REANALYSES AND OBSERVATIONS

What's the Difference?

BY WENDY S. PARKER

Differences between reanalysis datasets and familiar observations and measurements are not as deep as one might think, but there is still good reason for caution when using reanalysis data.

Reanalyses are among the most used datasets in the study of weather and climate. They provide comprehensive snapshots of conditions at regular intervals over long time periods—often years or decades. They are produced via data assimilation, a process that relies on both observations and model-based forecasts to estimate conditions. Despite these hybrid origins, practitioners frequently refer to reanalysis data as "observations" and use them for the same purposes as traditional observations. They have been used to study atmospheric dynamics (Kidston et al. 2010), to investigate climate variability (Kravtsov et al. 2014), to evaluate climate models (Gleckler et al. 2008), as data in which to look for the presence of greenhouse gas fingerprints (Santer et al. 2004), and for many other purposes. Recently, reanalysis data were even used to rebut skepticism about the reliability of thermometer-based estimates of twentieth-century global warming (Compo et al. 2013).

At the same time, some scientists warn that reanalysis data should not be equated with "real" observations and measurements (e.g., Schmidt 2011; Bosilovich et al. 2013). But if there are important differences between reanalysis data and familiar observations and measurements, such as those obtained from thermometers and rain gauges, what are these differences exactly? This essay examines four possible answers, considering how well each stands up to scrutiny. Some purported differences are shown to be illusory, while others are argued to be less significant than one might think. The most important difference is simply that errors and uncertainties associated with reanalysis results are often less well understood than those associated with observations. This difference can make it difficult to know what today's reanalysis datasets can—and cannot—be appropriately used for, and points to the need for increased efforts to understand and communicate the strengths and limitations of reanalysis systems.

AFFILIATION: PARKER—Department of Philosophy, Durham University, Durham, United Kingdom
CORRESPONDING AUTHOR: Dr. Wendy S. Parker, Department of Philosophy, Durham University, 50 Old Elvet, Durham DH1 3HN, United Kingdom
E-mail: wendy.parker@durham.ac.uk

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-14-00226.1

In final form 18 December 2015

DATA ASSIMILATION AND REANALYSIS.

In general terms, *data assimilation* can be characterized as a process in which available information is used to estimate as accurately as possible the state of a system (Talagrand 1997). In atmospheric data assimilation, this information typically includes both observations from a variety of sources—ground-based stations, ships, airplanes, and satellites—and forecasts from numerical weather prediction (NWP) models. The NWP forecast(s) provides a first-guess estimate of the atmospheric state, which is then updated in light of the observations. Different assimilation methods perform the updating somewhat differently (see, e.g., Kalnay 2003; Rabier 2005).

Data assimilation is a crucial part of operational NWP today. It is used to produce the *analysis* of current conditions that serves as the starting point for the next NWP forecast cycle; from gappy observations, the assimilation system delivers a complete gridded state estimate that provides values (initial conditions) for all NWP model variables at all grid points. Advances in data assimilation methodologies in recent decades have been credited with significantly improving NWP forecasts (Kalnay 2003).

Since the 1990s, data assimilation also has been used to construct long-term datasets for use in climate and other research, in a process known as retrospective analysis, or *reanalysis* (Trenberth and Olson 1988; Bengtsson and Shukla 1988). Reanalysis involves performing data assimilation for past periods, using a current NWP model and data assimilation method and some or all of the data that are now available for those past periods. It produces a long sequence of comprehensive snapshots (analyses) of atmospheric conditions—a reanalysis dataset.[1]

The first major atmospheric reanalysis project was the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) 40-Year Reanalysis (Kalnay et al. 1996), which delivered global analyses at subdaily intervals for the period 1957–96. Numerous other reanalysis projects have since been undertaken, covering various spatial and temporal domains (see, e.g., Bromwich et al. 2010; Rienecker at al. 2011; Dee et al. 2011; Compo et al. 2011). The University Corporation for Atmospheric Research (UCAR) Climate Data Guide (Dee et al. 2015) provides an overview of key features of recent atmospheric reanalysis projects, including the time periods they cover, the frequency of data they provide, and the types of assimilation methods they employ.

These reanalysis datasets are in heavy use. The Web of Science database currently reports more than 13,000 citations of the NCEP–NCAR 40-year reanalysis dataset. Reanalyses are attractive because they provide comprehensive, gridded estimates of atmospheric conditions at regular intervals over long time periods; this is a very convenient format for evaluating climate models, for climate change detection and attribution studies, and for many other purposes. Nevertheless, as noted above, some scientists warn that reanalysis data should not be equated with real observations. This raises the question: What important differences are there between reanalysis data and familiar observations and measurements?

THEORY-BASED INFERENCE VERSUS MIRRORING. One might think that reanalysis data and familiar observations and measurements differ as follows: Reanalysis datasets are produced via a complex inferential process that involves theory-based calculation, whereas familiar observations and measurements are obtained directly from instrument readings that mirror atmospheric conditions.

But this is just not correct. It is true that reanalysis results are inferred with the help of theory-based calculation: producing the first-guess forecast(s) involves calculating later conditions from earlier ones using NWP models that incorporate approximate laws of atmospheric motion as well as some empirical parameters and relationships. But many familiar observations and measurements also are inferred with the help of theory-based calculation; they are not obtained directly from instrument readings that mirror atmospheric conditions.

For instance, raw instrument readings often must be corrected for interfering factors. A raw rain gauge reading, for example, might need to be corrected for loss due to ambient wind interacting with the gauge; the final estimate of rainfall depth is not simply read off the gauge but is inferred using both the gauge reading and, say, an equation indicating how catch is reduced as a function of wind speed and precipitation type. This equation is likely to be informed by both theory and empirical data.

Some observations and measurements involve theory-based inference in an even more central way: results are derived from measurements of other physical variables or parameters, with the help of theory. For example, observations of relative humidity are

[1] The use of an unchanging NWP model and assimilation method avoids the artificial jumps and trends that appear in sequences of operational analyses as a result of periodic improvements to models and methods. Some artificial jumps and trends can still occur in reanalysis datasets, however, because of changes in the observational base over time (e.g., new instruments are added to the observing network and existing instruments are decommissioned and/or degrade in performance).

sometimes derived from a combination of wet-bulb temperature, dry-bulb temperature, and pressure measurements using an equation (or a psychrometric chart) that is at least partly theory based. Likewise, observations of storm radial velocity are obtained from radar with the help of Doppler effect calculations. Results like these are sometimes called *derived measurements* or *indirect observations*.

Some metrologists and philosophers of science argue that virtually all scientific measurement involves inference: we infer measurement outcomes from instrument indications, with the help of a *measurement model* that is very often informed by theory (see Mari 2005; Boumans 2006; Tal 2012). A measurement model is a conceptualization of i) the physical interactions that take place during a measuring process—both desired interactions and interfering ones—as well as ii) how the results of those interactions relate to values of the parameter(s) that we seek to measure. Such a model guides the inference from the rain gauge's indication to the final estimate of rainfall depth and from the thermometer and barometer indications to the final estimate of relative humidity. In some cases, the inference from instrument indication(s) to measurement outcome is rather trivial—we might have good reason to take a particular thermometer's indication at face value, for example—but in many other cases it is more complex, involving calculations informed by theory.

RELIANCE ON FORECASTS. A second, obvious way in which reanalysis results seem to differ from familiar observations and measurements is that reanalysis results are determined in part by forecasts, whereas familiar observations and measurements are not.

Reanalysis results do differ from familiar observations and measurements in this way. We saw above that both reanalysis results and many familiar observations and measurements are obtained with the help of theory-based calculations. In the case of reanalysis however, some of the equations used in the calculations relate variables at different times. These include the dynamical equations of NWP models that are used to produce the first-guess forecast(s) and, when variational assimilation methods are employed, to perform the assimilation. By contrast, theory-based calculations involved in familiar observation and measurement usually involve equations that relate variables at a single time. The equation used to correct for a rain gauge's wind loss over a period requires information about wind speed during that period, not an earlier or later period. Likewise, a value for relative humidity at time *t* is calculated from values for temperature and pressure at *t*.

Is this a deep or important difference? Not necessarily. What matters is that results have desired accuracy, that is, that they come close enough to the true values for variables and parameters, if there are such true values to be found. Results from equations that relate variables at different times can be just as accurate as results from equations that relate variables at a single time—it depends on the equations. So, the real question seems to be whether, in practice, the forecasts relied upon in reanalysis are as accurate as results calculated in the course of familiar observation and measurement. We return to this issue below, after considering one other candidate difference.

ILL-POSED INVERSE PROBLEMS. A third candidate for an important difference between reanalysis datasets and familiar observations and measurements is that producing reanalysis datasets requires solving an ill-posed inverse problem, whereas producing familiar observations and measurements does not.

The challenge confronted in reanalysis is to reconstruct the three-dimensional atmospheric state at *t* from gappy observations made around *t*—an *inverse problem*. The information provided by those observations, in conjunction with background theory, is insufficient to uniquely determine the state estimate, so the problem can be described as *ill posed*. Data assimilation is a particular approach to solving the problem: it brings in additional information, in the form of one or more NWP forecasts, that provides a first-guess state estimate or prior. By contrast, in most familiar observing and measuring practices, there is no need for a first-guess estimate or prior; the instrument readings made at *t*, in conjunction with background theory, are thought to be sufficient to determine a best-estimate value for the parameter/variable of interest. This is reflected in the measurement model.

It is worth noting, however, that some gridded datasets that are uncontroversially described as "observational"—surface temperature datasets like the Goddard Institute for Space Studies Surface (GISS) Temperature Analysis (GISTEMP; Hansen et al. 2010) and the Climatic Research Unit Temperature, version 4 (CRUTEM4; Jones et al. 2012)—also are produced by solving an ill-posed inverse problem very similar to that solved in reanalysis; the problem is merely solved differently. A spatial interpolation approach is used to infer values for gridpoint variables as a function of nearby observations, which in some cases are rather distant. While this approach does not involve NWP models, it is not model free: it relies on assumptions about the smoothness and typical structure of atmospheric fields. Moreover, data assimilation is considered superior to spatial interpolation when it comes to producing analyses in real-time NWP operations; the first-guess forecast(s) used in the assimilation contains additional valuable information derived from observations made at earlier times, which informed the initial conditions for the forecast(s).

Are results obtained by solving ill-posed inverse problems less accurate than results obtained from observing and measuring practices that do not require solving such problems? The answer again is "not necessarily." If the first-guess estimate or prior used to solve the inverse problem is itself very accurate, then the final results obtained can be very accurate too. In practice, however, there may be reason to believe that a first-guess estimate or prior, while adding substantial valuable information, also contains some significant errors. As explained below, this is a key concern in reanalysis.

CALIBRATION AND UNCERTAINTY. A fourth candidate for an important difference is that, whereas there is good reason to think that familiar observations and measurements are rather accurate, the same cannot be said of today's reanalysis datasets.

While this is an overgeneralization, it does point to an important difference between some reanalysis results and observations. Manufacturers of thermometers and barometers have established methodologies for *calibrating* their instruments before they are shipped to users: the instruments can be expected to give results that are free from significant systematic error, at least for some period of use under specified operating conditions. In addition, users are informed of the typical remaining *uncertainty* associated with results: for example, that the thermometer gives readings with a 2σ accuracy of ±0.15°C.[2] In the case of reanalysis, however, calibration remains very challenging, and results often are provided without any uncertainty information.

A common way to calibrate an instrument or system is with the help of accurate reference standards: the instrument's results are compared to the standards and, if necessary, adjustments are made to ensure a close enough fit; the remaining deviations are indicative of the uncertainty associated with results. But for reanalysis systems, comprehensive reference standards—that is, accurate estimates of full atmospheric fields—generally are not available. (If they were, there might be little need for reanalysis!) Observations used to produce the reanalysis cannot provide an independent basis for comparison. High-quality observations that are not assimilated, such as those obtained in special field campaigns, can be useful for learning about errors in reanalysis results in a piecemeal way, but they are available for a limited range of variables and times. The same is true for cross-validation techniques, in which some of the observations that would otherwise be assimilated are reserved for estimating analysis errors (e.g., Thorne and Vose 2010; De Pondeca et al. 2011). Cross-validation techniques also either reduce the information base on which the reanalysis is built (if the reanalysis is not repeated with the reserved observations added) or else provide information about errors in different reanalyses than those that are ultimately of interest (if the reserved observations are subsequently assimilated).

An alternative approach to calibration involves correcting for component sources of systematic error. This was illustrated earlier with the rain gauge example, where raw gauge readings were corrected for wind loss; if other interfering factors—evaporation, splashing of drops, or wetting of gauge sides—were thought to be significant in a particular case, then corrections for those factors would need to be applied as well. In the case of reanalysis, this approach requires correcting for, among other things, systematic error in the observations and in the first-guess forecast(s), and systematic error in raw reanalysis results that arises because assumptions of the assimilation algorithm are not perfectly met; the latter can include assumptions about how to map between observation space and model space (the observation operator), about the topology of the cost function in variational assimilation methods (see Talagrand 2010), and so on. Identifying and correcting for these errors is difficult. Systematic errors in observations and the first-guess forecast(s) have received the most attention, but there are a number of obstacles: metadata are quite limited for many historical observations (e.g., Kennedy 2014), high-quality observations that might be used to quantify forecast errors are available only for a limited

[2] In general, *uncertainty* refers to a lack of knowledge. In the context of measurement, a report of uncertainty indicates the possible error in an estimated parameter value (see JGCM 2008, section 2.2, for other definitions). When calibration is successful, uncertainty stems primarily from random error or "noise" and is indicative of the *precision* of the measurement. In practice, calibration is often incomplete, and uncertainty estimates also reflect some systematic effects, perhaps unrecognized.

Table 1. What important differences are there between reanalyses and observations?

Candidate difference	Conclusion
Reanalysis results are obtained by inference, while observations are not.	Not a real difference; both observations and reanalyses involve inference, often involving theory-based calculation.
Reanalysis relies on forecasts, while observation does not.	A real difference but not necessarily significant; what matters is whether results are sufficiently accurate.
Reanalysis involves solving an ill-posed inverse problem, while observation does not.	A real difference in many cases but not necessarily significant; what matters is whether the solution is sufficiently accurate.
Accuracy of reanalysis results is less well understood than that of observations.	A real difference in some cases and a significant one; makes it harder to judge appropriate use of reanalyses in those cases.

set of variables and times, and forecast errors are expected to be regime dependent. Errors in the first-guess forecast(s) due to inadequate representation of physical processes in NWP models (i.e., model error) remain a particular source of concern. For further details and some proposals for ways forward, see, for example, Dee (2005), Desroziers et al. (2005), Dee and Uppala (2009), and Peña and Toth (2014).

Many of the obstacles to calibrating reanalysis results—such as the lack of independent reference standards—create similar challenges when it comes to quantifying the uncertainty associated with those results. In part because of this, many of today's reanalysis datasets are not accompanied by uncertainty estimates at all; only a best-estimate value is given for each gridpoint variable. (An exception is discussed below.) This is why many reanalysis results, if viewed as putative measurements of atmospheric properties, must be considered incomplete: in parallel to the view in meteorology that no forecast is complete without an estimate of forecast uncertainty (Tennekes et al. 1987; Zhu et al. 2002; NRC 2006), there is a view in metrology that *no measurement is complete without an estimate of measurement uncertainty* (e.g., JGCM 2008).

So some observations and reanalysis results differ in the following way: while the observations are produced using instruments that have undergone a careful process of calibration and that give results with uncertainties that can be confidently quantified and are relatively small, the errors and uncertainties associated with the reanalysis results remain less well understood and are likely to be large in some cases. But it is important not to overgeneralize; not all observations and measurements are accompanied by well-motivated uncertainty estimates, and for specific reanalysis variables (e.g., surface temperature in regions where there are many assimilated observations) there is good reason to think that the results are typically quite accurate. The mere fact that a result is an observation or that it is a reanalysis result does not tell us how accurate we should expect it to be.

THE IMPORTANCE OF UNCERTAINTY ESTIMATION. Providing information about errors and uncertainties associated with results—whether those results are observations or measurements, analyses or reanalyses, or even forecasts—is important. First and foremost, it is important for drawing appropriate conclusions from those results. For example, whether an apparent trend in time-series data (i.e., a linear fit with nonzero slope) is good evidence of a real change in conditions depends on whether the uncertainties associated with the data imply that the actual slope could easily be zero. Without uncertainty information, it is unclear what results provide evidence for (or against). That said, conclusions that are sensitive even to very small changes in results obviously are more suspect than conclusions that are robust to very large changes.

Related considerations motivate recent calls for the production of "climate quality" reanalyses. Thorne and Vose (2010) propose that reanalyses be considered of climate quality only if we can confidently estimate their uncertainties to be less than 10% of the expected multidecadal climate change signal (as indicated by a suite of climate models) across a small range of important physical indicators, such as temperature, large-scale precipitation, etc. But while Thorne and Vose are right to emphasize robust uncertainty quantification, it is unclear what benefit is to be had by adding the "climate quality" label. On the contrary, such a label might obscure the fact that results for some variables or fields have larger (or unknown) errors and uncertainties.[3] It seems better

[3] Just as saying that a model is a "good" model—rather than saying what it is good enough for—can lead users to trust the model even where it is misleading, so could labeling some reanalyses "climate quality" lead to misplaced trust.

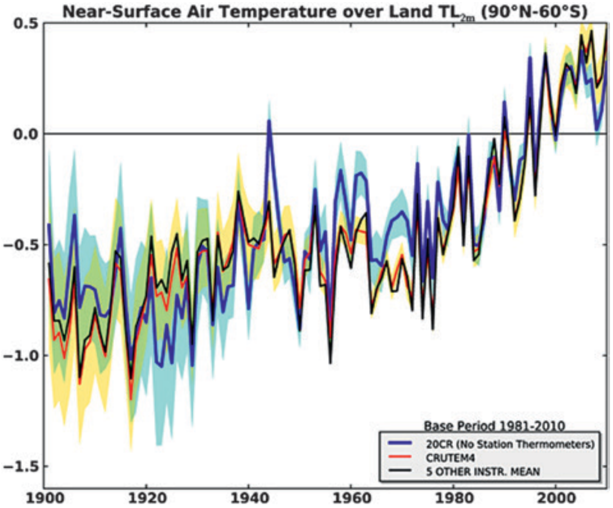

Fig. 1. Evolution of near-surface temperature anomaly over land from 20CR (blue), the station-based CRUTEM4 dataset (red), and the average of five other station-based datasets (black). The 95% uncertainty ranges for 20CR (blue shading) and CRUTEM4 (yellow shading), as well as areas of overlap (green shading), are indicated (after Fig. 1 of Compo et al. 2013).

to avoid such labels and to simply provide explicit uncertainty information.

Information about uncertainties is also important for evaluating whether techniques, instruments, and systems for producing results are working as expected. A set of results from different measuring techniques or instruments need not agree perfectly in their best-estimate values for parameters/variables, but if all goes well the differences between them should not exceed the differences that are expected, given their respective confidence intervals. Coming close to achieving this statistical consistency is a hallmark of successful measurement (Tal 2012). It is no guarantee that we are close to the true value of a parameter or variable, if there is a true value to be found, but a lack of consistency indicates that something has gone wrong somewhere. Without uncertainty estimates, it is impossible to check for consistency.

A good illustration of this sort of consistency check is found in a recent study involving the Twentieth-Century Reanalysis (20CR), a global reanalysis covering the period 1871–2011 (Compo et al. 2011). 20CR is one of the few reanalyses that does provide uncertainty estimates; they are generated by 20CR's ensemble Kalman filter assimilation methodology, which gives results in probabilistic form. Compo et al. (2013) compared global surface temperature changes derived from 20CR results with those derived from station-based thermometer data that were not assimilated in 20CR. They found that, while the datasets show rather good agreement (see Fig. 1), rigorous statistical consistency has not yet been achieved: "the mean square differences between [the datasets] are somewhat larger than expected from their respective confidence intervals… This suggests that the data sets underestimate their uncertainty, particularly 20CR during the periods of disagreement" (Compo et al. 2013, 3171–3172). This underestimation of uncertainty may stem in part from the calibration-related challenges noted in the last section, which in 20CR are partially addressed with the help of some "rather simplistic" (Compo et al. 2011, p. 21) assumptions about errors and uncertainties in the assimilated observations and forecasts.

Of course, even when reanalysis results lack rigorous uncertainty estimates, there can be useful insight into their accuracy. Some reanalysis results are known to be determined primarily by observations that we can expect to be rather accurate. In addition, it is known that the NWP models used in reanalysis tend to give more accurate forecasts for some physical quantities than for others; for example, temperature is typically easier to forecast than precipitation. De Pondeca et al. (2011) note the possibility of providing a "quality mark" array for precipitation analyses, communicating factors that can help users gauge the trustworthiness of results at each grid point (e.g., whether the grid point is outside an effective radar coverage area).[4] Comparing reanalysis datasets with one another is also useful, since the spread among results indicates, at least prima facie, a lower bound on uncertainty (see also, e.g., Buizza et al. 2005; Langland et al. 2008; Wei et al. 2010). The reanalyses.org website serves as a central clearinghouse for a range of intercomparison efforts and provides other useful resources, including plotting tools.

CONCLUSIONS. It is tempting to think that what justifies warnings about reanalysis datasets is the fact that they are determined in part by NWP forecasts. There is some truth in this, but it is not reliance on these forecasts per se—the results of theory-based calculation—that should prompt concern; after all, many trusted observations and measurements also are produced with the help of theory-based calculations. The real issue is that the errors and uncertainties

[4] Practitioners tend to treat as less reliable those reanalysis fields that are derived from the model's state variables and for which no conventional observations are assimilated [see, e.g., the distinction among the A, B, and C fields in Kalnay et al. (1996)], but this is just a reasonable rule of thumb. Information at the level of gridpoint variables, such as the arrays suggested by De Pondeca et al. (2011), would be more useful.

associated with NWP forecasts, and with some other aspects of the reanalysis process, are only partially understood, leaving it unclear just how accurate we can expect many reanalysis results to be (see Table 1). This is in contrast to some familiar observations and measurements, such as those made using today's standard thermometers and barometers, which have undergone a careful process of calibration, with uncertainties that can be confidently quantified and are relatively small. This is not, however, a universal difference between reanalysis data and observations; many observations are also reported without well-motivated uncertainty estimates.

For any type of result—an observation or measurement, an analysis or reanalysis, a forecast—understanding of errors and uncertainties is crucial for drawing appropriate conclusions about the system under investigation. Rather than an optional afterthought, characterization of uncertainties should be considered part and parcel of the processes of observation, measurement, data assimilation, and forecasting. Where uncertainty estimation cannot be performed in a rigorous and quantitative way, it may still be possible to provide information (e.g., quality mark arrays) that will aid users in judging the relative trustworthiness of conclusions suggested by results. Increased efforts to provide such information for reanalyses, and to provide well-motivated quantitative uncertainty estimates where possible, would be of significant value.

ACKNOWLEDGMENTS. Thanks to Leonard Smith, Reto Knutti, Gavin Schmidt, Zoltan Toth, and two anonymous BAMS referees for their helpful feedback. This material is based upon work supported by the U.S. National Science Foundation under Grant SES-1127710.

REFERENCES

Bengtsson, L., and J. Shukla, 1988: Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteor. Soc.*, **69**, 1130–1143, doi:10.1175/1520-0477(1988)069<1130:IOSAIS>2.0.CO;2.

Bosilovich, M. G., J. Kennedy, D. Dee, R. Allan, and A. O'Neill, 2013: On the reprocessing and reanalysis of observations for climate. *Climate Science for Serving Society: Research, Modeling and Prediction Priorities*, A. Ghassem and J. W. Hurrell, Eds., Springer, 51–71, doi:10.1007/978-94-007-6692-1_3.

Boumans, M. J., 2006: The difference between answering a 'why' question and answering a 'how much' question. *Simulation: Pragmatic Construction of Reality*, J. Lenhard, G. Küppers, and T. Shinn, Eds., Sociology of the Sciences Yearbook, Vol. 25, Springer, 107–124.

Bromwich, D. H., Y.-H. Kuo, M. Serreze, J. Walsh, L.-S. Bai, M. Barlage, K. Hines, and A. Salter, 2010: Arctic system reanalysis: Call for community involvement. *Eos, Trans. Amer. Geophys. Union*, **91**, 13–14, doi:10.1029/2010EO020001.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, doi:10.1175/MWR2905.1.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137A**, 1–28, doi:10.1002/qj.776.

——, P. D. Sardeshmukh, J. S. Whitaker, P. Brohan, P. D. Jones, and C. McColl, 2013: Independent confirmation of global land warming without the use of station temperatures. *Geophys. Res. Lett.*, **40**, 3170–3174, doi:10.1002/grl.50425.

Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323–3343, doi:10.1256/qj.05.137.

——, and S. Uppala, 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.*, **135**, 1830–1841, doi:10.1002/qj.493.

——, and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

——, J. Faullo, D. Shea, J. Walsh, and NCAR staff, Eds., 2015: The climate data guide: Atmospheric reanalysis; Overview and comparison tables. Accessed 30 March 2015. [Available online at https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.]

De Pondeca, M. S. F. V., and Coauthors, 2011: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612, doi:10.1175/WAF-D-10-05037.1.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, doi:10.1256/qj.05.108.

Gleckler, P. J., K. A. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.

JCGM, 2008: Evaluation of measurement data—Guide to the expression of uncertainty in measurement. Working Group 1 of the Joint Committee for Guides in Meteorology, JCGM 100:2008, GUM 1995 with Minor Corrections, 122 pp.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, **117**, D05127, doi:10.1029/2011JD017139.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 341 pp.

——, and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52**, 1–32, doi:10.1002/2013RG000434.

Kidston, J., D. M. W. Frierson, J. A. Renwick, and G. A. Vallis, 2010: Observations, simulations, and dynamics of jet stream variability and annular modes. *J. Climate*, **23**, 6186–6199, doi:10.1175/2010JCLI3235.1.

Kravtsov, S., M. G. Wyatt, J. A. Curry, and A. A. Tsonis, 2014: Two contrasting views of multidecadal climate variability in the twentieth century. *Geophys. Res. Lett.*, **41**, 6881–6888, doi:10.1002/2014GL061416.

Langland, R. H., R. N. Maue, and C. H. Bishop, 2008: Uncertainty in atmospheric temperature analyses. *Tellus*, **60A**, 598–603, doi:10.1111/j.1600-0870.2008.00336.x.

Mari, L., 2005: Models of the measurement process. *Handbook of Measuring Systems Design*, P. Sydenman and R. Thorn, Eds., Vol. 3, Wiley, 1–4.

NRC, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts.* National Academies Press, 124 pp., doi:10.17226/11699.

Peña, M., and Z. Toth, 2014: Estimation of analysis and forecast error variances. *Tellus*, **66A**, 21767, doi:10.3402/tellusa.v66.21767.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233, doi:10.1256/qj.05.129.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Santer, B. D., and Coauthors, 2004: Identification of anthropogenic climate change using a second-generation reanalysis. *J. Geophys. Res.*, **109**, D21104, doi:10.1029/2004JD005075.

Schmidt, G., 2011: Reanalyses 'R' us. Accessed 31 March 2015. [Available online at www.realclimate.org/index.php/archives/2011/07/reanalyses-r-us/.]

Tal, E., 2012: The epistemology of measurement: A model-based approach. Ph.D. dissertation, University of Toronto, 188 pp.

Talagrand, O., 1997: Assimilation of observations: An introduction. *J. Meteor. Soc. Japan*, **75**, 191–209.

——, 2010: Variational assimilation. *Data Assimilation: Making Sense of Observations*, W. A. Lahoz, B. Khattatov, and R. Ménard, Eds., Springer-Verlag, 41–67, doi:10.1007/978-3-540-74703-1_3.

Tennekes, H., A. P. M. Baede, and J. D. Opsteegh, 1987: Forecasting forecast skill. *Workshop on Predictability in the Medium and Extended Range*, Reading, United Kingdom, ECMWF, 277–302.

Thorne, P., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends. *Bull. Amer. Meteor. Soc.*, **91**, 353–361, doi:10.1175/2009BAMS2858.1.

Trenberth, K. E., and J. G. Olson, 1988: An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium Range Weather Forecasts. *Bull. Amer. Meteor. Soc.*, **69**, 1047–1057, doi:10.1175/1520-0477(1988)069<1047:AEAIOG>2.0.CO;2.

Wei, M., Z. Toth, and Y. Zhu, 2010: Analysis differences and error variance estimates from multi-centre analysis data. *Aust. Meteor. Oceanogr. J.*, **59**, 25–34.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73–83, doi:10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2.