

Spatial Coverage of Monitoring Networks: A Climate Observing System Simulation Experiment

ELIZABETH C. WEATHERHEAD,^a GREG E. BODEKER,^b ALESSANDRO FASSÒ,^c KAI-LAN CHANG,^d
JEFFREY K. LAZO,^e C. T. M. CLACK,^f DALE F. HURST,^{d,g} BIRGIT HASSLER,^b JASON M. ENGLISH,^{a,h}
AND SONER YORGUNⁱ

^a *University of Colorado Boulder, Boulder, Colorado*

^b *Bodeker Scientific, Alexandra, New Zealand*

^c *University of Bergamo, Bergamo, Italy*

^d *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*

^e *National Center for Atmospheric Research,^j Boulder, Colorado*

^f *Vibrant Clean Energy, Boulder, Colorado*

^g *NOAA/ESRL/Global Monitoring Division, Boulder, Colorado*

^h *NOAA/ESRL/Global Systems Division, Boulder, Colorado*

ⁱ *Monash University, Clayton, Victoria, Australia*


(Manuscript received 21 February 2017, in final form 12 June 2017)

ABSTRACT

Observing systems consisting of a finite number of in situ monitoring stations can provide high-quality measurements with the ability to quality assure both the instruments and the data but offer limited information over larger geographic areas. This paper quantifies the spatial coverage represented by a finite set of monitoring stations by using global data—data that are possibly of lower resolution and quality. For illustration purposes, merged satellite temperature data from Microwave Sounding Units are used to estimate the representativeness of the Global Climate Observing System Reference Upper-Air Network (GRUAN). While many metrics exist for evaluating the representativeness of a site, the ability to have highly accurate monthly averaged data is essential for both trend detection and climatology evaluation. The calculated correlations of the monthly averaged upper-troposphere satellite-derived temperatures over the GRUAN stations with all other pixels around the globe show that the current 9 certified GRUAN stations have moderate correlations ($r \geq 0.7$) for approximately 10% of the earth, but an expanded network incorporating another 15 stations would result in moderate correlations for just over 60% of the earth. This analysis indicates that the value of additional stations can be quantified by using historical, satellite, or model data and can be used to reveal critical gaps in current monitoring capabilities. Evaluating the value of potential additional stations and prioritizing their initiation can optimize networks. The expansion of networks can be evaluated in a manner that allows for optimal benefit on the basis of optimization theory and economic analyses.

1. Introduction

This paper offers one method for assessing the value of a monitoring network composed of a small set of stations. We use the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) as an example by evaluating the ability of individual stations to represent long-term variability in monthly averaged data for any other site globally. Merged satellite records from the Microwave Sounding Unit (MSU) datasets and statistical techniques

 Denotes content that is immediately available upon publication as open access.

^j The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author: E. C. (Betsy) Weatherhead, betsy.weatherhead@colorado.edu

that build from previous published efforts are used to evaluate the optimization of a valuable and robust Earth observing system. We acknowledge that monitoring networks often serve multiple purposes and that support of those efforts will require additional evaluations to fully optimize any network. In particular, observations that continue long-term, historical records are invaluable to climate studies as are all observations that support research into the dynamic interactions of climate parameters and climate impacts.

Efforts to optimize environmental monitoring networks have often focused on examining regional effects; McBratney et al. (1981) used statistical techniques to optimize sampling approaches for regional networks; Yost et al. (1982) identified “areas of influence” for monitoring networks to understand soil characteristics; Webster et al. (1989) identified cohesiveness of regions for reflected solar radiation values. Early efforts often focused on means at a particular sampling site to assure that all subregions would be appropriately monitored (e.g., Genton 1998). Indeed, some of the earliest advances in spatial statistical techniques, including variograms, were developed within the mathematical community with strong involvement and encouragement from the environmental community (Papritz and Flühler 1994; Weatherhead and Webb 1997; Wikle and Royle 1999; Kint et al. 2003; Dobbie et al. 2008). The motivation for understanding how much information could be derived from a finite monitoring network drove much of the development for now-standard techniques in spatial analysis.

From the earliest efforts, the environmental monitoring community recognized the need to optimize available resources and made use of available tools from optimization theory (e.g., Dantzig and Cottle 1963; Cressie 1985). These original efforts made extensive use of optimization algorithms with collected data and extrapolated relationships observed in datasets to identify needed additional monitoring stations. Todini and Ferraresi (1996) compared variogram techniques with kriging and variance estimates to better understand uncertainties based on incomplete sampling. Careful efforts in this area included looking at the dependence of results on the statistical behavior of the collected data, including van Groenigen (2000), who compared variogram techniques with different methods for optimizing networks. The computational complexity of testing for all possible optimization choices was recognized early (Nychka and Saltzman 1998; Royle 2002) and remains a challenge (Cressie and Wikle 2015).

Examining networks as an entity to be optimized, Nychka and Saltzman (1998) helped clarify the goals of optimizing networks when they addressed these

questions: What does it mean to measure the environment well? How many monitoring instruments are really necessary? They focused on the ability to estimate environmental parameters away from a monitoring station with a known level of certainty. In this paper, we expand on that goal, with a particular aim of capturing monthly variability, which is critical to tracking global patterns of climate variability such as El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO) as well as developing climatologies and detecting trends (Weatherhead et al. 1997; Messié and Chavez 2011). Girz et al. (2002) looked at the possibility of developing a fully adaptive observing system based on some of these principles.

Efforts for many environmental monitoring networks have turned to trend detection—both planning new networks and optimizing existing ones for the detection of trends (Weatherhead et al. 2005; MacDonald 2005). Royle and Nychka (1998) offered coverage designs to optimize distance-based metrics. Van Groenigen (2000) suggested the use of variogram patterns for optimizing sampling schemes. Royle (2002) expanded on earlier work to incorporate modifications to address some of the computationally challenging questions, showing that near-optimal results can be obtained if evaluation and optimization are intelligently planned. Weatherhead et al. (2002) outlined some of the general issues for detecting environmental change—including trends. Bodeker and Kremser (2015) identified methods for trend detection early in the planning of GRUAN and allowed network decisions to be guided by trend detection capabilities. Kreher et al. (2015) proposed an objective approach for network planning by identifying optimal locations for trend detection.

In addition to observing trends, some of the work looked for insights from network optimizing techniques. De Viron et al. (2013) looked at 25 climate indices for appropriate correlation on interannual time scales. Fischer et al. (2011) looked at the impact of time scales on trend estimates. Hunt et al. (2013) expanded on point correlation maps and worked with self-organizing maps to identify teleconnection patterns. For many of these studies, the complementary nature of understanding the global modes of variability and analyzing how networks should be planned led to mutually complementary goals of understanding the Earth system and improving network configurations.

While the introduction of satellite monitoring allows for global coverage, there remains a strong need for ground-based monitoring for both continuity and calibration. A clear strength of in situ or ground-based environmental observing systems is the ability to make measurements using regularly calibrated instruments at

selected locations (e.g., [Whiteman et al. 2011](#)). This advantage comes at the cost of having a finite number of locations with observations. The spatial representativeness of individual stations or a network of stations is addressed in this paper. One way to address spatial representativeness is to ask the question of how much of the surface of the earth is within a stated distance of a monitoring site. This approach is appropriate for some applications but ignores the often complex spatial coherence that exists in the Earth system. Another approach is to identify locations that represent different modes of variability. However, this approach might ignore important regions that are not well characterized by currently identified modes of variability. For this study, we have examined a number of metrics for evaluating individual sites and focused, as an example, on the ability of a monitoring site to reproduce monthly variability in upper-air temperature at nearby locations, using correlations as a primary way to describe the spatial representativeness. This technique proves to be appropriate because the correlation results display a smooth behavior; if an inhomogeneous situation were the case, the techniques would not be appropriate. We then expanded this approach to evaluate the network as a whole and discussed the need for redundancy and optimization techniques that can be applied.

2. Monthly means

Climate observations serve a multitude of purposes, including monitoring for long-term changes and extreme events, supporting climate services, and providing data for process studies to better understand the environment. The World Meteorological Organization (WMO) Integrated Global Observing System (WIGOS) is successfully encouraging network planning to include the service of networks to the full range of important scientific and societal goals. With this multiplicity of purposes, the requirements for any monitoring system can be carefully analyzed, and scientists may conclude that the requirements are “high spatial (e.g. every 10–100 km) resolution” and “high temporal resolution (e.g. every hour)” (e.g., [Wulfmeyer et al. 2015](#)). Such requirements can, in some circumstances, be logistically and economically prohibitive. In this paper, we address the role of finite networks in their ability to monitor variability and detect trends.

To monitor variability, we require observations that allow us to record the mean and variance of seasonal cycles and to produce some representation of diurnal cycles. To detect trends in mean values, we require monitoring efforts to be able to produce monthly averaged records that can be used to detect expected trends

with a stated level of efficiency. Monthly averaged values, and their associated uncertainties, are needed to both establish climatologies and detect trends. Little attention has been given in the literature to the appropriate way to derive monthly means, although thoughtful attention has been given to the subject within operational and research efforts. Temporal averaging is sensitive to temporal gradients of the parameter of interest that might exist throughout the chosen temporal averaging unit (e.g., one month). These variations will be smoothed with highly resolved measurements but not with measurements taken only sporadically or sparsely resolved. This challenge is similar to addressing sparse spatial sampling. This paper addresses both some of the temporal issues in creating monthly means as well as spatial issues.

Calculating unbiased monthly means

Many factors must be carefully considered when calculating monthly means at a given site; uncertainty of individual observations, local factors affecting instrumentation, and diurnal influences are among some of the factors. In this section, we address two issues that are sometimes ignored but may be relevant for data taken at less than daily resolution: the local seasonal cycle for the parameter being measured and the day-to-day autocorrelation of the data. Depending on the parameter, some of these issues may be minor and some may be highly influential on the resultant monthly average value. While for observations occurring every day at the same time, the impact of these factors is simplified, effects of these issues should still be critically evaluated.

1) FACTOR 1: SEASONAL CYCLE BIAS

A seasonal cycle can produce a bias in calculated monthly means if the measurements are not evenly distributed through the month or if efforts are not taken into account for such temporal sampling biases. This effect is strongest if the observations are not taken daily and even stronger if the observations are not taken frequently.

Consider the time series in [Fig. 1](#), which may represent temperature data from a springtime month, with the dark squares indicating when observations are taken. Even though most of the observations were above the seasonally normal level for that time of year, the simple mean of the data taken would result in a lower-than-normal monthly average because most of the data were collected in the early part of the month when seasonal temperatures are lower.

Even a regular sampling cycle, for instance, measurements every Monday, can create an unexpected bias in the monthly mean due to strong seasonal changes within a month. Measurements taken every Monday for 2013 would result in observations on 1, 8, 15, ... April

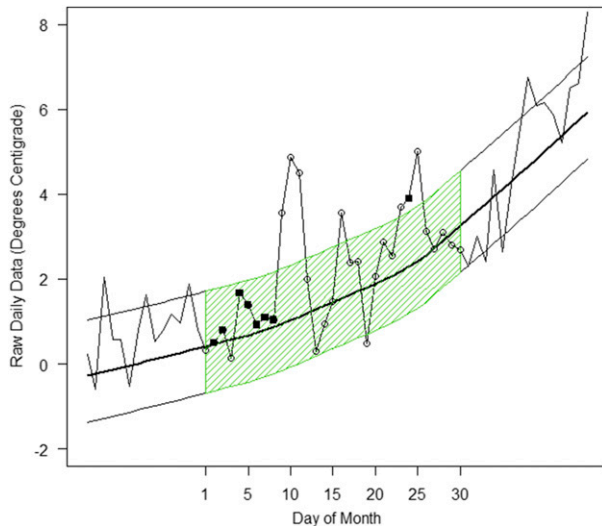


FIG. 1. Daily observations are taken frequently at the beginning of the month (black squares) when the seasonally expected values are relatively low and are taken less frequently during the end of the month. The long-term seasonal cycle is the dark solid line, and the two thin gray lines represent 1 standard deviation from historically expected levels. Without correcting for the changes due to the seasonal cycle, the calculated monthly mean would be much lower than seasonal norms even though the temperature on days when observations were made were at or above the seasonal expected level. Correcting daily data for seasonality is required to avoid this potential source of error. The green hatched area is the 1-month period being used to create a monthly average.

while the Monday measurements for 2014 would be taken on 7, 14, . . . April. The impact of this sampling program on monthly averages can be particularly important in polar regions where solar zenith angle and temperature can change notably over the period of a few days. Care with the monthly averaging of data collected at different times of the month can reduce variability, and at times influence trends, when the period of observation is short.

2) FACTOR 2: AUTOCORRELATION INFLUENCE

Irregularly sampled data are sometimes unavoidable given the availability of the instrumentation, schedules of personnel, and requirements for measurements (clear sky, etc.). The irregular sampling of data can result in clusters of observations, which may “oversample” one regime, for instance, clear-sky days, within the month. As an extreme example, consider observing the temperature nine times in a single day and then once two weeks later in the same month; computing a simple average of all ten numbers would not result in a robust estimate of the monthly mean temperature. The challenge for creating a best estimate when sporadic daily

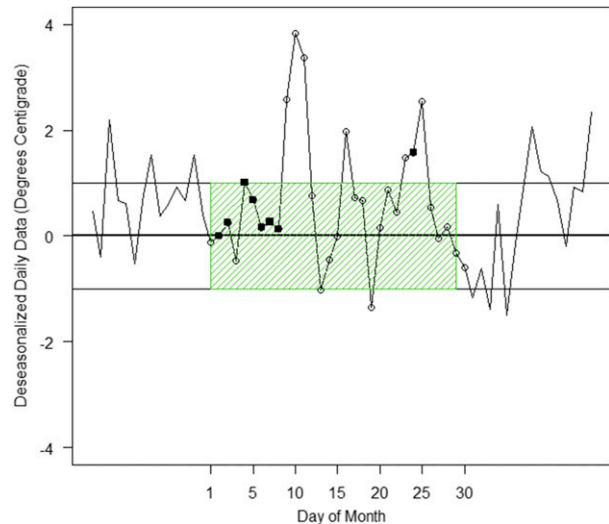


FIG. 2. Daily deseasonalized data for a single month (observations are the black squares). Oversampling in the early part of the month where the data are close to seasonal norms give the impression of a quiet, “normal” month even though there were three periods of fairly high values. The one sample taken during an unusually warm period would be outweighed by the multiple sequential observations at the beginning of the month unless a more intelligent way of averaging the autocorrelated data is used.

observations are taken implies that sequential days of data are, to some extent, redundant because of day-to-day autocorrelation.

In this study, we make the often-appropriate assumption that environmental data are autocorrelated and that the functional form of that autocorrelation is a first-order autoregressive [AR(1)] process both when daily data are considered and monthly data are considered. This AR(1) process does not imply that for daily data, an event, such as a warm set of days, has a time length of one day but rather that a single exponent can help explain the time scale of the event.

We introduce a weighted mean method that incorporates information on the day-to-day autocorrelation in the data. The motivation for this weighted mean approach is to assure that observations taken close together in time are weighted less than observations that are less clustered.

Consider the daily observations used in Fig. 1, presented in Fig. 2 as temperature anomalies, or deseasonalized daily data. The month roughly consists of three fairly brief warm periods and three cool periods. In the first part of the month, observations were taken frequently; during the last part of the month, only one observation was taken. The first part of the month coincided with a period of slightly higher-than-normal temperatures, while the middle and last part of the

month exhibited many notably higher-than-normal temperatures. If all observations are weighted equally, thus ignoring the redundancy in the day-to-day observations, the mean would be considerably lower than if the observations were weighted by their information content.

To address the issue of daily autocorrelation and potential oversampling of one period of the month, the monthly means are estimated by a weighted average of daily data, with weights based on the frequency of data collection. Consider the n deseasonalized daily observations Y_1, \dots, Y_n sampled at days t_1, \dots, t_n for a generic month. Assuming that daily data behave approximately as a first-order autoregressive process with daily autocorrelation denoted by φ , weights for each of the daily values can be computed as follows:

$$\text{Weight}_1 = \frac{1}{1 + \varphi^{t_2 - t_1}}, \quad (1)$$

$$\text{Weight}_i = \frac{1}{1 + \varphi^{t_i - t_{i-1}}} - \frac{\varphi^{t_{i+1} - t_i}}{1 + \varphi^{t_{i+1} - t_i}}, \quad \text{for } i=2, \dots, n-1, \quad (2)$$

$$\text{Weight}_n = \frac{1}{1 + \varphi^{t_n - t_{n-1}}}. \quad (3)$$

Hence, the weighted monthly mean for month m is defined by

$$y_m = \sum_{i=1}^n w_i \times Y_i, \quad (4)$$

where w_i are the normalized weights, summing to unity and given by

$$w_i = \frac{\text{Weight}_i}{\sum_{i=1}^n \text{Weight}_i}. \quad (5)$$

For cases where the observations occur daily at the same time each day and there are no missing values, in Eqs. (1)–(3), $t_i - t_{i-1} = \text{constant}$ and the corresponding normalized weights reduce to equal weights for observations $i = 2, \dots, n - 1$. As may be intuitive, the weights also reduce to equal values as autocorrelation approaches zero.

The effect of the weighted mean is to give greater value to observations that are temporally far from other observations. The impact on a collection of means may be small but removes some of the variability due to oversampling an unusual episode within the month. This effect can become even more important when, as in the case of trends, an unusually large or small value can have a large influence on the statistical results and their scientific interpretation.

We note that the autocorrelation observed in the data reflect only the information of the data collected. Biases in collecting of data, for instance, only launching sondes on clear-sky days, cannot be properly accounted for without independent data for corroboration of results.

3. Spatial averages

Assessment of regional climatologies and/or regional trends is often based on spatial averages. For example, in understanding the behavior of the Arctic atmosphere, some averaging of data collected from different stations located at high northern latitudes would be a natural choice. Usually, these stations are chosen to be representative of the geographic area but are not regularly spaced nor have they been optimized, in any strict sense, to cover spatial variability and spatial correlation. We can see that, similar to the impacts of uneven temporal sampling discussed in section 2, uneven spatial sampling can result in overemphasis of certain areas if geographic averages are not calculated with care.

Hence, if the interest is in estimating the spatial average of temperature, say θ , at, for example, 500 hPa, on a domain D given by the spherical cap over the Arctic, we commonly use n observations $\mathbf{y} = (y_1, \dots, y_n)$ made at locations (x_1, \dots, x_n) in domain D . Under suitable assumptions, there is an optimal estimator of θ given by a linear function $\hat{\theta} = \mathbf{A}\mathbf{y}$ where the matrix \mathbf{A} is the kriging matrix $\mathbf{K}(\mathbf{x})$ integrated over the domain D (see, e.g., Cressie and Wikle 2015). Note that matrix \mathbf{A} depends on a spatial covariance function among observations $c(x, x') = \text{cov}[y(x), y(x')]$ for any two points x and x' in domain D .

When the (x_1, \dots, x_n) are not fixed in advance, we are interested in network design. In this case, all design points $(x_1, \dots, x_{n-1}, x_n)$ or only the last design point x_n may be decided by minimizing the variance of $\hat{\theta}$.

In this context, network representativeness may be assessed by considering the uncertainty of the optimal estimate of \mathbf{y} at an unobserved location \mathbf{x} , namely, $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{K}(\mathbf{x})\mathbf{y}$, which is given by the root-mean-square error (RMSE), namely, $\text{RMSE}[\mathbf{y}(\mathbf{x}) - \hat{\mathbf{y}}(\mathbf{x})]$. It follows that a network gap may be defined for the region R where the uncertainty exceeds a certain threshold h :

$$R = \{\mathbf{x}: \text{RMSE}[\mathbf{y}(\mathbf{x}) - \hat{\mathbf{y}}(\mathbf{x})] > h\}. \quad (6)$$

An alternative to this approach is considered in section 9, which does not require detailed knowledge of the covariance function.

The approach discussed above is suitable for data at a fixed altitude. If the interest is on the spatial average of temperature, say θ , in a spherical shell, for example,

between 1000 and 500 hPa over the Arctic cap, a natural choice is to use data in the form of atmospheric profiles. In this case, the appropriate statistical modeling technique may be based on a functional data approach (see, e.g., Fassò et al. 2014; Ignaccolo et al. 2015). Now, θ is given by the integral over a volume, but its optimal estimator $\hat{\theta}$ may be expressed again as linear function $\hat{\theta} = \mathbf{A}\mathbf{y}$ where now matrix \mathbf{A} depends on the functional spatial covariance and is given by the integrated functional kriging estimator. Network design may similarly extend to functional data (e.g., Royle and Nychka 1998).

4. Representativeness of individual stations

The question of spatial representativeness is critical to planning any network. Even for well-behaved spatially smooth parameters such as wind, pressure, and humidity, the question of spatial representativeness in a climate observing network is difficult to address and depends strongly on the climate question being asked. For instance, monitoring for extreme events, such as strength of hurricanes or tornadoes, may require a completely different spatial sampling than monitoring for average air temperature. In this work, we focus on representativeness of individual stations for estimating monthly variability, such as the monthly variability of temperature and humidity, as well as possible changes in the monthly mean values. In this study, we look at how well an individual station might be able to represent changes taking place at other nearby locations. Common sense indicates that high-quality monitoring stations do not need to be placed very close to each other unless redundancy is desired in the network. The question addressed here is this: How close is redundant? This can best be addressed with long-term observations at many nearby sites, which is often not possible. Alternatively, this question can be addressed with satellite data, as we do here, or high-quality model output.

MSU temperature data are created from merged satellite observations going back to 1978. The data have been intensely studied, and corrections have been applied that account for a number of factors, including satellite drift, instrument degradation, new instruments coming on line, and diurnal sampling (Lanzante et al. 2003; Mears et al. 2003; Christy et al. 1998). The temperature values represent estimates of temperature in a broad vertical region based on the instruments' capabilities. In this paper, we look at two of the more robust MSU datasets: the upper-tropospheric and lower-stratospheric deseasonalized monthly averaged temperature data; we find similar results for spatial representativeness for both, with the differences discussed below. Comparison between

standard radiosondes and MSU tropospheric data consistently shows strong agreement for monthly averaged data with both the tropospheric (Christy et al. 2000) and stratospheric (Spencer and Christy 1993) datasets, with correlations between MSU and sondes often above 0.94 for monthly averaged data.

Although many metrics can be used, after consideration, we chose the Pearson correlation coefficient r to serve as an appropriate metric to describe representative observations for understanding global representativeness. The Pearson correlation coefficient r is the most commonly used correlation metric, is used in a wide range of fields, is easy to calculate, and is readily understood by many. For our spatial correlations, we consider temperature time series of N deseasonalized monthly averages $y_1(x), \dots, y_N(x)$ at locations or pixels x , which are identified by their coordinates, $x = (\text{Lat}, \text{Lon})$. In particular, we consider observations at a specific location x^* and calculate how they correlate with observations at all other locations x by means of correlation coefficient $r(x^*, x)$. Since monthly means are deseasonalized, they are zero mean time series, and the correlation coefficient $r(x^*, x)$ is given by

$$r(x^*, x) = \frac{\sum_{m=1}^N y_m(x^*)y_m(x)}{\sqrt{\sum_{m=1}^N y_m(x^*)^2 \sum_{m=1}^N y_m(x)^2}}. \quad (7)$$

We point out that the Pearson correlation coefficient has a number of properties that are helpful for interpreting the results presented later in the paper: 1) results are highly dependent on what time unit is used for averaging—for this work, we use monthly averaged deseasonalized values; 2) negative values can be interpreted as teleconnections with opposite signs for variations; 3) values ignore time-lagged relationships and only focus on concurrent values; 4) a high correlation can exist even if the two locations have very different magnitudes of variability; and 5) a high correlation can exist even if the two locations have very different mean values. These last two properties are very important because establishing a network based on correlations will allow estimates of relative variability in places not being monitored but will not allow estimates of means or absolute variability.

For this study, we examine GRUAN as a network of stations for which we would like to understand the spatial representativeness of the individual stations and we use available MSU monthly averaged deseasonalized observations as our reference dataset. The monthly averaged upper-tropospheric data span January 1981–February 2016, and the lower-stratospheric data span

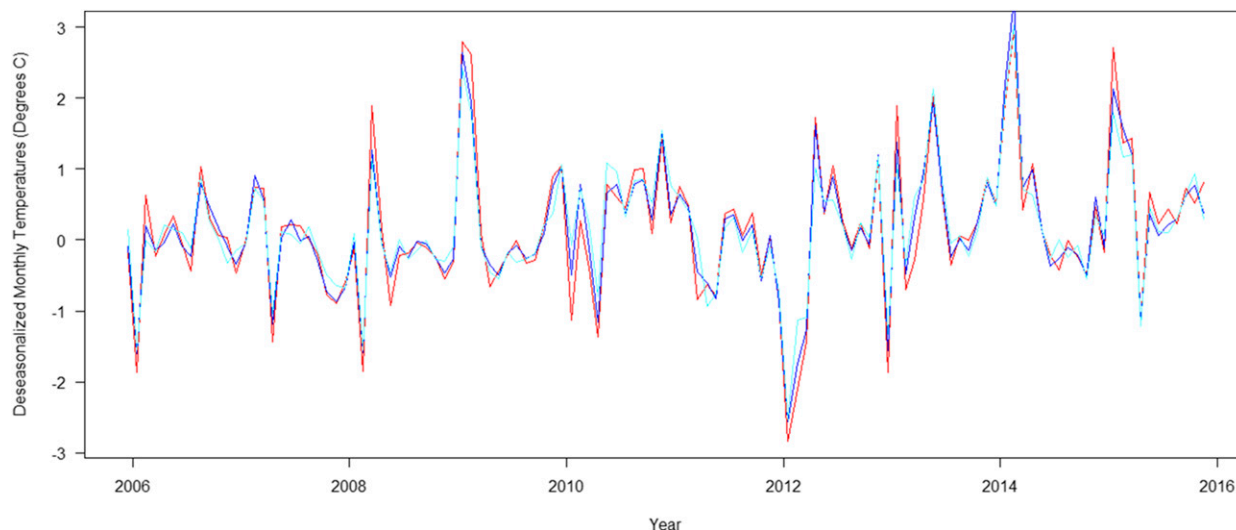


FIG. 3. Time series of monthly averaged, deseasonalized temperature data from MSU upper-tropospheric data for locations over Cabauw (red), Paris (purple), and Payerne (light blue). The data show such strong similarities in the time series for the 10 yr presented that the individual stations are hardly distinguishable.

November 1978–February 2016. If deseasonalized values were not used, most locations outside of the tropics would correlate strongly with each other—either positively for locations in the same hemisphere or negatively for locations in opposite hemispheres. If satellite data were not available, global model output may have been used, as is often the case for some climate observing system simulation experiments when no data exist (e.g., Andersen et al. 2006; Feldman et al. 2015).

To begin our exploration of spatial coherence using temperature data, we consider 10 yr of monthly averaged, deseasonalized, MSU data from three European locations: Cabauw, Netherlands; Paris, France; and Payerne, Switzerland, in Fig. 3. We see that the variability in the monthly averaged data shows great similarity for the locations Paris and Cabauw, with a correlation coefficient of 0.97 and a correlation of Paris and Payerne of 0.95; however, the correlation drops for Cabauw and Payerne to 0.91 because of the larger distance between these two locations.

Representation of common variability

Expanding this example of the spatial cohesiveness of these three locations, we look for the general relationship of an individual station's monthly averaged time series of upper-troposphere temperature at Cabauw with locations around the globe in a manner similar to Sofen et al. (2016). For the discussions in this paper, we refer to correlations of $r > 0.7$ as “well correlated” and $r > 0.9$ and above as “strongly correlated.”

Figure 4 shows results for the correlation of the deseasonalized monthly averages of upper-troposphere

MSU time series over Cabauw with the time series for the rest of the world. The results show high correlation between Cabauw and much of northern Europe. The areas around the globe where upper-tropospheric temperature data are strongly correlated with Cabauw ($r > 0.9$) cover $1.7 \times 10^6 \text{ km}^2$, and areas that are well correlated with Cabauw ($r > 0.7$) cover $6.8 \times 10^6 \text{ km}^2$. For the lower stratosphere, we see roughly similar spatial coherence, with the area of the earth well correlated with Cabauw as $5.7 \times 10^6 \text{ km}^2$ and the area strongly correlated with Cabauw as $1.6 \times 10^6 \text{ km}^2$.

The spatial extent of representativeness, as suggested by the correlation coefficients derived from MSU data,

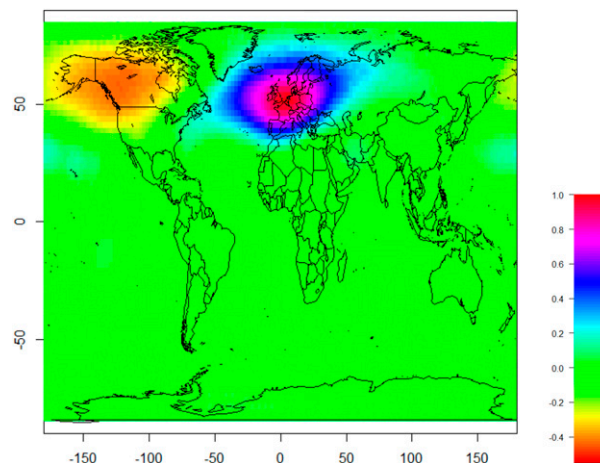


FIG. 4. Correlation of deseasonalized, monthly averaged MSU satellite temperature data for all pixels around the globe with the pixel over Cabauw.

TABLE 1. Value of confirmed GRUAN stations. Area represented by correlation of MSU temperature in the upper troposphere and lower stratosphere for each of the nine certified GRUAN stations. Results are shown for correlation of 0.7 and 0.9 when using the monthly deseasonalized temperature data. Each station is considered separately, without consideration for the other stations in the network.

GRUAN location	Upper troposphere		Lower stratosphere	
	Area represented by a correlation ≥ 0.7	Area represented by a correlation ≥ 0.9	Area represented by a correlation ≥ 0.7	Area represented by a correlation ≥ 0.9
Ny Ålesund (78.92°N)	14.3×10^6	4.2×10^6	8.6×10^6	2.9×10^6
Sodankylä (67.41°N)	9.3×10^6	2.7×10^6	8.4×10^6	2.4×10^6
Cabauw (51.97°N)	6.8×10^6	1.7×10^6	5.8×10^6	1.6×10^6
Lindenberg, Germany (47.60°N)	7.2×10^6	2.0×10^6	5.5×10^6	1.5×10^6
Payerne (46.82°N)	6.9×10^6	1.8×10^6	5.0×10^6	1.3×10^6
Potenza, Italy (40.64°N)	7.0×10^6	1.7×10^6	5.0×10^6	1.1×10^6
Boulder, Colorado (40.00°N)	8.6×10^6	2.0×10^6	4.6×10^6	1.1×10^6
Beltsville, Maryland (39.03°N)	6.1×10^6	1.1×10^6	4.1×10^6	1.2×10^6
Lauder (45.05°S)	10.3×10^6	2.3×10^6	9.1×10^6	2.1×10^6

can vary strongly by location. Figure 4 could erroneously be interpreted as implying that we need observing systems spaced roughly 1500 km apart to assure that all areas of the earth are well correlated ($r > 0.7$) with at least one monitoring station. Because of the many physical and chemical processes that drive variability as well as spatial correlations, some regions of the globe are correlated over a larger spatial extent than others. Results for spatial representativeness for all nine of the currently certified GRUAN stations are presented in Table 1.

Each of the certified GRUAN stations show areas of representativeness in the upper troposphere with areas well correlated with at least one GRUAN station as being between 6.1×10^6 and 14.3×10^6 km². For areas that are strongly correlated with at least one GRUAN station, the areas are reduced by roughly 75%. Relative to the upper troposphere, representativeness in the lower stratosphere is usually smaller by roughly 30% for both $r > 0.7$ and $r > 0.9$ criteria. The similarities between locations are due, in large part, to most of the current GRUAN stations being in or near the midlatitudes. Lauder, New Zealand;

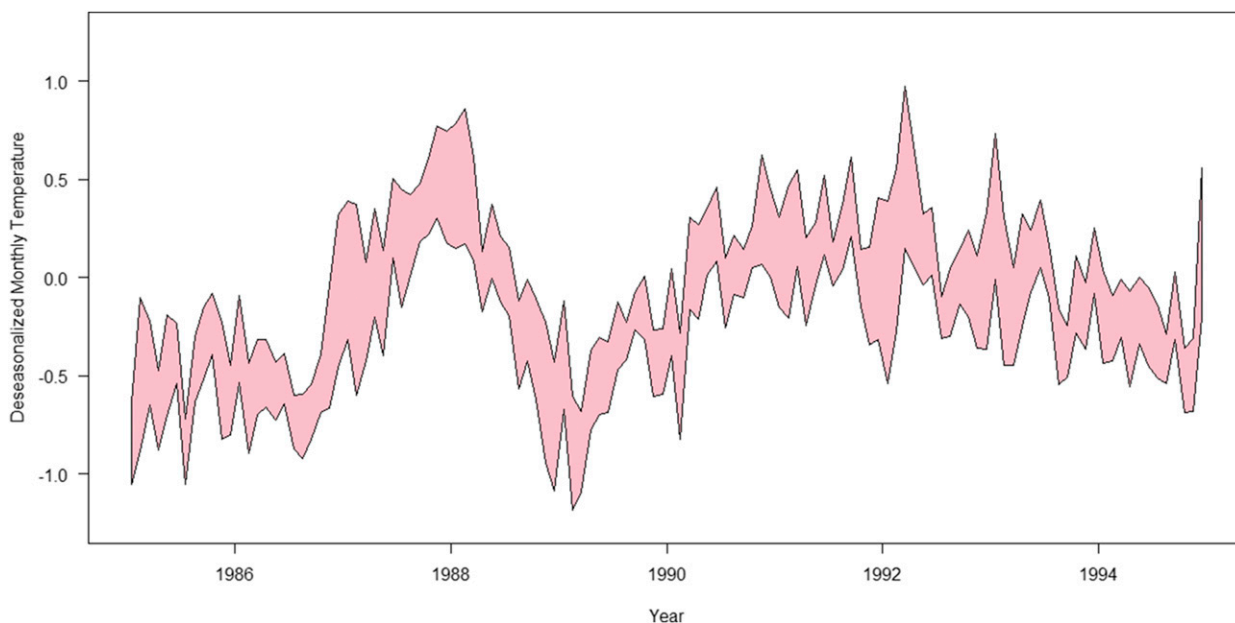


FIG. 5. Correlations of monthly averaged data across the tropics can be understood from this plot, which shows the range of temperature anomalies from MSU upper-tropospheric monthly averaged deseasonalized data across the full longitudinal range at the equator. The common variability results in high correlations among all locations in the tropics; this differs from other regions, such as the midlatitudes, where strong positive correlations do not exist across such long distances.

Ny Ålesund, Norway; and Sodankylä, Finland, have much larger correlation areas, potentially because Lauder is surrounded by the Southern Ocean and Ny Ålesund and Sodankylä are located in the Arctic, where orographic features are minimal and isolated air masses allow for larger spatial coherence.

Figure 5 shows how similar temperature data from different sites in the tropics behave when evaluated on a monthly basis. The figure shows the full range (maximum and minimum) of monthly averaged anomalies for all MSU upper-troposphere data at the equator. The large redundancy in information does not imply that only one location would be needed to monitor the tropics. Such minimal monitoring would limit data users' ability to address a large number of important scientific questions, including questions concerning the evolution and maximum strength of El Niño events, the potential expansion of the tropics, or the influence of regional activity such as volcanic eruptions, large spread fires, monsoons, or local ocean temperature changes.

It should be noted that the agreement between MSU-gridded data and GRUAN individual-site data are challenged by spatial averaging issues (Schutgens et al. 2016) as well as temporal sampling issues, including those issues discussed in sections 2 and 3 as well as fundamental measurement issues, some of which are discussed in Ladstädter et al. (2015), Sairanen et al. (2015), and Wulfmeyer et al. (2015). The advanced techniques for creating monthly averages presented in section 2 will help increase the agreement between MSU temperature observations and in situ GRUAN observations, although a detailed comparison is beyond the scope and intent of this paper. We also note that the results here can help guide the use of high-quality station data, such as GRUAN, in helping merge datasets from different platforms: the level of agreement between stations can be estimated with correlation studies and help identify agreement with other datasets. (Weatherhead et al. 2017)

5. Representativeness of a network of stations

Under the leadership of the WMO and other agencies, many individually maintained stations are coordinated in their operations and held to similar standards that allow consideration of the network as a whole for monitoring purposes, as opposed to simply individual efforts. GRUAN is an example where, under the auspices of UNEP, the Intergovernmental Oceanographic Commission, International Council for Science, and WMO, a network has been established to make reference-quality observations of the upper air to provide long-term, high-quality climate records, constrain and calibrate data from more spatially

comprehensive global observing systems (including satellites and current radiosonde networks), and fully characterize the properties of the atmospheric column. As of January 2017, there are nine certified GRUAN stations adhering to the requirements of reference-quality measurements. Reference measurements are traceable to an SI unit or an accepted standard, provide a comprehensive uncertainty analysis, can be traced back to an archive of raw data, include complete metadata descriptions, have processing techniques that are documented in accessible literature, and are validated, for example, by intercomparison with redundant observations. In addition to the 9 currently certified stations, there are a further 13 stations preparing for, or awaiting, certification.

We carry out the same calculations using MSU data as shown in Fig. 4 but now at the locations of all nine GRUAN stations. This gives the nine correlations $r(x_1^*, x), \dots, r(x_9^*, x)$ for each pixel x around the globe, and the maximum correlation map is given by the highest correlation with respect to the nine GRUAN stations:

$$r(x)^{\max} = \max[r(x_g^*, x) \quad \text{for } g = 1, \dots, 9]. \quad (8)$$

These calculations inform us of how well the network—as opposed to a single station—can correlate with any location around the world. Thus, stations that are very close to each other are observing similar air masses, and their coverages, in the context of this network analysis, are not counted twice. We present these results in Fig. 6 and can readily see that most developed regions are represented by at least one GRUAN site with a correlation of at least 0.2, but with the oceans well underrepresented.

The network calculations can give us some new insights into the network's representativeness. For this upper-air example, roughly 3% of the earth's surface is strongly correlated ($r > 0.9$) for upper-tropospheric temperature with at least one of the certified GRUAN stations and approximately 10% of the earth is well correlated ($r > 0.7$) for at least one GRUAN station. For the lower-stratospheric temperatures, we see similar results, with the percentage of the earth well correlated with a GRUAN station at 8% and the percentage of the earth strongly correlated with a GRUAN station at 3%. In general, we see similar results for the troposphere and stratosphere, with the results for individual stations summarized in Table 1.

Figure 6 reveals a number of features about the current configuration of GRUAN. First, the tropics are poorly represented by the current network. Africa, South America, Australia, and Asia are not well represented by the existing set of certified GRUAN stations. Although Sodankylä and Ny Ålesund provide coverage

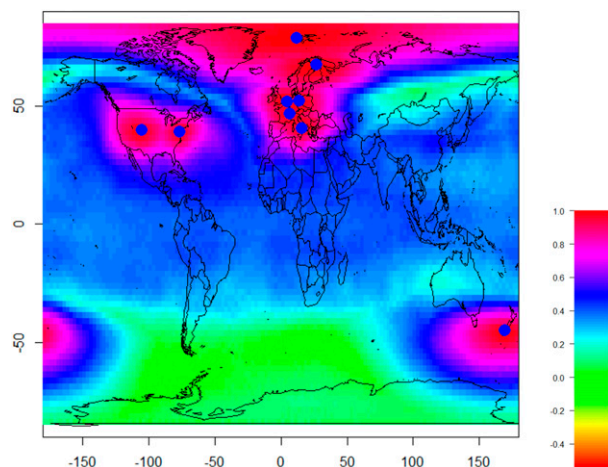


FIG. 6. Maximum correlation for each location around the globe with the single highest correlated GRUAN station when the current nine certified GRUAN stations are considered.

of the Arctic, given the Arctic Oscillation and potential future changes in the Arctic, sites in the Alaskan Arctic may be warranted.

Correlating monthly averages results in some of the major climate indices being represented, such as El Niño–Southern Oscillation, the Pacific decadal oscillation, and the Arctic Oscillation, because these features can dominate some of the upper-troposphere temperature variations. However, the ability to monitor these features, and their potential changes, would require additional climate observing system simulation experiments to optimize monitoring to be able to assess any future changes in intensity or areal coverage of these phenomena.

6. Expansion and contraction of networks

Some of the strengths and challenges of networks composed of a small set of isolated stations are the possibilities for expansion or contraction of the network. Studies, such as this one, that make use of available satellite or model output can allow for insights into the value of proposed changes to a network. In this section, we look at the potential impact of additions to the current set of nine certified GRUAN stations. Figure 7 shows the same results presented in Fig. 6 but now uses all locations formally in GRUAN, including the 9 certified stations, 13 stations awaiting certification, and 2 stations no longer in operation. Table 2 shows the impact to the network coverage of adding each of the 13 stations awaiting certification to the network of 9 certified stations.

Most notably, MSU upper-tropospheric temperature data at Darwin, Northern Territory, Australia, a proposed GRUAN station, are well correlated ($R = 0.7$) with

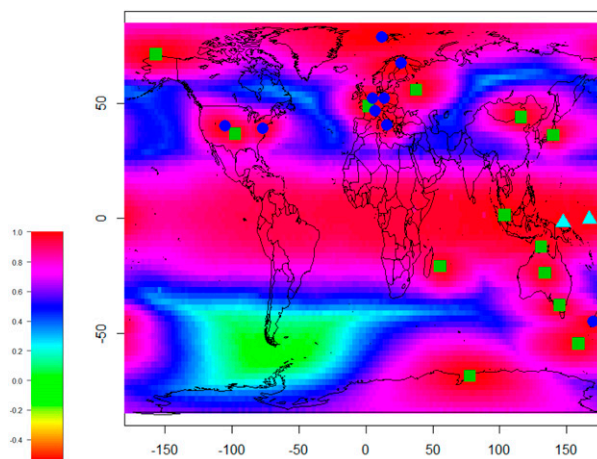


FIG. 7. Maximum correlation of upper-tropospheric temperature is shown for the complete 24 GRUAN stations (including 9 certified, 13 yet to be certified, and 2 now-inactive sites). Results are for the maximum correlation at each pixel with any one GRUAN station. Dark blue dots indicate existing certified stations; green squares indicate sites preparing for, or awaiting, certification; and the light blue triangles indicate GRUAN stations no longer submitting data (Manus and Nauru).

$177 \times 10^6 \text{ km}^2$ in the troposphere and slightly less area in the lower stratosphere. When strongly correlated criteria are considered ($R > 0.9$), these areas become $23 \times 10^6 \text{ km}^2$ and $53 \times 10^6 \text{ km}^2$ in the troposphere and stratosphere, respectively. Darwin correlates strongly with most sites in the tropics in part because ocean-driven phenomena such as ENSO have a widespread impact on the tropical atmosphere, as illustrated in Fig. 5.

If the entire set of proposed GRUAN stations is adopted, the full network will result in roughly 26% of the area of the earth strongly correlating, in the upper troposphere, with at least one of the GRUAN stations. The geographic representation of this network (Fig. 7) shows the necessity for stations beyond the 13 current awaiting certification, particularly stations in South America, Antarctica, northern Africa, Canada, central Asia, and the Pacific Ocean.

In recent years, Nauru and Manus have exited from GRUAN. The impact to the global coverage by losing those two stations has been large. An 11-station GRUAN (current 9 certified stations plus Nauru and Manus) offered coverage of the upper troposphere for $234 \times 10^6 \text{ km}^2$ when correlations of 0.7 were considered, while the current 9 GRUAN stations only offer coverage of $52 \times 10^6 \text{ km}^2$, with similar degradation observed for the coverage of the stratosphere. Much of that lost area is due to the poor coverage of the tropics with the current network and the strong coherence of temperature data in the tropics, allowing for Manus

TABLE 2. Value of proposed GRUAN stations: area (10^6 km^2) represented by correlation of MSU temperature in the upper troposphere and lower stratosphere for each of the 13 stations being considered for GRUAN certification. Four results are shown for each of the troposphere and stratosphere: 1) the area that is moderately correlated with the station, 2) the area that is strongly correlated with the station, 3) the additional area that would be added to the network coverage when moderate correlation ($r > 0.7$) is considered, and 4) the additional area that would be insured against dropouts when moderate correlation ($r > 0.7$) is considered.

Station name	Upper troposphere				Lower stratosphere			
	Area represented by a correlation ≥ 0.7	Area represented by a correlation ≥ 0.9	Additional area of network coverage at 0.7	Additional area of insured coverage at 0.7	Area represented by a correlation ≥ 0.7	Area represented by a correlation ≥ 0.9	Additional area of network coverage at 0.7	Additional area of insured coverage at 0.7
Barrow (71.29°N)	9.8	2.7	7.4	2.4	9.7	2.8	9.7	0.0
Dolgoprudny (55.95°N)	7.1	1.9	2.2	1.8	5.4	1.5	2.1	1.8
Paris (48.86°N)	6.6	1.6	0.3	0.6	5.4	1.4	0.1	0.9
Xilinhot (43.93°N)	9.7	2.3	9.7	0.0	8.3	2.1	8.3	0.0
Tateno (36.72°N)	8.4	1.1	8.4	0.0	6.1	1.5	6.1	0.0
Lamont (35.26°N)	9.0	1.5	1.1	5.8	4.3	1	1	3.3
Singapore (1.35°N)	185.3	83.7	185.3	0.0	136.4	82.6	136.4	0
Darwin (12.46°S)	177.1	23	177.1	0.0	164.0	53.4	164.0	0.0
Réunion (21.11°S)	93.7	3.8	93.7	0.0	49.5	2.7	49.5	0.0
Alice Springs (23.70°S)	15.8	3.0	15.8	0.0	13.6	2.3	13.6	0.0
Macquarie (33.77°S)	12.6	3.2	5.3	7.3	12.3	3.3	5.7	6.6
Melbourne (37.81°S)	9.2	1.4	6.3	2.9	6.3	1.3	4.7	1.6
Davis (68.58°S)	15.2	3.8	15.2	0.0	13.8	3.3	13.8	0.0

and Nauru to offer significant information for the large tropical region.

7. Optimization of additional stations

With a finite number of proposed stations—in the case of GRUAN, 13 stations currently await certification and formal inclusion in the network—it is relatively easy to quantify the impact of each additional site, as shown in [section 6](#). If a metric of success is agreed to, such as the sum of areas of interest covered by moderate or high correlation for a specific parameter with at least one network station, proposed stations can be ordered by impact on that metric. The proposed stations can be considered both in isolation or, potentially more relevant, as incremental additions to the existing network. In performing this ordering, two complications emerge for placing a priority on one proposed site over another: 1) there will always be more than one metric by which to evaluate a site, and 2) often, the optimal network will not be achieved by addressing each proposed site individually and sequentially.

Most monitoring networks serve multiple purposes. Even networks focused on climate, such as GRUAN, will serve multiple functions (such as monitoring temperature, humidity, and winds) and will serve different communities (such as the climate trends detection, numerical weather prediction, atmospheric process studies, and satellite calibration–validation communities). These multiple functions will have different metrics for evaluation, and each can be weighted for their relative usefulness in obtaining an optimal configuration of stations based on a balanced approach with limited resources. However, this will require a subjective decision on how to determine the importance of each function within the monitoring network. WIGOS continues to lead critical conversations both to identify the highest priorities for observations and to help decision-makers balance the many factors of importance to participating parties. While prioritization is challenging, monitoring climatological levels of key parameters is critical to addressing the broad range of science goals.

Examining the value of individually proposed stations, and always adding the station of highest value, can result in a poorly designed network. For instance, if only one monitoring site is to be added to the Pacific Ocean, evaluation of proposed stations might result in an addition near the middle of the Pacific. Then, if at some later date, another Pacific Ocean site is desired for inclusion, the new site may be forced to be placed in a suboptimal location because of the choice for the first site being in the middle of the Pacific Ocean. However, if it is known that the possibility exists for a second site at some later date when the first one is being decided, a

better choice may emerge. Essentially, if an efficient network is to be constructed, with uncertainty being a part of that planning process, multiple choice avenues and risk functions need to be examined within the context of an optimization approach.

For a relatively low number of site choices, calculations can be performed easily. For example, with the existing GRUAN of 9 certified stations, if it was decided that only 2 of the additional 13 stations were to be confirmed, only 78 calculations ($\frac{1}{2} \times 13 \times 12$) would need to be made. Just by deciding to include a third site in the same set would result in 858 calculations. If we now assume a much larger set of possible stations, say 100 possible new locations, and we want to select 10 stations, there are more than 17×10^{12} combinations to consider. This number may be too high for direct calculation ([Schrijver 1998](#)). Expansion of this larger network is much better computed using optimization theory (see, e.g., [Dantzig and Cottle 1963](#); [Schrijver 1998](#); [Boyd and Vandenberghe 2004](#); [Clack et al. 2015](#)), which identifies strategies likely to result in the optimal results but does not directly test each possible configuration.

When optimization theory is considered, defining the metric of success is of paramount importance, pointing again to the important conversations carried out about prioritization of environmental observations. The choices will constrain the number of combinations that are necessary to be solved for an optimized network. Additional information can be provided, such as the limited distance between stations, locational-dependent variables that must be considered, redundancy values, time series for each possible site, and uncertainty fields. Economic constraints will be considered in [section 10](#). The more information that is provided to the optimization, the more robust the output will be. Indeed, the impact of each constraint can be evaluated in such an approach.

8. Network robustness

Current and future networks are never guaranteed to continue; unexpected disruptions occur in one or more single stations for a ground-based network. While [section 7](#) discussed planned expansion and contraction of a network to allow network coverage, this section addresses the need for redundancy in planned monitoring networks, particularly when those networks are extremely important to scientific and societal goals. Redundancy in monitoring can be valuable both to cover the possibility of a station becoming inactive, as well as to allow quality assurance of unusual observations. Using similar “areas of correlation” analyses as presented in [section 5](#), we can ask how robust areas of correlation are to removal of a single station.

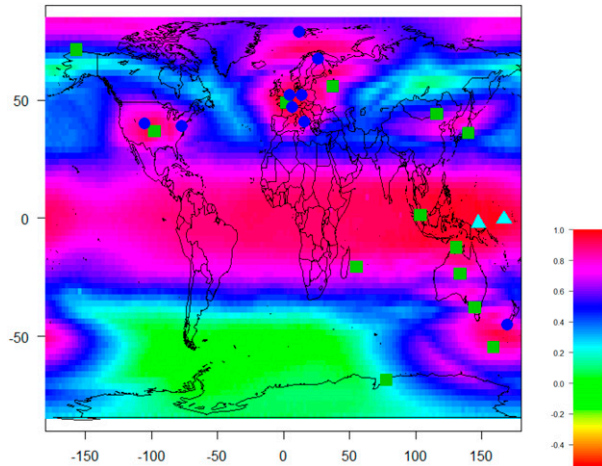


FIG. 8. Level of redundancy of representativeness for each location around the world based on the proposed 24-site GRUAN. The correlation listed is the highest correlation that would exist if the single most important station for that location were removed. Symbols are as in Fig. 7.

We can, with the small number of stations in GRUAN, identify the impact of removal of any one station. We can identify the largest impact to each pixel around Earth resulting from the removal of any one station using

$$r(x)^{\max_2} = \max_2[r(x_g^*, x) \text{ for } g = 1, \dots, 9], \quad (9)$$

where \max_2 refers to the second-highest values in this series of nine values. In this sense, we identify the level of redundancy of information, or “insurance,” for each location around the world. Figure 8 shows the results of these calculations for the full 24-station GRUAN. Note that if the full 24-station monitoring network existed, Davis, Antarctica, as well as much of the far southern latitudes, would not be redundantly covered, meaning that unusual observations could not be confirmed and representativeness of the network would be strongly affected if a southern-latitude site dropped from the network. Similar calculations carried out for just the current nine certified GRUAN stations indicate that only parts of Europe and North America would be well represented with the loss of any one of the nine stations. Ideally, a robust network would be highly resilient to the loss of a single station and the percentage of the earth’s surface well correlated with a nearby station would change very little with the loss of any one station.

We consider the percentage of the earth that would still be well correlated with a GRUAN station after the loss of a single station as the “insured” coverage, a gauge of network robustness against station dropouts. The percentage of the earth that would be well correlated with the full 24-station GRUAN is 61%; however, only 44% of

the earth would be robustly covered—that is, only 35% of Earth would still be well covered if one of the nearby stations dropped out of the network. Coverage in the lower stratosphere for the full 24-site network would be 52%, but only 35% of the earth is immune from fallout of a single station. Because the value of redundancy in the network is both to assure against dropouts of stations but also to assure that unusual events can be confirmed by more than one site, we may, for some networks, choose to look at a higher-level insurance. That is, we may ask how robustly Earth would be represented with moderate correlation if two stations dropped out of a network; such results are not presented in this paper. This examination of insurance may be even more important for network planning after large coverage is already obtained.

As the working group on GRUAN and individual countries consider decisions on additional stations, among other factors to consider is the value of those stations to the global coverage by GRUAN. Table 2 summarizes results for the 13 stations being considered for certification as GRUAN stations. For each of these stations, we calculate 1) the area represented by that station, based on available MSU data, 2) the additional area that the site would bring to the GRUAN, and 3) the insurance by offering redundancy to regions of the earth that are currently only well represented by a single GRUAN station. Table 2 summarizes these results for both the upper troposphere and lower stratosphere.

Note that some locations in the tropics, such as Darwin and Singapore, represent large areas both individually and as additions to the current network, because much of the variability in monthly averaged levels is from ENSO and shows a signal throughout the tropics where the current GRUAN is lacking coverage. For some stations, such as Réunion, France, the addition of that station offers little information as insurance against the dropout of any single station, but they are still critical monitoring stations for the area that they do represent; in the case of Réunion, the addition of that station would add $93.7 \times 10^6 \text{ km}^2$ of coverage to GRUAN that would not otherwise be represented.

9. Alternative metrics for evaluating a network

Given the multiple communities any one monitoring network may serve, including weather, climate, hydrology, research, and aviation, there can be a large number of metrics for evaluating the adequacy of a network. For climate purposes, another valid approach to evaluating sites for an existing or proposed monitoring network is to determine whether different locations can be expected to reproduce similar global means (Chang et al. 2015) or to observe similar trends and whether those trends will be efficiently detected (Weatherhead et al. 2002).

Monitoring and understanding major structures of the climate system, including Hadley circulation, ENSO events, Pacific decadal oscillation (PDO), and cyclone development will be key to improving our ability to project future climate changes (de Viron et al. 2013; Messié and Chavez 2011). Designing a network that will adequately monitor these aspects of the climate system will be challenging and is beyond the scope of this paper. However, a basic understanding of these phenomena pushes further for more stations in the tropics, Southern Hemisphere, Arctic, and over the oceans.

Even for identifying the spatial coverage needed, there are a number of approaches beyond the approaches presented in sections 4–6. In this study, we examined the possibility of using lagged correlations: for two time series y_t and y'_t , an approach finds the time lag d that maximizes the cross-correlation coefficient $r(y_t, y'_t)$, which is equivalent to the approaches presented in section 4 for the case when $d = 0$. We found substantially similar results for the lagged correlations as for the non-lagged correlations.

Further correlation approaches, such as the Shannon information index (Silva and Quiroz 2003; Madonna et al. 2014) or directional information transfer index (Alfonso et al. 2010), can measure the more general nonlinear information. It is computationally intensive for a large dataset and thus unsuitable for the application of global network assessment.

In this paper, we quantify the representativeness of a station in GRUAN (including certified, yet to be certified, and inactive stations). In the case where the locations of potential new stations are undetermined, we can adopt the space-filling approach, based on maximum or minimax criterion (Royle 2002; Tan 2013), to improve the network coverage. For instance, we can use such techniques to determine the next most valuable location across South America, Antarctic, northern Africa, the Arctic (particularly Greenland), and the Pacific Ocean to be added in GRUAN.

We carried out a number of variations of the techniques presented in sections 4–6 and found agreement with many alternative techniques, including covariance, time-lagged correlations, and seasonal correlations. For those techniques tested, we found little change from our fundamental conclusions. We consider one strength of our current approach to be the ease of calculating and interpreting different options for network configurations.

The techniques presented are useful when there is coherence to the variations in the parameter being considered, even if that coherence has a relatively short spatial scale. For instance, a network of soil moisture monitors in a complex terrain situation may

have correlations of a few kilometers or less; potential contaminants may have a spatial correlation of meters. Even in these complex situations, it is unlikely that a single spatial length will be adequate to describe the coherence in the region and the need for individual monitoring. The combined use of scientific insight and available data will likely give the best guidance for network design.

These techniques do not work well if the spatial scale is fractal in nature and spatial distance is not a good indicator of common variability. This can happen in near-surface phenomena depending on the spatial scale of interest. For instance, mountains are often identified as fractal (Prusinkiewicz and Hammel 1993) and may require special techniques (e.g., Strachan et al. 2016; Mountain Research Initiative Elevation-Dependent Warming Working Group 2015; Daly et al. 2010). Yorgun and Rood (2015) investigated the spatial continuity of orographic precipitation using climate model simulations and synthetic data using variogram analysis. The results of this study indicated significant differences in the behavior of the variograms of orographic precipitation when compared with that of the variables with spatially coherent nature such as temperature. A more focused study, including case studies with varying network station locations (e.g., investigation of the spatial correlation when a station is located in the windward vs leeward side of a mountain to quantify the differences in wet vs dry regimes) will be valuable in order to design an efficient observing network representing mountainous regions. Depending on the network purpose, complex terrain may also require special considerations (Abatzoglou 2013; Pepin and Lundquist 2008). This technique does require available information for estimating coherence scales, and short-term campaigns may be needed to help establish those spatial scales.

10. Economic analyses of network decisions

Decisions to improve, add, or change locations of stations should be driven by considerations of the societal benefits and costs of network design. Benefit–cost analysis (BCA) approaches should be implemented for any program with significant costs or potentially significant societal benefits or impacts such as global observing systems (WMO 2015). In BCA, the present value of costs (i.e., the discounted value of the flow of costs) is subtracted from the present value of benefits to indicate whether a program is economically viable (i.e., benefits exceed costs). BCA analysis of different network options can also be compared to identify options that maximize socioeconomic benefits.

Costs are generally borne by the network sponsors and may be readily apparent in part as they tend to be near term and “upstream.” On the cost side, according to the 2010 U.S. GRUAN implementation plan (NOAA/National Climatic Data Center 2010), total costs, including equipment, sondes, labor, data management, and overall system coordination, over 5 yr for 7 sites were projected to be $\$8.325 \times 10^6$ (Tables 1 and 2), averaging to about \$240,000 per site per year for 1 weekly and 1 monthly additional launch of high-quality radiosondes. While this cost estimate included weekly sonde deployment, GRUAN sites may cost up to \$1 million per year with more frequent launches. Additional costs may be incurred for further data processing, storage, communication, and use in specific models or applications, but these are likely minor relative to capital costs.

On the benefit side, data-quality improvements should be evaluated in terms of potential changes in societal outcomes from the use of improved information [e.g., using a “value of information” (VOI) approach; Laxminarayan and Macauley 2012]. For VOI analysis, it is necessary to determine how improvements in observing systems translate into improved information for decision-makers and potentially generate socioeconomic benefits. As benefits tend to accrue “downstream” to a spatially, temporally, and societally diverse decision-makers, benefits may be more difficult to identify and quantify (Morss et al. 2005).

Currently, there is fairly limited literature on the value of Earth observations and climate modeling for policy on which to build such economic analysis (e.g., Williamson et al. 2002; Macauley 2006a,b; Weaver et al. 2013; Cooke et al. 2014; Donaldson and Storeygard 2016). A recent review of the literature on the economic value of climate services found a fairly limited number of studies, with most of those focusing on benefits in agriculture and mostly for developed countries (Clements et al. 2013). More work in this area would be important to support policy analysis for optimization of observing systems, and, in comparison with the costs of implementing observing systems, the economic analysis itself is relatively inexpensive.

11. Conclusions

Observing networks are critical to improving our understanding of the Earth system. In the case of climate, multiple scientific questions can be addressed with any one observing system. In many cases, a global observing system will benefit from being representative of as much of the globe as possible. In this paper, we propose a technique for evaluating the representativeness of a network to address one of the basic requirements for monitoring the climate: establishing monthly means

across the globe. This technique allows the cohesiveness of the system to dictate the spatial distance between stations as opposed to an “even sampling” method or similar approach. The technique is most useful when there is some cohesiveness within the system to be observed, but the characteristics of cohesion vary by location, as with the upper-tropospheric temperature data. The same technique, however, can be applied to smaller-scale problems, including observations within a parcel of a few acres as long as some information is available to allow derivation of the spatial coherence.

Using the GRUAN site locations as an example, we calculate correlation coefficients on MSU monthly averaged data to understand the global representativeness of the network. We note that the current set of 9 certified GRUAN stations offers representativeness (correlation at the 0.7 level) for the upper-troposphere and lower-stratosphere temperatures for 10% of the earth, while expansion of the network to include the 13 stations awaiting certification and reinstating Manus and Nauru would offer representativeness (correlation at the 0.7 level) in the upper troposphere for roughly 61% of the globe. The analysis points out how underrepresented certain areas of Earth will be even with the more expanded GCOS network: South America, Antarctica, northern Africa, the Pacific Ocean, and parts of the Arctic all require additional monitoring to allow GRUAN to more completely monitor the globe. Without this analysis, the lack of observations in South America may have pointed to an obvious need for the network; with this analysis, it becomes clear that a station in the lower half of South America is highly important to allow global representativeness and that a station in the northern half of South America would add significantly less value.

Multiple decisions may be made in an effort to improve a network. Understanding the complexity of science questions and estimating societal benefits of high-quality monitoring is daunting, but significant improvements may be gained through objective evaluation as well as economic analyses with considerations of costs, benefits, and alternative choices under the constraints of fixed budgets. As informative as the analyses presented in this paper may be for showing spatial representativeness, it is equally informative for cautioning against oversimplifying network decisions. Key climate questions, such as changes in circulation, impacts of oceans, and water cycle questions, would require different climate observing system simulation experiments than those presented here. We offer this initial evaluation as a starting point for discussions on planning the climate observing systems of the future.

Acknowledgments. Funding for this work was provided by NOAA through a grant to the University of

Colorado Boulder. The work was partially funded by the European Union's Horizon 2020 Programme for Research and Innovation under Grant Agreement 640276. The MSU dataset that was used in this study is the result of the hard work of Zou et al. (2015).

REFERENCES

- Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **33**, 121–131, doi:10.1002/joc.3413.
- Alfonso, L., A. Lobbrecht, and R. Price, 2010: Information theory-based approach for location of monitoring water level gauges in polders. *Water Resour. Res.*, **46**, W03528, doi:10.1029/2009WR008101.
- Andersen, S. B., and Coauthors, 2006: Comparison of recent modeled and observed trends in total column ozone. *J. Geophys. Res.*, **111**, D02303, doi:10.1029/2005JD006091.
- Bodeker, G. E., and S. Kremser, 2015: Techniques for analyses of trends in GRUAN data. *Atmos. Meas. Tech.*, **8**, 1673–1684, doi:10.5194/amt-8-1673-2015.
- Boyd, S., and L. Vandenberghe, 2004: *Convex Optimization*. Cambridge University Press, 727 pp.
- Chang, K. L., S. Guillas, and V. E. Fioletov, 2015: Spatial mapping of ground-based observations of total ozone. *Atmos. Meas. Tech.*, **8**, 3967–4009, doi:10.5194/amt-8-3967-2015.
- Christy, J. R., R. W. Spencer, and E. S. Lobl, 1998: Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate*, **11**, 2016–2041, doi:10.1175/1520-0442-11.8.2016.
- , —, and W. D. Braswell, 2000: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons. *J. Atmos. Oceanic Technol.*, **17**, 1153–1170, doi:10.1175/1520-0426(2000)017<1153:MTTDC>2.0.CO;2.
- Clack, C. T. M., Y. Xie, and A. E. MacDonald, 2015: Linear programming techniques for developing an optimal electrical system including high-voltage direct-current transmission and storage. *Int. J. Electr. Power Energy Syst.*, **68**, 103–114, doi:10.1016/j.ijepes.2014.12.049.
- Clements, J., A. Ray, and G. Anderson, 2013: The value of climate services across economic and public sectors: A review of relevant literature. U.S. Agency for International Development Tech. Rep., 54 pp., http://pdf.usaid.gov/pdf_docs/PA00KJXW.pdf.
- Cooke, R., B. A. Wielicki, D. F. Young, and M. G. Mlynczak, 2014: Value of information for climate observing systems. *Environ. Syst. Decis.*, **34**, 98–109, doi:10.1007/s10669-013-9451-8.
- Cressie, N., 1985: Fitting variogram models by weighted least squares. *J. Int. Assoc. Math. Geol.*, **17**, 563–586, doi:10.1007/BF01032109.
- , and C. K. Wikle, 2015: *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 512 pp.
- Daly, C., D. R. Conklin, and M. H. Unsworth, 2010: Local atmospheric decoupling in complex topography alters climate change impacts. *Int. J. Climatol.*, **30**, 1857–1864, doi:10.1002/joc.2007.
- Dantzig, G. B., and R. W. Cottle, 1963: Positive (semi-) definite matrices and mathematical programming. University of California, Berkeley, Operations Research Center Rep. ORC-63-18-RR, 27 pp.
- de Viron, O., J. O. Dickey, and M. Ghil, 2013: Global modes of climate variability. *Geophys. Res. Lett.*, **40**, 1832–1837, doi:10.1002/grl.50386.
- Dobbie, M. J., B. L. Henderson, and D. L. Stevens Jr., 2008: Sparse sampling: Spatial design for monitoring stream networks. *Stat. Surv.*, **2**, 113–153, doi:10.1214/07-SS032.
- Donaldson, D., and A. Storeygard, 2016: The view from above: Applications of satellite data in economics. *J. Econ. Perspect.*, **30**, 171–198, doi:10.1257/jep.30.4.171.
- Fassò, A., R. Ignaccolo, F. Madonna, B. B. Demoz, and M. Franco-Villoria, 2014: Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles. *Atmos. Meas. Tech.*, **7**, 1803–1816, doi:10.5194/amt-7-1803-2014.
- Feldman, D. R., W. D. Collins, and J. L. Paige, 2015: Pan-spectral observing system simulation experiments of shortwave reflectance and longwave radiance for climate model evaluation. *Geosci. Model Dev.*, **8**, 1943–1954, doi:10.5194/gmd-8-1943-2015.
- Fischer, E. V., D. A. Jaffe, and E. C. Weatherhead, 2011: Free tropospheric peroxyacetyl nitrate (PAN) and ozone at Mount Bachelor: Potential causes of variability and timescale for trend detection. *Atmos. Chem. Phys.*, **11**, 5641–5654, doi:10.5194/acp-11-5641-2011.
- Genton, M. G., 1998: Highly robust variogram estimation. *Math. Geol.*, **30**, 213–221, doi:10.1023/A:1021728614555.
- Girz, C. M. I. R., A. E. MacDonald, F. Caracena, R. L. Anderson, T. Lachenmeier, B. D. Jamison, R. S. Collander, and E. C. Weatherhead, 2002: GAINS—A global observing system. *Adv. Space Res.*, **30**, 1343–1348, doi:10.1016/S0273-1177(02)00540-9.
- Hunt, F. K., J. J.-M. Hirschi, B. Sinha, K. Oliver, and N. Wells, 2013: Combining point correlation maps with self-organising maps to compare observed and simulated atmospheric teleconnection patterns. *Tellus*, **65A**, 20822, doi:10.3402/tellusa.v65i0.20822.
- Ignaccolo, R., M. Franco-Villoria, and A. Fassò, 2015: Modelling collocation uncertainty of 3D atmospheric profiles. *Stochastic Environ. Res. Risk Assess.*, **29**, 417–429, doi:10.1007/s00477-014-0890-7.
- Kint, V., M. Van Meirvenne, L. Nachtergale, G. Geudens, and N. Lust, 2003: Spatial methods for quantifying forest stand structure development: A comparison between nearest-neighbor indices and variogram analysis. *For. Sci.*, **49**, 36–49.
- Kreher, K., G. E. Bodeker, and M. Sigmond, 2015: An objective determination of optimal site locations for detecting expected trends in upper-air temperature and total column ozone. *Atmos. Chem. Phys.*, **15**, 7653–7665, doi:10.5194/acp-15-7653-2015.
- Ladstädter, F., A. K. Steiner, M. Schwarz, and G. Kirchengeast, 2015: Climate comparison of reference upper-air measurements: GPS radio occultation and GRUAN radiosondes. *European Geosciences Union General Assembly*, Vienna, Austria, 17.1918L.
- Lanzante, J. R., S. A. Klein, and D. J. Seidel, 2003: Temporal homogenization of monthly radiosonde temperature data. Part II: Trends, sensitivities, and MSU comparison. *J. Climate*, **16**, 241–262, doi:10.1175/1520-0442(2003)016<0241:THOMRT>2.0.CO;2.
- Laxminarayan R., and M. K. Macauley, Eds., 2012: *The Value of Information: Methodological Frontiers and New Applications in Environment and Health*. Springer, 304 pp., doi:10.1007/978-94-007-4839-2.
- Macauley, M. K., 2006a: Ascribing societal benefit to environmental observations of Earth from space: The Multiangle Imaging SpectroRadiometer (MISR). Resources for the Future Discussion Paper 06-09, 30 pp.
- , 2006b: The value of information: Measuring the contribution of space-derived Earth science data to resource management. *Space Policy*, **22**, 274–282, doi:10.1016/j.spacepol.2006.08.003.
- MacDonald, A. E., 2005: A global profiling system for improved weather and climate prediction. *Bull. Amer. Meteor. Soc.*, **86**, 1747–1764, doi:10.1175/BAMS-86-12-1747.

- Madonna, F., M. Rosoldi, J. Güldner, A. Haeefe, R. Kivi, M. P. Cadeddu, D. Sisterson, and G. Pappalardo, 2014: Quantifying the value of redundant measurements at GCOS Reference Upper-Air Network sites. *Atmos. Meas. Tech.*, **7**, 3813–3823, doi:[10.5194/amt-7-3813-2014](https://doi.org/10.5194/amt-7-3813-2014).
- McBratney, A. B., R. Webster, and T. M. Burgess, 1981: The design of optimal sampling schemes for local estimation and mapping of regionalized variables—I: Theory and method. *Comput. Geosci.*, **7**, 331–334, doi:[10.1016/0098-3004\(81\)90077-7](https://doi.org/10.1016/0098-3004(81)90077-7).
- Mears, C. A., M. C. Schabel, and F. J. Wentz, 2003: A reanalysis of the MSU channel 2 tropospheric temperature record. *J. Climate*, **16**, 3650–3664, doi:[10.1175/1520-0442\(2003\)016<3650:AROTMC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3650:AROTMC>2.0.CO;2).
- Messié, M., and F. Chavez, 2011: Global modes of sea surface temperature variability in relation to regional climate indices. *J. Climate*, **24**, 4314–4331, doi:[10.1175/2011JCLI3941.1](https://doi.org/10.1175/2011JCLI3941.1).
- Morss, R., K. A. Miller, and M. S. Vasil, 2005: A systematic economic approach to evaluating public investment in observations for weather forecasting. *Mon. Wea. Rev.*, **133**, 374–388, doi:[10.1175/MWR-2857.1](https://doi.org/10.1175/MWR-2857.1).
- Mountain Research Initiative Elevation-Dependent Warming Working Group, 2015: Elevation-dependent warming in mountain regions of the world. *Nat. Climate Change*, **5**, 424–430, doi:[10.1038/nclimate2563](https://doi.org/10.1038/nclimate2563).
- NOAA/National Climatic Data Center, 2010: The United States national plan for implementing the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) for FY2011 and beyond. NOAA/National Climatic Data Center Rep., 27 pp., [ftp://ftp.atdd.noaa.gov/pub/GCOS/Reference_Observing/US%20GRUAN%20Planning%20Document%20for%20FY11-15.pdf](http://ftp.atdd.noaa.gov/pub/GCOS/Reference_Observing/US%20GRUAN%20Planning%20Document%20for%20FY11-15.pdf).
- Nychka, D., and N. Saltzman, 1998: Design of air-quality monitoring networks. *Case Studies in Environmental Statistics*, Springer, 51–76, doi:[10.1007/978-1-4612-2226-2_4](https://doi.org/10.1007/978-1-4612-2226-2_4).
- Papritz, A., and H. Flüßler, 1994: Temporal change of spatially autocorrelated soil properties: Optimal estimation by cokriging. *Geoderma*, **62**, 29–43, doi:[10.1016/0016-7061\(94\)90026-4](https://doi.org/10.1016/0016-7061(94)90026-4).
- Pepin, N. C., and J. D. Lundquist, 2008: Temperature trends at high elevations: Patterns across the globe. *Geophys. Res. Lett.*, **35**, L14701, doi:[10.1029/2008GL034026](https://doi.org/10.1029/2008GL034026).
- Prusinkiewicz, P., and M. Hammel, 1993: A fractal model of mountains and rivers. *Proc. Graphics Interface '93*, Toronto, Ontario, Canada, Canadian Information Processing Society, 174–180.
- Royle, J. A., 2002: Exchange algorithms for constructing large spatial designs. *J. Stat. Plann. Inference*, **100**, 121–134, doi:[10.1016/S0378-3758\(01\)00127-6](https://doi.org/10.1016/S0378-3758(01)00127-6).
- , and D. Nychka, 1998: An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Comput. Geosci.*, **24**, 479–488, doi:[10.1016/S0098-3004\(98\)00020-X](https://doi.org/10.1016/S0098-3004(98)00020-X).
- Sairanen, H., M. Heinonen, and R. Högström, 2015: Validation of a calibration set-up for radiosondes to fulfil GRUAN requirements. *Meas. Sci. Tech.*, **26**, 105901, doi:[10.1088/0957-0233/26/10/105901](https://doi.org/10.1088/0957-0233/26/10/105901).
- Schrijver, A., 1998: *Theory of Linear and Integer Programming*. Wiley & Sons, 484 pp.
- Schutgens, N. A. J., E. Gryspeerdt, N. Weigum, S. Tsyro, D. Goto, M. Schulz, and P. Stier, 2016: Will a perfect model agree with perfect observations? The impact of spatial sampling. *Atmos. Chem. Phys.*, **16**, 6335–6353, doi:[10.5194/acp-16-6335-2016](https://doi.org/10.5194/acp-16-6335-2016).
- Silva, C., and A. Quiroz, 2003: Optimization of the atmospheric pollution monitoring network at Santiago de Chile. *Atmos. Environ.*, **37**, 2337–2345, doi:[10.1016/S1352-2310\(03\)00152-3](https://doi.org/10.1016/S1352-2310(03)00152-3).
- Sofen, E. D., D. Bowdalo, and M. J. Evans, 2016: How to most effectively expand the global surface ozone observing network. *Atmos. Chem. Phys.*, **16**, 1445–1457, doi:[10.5194/acp-16-1445-2016](https://doi.org/10.5194/acp-16-1445-2016).
- Spencer, R. W., and J. R. Christy, 1993: Precision lower stratospheric temperature monitoring with the MSU: Technique, validation, and results 1979–1991. *J. Climate*, **6**, 1194–1204, doi:[10.1175/1520-0442\(1993\)006<1194:PLSTMW>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1194:PLSTMW>2.0.CO;2).
- Strachan, S., and Coauthors, 2016: Filling the data gaps in mountain climate observatories through advanced technology, refined instrument siting, and a focus on gradients. *Mt. Res. Dev.*, **36**, 518–527, doi:[10.1659/MRD-JOURNAL-D-16-00028.1](https://doi.org/10.1659/MRD-JOURNAL-D-16-00028.1).
- Tan, M. H., 2013: Minimax designs for finite design regions. *Technometrics*, **55**, 346–358, doi:[10.1080/00401706.2013.804439](https://doi.org/10.1080/00401706.2013.804439).
- Todini, E., and M. Ferraresi, 1996: Influence of parameter estimation uncertainty in kriging. *J. Hydrol.*, **175**, 555–566, doi:[10.1016/S0022-1694\(96\)80024-2](https://doi.org/10.1016/S0022-1694(96)80024-2).
- van Groenigen, J. W., 2000: The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, **97**, 223–236, doi:[10.1016/S0016-7061\(00\)00040-9](https://doi.org/10.1016/S0016-7061(00)00040-9).
- Weatherhead, E. C., and A. R. Webb, 1997: International response to the challenge of measuring solar ultra-violet radiation. *Radiat. Prot. Dosim.*, **72**, 223–229, doi:[10.1093/oxfordjournals.rpd.a032094](https://doi.org/10.1093/oxfordjournals.rpd.a032094).
- , G. C. Tiao, G. C. Reinsel, J. E. Frederick, J. J. DeLuisi, D. Choi, and W.-K. Tam, 1997: Analysis of long-term behavior of ultra-violet radiation measured by Robertson–Berger meters at 14 sites in the United States. *J. Geophys. Res.*, **102**, 8737–8754, doi:[10.1029/96JD03590](https://doi.org/10.1029/96JD03590).
- , A. J. Stevermer, and B. E. Schwartz, 2002: Detecting environmental changes and trends. *Phys. Chem. Earth*, **27**, 399–403, doi:[10.1016/S1474-7065\(02\)00019-0](https://doi.org/10.1016/S1474-7065(02)00019-0).
- , and Coauthors, 2005: Ozone and ultraviolet radiation. *Arctic Climate Impact Assessment*, C. Symon, L. Arris, and B. Heal, Eds., Cambridge University Press, 151–182, <https://www.amap.no/documents/doc/arctic-arctic-climate-impact-assessment/796>.
- , and Coauthors, 2017: How long do satellites need to overlap? Evaluation of climate data stability from overlapping satellite records. *Atmos. Chem. Phys.*, doi:[10.5194/acp-2016-1156](https://doi.org/10.5194/acp-2016-1156), in press.
- Weaver, C. P., R. J. Lempert, C. Brown, J. A. Hall, D. Revell, and D. Sarewitz, 2013: Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks. *Wiley Interdiscip. Rev.: Climate Change*, **4**, 39–60, doi:[10.1002/wcc.202](https://doi.org/10.1002/wcc.202).
- Webster, R., P. J. Curran, and J. W. Munden, 1989: Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry. *Remote Sens. Environ.*, **29**, 67–78, doi:[10.1016/0034-4257\(89\)90079-5](https://doi.org/10.1016/0034-4257(89)90079-5).
- Whiteman, D. N., K. C. Vermeesch, L. D. Oman, and E. C. Weatherhead, 2011: The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor. *J. Geophys. Res.*, **116**, D21118, doi:[10.1029/2011JD016610](https://doi.org/10.1029/2011JD016610).
- Wikle, C. K., and J. A. Royle, 1999: Space–time dynamic design of environmental monitoring networks. *J. Agric. Biol. Environ. Stat.*, **4**, 489–507.
- Williamson, R. A., H. R. Hertzfeld, J. Cordes, and J. M. Logsdon, 2002: The socioeconomic benefits of Earth science and applications research: Reducing the risks and costs of natural disasters in the USA. *Space Policy*, **18**, 57–65, doi:[10.1016/S0265-9646\(01\)00057-1](https://doi.org/10.1016/S0265-9646(01)00057-1).
- WMO, 2015: Valuing weather and climate: Economic assessment of meteorological and hydrological services. WMO Rep. 1153,

- 308 pp., http://www.wmo.int/pages/prog/amp/pwsp/documents/wmo_1153_en.pdf.
- Wulfmeyer, V., and Coauthors, 2015: A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles. *Rev. Geophys.*, **53**, 819–895, doi:[10.1002/2014RG000476](https://doi.org/10.1002/2014RG000476).
- Yorgun, M. S., and R. B. Rood, 2015: An object-based approach for quantification of GCM biases of the simulation of orographic precipitation. Part II: Quantitative analysis. *J. Climate*, **28**, 4863–4876, doi:[10.1175/JCLI-D-14-00730.1](https://doi.org/10.1175/JCLI-D-14-00730.1).
- Yost, R. S., G. Uehara, and R. L. Fox, 1982: Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. *Soil Sci. Soc. Amer. J.*, **46**, 1028–1032, doi:[10.2136/sssaj1982.03615995004600050028x](https://doi.org/10.2136/sssaj1982.03615995004600050028x).
- Zou, C.-Z., W. Wang, J. Li, and NOAA Climate Data Record Program, 2015: NOAA Climate Data Record of Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU-A) mean layer temperature, version 3.0. NOAA/National Centers for Environmental Information. Subset used: upper troposphere and lower stratosphere, accessed 15 December 2016, doi:[10.7289/V5QF8QTK](https://doi.org/10.7289/V5QF8QTK).