




RESEARCH ARTICLE

Towards a global land surface climate fiducial reference measurements network

P. W. Thorne¹  | H. J. Diamond² | B. Goodison³ | S. Harrigan^{4,5}  | Z. Hausfather⁶  |
N. B. Ingleby⁵ | P. D. Jones^{7,8} | J. H. Lawrimore⁹ | D. H. Lister⁷ | A. Merlone¹⁰ | T. Oakley^{11,12}
| M. Palecki⁹ | T. C. Peterson¹³ | M. de Podesta¹⁴ | C. Tassone¹² | V. Venema¹⁵ | K. M. Willett¹¹

¹Irish Climate Analysis and Research Unit,
Department of Geography, Maynooth University,
Maynooth, Ireland

²National Oceanic and Atmospheric
Administration's Air Resources Laboratory, Silver
Spring, Maryland

³WMO Global Cryosphere Watch, Ottawa,
Ontario, Canada

⁴Centre for Ecology and Hydrology,
Wallingford, UK

⁵European Centre for Medium-Range Weather
Forecasts, Reading, UK

⁶Energy and Resources Group, University of
California, Berkeley, California

⁷Climatic Research Unit, School of Environmental
Sciences, University of East Anglia, Norwich, UK

⁸Center of Excellence for Climate Change
Research, Department of Meteorology, King
Abdulaziz University, Jeddah, Saudi Arabia

⁹National Oceanic and Atmospheric
Administration's National Centers for
Environmental Information, Asheville, North
Carolina

¹⁰Istituto Nazionale di Ricerca Metrologica and
Istituto di Scienze dell'Atmosfera e del Clima del
Consiglio Nazionale delle Ricerche, Torino, Italy

¹¹Met Office, Exeter, UK

¹²Global Climate Observing System Secretariat,
Geneva, Switzerland

¹³WMO Commission for Climatology, Asheville,
North Carolina

¹⁴National Physical Laboratory, Teddington, UK

¹⁵Meteorological Institute, University of Bonn,
Bonn, Germany

Correspondence

P. W. Thorne, Irish Climate Analysis and
Research Unit, Department of Geography,
Maynooth University, Maynooth, Ireland.
Email: peter@peter-thorne.net

There is overwhelming evidence that the climate system has warmed since the instigation of instrumental meteorological observations. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change concluded that the evidence for warming was unequivocal. However, owing to imperfect measurements and ubiquitous changes in measurement networks and techniques, there remain uncertainties in many of the details of these historical changes. These uncertainties do not call into question the trend or overall magnitude of the changes in the global climate system. Rather, they act to make the picture less clear than it could be, particularly at the local scale where many decisions regarding adaptation choices will be required, both now and in the future. A set of high-quality long-term fiducial reference measurements of essential climate variables will enable future generations to make rigorous assessments of future climate change and variability, providing society with the best possible information to support future decisions. Here we propose that by implementing and maintaining a suitably stable and metrologically well-characterized global land surface climate fiducial reference measurements network, the present-day scientific community can bequeath to future generations a better set of observations. This will aid future adaptation decisions and help us to monitor and quantify the effectiveness of internationally agreed mitigation steps. This article provides the background, rationale, metrological principles, and practical considerations regarding what would be involved in such a network, and outlines the benefits which may accrue. The challenge, of course, is how to convert such a vision to a long-term sustainable capability providing the necessary well-characterized measurement series to the benefit of global science and future generations.

Funding information

German Science Foundation, Grant/Award number: VE366/8; EMRP, Grant/Award number: ENV58; BEIS/DEFRA, Grant/Award number: GA01101; GAIA-CLIM, Grant/Award number: 640276

1 | INTRODUCTION: HISTORICAL OBSERVATIONS, DATA CHALLENGES, AND HOMOGENIZATION

A suite of meteorological parameters has been measured using meteorological instrumentation for more than a century (e.g., Becker et al., 2013; Jones, 2016; Menne, Durre, Vose, Gleason, & Houston, 2012; Rennie et al., 2014; Willott et al., 2013, henceforth termed “historical observations”). Numerous analyses of these historical observations underpin much of our understanding of recent climatic changes and their causes (Hartmann et al., 2013). Taken together with measurements from satellites, weather balloons, and observations of changes in other relevant phenomena, these observational analyses underpin the Intergovernmental Panel on Climate Change conclusion that evidence of historical warming is “unequivocal” (Intergovernmental Panel on Climate Change, 2007, 2013).

Typically, individual station series have experienced changes in observing equipment and practices (Aguilar, Auer, Brunet, Peterson, & Wieringa, 2003; Brandsma & van der Meulen, 2008; Fall et al., 2011; Mekis & Vincent, 2011; Menne, Williams Jr., & Palecki, 2010; Parker, 1994; Sevruk, Ondrás, & Chvíla, 2009). In addition, station locations, observation times, instrumentation, and land use characteristics (including in some cases urbanization) have changed at many stations. Collectively, these changes affect the representativeness of individual station series, and particularly their long-term stability (Changnon & Kunkel, 2006; Hausfather et al., 2013; Karl, Williams Jr., Young, & Wendland, 1986; Quayle, Easterling, Karl, & Hughes, 1991). Metadata about changes are limited for many of the stations. These factors impact our ability to extract the full information content from historical observations of a broad range of essential climate variables (ECVs) (Bojinski et al., 2014). Many ECVs, such as precipitation, are extremely challenging to effectively monitor and analyse due to their restricted spatial and temporal scales and globally heterogeneous measurement approaches (Goodison, Louie, & Yang, 1998; Sevruk et al., 2009).

Changes in instrumentation were never intended to deliberately bias the climate record. Rather, the motivation was to either reduce costs and/or improve observations for the primary goal(s) of the networks, which was most often meteorological forecasting. The majority of changes have been localized and quasi-random in nature and so are

amenable to statistical averaging of their effects. However, there have been regionally or globally systemic transitions specific to certain periods of time whose effect cannot be entirely ameliorated by averaging. Examples include:

- Early thermometers tended to be housed in polewards facing wall screens, or for tropical locales under thatched shelter roofs (Parker, 1994). By the early 20th century better radiation shielding and ventilation control using Stevenson screens became ubiquitous. In Europe, Böhm *et al.* (2010) have shown that pre-screen summer temperatures were about 0.5 °C too warm.
- In the most recent 30 or so years a transition to automated or semi-automated measurements has occurred, although this has been geographically heterogeneous.
- As highlighted in the recent World Meteorological Organization (WMO) SPICE intercomparison (<http://www.wmo.int/pages/prog/www/IMOP/intercomparisons/SPICE/SPICE.html>) and the previous intercomparison (Goodison et al., 1998), measuring solid precipitation remains a challenge. Instrument design, shielding, siting, and transition from manual to automatic all contribute to measurement error and bias and affect the achievable uncertainties in measurements of solid precipitation and snow on the ground.
- For humidity measurements, recent decades have seen a switch to capacitive relative humidity sensors from traditional wet- and dry-bulb psychrometers. This has resulted in a shift in error characteristics that is particularly significant in wetter conditions (Bell, Carroll, Beardmore, England, & Mander, 2017; Ingleby, Moore, Sloan, & Dunn, 2013).

As technology and observing practices evolve, future changes are inevitable. Imminent issues include the replacement of mercury-in-glass thermometers and the use of third party measurements arising from private entities, the general public, and non-National Met Service public sector activities.

From the perspective of climate science, the consequence of both random and more systematic effects is that almost invariably a post hoc statistical assessment of the homogeneity of historical records, informed by any available metadata, is required. Based on this analysis, adjustments must be applied to the data prior to use. Substantive efforts have been made to post-process the data to create homogeneous long-term records for multiple ECVs

(Mekis & Vincent, 2011; Menne & Williams, 2009; Rohde et al., 2013; Willett et al., 2013, 2014; Yang, Kane, Zhang, Legates, & Goodison, 2005) at both regional and global scales (Hartmann et al., 2013). Such studies build upon decades of development of techniques to identify and adjust for breakpoints, for example, the work of Guy Callendar in the early 20th century (Hawkins & Jones, 2013). The uncertainty arising from homogenization using multiple methods for land surface air temperatures (LSAT) (Jones et al., 2012; Venema et al., 2012; Williams, Menne, & Thorne, 2012) is much too small to call into question the conclusion of decadal to centennial global-mean warming, and commensurate changes in a suite of related ECVs and indicators (Hartmann et al., 2013, their FAQ2.1). Evidence of this warming is supported by many lines of evidence, as well as modern reanalyses (Simmons et al., 2017).

The effects of inhomogeneities are stronger at the local and regional level, may be impacted by national practices complicating homogenization efforts, and are more challenging to remove for sparse networks (Aguilar et al., 2003; Lindau & Venema, 2016). The effects of inhomogeneities are also manifested more strongly in extremes than in the mean (e.g., Trewin, 2013) and are thus important for studies of changes in climatic extremes. State-of-the-art homogenization methods can only make modest improvements in the variability around the mean of daily temperature (Killick, 2016) and humidity data (Chimani et al., 2017).

In the future, it is reasonable to expect that observing networks will continue to evolve in response to the same stakeholder pressures that have led to historical changes. We can thus be reasonably confident that there will be changes in measurement technology and measuring practice. It is possible that such changes will prove difficult to homogenize and would thus threaten the continuity of existing data series. It is therefore appropriate to ask whether a different route is possible to follow for future observational strategies that may better meet climate needs, and serve to increase our confidence in records going forwards. Having set out the current status of data sets derived from ad hoc historical networks, in the remainder of this article, we

propose the construction of a different kind of measurement network: a reference network whose primary mission is the establishment of a suite of long-term, stable, metrologically traceable, measurements for climate science.

The remainder of the article is structured as follows. We begin in section 2 by articulating a view of the observing system as a system of systems and outlining how the reference network adds value within such a framework. We then discuss what the defining features of such a reference network for the land surface would be, touching upon: metrological principles (section 3); station characteristics (section 4); network configuration (section 5); governance and coordination (section 6); and likely financial support requirements (section 7). Having outlined the what, the how, and the how much we then make recourse to a number of existing usage examples from similar networks in section 8 to justify the why. Finally, section 9 considers next steps.

2 | THE SCIENTIFIC RATIONALE FOR A LAND SURFACE FIDUCIAL REFERENCE MEASUREMENTS NETWORK

Climate changes will undoubtedly occur in the future due to both human and natural factors (Intergovernmental Panel on Climate Change, 2013). It is important that we can monitor these changes adequately, so as to enable relevant, timely, responses, and to understand the extent to which mitigation strategies are working. The greater the confidence that society has in these measurements, the more impact they will have. We propose to create a network that is uniquely suited to this purpose.

Our proposal is that a surface climate fiducial reference measurement network will enhance the value of existing observing networks. To see how this would work it is helpful to envisage global measurements as consisting of three tiers (Figure 1) (Thorne et al., 2017): fiducial reference measurement networks; baseline networks; and comprehensive networks.

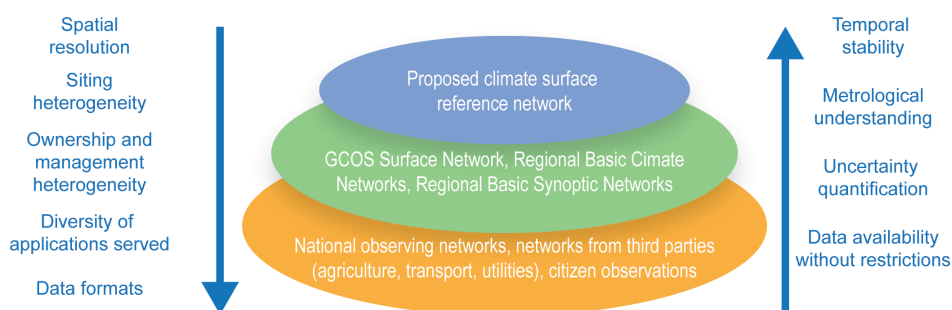


FIGURE 1 Conceptual outline of how surface observational capabilities for climate map onto the tiered system of systems approach of Thorne *et al.* (2017). The tiers from top to bottom are reference, baseline, and comprehensive. Arrows and associated text denote important facets of the measurements that increase as you move down tiers (left-hand side) or up tiers (right-hand side). The network types given for each tier are solely exemplars. To decide where a given network resides requires an in-depth assessment (Thorne et al., 2017). There are, for example, very high quality agricultural networks in some countries, which may fall into higher tiers than indicated in this oversimplified example

The fiducial reference measurement network—which does not currently exist—would provide measurements that are metrologically traceable, with full metadata. It would not only provide unambiguous high-quality time series but serve to validate and enhance trust in the quality of the other networks, as outlined later in the article.

The baseline network consists of a lower long-term quality but much higher-density subset of stations with long-term operational commitment. Of the ~11,000 WMO reported stations, ~4,000 comprise regional basic synoptic networks (RBSNs), and ~3,000 comprise the regional basic climatological networks (RBCNs). RBSN and RBCN overlap substantively in almost all regions and both will possibly eventually be replaced by a regional basic observing networks designation (World Meteorological Organization, 2016). Of the RBCN network stations, ~1,000 are also part of the long-standing GCOS surface network (GSN) (GCOS, 2010). Thus, several thousand stations might be considered as belonging to this tier, and they would deliver well-managed spatially representative measurements.

The comprehensive network would consist of “everything else.” It would include those WMO stations not in the baseline network, and non-national meteorological and hydrological service networks such as agricultural, urban, or transport networks, and citizen observer networks. These might comprise several tens of thousands of stations, most of which would have poor documentation and traceability, but which would show spatial detail and spatio-temporal details beyond the capability of the baseline networks. For monthly temperatures the complete network likely consists of in excess of 30,000 station records (Rennie et al., 2014; Rohde et al., 2013), while for surface precipitation daily records there are likely in excess of 100,000 (Menne et al., 2012, and updates).

The tiered network approach is based on the insight that it is not necessary, economically viable, or technically practical to have reference quality measurements everywhere. Rather, they are required in sufficient locations to build confidence in the remaining observations. The surface fiducial reference measurements network would provide the temporally stable high-quality backbone of the global climate observing system. Aspects of such a tiered network structure are reflected in both the GCOS adequacy report (GCOS, 2015) and the WIGOS manual (WIGOS, 2015). Such a tiered network design is also alluded to in the WMO vision for 2040 (World Meteorological Organization, 2015).

Several exemplars of fiducial reference networks currently exist. Firstly, at a national/continental level, there is the U.S. climate reference network (USCRN) (Diamond et al., 2013). It consists of well-sited surface stations in stable areas away from artificial heat sources with triple-redundant measurements of several ECVs, involving instruments that are calibrated to SI-traceable standards and continuously monitored so that problems can be addressed

quickly. A limited suite of surface ECVs are measured that involve air temperature, precipitation, soil moisture and temperature, relative humidity, and surface radiation. There are other more recently instigated similar national programs such as for Canada (Milewska & Vincent, 2016). A second example is the GCOS reference upper air network (GRUAN) (Bodeker et al., 2016; Seidel et al., 2009). The GRUAN network aims to become a network of 30–40 sites making traceable measurements with quantified uncertainties of the atmospheric column properties. Products are being developed for various radiosondes (Dirksen et al., 2014), frostpoint hygrometers, ozonesondes, and ground-based remote sensing techniques. Finally, the global cryosphere watch (GCW), a component observing system of WIGOS, has recently instigated tiered surface observing networks. Their cryonet sites and stations (Schöner et al., 2016, https://globalcryospherewatch.org/cryonet/site_types.html provides information on the distinction) form the core network, are representative of the surrounding region, and must meet a set of minimum requirements. They measure relevant cryosphere variables to the highest measurement standards currently attainable following documented best practices and with specified data curation and access and exchange protocols. In the following sections, further references will be made to lessons learned from establishing and operating these networks.

3 | METROLOGICAL CONSIDERATIONS OF MEASUREMENT PRACTICES AND ANALYSES

3.1 | Traceability

The *Vocabulaire International de Métrologie* (JGCM, 2012) defines metrological traceability as “the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.”

The absolute requirement of a reference-grade measurement is thus that it be made in such a way that after accounting for all sources of uncertainty it can be concluded that the true value of the measurand lies within the reported uncertainty interval with specified confidence, and that the measurement result is traceable to standards of the system of units (SI) or other standards.

To obtain full measurement traceability, the complete measurement uncertainty must be evaluated through the quantification of the contribution of all sources of uncertainty. Critically, for environmental measurements the uncertainty includes quantities of influence, such as the site effect, shielding, instrument ageing, etc. A fundamental problem with meteorological measurements is that—unlike physical measurements in a laboratory—the “right answer” or “true value” is defined operationally rather than through

laws of physics. For example, LSAT and precipitation vary with height above the ground and thus are not uniquely defined. Furthermore, temperature measurements need to take place within a screen, which reduces measurement errors, but results in characteristic temporal and remaining thermal artefacts on the measured data. For any reference quality measurement, all such factors need to be adequately quantified.

3.2 | Comparability

Measurements of the same set of measurands at multiple locations and/or at different times, even if individually traceable, will still not necessarily be directly comparable. Comparability is attained when two reported measurements differ solely due to any difference in the measurand, independently from the measurement techniques or instruments or siting. Even nominally identical equipment may have variations within and between production batches and different ageing drifts due to differential exposure to environmental conditions. The challenge is to create a diverse network within which comparability of measurements is ensured.

3.3 | Representativity

Representativeness is a key property of a reference measurement. A representative measurement reflects the nature of the measurand across a broader spatial and temporal domain than the immediate measurement location. If a fiducial reference measurement network's purpose is to help constrain and validate more regional measurements from other networks, or measurements from satellites, then it is important to choose sites which optimize the spatial representativeness of the measurements.

3.4 | Practical pathways to implementation of metrological best practices

Certain properties of any reference grade measurement follow from the traceability requirement. The raw data and full metadata must be available and retained to permit reprocessing as required. All processing software should be available under an appropriate access model (ideally fully open-source) that enables full understanding of the processing steps. The uncertainty in each step of the chain must be quantified and reported and any correlations or co-dependencies in uncertainties adequately accounted for. And, finally, there can be no proprietary black-box process which breaks the traceability chain.

Measurement redundancy is one way to assess aspects of both traceability and comparability. By using multiple, co-located traceable instruments to measure the same parameter the resultant data series can be compared. Disagreement between the data series can highlight measurement problems which would be undetectable with a single

sensor, and agreement results in a lower statistical measurement uncertainty. In the USCRN, for some ECVs, the same instruments are used with triple redundancy, with one sensor being replaced annually, so that inaccurate readings or drift from any single instrument can be identified and addressed. USCRN includes a number of 'paired' sites within reasonably close proximity, and also maintains individual sites in Canada and Russia alongside their reference stations, that aid quantification of comparability to their sites (Diamond et al., 2013). Within GRUAN it is measurements of the same measurand with different measurement techniques that is pursued. In both cases the end result is a degree of redundancy, or to use a less pejorative term complementarity, in measurements that builds confidence in the metrological verity of the resulting series.

Regular calibration is another key aspect of the implementation of metrological best practice in a fiducial reference network. Documented and in some cases dedicated calibration procedures must be adopted by all network stations. Knowledge of the components of measurement uncertainty due to the site characteristics and the local quantities of influence also require dedicated research and field campaigns.

Finally, in view of continuous improvements in the state of the art instrumentation, procedures to implement managed changes of the instruments must be adopted. The network should as a result be a key resource to study the value of new and emerging technologies using innovative measuring principles. A number of "super-sites" (sites with extra instrumentation and perhaps associated directly with national observatories and/or national measurement institutes) in the network should therefore be devoted also to continuous research efforts.

4 | WHAT COULD A SURFACE CLIMATE FIDUCIAL REFERENCE NETWORK SITE LOOK LIKE AND HOW WOULD IT OPERATE?

4.1 | Instrumentation

The exact instrumentation deployed at each station would depend upon the agreed remit of the network. From a purely scientific perspective, surface climate fiducial reference network sites should eventually aim to measure all surface atmospheric, atmospheric composition and terrestrial ECVs capable of being measured to reference quality (Table 1). However, we note that the variables to be measured will almost certainly vary geographically.

In terms of practical implementation, there would be considerable value in following the GRUAN ethos of "start small, but start" and thus instigating the network to measure in the first instance a subset of atmospheric ECVs for which metrologically well-characterized measurements are either

TABLE 1 Selected ECVs (from the list published and periodically updated by GCOS), which may be measurable at some or all land climate reference network sites. An initial set of ECVs for which current measurement understanding is high are highlighted by text entries in italics and underlined. This may constitute a starting set of measurements. There may be additional terrestrial ECVs that are measurable at a small subset of selected sites. Note that precipitation includes both liquid and solid precipitation

Domain	Selected GCOS ECVs of relevance
Atmospheric	<i><u>Air temperature</u></i> <i><u>Wind speed and direction</u></i> <i><u>Water vapour</u></i> <i><u>Pressure</u></i> <i><u>Precipitation</u></i> <i><u>Surface radiation budget</u></i>
Composition	<i><u>Carbon dioxide</u></i> <i><u>Methane</u></i> Other long-lived greenhouse gases Ozone (and precursors) Aerosols (and precursors)
Terrestrial	River discharge <i><u>Snow cover</u></i> Permafrost <i><u>Albedo</u></i> Land cover (including vegetation type) Fraction of absorbed photosynthetically active radiation (FAPAR) Leaf area index (LAI) Above-ground biomass Soil carbon Fire disturbance <i><u>Soil moisture</u></i> <i><u>Land surface temperature</u></i>

available or likely to be achievable for little additional effort (highlighted in Table 1). However, it is only with the full suite of ECVs being monitored that full understanding can accrue. Given their significance to multiple application areas, an initial network roll-out should consider at a minimum temperature and total precipitation (both liquid and solid) measurements as mandatory core measurements at every site, along with as many other highlighted ECVs from Table 1 as are practicable. In the longer-term efforts should be encouraged to develop and roll-out instrumentation to measure the full suite of relevant ECVs at each site, including terrestrial and atmospheric composition ECVs at least for a subset of the sites. However, additional effort will be needed to reconcile ECVs that are spatially integrated over large areas of the land (sub)surface, such as river discharge, with point-based observations and joint links with current domain-specific initiatives will be an imperative (e.g., reference hydrologic networks [RHNS]) (Whitfield et al., 2012).

Fiducial reference stations should utilize high-quality instruments that enable traceability and comparability (section 3). However, instrumentation need not be identical across the network. Indeed, requiring identical instrumentation across the network may reduce resilience to potential disruptive changes in instrumentation supply as well as reduce competition among manufacturers to provide improvements in instrumentation. Therefore, the focus should not be on the actual instrumentation, but on the set

of observational requirements that the instrumentation should meet.

Any fiducial reference network should have operating procedures for changing instrumentation as technology advances. Without incorporating “upgradeability” into the design, the network could eventually find itself in a technological “dead end,” with equipment that is hard to maintain and not state-of-the-art. In line with the GCOS climate monitoring principles (World Meteorological Organization, 2017), the value of the long-term data record can be preserved by minimizing changes and using a robust program of change management. This will include substantive periods of overlapping observations to quantify the uncertainty in the change and thus ensure long-term time series comparability across all scales, from the individual measurement to multi-decadal trend, and a program of continual evaluation and comparison of new and evolving technologies (in conjunction with other relevant bodies) as is performed with the USCRN program.

4.2 | Siting considerations

Each site will need to be large enough to house all instrumentation without adjacent instrumentation interfering with one another, with no shading or wind-blocking vegetation or localized topography, and at least 100 m from any artificial heat sources. Figure 2 provides a site schematic for USCRN stations that meets this goal. The siting should strive to adhere to Class 1 criteria detailed in guidance from the WMO Commission for Instruments and Methods of Observations (World Meteorological Organization, 2014, part I, chap. I). This serves to minimize representativity errors and associated uncertainties. Sites should be chosen in areas where changes in siting quality and land use, which may impact representativity, are least likely for the next century. The site and surrounding area should further be selected on the basis that its ownership is secure. Thus, site selection requires an excellent working and local knowledge of items such as land/site ownership proposed, geology, regional vegetation, and climate. As it cannot be guaranteed that siting shall remain secure over decades or centuries, sites need to be chosen so that a loss will not critically affect the data products derived from the network. A partial solution would be to replace lost stations with new stations with a period of overlap of several years (Diamond et al., 2013). It should be stressed that sites in the fiducial reference network do not have to be new sites and, indeed, there are significant benefits from enhancing the current measurement program at existing sites. Firstly, co-location with sites already undertaking fiducial reference measurements either for target ECVs or other ECVs, such as GRUAN or GCW would be desirable. Secondly, co-location with existing baseline sites that already have long records of several target ECVs has obvious climate monitoring, cost and operational benefits.

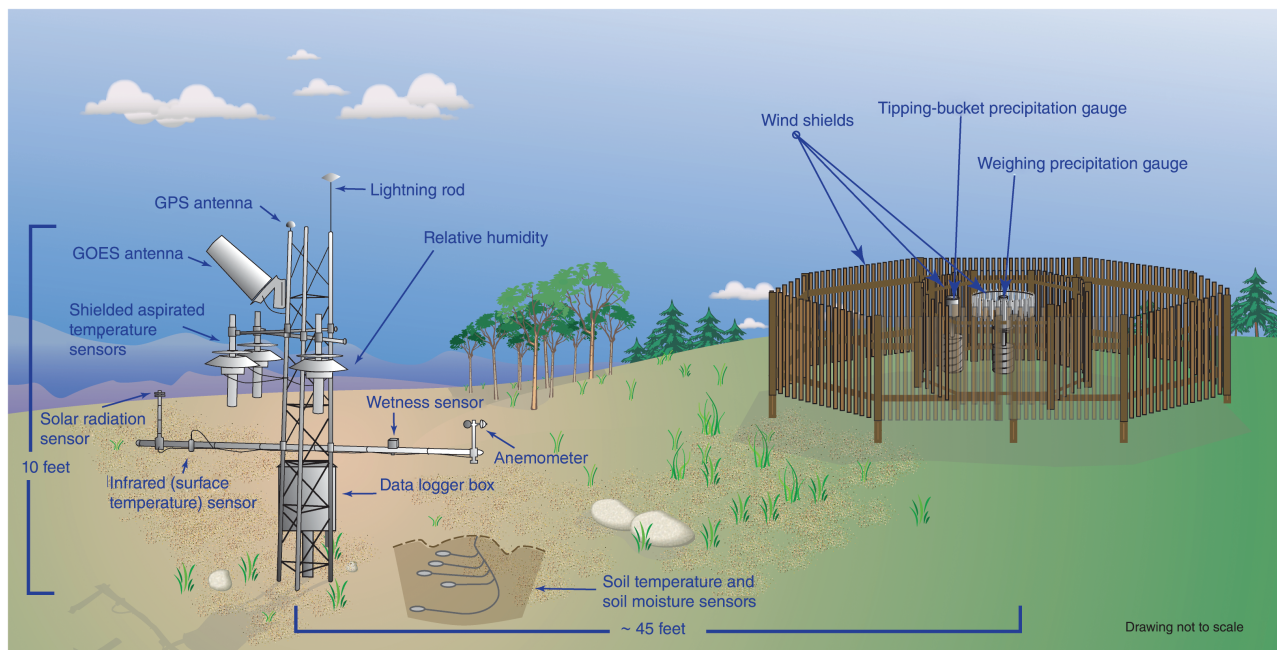


FIGURE 2 Schematic of the instrumentation at a typical USCRN station in the CONUS. The triplicate configuration of temperature sensors is repeated in the three precipitation gauge weighing mechanisms and in the three sets of soil probes located around each tower (taken from Diamond et al., 2013)

Siting considerations should be made with accessibility in mind both to better ensure uninterrupted operations and communications, and to enable both regular and unscheduled maintenance/calibration operations. If a power supply and/or wired telecommunication system is required then the site will need to provide an uninterrupted supply, and have additional redundancy in the form of a back-up generator or batteries. For many USCRN sites the power is locally generated via the use of a combination of solar, wind, and/or methane generator sources, and the GOES satellite data collection system provides one-way communication from all sites.

For a reference grade installation, an evaluated uncertainty value should be ascertained for representativeness effects which may differ synoptically and seasonally. Techniques and large-scale experiments for this kind of evaluation and characterization of the influences of the siting on the measured atmospheric parameters are currently in progress (Merlone et al., 2015).

Finally, if the global surface fiducial reference network ends up consisting of two or more distinct set-ups of instrumentation (section 4.1), there would be value in side-by-side operations of the different configurations in a subset of climatically distinct regions to ensure long-term comparability is assured (section 3). This could be a task for the identified *super-sites* in the network.

4.3 | Data and metadata reporting requirements

Data collected at fiducial reference stations should ideally be transmitted and exchanged in near-real-time. This would permit use in real-time applications, but more pertinently,

enable pro-active quality control/quality assurance and issue tracking and resolution to be enacted and enforced. This is a key facet of high-quality networks such as USCRN, and the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) program network that contributes to GRUAN.

The data transmitted must be at the basic instrument measurement frequency. Additionally, time-averaged series over, for example, sub-hourly, hourly, daily, or monthly periods should be transmitted. Data should be archived at the native data reporting resolution and made freely available and accessible in order to enable subsequent analysis. If data processing occurs at the measurement station, then in addition to the processed data the original measured series (i.e., a digital count) should be transmitted. This permits subsequent reprocessing of the entire record from the fundamental measurement data if required. For measurements that require substantive post-processing, the GRUAN model of a centralized data collection and processing facility that collects the fundamental measurement series (including all relevant metadata) and applies a consistent set of processing would be advisable. This structure ensures consistency and comparability in the resulting data products (Bodeker et al., 2016), and additionally guarantees that the original measured data are retained. However, implementation of interoperability among distributed data archives/centres through existing portals allows metadata and data to be easily exchanged and offers an effective alternative to a centralized facility and still can provide consistent data processing.

Comprehensive metadata are key to making a surface climate fiducial reference network a success. Such metadata are required to be collected and curated to enable

unambiguous subsequent use by the research and operations communities. This includes discovery metadata (sometimes termed collection level metadata), file level metadata, and documentation metadata. Metadata collection should follow appropriate agreed terminology such as the emerging WIGOS metadata standard (WIGOS, 2015).

All fiducial reference stations should meet minimum criteria for discovery metadata including:

- Longitude and latitude as decimal degrees (with at least three decimal places).
- Elevation to at least 1 m precision.
- Station name and unique network identifier.
- All known additional identification codes.
- Site photographs taken at a minimum annually and preferably seasonally to track environmental changes around the site.
- Site and locale description using a standard template.
- Instrument types description.
- Instrument screen description.

File level metadata includes aspects such as:

- Instrument.
- Location.
- Time of observation.
- Calibration applied.

Such metadata should be exchanged via BUFR file formats for real-time applications.

Documentation metadata may include items such as:

- Instrument manuals.
- Documentation of methods of observation and data transmission.
- Details of quality assurance and quality control procedures applied.
- Calibration records.
- Full documentation of any changes and evaluation thereof.

These metadata should also be stored and archived indefinitely.

5 | WHAT WOULD A GLOBAL LAND SURFACE CLIMATE FIDUCIAL REFERENCE NETWORK LOOK LIKE GEOGRAPHICALLY?

The optimum size and composition of a network are inescapably intertwined with its purpose as discussed in section 2. It is therefore necessary first to define the specific purpose(s) of the proposed fiducial reference network. We would suggest that a global surface climate fiducial

reference network must serve at least two principal purposes. First, it should provide a high-quality, stable and independent estimate of hemispheric and global-scale changes in air temperature to ascertain the effectiveness of internationally agreed mitigation measures. Second, it should enable understanding of regional level observations arising from the relevant baseline and comprehensive networks.

5.1 | Monitoring global warming

Following COP-21, signatories to the UNFCCC unanimously committed to avoiding ‘dangerous’ climate change, defined as avoiding breaching certain global mean surface temperature thresholds relative to a “pre-industrial” baseline (Hawkins et al., 2017). A fiducial reference network would enable monitoring of the effectiveness of agreed mitigation measures in future. Of course, it cannot inform us on the c. 70% of the surface domain covered by oceans (for which a similar network may be possible but is outside the scope of the current article), but it can inform us about the land response to anthropogenic climate forcing, if sufficiently globally representative.

Air temperature series at monthly to annual scales have the longest spatial correlation lengths of all atmospheric surface ECVs (Peterson, Daan, & Jones, 1997), and Jones (1995) showed that a well-spaced network of 170 representative sites could be used to estimate the global mean LSAT series on monthly to annual timescales with reasonable fidelity. This analysis is repeated and updated here using CRUTEM4.5 (Climatic Research Unit Temperature data set version 4.5) in Figure 3 using five unique subsets of 163 well-separated long-term stations. The spatial correlation length of annual average temperature is about 2,000 km, corresponding to about 85 evenly spaced stations (Briffa & Jones, 1993). In practice, achieving evenly spaced stations would be impractical, and thus c. 160 stations that are free of inhomogeneities would provide a sufficient sample for the annual means.

Similarly, an analysis showed that approximately 135 evenly spaced stations would be needed to characterize annual air temperature and precipitation trends on a national scale across the conterminous United States (Vose, 2005; Vose & Menne, 2004). Further refinement of this work using actual station locations reduced the requirement to the 114 stations finally deployed in the USCRN network in the conterminous United States, while an additional 29 stations are being deployed to characterize climate change across the state of Alaska, (Diamond et al., 2013) (Figure 4).

Typically, higher-frequency timescales (e.g., daily and sub-daily) would require many more stations. Also, the number of stations required for a reference network is driven by observation variables exhibiting lower spatial autocorrelation and/or more temporal variability such as precipitation and related snow variables (e.g., snowfall and

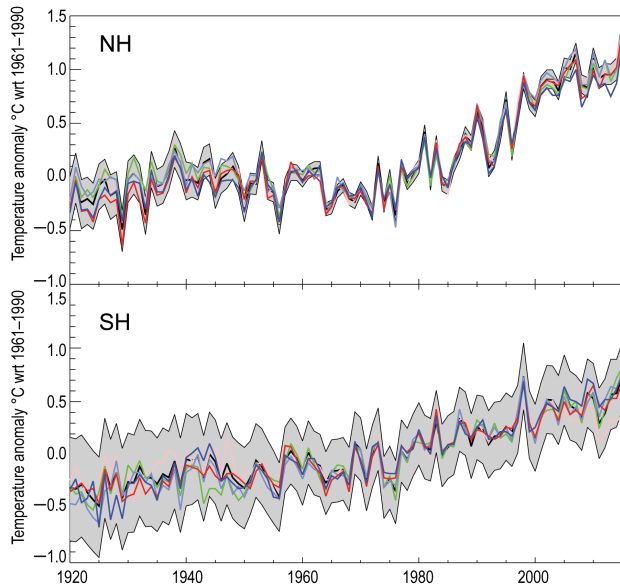


FIGURE 3 Hemispheric time series of the full network of CRUTEM4.5 and 5 unique well-spaced subsets of long-term source station records since 1920 (data prior to 1920 becomes sufficiently sparse that distinctions arising from station drop-out that would not pertain to a stable reference network become important). The thick black line shows CRUTEM 4.5 annual anomalies taken from the Hadobs website, along with the 95% confidence limits from the same source (shaded grey). The coloured lines show the smoothed annual gridded-anomalies from the respective hemispherical component of five unique 163-member subsets of CRUTEM 4.5 well spread over the globe

snow depth) (Vose & Menne, 2004). In regions of complex topography consideration of sampling different altitudes would be important for aspects such as water supply.

Further refinement of station site selection can be achieved through consideration of the impacts of modes of

variability such as El Niño-Southern Oscillation (ENSO). Placing measuring stations near the nodes of the modes of variability will have greatest explanatory power over the broadest possible regions (Kreher, Bodeker, & Sigmund, 2015), but at the same time make the network particularly susceptible to the loss of such stations. It is possible to run observing system simulation experiments (OSSEs) to explicitly design an optimal set of observing locations (Kreher et al., 2015). However, such design would also need to account for possible future large-scale changes such as those related to changing seasonal coverage of sea-ice and snow and related feedbacks.

5.2 | Constraining regional observing networks

In a tiered networks concept (Figure 1) (Thorne et al., 2017) the fiducial reference network measurements provide a potential means through which to characterize and make sense of the remainder of the observing system for the full range of target ECVs (section 4.1). They therefore must be sufficiently regionally representative and located so as to provide a meaningful cross-validation to as many other nearby sites as possible. It may be necessary to consider, for example, sampling a representative range of surface types within a given region as was performed for the GSN when choosing mountainous region sites (Peterson et al., 1997). This is particularly important for those potential applications that require a consideration of absolute rather than anomaly values or which require data to characterize satellite performance. This requirement, coupled with the need for redundancy to improve the resilience of the network against individual site losses, may increase, slightly, the required

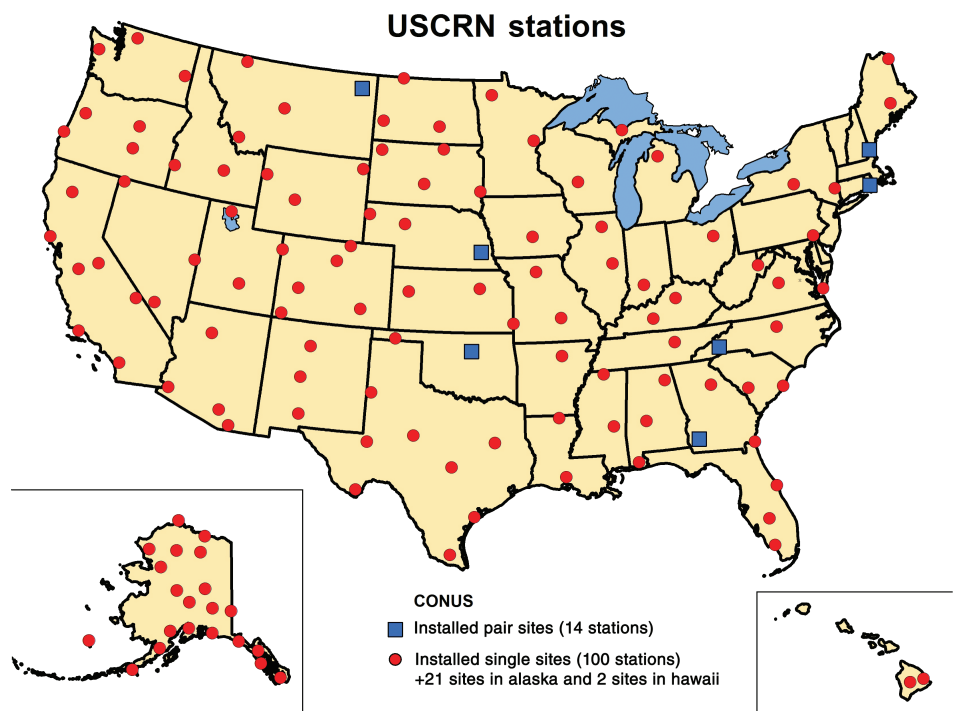


FIGURE 4 Locations of the paired and single site USCRN stations in the conterminous United States, Alaska, and Hawaii

station count, and shall certainly affect the considerations of where to place individual sites.

6 | HOW WOULD A CLIMATE SURFACE FIDUCIAL REFERENCE NETWORK BE INSTIGATED, MANAGED, AND COORDINATED?

Stations contributing to a global surface fiducial reference network would first and foremost be hosted, financially supported and practically implemented at the national and/or regional level, while leaving room for mechanisms such as pairing to support installations in developing nations. As an example of the pairing concept, Meteoswiss has an active twinning with the Kenyan Meteorological Department which assures high-quality radiosonde measurements from Nairobi. Similar partnerships could be adopted for the surface climate reference network. However support is realized, it would undoubtedly require global governance and coordination if the network were to be effective. Such global oversight would help ensure comparability and interoperability of the national and regional contributions and seamless user access to the data. The global governance would have to strongly and appropriately recognize the underlying national and regional contributions to the network. Global governance will also require comprehensive recognition and participation by many national entities. This points to the need for recognition and coordination under one or more appropriate programmatic sponsors constituting a recognized and respected international program.

Any network, if it is to be used as a network, requires a degree of standardization to ensure the data can be drawn seamlessly from across the network. Where the network is supported and managed by a single national entity, as is the case with USCRN, the management and standardization can be rigorously enforced. In the absence of a supra-national management, oversight, and funding mechanism (which seems unlikely on practical and political grounds) a confederated approach as adopted in GRUAN or Global Cryosphere Watch would appear more tractable. Key facets which would be required to ensure success of a surface reference network are:

1. A scientific working group providing leadership and oversight in implementation that answers to the sponsoring program(s). This working group should have the correct mix of instrument experts, users, scientific experts in relevant multidisciplinary fields, etc. to provide varied perspectives and long-term oversight of network operations.
2. A dedicated (set of) coordination/monitoring facility(s) charged with overseeing the day-to-day operations of the network, coordinating network activities and providing regular reports to the working group. This must be

adequately resourced to enable pro-active management of the network including scheduled and unscheduled maintenance, exception reporting and resolution, and ensuring innovations are adopted seamlessly.

3. Identification of an initial selection of contributors of sites to the network willing to help develop and test implementation of the protocols and practices, and to build the necessary data protocols and data exchange structures.
4. Sites should undergo a rigorous assessment process to ensure that the network is sufficiently similar to ensure comparability between sites.
5. Data and metadata streams should be verified as being reference quality, which implies a high level of metrological understanding, which is well documented via both the peer-reviewed literature and instrument technical documentation.
6. Centralized or coordinated processing of data streams should serve data in a consistent format through a dedicated (set of) portal(s) to enable ease of use.
7. Active quality control should pro-actively identify data issues as they arise. This should be accompanied by a resolution system that quickly fixes technical and instrumentation issues as they arise. A target availability would be, for example, >99%, that is, fewer than 10 lost days every 3 years. This may have to be relaxed for stations in more challenging and inaccessible environments, such as high latitudes and high altitudes.

Specific governance and management protocols would need to be developed and adopted. Experience of existing similar networks suggests that this is an iterative and ongoing process best facilitated by annual meetings of stakeholders in the network (at least during development and initial implementation) including users, managers and observers, supplemented by more frequent remote-participation meetings and discussions.

7 | NETWORK COSTS

As detailed in preceding sections the fiducial reference network would consist of a set of sites, facilitated on a day-to-day basis by a dedicated coordination mechanism overseeing data flow and pro-actively assessing data quality, and guided by an appropriate international oversight body. Considering each of these facets in turn allows us to make a very approximate cost estimate.

The USCRN network provides an initial rough estimate of the installation and routine maintenance costs of surface sites similar to those envisaged. A typical figure would be \$50K per station in an easy location that is not climatically challenging and easily accessible, while for more remote or harsh locations (e.g., Arctic, high-elevation, inaccessible, etc.) costs may be closer to or even exceed \$100K per

station. Keeping remote and cold-climate stations operational comes with substantial logistical and technical challenges, yet it is critical to monitor such areas. Here much can be learnt from the GCW activities which have been grappling with these challenges for decades. These costs include the initial installation, and the first decade of recurring costs of maintenance, communications, data access, archival, quality control (QC), etc. Obviously, the site costs will depend upon the range of ECVs to be monitored and the measurement techniques involved. A key lesson arising from USCRN is that whatever it costs to initially install instrument hardware, it is necessary to budget at least that much again to maintain the operational status from that point on.

Given the size and the international nature of the network and experiences from GRUAN, GCW, and USCRN, we would contend *at least* 10 full time equivalent (FTE) staff would be required to fulfil posited roles of network-wide coordination, archival, quality assurance, etc. to an adequate degree. The more stations and ECVs in the network, the greater the staffing resource required to carry out network coordination and quality assurance functions.

The governance and oversight should be an international committee activity. Such activities have typically hidden costs either in de facto time given by employers to enable participation or in other tasks not taken on. The only direct cost associated with this group is facilitation of annual meetings with the coordination centre(s) and representatives from sites. Assuming a degree of self-funding of attendance by sites and coordination centre(s) somewhere of the order \$50K per annum should facilitate attendance by all those who would require support (including site representatives from developing regions).

Overall, therefore, the order of magnitude cost of instigation and ongoing management of network will likely be somewhere in the low tens of millions of U.S. dollars per

decade. For context, this is considerably less than a single satellite mission with an expected lifetime of 5–10 years, which measures far fewer ECVs, and lacks the metrological traceability that would be possible in a global land surface fiducial reference network.

8 | WHAT SCIENTIFIC AND SOCIETAL BENEFITS WOULD ACCRUE?

If the climate community, and the wider community they represent, is to know whether international targets are met regarding limiting human induced climatic change it is necessary to monitor the global climate system sufficiently well to determine when any threshold is breached. If “dangerous” climate change is to be avoided (where dangerous is defined as some net change from a baseline state), it is necessary to quantify in a rigorous manner the change since the baseline period. Beyond recognized uncertainties in defining the pre-industrial baseline (Hawkins et al., 2017), a fiducial reference network would take measurements in such a way that there would be high confidence in stating when certain thresholds of change since network instigation are exceeded.

Indeed, the primary motive for the establishment of the USCRN was to answer questions about how much climate has changed over the United States (Diamond et al., 2013) with a high degree of certainty. Already after 10 years, the USCRN has been used to validate United States annual surface air temperature anomalies as determined by homogenized standard network observations (Hausfather, Cowtan, Menne, & Williams Jr., 2016; Menne et al., 2010). However, as Figure 5 shows, in 2015 and 2016, small differences have been observed. The origin of these differences is a topic of current research, but one could not possibly know that such potential effects existed without a reference network.

For a fiducial reference network that measures multiple ECVs (Table 1), it will be possible to determine trends for each ECV at the site and to determine trends around each site commensurate with the correlation scales of each ECV and the representativeness of the site with its regional surroundings. These high-quality observations (in addition to temperature) will likely become increasingly politically important because of agreements on burden sharing for climate impacts and adaptation. In providing well-characterized, homogeneous series, data from fiducial reference networks will aid future comparisons between climate model outputs and observations, enhancing our ability to detect and attribute emerging signals.

However, the value of a fiducial reference site is not limited solely to the ability to determine and understand long-term trend behaviour. Indeed, if this were the sole purpose of the network it would have to be in operation for multiple decades before it gives a return on investment.

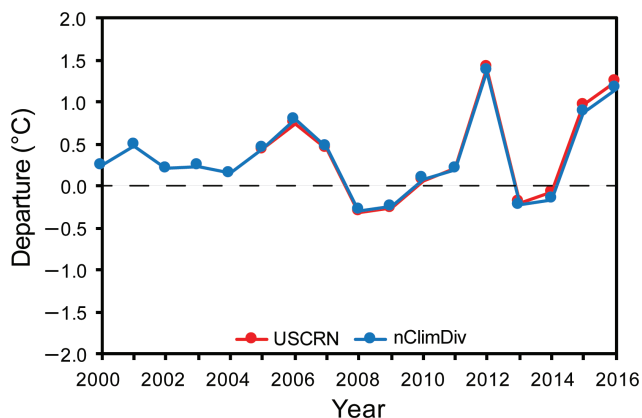


FIGURE 5 Comparison of the temperature anomaly in air temperature in the lower 48 states of the United States of America determined by analysis of the homogenized “baseline” network consisting of several thousand stations and the unhomogenized USCRN with 132 stations. The agreement is generally excellent, but small disagreements are visible in some years, most notably in 2015 and 2016

Given the relative expense of maintaining metrologically well-characterized instruments and measurement procedures in stable and regionally representative locations, it is imperative that nearer-term scientific, technological, and societal returns on investment accrue. We cannot wait decades to realize the benefits. Fortunately, there are many potential applications for, and benefits of, fiducial reference quality observations which can be achieved far more immediately. Table 2 provides a summary, based upon citations, of current usage of the USCRN and GRUAN networks over recent years showing a broad range of applications of these data to address numerous research questions.

In the shorter-term a fiducial reference network can help confirm our understanding of instrumentation from other networks. Arguably, the existing comprehensive network requires such calibration in order for it to be used robustly to study at the level of detail relevant to people, that is, at the hourly individual station level. For example, Otkin, Anderson, Mecikalski, and Diak (2005) used hourly global solar radiation observations from USCRN stations to validate GOES surface insolation estimates used in hydrologic modelling. Work under the Horizon 2020 GAIA-CLIM (gap analysis and impacts assessment for climate) project is investigating how data assimilation techniques can spread the information from reference sites to broader geographical inferences (Noh, Sohn, Kim, Joo, & Bell, 2016). Hence, a fiducial reference network would be of high value to reanalyses developers by reducing vulnerability to observational biases and improving long-term homogeneity, and given the data will be available in real-time will be welcomed too by operational forecasting. Fiducial reference network observations could also be used to robustly characterize satellite observations, as is already being carried out for upper-air ECVs by GAIA-CLIM using GRUAN and similar measurements. Furthermore, a fiducial reference network should be a useful validation tool for both large-scale and downscaled climate model reconstructions, ultimately enabling advances in model development. In the short-term it can help in validating diurnal, seasonal and process scales, longer-term it can help validate climate-timescale processes and trends.

TABLE 2 Summary of citations by primary application area based upon title and abstract for USCRN (Diamond et al., 2013) and GRUAN (Bodeker et al., 2016; Seidel et al., 2009) taken from the citations listed by the Bulletin of the American Meteorological Society as at October 14, 2017. Numerous applications already have benefitted from these networks despite only now being a decade or so old

Analysis type	USCRN	GRUAN
Process understanding	10	1
Trend detection	7	13
Data analysis	12	5
Instrument development	6	9
Satellite cal/val	11	14
NWP/real-time	6	1
Applications	9	4

A fiducial reference network would enable us to improve our understanding of fundamental climate processes through observing multiple ECVs to high quality and with high temporal resolution. In addition, with high-quality observations on a continuous basis, reference sites would constitute desirable locations to base future field campaigns as they provide a pre-existing capability and a longer-term context in which to interpret the results including how climatologically representative the period of the campaign was.

Because fiducial reference observations are striving for improved metrological characterization and understanding there are potentials for trickle-down to other networks of either improved instrumentation or improved practices. For example, work within GRUAN to characterize radiosondes has led to changes in protocols and processing, which has served to improve the Modem M10 sonde throughout the baseline and comprehensive radiosonde sounding network.

Finally, fiducial reference quality observations can also improve our ability to interpret historical observational records. For example, an important advantage of automated measurements at USCRN has been to accurately record precipitation and other variables in specific time intervals, avoiding issues of time-of-observation inconsistencies in manual observations (Leeper, Rennie, & Palecki, 2015).

9 | NEXT STEPS

While it is beyond question that the climate system has changed since instrumental records were instigated, we can improve our collective ability to characterize these changes through instigating and maintaining a global surface fiducial reference network. If such a network is to become a reality then it needs to be formalized and adopted by a relevant sponsor(s) and then accepted by those national agencies likely to contribute to it. Preceding sections have provided an outline upon which much specific detail must be built if a network is to be adopted and then to become successful. This progress will only be assured with strong backing from the World Meteorological Organization and its members. This article was requested by the Global Climate Observing System and the Commission for Climatology to outline a vision and provide a coherent perspective on what would be required and how it could be achieved. It is now up to these and other relevant parties to decide how to take this forward. Logically this may be achieved via some variant on the following steps:

1. Development of more detailed concept and formal guidance materials.
2. Agreeing to a governance and oversight structure and ensuring appropriate resourcing including staffing of a lead centre(s).

3. Recruitment of national and regional contributions to the network.

Furthermore, all aspects of a successful network would require strong and sustained stakeholder engagement. This article not only serves as a means to collate and peer-review a vision, but also to begin the long process of creating a network that has strong commitment from data owners and data users.

There are many possible metrics for determining the success of a global land surface fiducial reference climate network as it evolves, such as the number and distribution of fiducial reference climate stations or the percent of stations adhering to the strict reference climate criteria described in this article. However, in order to fully appreciate the significance of the proposed global climate surface fiducial reference network, we need to imagine ourselves in the position of scientists working in the latter part of the 21st century and beyond. However, not just scientists, but also politicians, civil servants, and citizens faced with potentially difficult choices in the face of a variable and changing climate. In this context, we need to act now with a view to fulfilling their requirements for having a solid historical context they can utilize to assist them making scientifically vetted decisions related to actions on climate adaptation. Therefore, we should care about this now because those future scientists, politicians, civil servants, and citizens will be—collectively—our children and grandchildren, and it is—to the best of our ability—our obligation to pass on to them the possibility to make decisions with the best possible data. Having left a legacy of a changing climate, this is the very least successive generations can expect from us in order to enable them to more precisely determine how the climate has changed.

ACKNOWLEDGEMENTS

We thank John Eyre for comments on an early draft. The draft benefitted from discussions by the GCOS Atmospheric and Terrestrial Observation Panels for Climate. Craig Donlon and participants at an ESA-sponsored meeting on fiducial reference temperature measurements provided insightful feedback. Two anonymous reviewers provided constructive review comments. We also thank NOAA NCEI's graphics team for support in providing the final production versions of the graphics. Part of this work is being developed within the frame of the European Metrology Research Program (EMRP) joint research project ENV58 "MeteoMet2." The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. K.M.W. was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). V.V. was supported by the German Science Foundation (grant no. VE366/8). P.W.T. was supported by GAIA-CLIM under

the European Union's Horizon 2020 research and innovation programme under grant agreement no. 640276.

ORCID

P. W. Thorne  <http://orcid.org/0000-0003-0485-9798>

S. Harrigan  <http://orcid.org/0000-0002-0992-3667>

Z. Hausfather  <http://orcid.org/0000-0002-2926-0581>

REFERENCES

- Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., & Wieringa, J. (2003). *Guidelines on climate metadata and homogenization* (WMO-TD No. 1186, WCDMP No. 53). Geneva, Switzerland: World Meteorological Organization.
- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., & Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth System Science Data*, 5, 71–99. <https://doi.org/10.5194/essd-5-71-2013>
- Bell, S. A., Carroll, P. A., Beardmore, S. L., England, C., & Mander, N. (2017). A methodology for study of in-service drift of meteorological humidity sensors. *Metrologia*, 54(3), S63. <https://doi.org/10.1088/1681-7575/aa6dd0>
- Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., ... Wang, J. (2016). Reference upper-air observations for climate: From concept to reality. *Bulletin of the American Meteorological Society*, 97, 123–135. <https://doi.org/10.1175/BAMS-D-14-00072.1>
- Böhm, R., Jones, P. D., Hiebl, J., Frank, D., Brunetti, M., & Maugeri, M. (2010). The early instrumental warm-bias: A solution for long central European temperature series, 1760–2007. *Climate Change*, 101, 41–67.
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, 95, 1431–1443. <https://doi.org/10.1175/BAMS-D-13-00047.1>
- Brandtsma, T., & van der Meulen, J. P. (2008). Thermometer screen intercomparison in De Bilt (the Netherlands)—Part II: Description and modeling of mean temperature differences and extremes. *International Journal of Climatology*, 1, 389–400.
- Briffa, K. R., & Jones, P. D. (1993). Global surface air temperature variations during the twentieth century: Part 2: Implications for large-scale high-frequency palaeoclimatic studies. *The Holocene*, 3, 77–88. <https://doi.org/10.1177/095968369300300109>
- Changnon, S. A., & Kunkel, K. E. (2006). Changes in instruments and sites affecting historical weather records: A case study. *Journal of Atmospheric and Oceanic Technology*, 23, 825–828.
- Chimani B, Venema VKC, Lexer A, Andre K, Auer I, Nemeč J. (2017). Inter-comparison of methods to homogenise daily relative humidity. Manuscript submitted for publication.
- Diamond, H., Karl, T., Palecki, M., Baker, C., Bell, J., Leeper, R., ... Thorne, P. (2013). U.S. climate reference network after one decade of operations: Status and assessment. *Bulletin of the American Meteorological Society*, 94, 485–498. <https://doi.org/10.1175/BAMS-D-12-00170.1>
- Dirksen, R. J., Sommer, M., Immeler, F. J., Hurst, D. F., Kivi, R., & Vömel, H. (2014). Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmospheric Measurement Techniques*, 7, 4463–4490. <https://doi.org/10.5194/amt-7-4463-2014>
- Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J., & Pielke Sr., R. A. (2011). Analysis of the impacts of station exposure on the U.S. historical climatology network temperatures and temperature trends. *Journal of Geophysical Research*, 116, D14120.
- GCOS. (2010). *Guide to the GCOS surface network (GSN) and GCOS upper-air network (GUAN): 2010 update of GCOS-73* (WMO/TD No. 1558, GCOS No. 144). Geneva, Switzerland: World Meteorological Organization. Retrieved from https://library.wmo.int/opac/index.php?lvl=notice_display&id=12885
- GCOS. (2015). *Status of the global observing system for climate* (GCOS No. 195). Geneva, Switzerland: World Meteorological Organization.

- Retrieved from https://library.wmo.int/opac/index.php?lvl=notice_display&id=18962
- Goodison, B. E., Louie, P. Y. T., & Yang, D. (1998). *WMO solid precipitation measurement intercomparison* (WMO/TD No. 872, IOM No. 67). Geneva, Switzerland: World Meteorological Organization.
- Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., ... Zhai, P. M. (2013). Observations: Atmosphere and surface. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 159–254). Cambridge, UK and New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.008>
- Hausfather, Z., Cowtan, K., Menne, M. J., & Williams Jr., C. N. (2016). Evaluating the impact of U.S. historical climatology network homogenization using the U.S. climate reference network. *Geophysical Research Letters*, *43*, 1695–1701. <https://doi.org/10.1002/2015GL067640>
- Hausfather, Z., Menne, M. J., Williams Jr., C. N., Masters, T., Broberg, R., & Jones, J. (2013). Quantifying the effect of urbanization on U.S. historical climatology network temperature records. *Journal of Geophysical Research*, *118*(2), 481–494.
- Hawkins, E., & Jones, P. D. (2013). On increasing global temperatures: 75 years after Callendar. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1961–1963. <https://doi.org/10.1002/qj.2178>
- Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., ... van Oldenborgh, G. J. (2017). Estimating changes in global temperature since the pre-industrial period. *Bulletin of the American Meteorological Society*, *98*, 1841–1856. <https://doi.org/10.1175/BAMS-D-16-0007.1>
- Ingleby, B., Moore, D., Sloan, C., & Dunn, R. (2013). Evolution and accuracy of surface humidity reports. *Journal of Atmospheric and Oceanic Technology*, *30*, 2025–2043. <https://doi.org/10.1175/JTECH-D-12-00232.1>
- Intergovernmental Panel on Climate Change (2013). In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change* (p. 1535). Cambridge, UK and New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324>
- Intergovernmental Panel on Climate Change (2007, 2007). In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, et al. (Eds.), *Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY: Cambridge University Press.
- JCGM. (2012). *200:2012 International vocabulary of metrology—Basic and general concepts and associated terms (VIM)* (3rd ed.). Author. Retrieved from <https://www.bipm.org/en/publications/guides/vim.html>
- Jones, P. (2016). The reliability of global and hemispheric surface temperature records. *Advances in Atmospheric Sciences*, *33*, 269–282. <https://doi.org/10.1007/s00376-015-5194-4>
- Jones, P. D. (1995). Land surface temperatures—Is the network good enough? *Climate Change*, *31*, 545–558.
- Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*, D05127. <https://doi.org/10.1029/2011JD017139>
- Karl, T. R., Williams Jr., C. N., Young, P. J., & Wendland, W. M. (1986). A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *Journal of Applied Meteorology and Climatology*, *25*, 145–160.
- Killick, R. E. (2016). *Benchmarking the performance of homogenisation algorithms on daily temperature data*. (Unpublished doctoral dissertation). University of Exeter, Exeter, UK. Retrieved from <http://hdl.handle.net/10871/23095>
- Kreher, K., Bodeker, G. E., & Sigmond, M. (2015). An objective determination of optimal site locations for detecting expected trends in upper-air temperature and total column ozone. *Atmospheric Chemistry and Physics*, *15*, 7653–7665. <https://doi.org/10.5194/acp-15-7653-2015>
- Leeper, R. D., Rennie, J., & Palecki, M. A. (2015). Observational perspectives from U.S. climate reference network (USCRN) and cooperative observer program (COOP) network: Temperature and precipitation. *Journal of Atmospheric and Oceanic Technology*, *94*, 703–721. <https://doi.org/10.1175/JTECH-D-14-00172.1>
- Lindau, R., & Venema, V. (2016). The uncertainty of break positions detected by homogenization algorithms in climate records. *International Journal of Climatology*, *36*(2), 576–589. <https://doi.org/10.1002/joc.4366>
- Mekis, E., & Vincent, L. A. (2011). An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean*, *49*, 163–177. <https://doi.org/10.1080/07055900.2011.583910>
- Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, *29*, 897–910. <https://doi.org/10.1175/JTECH-D-11-00103.1>
- Menne, M. J., & Williams, C. N. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, *22*, 1700–1717.
- Menne, M. J., Williams Jr., C. N., & Palecki, M. A. (2010). On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research*, *115*, D11108. <https://doi.org/10.1029/2009JD013094>
- Merlone, A., Lopardo, G., Sanna, F., Bell, S., Benyon, R., Bergerud, R. A., ... Underwood, R. (2015). The MeteoMet project—Metrology for meteorology: Challenges and results. *Meteorological Applications*, *22*(1), 820–829. <https://doi.org/10.1002/met.1528>
- Milewska, E. J., & Vincent, L. A. (2016). Preserving continuity of long-term daily maximum and minimum temperature observations with automation of reference climate stations using overlapping data and meteorological conditions. *Atmosphere-Ocean*, *54*(1), 32–47.
- Noh, Y. C., Sohn, B. J., Kim, Y., Joo, S., & Bell, W. (2016). Evaluation of temperature and humidity profiles of unified model and ECMWF analyses using GRUAN radiosonde observations. *Atmosphere*, *7*, 94.
- Otkin, J., Anderson, M. C., Mecikalski, J. R., & Diak, G. R. (2005). Validation of GOES-based insolation estimates using data from the U.S. climate reference network. *Journal of Hydrometeorology*, *6*, 460–475.
- Parker, D. E. (1994). Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, *14*, 1–31.
- Peterson, T. C., Daan, H., & Jones, P. D. (1997). Initial selection of a GCOS surface network. *Bulletin of the American Meteorological Society*, *78*, 2145–2152.
- Quayle, R. G., Easterling, D. R., Karl, T. R., & Hughes, P. Y. (1991). Effects of recent thermometer changes in the cooperative station network. *Bulletin of the American Meteorological Society*, *72*, 1718–1723.
- Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., ... Yin, X. (2014). The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. *Geoscience Data Journal*, *1*(2), 75–102.
- Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., ... Wickham, C. (2013). A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinfor Geostat: An Overview*, *1*, 1.
- Schöner, W., Key, J., Fierz, C., Goodison, B., Godøy, Ø., Citterio, M., ... Ondráš, M. 2016. *The global cryosphere watch surface network in the Arctic and beyond*. Paper presented at the Arctic Observing Summit 2016: 3rd Biennial Arctic Observing Summit, Fairbanks, AK. Retrieved from http://www.arcticobservingsummit.org/sites/arcticobservingsummit.org/files/Schoener_REVISEDRC--GCW_whitepaper_AOS2016.pdf
- Seidel, D. J., Berger, F. H., Immler, F., Sommer, M., Vömel, H., Diamond, H. J., ... Wang, J. (2009). Reference upper-air observations for climate: Rationale, progress, and plans. *Bulletin of the American Meteorological Society*, *90*, 361–369. <https://doi.org/10.1175/2008BAMS2540.1>
- Sevruk, B., Ondráš, M., & Chvíla, B. (2009). The WMO precipitation measurement intercomparisons. *Atmospheric Research*, *92*, 376–380. <https://doi.org/10.1016/j.atmosres.2009.01.016>
- Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S., & Thépaut, J. N. (2017). A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quarterly Journal of the Royal Meteorological Society*, *143*, 101–119. <https://doi.org/10.1002/qj.2949>
- Thorne, P. W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., ... Pappalardo, G. (2017). Making better sense of the mosaic of environmental measurement networks: A system-of-systems approach and quantitative assessment. *Geoscientific Instrumentation, Methods and Data Systems*, *6*, 453–472. <https://doi.org/10.5194/gi-2017-29>
- Trewin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of Climatology*, *33*, 1510–1529.

- Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., ... Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8, 89–115. <https://doi.org/10.5194/cp-8-89-2012>
- Vose, R. S. (2005). Reference station networks for monitoring climatic change in the conterminous United States. *Journal of Climate*, 18, 5390–5395. <https://doi.org/10.1175/JCLI3600.1>
- Vose, R. S., & Menne, M. J. (2004). A method to determine station density requirements for climate observing networks. *Journal of Climate*, 17, 2961–2971. [https://doi.org/10.1175/1520-0442\(2004\)017<2961:AMTDS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2961:AMTDS>2.0.CO;2)
- Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T., & Looser, U. (2012). Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal*, 57(8), 1562–1579. <https://doi.org/10.1080/02626667.2012.728706>
- WIGOS. (2015). *Manual on the WMO Integrated Global Observing System*. Geneva, Switzerland: World Meteorological Organization. Retrieved from http://www.wmo.int/pages/prog/www/wigos/documents/WIGOS-RM/1160_en.pdf
- Willett, K. M., Dunn, R. J. H., Thorne, P. W., Bell, S., de Podesta, M., Jones, P. D., ... Williams Jr., C. N. (2014). HadISDH land surface multi-variable humidity and temperature record for climate monitoring. *Climate of the Past*, 10, 1983–2006. <https://doi.org/10.5194/cp-10-1983-2014>
- Willett, K. M., Williams Jr., C. N., Dunn, R. J. H., Thorne, P. W., Bell, S., de Podesta, M., ... Parker, D. E. (2013). HadISDH: An updateable land surface specific humidity product for climate monitoring. *Climate of the Past*, 9, 657–677. <https://doi.org/10.5194/cp-9-657-2013>
- Williams, C. N., Menne, M. J., & Thorne, P. W. (2012). Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *Journal of Geophysical Research*, 117, D05116. <https://doi.org/10.1029/2011JD016761>
- World Meteorological Organization. (2014). *Guide to meteorological instruments and methods of observation*. Geneva, Switzerland: Author. Retrieved from <http://www.wmo.int/pages/prog/www/IMOP/IMOP-home.html>
- World Meteorological Organization. (2015). *Vision for the WIGOS component systems in 2040* (IPET-SUP-1/Doc. 7). Geneva, Switzerland: Author. Retrieved from http://www.wmo.int/pages/prog/sat/meetings/documents/IPET-SUP-1_Doc_07_Vision2040.pdf
- World Meteorological Organization. (2016). *Regional basic observing network*. Geneva, Switzerland: Author. Retrieved from https://www.wmo.int/pages/prog/www/.../RBON-wksp_Final-Report_2016.docx
- World Meteorological Organization. (2017). *GCOS monitoring principles*. Geneva, Switzerland: Author. Retrieved from <https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables>
- Yang, D., Kane, D., Zhang, Z., Legates, D., & Goodison, B. (2005). Bias-corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophysical Research Letters*, 32, L19501. <https://doi.org/10.1029/2005GL024057>

How to cite this article: Thorne PW, Diamond HJ, Goodison B, et al. Towards a global land surface climate fiducial reference measurements network. *Int. J. Climatol.* 2018;1–15. <https://doi.org/10.1002/joc.5458>