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.

Interactions with other activities

Peter Thorne, Elizabeth Kent, Nick Rayner, Stephan Bojinski, Thomas Peterson, Dick Dee, John Christy

Background

The rationale behind the initial focus on land surface temperatures was outlined in a proposal for the workshop (Met Office 2010). Subsequent discussions with the World Climate Research Programme (WCRP) and the Global Climate Observing System (GCOS) emphasized wider applicability, but recognized the practical difficulties of initiating a broader activity. Planning that explicitly incorporates the requirements for marine temperatures, precipitation, surface pressure and other variables from the start will ensure that the benefits of an inclusive approach can be realized, once resources for these activities become available.

Much of the work that is planned to be addressed by this initiative has implications for a wide variety of other activities. For example, if a group is digitizing paper archives of temperature data it would be illogical not to digitize the precipitation data, or other variables, at the same time. Therefore, from the outset it is important to recognize the need to reach out to the wider community. Whilst the primary objective is a suite of consistently assessed and accessible global surface land temperature record datasets and resulting applications tools, we must ensure that governance, scientific methodology and technology (processes, standards, and formats) are appropriate for a multivariate approach and thereby facilitate wider interactions.

In order to launch the discussion on better transparency, traceability and scientific analysis of climate data records in general, GCOS and WCRP produced guidance (GCOS, 2010) to institutions involved in climate data generation and archiving. Such institutions, involving meteorological services, oceanographic services and space agencies, were also asked by GCOS and WCRP, through a May 2010 letter, to give stronger support to international scientific working groups involved in scientific analysis, intercomparison and effective peer-review of climate datasets.

This white paper considers specifically how the concepts discussed within the previous white papers may most optimally be considered in a global multivariate framework and expanded to other observing technologies. It is expected that the approach outlined in the previous white papers, if successfully applied, should be broadly applicable and represent scientific best practice. This will ensure that our understanding of the climate system is scientifically robust and that we are creating products that underpin society's needs for climate services, as promoted by the 2009 World Climate Conference-3.

Lessons from the marine experience

The land data management effort can learn from the marine experience, which is more mature. The marine databank (the International Comprehensive Ocean-Atmosphere Data Set, ICOADS) is already open access. ICOADS is managed and delivered by the three US founding partners (NOAA, NCEP and NCAR) but also benefits from international contributions of expertise, data and metadata, as recognized in the name. ICOADS is multivariate, ingesting the full ship report. This has been particularly helpful in the study of air-sea interaction and the use of multivariate quality assurance. Additionally, the different nature of bias and uncertainty in marine air temperature and SST provides increased confidence in analyses of each and hence in the global temperature record. That is not to say that everything is necessarily perfect in the marine world. The ICOADS effort has arguably never been run with sufficient budget to do everything the databank owners and users would like it to. Ensuring adequate resources for the development and maintenance of databanks is clearly a challenge that must be addressed. Further, commercial considerations have led to a reduction in the number of vessels undertaking marine meteorological observations – an analogy to the challenges faced in the land arena where commercial imperatives confuse the picture over data availability substantially (see White Paper 5).

The marine community is attempting to tackle some of the same or similar issues as this workshop. Although ICOADS was conceived as a surface marine dataset, the need for compatibility with subsurface ocean datasets is recognized, but not yet addressed. The need for comparison, validation, calibration and joint analyses among in situ and satellite datasets presents much the same practical difficulties for both land and marine data. Lessons should be learnt and experience shared as we look to develop a similar databank for land data. Efforts are already underway to ensure a better degree of coordination. Several experts from the marine community will be attending the workshop. There are also plans to address coordination at the MARCDAT-III meeting, in May 2011, and the organizing committees for the two workshops overlap ensuring a degree of consistency. There, the discussion started at this workshop will be picked up, with a focus on the marine aspects.

The need for global and multivariate analyses including dynamical reanalyses

The advantages of a multivariate approach include: the inherent advantages of managing variables collected together in a consistent way, the potential for consistent and multivariate quality assurance; ease of input to multivariate analyses including dataset construction techniques and dynamical reanalyses; and improved climate system understanding.

Estimates of the global surface temperature record will require combined analysis of land air temperature and sea surface temperature measurements. These marine measurements are made in a fundamentally different way to land measurements and creating climate datasets from these data have very different challenges. Historically, the problems in land and marine observations have been considered in isolation from one another with any merging performed over geographically congruent areas after the fact. Currently available monthly global products that follow this approach are HadCRUT, NCDC and GISS. This non-integrated approach, almost inevitably, has led to issues over interpretation. At different points in time the maturity of our estimates for one or other of the land or SST components has differed from the other making interpretation sub-optimal. For example, the new Hadley Centre SST product HadSST3 will consist of an equi-probable solution set allowing for assessment of uncertainty at all space and time scales. However, it will be merged with a single land dataset with an underlying error model that is fundamentally different. How to merge and interpret these different products is far from trivial. We must ensure that the future evolution of SST and land surface air temperature analyses is coordinated to provide compatible products.

Recognition of these issues and the parallel planning undertaken in this workshop and MARCDAT-III will facilitate this. But practically, this can only be achieved through greater collaboration from the start of dataset development. Funding opportunities need to be aligned.

Meeting the challenges of producing consistent global analyses from disparate observations may require the application of analysis techniques beyond purely statistical techniques. Model-based dynamic reanalysis provides a framework for accommodating and integrating different overlapping sources of information, and for producing physically consistent datasets (Trenberth et al. 2010). Reanalysis can incorporate uncertainty information such as will be provided with HadSST3. In fact, the EU-funded ERA-CLIM project, which will start early next year, will use multiple HadISST2 realizations to produce an ensemble of global atmospheric reanalyses of the 20th century, as will the Twentieth Century Reanalysis Project (Compo et al, 2010).

Reanalyses form an increasingly important component of our monitoring capabilities and constitute the most utilized data in the climate sphere. They integrate data from many technologies and utilize model physics as well as statistics to infer values where none were observed or for parameters which are unobservable. They also continue to improve as successive efforts learn from previous efforts and build upon developments in numerical forecasting. Reanalyses themselves constitute a surface temperature analysis. Although they cannot by design be run against the benchmarks (White paper 9) or realistically be entirely reproducible by a third party they will have substantial value. However, reanalyses also suffer from having the same incomplete and imperfect data as the more classical climate datasets, and there is inevitable uncertainty.

Of particular value are reanalysis "feedback files" which provide information regarding analysis increments and fundamental quality. Reanalysis fields, particularly from systems that do not directly ingest the variable of interest, can be used as a background for comparison for homogenization and quality control. This approach has proved successful for the analysis of bias in radiosonde observations (Haimberger, 2007). Reanalyses could also be used to help assess for the presence of any inconsistencies between land and marine data.

An improved databank would be a natural primary data source for ingest to reanalyses. Reanalysis could then take advantage of data quality assurance and metadata produced by the project if it were provided in a consistent format. That is to say that the reanalysis could make use of the quality control flags and bias estimates returned by the various dataset creation algorithms to inform in some sense the assimilation step. Finally, and perhaps most easily, the data presently ingested by reanalyses may themselves provide data to populate the databank (White paper 3).

Challenges of extension to satellite observations

Working with satellite data poses very different obstacles and opportunities than working with in situ data. Polar orbiting satellites provide near-global coverage from one data source. International exchange of satellite data involves just a few agencies or countries. Satellite observations are indirect measurements of the variables of interest. And satellite data present many orders of magnitude more bits and bytes to archive, exchange and transport. Yet there are lessons and collaborations that can be gained from producers and users of satellite data.

Many space agencies are giving increased attention to maintaining archives of the basic satellite data records and metadata from past and current missions and the reprocessing of datasets for climate monitoring and climate research. Guidance has been provided by the 2006 GCOS satellite requirements for "Fundamental Climate Data Records" (generally, calibrated radiances or level 1) and Essential Climate Variable (ECV) satellite products (meteorological or geophysical fields/variables). For example, ESA member states have launched the ESA Climate Change Initiative, initially addressing 11 ECVs from their past and future missions, with full transparency, traceability, and estimating uncertainties as principal objectives. EUMETSAT, in conjunction with NOAA, JMA and WMO, has been supporting the SCOPE-CM initiative (Sustained Coordinated Processing of Environmental Satellite Data for Climate Monitoring), with an initial focus on the generation of long-term ECV products for upper-tropospheric humidity, clouds, aerosols, water vapour, surface albedo, atmospheric motion vectors and clear-sky radiance. Many agencies are undertaking a range of reprocessing activities, although attention to ensuring common product benchmarks between the activities of different agencies is sometimes inadequate. Some level of international coordination of these activities by space agencies has been occurring at the level of the Committee on Earth Observation Satellites (CEOS) and the Coordination Group for Meteorological Satellites (CGMS). Recently, CEOS decided to set up an ad-hoc working group to enhance coordination among agencies on the generation of "Fundamental Climate Data Records" and ECV satellite products.

Notwithstanding the progress made, there is an ongoing need for space agencies to achieve the funding needed to sustain both the required future missions and the production of the resulting FCDRs. There is also a recognition that their value will be greater when combined with or compared to in-situ records that provide a synergistic check. Here, we could learn from the GHRSST activity that has undertaken this for Sea Surface temperatures. The exchange between agencies and the lessons learned in the construction of integrated climate databases and their exploitation could be useful. Efforts will be needed to ensure that the data from the surface network can be easily compared to that from other networks / platforms more widely than just the marine component although this may take longer. This may be best achieved through some common observational portal for all types of data as has been discussed to our knowledge in various fora but has yet to mature into a concrete activity. This portal would in all probability need rigid naming conventions, version control and metadata to be useful. Another option to enhance the value of observational datasets may be the adoption of common standards in data formatting and documentation, as exemplified in GHRSST and in the coordinated modelling activities under the Coupled Model Intercomparison Project Phase 5 (CMIP5) that are closely linked to the IPCC AR5 process.

Requirements for a databank for multivariate and global applications

The precise details of how the concepts outlined in the white papers should be applied to other parameters and domains will differ depending upon the nature of the underlying data, the variable itself, the measuring technology and the maturity of our understanding. For example, for marine data and data from polar orbiting satellites the fact that the instruments are constantly on the move presents a very different set of challenges. Careful thought will be required as to how to entrain efforts on other variables and observing technologies and learn from their experiences to date. Further, it is imperative, as discussed in the databank white papers, to collate all information from a station, measurement platform (e.g. ship, satellite) or instrument and not just temperatures. Compatibility with existing databanks is essential.

On a practical level there could be efforts made to enable more optimal use of both marine and land temperature data. This could include, for example, efforts to create match-up databases where marine and land data are in close proximity in space and time and are fully characterized by respective metadata to the extent practical. Also, ensuring that the databanks and resulting value added products are sufficiently consistent to enable easy use in studies that wish to consider both land and marine data is essential. In this respect we may wish to consider the “climate ICOADS” model which aims to create a version of ICOADS with optional bias adjustments and uncertainty estimates appended on a per observation basis.

Recommendations

1) The development of a land surface temperature databank should allow for the extension to multivariate datasets and compatibility with global databanks from the outset.
2) Digitization activities should take a multivariate approach and always incorporate all metadata.
3) Emerging links between land and ocean dataset developers and researchers should be fostered and facilitated by the development of compatible databanks, data products and joint research projects.
4) Appropriate linkage to the activities supported by space agencies aimed at the generation of long-term climate data records addressing the GCOS Essential Climate Variables should be established when developing the surface temperature databank. The experience of GHRSST in the combined management and use of satellite and in situ data should be exploited.
5) Research is needed to improve multivariate analysis methods and to develop techniques to produce consistent global data products.
6) The important role of reanalysis in providing global multivariate analyses with wide application including quality assessment should be recognized. These products will form an essential part of a successful surface temperatures project providing both a set of estimates and a wealth of metadata regarding the data quality.
7) Funding agencies should recognize that an internationally coordinated and sustained approach to the development, maintenance and improvement of climate databanks and derived data products will have wide benefits. This most logically includes a CMIP type portal for climate data records from all observing platforms with common formats and strong naming conventions to enable ease of intercomparison.

References

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M.C. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, S.J. Worley, 2010, The Twentieth Century Reanalysis Project. Quart. J. Roy. Meteor. Soc., submitted.

GCOS, 2010, Guideline for the Generation of Datasets and Products Meeting GCOS Requirements, May 2010, GCOS-143 WMO/TD No. 1530.

Haimberger, L., 2007, Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. Journal of Climate, 20, 1377-1403

Met Office, 2010, Proposal for a new international analysis of land surface air temperature data, WMO Commission for Climatology, Fifteenth session, Antalya, Turkey, 19-24 February 2010, CCl-XV/INF. 15, available from http://www.surfacetemperatures.org/background.

Trenberth, K.E., R. Dole, Y. Xue, K. Onogi, D. Dee, M. Balmaseda, M. Bosilovich, S. Schubert, W. Large, 2010, Atmospheric Reanalyses: A major resource for ocean product development and modeling, " in Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. & Stammer, D., Eds., ESA Publication WPP-306.