these products to only including station data that can be freely exchanged would be detrimental to the products in many parts of the world

[3] see: http://www.epsiplatform.eu/news/news/wmo_ra_vi_considers_private_sector

(Figure 4) while including them would limit the derived information’s transparency and reproducibility. In recent years, the data policy has changed in some countries and several NMHSs now provide additional climate data through their websites. However, these are often difficult to use since many data series refer to national numbering systems that must be related back to WMO Station Identifiers, and in addition, the metadata are not standardized and are either missing or in a non-English language. Metadata are particularly important, as, for example, it is vital to know if the data have been adjusted for well-known inhomogeneities.

In an attempt to circumvent data policy issues, a series of hands-on workshops (see Peterson and Manton, 2008) have been held in diverse regions of the world to produce analyses of trends in extremes. Participants brought their daily station data which was quality-controlled and analyzed as part of the workshop. Often these workshops were the first time individuals had analyzed how the climate in their country was changing. The NMHSs involved are generally happy to release the derived indices from their data which were used in the analysis, even if they restrict access to their digital climate archives. As this work was coordinated by a WMO Expert Team (Klein Tank et al., 2009), the analyses for each part of the world could fit together seamlessly[4]. This implies that though the original sources cannot be released, the information may be used to develop derived international datasets (e. g., see Caesar et al., 2006). The workshops also support the development of local climate services. However, in following this avenue, it must be accepted that the derived datasets will not be traceable to the primary data source. This workshop series have served climatology well, but must be rerun every few years to access the latest data. Subsequent workshops provide continuity (for NMHSs enhancing capacity building) but getting resources a second or third time is even more difficult than obtaining resources for the first workshop in an area.

A more sustainable activity has been set up in Europe[5], which is now being implemented for the Indonesian region too. The lessons learned from this project, in which the NMHSs and some universities from over 60 countries in the region co-operate, show the necessity for combining data collection, archiving, quality control, analysis and dissemination in an operational process. This goes beyond setting up and populating a centralised databank alone. An end-to-end approach, in which data providers are engaged in the construction of value-added products such as daily gridded datasets and user oriented indices products, makes it easier to overcome access restrictions to the original data. Also, the necessary knowledge about the procedures and circumstances under which the observations have been made resides locally. Although not all raw data can be passed onto third parties, this European initiative is open about the details of the non-public data used in the derived products and about whom to contact to obtain the underlying raw data. Establishing close co-operation between data providers and scientists, and jointly analysing the data and developing value-added products, will increase the chances of success, also for a global high resolution surface air temperature databank.

Recommendations

[4] see: http://www.clivar.org/organization/etccdi/etccdi.php
[5] see: http://eca.knmi.nl

1. Enhance data availability
    a. Build a central databank in which both the original temperature observations as well as multiple versions of the value-added datasets, i.e., quality controlled, homogenized and gridded products, are stored and documented together (including version control). The opportunity to repeat any enhanced analysis should exist. Not only will the methods used for adding value change over time and between scientists, but the data policy will change as well.
    b. Provide support for digitization of paper archives wherever they may exist with the proviso that any data (and metadata) digitized under this program be made available to the central databank.
    c. Enhance the international exchange of climate data by linking this activity to joint projects of global and regional climate system monitoring and by promoting the free and open access of existing databanks in accordance with set principles, e.g., those of the GEO.

2. Enhance derived product availability
    a. Accept that there is a trade off between transparency and data quantity used for derived products. Transparency and openness, which scientists (including the authors) advocate, are hampered by the data policies of national governments and their respective NMHSs. Data policy issues are persistent and unlikely to change in the near future.
    b. Hold a series of workshops to homogenize data and produce a gridded dataset. The original and adjusted data might not be able to be released but the gridded dataset and information on the stations that contributed to each grid box value would be released. These gridded datasets could be used by NMHSs to monitor their climate and fit together seamlessly into a global gridded dataset.
    c. Ensure that the datasets are correctly credited to their creators and that related rights issues on the original data and the value-added products are observed and made clear to potential users. The conditions will be different for bona fide research and commercial use of data.

3. Involve NMHSs from all countries
    a. Acknowledge that involvement of data providers (mainly NMHSs) from countries throughout the world is essential for success, and involves more than simply sending the data to an international data centre. For all nations contributing station records to benefit from this exercise, the scientific community needs to also deliver derived climate change information which can be used to support local climate services by the NMHSs. This return of investment is of particular importance for developing countries.
    b. Adopt an end-to-end approach in which data providers are engaged in the construction and use of value-added products, not only because it is at the local level where the necessary knowledge resides on the procedures and circumstances under which the observations have been made, but also because this will make it easier to overcome access restrictions to the original data.
    c. Increase the pressure on those countries not inclined to follow a more open data policy by engaging with institutions widely beyond the

community of research scientists, including funding bodies, the general public, policy makers and international organisations.

Acknowledgement

We thank … (NCDC) and Else van den Besselaar (KNMI) for producing the figures.

References

Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded dataset, J. Geophys. Res., 111, D05101, doi:10.1029/2005JD006280.

GCOS, 2010: Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC, 2010 Update, DRAFT v1.0, 13 November 2009 (available from www.wmo.int/pages/prog/gcos/documents/GCOSIP-10_DRAFTv1.0_131109.pdf).

Jones PD , Raper SCB, Cherry BSG, Goodess CM, Wigley TML, Santer B, Kelly PM, Bradley RS, Diaz HF, (1985) TR022 A Grid Point Surface Air Temperature Dataset for the Northern Hemisphere. Office of Energy Research , Carbon Dioxide Research Division, US Department of Energy. Under Contract No. DE-ACO2-79EV10098.

Klein Tank, A.M.G., F.W. Zwiers and X. Zhang, 2009: Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. WMO-TD No. 1500, 56 pp.

Le Treut, H., Somerville, R., U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson and M. Prather, 2007: Historical Overview of Climate Change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Peterson, T.C., and M.J. Manton, 2008: Monitoring changes in climate extremes - A tale of international collaboration. BAMS, DOI:10.1175/2008BAMS2501.1.

Peterson, Thomas C. and Russell S. Vose, 1997: An overview of the Global Historical Climatology Network temperature data base. *Bulletin of the American Meteorological Society*, **78**, 2837-2849.

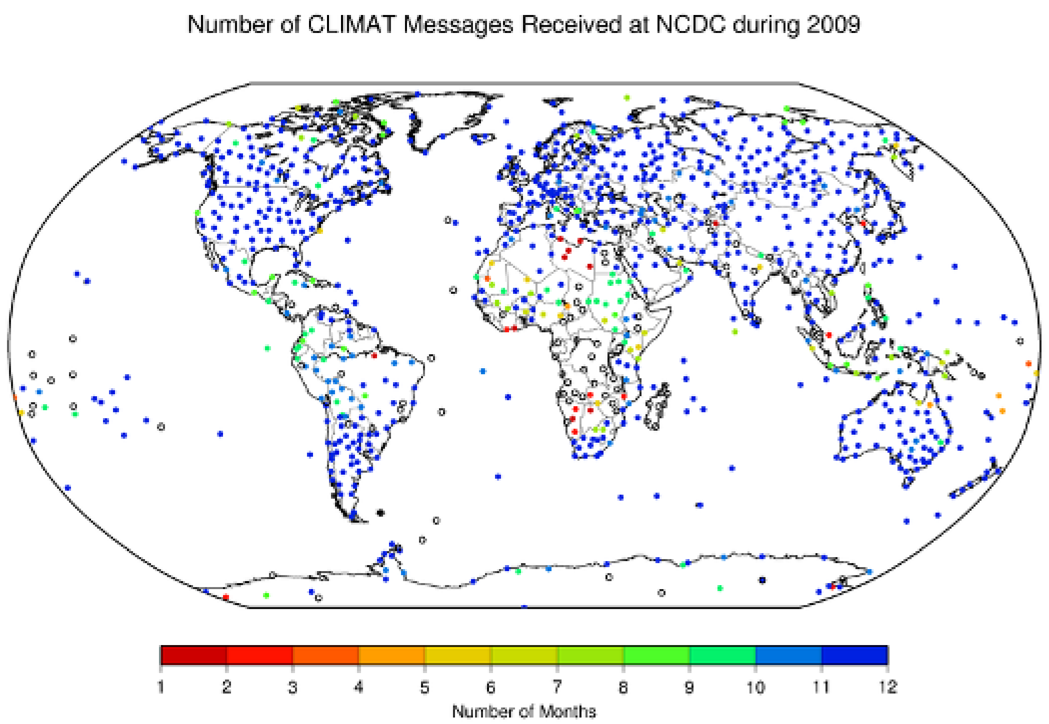

Figure 1. Number of CLIMAT messages received at NCDC in 2009 from GCOS Surface Network stations (black open circles indicate no message received/no information available). Large portions of the world transmit few CLIMAT messages from their GCOS Surface Network stations. Monthly CLIMAT data are available from http://cdo.ncdc.noaa.gov/pls/plclimprod/cdomain.DS3500

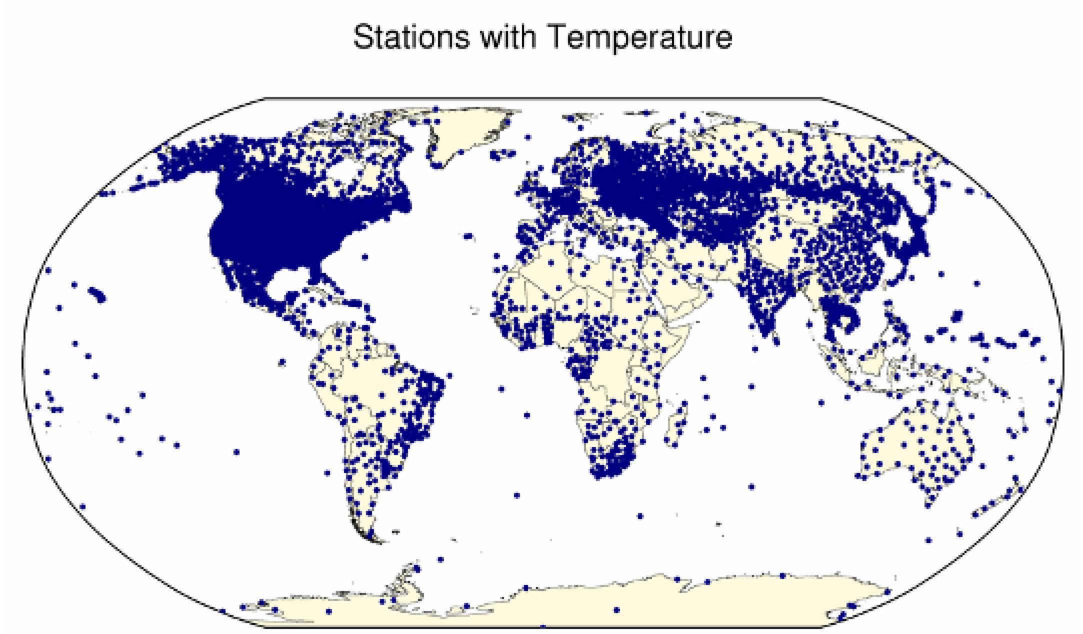

Figure 2. Map showing the location of the 23,000 GHCN-Daily stations with daily temperature data. An additional 20,000 GHCN-Daily stations are available that only have daily precipitation data. This high number is dominated by contributions of essentially entire national archives from a few countries. GHCN-Daily data are available from http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/index.php

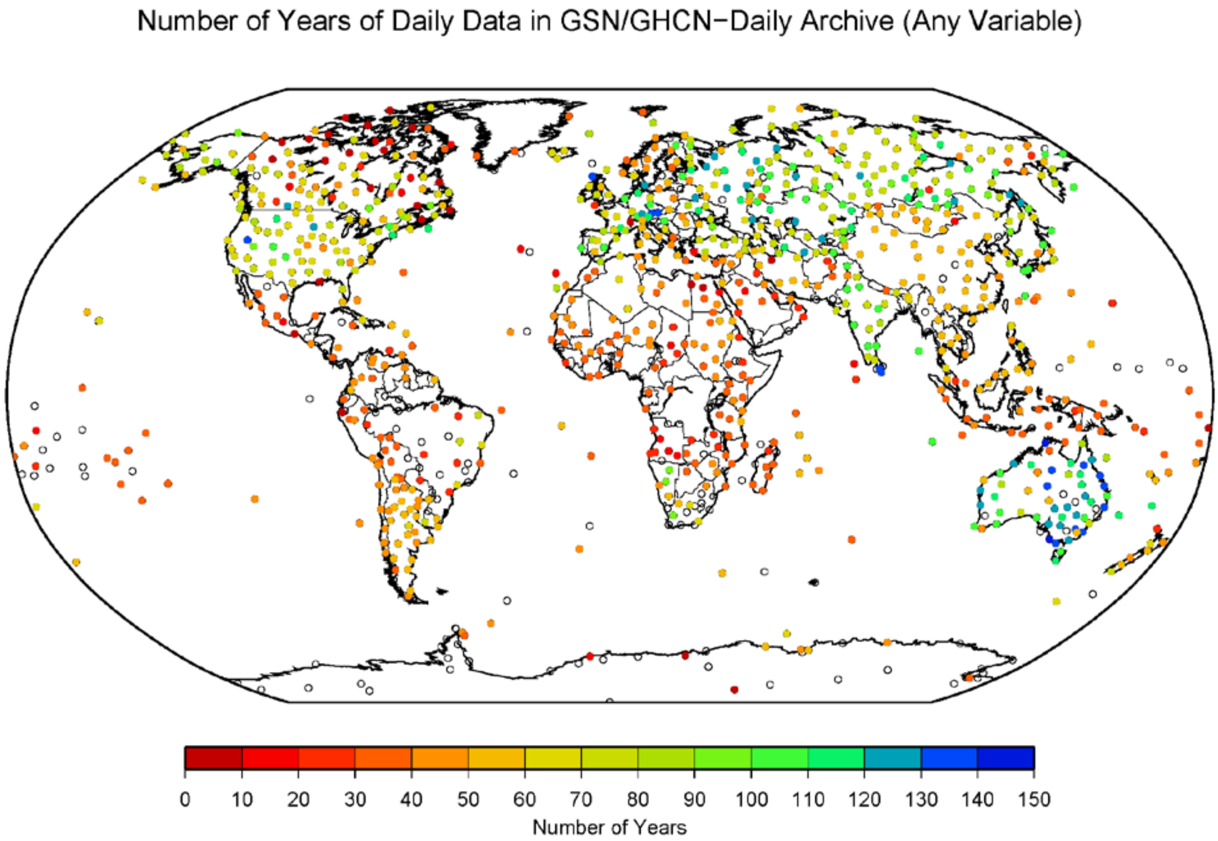

Figure 3. Number of years of daily data from GSN stations in the GHCN-Daily archive as of March 2010. GSN stations have been selected on the basis of available long-term observation series, but the GSN archive is incomplete despite years of concerted efforts to obtain contributions to it. Daily GSN data are available from http://gosic.org/gcos/GSN-data-access.htm

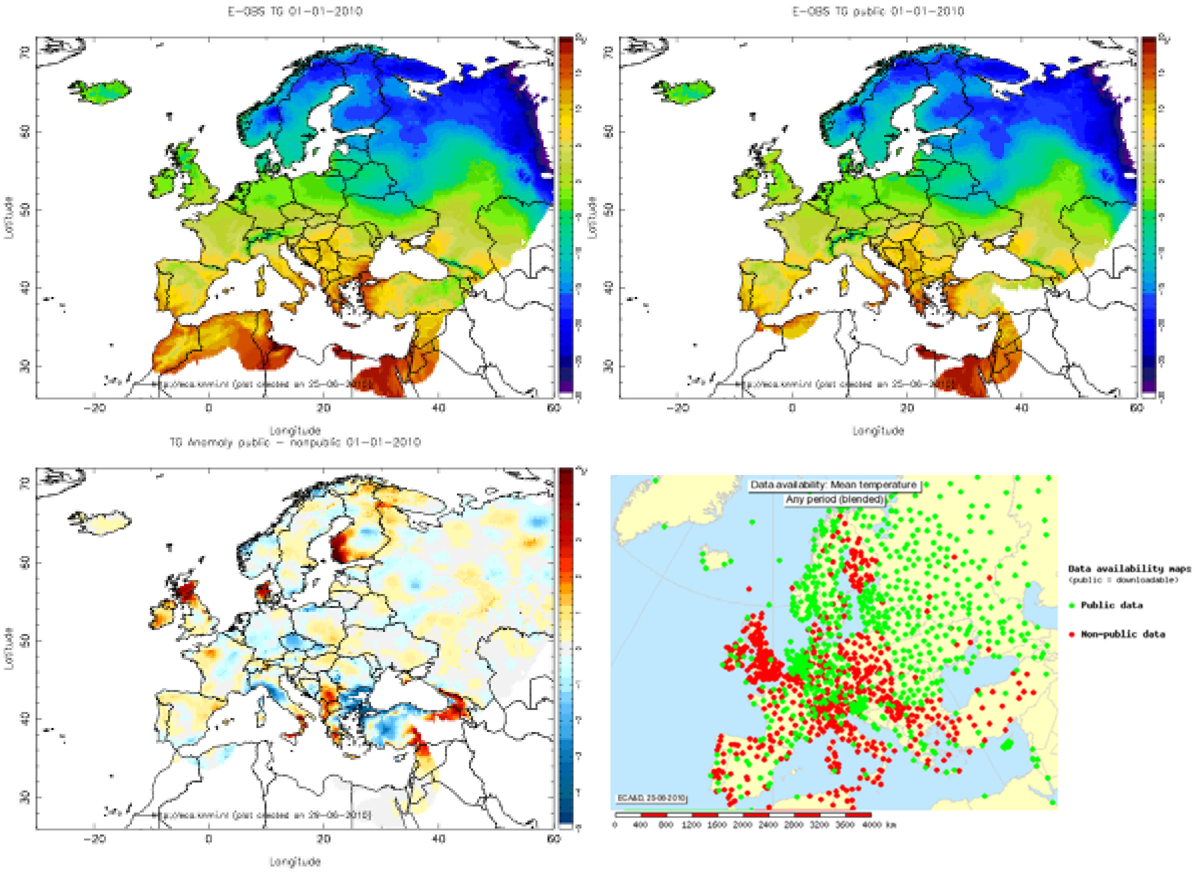

Figure 4. Daily temperature field observed in Europe on 1 Jan 2010 as generated using all available stations (a), or public stations only (b). The difference between (a) and (b) is given in (c). The station network underlying these products is shown for comparison (d). The differences between the two grids illustrate the effect of leaving out non-public data on the quality of the daily gridded products. Gridded products and public data (plus metadata for all stations) are available from http://eca.knmi.nl