



AMERICAN METEOROLOGICAL SOCIETY

Bulletin of the American Meteorological Society

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: [10.1175/BAMS-D-16-0165.1](https://doi.org/10.1175/BAMS-D-16-0165.1)

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Thorne, P., R. Allan, L. Ashcroft, P. Brohan, R. Dunn, M. Menne, P. Pearce, J. Picas, K. Willett, M. Benoy, S. Bronnimann, P. Canziani, J. Coll, R. Crouthamel, G. Compo, D. Cuppett, M. Curley, C. Duffy, I. Gillespie, J. Guijarro, S. Jourdain, E. Kent, H. Kubota, T. Legg, Q. Li, J. Matsumoto, C. Murphy, N. Rayner, J. Rennie, E. Rustemeier, L. Slivinski, V. Slonosky, A. Squintu, B. Tinz, M. Valente, S. Walsh, X. Wang, N. Westcott, K. Wood, S. Woodruff, and S. Worley, 2017: Towards an integrated set of surface meteorological observations for climate science and applications. *Bull. Amer. Meteor. Soc.* doi:10.1175/BAMS-D-16-0165.1, in press.



Towards an integrated set of surface meteorological observations for climate science and applications

P. W. Thorne(1), R. J. Allan(2), L. Ashcroft(3), P. Brohan(2), R.J.H Dunn(2), M. J. Menne(4), P. R. Pearce(5), J. Picas(6), K. M. Willett(2), M. Benoy(7), S. Bronnimann(8), P. O. Canziani(9), J. Coll(1), R. Crouthamel(10), G. P. Compo(11), D. Cuppett(10), M. Curley(12), C. Duffy(2), I. Gillespie(1), J. Guijarro(13), S. Jourdain(14), E. C. Kent(15), H. Kubota(16), T. P. Legg(2), Q. Li (17), J. Matsumoto(18), C. Murphy(1), N. A. Rayner(2), J. J. Rennie(4,19), E. Rustemeier(20), L. C. Slivinski(11), V. Slonosky(21), A. Squintu(22), B. Tinz(20), M. A. Valente(23), S. Walsh(12), X. L. Wang(24), N. Westcott(25), K. Wood(26), S. D. Woodruff(4,11), S. J. Worley(27)

Corresponding Author:

Peter Thorne,

ICARUS

Department of Geography

Maynooth University

Maynooth

Ireland

peter@peter-thorne.net

- (1) Irish Climate Analysis Research UnitS, Department of Geography, Maynooth University, Maynooth, Ireland
- (2) Met Office Hadley Centre, Exeter, United Kingdom
- (3) Centre for Climate Change, Universitat Rovira i Virgili, Tarragona, Spain
- (4) National Oceanic and Atmospheric Administration's National Centers for Environmental Information, Asheville, NC, USA
- (5) National Institute of Water and Atmospheric Research, Auckland, New Zealand
- (6) University of the Witwatersrand, Johannesburg, South Africa
- (7) Citizen Science Unit, Australian Meteorological Association / Australian Government Bureau of Meteorology, Adelaide, South Australia, Australia.
- (8) Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Switzerland
- (9) CONICET, National Scientific and Technical Research Council, Argentina
- (10) International Environmental Data Rescue Organization, Deale, MD, USA
- (11) Cooperative Institute for Research in Environmental Sciences, University of Colorado, & NOAA Earth System Research Laboratory Physical Sciences Division, Boulder, CO, USA

- (12) Met Eireann, Glasnevin, Dublin, Ireland
- (13) AEMET, Spain
- (14) Meteo-France, France
- (15) National Oceanography Centre, Southampton, UK
- (16) Atmosphere and Ocean Research Institute, University of Tokyo, Chiba, Japan
- (17) School of Atmospheric Sciences, SUN -Yat Sen University, Guanzhou, China
- (18) Department of Geography, Tokyo Metropolitan University, Tokyo, Japan
- (19) CICS-NC, North Carolina State University, NC
- (20) Deutscher Wetterdienst, Offenbach and Hamburg, Germany
- (21) McGill University, McGill, Canada
- (22) KNMI, De Bilt, Netherlands
- (23) IDL, Faculty of Science of the University of Lisbon, Portugal
- (24) Climate Research Division, Environment and Climate Change Canada, Ontario, Canada
- (25) University of Illinois, Midwestern regional climate centre, IL, USA
- (26) JISAO, University of Washington, WA, USA
- (27) National Center for Atmospheric Research, Boulder, CO, USA

Capsule

We cannot predict what is not observed. And, we cannot analyze what is not archived.

Fully integrated land meteorological records are essential to advance understanding of weather and climate.

Abstract

Observations are the foundation for understanding the climate system. Yet, currently available land meteorological data are highly fractured into various global, regional and national holdings for different variables and timescales, from a variety of sources, and in a mixture of formats. Added to this, many data are still inaccessible for analysis and usage. To meet modern scientific and societal demands as well as emerging needs such as the provision of climate services, it is essential that we improve the management and curation of available land-based meteorological holdings. We need a comprehensive global set of data holdings, of known provenance, that is truly integrated both across Essential Climate Variables (ECVs) and across timescales to meet the broad range of stakeholder needs. These holdings must be easily discoverable, made available in accessible formats, and backed up by multi-tiered user support. The present paper provides a high level overview, based upon broad community input, of the steps that are required to bring about this integration. The significant challenge is to find a sustained means to realize this vision. This requires a long-term international program. The database that results will transform our collective ability to provide societally relevant research, analysis and predictions in many weather and climate related application areas across much of the globe.

1. Background and current status of land meteorological holdings

Meteorological observations at stations over land areas have been taken for several centuries. Early measurements were performed by the scientists of the enlightenment exploring their environment, by medical doctors trying to understand illnesses, agronomists aiming at improving yields, or colonial administrators documenting the wealth of their assigned colonies. These early observations used a broad range of instrumentation, emerging measurement systems, and disparate measurement scales (e.g. Reaumur, Fahrenheit, Celsius, Paris Lines, Inches, Millimetres, Hectopascals; Brázdil et al., 2010). From the mid-19th century onwards, many nation states started the operation of meteorological networks in the framework of a new administration style, new responsibilities, and technical innovations (e.g., telegraph) (Edwards, 2011). By the late nineteenth century, station coverage (albeit sparse) extended across all inhabited continents sufficient that, today, we can calculate estimates of global mean change in several important parameters, such as temperature (e.g. Lawrimore et al., 2011, Morice et al., 2012), dating back to the mid-to-late nineteenth century. The late nineteenth and early twentieth centuries saw a drive to standardize methods of observations, data formats, and metadata under the World Meteorological Organization (WMO) and its predecessors as the scientific and societal value of these observations became more and more apparent (Parker, 1994). But change continues to this day, for example, with an increased propensity to unmanned methods of observation.

These meteorological data constitute the foundation of our knowledge of the climate system. Without long records of observations, there can be no viable pathway to understanding climatic processes, climate variability, climate and weather extremes, or climate change. Historical observations have been used to create datasets for many Essential Climate Variables (ECVs, Bojinski et al., 2014) that have enabled assessments of changing climate system properties (e.g. Blunden and Arndt, 2016, Hartmann et al., 2013). They tell us about a world that has changed and warmed over the last 150 or so years (Kennedy et al., 2010), and their continuation is important in monitoring effectiveness of the recent Paris agreement (Dolman et al., 2016). They can also be used (along with many other observations) to derive reanalysis products (Saha et al., 2010; Compo et al., 2011; Rienecker et al., 2011; Dee et al., 2011, 2014; Kobayashi et al., 2015) and assess performance of climate models (e.g. Flato et al., 2013 and references therein).

Despite the importance of land-based meteorological observations, their data management is currently very fragmented. While for marine observations the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al., 2016) provides integrated access to a wide range of surface marine data, the same cannot be said for the land-based observations. The absence of a coordinated global program for data rescue and provision, data management, data curation, and data usage means that we are not currently extracting the full scientific and societal benefits from the observations that we know of (and that ignores the unknown knowns — the uncatalogued, the unshared, and the forgotten). Such data have potential to be used seamlessly across a range of local, regional and global products and applications.

However, this problem is multifaceted with no simple solutions. Data sharing is complex, involving both data rescue and data policy. National and global meteorological data management has historically been highly fractured, such that today we have distinct holdings for hourly, daily, and monthly data managed by different groups in disparate ways. We have also developed several ECV-specific sets of holdings for many ECVs such as surface pressure (Cram et al., 2015), temperature (Rennie et al., 2014) and precipitation (Adler et al., 2003; Schneider et al., 2014). These holdings use distinct data formats, carry different station identifiers, often disagree in simple facets such as station names and even coordinates, and have differing levels of data completeness. Different sources may have had distinct quality control applied. Even more confusingly, daily and monthly averages derived from the same underlying data may differ if processed in different ways. Further compounding the issues, available data discovery tools are often rudimentary, and traceability to the original source is lacking. Hence, it is at best hard and at worst impossible for users to reconcile available holdings.

While some efforts to create merged holdings exist regionally, e.g. ECA&D (European Climate Assessment & Dataset; Klein Tank et al., 2002; Klok and Klein Tank, 2009) and ICA&D components (International Climate Assessment & Dataset; Van Den Besselaar et al., 2015), and globally, e.g. GHCN-D (Global Historical Climatology Network-Daily; Menne et al., 2012), they are limited in geographical domain (ECA&D, ICA&D), temporal integration (GHCN-D), and often to a narrow subset of variables (all to some extent). This globally piecemeal approach to data management is a major impediment to optimal usage of historical land meteorological holdings. What we need is a

coordinated, sustained, and international effort to better access, manage, and integrate these holdings.

The challenge is not simply that the holdings are sub-optimally managed for present-day applications, but also that there are emerging needs that are currently ill-served (Allan et al., 2012). In the era of climate services, there is a growing and changing demand for local information, for information on daily and sub-daily aspects of climate change (not least extremes, impacts and risks), and for data that are 'open' and free from restrictions and conditions. Many of these emerging applications also require consideration of the full suite of surface ECVs (Bojinski et al., 2014), as well as measurements which are not currently routinely captured or used from multiple sources beyond the remit of many National Meteorological Services (NMS) (e.g., agriculture, transport, energy, amateur observers). Furthermore, there is an increasing need for high-quality discovery and observational metadata: indicators of quality and uncertainties, in addition to known changes in measurement techniques, practices, locations and siting environments. Such data and metadata are essential in provision of scientifically robust climate services (Van Den Besselaar et al., 2015).

Put bluntly, we keep being confronted with climate situations which are seemingly unprecedented. This is likely in part because we are currently unable to comprehensively represent the climate of our recent past due to the present disorganized state of land-based holdings. In order to be able to better assess and respond to climate-induced challenges, it is crucial that we gain a better understanding not only of large-scale climatic trends, but also of local climate, past conditions and

variability, and short-term events such as flooding, heat waves, and storms. We are in a fortunate position of being able to draw upon decades, and in many cases, centuries of scientific heritage in the form of rich meteorological archives, but this invaluable heritage is not currently capable of being used to its full potential. In order to be able to research the scope and causes of climatic events effectively, it is imperative that the many millions of observations taken at some tens of thousands of locations, including emerging methods and modes of observation, be integrated and consolidated into a coherent database.

In this paper we outline a high-level roadmap for a Comprehensive Land-based International Meteorological Observation Databank (CLIMOD herein), and offer examples (via boxes) of how valuable a database like this could be. We also consider the numerous logistical hurdles to overcome. We begin by outlining the high-level vision of the approach to be undertaken. Then, in turn, we discuss in more detail aspects around data sourcing, data rescue and collection (including metadata), data management, data serving, and governance (including data policy).

2. High level vision for CLIMOD

The combined policies, practices, and projects since systematic observations of weather were instigated have led to a massively fractured set of land-based meteorological archives ranging from single-station, single-ECV, single time resolution series to large global multi-ECV, multi-timescale holdings. Stations often have contributed to multiple archives, sometimes at a range of time averages, such that there

is gross duplication and resulting confusion in the available records. CLIMOD will first collect all digital data holdings and associated metadata, augment them with rescued and newly-discovered data as they become available, and combine them into a single database of all commonly observed ECVs at land meteorological observing sites. To ensure provenance (data lineage), these collected data should be as close to their original format and source as possible. This basic consolidation is a key and necessary first phase, but it is not sufficient to facilitate widespread usage, except perhaps by a handful of experts who can properly assess and use such complex holdings.

What is next required is analogous to a satellite “Level 1b” product (Campbell and Wynne, 2011), that is geo-located, with duplicates removed, in consistent geophysical units, with metadata directly associated and with gross errors flagged. This would provide a set of openly available land-meteorological holdings that is integrated, consistently-formatted and discoverable. In this phase, the different ECVs should be collated and served together; to quote Aristotle, “the whole is greater than the sum of its parts” (as illustrated by the three boxes outlining potential applications). When these data are served in this manner, their potential utility will be substantially increased.

Figure 1 illustrates the change between a set of consolidated (all available) holdings and a set of merged (unique only) holdings for the particular case of monthly-averaged surface temperatures. Note the reduction in apparent station count that relates to duplicates and records of uncertain quality or provenance.

Another problem is that records exist in the archives at a mixture of sub-daily, daily, and monthly-averaged timescales. Hence, next it is required to ensure timescale

consistency. The sub-daily data — where they exist — should be consistent with the derived daily values, which should in turn be consistent with the monthly values. Consistency implies that, for example, daily averages and ranges are consistent with the respective reported maxima and minima recognising national reporting norms that vary over time and between countries. This will allow analyses of extremes and events to be consistent with studies of mean change. CLIMOD should thus, in addition, contribute to envisaged cross-domain and cross-ECV analysis systems (Kent et al., 2016; their recommendation 7). Figure 2 provides a schematic overview of the proposed CLIMOD holdings and their subsequent analysis.

3. Data sourcing

Historical data can be considered in three classes: i) data that are digitized and are already made available with workable or no usage restrictions; ii) data that are digitized but are either not made available or made available with restrictive usage conditions; and iii) data that are only available presently as either hard copy or, at best, non-machine-readable image formats. Imperfect individual and community knowledge leads to a fourth class of ‘unknown knowns’ or ‘forgottens’ — data that exist but that the present-day scientific community, for a variety of reasons, do not know about. These data should become substantively less elusive with a successful CLIMOD simply as a result of raised community awareness as to the existence of a recognised repository for such data.

Data that are digitized and available for, at a minimum, academic and research usage are served from a broad variety of local, national, regional and international repositories. This includes National Meteorological Services, regional efforts such as ECA&D, and global repositories such as GHCN-D. Experience with the International Surface Temperature Initiative (ISTI, Rennie et al., 2014) and the Atmospheric Circulation Reconstructions over the Earth initiative (ACRE, Allan et al., 2016) points to additional, non-traditional, sources which could greatly improve the density of available observations. For example, through Argentinean members of the ISTI databank working group, access was secured to a number of long-term rural stations from the Argentinean Agricultural Ministry.

Many of the currently served holdings are actually amalgamations of a broad range of underlying sources. For example, GHCN-D is an amalgamation of over 50 sources many of which themselves (such as ECA&D) are the result of amalgamating underlying sources. It would be preferable to start from the underlying sources in subsequent efforts, described in the next Section, to merge holdings together. Starting from the underlying sources provides better provenance and data processing transparency.

Initially, CLIMOD would start from as many of the available holdings with minimal usage restrictions as could be collected. Efforts would be made to capture the contributing data sources from amalgamated holdings. This collation and cataloguing of available data will require substantive international efforts and cooperation, and will build on the partnerships already developed by the currently served holdings.

There are several known data holdings which are available digitally but not currently available for academic and research use. The prevalence of such holdings has strong geographical patterns, with data policy being a particular issue for many of the areas where the community currently has access to the fewest data. In some cases, such as in the Pacific and South East Asian regions (Page et al., 2004), these amount to several thousand stations. Efforts are already underway to advocate for more relaxed data policies and for allowing access to these data (Lorrey, 2011 [<http://www.apn-gcr.org/resources/items/show/1666>], Williamson et al., 2015). CLIMOD could facilitate such efforts, but the issue is vexing and could take decades to resolve. We return to data policy in Section 6.

Data rescue activities have provided major improvements in the availability of historical meteorological data and must be sustained. There are, by conservative estimates, at least as many data to be rescued as are currently available in digital archives for the period prior to 1950 (Allan et al., 2011). There are growing international efforts such as ACRE and the International Environmental Data Rescue Organization (IEDRO), and funded projects such as European Reanalysis of the Global Climate System (ERA-CLIM2) and Uncertainties in Ensembles of Regional ReAnalyses (UERRA) which have attempted to mobilise, catalogue, and coordinate data rescue activities. For example, in 2014, WMO, ACRE, and the Global Framework for Climate Services (GFCS) developed the Indian Ocean Data Rescue (INDARE) initiative (<http://www.wmo.int/pages/prog/wcp/wcdmp/documents/INDAREImplementationPlan.pdf>). More recently, the WMO Commission for Climatology formed a task team on data rescue and instigated a portal to encourage coordination hosted by the Dutch

Meteorological Service KNMI (www.idare-portal.org/). These efforts have substantially increased historical meteorological digital data availability. They also identified key requirements essential to long-term preservation of the rescued data, including: cataloguing; having an institutional champion; and provision of necessary equipment to capture the data (Brönnimann et al., 2006; Tan et al., 2004).

Many different types of data can be valuable in addition to NMS stations (Allan et al., 2016). For example, data have been rescued from a French 18th Century Doctor's network (Yiou et al., 2014), and colonial administrations (White et al., 2016). Early records, and those from countries without a mature NMS, are most likely to come from such non-traditional sources, e.g., the National Oceanic and Atmospheric Administration (NOAA)'s Climate Data Modernization Program resulted in the keying of 450 stations across the United States based on data from the US Army Surgeon network, the Smithsonian Institute, and the US Signal Service spanning the 1820–1892 period. Discovering and securing access to such data is a long-term process that is already ongoing and mature. The broader aspects of CLIMOD can add enormously to the long-term value of data rescue work, by providing a recognised repository that is well used, permitting the easy identification of data that have already been digitised, and highlighting vital applications for the rescued data. Experience from the marine ICOADS effort attests to the value of such holdings in soliciting additional contributions.

4. Data source amalgamation

To ensure consistency of processing for derived analyses and applications, CLIMOD will work from the most original basic data sources available. Where this consists of images of original data, a link back to the image and source should be retained and such 'gold standard' data provenance should be an aspiration for all sources. To minimise transcription errors occurring on conversion to the archive data model, data and metadata should be ingested and archived in perpetuity in the native format of the data supplier with relevant provenance metadata to ensure future usability. CLIMOD should take all relevant data regardless of quality or format. However, certain minimum data standards are likely to be required for inclusion in public-facing holdings — at least sufficient Discovery metadata (name, latitude, longitude, and ideally elevation) to enable users to appropriately contextualise the data. Such Discovery metadata might become available later so it is important that all data are retained and archived even if not currently useable. This data ingest model permits reprocessing and reconversion of the data, and also sets the lowest possible hurdle to data submission. This is a substantial advantage because data will be submitted from a wide variety of organisations and programs. If higher level products exist (e.g. post quality control or homogenisation) these should be stored in parallel. However, preference should be given to data closer to 'original', such that multiple groups of analysts can subsequently consider the challenging issues of quality control, homogenisation, and interpolation (e.g. Venema et al. 2012), to create datasets and data products.

The data lineage (provenance) of observations should be carried through the entire data archive, and each stage of processing recorded in the metadata associated with the observation. The next logical step is thus to convert the multiple sources to a flexible and extendable common data model of consistent formatting and metadata conventions, that retains all metadata and allows full provenance tracking through the use of Unique Identifiers (UIDs). Each separately provided element of data or information should be assigned a unique identifier (data, station metadata, general source format documentation, etc.) and all associated information linked. This should indicate alternative sources of data at a given timestep if they exist. Converting to a common data model (such as espoused by the Copernicus Climate Change Service, <https://climate.copernicus.eu>) makes intercomparison of the various sources easier but still requires the user to understand the different data sources and use them appropriately. While a necessary step, stopping at this stage is an impediment to usage for all but the most expert users. We then need to create a merged set of holdings.

A large fraction of observing platforms are found in more than one existing data holding, because of past splitting of multivariate observations into different archives and subsequent repeated splitting and merging of different archives (e.g. Rennie et al., 2014). In creating the merged CLIMOD data archive, the source holdings will be prioritised and then merged sequentially. In-depth understanding of each data source will be required to enable a clear decision on the priority order to be made. This may depend on some or all of the following non-exhaustive list: data provenance, data quality level, temporal completeness, number of ECVs supplied, reporting resolution, metadata detail, most recent update, consistency over time, quality control applied. Any

duplication of particular observing platforms across sources will allow quality to be assessed and might fill gaps in the record.

Merging of the source data will be a vital step in producing a final archive valuable to the wider community. It will by necessity be iterative, merging each prioritised source in turn (Figure 3). A high level of automation will be required, using the metadata and data similarity across all variables to appropriately decide whether a given platform is unique or to be merged. However some manual intervention is foreseen, using local knowledge wherever possible. The merging process will capture the decisions made into the metadata, to keep the data provenance and audit trail, including for any manual process. Knowledge gained during manual intervention will be used to update the automatic procedure. Experts would be able to use the resulting metadata to perform a different merge or ensemble of merges matching their own specific requirements.

Merging ECVs and merging time resolutions are both challenging. Merging ECV holdings will require careful attention to ensure physical consistency between variables (supersaturation, rain at $<0^{\circ}\text{C}$ or snow at $>10^{\circ}\text{C}$ etc.). To have consistency across the temporal range (where monthly values are to be based on daily values, which themselves are based on the synoptic reports), extra checks will be required, e.g. the sub-daily report values of temperature do not exceed the daily minimum / maximum. These will be easier when additional metadata have been included with the original reports (e.g., the definition of the climatological/meteorological day, daily averaging method, details on instrumentation). Data supplied as temporal averages will only be

taken as primary sources in the absence of any information at higher temporal resolution.

The final output will undergo quality assurance for gross errors. Any errors identified (e.g., unphysical values) will be removed from the primary data stream but retained within the archive. However, a comprehensive quality control system is not envisaged as part of this primary data product. Quality control processes (manual or automated, e.g. Dunn et al., 2012, Durre et al., 2010, Schneider et al., 2014) are not able to correctly classify an observation in every instance. No matter how carefully the process is designed, errors which either retain bad data points or exclude valid data points will occur, even if infrequently. Furthermore, in some instances the available information will not allow clear decisions to be made. Enforcing one specific set of choices (e.g. 4 sigma vs. 4.5 sigma trimming) removes an important means of quantifying uncertainty, and risks biasing all downstream applications. Each user of the data product will have different requirements for their application of the data. We encourage the development of multiple quality control suites, each tailored to the end data use, which should be made available alongside the primary CLIMOD holdings.

5. Data serving

CLIMOD will become the one-stop-shop for all land meteorological data, ready for scientists and other interested parties to develop climate data records and derived products. It is expected to be a dataset roughly the same size as the current ICOADS holdings. As such, it is essential that the combined CLIMOD be built on a strong base

of machine readable metadata to enable powerful searching and subsetting capabilities across timescales and ECVs. The ability to ingest, update, and communicate a range of new data quickly will facilitate the development of timely climate data products and services. This includes access to and timely (with little delay) quality assessments of synoptic reports, daily summaries (including Daily CLIMATs when these become available), and monthly CLIMAT summaries exchanged via the WMO Information System. CLIMOD should be capable of providing multiple formats to suit user needs. To build user confidence in the database, full provenance metadata will need to be servable alongside the data.

The proposed CLIMOD holdings represent a requisite high-quality database for the production of climate datasets and value-added products (see Boxes). These subsequent uses would have to consider questions of quality control, quality assurance, and record homogeneity, as discussed in section 4. Using the database without taking into account these issues would be actively discouraged. To aid such applications, metadata may be appended where available, giving measurement uncertainties, findings of analysts and other aspects similar to the ICOADS Value Added Database (IVAD, Freeman et al. 2016).

To be sustainable CLIMOD will need to be well-used. Primary users are likely to be experts and practitioners accessing the raw observations and deriving data products suitable for specific applications. These users should be engaged from the outset to ensure that CLIMOD meets their needs. Users may inform prioritisation with regards to data rescue and sourcing, data formats, data access options, and data provenance.

Broad usage will require data discovery metadata and associated data exploration tools. This may include UIDs, tools to subset and visualize holdings by region, timescale, ECV, source, and provision of inventories. Discovery metadata should be ISO19115 compliant. Existing resources such as the National Center for Atmospheric Research (NCAR) Research Data Archive (<http://rda.ucar.edu/>) would be able to perform user-defined sub-setting operations. Metadata exploration could be enabled through the WMO Observing Systems Capability Analysis and Review tool (<https://oscar.wmo.int/surface>) which would also ensure compliance with emerging WMO Integrated Global Observing System (WIGOS) metadata standards, or the NOAA National Centers for Environmental Information (NCEI) Historical Observing Metadata Repository (<http://www.ncdc.noaa.gov/homr>) interface. Although initially machine readable metadata may be extremely limited for many stations, in the longer-term such tools may facilitate appending of richer metadata for stations.

A range of download tools will be required to cater to diverse users. Regional mirrors would aid data provision and also ensure accessibility. Some users will prefer web-based access tools, whereas frequent users are likely to need service via ftp, OPeNDAP, or similar. Many applications will require timely access to data updates, so a regular update cycle will be needed. An increasing number of applications for climate services require timely updates (within days to a month).

It is likely that users will find errors in the data and its presentation. A mechanism will be required for users to be able to report data issues. This should include a response system that enables innovations to improve the database to be enacted in a timely

manner. The NOAA NCEI Datzilla ticket system (<https://datzilla.srcc.lsu.edu/datzilla/>) is an example of such a facility and has led to corrections including poor station merge decisions and bad data segments. Users also need to be made aware of updates. This requires a clear version control protocol. One or more of monthly status reports, blog posts highlighting updates and applications, and provision of educational resources would help build a loyal user community.

6. Governance and modus operandi

CLIMOD clearly constitutes a long-term effort which would benefit substantially from sustained international cooperation. Such an explicitly international effort would reduce the risk of a single point of failure (whereby the decision of a single funder or host organisation may imperil continued function) and also make it clear that the set of holdings belonged to the global research and applications communities. This might reduce anxieties over data intellectual property rights amongst potential data contributors. A broadly-based international steering committee of recognized experts would be advisable, answering to a body such as the Global Climate Observing System (GCOS), World Climate Research Program or WMO Commission for Climatology. The steering committee should provide advice on data rescue, policy, processing and serving to cover the full range of activities.

Such international and sustained collaboration is only likely to result if sufficient resource support is forthcoming to facilitate the work on a long-term basis. Regular forums or meetings of participants would be required to ensure a coordinated approach,

to discuss progress and challenges, and to assure continued cooperation. The effort will require the input of scientists, software engineers, data access rights experts, data base / data life cycle experts, project managers and associated support staff. Although initial work may benefit from a research project, long-term continuity will be best achieved through sustained support from NMSs or large international programs, such as Copernicus. To that end, we strongly welcome the recent Copernicus Climate Change Service Contract awarded (led by the lead author on this paper and in partnership with NOAA NCEI), which shall, over the coming four years, start to make CLIMOD a reality and as a further benefit provide integrated access to land and surface marine observations using a common data model.

Data policy needs to be clearly articulated and to have broad buy-in. Internationally, efforts have been made to secure access to observations resulting in WMO Resolutions 40 and 60 which pertain respectively to real-time and historical data exchange. These govern the routine global exchange of meteorological data by all nations, but are far from perfect since they allow institutions to place restrictions on data use and reuse. CLIMOD must operate within these boundaries while working with others to advocate for improved access to data. Usage policy must be agreed upon and clearly articulated to end-users. It may be that different data streams could have different levels of restriction on use and reuse. There is no value in publically serving data that are not useable at a minimum for academic and research purposes. While recognising the substantial data policy hurdles to be overcome, the long-term aspiration should be that data be shared openly and without restriction, with a share-and-share-alike onwards usage policy. Regardless, acknowledgement by citation and DOI (to a list of

contributors), and where necessary to relevant citations of value-added products, will be required.

7. Summary

We have outlined a framework for the construction of a comprehensive database of land meteorological holdings that are integrated across timescales and ECVs. Such integration is essential to understand and address the challenges of climate change and variability. Observations are the foundation upon which datasets, products and services are built, and based upon which decisions can be made. The time is right to address the challenge. However, outlining and agreeing on a broad-scale vision is necessary but not sufficient. A sustained and sufficiently-resourced effort is needed, covering the entire processing chain from data sourcing (including rescuing) through to amalgamation and provision to the community. The need is clearly articulated in the latest GCOS Implementation Plan (GCOS, 2016). The question that remains is whether sufficient organisations will ‘step up to the plate’ on a sustained basis to take on the meteorological databank challenge.

Acknowledgments

The work was facilitated by grant 16/CW/3801 from Science Foundation Ireland. Dr. Jan Rigby (Maynooth University) provided expert advice on the mosquito borne diseases box. Prof. Aiguo Dai (SUNY-Albany) provided useful input and a figure for the drought box. Deb Misch (Telesolv Consulting) and Sara Veasey of NOAA NCEI graphics team are acknowledged for substantial help in creating publication-ready figures. Prof. Rob Allan is supported by a combination of funding from the Joint BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101), the European Union's Seventh Framework Programme (FP7) European Reanalysis of Global Climate Observations 2 (ERA-CLIM2) project and the Climate Science for Service Partnership (CSSP) China under the Newton Fund. He also acknowledges the University of Southern Queensland, Toowoomba, Australia, where he is an Adjunct Professor. Robert Dunn and Kate Willett were supported by the Joint BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). Linden Ashcroft is supported by EU-FP7- SPACE-2013- 1 project Uncertainties in Ensembles of Regional Reanalyses (UERRA, Grant agreement no.: 607193). Elizabeth Kent was supported by NERC by grant NE/J020788/1. John Coll acknowledges funding provided by the Irish Environmental Protection Agency under project 2012-CCRP-FS.11. Gilbert Compo is supported by the NOAA Climate Program Office and the US Department of Energy Office of Science Biological and Environmental Research Program. Petra Pearce was supported by NIWA Core Funding under the Climate Present and Past project (CAOA1701), the New Zealand Deep South National

Science Challenge (BDS16101), and NOAA. Stefan Bronnimann acknowledges funding from the Swiss National Science Foundation (Project 205121_169676).

References

- Adler, R.F. et al., 2003: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, **4**(6): 1147–1167. DOI: 10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.
- Allan, R., P. Brohan, G.P. Compo, R. Stone, J. Luterbacher, and S. Brönnimann, 2011: The International Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Amer. Met. Soc.*, **92**, 1421-1425. doi: 10.1175/2011BAMS3218.1.
- Allan, R., P. Brohan, S. Leduc, and K.M. Willett, 2012: Climate data rescue and stewardship: Filling the gaps in 21st Century needs [in “State of the Climate in 2011”]. *Bull. Amer. Met. Soc.*, **93** (7), S12-S13.
- Allan, R. et al., 2016: Towards integrated historical climate research: the example of ACRE (Atmospheric Circulation Reconstructions over the Earth). *WIREs Clim. Change*, **7**, 164–174.
- Berry, I.L., M.D. Shanklin, and H.D. Johnson, 1964: Dairy shelter design based on milk production decline as affected by temperature and humidity *Trans. Am. Soc. Ag. Eng.*, **7** 329–31
- Blunden, J., and D.S. Arndt, 2016: State of the Climate in 2015. *Bull. Amer. Meteor. Soc.*, **97**, Si–S275, doi: 10.1175/2016BAMSSStateoftheClimate.1.

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zem, 2014: The concept of Essential Climate Variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, **95**, 9, 1431-1443. DOI: 10.1175/BAMS-D-13-00047.1.

Brázdil R., P. Dobrovolný, J. Luterbacher, A. Moberg, C. Pfister, D. Wheeler, and E. Zorita, 2010: European climate of the past 500 years: new challenges for historical climatology. *Climatic Change* **101**(1-2): 7–40, DOI:10.1007/s10584-009-9783-z.

Brönnimann, S. et al., 2006: A guide for digitising manuscript climate data, *Clim. Past*, **2**, 137-144, doi:10.5194/cp-2-137-2006.

Campbell, J.B. and R.H. Wynne, 2011: Introduction to remote sensing. Guilford Press.

Compo, G.P., et al., 2011: [The Twentieth Century Reanalysis Project](#). *Quarterly J. Roy. Meteorol. Soc.*, **137**, 1-28. DOI: 10.1002/qj.776.

Cram, T.A., et al., 2015: The International Surface Pressure Databank version 2. *Geoscience Data Journal*, **2**, 31-46. DOI: 10.1002/gdj3.25.

Dai, A., 2011: Drought under global warming: a review, *WIREs Clim. Change*, **2**, 45-65, 10.1002/wcc.81

Dai, A. and Zhao, T., 2016: Uncertainties in historical changes and future projections

of drought. Part I: estimates of historical drought changes, *Climatic Change*,

doi:10.1007/s10584-016-1705-2

Dee, D. P. et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553–597. doi:10.1002/qj.828

Dee, D.P., M. Balmaseda, G. Balsamo, R. Engelen, A.J. Simmons, and J.N. Thépaut, 2014: Toward a consistent reanalysis of the climate system. *Bull. Amer. Meteor. Soc.*, **95**(8), 1235-1248.

Dolman, A. J, A. Belward, S. Briggs, M. Dowell, S. Eggleston, K. Hill, C. Richter and A. Simmons, 2016: A post-Paris look at climate observations. *Nature Geo.*, **9**, 646, doi:10.1038/ngeo2785

Dunn, R.J.H., K.M. Willett, P.W. Thorne, E.V. Woolley, I. Durre, A. Dai, D.E. Parker, and R.S. Vose, 2012: HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Clim Past*, **8**(5), 1649–1679, DOI: 10.5194/cp-8-1649-2012.

Dunn, R.J.H., N.E. Mead, K.M. Willett, and D.E. Parker, 2014: Analysis of heat stress in UK dairy cattle and impact on milk yields. *Environmental Research Letters*, **9**, 064006.

Durre, I., M.J. Menne, B.E. Gleason, T.G. Houston, and R.S. Vose, 2010: Comprehensive Automated Quality Assurance of Daily Surface Observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, doi: 10.1175/2010JAMC2375.1.

Edwards, P. N. 2011: A Vast Machine. Computer Models, Climate Data and the Politics of Global Warming. MIT Press, Cambridge.

Flato, G., et al., 2013: Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K.

Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 741–866, doi:10.1017/CBO9781107415324.020

Freeman, E. et al., 2016: ICOADS Release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.* doi:10.1002/joc.4775.

GCOS, 2016, The Global Observing System for Climate: Implementation Needs, GCOS-200/GOOS-214, available via <http://www.wmo.int/pages/prog/gcos/index.php?name=Publications> (accessed 11/4/16)

Hales, S., P. Weinstein, and A. Woodward, 1996, Dengue fever epidemics in the South Pacific: driven by El Niño Southern Oscillation?, *The Lancet*, **348**, 1664-1665, DOI: 10.1016/S0140-6736(05)65737-6

Hartmann, D.L. et al., 2013: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 159–254, doi:10.1017/CBO9781107415324.008.

Hay, S.I., et al., 2002: Climate change and the resurgence of malaria in the East African highlands. *Nature*, **415**, 905–909

Kennedy, J.J. et al., 2010: How do we know the world has warmed? in: 2. "Global Climate," in: "State of the Climate in 2009". *Bull. Amer. Meteor. Soc.*, **91** (7), S26

Kent, E.C. et al., 2016: A call for new approaches to quantifying biases in observations of sea-surface temperature, *Bull. Amer. Meteor. Soc.*, early online release. DOI: <http://dx.doi.org/10.1175/BAMS-D-15-00251.1>.

Klein Tank, A.M.G. et al., 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.*, **22**, 1441-1453.

Klok, E.J., and A.M.G. Klein Tank, 2009. Updated and extended European dataset of daily climate observations. *Int. J. Climatol.*, **29**, 1182–1191. DOI: 10.1002/joc.1779.

Kobayashi, S. et al., 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Japan*, **93**, 5-48, doi: 10.2151/jmsj.2015-001.

Lawrimore, J.H., M.J. Menne, B.E. Gleason, C.N. Williams, D.B. Wuertz, R.S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3, *J. Geophys. Res.*, **116**, D19121

Lorrey, A. 2011: Improving Pacific Island meteorological data rescue and data visualization capabilities through involvement in emerging climate research programmes. *APN Science Bulletin*, **1**, 64.

McMichael, A. J., R.E. Woodruff and S. Hales, 2006: Climate change and human health: present and future risks, *The Lancet*, **367**, 859-869, DOI: [http://dx.doi.org/10.1016/S0140-6736\(06\)68079-3](http://dx.doi.org/10.1016/S0140-6736(06)68079-3)

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, **29**, 897-910, doi:10.1175/JTECH-D-11-00103.1.

Morice, C.P., J.J. Kennedy, N.A. Rayner, and P.D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, *J. Geophys. Res.*, **117**, D08101, doi:10.1029/2011JD017187.

Page C.M. et al., 2004: Data rescue in the Southeast Asia and South Pacific Region.

Bull. Amer. Meteor. Soc., **85**: 1483–1490, DOI: 10.1175/BAMS-85-10-1483.

Parham, P.E. and E. Michael, 2010, Modeling the Effects of Weather and Climate

Change on Malaria Transmission, *Environ Health Perspect*, **118**, 620-626.

<http://dx.doi.org/10.1289/ehp.0901256>

Parker, D. E., 1994: Effects of changing exposure of thermometers at land stations. *Int.*

J. Climatol., **14**: 1–31. doi:10.1002/joc.3370140102

Rennie, J.J. et al., 2014: The international surface temperature initiative global land

surface databank: monthly temperature data release description and methods. *Geosci.*

Data Journal, DOI: 10.1002/gdj3.8

Rienecker, M.M. et al., 2011: MERRA: NASA's modern-era retrospective analysis for

research and applications. *J. Clim.* **24**, 3624–3648. [http://dx.doi.org/10.1175/JCLI-D-11-](http://dx.doi.org/10.1175/JCLI-D-11-00015.1)

00015.1.

Saha, S., et al., 2010: The NCEP climate forecast system reanalysis. *Bull. Am.*

Meteorol. Soc. **91**, 1015–1057. <http://dx.doi.org/10.1175/2010BAMS3001.1>

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCP's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology* **115**(1): 15–40. DOI: 10.1007/s00704-013-0860-x.

Tan, L.S., S. Burton, R. Crouthamel, A. Van Engelen, R. Hutchinson, L. Nicodemus, T.C. Peterson, and F. Rahimzadeh, 2004: Guidelines on Climate Data Rescue - World Meteorological Organization - Technical Document No. 1210.

Trenberth K. E., A. Dai, G. van der Schrier, P.D. Jones, J. Barichivich, K.R. Briffa, and J. Sheffield, 2014: Global warming and changes in drought, *Nature Climate Change*, **4**, 17–22, doi:10.1038/nclimate2067

Van Den Besselaar, E.J.M., et al., 2015: International Climate Assessment & Dataset: Climate Services across Borders. *Bull. Amer. Met. Soc.*, **96**(1), 16–21. DOI: 10.1175/BAMS-D-13-00249.1.

Venema, V.K.C. et al., 2012: Benchmarking homogenization algorithms for monthly data. *Clim. Past*, **8**, 89–115, doi:10.1063/1.4819690

White, S, Pfister, C. and Mauelshagen, F. (Eds), 2016: *The Palgrave Handbook of Climate History*. Palgrave Macmillan UK (In press), 544 pp. ISBN 978-1-137-43020-5.

Williamson, F., R. Allan, A.D. Switzer, J.C.L. Chan, R.J. Wasson, R. D'Arrigo, and R. Gartner, 2015: New Directions in Hydro-Climatic Histories: Observational Data Recovery, Proxy Records and the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative in Southeast Asia. *Geoscience Letters*, **2:2**, <http://link.springer.com/content/pdf/10.1186/s40562-015-0018-z.pdf> DOI 10.1186/s40562-015-0018-z.

Yiou, P., M. Boichu, R. Vautard, M. Vrac, S. Jourdain, E. Garnier, F. Fluteau, and L. Menut, 2014: Ensemble meteorological reconstruction using circulation analogues of 1781–1785. *Clim. Past*, **10**, 797-809, doi:10.5194/cp-10-797-2014. Images available: <http://meteo.academie-medecine.fr/>

Zargar, A., Rehan Sadiq, Bahman Naser, Faisal I. Khan, 2011: A review of drought indices, *Environmental Reviews*, **19**, 333-349, 10.1139/a11-013

Sidebars (boxes)

[Editorial note: Note that these boxes were designed to support the totality of the piece and therefore their placement can be made on aesthetic grounds anywhere after first mention in Section 2 were the piece to be published.]

Box 1 - Mosquito-borne disease research

Mosquito-borne diseases such as Dengue fever, Zika virus, West Nile virus, and Malaria are each transmitted by different mosquito species. Each species has a certain climatic range which serves to limit the disease range. Mosquitos require standing water sources to reproduce and can only survive in certain thermal ranges. Hence, at a minimum, temperature and precipitation data are required (at least away from areas with human mediated standing water sources) to help understand disease outbreaks. Studies have investigated links between meteorology, climate and outbreaks of various mosquito-borne diseases (e.g. Parham and Michael, 2010, McMichael et al., 2006, Hales et al., 1996, Hay et al., 2002). However, climate is only one factor in historical (and future) prevalence of such diseases. Epidemiological studies would substantively benefit from CLIMOD enabling access to and an analysis of historical frequency of favourable conditions for the different mosquito species to improve understanding of past outbreaks. This does not address remaining issues surrounding access to other required information such as disease occurrence monitoring that also need to be addressed by other expert communities.

Box 2 - Drought monitoring

The study of meteorological drought is a complex issue. Drought is directly caused by a deficit of rainfall, but is more than simply a rainfall deficit. There are a large variety of drought definitions, e.g., meteorological, agricultural, hydrological, or socio-economic drought, each of which has distinct impacts (Dai, 2011). Observational records of drought are uncertain, and the Fifth Assessment Report of IPCC concluded that *“Confidence is low for a global-scale observed trend in drought or dryness (lack of rainfall) since the middle of the 20th century, owing to lack of direct observations, methodological uncertainties and geographical inconsistencies in the trends.”* (Hartmann et al., 2013). Most drought indices make use of two or more ECVs (Zargar et al., 2011). The availability of multivariate surface holdings through CLIMOD that included drought-relevant parameters such as temperature, humidity, and winds in addition to rainfall would greatly facilitate historical and ongoing observationally-based drought research activities and help to reconcile apparent discrepancies in historical estimates of drought prevalence which depend both on data sources considered (Trenberth et al., 2014, Dai and Zhao, 2016) and choice of specific indices (Dai and Zhao, 2016; Figure 4).

Box 3 - Heat Stress

The effect of heat waves on flora and fauna can be studied using solely dry-bulb air temperature. However, the effect of moisture in the air is an important measure of the thermal comfort of animals (including humans) who rely on the evaporation of water (sweating or panting) to thermo-regulate. High relative humidity at high temperatures limits the rate at which evaporation can occur, leading to heat stress or in extreme circumstances, fatal over-heating. Conversely at low temperatures, moister air can make the body feel cooler through more efficient conduction of heat away from the skin and more energy required to warm the moist air close to the skin. An additional ECV which has value in the study of heat stress is the local windspeed.

A common measure of heat stress is the Temperature-Humidity Index (THI), which was initially developed for humans but has subsequently been shown to correlate strongly with the impact of heat stress on dairy cattle (Berry et al., 1964). By combining air and dewpoint temperature from the sub-daily multivariable HadISD dataset, Dunn et al. (2014) investigated the changes in the THI for UK dairy cattle since 1973. Although on average there are only a few days per year where cattle would experience heat stress, during the heat wave years of 2003 and 2006, this exceeded 5 days or more (Figure 5). A shift in the distribution of THI between the early and late periods was also shown. CLIMOD would allow these kinds of studies to be conducted on a larger scale for a wider range of environments and adaptations, including the assessment of urban heat stress, heat sensitivity in crops, and vulnerability thresholds for plants and animals.

Figure captions

Figure 1. Summary of the International Surface Temperature Initiative merging of 67 sources ranging from single sites to large collections and short records to centennial records (small maps a representative subset of these, additional maps available via www.surface temperatures.org) to a single merged set of holdings of >35,000 unique stations (large map). For further details see Rennie et al., 2014.

Figure 2. Pictorial representation of CLIMOD in the broader context. The left hand side funnel has disparate data sources of different temporal resolution (sub-daily, daily, monthly; colors) and number of records / ECV completeness of holdings (sizes). Within CLIMOD is the processing of these disparate sources to a common format and their merging across ECVs and within and between timescales that leads to the production of a set of unified basic data holdings (black, grey and white stacked blocks each representing integrated holdings at distinct reporting timescales). These integrated holdings will enable development of high-quality downstream products that consider issues such as quality control, homogenisation etc. and applications to end-users (right hand side).

Figure 3. Example for an Argentinian station (Posadas) of results of merging several sources of monthly mean surface air temperature as described in Rennie et al. (2014). Different colors denote distinct source decks and their identifiers in these source decks are given in the inline key. Note that source deck lines often overplot one another.

Figure 4. Illustration of the effects on the Palmer Drought Severity Impact of inclusion (purple) and exclusion (blue) of temperature and vapour pressure data. More negative

values denote an increase in drought prevalence. Also shown are available observations of soil moisture from the top 1m which more closely track an index which takes a multivariate approach. Figure courtesy A. Dai.

Figure 5. Normalized distributions of daily Temperature-Humidity Index (THI) during June–August 1993 (a cool summer), 2003 and 2006 (warm summers), compared to the distribution over 1995–2009. The fitted curves are Gaussian with skew and kurtosis components estimated using a Gauss-Hermite series. (Figure 5 from Dunn et al. 2014).

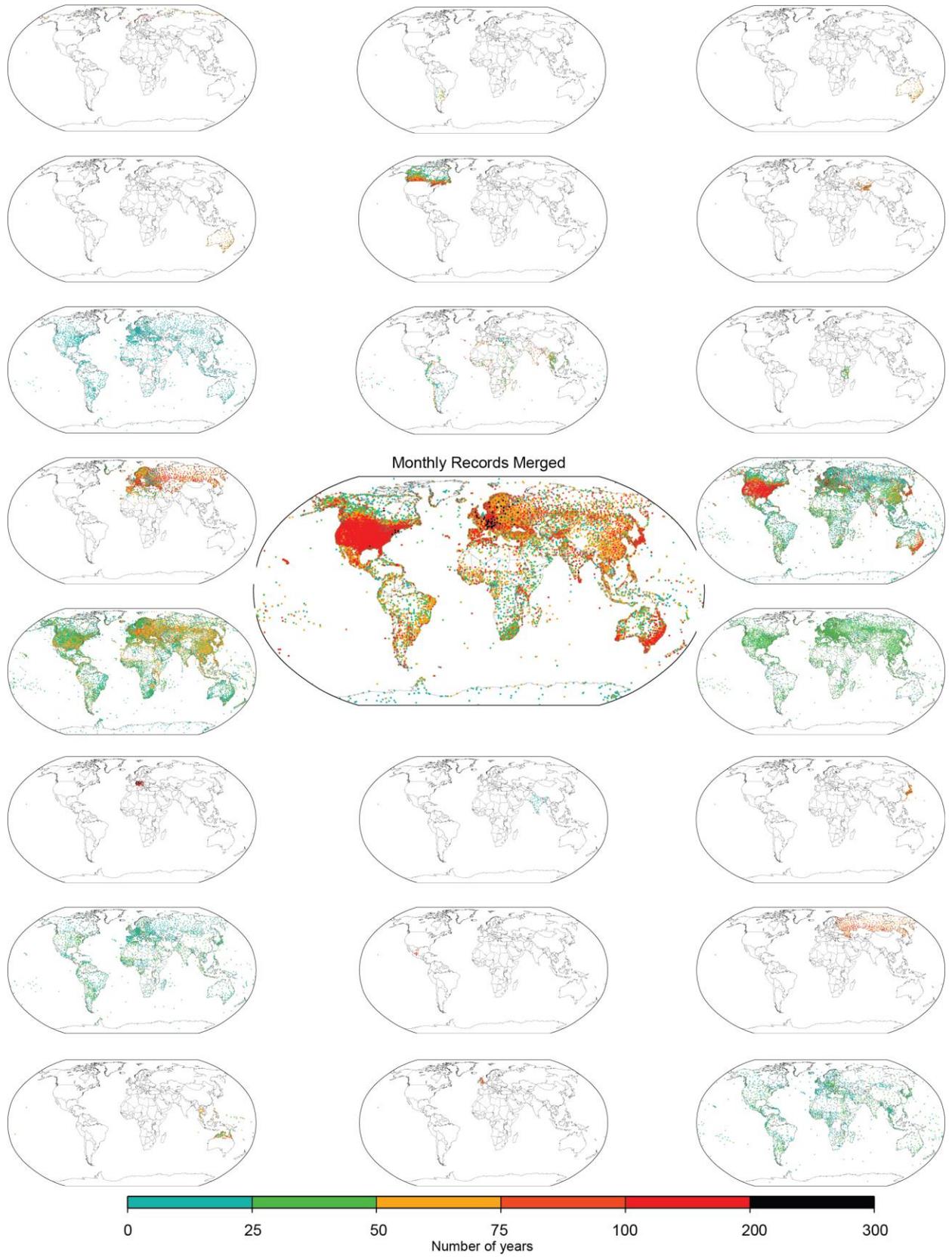


Figure 1. Summary of the International Surface Temperature Initiative merging of 67 sources ranging from single sites to large collections and short records to centennial records (small maps a representative subset of these, additional maps available via www.surface temperatures.org) to a single merged set of holdings of >35,000 apparently unique stations (large map). For further details see Rennie et al., 2014.

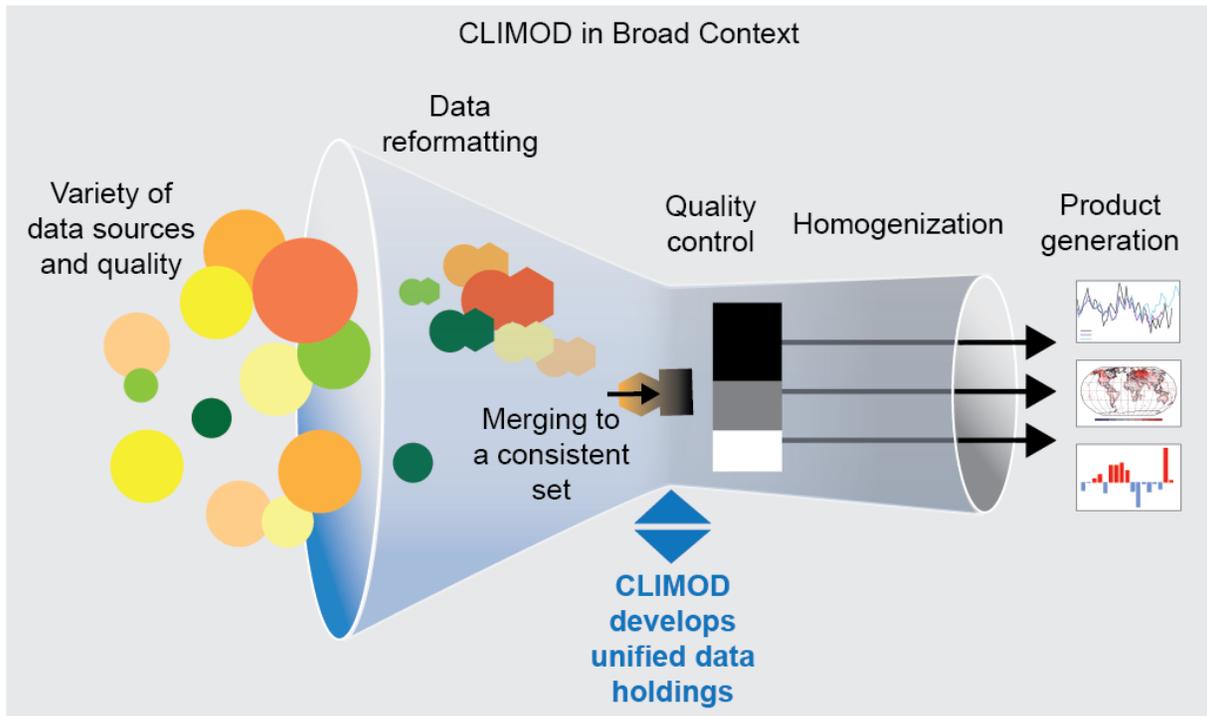


Figure 2. Pictorial representation of CLIMOD in the broader context. The left hand side funnel has disparate data sources of different temporal resolution (sub-daily, daily, monthly; colors) and number of records / ECV completeness of holdings (sizes). Within CLIMOD is the processing of these disparate sources to a common format and their merging across ECVs and within and between timescales that leads to the production of a set of unified basic data holdings (black, grey and white stacked blocks each representing integrated holdings at distinct reporting timescales). These integrated holdings will enable development of high-quality downstream products that consider issues such as quality control, homogenisation etc. and applications to end-users (right hand side).

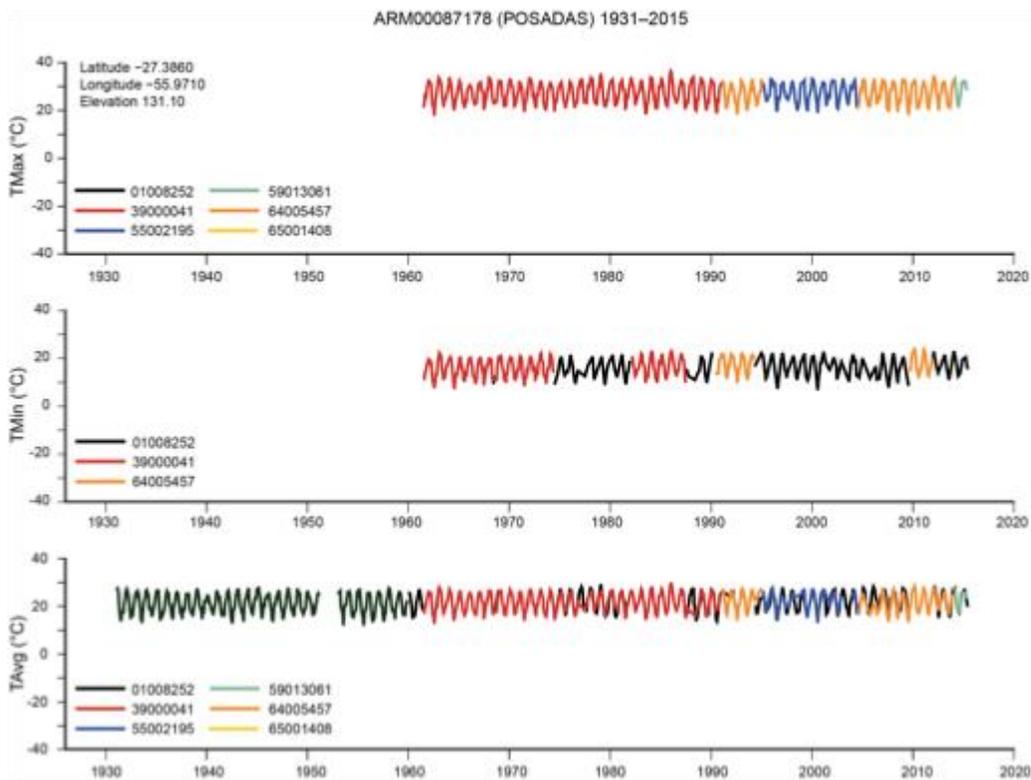


Figure 3. Example for an Argentinian station (Posadas) of results of merging several sources of monthly mean surface air temperature as described in Rennie et al. (2014). Different colors denote distinct source decks and their identifiers in these source decks are given in the inline key. Note that source deck lines often overplot one another.

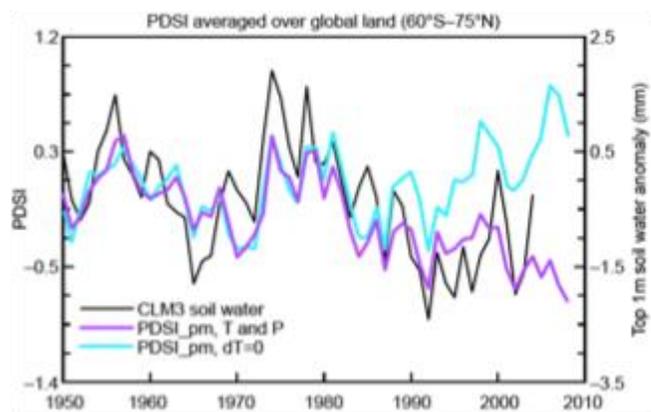


Figure 4. Illustration of the effects on the Palmer Drought Severity Index of inclusion (purple) and exclusion (blue) of temperature and vapour pressure data. More negative values denote an increase in drought prevalence. Also shown are available observations of soil moisture from the top 1m which more closely track an index which takes a multivariate approach. Figure courtesy A. Dai.

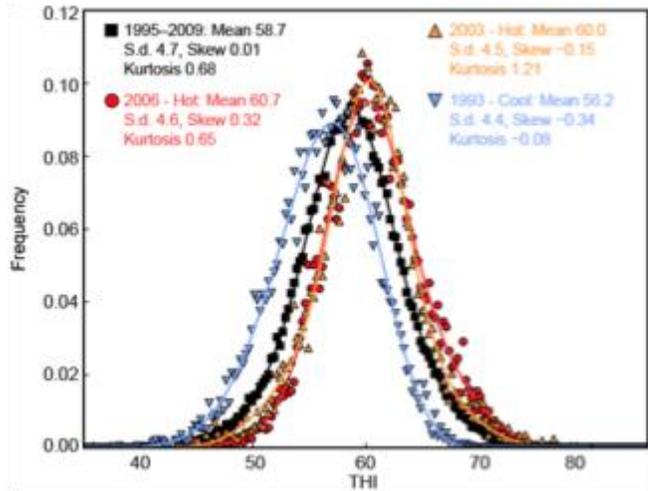


Figure 5. Normalized distributions of daily Temperature-Humidity Index (THI) during June–August 1993 (a cool summer), 2003 and 2006 (warm summers), compared to the distribution over 1995–2009. The fitted curves are Gaussian with skew and kurtosis components estimated using a Gauss-Hermite series. (Figure 5 from Dunn et al. 2014).