WMO/WCP/WCDMP MEDARE International workshop on rescue and digitization of climate records in the Mediterranean Basin

Summary Report

Introduction

The first meeting of the MEDARE initiative group was held at University Rovira i Virgili (Tarragona, Spain) on November 28-30, 2007 with 45 attendees. The long-term goal of the project is to develop a high quality Mediterranean climate dataset, and the meeting laid out the initial plans for this undertaking. The first two days were given over to presentations – on the first day a number of invited experts spoke about dataset development, data rescue activities, data digitization efforts and homogeneity assessment of various digitized time series. The second day saw presentations by most of the National Meteorological and Hydrological Services (NMHSs) that encompass the Greater Mediterranean Region (GMR). The final day saw discussions about the best way to achieve the ambitious goal of the project. This summary report briefly discusses some of the common threads evident through the two sets of presentations, and lays out the decisions taken on the final day of the meeting.

The extended versions of all the papers presented will be in the proceedings of the workshop, which will be hung off of the MEDARE web portal. The presentations as given during the meeting can be downloaded from the MEDARE web portal (http://www.omm.urv.cat/MEDARE-workshop-outcomes/index.html).

Summary of the Presentations from Day 1 from the invited experts

Each expert was asked to talk on a specific aspect of data rescue. These aspects included, the need for data rescue and digitization, the apparent lack of digitized data in some parts of the GMR in international databases (e.g. the European Climate Assessment and Dataset, ECA&D and the Global Historical Climatology Network, GHCN and the latter's daily version, GDCN), scanning options, possible ways of improving the efficiency of digitization and the best techniques for homogenizing the resulting long climatic series. The best illustration of the need for the complete set of procedures was illustrated by Olivier Mestre's figure for France, which showed the spatial pattern of temperature trends for the 1901-2000 period before and after homogenization. Without the final homogeneity step, the digitization efforts in France would not have produced a coherent picture of the temperature increase. We show this figure as it is an ideal for all countries to aspire to (Figure 1, from Caussinus and Mestre, 2004). An example of what can be achieved for a single station was shown for Gibraltar (Dennis Wheeler), which is possibly the longest site in the southern part of the GMR. Additionally, a number of speakers discussed the MEDARE efforts in the context of wider international efforts, such as those envisaged by ACRE (Atmospheric Circulation Reconstructions over the Earth: http://brohan.org/hadobs/acre/acre.html), RECLAIM (Recovery of Logbooks and International Marine Data: http://icoads.noaa.gov/reclaim/index.html), GDCN

(http://www.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html) and IEDRO (International Environmental Data Rescue Organization: http://www.iedro.com).

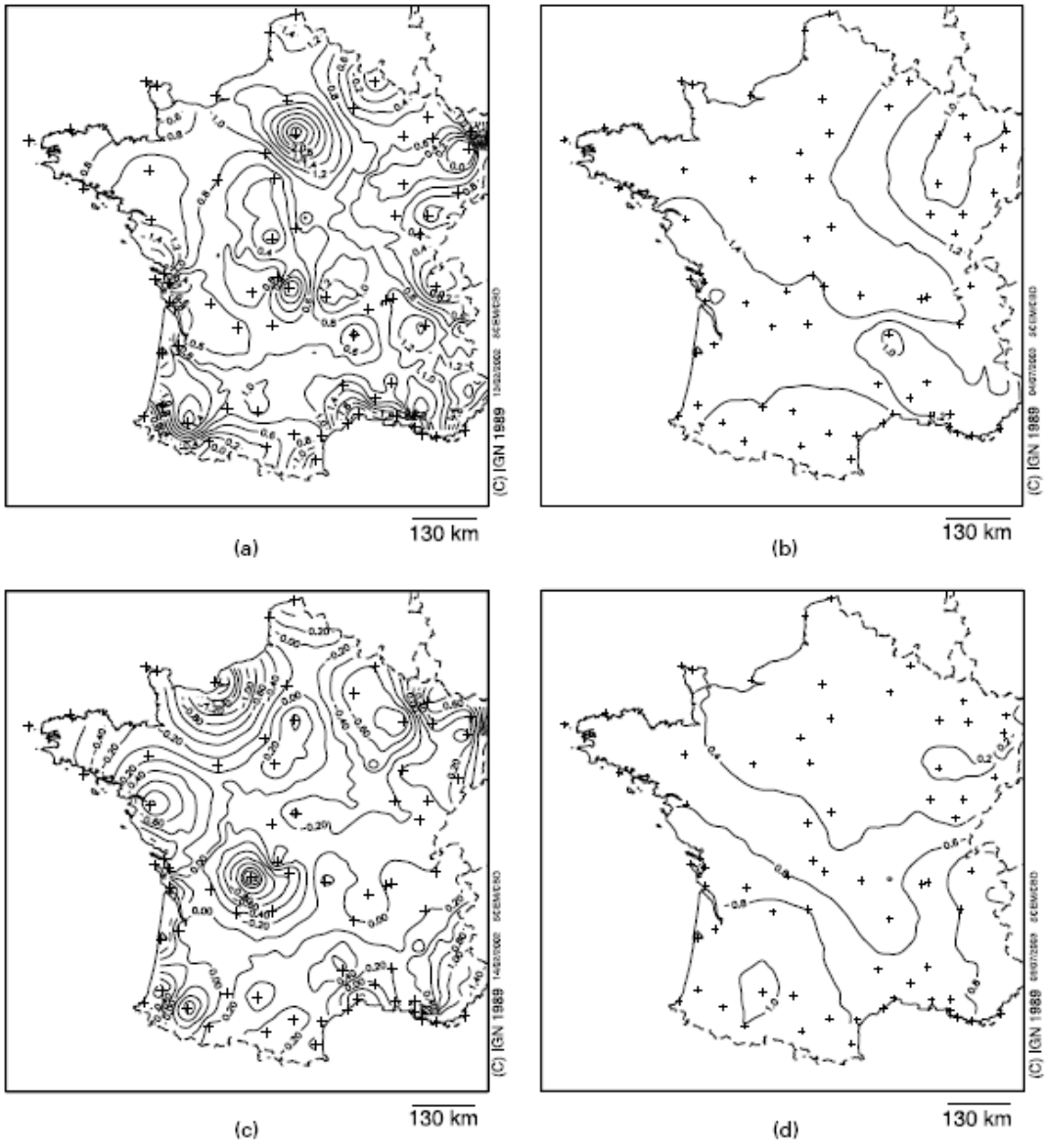

Figure 1: Maps showing smoothed contour plots of station temperature trends for 1901-2000 for a) minimum temperature as measured, b) as a) but after homogenization, c) as a) but for maximum temperature and d) as c) but after homogenization.

Summary of the Presentations from Day 2 for NMHS representatives

Most countries from the GMR region gave summary presentations on data rescue and recovery projects in their respective countries. There were a number of common recurring themes in many of the presentations:

- *Costs/access of the data*: Each country has different policies with respect to this, ranging between free data access to reduced prices for non-commercial

purposes like climate research. The general opinion was that historical climate data should be freely accessible for climate research, but commercial use is still an issue in many GMR countries.

- *Availability/access to the data*: In each country it is important to easily know what data have been digitized and how the national and international climate community can gain access. For each country there should be a contact person for assistance with access to the data. It is recommended that a list be made with e-mail addresses of NMHS contact persons, together with a list of other more European and Global sources of data (e.g. ECA&D, GHCN, etc.).
- *Resources for digitization*: It is clear that few NMHSs in the GMR have resources available to make much progress in achieving the goals of MEDARE. Some are routinely digitizing current observations, so can get some digitizing done (albeit slowly) through this process. The resource issue will be discussed later in the MEDARE Implementation Plan.
- *Digitizing data and metadata in NHMS and national archives*: It is recommended that with the digitized data not only the common metadata should be made available, but also articles or reports dealing with the climatological time series be made available (via pdfs). Standards need to be adopted for content of both the metadata and the data themselves. NMHSs should use their existing data formats when digitizing early data and adapt if necessary. For metadata, there is a WMO/CCl publication with a number of possibilities (http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP-53.pdf)
- *Digital and paper/printed-archival sources*: Regional problems with data rescue were also discussed. Many of the concerns raised cross current political borders. There are also problems with "colonial" data rescue and digitization - where are the data and how do we gain access to them? A modern day political border issue for many of the new Balkan states is: How does an NMHS gain access to historical station data that are currently held by a different national NHMS? What type of agreements may need to be set up to streamline this process? Related to this, for some countries, images of early instrumental data are available on the internet. Examples of this are the images made by the NOAA Central Library Climate Data Imaging Project using publicly available meteorological yearbooks for these countries. It is recommended that countries check the website of this project (http://docs.lib.noaa.gov/rescue/data_rescue_home.html) to see if digital images of their data (under previous colonial and current national administrations) are available there before making images themselves.
- *Assistance in digitizing data*: The EU-CIRCE project (http://www.circeproject.eu/) has 60Keuros available for digitization of historical data. Olivier Mestre (Meteo-France) is co-ordinating this to digitize historical climate data from some of the present GMR. Tom Ross indicated that the Climate Database Modernization Program (CDMP) of the National Climatic Data Center (NCDC) (http://www.ncdc.noaa.gov/oa/climate/cdmp/cdmp.html) in the US can be approached to digitise important data sets for MEDARE
- *Working with data rescue, imaging and digitisation activities of existing projects and initiatives*. The representatives of a number of existing regional to international initiatives at the MEDARE workshop, such as ACRE, CIRCE, ECA&D, IEDRO, MedCLIVAR and RECLAIM, all indicated a keenness to

work with the MEDARE NMHS's to rescue, recover, image and digitise historical to contemporary data series for the GMR.

- *Quality control of the digitized data and homogeneity assessment of the long time series*: Existing NMHS standards and software should suffice for quality control. Some NHMSs have experience of different homogeneity software packages, but more need to be aware of developments in this area.

Preamble to the Implementation Plan

The common thread from most of the NMHS presentations was that more resources (both financial and personnel) would be needed before significant progress could be made. It is also extremely unlikely that potential funding agencies would consider a proposal just for data rescue and digitization activities. It is necessary, therefore, to emphasize all the potential uses for which the instrumental dataset will be essential.

These include:

- Placing extreme events in a long context
- Enhancing knowledge about instrumental climate variability and change, and the possible factors causing these changes across the region
- Contributing to further advancement in climate change detection and attribution studies
- Enhancing inputs for defining/adopting the best strategies to mitigate climate change over the GMR
- Improving adaptation to climate change impacts, by developing longer series for assessing impact sector models
- Developing climate change scenarios by combining observational climate measurements with projections from Regional Climate Model simulations
- Enhancing the ability for contribution to the climate component of large field experiments/programmes
- Providing input to extended historical reanalysis (i.e. reanalyses prior to 1948)
- Calibrating natural/documentary proxies, for potential further extension of the climatic history of a country/region
- Calibrating satellite estimates of surface variables
- Providing better observation data for the validation of climate model outputs (both RCMs and GCMs)
- Performing more robust analysis of climate and applied climatological studies

All of these are discussed further in the papers at the Proceedings.

Potential Funding Sources

The group discussed numerous potential funding sources, from the regional (e.g. the European Union, the World Bank, the African Development Bank) to the national (e.g. National Governments and Research Councils) and the Private Sector (e.g. Google World). The basic limitation of all is that scanning and digitization efforts are not considered high-profile science, even though the requirement for credible climate data is strongly emphasised by various international organizations and fora

(i.e. G8 2005; GCOS 2003, 2004, 2006; GEO 2005, etc.). In addition, we all consider such efforts as essential requirements forming the backbone of our discipline.

Essential components are endorsement of the needs of these activities by the GMR as whole, and by WMO and other relevant intergovernmental bodies.

Implementation Plan

Date Rescue Activities

All NMHSs have digitized most of their recent records, with many having most of the period from the 1950s digitized. All NMHSs, however, have many old and original paper records, together with much non-digitized material in the old year books. This plan seeks to both preserve this material and digitize as much of the useful non-digitized material available to extend climatologically important time series.

Undertaking Data Rescue (DARE) activities involves rescuing both the data and metadata. The first step is to locate the original records and ensure their preservation for future studies. In many countries, old paper records need copying. The ideal is to not only scan the original material but also achieve long-term preservation by producing poly-acetate films. The latter is expensive, and will only be required where there is a serious risk of the records disintegrating or being destroyed. Storing for the future as paper records is adequate, provided recommendations for preservation from, for example, the European Commission Preservation and Access (EPCA) are followed. Digitizing of the material is the second and more important phase of any DARE project. This can proceed from either the original paper or year-book material or from scanned images. It is also possible that in some countries, the private sector may be able to help with the scanning and filming. Climatological insight is necessary in deciding what needs to be scanned, but much of the work could be achieved through the use of students and well-motivated private individuals (e.g. recent retirees). All DARE activities should be considered long-term, so there is a need to prioritize efforts, as well as continuing to look for sources of older material, looking particularly for measurements made prior to the founding of the NMHS.

Within the GMR there are many countries which have only become independent during the last 50 years. It is important, therefore, to consult the MEDARE community concerning archives held by other nations during colonial periods. Throughout the GMR, it should be possible to develop a few series for every country back at least into the mid to late-19th century. More extensive and spatially complete series will be available for later decades of the 20th century.

Variables

The ideal would be to digitize all material, but resources will always be limited. The recommendation from the meeting is to emphasize the Essential Climate Variables (ECVs) for the surface given in numerous Global Climate Observing System (GCOS) publications (see e.g. the GCOS Implementation Plan). These are, in priority order:

- Air temperature at 1.5-2m – including mean, maximum and minimum values
- Precipitation amounts
- Atmospheric Pressure – corrected to mean sea level

- Surface humidity – ideally vapour pressure, but also dewpoints, specific humidity and relative humidity
- Wind speed and direction
- Surface radiation – ideally measurements from radiometers, but sunshine records can be converted to the above with simple algorithms

Temporal Resolution

Again the ideal would be to develop a database down to the shortest temporal scale, but the minimum recommendation would be the daily timescale. Series should also be aggregated up to the monthly timescale as well.

Digitizing

This aspect is potentially the most time consuming and expensive part of the work, as it requires the development of digitizing software and climatological experience and input at all stages. In almost all countries such software and expertise has been developed for local needs, but it generally requires more resources to cater for the potential volumes of material involved. Manual digitizing is the only approach for hand-written material, but Optical Character Recognition (OCR) techniques should be investigated for all printed material (e.g. in early year books). There is a lot of experience of OCR across the GMR and also in many countries in Northern Europe. In addition, work is also progressing on electronic means of digitising strip charts, such as thermohygrograph or barograph traces. In many NMHSs the use of well-supervised students has been found to be a very cost-effective way of achieving the best results – in terms of the amount of data digitized.

Quality Control

Every NMHS has quality control (QC) procedures for assessing current data entering national NMHS digital archives. Once digitized, the extended series should be passed through these procedures taking advantage of the experience of trained staff.

Homogenization

Over the last 20 years, a number of software packages have been developed to assess the long-term homogeneity of climatic time series. At present, there isn't a best method, but interactions with the COST Action HOME (http://www.homogenisation.org/), which will run over the next four years, is highly recommended. However, MEDARE will not be able to rely entirely on links to activities such as HOME, so funding for a MEDARE training workshop, where participants can come with data and learn about experiences across the GMR with some of the more well-used approaches is recommended. In support of such activities, MEDARE should also look to interact with the CCl/CLIVAR/JCOMM Expert Team

on Climate Change and Indices (ETCCDI) in association with WMO, NMHSs, and other co-sponsors such as GCOS, IPCC, START, who are organising roving regional workshops on Climate Data Homogenization and Climate Change Indices (http://www.clivar.org/organization/etccdi/activities.php). The ETCCDI has accomplished a series of regional workshops, covering SE Asia (Nov 07), Africa (central April 07), southern Asia (Feb 05), Central America (Nov 04), SW Asia (Oct 04), South America (Aug 04) and southern Africa (May 04). These workshops have promoted regional climate change detection activities and filled in gaps in global climate data sets.

Development of Data Inventory

Mindful of the time such a project will take, and the variety of rates of achievement of the aims across the GMR, the group proposed a data inventory to enable some progress to be made within the first few months. Here, the MEDARE community would develop lists for their country of the longest times series data for the ECVs. The lists would not include the data, but give sufficient details on what metadata are available and what part of the record is digitally available (from whom), what part still needs recovering and digitizing and what homogeneity assessments have been carried out on the series. The national lists would be combined into a GMR inventory of source availability of the long and potentially long records (> 50 years).

Development of the Web Portal

The University Rovira i Virgili in Tarragona agreed to host the site, where information on goals, people, contact lists, working groups, documentation, inventory of the longest climate records, and other MEDARE activities will be provided. It will include restricted areas for the MEDARE community and its working groups.

Set up a number of email accounts for contacting working groups with specific questions and general advice:

- Where might early colonial material be held?
- What are the best scanners to purchase?
- Which is the best OCR software for printed material?
- Which homogenization software is best for specific variables?

Next meeting!

Greece!

References

Caussinus, H. and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Applied Statistics* **53**, 405-425.

G8, 2005: Gleneagles plan of action on climate change, clean energy and sustainable development (http://en.g8russia.ru/g8/history/gleneagles2005/7/)

GCOS, 2003: *Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCCC*". GCOS-82, WMO/TD-No. 1143 (http://www.wmo.int/pages/prog/gcos/Publications/gcos-82_2AR.pdf)

GCOS, 2004: *GCOS Implementation Plan for Global Observing System for Climate in support of UNFCC*. GCOS-92, WMO/TD 1219, 136 pp. (http://www.wmo.int/pages/prog/gcos/Publications/gcos-92_GIP.pdf)

GCOS, 2006: GCOS Regional Action Plan for the Mediterranean Basin. September 2006. (http://www.wmo.ch/pages/prog/gcos/documents/GCOS_MED_RAP_FINALDRAFT_Sep06eng.pdf)

GEO, 2005: Global Earth Observation System of Systems - GEOSS 10-Year Implementation Plan. Reference Document. The Netherlands. ISBN No.: 92-9092-986-3