Received: 13 August 2021 | Revised: 24 November 2021 | Accepted: 30 November 2021 | Published on: 10 January 2022

DOI: 10.1002/qj.4226

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Assessing observation network design predictions for monitoring Antarctic surface temperature

Robert Tardif[1] | **Gregory J. Hakim**[1] | **Karin A. Bumbaco**[2] | **Matthew A. Lazzara**[3,4] | **Kevin W. Manning**[5] | **David E. Mikolajczyk**[3] | **Jordan G. Powers**[5]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, Washington

[2]Cooperative Institute for Climate, Ocean and Ecosystem Studies, University of Washington, Seattle, Washington

[3]Antarctic Meteorological Research Center, Space Science and Engineering Center, University of Wisconsin–Madison, Madison, Wisconsin

[4]Department of Physical Sciences, School of Engineering, Sciences, and Mathematics, Madison Area Technical College, Madison, Wisconsin

[5]National Center for Atmospheric Research, Boulder, Colorado

Correspondence
R. Tardif, Vaisala Inc., Louisville, CO 80027, USA.
Email: robert.tardif@vaisala.com

Present Address
Robert Tardif, Vaisala Inc., Louisville, Colorado

Funding information
National Science Foundation, Office of Polar Programs (Grant 1542766 to the University of Washington, Grants 1542789, 1543305, 1924730 to the University of Wisconsin–Madison).

Abstract

Networks of observations ideally provide adequate sampling of parameters to be monitored for climate and weather forecasting applications. This is a challenge for any network, but is particularly difficult in the harsh environment of the Antarctic continent. We evaluate a network design method providing objective information on station siting for optimal sampling of a variable, here taken to be surface air temperature. The method uses the concept of ensemble sensitivity to predict locations reducing the most total ensemble variance, that is, uncertainty, across the continent. The method is applied to a network of frequently-reporting stations, and validation is performed using results from assimilating station observations. A cost-efficient "offline" data assimilation framework is used to allow testing over a large sample of experiments, including a large number of randomly chosen networks that serve as a null hypothesis. Network design predictions agree well with observed error reductions from assimilation. The important role of stations on the East Antarctic Plateau in monitoring surface air temperature is evident in network design and data assimilation results, followed by stations in West Antarctica and the Ross Ice Shelf region. Antarctic coastal and Peninsula stations are found to provide the smallest information content integrated over the continent. Validation results are also robust to covariance localization, an essential factor for ensemble methods. Optimal networks outperform randomly chosen-networks in all cases, by up to nearly 50%, depending on the size of the network and the covariance localization distance.

KEYWORDS

Antarctica, data assimilation, network design, optimization, surface observations, validation

1 | INTRODUCTION

Observing networks are critical components of the infrastructure established to monitor and predict the evolution of the Earth system. A large number of observations are taken to support a wide variety of activities, from the long-term monitoring of key climate variables to the production of accurate weather forecasts. Decisions on which observation platforms to deploy and where to deploy them depend on the specific objectives pursued,

ranging in scope from investigating important scientific questions about the climate system (e.g., Weatherhead *et al.*, 2018) to practical issues such as ensuring safe aircraft operations at airfields. Typically, these decisions are made subjectively, or dictated by practical considerations (e.g., cost) and/or operational needs (e.g., observing visibility at airfields). Despite the significant resources devoted to the global observing system, gaps in direct observations remain, not only in the oceans (e.g., Fujii *et al.*, 2019), but also in the atmosphere, particularly over the polar regions (Jung *et al.*, 2016). The rising costs of operating extensive networks of observations also pose a significant challenge in maintaining an infrastructure that best supports the growing need for environmental observations. We posit that an objective network-design approach can provide useful information for the optimal and cost-efficient deployment of assets. Here we test the skill of a network design algorithm in predicting the impact of actual observations using observing-system experiments.

Optimal network design is a methodology developed to identify objectively those observations whose inclusion results in the largest change in a measurable objective, defined by a metric. In the atmospheric sciences, the method originated in the adaptive observation strategy (e.g., Snyder, 1996; Morss *et al.*, 2001), aimed at reducing numerical weather forecast errors. Adjoint (e.g., Langland and Baker, 2004) and ensemble (e.g., Bishop *et al.*, 2001) implementations were developed to identify the locations of input observations providing the largest improvement in weather forecasts. These methods have been tested in the context of targeted observations in the Fronts and Atlantic Storm Track Experiment (FASTEX: e.g., Bergot, 1999; Szunyogh *et al.*, 1999), the North Pacific Experiment (NORPEX-98: e.g., Langland *et al.*, 1999), the Winter Storm Reconnaissance (WSR) program (e.g., Hamill *et al.*, 2013), and the Observing System Research and Predictability Experiment (THORPEX) program (see Majumdar, 2016, for a review). Similar approaches have been proposed for the design of fixed networks of observations (e.g., Khare and Anderson, 2006), which is the context we consider here.

One particularly challenging environment for weather observations is the remote Antarctic continent, the setting of a spectrum of demanding scientific investigations and the logistical operations needed to support them (Powers *et al.*, 2012). Routine weather observations in Antarctica were first taken at staffed stations during the 1957–1958 International Geophysical Year (Odishaw, 1959), subsequently augmented with automated weather stations (AWS) able to withstand harsh weather conditions (Lazzara *et al.*, 2012). Despite the increase in the number of stations, gaps in the network remain, particularly in the interior of the continent (e.g., Chapman and Walsh, 2007; Lazzara *et al.*, 2012; Bumbaco *et al.*, 2014), leading to incomplete characterization of extreme conditions, long-term trends (e.g., Wei *et al.*, 2019; Clem *et al.*, 2020), and variability at the mesoscale (e.g., Gonzalez *et al.*, 2021). These studies underline the need for a more effective network. From the perspective of weather forecasting, the value of additional observations has already been demonstrated using data from special field campaigns (see, e.g., Rabier *et al.*, 2013; Bromwich *et al.*, 2020), further motivating efforts toward the optimization of Antarctica's observing networks.

An ensemble observational network design method has been proposed and applied to the monitoring of surface air temperature by Hakim *et al.* (2020, H20 hereafter). The method uses the concept of ensemble sensitivity (Ancell and Hakim, 2007; Torn and Hakim, 2008) to determine optimal sensor placement, by using ensembles composed of samples from model simulations (here surface air temperature) and finding the location where an observation reduces the variance (i.e., uncertainty) in a chosen metric the most. Earlier implementations of this approach have been applied to a climate observing network for the Pacific Northwest by Mauger *et al.* (2013), and in an example application in Bumbaco *et al.* (2014) for the monitoring of regionally-averaged surface temperature in Antarctica. H20 have applied the technique to identify locations that maximize temperature information, either from the hypothetical perspective of a continent without observations (a "blank slate") or to complement an existing network of surface stations. Their study has identified the Megadunes region of the Antarctic Plateau as the most important for monitoring temperature and reducing model forecast errors in the blank-slate network scenario, while Queen Maud Land, Ellsworth Land, and the Adélie Coast are regions not well observed by an existing network of stations frequently reporting valid observations. However, these network design results are predictions, in that actual observations are not used in determining the optimal monitoring locations. Therefore, validation of the results is required to test the predictions. This is performed here by comparing predicted reductions in metric variance associated with a number of locations, corresponding to a network of existing stations, with reductions in ensemble-mean analysis errors from the assimilation of actual observations from the same network.

The remainder of this article is organized as follows. Section 2 summarizes the network design method, the data assimilation experiments used in the validation, and the observational data. In Section 3, we present results of the network design applied to the locations of an existing network and describe comparisons with error reduction statistics diagnosed in our data assimilation experiments. Conclusions are provided in Section 4.

2 | METHODS AND DATA

2.1 | Network design method

The network design (ND) method described in H20 is based on a generalization of the ensemble sensitivity approach (see Ancell and Hakim, 2007; Torn and Hakim, 2008), involving the sensitivity of a multivariate metric $\mathbf{J}$ of dimension n to a state vector $\mathbf{x}$ of dimension m. The sensitivity of the metric is estimated using ensemble samples for $\mathbf{J}$ and $\mathbf{x}$ taken from a numerical model (see Section 2.3). The covariance of the metric $\mathbf{J}$ about the mean, $\mathbf{\Sigma}^2 = \{\delta\mathbf{J}\delta\mathbf{J}^{\mathrm{T}}\}$, where $\{\}$ denotes an expectation operator that has input from the finite ensemble samples and superscript T represents the transpose operation, can be approximated as

$$\mathbf{\Sigma}^2 \approx \left[\frac{\partial \mathbf{J}}{\partial \mathbf{x}}\right]^{\mathrm{T}} \{\delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}}\} \left[\frac{\partial \mathbf{J}}{\partial \mathbf{x}}\right], \tag{1}$$

where $\partial\mathbf{J}/\partial\mathbf{x}$ is the sensitivity of the metric $\mathbf{J}$ to the state vector $\mathbf{x}$. The change in the metric covariance, expressed as $\delta\mathbf{\Sigma}^2$, for a single measurement i can be expressed as

$$\delta\mathbf{\Sigma}_i^2 = -E^{-1}\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}\left[\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{2}$$

where $\delta\mathbf{J}$ is the deviation in metric $\mathbf{J}$ about the mean and $\mathbf{H}\delta\mathbf{x}$ is the deviation from the mean of the state vector $\mathbf{x}$ projected in observation space through the forward operator $\mathbf{H}$. We see that Equation 2 expresses the squared covariance between the state at the potential observation location i and the metric $\mathbf{J}$, weighted by scalar $E = var(\mathbf{H}\delta\mathbf{x}_i) + r_i$, where r_i is the element of the observation-error covariance matrix $\mathbf{R}$ corresponding to measurement i. For the multivariate $\mathbf{J}$ used here, temperature values sampled on a grid across the Antarctic continent (defined in Section 2.3), we seek the largest negative value of the trace of the $\delta\mathbf{\Sigma}^2$ $n \times n$ matrix, as this scalar value would represent the largest cumulative reduction in the variance of $\mathbf{J}$. This identifies the location of the observation giving the greatest impact (in terms of minimizing the variance of $\mathbf{J}$) for monitoring continent-wide temperatures. We note that other choices can be made for summarizing $\mathbf{J}$ in terms of a scalar, and that this effectively reduces the calculation back to the traditional scalar form (e.g., Majumdar, 2016).

Once the most impactful observation is identified, the ensemble Kalman filter update step is used to revise $\delta\mathbf{x}$ and $\delta\mathbf{J}$ to reflect the impact of the selected measurement. Equation 2 is then evaluated with the updated matrices to find the next most impactful measurement, reflecting its particular location, conditional on the previous selected location. This process is repeated until all the desired observation locations have been considered. To account for sample error, a Monte Carlo (MC) approach is applied in which the above-described process is repeated 10,000 times, with different random draws of the model temperatures that form the $\mathbf{J}$ and $\mathbf{x}$ ensembles. H20 provide a more detailed derivation of Equation 2 and description of the algorithm. We note that this algorithm only requires knowledge of observation errors, not observed values, and thus can be used in advance to predict the impact of hypothetical observations.

In Equation 2, the central component of the numerics of the network design method is $\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}$, the covariance between metric $\mathbf{J}$ and single observation $(\mathbf{H}\delta\mathbf{x})_i$. Since we estimate this covariance with an ensemble sample, covariance localization must be considered, as it is essential to all implementations of ensemble techniques. Spatial localization is based on the assumption that covariability between variables decreases over physical distance, so that noise dominates the signal at large distances. This noise can be mitigated by applying a smooth localizing function to the estimated covariances to reduce long-distance covariance estimates (see Houtekamer and Zhang, 2016). In this study, localization is applied directly to Equation 2 as follows:

$$\delta\mathbf{\Sigma}_i^2 = -E^{-1}w_{\mathrm{loc}} \circ \left[\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}\left[\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}\right]^{\mathrm{T}}\right], \tag{3}$$

where w_{loc} represents the localization weights determined from the distances between elements in the metric vector $\mathbf{J}$ and state variables in $\mathbf{x}$, and $\circ$ denotes the element-wise product. We use the Gaspari–Cohn fifth-order piecewise localization function (Gaspari and Cohn, 1999) to define w_{loc}. This implementation differs from H20, where the Gaspari–Cohn function is applied to the covariance itself (i.e., $w_{\mathrm{loc}} \circ \delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})_i^{\mathrm{T}}$), leading to a $(w_{\mathrm{loc}})^2$ dependence in the predicted change in variance $\delta\mathbf{\Sigma}_i^2$ (see Equation 2). Localization is ad hoc, as it is not part of the Kalman filter theory, so this is essentially an empirical choice, made here in favor of Equation 3, which provides better results (see Section 3.1.2).

Our goal is to test the predictions of the ND method with actual observations. Specifically, the ND method is applied here to the subset of locations corresponding to the existing stations that reported temperature on at least 90% of the days during the period of interest, as identified by Bumbaco *et al.* (2014). This application of the method differs from H20, where all possible locations in a theoretical observation network covering the entire continent are considered. We will refer to the 18 selected stations as "CD90", for 90% "complete data". The CD90 stations are listed in Table 1 and their locations shown in Figure 1.

The ND method is first applied to the CD90 locations individually to evaluate their impact as "stand-alone"

TABLE 1 List and coordinates of the CD90 stations during the 2008–2012 study period, including staffed and automatic weather stations (AWS)

Station name	Latitude	Longitude	Elevation (m)	Type
Amundsen–Scott	90°S	0°E	2,800	Staffed
Bellingshausen	62.2°S	58.9°W	49	Staffed
Cape Ross*	76.7°S	163.0°E	150	AWS
Casey	66.3°S	110.5°E	50	Staffed
Davis	68.6°S	78.0°E	13	Staffed
Dome A	80.4° S	77.4° E	4,084	AWS
Elizabeth	82.6°S	137.1°W	519	AWS
Ferrell	77.8° S	170.8° E	45	AWS
Halley	75.6°S	26.7°W	30	Staffed
Harry	83.0°S	121.4°W	945	AWS
Marble Point	77.4°S	163.8°E	108	AWS
Marilyn	79.9°S	165.5°E	63	AWS
Mawson	67.6°S	62.9°E	16	Staffed
Mirnyj	66.6°S	93.0°E	30	Staffed
Neumayer	70.7°S	8.3°W	40	Staffed
Rothera	67.6°S	68.1°W	15	Staffed
Syowa	69.0°S	39.6°E	21	Staffed
Vostok	78.5°S	106.9°E	3,420	Staffed

* Also known as Arelis.

stations for monitoring continental temperature over Antarctica. For this test, Equation 2 (or Equation 3 when localization is applied) is evaluated at each CD90 location, and stations are ranked by the resulting reductions in the metric variance. A second series of results is generated by applying the ND method to the entire CD90 network. Similarly to H20, the most impactful station is then identified conditional on the information (i.e., reduction in variance) from the previous observation. This process is repeated until all CD90 stations have been considered, resulting in a rank ordering of importance in the chosen network.

2.2 | Data assimilation experiments

The assessment of ND results is performed by comparing the predicted reductions in metric variance with the actual reductions of ensemble-mean errors in analyses of surface air temperature that involve the assimilation of Antarctic temperature observations using the ensemble square-root Kalman filter (EnSRF) approach of Whitaker and Hamill (2002). In essence, it is the agreement of the Kalman update equations for the ensemble mean (update from observations) and ensemble variance (the basis of ND predictions). While they should agree in theory, in practice many factors complicate potential agreement, such as observation error, sample error, and covariance localization. As a result, numerical verification is needed.

Here, data assimilation (DA) is performed with an "offline" application of the Kalman filter (see Evensen, 2003), a method used for paleoclimate reconstructions (e.g., Hakim *et al.*, 2016), ocean DA (e.g., Oke *et al.*, 2002) and, more closely related to our work, the generation of land-surface analyses (e.g., Devers *et al.*, 2020). This is a cost-efficient alternative to full observation-system experiments (OSE) with an operational data assimilation system. In an offline DA application, the prior ensemble is composed of states taken at random times from a single model simulation or from an ensemble of model simulations, and the same prior ensemble is used at all analysis times. This prior ensemble samples the model climatology and is "uninformed" with respect to temperature across the Antarctic continent at any particular time, thus forming a reference that is equivalent to the prior ensemble used in the ND algorithm. The assimilation is performed in anomaly space (i.e., deviations from the temporal mean are used as prior states and also for observations) to minimize the impact of possible biases in the prior. As in H20,

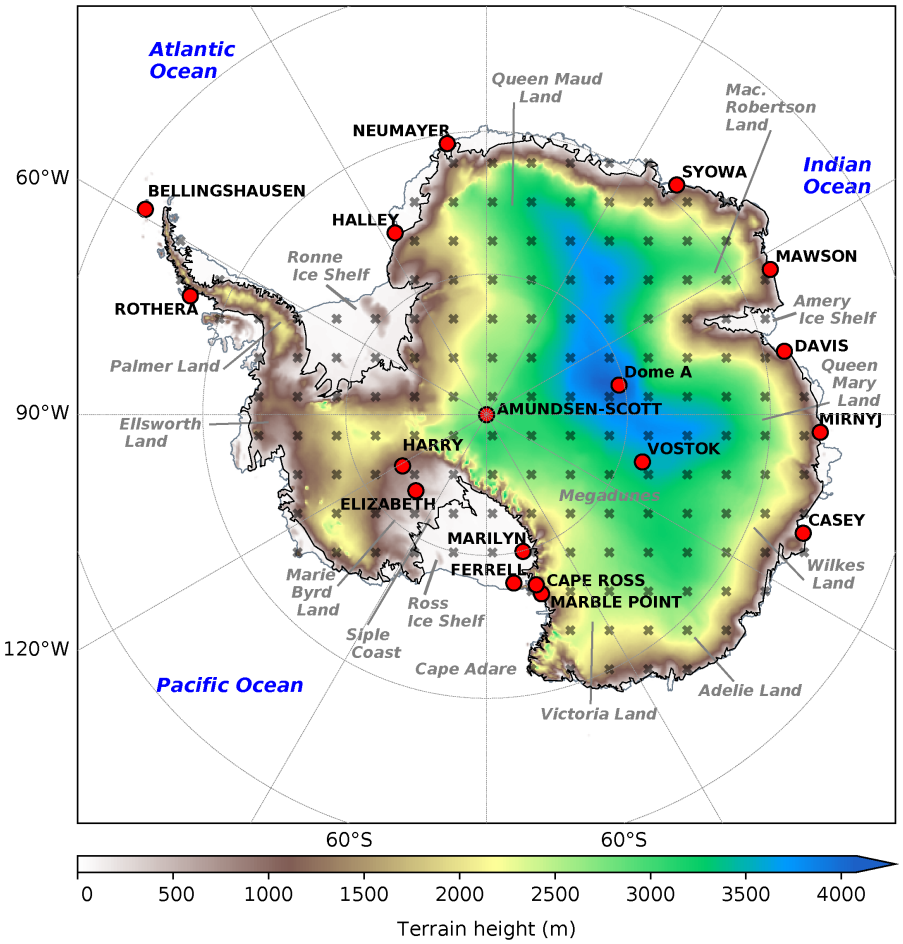

FIGURE 1 Locations of the CD90 stations, topography of the Antarctic continent, and regions mentioned in the text. The **X** symbols on the map define the ND metric grid points, and the locations where mean squared error reductions in offline analyses are evaluated [Colour figure can be viewed at wileyonlinelibrary.com]

we use an ensemble size of 250 members, and uncertainties in resulting analyses are estimated through 100 MC realizations, each using a different set of randomly-drawn states populating the prior ensemble. Analyses are generated daily at 0000 UTC over the period of October 1, 2008–October 31, 2012, representing a total of 149,200 analyses (i.e., daily over 49 months and 100 MC realizations), which is not feasible with an online DA system. The chosen period is one during which the configuration of the operational Antarctic Mesoscale Prediction System (AMPS: Powers *et al.*, 2003), used as the source of prior data in ND calculations and in our offline DA experiments (see Section 2.3), was stable (e.g., grid sizes remained unchanged). This is an important condition to consider for our experiments, as modifications to the system could lead to changes in the spatial covariances characterizing the modelled temperature field.

For comparison with ND results where stations are considered individually, a series of DA experiments is carried out, in which observations from a single station are assimilated, producing 18 sets of analyses, one for each of the CD90 stations. Resulting analysis errors are compared against the network design estimates (see Section 3.1). Another set of data assimilation experiments is then performed by sequentially assimilating the observations from all of the CD90 stations, and results are compared with the conditional ranking of stations from the second series of ND calculations (see Section 3.2). The evaluation of ND predictions is performed by comparing reductions of mean squared errors of the target analysis quantity (e.g., surface air temperature) in the offline analyses with the reductions in metric variance obtained from the ND algorithm. A good correspondence between error reductions from the assimilation of actual Antarctic observations and ND results validates the predicted optimal placement of observations.

Errors in our surface air-temperature analyses are evaluated with respect to an external gridded reanalysis product, which allows for measuring errors throughout the continent on a daily time-scale. We compare predictions of changes in error variance using the mean squared error (MSE), calculated as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(T_{\text{offl}} - T_{\text{RNL}})^2, \tag{4}$$

where T_{offl} represents the ensemble-mean temperature anomalies from our DA experiments (either based on the

analysis or the prior field) and T_{RNL} represents the temperature anomalies from the reanalysis used as the reference. Anomalies are obtained by removing the mean over the period of interest at every grid point. N is the number of data points along the time dimension. MSE values are calculated at every analysis grid point over the continent. Reductions in errors are determined by taking the difference between the MSE in the posterior field (i.e., the analyses) and the uninformed prior $\Delta MSE = MSE^{\text{posterior}} - MSE^{\text{prior}}$, summing over the n grid points defining the multivariate metric used in our ND calculations and then normalizing by the sum of the errors of the prior:

$$S_{DA} = \frac{\sum_{i=1}^{n} \Delta MSE_i}{\sum_{i=1}^{n} MSE_i^{\text{prior}}}. \quad (5)$$

S_{DA} is then compared with the reduction in the metric variance obtained from the ND calculations normalized by the total initial variance:

$$S_{ND} = \frac{Tr\left(\delta \mathbf{\Sigma}^2\right)}{Tr\left(\mathbf{\Sigma}_0^2\right)}, \quad (6)$$

where $\mathbf{\Sigma}_0^2$ is the initial covariance matrix in ND calculations.

We emphasize that S_{DA} and S_{ND} are not necessarily expected to behave in similar ways as station data are processed. Most importantly, a property of ensemble data assimilation is that it invariably leads to a reduction in ensemble variance as observations are assimilated (see, e.g., Houtekamer and Zhang, 2016), whereas a decrease in errors in the ensemble mean is not guaranteed for all analysis grid points as a result of observation assimilation. A good correspondence between the variance-based S_{ND} metric and S_{DA} summarizing ensemble-mean errors indicates a consistent choice for the observation-error variance as well as an accurate representation of prior covariances. The latter would be particularly valuable, as ensemble-mean errors are evaluated with respect to an independent reference reanalysis, itself based on a different numerical model and a much larger number of assimilated observations.

2.3 | Data

Output from the Weather Research and Forecasting (WRF) model (Skamarock *et al.*, 2019) that was running in the AMPS (Powers *et al.*, 2003; 2012) is used as input in the ND calculations, as described in H20. State $\mathbf{x}$ and metric $\mathbf{J}$ are derived from the 0000 UTC analyses of surface air temperature over the continent from the WRF 15-km grid at every fifth and 20th grid point, respectively. Reductions in metric variance (Section 2.1) are therefore evaluated over 155 grid points across the continent (see Figure 1). The 15-km gridded data are preferred over another, lower resolution WRF grid running in AMPS (namely, 45-km spacing), due to the improved accuracy in the 15-km simulations (see, e.g., Bromwich *et al.*, 2013; Bozkurt *et al.*, 2020) and because this 15-km grid still covers the entire continent. Gridded WRF surface air-temperature analyses are also the source data for the prior ensembles used in offline data assimilation. As in the ND calculations, the prior is composed of gridded surface air-temperature data at every fifth grid point on the WRF grid. Observations are assimilated at 0000 UTC daily during the October 1, 2008–October 31, 2012 period.

Quality-controlled Antarctic observations from the archives of the National Centers for Environmental Information (NCEI) Integrated Surface Database (Smith *et al.*, 2011)[1] are assimilated in our experiments. Those observations taken at 0000 UTC are used, and deviations from the temporal mean over the entire period of interest are assimilated. Observation-error variance is taken uniformly across stations and set to $4\,K^2$ as in the ND calculations. This represents an increase by roughly a factor of two from the value typically used in data assimilation systems (see, e.g., Ha and Snyder, 2014), reflecting the added challenges in obtaining accurate temperature observations in Antarctica's harsh environment (see, e.g., Lazzara *et al.*, 2012, and references therein).

Reanalysis surface air temperatures are taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5) dataset (Hersbach *et al.*, 2020), available on a 31-km grid. Data from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2: Gelaro *et al.*, 2017) and Japanese 55-year Reanalysis (JRA-55: Kobayashi *et al.*, 2015) are also considered, as available reanalyses are characterized by significant uncertainties in their representation of surface conditions over Antarctica. For example, reanalysis biases in surface air temperature have been identified over the Antarctic Plateau, particularly a warm bias in the winter season (e.g., Bracegirdle and Marshall, 2012; Fréville *et al.*, 2014; Jones and Lister, 2015). While these errors are present to an extent in ERA5 (Gossart *et al.*, 2019), studies show that ERA5 tends to perform best (e.g., Gonzalez *et al.*, 2021; Zhu *et al.*, 2021). We therefore use ERA5 as the primary source for comparison with our offline analyses, while results based on MERRA2 and JRA55, shown in Supplementary Appendix S1, are used to assess the sensitivity of our results to the reference dataset. Reanalysis temperatures are interpolated to the offline analysis grid using nearest-neighbor interpolation.

[1] https://www.ncdc.noaa.gov/isd/data-access

3 | RESULTS

The network design method is tested from two perspectives. We first examine the CD90 stations individually to compare the predicted and actual impact of each station on the continent-wide temperature analysis. We follow this assessment with the more complex analysis of CD90 stations as rank-ordered contributors to a conditionally-sampled observing network. That is, we evaluate the information provided by stations conditional on the others in rank order, so that more impactful stations have priority. This more demanding test reveals the ability of the network-design predictions to provide quantitative guidance on the relative importance of observations for monitoring continent-wide Antarctic temperatures.

3.1 | Single station observations

In this section, we present and validate results from the calculations of the change in metric variance (Equation 2) on the 18 CD90 locations taken individually, without taking into account the other 17 stations. That is, the extent to which single station observations inform on temperatures over the continent is quantified objectively, and stations are ranked accordingly. This ranking is determined to a significant extent by the spatial correlations characterizing the temperature field around each station (H20). The spatial extent of that footprint determines the effectiveness of each station in providing information on temperatures over surrounding regions of the continent. The accurate estimation of spatial covariances is therefore germane to ND, prompting us also to consider the role of covariance localization.

3.1.1 | Network design applied to CD90 stations

Figure 2 shows the CD90 stations ranked by their predicted reductions in metric variance. Without covariance localization (Figure 2a), Vostok is chosen as the most impactful location nearly 100% of the time in the 10,000 MC realizations, while Syowa, Rothera, and Bellingshausen are consistently ranked with the least information on continental temperatures. The median total reduction in metric variance ranges from about 52% (5,800 K^2) of the total prior variance for Vostok to below 20% of the variance (less than 2,000 K^2) for the two lowest-ranking stations, Rothera and Bellingshausen (Figure 2b). Amundsen–Scott, Dome A, and Ferrell rank as the next most impactful locations, with Amundsen–Scott ranking second most often, and Dome A chosen in some MC realizations. Ferrell also ranks second, but in less than 10% of the MC realizations. It most often ranks as the fourth most influential station. These three stations, with variance reductions in the 45% range (5,000–5,300 K^2), are followed by Cape Ross and Marilyn, ranking equally likely as the fifth and sixth locations. Marble Point and Elizabeth are ranked with greater confidence as the seventh and eight locations, respectively, all with variance reductions remaining above 35% (4,000 K^2). Apart from the three lowest-ranking stations, greater uncertainty characterizes the order of stations among the lower rankings. The coastal stations of Mawson, Neumayer, Halley, Mirnyj, Casey, and Davis are generally ranked between ninth and 15th, with varying orderings possible within the uncertainty, with a number of stations showing a range of frequencies for any given rank (spread along the rank axis). For example, five locations are selected in 10% or more of the MC realizations as the 11th or 12th stations. Similarly, Halley is chosen with similar frequencies between the 10th and 14th spot. All of these locations have variance reductions between 25% and 35% (3,000–4,000 K^2). A notable shift toward smaller variance reductions is observed for the three lowest-ranking stations, particularly for the two stations located on the Antarctic Peninsula, Rothera and Bellingshausen.

The ND experiments in H20 show that different regions are highlighted when the influence of observations is localized in space to mitigate the effects of spurious spatial covariances from the limited number of samples used in our ensemble estimates. To measure the sensitivity to covariance localization, we calculate Equation 3 with different cut-off localization length-scales. Two experiments are reported here, the first performed with a cut-off length-scale of 3,000 km (Figures 2c,d), a value corresponding to the average correlation *e*-folding length-scale found in observations by Bumbaco *et al.* (2014), and a second using a more restrictive 1,000-km length-scale (Figure 2e,f).

With a cut-off distance of 3,000 km, Vostok ranks decisively as the most impactful location once again, while Marilyn ranks fourth with high confidence, a higher rank than without localization. Elizabeth also ranks higher, between the fifth and eighth position. The opposite effect is observed for Cape Ross, Mirnyj, Neumayer, and Mawson, now ranked lower in the list. The sharp contrast in variance reduction between continental and coastal locations is again present, while Peninsula stations again exhibit the least amount of variance explained.

For the 1,000-km length-scale, the reductions in metric variance are more than an order of magnitude smaller than without localization, as the spatial extent of stations' influence is constrained to much smaller scales. Elizabeth compares closely with Vostok as the most impactful station. Constraining observation impact to shorter scales

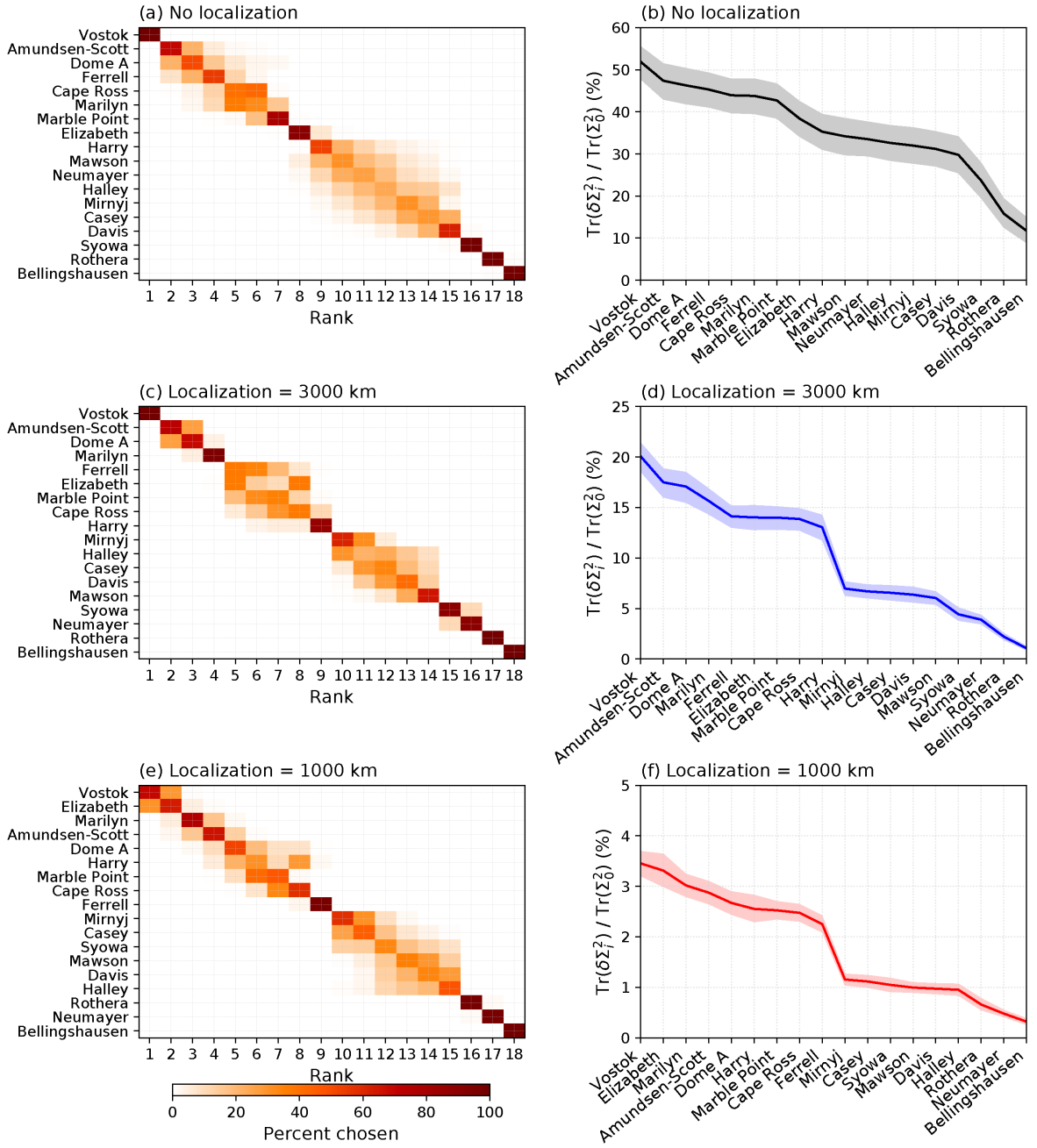

FIGURE 2 (a, c, e) Percentage of samples in which CD90 stations are chosen in each rank in the Monte Carlo realizations (color bar), and (b, d, f) the corresponding distributions of the change in metric variance, ranked from the highest to lowest median values (solid lines). Panels (a) and (b) show results without covariance localization, (c) and (d) with covariance localization with a cut-off distance of 3,000 km, (e) and (f) with covariance localization with a cut-off distance of 1,000 km. Distributions are represented by the median (solid lines) and 5th and 95th percentiles (shading). Notice the different ranges of values on the *y*-axes among plots in the right column, and station rankings depending on covariance localization parameters [Colour figure can be viewed at wileyonlinelibrary.com]

also elevates the importance of the Marie Byrd Land and Siple Coast regions, which are characterized by larger temperature variability and shorter correlation length-scales (see Figure 3 in H20). A sharp contrast between continental and coastal stations is again evident.

The dependence of station rankings on localization distance suggests the possible influence of how long-range covariances, which are particularly sensitive to sampling error, are estimated. The localization distance influences how much long-range covariance is captured in the ranking process. Figure 3 shows the footprint characterizing the variance reductions from each station predicted without covariance localization. Other than the more extensive areas with the largest variance reductions that characterize

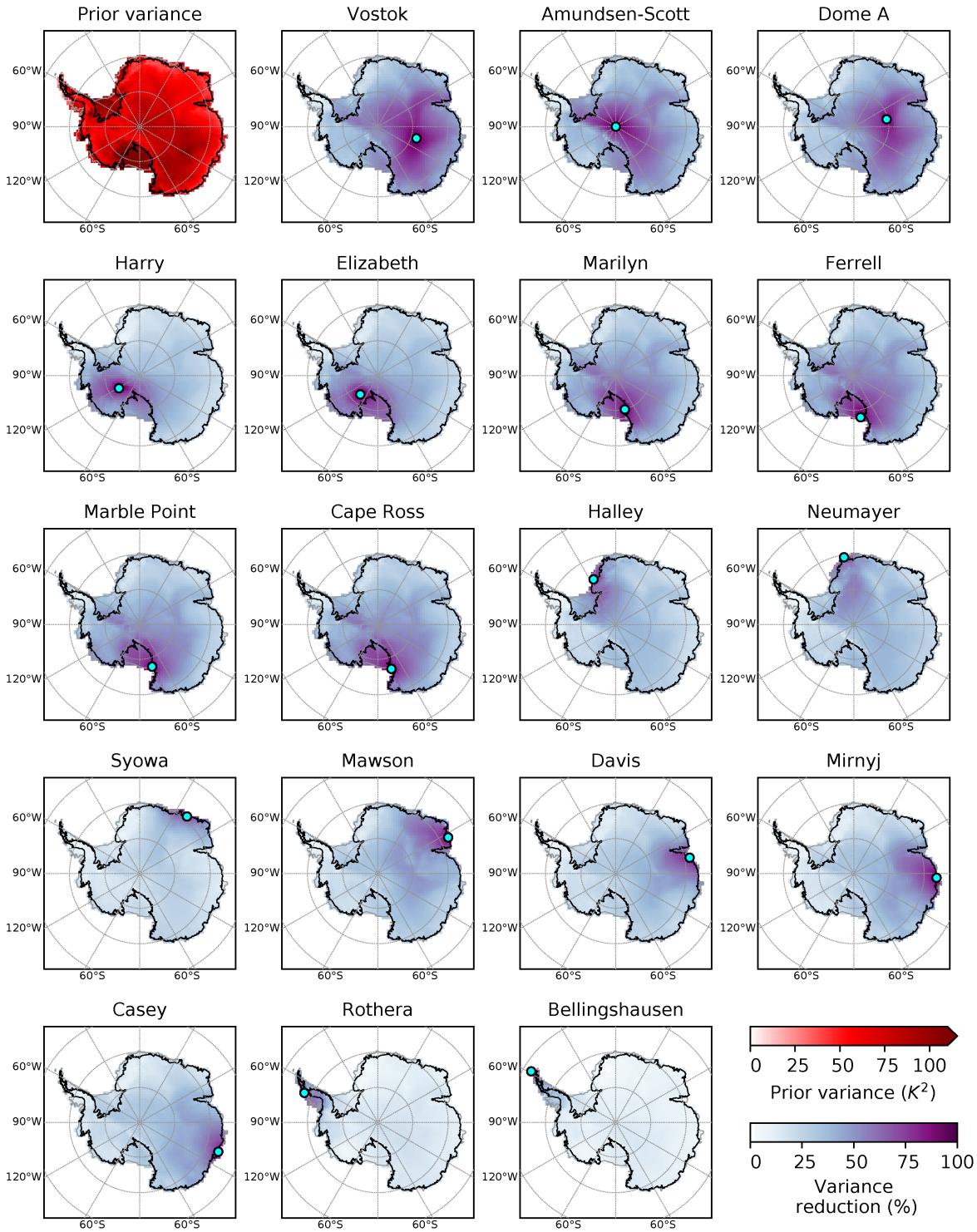

FIGURE 3 Spatial distributions of (upper-left frame) the median of the ensemble prior variance, and relative changes in median variance (in % with respect to the prior variance, all other frames) predicted with the ND method for each individual station (locations shown by the dots), without the use of covariance localization [Colour figure can be viewed at wileyonlinelibrary.com]

the top stations, we note that most stations produce sizable reductions across long distances over the continent. Exceptions are Rothera and Bellingshausen, the impacts of which are mainly limited to the Antarctic Peninsula, explaining why they rank the lowest among the CD90 stations. That the observations from most stations appear to provide information on temperatures across the continent is unrealistic. Examples are Ferrell, Cape Ross, and Marble Point, which are associated with variance reductions in excess of 20% over Mac. Robertson Land, west

of the Amery Ice Shelf on the opposite side of the continent. Use of the Gaspari–Cohn localization function with a realistic cut-off distance of 3,000 km (see Figure S1 in Appendix S1), however, effectively eliminates these remote and spurious numerical influences.

We also note similarities between footprints from various stations. The closely spaced stations Marilyn, Ferrell, Cape Ross, and Marble Point all project predominantly on areas of the Ross Ice Shelf and Victoria Land (see Figures 3 and S1). The similar footprints lead to similar changes in metric variance, and hence rankings, for these stations, as evidenced by the relatively flat portions of the curves corresponding to these stations in Figures 2b,d. A significant overlap in the spatial patterns from Dome A and Vostok is also observed, specifically over the East Antarctic Ice Sheet and extending to Mac. Robertson Land. These similarities will be referenced in Section 3.2 to explain some characteristics in the conditional ranking of stations. Elizabeth and Harry also share similar spatial footprints, although Elizabeth projects more strongly on the high-variance areas of the Siple Coast and Ross Ice Shelf. In contrast, coastal stations are shown to have mostly complementary influences over the outer areas of the continent.

Although stations are considered here as hypothetical stand-alone measurement sites and not as a network of locations complementing each other (see Section 3.2 for these results), the results are broadly consistent with those presented in H20. In particular, the primary importance of locations on the East Antarctic Plateau in monitoring temperatures at the continental scale is highlighted, complemented by locations in West Antarctica.

3.1.2 | Comparison with single-station offline DA

Here we assess ND predictions of observation impact through comparisons with error reductions from the assimilation of observations. We evaluate CD90 stations individually through 18 sets of analyses, each produced by assimilating observations from a single station. Results are compared in Figure 4 by plotting the summary diagnostics S_{DA} versus S_{ND} (see Equations 5 and 6).

A strong correlation ($r = 0.98$) is obtained between predicted (S_{ND}) and diagnosed (S_{DA}) observation impacts when covariance localization is not applied (Figure 4a). The ranking of stations between ND and DA is generally consistent, particularly when groups of stations are considered. For example, a group located on the East Antarctic Plateau ranks as the set of most impactful stations, followed by those stations located on or near the Ross Ice Shelf. Coastal and West Antarctica stations are generally in the middle of the ranking, while Peninsula stations, not surprisingly, provide the least amount of information on temperatures across the continent. This group ranking is found in both ND and DA results. The ranking of individual stations within groups from DA experiments does not match the ND results as closely, but some significant similarities are observed. Among the Plateau

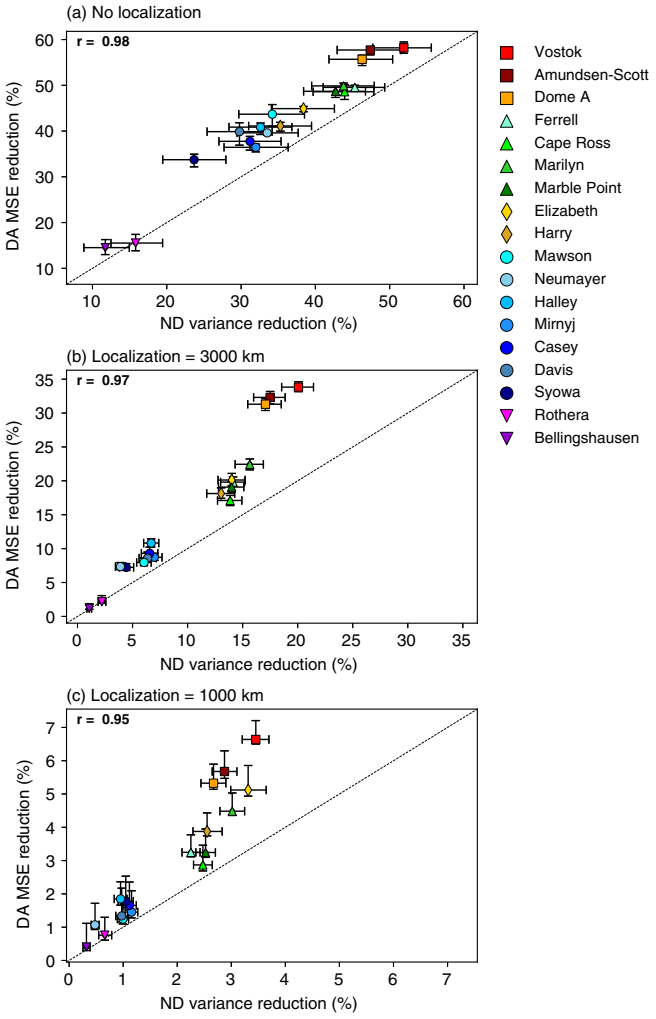

FIGURE 4 MSE reductions, in percent relative to prior MSE, (S_{DA}, Equation 5, *y*-axis) from assimilation of observations, compared with reductions in variance relative to the variance in the prior from the network design method (S_{ND}, Equation 6, *x*-axis) for the CD90 stations considered as individual observations. (a) Experiments without covariance localization, (b) with localization using a cut-off distance of 3,000 km, and (c) with localization using a cut-off distance of 1,000 km. The continental stations on the East Antarctic Plateau are shown as squares, coastal stations as circles, stations clustered near the Ross Ice Shelf as triangles, West Antarctica stations as diamonds, and Peninsula stations using the downward-pointing triangle symbol. The 1:1 line is shown by the dotted line. Error bars represent the 5th–95th percentiles of distributions from the Monte Carlo realizations, and symbols indicate the median values. Note the different ranges on the *y*-axis [Colour figure can be viewed at wileyonlinelibrary.com]

stations, Vostok ranks as the first and Dome A as the third stations in both DA and ND, but the delineation between Vostok and Amundsen–Scott is not as clear in the DA results compared with the ND results. Another common feature between DA and ND is Elizabeth ranking higher than Harry in West Antarctica. Also, Mawson ranks highest and Syowa lowest among the coastal stations. For the two Peninsula stations, Rothera ranks over Bellingshausen in ND, but their difference is not as distinguishable when DA results are considered. Other relative rankings within coastal and Ross Ice Shelf stations cannot be clearly defined, given uncertainties estimated from the MC realizations in DA and ND experiments.

Next we turn to the effect of covariance localization on the results (Figures 4b,c). With localization at 3,000 km, a strong correlation ($r = 0.97$) between DA diagnostics and ND predictions is observed once again, with the same group ranking as with the configuration without covariance localization: continental stations are at the top of the list, followed by Ross Ice Shelf and West Antarctica stations, and then by coastal and Peninsula stations. On the Plateau, Vostok ranks as the top location in DA and ND once again, followed by Amundsen–Scott and Dome A, having similar impact. On the Ross Ice Shelf and West Antarctica, Marilyn ranks the highest in the group of stations by a small margin in both DA and ND. Other locations in the region are statistically indistinguishable in ND results, while only small differences in MSE reductions between Elizabeth, Ferrell, Marble Point, Harry, and Cape Ross are obtained in DA experiments. For coastal stations, Halley stands out in DA results as the station providing the largest reduction in error, while Neumayer and Syowa share the smallest reductions, as in ND results. On the Peninsula, Bellingshausen has the smallest impact in both the DA and ND results. With a shorter localization length-scale (1,000 km: Figure 4c), correlation is reduced to $r = 0.95$ and the ranking among locations within geographical regions is not as clear, particularly between Plateau, West Antarctica, and Ross Ice Shelf stations. Vostok remains the most impactful in the ND and DA results, while Elizabeth in West Antarctica and Marilyn on the Ross Ice Shelf are featured more prominently in the ND results, in contrast to the other interior stations of Amundsen–Scott and Dome A in the DA results. Again, coastal and Peninsula stations occupy the bottom of the ranking, with coastal Neumayer and Bellingshausen and Rothera on the Peninsula appearing as the three lowest-ranked stations in both the ND and DA results.

A notable feature in the comparisons is the general tendency for larger MSE reductions in the DA experiments compared with the ND predictions of total variance, especially when covariance localization is applied. These discrepancies have at least two two contributing factors. One possibility is that the reanalysis product used for verification has different error characteristics. For example, a different representation of seasonal variability in the reference reanalysis will be reflected in the verification rather than the prediction. A second possibility is the exact implementation of covariance localization. As discussed in Section 2, localizing the $\delta\mathbf{J}(\mathbf{H}\delta\mathbf{x})^{\mathrm{T}}$ covariance induces a dependence in calculated changes in variance of $\delta\boldsymbol{\Sigma}^2 \sim w_{\mathrm{loc}}^2$. We found that this formulation is responsible for degraded comparisons against the reductions in variance obtained from updating ensemble perturbations in the ensemble square-root filter in our DA experiments. The squaring of w_{loc} leads to reduced predictions of changes in variance, which do not compare as well with those obtained from the Kalman update in DA. The comparison is improved when the implementation described by Equation 3 is used (see Figure S2 in Appendix S1). We note that this difference in localization influence between ensemble members and the covariance they approximate is generic to most implementations of ensemble filters. It is aggravated here by the squared covariance in $\delta\boldsymbol{\Sigma}^2$ (see Equation 2).

3.2 | Stations as a network (conditional ranking)

A more challenging test consists of assessing the ND conditional selection of stations to build out a network in priority order. As described in H20, the selection of stations beyond the first can be performed once the impact of the previous station(s) has been taken into account. Here we show results applied to the subset of CD90 stations. This will be followed by an assessment of these results using data assimilation experiments.

3.2.1 | Network design applied to CD90 network

The conditional ND ranking of the CD90 stations is shown in Figure 5 and also summarized in Table 2, as quantified by the percentage of samples in the 10,000 MC realizations where the stations are chosen in each rank. In the absence of covariance localization (Figure 5a), Vostok is chosen as the most impactful station nearly 100% of the time, consistent with the approach described in the previous section. Here the impact of Vostok is taken into account by updating the metric and state before choosing the second station. The second station chosen most frequently is Elizabeth in West Antarctica (over 60% of the time), as well as Cape Ross in Victoria Land, and

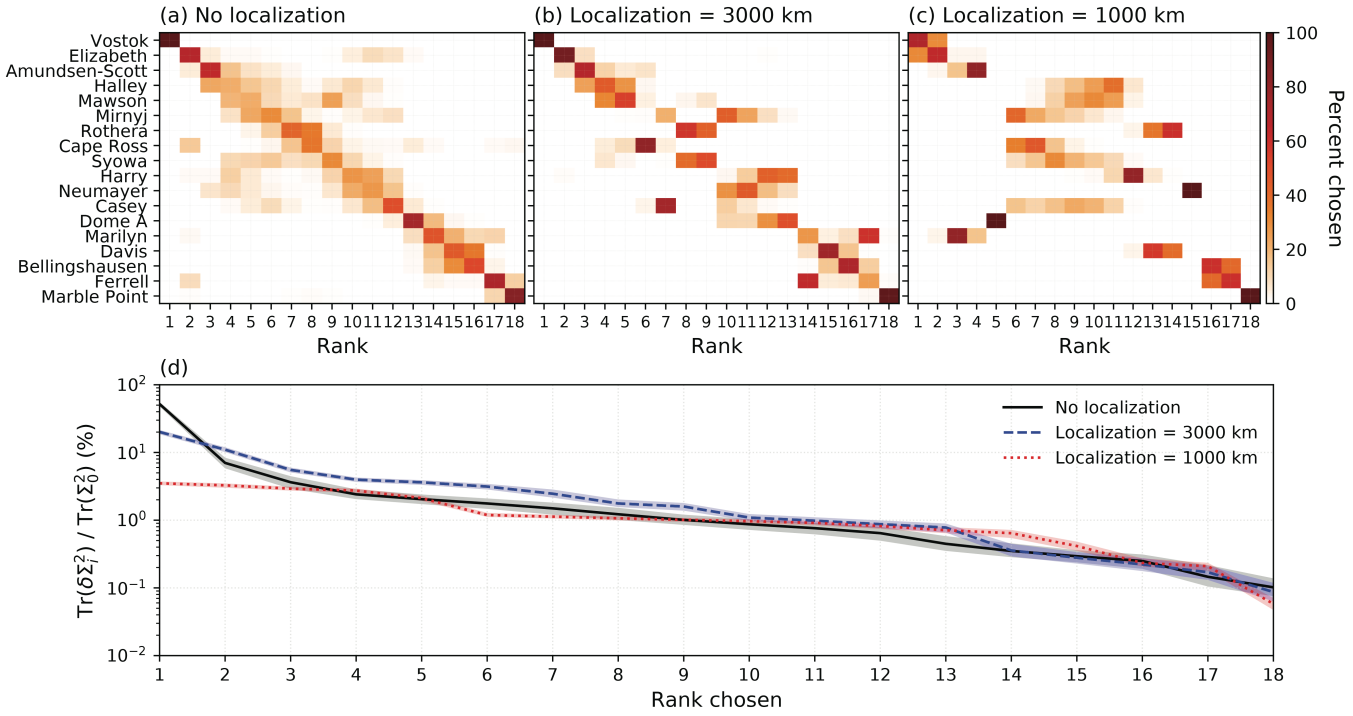

FIGURE 5 Similar to Figure 2, but for the conditional selection of stations. The percentage of samples for which stations are chosen in each rank is shown in the top frames for (a) no covariance localization, (b) covariance localization using a cut-off distance of 3,000 km, and (c) with a cut-off distance of 1,000 km. Lines in (d) show the median of the distributions of the change in the metric variance from the 10,000 Monte Carlo realizations, while shading illustrates the 5th–95th percentiles. Note the logarithmic scale on the *y*-axis in (d). The *x*-axis in all frames here represents station rank, not specific stations as in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

Ferrell (Ross Ice Shelf) and Amundsen–Scott, but at much lower rates. Repeating the process for subsequent stations, Amundsen–Scott is chosen most often as the third most impactful station, although with Halley also sometimes selected third in a small fraction of the realizations. Dome A now ranks among the lower tier of stations in terms of influence, in contrast to when stations are considered as stand-alone independent observations, due to significant covariability with Vostok (i.e., factoring Vostok removes a large amount of the variance explained by Dome A, as highlighted in Section 3.1). Also, Ferrell and Marble Point are now ranked most often as the two least-influential stations, a consequence of the strong covariability with the nearby Cape Ross station that now occupies the upper half of the rankings. Marilyn, Davis, and Bellingshausen also rank in the bottom tier. Coastal stations occupy ranks in the middle tier. The stations with some counts in ranks off the diagonal highlight the importance of uncertainties in network design and the value of adopting a probabilistic perspective.

The distributions of metric reduction corresponding to each rank, relative to the prior metric, are shown in Figure 5d. Vostok, as the first station, explains more than 50% of the total variance, with the second and third stations explaining another 7% and 4%, respectively. Therefore, the first three stations explain more than 60% of the total error variance, with the remaining stations only contributing incremental variance reductions. As described in H20, the sharp decrease in variance from the first few stations reflects an unrealistic removal of ensemble spread by spurious long-distance correlations. The impact of using covariance localization on CD90 station ranking is thus assessed next.

Figures 5b,c show the station rankings obtained with covariance localization cut-off distances of 3,000 and 1,000 km, respectively. With 3,000 km, Vostok and Elizabeth stand out once again as the number one and number two stations, respectively, with increased confidence in Elizabeth's ranking (chosen nearly 100% of the time in the MC realizations). The median change in metric variance from Vostok is reduced from 52% to 20%. Changes are larger with localization for the second to the 13th stations (Figure 5d), a result of the reduced impact of the top station. Amundsen–Scott, Halley, and Mawson again predominantly rank as the third, fourth, and fifth stations, respectively. Also, similar to results obtained without localization, coastal Davis, Bellingshausen on the Peninsula, and Marble Point near Ross Island occupy the bottom positions. Highlighted positions off the diagonal in Figure 5b indicate stations for which localization has

TABLE 2 Conditional ranking of CD90 stations, from most to least impactful, obtained from the ND method for runs (a) without covariance localization and (b) with localization applied with 3,000- and 1,000-km cut-off distances

Rank	No localization	Localization (3,000 km)	Localization (1,000 km)
1	Vostok	Vostok	Vostok
2	Elizabeth	Elizabeth	Elizabeth
3	Amundsen–Scott	Amundsen–Scott	Marilyn
4	Halley	Halley	Amundsen–Scott
5	Mawson	Mawson	Dome A
6	Mirnyj	Cape Ross	Mirnyj
7	Rothera	Casey	Cape Ross
8	Cape Ross	Rothera	Syowa
9	Syowa	Syowa	Casey
10	Harry	Mirnyj	Mawson
11	Neumayer	Neumayer	Halley
12	Casey	Harry	Harry
13	Dome A	Dome A	Davis
14	Marilyn	Ferrell	Rothera
15	Davis	Davis	Neumayer
16	Bellingshausen	Bellingshausen	Bellingshausen
17	Ferrell	Marilyn	Ferrell
18	Marble Point	Marble Point	Marble Point

Note: The rankings also correspond to the order in which station data are assimilated in the "optimal" offline DA experiments.

a significant influence on the ranking. This is the case for Cape Ross, Casey, and Ferrell, which are ranked higher with localization, while Mawson, Harry, and Marilyn rank lower.

With a 1,000-km localization length-scale, the rank order of stations is greatly modified, as shown by stations with large counts appearing in off-diagonal positions in Figure 5c. Changes in metric variance are reduced further, particularly for the top stations (see Figure 5d). This leads to similar changes in variance for the top five stations, indicating a loss in ability to discriminate confidently among the top-tier stations. This is also illustrated by the smaller frequencies with which stations are chosen in the top ranks (Figure 5c). Vostok and Elizabeth are interchangeably chosen as the number one and two stations, while Marilyn is now ranked third most frequently, followed closely by Amundsen–Scott and Dome A as the fourth and fifth stations, respectively. Some delineation remains between the top stations and the coastal and Peninsula stations, which again rank in the bottom tier.

As discussed in H20, and again demonstrated here, covariance localization has a significant influence on network design results. The results indicate that an overestimation of the impact of the first station occurs in the absence of covariance localization, while short localization length-scales lead to a weakened ability to discriminate among the most impactful stations within the CD90 network. A localization cut-off distance of 3,000 km, consistent with the climatological correlation length-scale found by Bumbaco *et al.* (2014), appears to provide a good compromise between mitigating noisy long-distance covariances from sampling errors and maintaining real teleconnections across the continent.

The spatial patterns defining station impact are shown for a 3,000-km covariance localization radius in Figure 6. Vostok reduces variance over a substantial portion of East Antarctica, explaining the largest total metric variance reduction, as discussed above. Vostok is complemented by Elizabeth and Amundsen–Scott, contributing respectively to reductions over West Antarctica and over the Antarctic Plateau around the South Pole and extending toward the Ronne Ice Shelf. The fourth (Halley) and fifth (Mawson) stations project on portions of Queen Maud Land and Mac. Robertson Land, respectively. Cape Ross as the sixth station projects on a more localized area of Victoria Land and, specifically, Cape Adare, while Casey (ranked seventh) provides information on the outer areas of Wilkes Land. The eighth (Rothera) and ninth (Syowa) stations cover portions of the Peninsula and Queen Maud Land, respectively. Weak, more localized, variance reduction is observed with the remaining stations.

3.2.2 | Validation with DA of CD90 network observations

As in Section 3.1.2, offline assimilation of real observations is used to verify ND results. In contrast to the previous section on ND predictions, DA experiments are performed by assimilating all of the CD90 stations sequentially using the EnSRF of Whitaker and Hamill (2002). An ensemble size of 250 is again used, and a total of 100 MC realizations are performed, each with a different random draw of the 250 ensemble members. Here the intermediate analyses at every analysis time (i.e., reflecting updates to the state as data from each station is assimilated sequentially) are saved and compared with reanalysis data (i.e., ERA5). As in the previous section, the ensemble-mean MSE is used as the basis for validating ND predictions.

In a first set of experiments, the sequence of assimilated stations follows the ranking determined by the ND algorithm; that is, stations ranked using the highest selection frequencies over all ND MC realizations (see

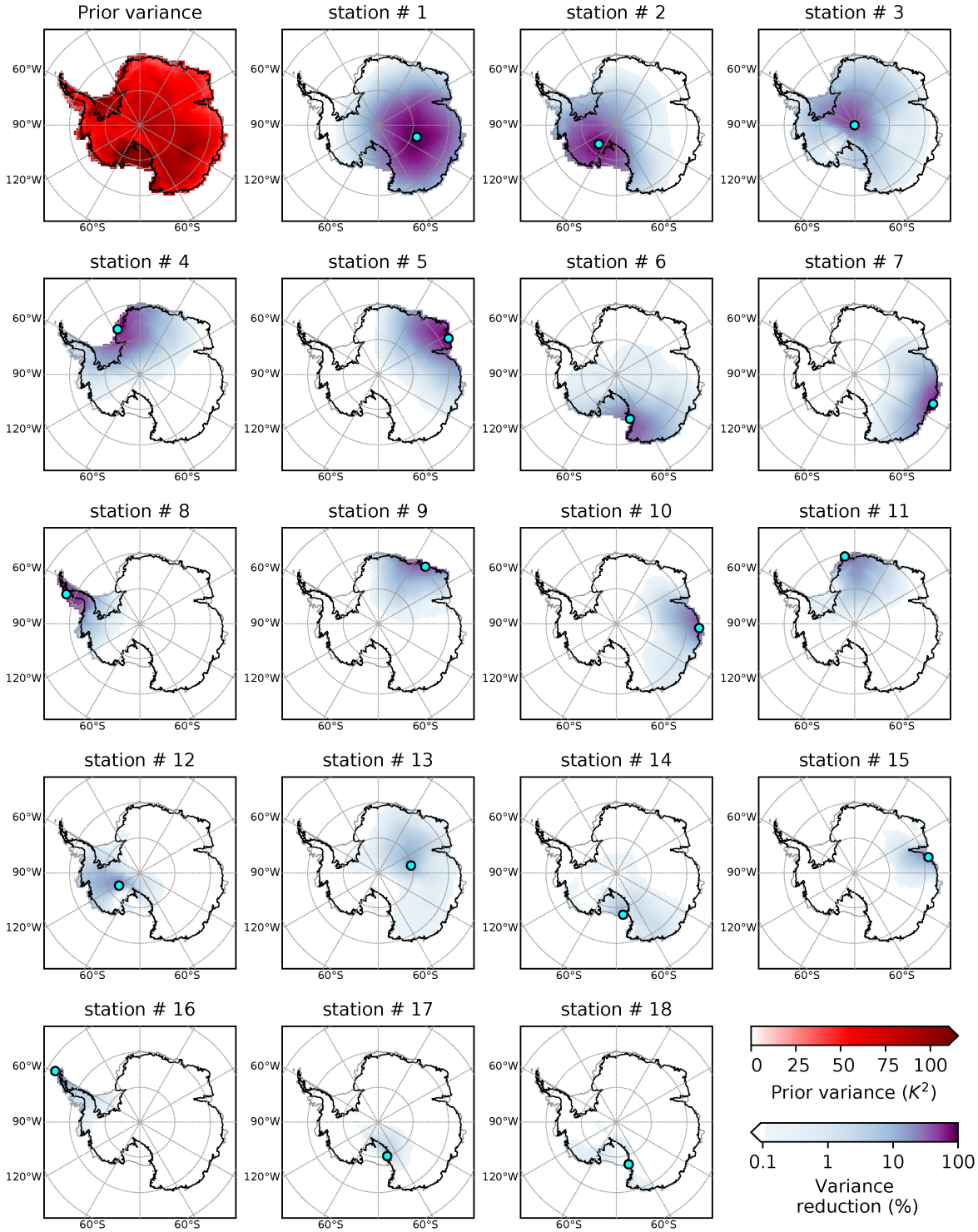

FIGURE 6 Spatial distribution of median prior variance (upper left frame) and conditional contributions to relative changes in variance (in % with respect to the prior) predicted with the ND method for each station (shown by the dot) following the conditional ranking shown in Table 2, using a covariance localization radius of 3,000 km. Note the use of a logarithmic scale in the color bar for variance reduction [Colour figure can be viewed at wileyonlinelibrary.com]

Table 2). The incremental reductions in analysis MSE for each station are averaged over all 0000 UTC analyses, and results are compared with the predicted ND changes in metric variance (Figure 7). From this we find high correlations ($r > 0.95$) between median DA reductions in MSE and ND predictions for each CD90 station, whether covariance localization is used or not. For the experiment without covariance localization (Figure 7a), reductions in

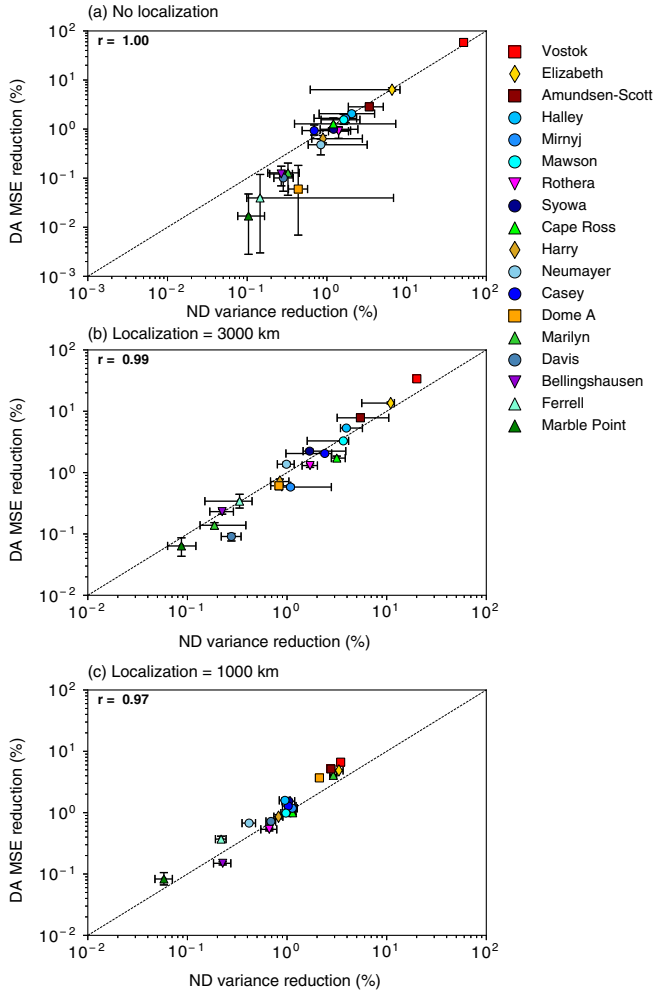

FIGURE 7 Similar to Figure 4, but comparing conditional reductions in MSE from DA with ND-estimated changes in ensemble variance for experiments (a) without covariance localization, (b) with a 3,000 km covariance localization cut-off radius, and (c) with a cut-off radius of 1,000 km. Note the use of logarithmic scales on all axes and the different range along the *x*-axis on (b) and (c) [Colour figure can be viewed at wileyonlinelibrary.com]

analysis errors correspond well to ND predictions for the highest-ranking stations, with a tendency to underestimate the impact of the last six stations. When localization is applied with a cut-off radius of 3,000 km (Figure 7b), good correspondence between DA and ND measures of station impact is again obtained, and the match is particularly improved for the lowest-ranking stations. However, MSE reductions in DA results are now somewhat larger than predicted for the top stations. This effect is also present with 1,000 km localization (Figure 7c), but a good station-to-station correspondence remains between DA and ND estimates.

We note also that the error bars in Figure 7, representing the 5th and 95th percentiles of the distributions, show that measures of station impact can cover a wide range of values. This spread is generally not as pronounced when localization is applied, indicating that the influence of station observations over long distances contributes to uncertainty in continent-wide temperature estimates. Also, tighter distributions tend to characterize error reductions derived from DA, possibly resulting from the averaging of errors in the time domain.

Next, we assess ND guidance from an integrated network perspective by examining whether the sequential assimilation of station data using an optimal network outperforms a randomly-chosen network. We do this by comparing results obtained from optimal (see Table 2) and random station ordering for networks of a given size. The same Monte Carlo approach is employed, with 100 realizations of prior ensembles, with the addition of different realizations of station orders for the random network experiments.

The reduction in analysis MSE, summarized with the S_{DA} metric, is shown in Figure 8. A monotonic reduction of errors is apparent, converging to a common endpoint after all 18 stations have been assimilated. The narrow uncertainty distributions for the optimal station order (too small to be clearly visible on the figure) are compared with the much wider distributions for randomly-chosen stations, denoting the negligible contribution of uncertainties in prior states to the posterior ensemble spread. Without covariance localization (Figure 8a), nearly 60% of errors are eliminated with the first optimal station, reaching 70% by the fourth station, and error reductions are within the uncertainty after the 11th station. Error reductions from optimal stations are significantly larger than those from randomly-selected stations. Differences become small after the fifth station, and then slowly converge to 78% error reduction as more stations are included. With a 3,000-km covariance localization cut-off distance (Figure 8b), a similar level of error reduction as in the case without localization is ultimately reached once all stations have been assimilated (75% of prior errors compared with 78%), but with a slower convergence rate. Randomly-ordered stations yield significantly smaller error reductions, reaching the 60% level only by inclusion of the eighth station, as compared with the fourth station in the optimal case. Similar behavior is found for 1,000-km localization (Figure 8c), but with this artificially short influence radius the total error reduction only reaches 35%.

In order to test the sensitivity of the error analysis to the verification dataset, we repeated the calculations using the MERRA2 and JRA55 reanalyses. The results, reported in Appendix S1, confirm the conclusions drawn using ERA5. The good correspondence between the ND variance reductions and the MSE reductions from DA is also found for both MERRA2 and JRA55 used as the verification dataset (Figure S3). Similar correlations in

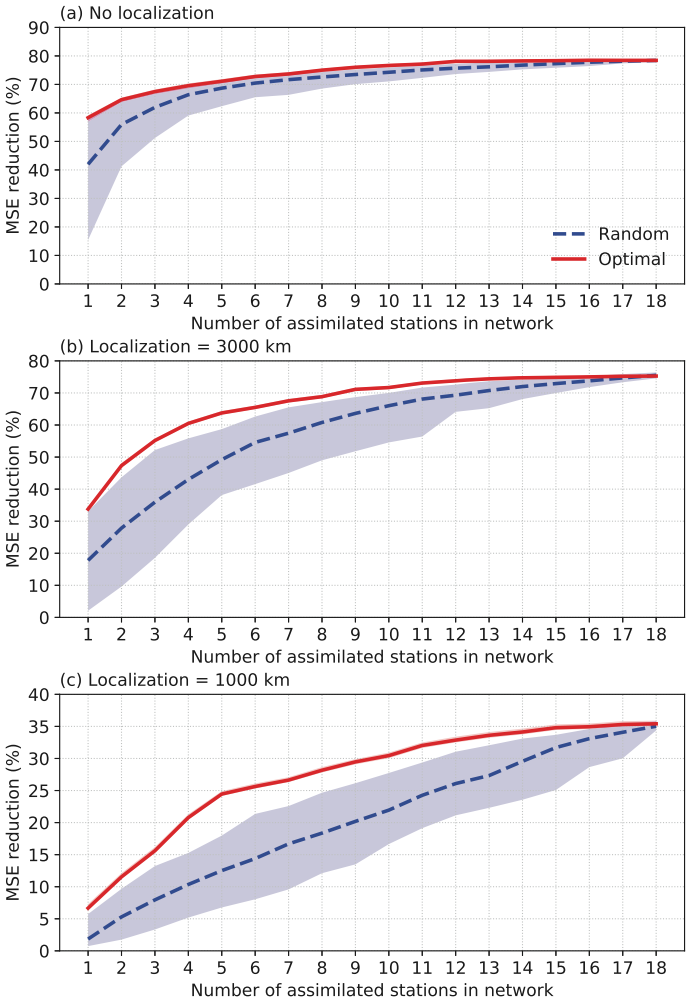

FIGURE 8 Cumulative change in analysis MSE with respect to ERA5 (summarized with the S_{DA} metric, Equation 5), as station data are assimilated: (a) without covariance localization, (b) with 3,000-km localization, (c) with 1,000-km localization. Solid lines and shading represent the median and the 5th and 95th percentiles of the 100 Monte Carlo realizations [Colour figure can be viewed at wileyonlinelibrary.com]

station-to-station comparisons between DA MSE and ND variance reductions are also obtained, as are the greater MSE reductions from the assimilation of stations ordered according to ND predictions, as opposed to random selections (Figure S4). The validation results with covariance localization are also found to be robust to the choice of reference reanalysis.

Once all CD90 stations have been assimilated using a realistic 3,000-km covariance localization (Figure 9a), the regions with the most and least error reductions correspond well with the ND predictions (Figure 9b). Differences between DA and ND results (Figure 9c) confirm that MSE reductions are a few percentage points larger than ND predictions over most of the continent, with the exception of an area around Vostok and near Cape Adare. Despite these differences, MSEs are reduced most over the higher terrain of the East Plateau, the Megadunes area, and central and West Antarctica and the Ross Ice Shelf region, in good agreement with ND predictions. MSEs are reduced least over Palmer Land and Ellsworth Land, the coastal areas of Victoria Land and Adélie Land, and in a region extending from the coast toward the interior of Queen Maud Land. These correspond to the largest areas missing observations within the CD90 network. We note that the aforementioned regions correspond to locations that complement the existing CD90 network optimally, as found by H20 (see their Figure 8a–c). We also note the good agreement with the patterns in variance explained for weekly and monthly time-scales presented in Figure 6 of Bumbaco *et al.* (2014).

4 | SUMMARY AND CONCLUSIONS

An objective method for the optimal siting of stations for monitoring surface air temperature in Antarctica has been proposed and demonstrated by Hakim *et al.* (2020) (H20). Preferred regions were clearly identified for siting observations, either in designing a theoretical network from scratch or for augmenting existing networks for improved monitoring of continental surface temperatures. The deployment and maintenance of observational assets in the context of the harsh, difficult-to-access environment of Antarctica motivates validation of this approach for

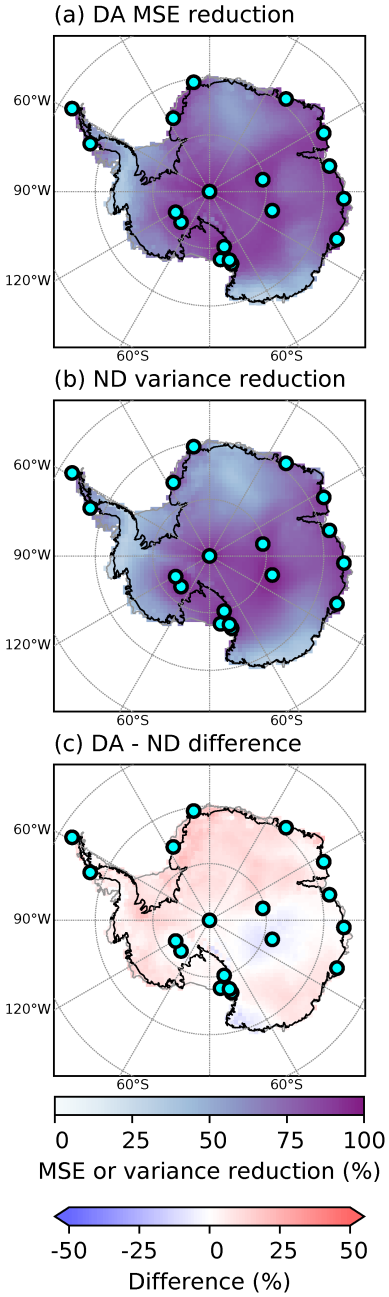

FIGURE 9 Total relative reduction in (a) MSE from the assimilation of CD90 observations, (b) variance from the ND method after all 18 stations (dots) have been included in respective calculations, and (c) the difference between DA error reductions and ND predictions. Results obtained using a 3,000-km covariance localization cut-off radius [Colour figure can be viewed at wileyonlinelibrary.com]

practical applications. Here we perform such a check by assimilating observations from existing surface stations using an offline ensemble Kalman filter and then comparing results with predictions from the optimal siting approach.

This validation has been performed by quantifying the contributions of real station observations to reducing errors from a climatological prior analysis, where the errors are determined relative to an independent verification reference. Offline data assimilation experiments provide a cost-efficient alternative to full observation-system experiments (OSE), with the former necessary to evaluate the 149,200 experiments (daily analyses over 49 months and 100 MC realizations) considered here for individual station and full-network configurations. We consider observations from stations with high levels of data availability (i.e., ≥90% reportage in our "CD90" network) and reductions in ensemble-mean MSEs. Validation of the ND method is performed by comparing these error diagnostics with the predictions of variance reductions from the same set of stations. The MSE statistics are sampled across the entire continent and compiled over the 149,200 analyses.

The comparisons show that the ND predictions agree with the station-to-station reductions in MSE, indicating that the ND method is able to discriminate between error reduction for different stations, either as individual observing locations or more realistically as part of a network. The greater importance of stations located on the East Antarctic Plateau is clearly identified in both the ND and DA results. This location is followed by stations in West Antarctica and the Ross Ice Shelf region, while coastal and Peninsula CD90 stations have the least extensive impact. This is consistent with "blank slate" results in H20, who found that the Megadunes area in East Antarctica and the Siple Coast region in West Antarctica are key observation locations. Also, the results from our DA experiments demonstrate that equivalent levels of error reduction can be achieved with fewer stations when observations from the ND top-ranked stations are assimilated, providing further evidence of the usefulness of the ND method toward optimizing temperature monitoring in Antarctica. In addition, our study has shown that ND validation results are robust to the application of covariance localization. The agreement between the DA estimates of error reductions and ND predictions is found for a range of localization length-scales.

In addition to the successful validation of the ND predictions, our tests here highlight important factors in the application of the method. The use of ND in testing CD90 stations considered as single sites informing on temperatures across the continent clearly reveals the presence of long-distance covariances estimated from samples of WRF-modelled temperatures. As discussed in H20, covariance localization is an important component of this method, and one that requires careful consideration. Our results confirm the importance of mitigating the effects of uncertain long-distance covariances, while preventing the suppression of measurement information and the inability to discriminate between stations when spatial covariance

localization is overly restricted. Our results also show strong similarities in the spatial influence between certain stations, suggesting redundant contributions of sites to the observation/monitoring of regional temperature. In turn, this suggests the value of adopting a network perspective in which observation sites are evaluated conditionally upon the presence and input of others, where such redundancy is intrinsically taken into account.

Adopting a network perspective with a realistic 3,000-km localization radius, we find that Vostok in East Antarctica, Amundsen–Scott at the South Pole, and Elizabeth in West Antarctica combine to reduce total errors by more than 25%. Halley in the Atlantic Ocean sector and Mawson in the Indian Ocean sector figure prominently among coastal stations with contributions in the 5% range, while Cape Ross in Victoria Land accounts for an additional few percentage points of error reduction. While many other stations have been deployed beyond those of the CD90 set considered here, our tests serve to demonstrate the quantitative information that can be gained using the objective ND method in the Antarctic setting. It has been shown that areas with the least amount of information provided by a network can be identified with confidence, and thus possibly targeted for the deployment of additional assets.

Successful validation of the ND method provides added motivation for exploiting information in decisions to establish an effective observing network in support of science and logistics in Antarctica. We emphasize, however, that the information provided by the ND method provides only one contributing element in an observational decision-making process, which can involve varying aims, such as answering specific science questions or ensuring safe aircraft operations. We believe the method presented offers the necessary flexibility for addressing different monitoring goals, whether the focus is on monitoring a climate variable, capturing weather conditions, or reducing numerical model forecast errors. Whatever the end purpose, optimal design can be a useful tool toward efficient allocation of the resources applied to it.

ACKNOWLEDGEMENTS

The authors thank the National Science Foundation Office of Polar Programs for supporting this research through Grants 1542766 to the University of Washington and Grants 1542789, 1543305, and 1924730 to the University of Wisconsin–Madison. We gratefully acknowledge conversations on localization and ensemble filters with Chris Snyder.

AUTHOR CONTRIBUTIONS

Robert Tardif: conceptualization; data curation; formal analysis; methodology; validation; visualization; writing – original draft. **Gregory J. Hakim:** conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; writing – original draft; writing – review and editing. **Karin A. Bumbaco:** methodology; project administration; resources; writing – review and editing. **Matthew A. Lazzara:** conceptualization; funding acquisition; project administration; resources; supervision; writing – review and editing. **Kevin W. Manning:** data curation; resources; writing – review and editing. **David E. Mikolajczyk:** data curation; resources; writing – review and editing. **Jordan G. Powers:** conceptualization; formal analysis; funding acquisition; investigation; project administration; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Robert Tardif https://orcid.org/0000-0003-1313-0878
Gregory J. Hakim https://orcid.org/0000-0001-8486-9739
Matthew A. Lazzara https://orcid.org/0000-0002-4343-498X
David E. Mikolajczyk https://orcid.org/0000-0001-7931-5504

REFERENCES

Ancell, B. and Hakim, G.J. (2007) Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Monthly Weather Review*, 135, 4117–4134. https://doi.org/10.1175/2007MWR1904.1.

Bergot, T. (1999) Adaptive observations during FASTEX: a systematic survey of upstream flights. *Quarterly Journal of the Royal Meteorological Society*, 125, 3271–3298. https://doi.org/10.1002/qj.49712556108.

Bishop, C.H., Etherton, B.J. and Majumdar, S.J. (2001) Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Monthly Weather Review*, 129, 420–436. https://doi.org/10.1175/1520-0493(2001)129,0420:ASWTET.2.0.CO;2.

Bozkurt, D., Bromwich, D.H., Carrasco, J., Hines, K.H., Maureira, J.C. and Rondanelli, R. (2020) Recent near-surface temperature trends in the Antarctic Peninsula from observed, reanalysis and regional climate model data. *Advances in Atmospheric Sciences*, 37, 477–493. https://doi.org/10.1007/s00376-020-9183-x.

Bracegirdle, T.J. and Marshall, G.J. (2012) The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *Journal of Climate*, 25, 7138–7146. https://doi.org/10.1175/JCLI-D-11-00685.1.

Bromwich, D.H., Otieno, F.O., Hines, K.M., Manning, K.W. and Shilo, E. (2013) Comprehensive evaluation of polar Weather Research and Forecasting model performance in the Antarctic. *Journal of Geophysical Research: Atmospheres*, 118, 274–292. https://doi.org/10.1029/2012JD018139.

Bromwich, D.H., Werner, K., Casati, B., Powers, J.G., Gorodetskaya, I.V., Massonnet, F., Vitale, V., Heinrich, V.J., Liggett, D., Arndt, S., Barja, B., Bazile, E., Carpentier, S., Carrasco, J.F., Choi, T., Choi, Y., Colwell, S.R., Cordero, R.R., Gervasi, M., Haiden, T., Hirasawa, N., Inoue, J., Jung, T., Kalesse, H., Kim, S.-J., Lazzara, M.A., Manning, K.M., Norris, K., Park, S.-J., Reid, P., Rigor, I., Rowe, P.M., Schmithüsen, H., Seifert, P., Sun, Q., Uttal, T., Zannoni, M. and Zou, X. (2020) The year of polar prediction in the southern hemisphere (YOPP-SH). *Bulletin of the American Meteorological Society*, 101, E1653–E1676. https://doi.org/10.1175/BAMS-D-19-0255.1.

Bumbaco, K.A., Hakim, G.J., Mauger, G.S., Hryniw, N. and Steig, E.J. (2014) Evaluating the Antarctic observational network with the Antarctic mesoscale prediction system (AMPS). *Monthly Weather Review*, 142, 3847–3859. https://doi.org/10.1175/MWR-D-13-00401.1.

Chapman, W.L. and Walsh, J.E. (2007) A synthesis of Antarctic temperatures. *Journal of Climate*, 20, 4096–4117. https://doi.org/10.1175/JCLI4236.1.

Clem, K.R., Fogt, R.L., Turner, J., Lintner, B., Marshall, G.J., Miller, J.R. and Renwick, J.A. (2020) Record warming at the South Pole during the past three decades. *Nature Climate Change*, 10, 762–770. https://doi.org/10.1038/s41558-020-0815-z.

Devers, A., Vidal, J.-P., Lauvernet, C., Graff, B. and Vannier, O. (2020) A framework for high-resolution meteorological surface reanalysis through offline data assimilation in an ensemble of downscaled reconstructions. *Quarterly Journal of the Royal Meteorological Society*, 146, 153–173. https://doi.org/10.1002/qj.3663.

Evensen, G. (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343–367. https://doi.org/10.1007/s10236-003-0036-9.

Fréville, H., Brun, E., Picard, G., Tatarinova, N., Arnaud, L., Lanconelli, C., Reijmer, C. and Broeke, M.V.D. (2014) Using MODIS land surface temperatures and the Crocus snow model to understand the warm bias of ERA-Interim reanalyses at the surface in Antarctica. *Cryosphere*, 8, 1361–1373. https://doi.org/10.5194/tc-8-1361-2014.

Fujii, Y., Rémy, E., Zuo, H., Oke, P., Halliwell, G., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J., Xue, Y., Masuda, S., Smith, G.C., Balmaseda, M., Germineaud, C., Lea, D.J., Larnicol, G., Bertino, L., Bonaduce, A., Brasseur, P., Donlon, C., Heimbach, P., Kim, Y.-H., Kourafalou, V., Traon, P.-Y.L., Martin, M., Paturi, S., Tranchant, B. and Usui, N. (2019) Observing system evaluation based on ocean data assimilation and prediction systems: on-going challenges and a future vision for designing and supporting ocean observational networks. *Frontiers in Marine Science*, 6, 417. https://doi.org/10.3389/fmars.2019.00417.

Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. https://doi.org/10.1002/qj.49712555417.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M. and Zhao, B. (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Gonzalez, S., Vasallo, F., Sanz, P., Quesada, A. and Justel, A. (2021) Characterization of the summer surface mesoscale dynamics at Dome F, Antarctica. *Atmospheric Research*, 259, 105699. https://doi.org/10.1016/j.atmosres.2021.105699.

Gossart, A., Helsen, S., Lenaerts, J.T.M., Broucke, S.V., van Lipzig, N.P.M. and Souverijns, N. (2019) An evaluation of surface climatology in state-of-the-art reanalyses over the Antarctic Ice Sheet. *Journal of Climate*, 32, 6899–6915. https://doi.org/10.1175/JCLI-D-19-0030.1.

Ha, S.-Y. and Snyder, C. (2014) Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Monthly Weather Review*, 142, 1489–1508. https://doi.org/10.1175/MWR-D-13-00108.1.

Hakim, G.J., Bumbaco, K.A., Tardif, R. and Powers, J.G. (2020) Optimal network design applied to monitoring and forecasting temperature in Antarctica. *Monthly Weather Review*, 148, 857–873. https://doi.org/10.1175/MWR-D-19-0103.1.

Hakim, G.J., Emile-Geay, J., Steig, E.J., Noone, D., Anderson, D.M., Tardif, R., Steiger, N.J. and Perkins, W.A. (2016) The last millennium climate reanalysis project: framework and first results. *Journal of Geophysical Research: Atmospheres*, 121, 6745–6764. https://doi.org/10.1002/2016JD024751.

Hamill, T.M., Yang, F., Cardinali, C. and Majumdar, S.J. (2013) Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Monthly Weather Review*, 141, 2058–2065. https://doi.org/10.1175/MWR-D-12-00309.1.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Houtekamer, P.L. and Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1.

Jones, P.D. and Lister, D.H. (2015) Antarctic near-surface air temperatures compared with ERA-Interim values since 1979. *International Journal of Climatology*, 35, 1354–1366. https://doi.org/10.1002/joc.4061.

Jung, T., Gordon, N.D., Bauer, P., Bromwich, D.H., Chevallier, M., Day, J.T., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H.D., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I.A., Smith, G., Svensson, G., Tolstykh, M. and Yang, W. (2016) Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, 97, 1631–1647. https://doi.org/10.1175/BAMS-D-14-00246.1.

Khare, S. and Anderson, J. (2006) A methodology for fixed observational network design: theory and application to a simulated global prediction system. *Tellus A*, 58A, 523–537. https://doi.org/10.1111/j.1600-0870.2006.00200.x.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93, 5–48. https://doi.org/10.2151/jmsj.2015-001.

Langland, R.H. and Baker, N.L. (2004) Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, 56A, 189–201. https://doi.org/10.1111/j.1600-0870.2004.00056.x.

Langland, R.H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M.A., Majumdar, S.J., Morss, R.E., Rohaly, G.D., Velden, C., Bond, N. and Bishop, C.H. (1999) The North Pacific Experiment (NORPEX-98): targeted observations for improved North American weather forecasts. *Bulletin of the American Meteorological Society*, 80, 1363–1384. https://doi.org/10.1175/1520-0477(1999)080<1363:TNPENT>2.0.CO;2.

Lazzara, M.A., Weidner, G.A., Keller, L.M., Thom, J.E. and Cassano, J.J. (2012) Antarctic automatic weather station program: 30 years of polar observation. *Bulletin of the American Meteorological Society*, 93, 1519–1537. https://doi.org/10.1175/BAMS-D-11-00015.1.

Majumdar, S.J. (2016) A review of targeted observations. *Bulletin of the American Meteorological Society*, 97, 2287–2303. https://doi.org/10.1175/BAMS-D-14-00259.1.

Mauger, G.S., Bumbaco, K.A., Hakim, G.J. and Mote, P.W. (2013) Optimal design of a climatological network: beyond practical considerations. *Geoscientific Instrumentation Methods and Data Systems*, 2, 199–212. https://doi.org/10.5194/gi-2-199-2013.

Morss, R.E., Emanuel, K.A. and Snyder, C. (2001) Idealized adaptive observation strategies for improving numerical weather prediction. *Journal of the Atmospheric Sciences*, 58, 210–232. https://doi.org/10.1175/1520-0469(2001)058<0210:IAOSFI>2.0.CO;2.

Odishaw, H. (1959) International geophysical year. *Science*, 129, 14–25. https://doi.org/10.1126/science.129.3340.14.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D. and Kosro, P.M. (2002) Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research: Oceans*, 107, 3122. https://doi.org/10.1029/2000JC000511.

Powers, J.G., Manning, K.W., Bromwich, D.H., Cassano, J.J. and Cayette, A.M. (2012) A decade of Antarctic science support through AMPS. *Bulletin of the American Meteorological Society*, 93, 1699–1712. https://doi.org/10.1175/BAMS-D-11-00186.1.

Powers, J.G., Monahan, A.J., Cayette, A.M., Bromwich, D.H., Kuo, Y. and Manning, K.W. (2003) Real-time mesoscale modeling over Antarctica: the Antarctic Mesoscale Prediction System. *Bulletin of the American Meteorological Society*, 84, 1533–1546. https://doi.org/10.1175/BAMS-84-11-1533.

Rabier, F., Cohn, S., Cocquerez, P., Hertzog, A., Avallone, L., Deschler, T., Haase, J., Hock, T., Doerenbecher, A., Wang, J., Guidard, V., Thépaut, J.-N., Langland, R., Tangborn, A., Balsamo, G., Brun, E., Parsons, D., Bordereau, J., Cardinali, C., Danis, F., Escarnot, J.-P., Fourrié, N., Gelaro, R., Genthon, C., Ide, K., Kalnajs, L., Martin, C., Meunier, L.-F., Nicot, J.-M., Perttula, T., Potts, N., Ragazzo, P., Richardson, D., Sosa-Sesma, S. and Vargas, A. (2013) The Concordiasi field experiment over Antarctica: first results from innovative atmospheric measurements. *Bulletin of the American Meteorological Society*, 94, ES17–ES20. https://doi.org/10.1175/BAMS-D-12-00005.1.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D. and Huang, X.-Y. (2019). A description of the Advanced Research WRF Model Version 4. NCAR Technical Note NCAR/TN-556+STR, UCAR/NCAR. https://doi.org/10.5065/1dfh-6p97.

Smith, A., Lott, N. and Vose, R. (2011) The integrated surface database: recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92, 704–708. https://doi.org/10.1175/2011BAMS3015.1.

Snyder, C. (1996) Summary of an informal workshop on adaptive observations and FASTEX. *Bulletin of the American Meteorological Society*, 77, 953–961. https://doi.org/10.1175/1520-0477-77.5.953.

Szunyogh, I., Toth, Z., Emanuel, K.A., Bishop, C.H., Snyder, C., Morss, R.E., Woolen, J. and Marchok, T. (1999) Ensemble-based targeting experiments during FASTEX: the effect of dropsonde data from the LearJet. *Quarterly Journal of the Royal Meteorological Society*, 125, 3189–3217. https://doi.org/10.1002/qj.49712556105.

Torn, R.D. and Hakim, G.J. (2008) Ensemble-based sensitivity analysis. *Monthly Weather Review*, 136, 663–677. https://doi.org/10.1175/2007MWR.2132.1.

Weatherhead, E.C., Wielicki, B.A., Ramaswamy, V., Abbott, M., Ackerman, T.P., Atlas, R., Brasseur, G., Bruhwiler, L., Busalacchi, A.J., Butler, J.H., Clack, C.T.M., Cooke, R., Cucurull, L., Davis, S.M., English, J.M., Fahey, D.W., Fine, S.S., Lazo, J.K., Liang, S., Loeb, N.G., Rignot, E., Soden, B., Stanitski, D., Stephens, G., Tapley, B.D., Thompson, A.M., Trenberth, K.E. and Wuebbles, D. (2018) Designing the climate observing system of the future. *Earth's Future*, 6, 80–102. https://doi.org/10.1002/2017EF000627.

Wei, T., Yan, Q. and Ding, M. (2019) Distribution and temporal trends of temperature extremes over Antarctica. *Environmental Research Letters*, 14, 084040. https://doi.org/10.1088/174.8-9326/ab33c1.

Whitaker, J.S. and Hamill, T.M. (2002) Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130, 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Zhu, J., Xie, A., Qin, X., Wang, Y., Xu, B. and Wang, Y. (2021) An assessment of ERA5 reanalysis for Antarctic near-surface air temperature. *Atmosphere*, 12, 217. https://doi.org/10.3390/atmos12020217.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Tardif, R., Hakim, G.J., Bumbaco, K.A., Lazzara, M.A., Manning, K.W., Mikolajczyk, D.E. *et al.* (2022) Assessing observation network design predictions for monitoring Antarctic surface temperature. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 727–746. Available from: https://doi.org/10.1002/qj.4226