REANALYSES SUITABLE FOR CHARACTERIZING LONG-TERM TRENDS

Are They Really Achievable?

BY P. W. THORNE AND R. S. VOSE

To facilitate long-term trend studies, we need a substantial reappraisal of the overall reanalysis strategy, with attention to methodology, verification, and uncertainties, among other things.

A reanalysis is a fixed numerical weather prediction (NWP) system run in hindcast rather than forecast mode, ingesting the available historical observations presented to it (left-hand side of Fig. 1 describes the typical steps). For climate monitoring and research applications, it has several distinct potential advantages over more traditional climate datasets in that it synthesizes all observations in a manner consistent with model (and therefore atmospheric) physics: it provides complete spatial and temporal coverage with physical rather than statistical interpolation into data void regions, and it provides information on unobservable parameters (e.g., potential vorticity). Reanalysis centers make these data available as gridded fields (e.g., temperatures at the surface and pressure levels, vorticity, precipitation, etc.) for bona fide research. They also save ancillary information, which has proven useful for some applications (e.g., Haimberger et al. 2008).

That the paper describing the first multidecadal reanalysis system (Kalnay et al. 1996), expanding on earlier shorter-period efforts (Bengtsson et al. 1982; Schubert et al. 1993), constitutes the most cited paper in climate science over the last decade attests to

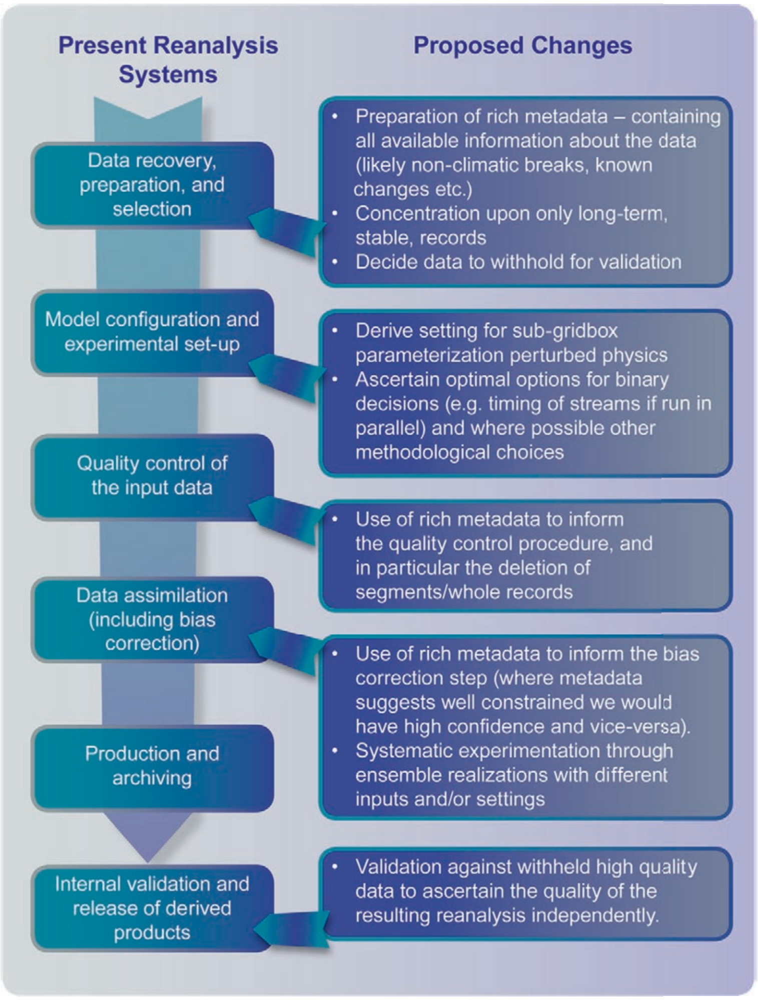

FIG. 1. (left) Flow diagram of typical steps currently undertaken in a reanalysis and (right) a very brief synopsis of those aspects that are either entirely new or relatively novel that we are proposing here. See text for further details.

their high utility to a large range of applications. Since then, many reanalyses using a variety of methodological assumptions have been completed (Uppala et al. 2005; Onogi et al. 2007), are in process (see online at http://gmao.gsfc.nasa.gov/merra/; Bosilovich et al. 2006; Kistler et al. 2008; see online at www.ecmwf.int/research/era/do/get/era-interim), or are in the planning stage.

However, the reanalyses completed to date have undesirable and in some cases very obvious (e.g., Fig. 1 in Bosilovich et al. 2006) and unphysical time-varying biases, which at best reduce their utility for long-term trend monitoring (but not real-time monitoring, for which they are an undoubtedly valuable tool) and at worst make them useless for such activities, depending on the region, variable of interest, and application (e.g., Bengtsson et al. 2004; Karl et al. 2006; Thorne 2008). This is definitively not to say that they are inadequate everywhere or useless for all such long-term behavior applications (e.g., Simmons et al. 2004, 2009), but great care must be taken, care that cannot be assumed of end users, occasionally resulting in high-profile findings upon which considerable doubt has been cast (e.g., Graversen et al. 2008; Thorne 2008; Grant et al. 2008; Bitz and Fu 2008).

It would be wrong not to explicitly recognize two things at this point in our discussion: 1) that very considerable efforts are made in each reanalysis to minimize nonclimatic influences and the results are much better than they would otherwise have been and 2) that reanalyses were never primarily constructed to be long-term homogeneous (free of nonclimatic influences) records but rather to provide the best possible analysis at each time step. The presence of residual inhomogeneities certainly serves to illustrate the enormity of the challenge facing those undertaking a reanalysis effort and the substantial time, effort, and resources required. Arguably, it also reflects upon the desire to get a best analysis at each time step, which we contend herein may not be the optimal approach for a long-term homogeneous product.

The presence of residual nonclimatic behavior in reanalyses to date has led several to call for "climate quality" reanalyses to be created that explicitly retain long-term fidelity (Karl et al. 2006; Bengtsson et al. 2007; Uppala et al. 2008). Although welcome, these calls are often contradictory both in their definition of what climate quality would entail and how methodologically one would go about achieving it. We believe that this is because, to date, the basic problem of creating a long-term homogeneous product has been poorly documented and has not been posed in a straightforward way. We aim here to clarify the issue in the hope that we can engender a robust methodological framework against which rational decisions can be made to advance toward this goal in as expeditious a manner as possible.

We largely concentrate on atmosphere-only reanalyses here to make the issues more intuitive to the reader. The coupled ocean–atmosphere reanalysis problem, which is just starting to be pursued (Kistler et al. 2008), will have to overcome essentially the same issues. However, it faces the additional complications of addressing how to interface the two components and how to cope with the even more dramatic shifts in subsurface ocean data availability and quality through time than is the case for atmospheric observational data (Willis et al. 2007; Gouretski and Koltermann 2007).

DEFINING THE CLIMATE QUALITY REANALYSIS PROBLEM. Methodologically, we believe that the reanalysis output at each time step can most easily be considered as follows:

$$R = t + f(O_e, M_e, A_e, B_e), \quad (1)$$

where R is the reanalysis output; t is the true climate system state, which is unknown; and the terms in the parentheses are error terms, which are outlined in this section. These error terms will interact in a nonlinear way, making it mathematically impossible after the event to unambiguously disentangle their contributions and thereby retrieve t or from a single reanalysis to unambiguously ascertain what the causes of any apparent nonclimatic effects in that analysis are.

The observational error term O_e incorporates the spatiotemporal incompleteness of the observational field and any absolute biases that exist in the observations. These biases are very likely to incorporate

AFFILIATIONS: Thorne—Met Office Hadley Centre, Exeter, United Kingdom; Vose—NOAA/National Climatic Data Center, Asheville, North Carolina
CORRESPONDING AUTHOR: Peter Thorne, Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, United Kingdom
E-mail: peter.thorne@metoffice.gov.uk

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/2009BAMS2858.1

In final form 19 August 2009

random and systematic components and will almost certainly vary both on synoptic time scales and over the longer term with instrument changes, drifts, algorithm changes, etc. The best-documented issues with early reanalysis efforts are apparent changes in global or large-scale mean properties coinciding with the large observational shocks associated with the introduction of new satellite observations (Pawson and Fiorino 1999; Bengtsson et al. 2004). However, this is simply indicative of O_e issues that must pertain at all scales. In this context, the nonclimatic influences in in situ point observations will likely prove to be a subtle and important problem to solve. This is particularly so because we are switching our attention toward changes at smaller space and time scales as scientific focus moves toward impacts, extreme events, and adaptation and away from global-mean changes and their probable causes.

The model error term M_e encompasses any physical shortcomings of the NWP model being used. For example, in many models, the troposphere has a substantial cold bias (John and Soden 2007). The term also incorporates the fact that, even at NWP scales, a large number of physical processes that occur in the real atmosphere need to be parameterized. The Quantifying Uncertainties in Model Predictions (QUMP) project applied to a climate model attests to the very real impact that these uncertain choices can have (Murphy et al. 2004), although such effects will of course be mitigated by constant nudging of the field by observations.

The unintentional errors that can be imparted because of the chosen methodological approach are represented by A_e. There are a number of essentially subjective decisions that must be made in the process of setting up and running a reanalysis system that are clearly independent of either the model error or the observational error in the strictest sense. For example, if the reanalysis needs to be run in parallel streams (e.g., 1979→, 1985→, etc.), when should these streams be started and how much overlap should there be? At what model spatial resolution should the reanalysis be run? What, if any, variational bias correction should be applied? What variational assimilation scheme should be used? In addition, A_e will include a number of decisions that initially seem to be under the previous two classes [i.e., run with slightly perturbed sea surface temperature (SST) boundary, run without Microwave Sounding Unit (MSU; O_e) data, and run with tweaked model physics (M_e)] but are also decisions that in reality we have made subconsciously. At least some of these decisions will have a substantial impact on certain characteristics of the output, as has been conclusively shown to be the case in climate dataset construction (Thorne et al. 2005a; Titchner et al. 2009).

The last error term is the background error, which is the error in the forecast field from the last analysis step and therefore by logical inference must be a time integral of a priori unknown form of the remaining error terms. That is, in a perfect reanalysis system presented with perfect, globally complete, error-free input data, this term would be zero. Therefore, we do not discuss it further herein.

SOME SUGGESTIONS FOR ATTAINING CLIMATE QUALITY REANALYSES. We do not presume to have definitive answers as to how to make a climate quality reanalysis. However, in view of the previous considerations, there immediately arise several possible avenues to improve long-term homogeneity that we believe should be considered and in some cases are already being actively considered (we note cases of which we are aware, a list that is unlikely to be comprehensive).

Minimizing observational errors. Selectively thin the ingest data field to minimize observational shocks. For the in situ record, short records and changes in observing technology (e.g., introduction of radiosondes) will impact the analysis and hence potentially homogeneity if they are included, particularly for remote sites and in the presatellite era when only in situ data are available to constrain the analysis. Logically, long-term continuity will likely be best gained through considering a stable long-term subset of the in situ network rather than incorporating short or intermittent records, particularly because these records are likely to be poorly studied (see next section). The twentieth-century reanalysis (Compo et al. 2006), which ingests solely surface observational data (specifically surface pressure and SST), goes some way toward this end but still has an order of magnitude increase in station count over the period of record that is likely to have an impact on homogeneity.

For the satellite record, it would likewise make sense to only ingest the long-term operational satellite radiances and not those from shorter-term experimental platforms that drop in and out of the record. This will particularly be the case above the upper troposphere, where in situ observations are either absent or of dubious quality (Karl et al. 2006) and the climate system is largely physically decoupled from the surface observations. At the very least, because these operational satellites form the backbone of the operational system, this provides a necessary focus for our efforts to clean up and understand the satellite component of the input data. The planned Modern Era Retrospective-analysis for Research and Applications (MERRA) Reduced Observing System Baseline (ROSB) run (Bosilovich et al. 2006) product will go a significant way toward achieving this aim.

Notwithstanding the previous considerations, we do recognize that because the observational system has fundamentally changed over time, particularly away from the surface, we need to permit some degree of varying input field. We are proposing minimizing this effect through only undertaking data rescue for and ingesting long-term or vital components of each data source. A strategy of pursuing every scrap of information may serve to divert resources from other more tractable areas of investigation and to add unnecessary inhomogeneities.

Recast the data assimilation step to utilize raw data and all available metadata. Before addressing how, a note of caution is required regarding an alternative and intuitively appealing strategy of ingesting homogenized datasets instead of the raw data and metadata as a long-term solution. We note that European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA)-Interim and MERRA (Bosilovich et al. 2006) already use a homogenized radiosonde temperature product (Haimberger et al. 2008), and such a strategy is likely better than using raw data in such cases where the raw data contain gross biases. However, although a single homogenized version of the data will, on average at least, likely be better than the raw observations, it will not be free of error. All of these climate datasets are based largely or wholly on statistical rather than physical inferences. Worse, residual error will tend to be systematic in nature, particularly where some form of geographical background expectation field has been implicitly or explicitly used in finding and adjusting for nonclimatic breakpoints in the series, as is almost always the case. It is also likely to project most strongly on precisely those long time scales of most interest in creating a climate quality reanalysis (Thorne et al. 2005a). The effect of using bias-adjusted datasets therefore is to substantially modify the O_e error term, making the O_e covariance matrix less diagonal (an implicit assumption in the assimilation schemes) and hence potentially confounding the reanalysis system if many such datasets with insidious residual errors are utilized. Furthermore, choosing one dataset precludes the use of information from the others where multiple datasets exist, and we have no a priori robust reason in general for such a selection.

We contend instead that a combination of the raw data, the available metadata, and the background departure of the observation can be used to make a physically sensible decision in the assimilation step that minimizes the impact of poor quality or systematically biased observations. The data assimilation step produces an optimal (minimum absolute error) blend of the background field from the last time step and all available observations presented to it from a diverse range of sources (surface, satellite, radiosonde, etc.). Typically nowadays, it uses a four-dimensional variational data assimilation (4DVAR) approach whereby all data within ± a specified period are ingested and (increasingly common) a variational bias adjustment to the data (e.g., Desroziers et al. 2005; Li et al. 2009a; Dee 1995) is made. The ERA-Interim product has applied such a bias adjustment scheme to the satellite radiance data (Dee and Uppala 2009). However, these schemes at present have solely the model analysis increments and the combination of raw data presented at their disposal. Generally, they also make assumptions regarding the distribution of observational errors, which are not strictly valid in the real world (Desroziers et al. 2005).

There is a substantial heritage of efforts to investigate the homogeneity of much of the data that are ingested and form climate data records from there (e.g., Durre et al. 2006; Thorne et al. 2005b; Free et al. 2005; Sherwood et al. 2008; Haimberger et al. 2008; Rayner et al. 2003; Mears and Wentz 2009) as well as feedback files (observations minus background expectations) from previous reanalysis efforts (Haimberger et al. 2008). We believe that the resulting diverse "rich metadata" could and should be used in combination with the raw data to inform the data assimilation step. Most of these data relate to the long-term average biases present; it is of limited utility to the quality control of individual point observations, but it may help in any quality control blacklisting of segments of series and/or entire series. The main benefit would largely derive from informing the estimation of the data bias within the assimilation step. The utility of using multiple pieces of independent evidence of likely bias structure is lost when using either the raw data or a homogenized dataset in isolation, but it clearly has massive potential to impact the long-term reanalysis homogeneity by helping to inform where biases are robustly known and where they are not.

By recasting the assimilation step in this way, we are taking advantage of the reanalysis system being an optimal physically consistent blending procedure and effectively asking it to build on all of our knowledge regarding the data and its adequacy (rich metadata), rather than the very small subset of this information that has been presented to it to date. We are very likely to have been handicapping the present reanalysis systems from retaining trend fidelity by not allowing them to make optimal use of all potential information on data provenance. If such an alternative strategy is to be followed, then the major data effort required will be on collation of the metadata describing our knowledge of the raw data from numerous past studies. Specific modifications required to the data assimilation scheme are not an area that the authors have the expertise to discuss further. However, it may logically infer a more temporally relaxed 4DVAR requirement than is typical to NWP schemes.

Minimizing model errors. WHERE LARGE-SCALE INGEST FIELD CHANGES ARE INEVITABLE, UNDERSTAND HOW THESE IMPACT M_e. The model error term will be most important when there are substantial observational "shocks" to the system. For example, Fig. 1 of Bosilovich et al. (2006) shows the dramatic change in precipitation in the 25-yr Japanese Reanalysis (JRA-25) system after the introduction of Special Sensor Microwave Imager (SSM/I) satellite data. A first step is to run observing system experiments (OSEs) over the period of major observational shocks with and without the new data to ascertain their impact (or a parallel suite with a more fixed observational constraint; Bosilovich et al. 2006). Indeed, there have already been substantial efforts to understand the model error term in a data assimilation context (e.g., Dee and da Silva 1998; Danforth et al. 2007; Li et al. 2009b). This, however, presages a harder choice as to how to account for the effects of an inevitably changing data mask. One way, for example, would be to nudge the analysis with simulated Radiative Transfer for Television and Infrared Observation Satellite Operational Vertical Sounder (RTTOV) satellite radiances calculated from the previous time step (Matricardi et al. 2004) when satellite data are absent. Another way would be to remove the average impact when they are present. Neither way may work in practice.

Neither of these is particularly satisfactory; clearly, the more advantageous way forward would be to minimize M_e before running the reanalysis by learning from the OSEs and if necessary changing the model configuration to a setup that minimizes the effect of the shocks. Modern NWP models, from which reanalyses are derived, are inevitably (subconsciously) configured in such a way as to minimize the model error term with a present-day observation-rich input to support operational weather forecasting. It does not logically follow that this will be an optimal setup to minimize the long-term biases in reanalyses that must cope with a substantially evolving observational constraint. Therefore, a different basic model configuration than that employed in modern-day NWP may well be required to minimize nonclimatic biases in a run extending from radiosondes to early satellite through to a modern-day data-rich observational mask.

ASCERTAIN WHICH SUBGRID-SCALE PARAMETERIZATIONS HAVE THE LARGEST IMPACT USING DETUNED NWP VERSIONS. The M_e term could also be better understood by running, likely at reduced model spatial resolution, an ensemble of OSEs where the subgrid-scale parameterizations are detuned in a similar vein to the QUMP experiment for climate models (Murphy et al. 2004). This would ascertain which of these parameterizations had the greatest impact in a reanalysis setting. Those that were found to have an impact on the analysis behavior could then be optimized to minimize their impact before running the full reanalysis system. Although desirable to run over the whole period, running such a suite over one or several carefully selected subperiods of the final reanalysis could reduce the large overhead such a strategy would likely introduce.

Minimizing methodological errors. A necessary first step to minimizing A_e is an honest appraisal of all the ad hoc decisions that have to be made that may impart biases to the final reanalysis output. These should then be assessed in a systematic manner and documented wherever possible. Two fundamental classes of choices are bound to arise: those for which an unambiguous optimal value can be ascertained and those for which only a range of plausible settings can be found.

CHOOSE BEST SETTINGS WHEREVER POSSIBLE. That decisions and algorithms play a role in differences between analyses is well known (e.g., Lorenc and Hammon 1988; Hollingsworth et al. 1985). Factors that are largely binary decisions, such as the choice of data stream timing, should be assessed carefully and chosen such as to minimize the chances of problems arising. For example, one 40-yr ERA (ERA-40) stream was started around the timing of the *National Oceanic and Atmospheric Administration-9* (*NOAA-9*) satellite—known to be an issue in long-term record continuity (Karl et al. 2006)—and this negatively impacted the resulting reanalysis (Uppala et al. 2005). Through a careful and critical assessment of all such potential issues prior to running a reanalysis system, their potential impact can be minimized or eliminated.

It may also be possible to optimize some further decisions that are more of a continuum of possibilities. For example, how long an in situ station must report to be incorporated in the data ingest is a methodological approach decision. This could, at least in theory, be optimized through appropriately targeted OSEs.

Use an ensemble approach to bracket remaining uncertainties in a robust manner. It is unlikely that all choices that may impart biases can be optimized. It would therefore seem logical, as resources permit, to run an ensemble where a range of seemingly sensible decisions is made for each choice. This is fundamentally distinct from an ensemble Kalman filter approach where the aim is to assess random rather than systematic error sensitivity and initial conditions are only relatively subtly changed. Instead, any ensemble would need to span long-term structural uncertainties in both models and observations that can affect the long-term system evolution. The ensemble would logically include perturbed model physics (Murphy et al. 2004; see earlier discussion), different plausible quality control and assimilation schemes, different equiprobable realizations of boundary forcings such as SSTs, and different withholding of data (e.g., some members not ingesting MSU satellite data). Careful experimental design would be required to optimize the utility that would result to explore the interdependency between sources of error.

It is probably only through such an ensemble approach that we can build the confidence in what the robust features are and what we should treat with extreme caution. At least in theory there is no reason why such an ensemble needs to be run at the fine spatial resolution of the flagship reanalysis, as long as the effect of coarser resolution can be explicitly quantified. This is similar to and compliments the idea of a "fleet" of model resolutions rather than a single flagship model for climate models for different applications under active development at the Met Office Hadley Centre and in the World Climate Research Program strategic framework (available online at http://wcrp.ipsl.jussieu.fr/). In fact, running a reanalysis mode across such a fleet would probably yield useful information as to the viability and limitations of such an approach that would compliment the runs in data assimilation free mode. These ensembles would not necessarily have to span the whole reanalysis period to be useful, although this would clearly be desirable. They would also undoubtedly have the additional benefit of pointing to further development pathways for the next generation of reanalyses.

DEFINING CLIMATE QUALITY ACCEPTANCE CRITERIA.

Finally, in the absence of formal criteria for a label of climate quality to be attached to a given reanalysis, there will always be arguments about whether this has been met. To expect any dataset either from classical observations only or from a reanalysis to be perfect in every aspect is unrealistic given that the observing system is not spatially or temporally complete, it was never designed for climate, and spurious nonclimatic influences are ubiquitous in all except perhaps a handful of National Observatory/Atmospheric Radiation Measurement Program (ARM) types of facilities. Thus, the climate quality assessment must include some leeway to account for this. Because these criteria would require cross-community sign on, it is inappropriate to be prescriptive here, but some general discussion is warranted.

We propose that a reanalysis be accepted as climate quality if its uncertainty is robustly constrained to be less than 10% of the expected multidecadal climate change signal (presumably from the average of a suite of climate model runs) across a small range of quasi-orthogonal indicators that encompass the climate system behavior. These indicators may reasonably be expected to include directly relevant impacts such as temperature and large-scale precipitation. So, for example, if a multiclimate model mean expected trend is 0.2 K decade^{-1} in global surface temperature, then the global surface temperature would need to be known within ±0.02 K decade^{-1}. For the observations themselves, 10% of the expected signal is typical of requirements tables (e.g., GCOS 2007). By requiring uncertainty to be constrained to within 10% of the expected emerging signal of climate change, the data can be used with a high degree of confidence that false positive conclusions about the causes of changes and event attribution can be largely avoided.

Selection of the validation indicators will undoubtedly depend on the validation sets that are available. The obvious candidate sets for validation are those observations and climate datasets that are deliberately withheld from the reanalyses. Given the richness and quality of the data from research satellite missions, intensive field campaigns, and National Observatory/ARM program data, these data would seem ideal for such purposes. Some current climate datasets may also qualify if their quality can be robustly ascertained. The amount of work required to identify and collate such an independent validation set should not be underestimated, nor should the fact that such data will be increasingly sparse prior to the satellite era and therefore we may always have to live with significant ambiguity in all except the most recent portion of any reanalysis record. The use of an ensemble approach, if undertaken, may minimize this impact and in itself prove an acceptable robust uncertainty estimate.

The use of validation data raises an interesting contention, raised by internal and external reviewers, as to whether it would be more valuable if ingested by the reanalysis system. We believe that fundamentally you need out-of-sample data to perform validation and that validation leads to much greater intrinsic value. Using within-sample data to validate any approach is well known to be highly dangerous and statistically invalid.

CONCLUSIONS. In recasting the reanalysis problem in a methodologically tractable way, we conclude that it is unlikely that continuing with the current strategies will produce a true climate quality reanalysis any time soon. Current reanalyses consist of a combination of the true climate signal and some nonlinear interaction among observational errors, model errors, and errors arising through the choice of methodological approach. With this in mind, a number of obvious potential avenues (that are fundamentally different to the current approaches in many but not all cases) arise to maximize the true climate signal component in the final reanalysis and robustly bracket the error terms (see also Fig. 1):

1) Restrict the data input solely to longer-term and/or essential components of the observing system rather than trying to ingest all data.
2) Ingest the raw data and all rich metadata regarding those data (quality control, homogenization, feedback files, instrumentation metadata, etc.) and utilize this in an optimal way in the data assimilation step.
3) Where large-scale changes to the input data are unavoidable, understand and minimize their impact on the model behavior ahead of time.
4) Understand which model parameterizations have the largest impact and optimize these to minimize their effects.
5) Critically assess the methodological assumptions that underpin the reanalysis and minimize their impact.
6) Run an ensemble of reanalyses at reduced resolution to address those methodological choices that have no rigorous basis and to understand strengths and limitations by establishing quantitative uncertainty estimates.

There are likely to be other possibilities that we have not considered here. Because our suggestions in many cases are fundamentally different from current strategies and will, on average, yield a poorer instantaneous field than current products, there is a strong case for creating two mutually exclusive classes of reanalysis: one that retains long-term fidelity and one that gives the best instantaneous field estimate. Together, these would encompass the range of expected applications, whereas it is extremely doubtful that a single reanalysis approach can optimally serve all constituencies. Certainly, splitting the requirement in this way would simplify the seemingly fundamental issue of current reanalyses having to try to fulfill multiple very different purposes.

The necessary complement to undertaking a fundamentally different reanalysis approach is to construct a set of robust acceptance criteria by which the scientific community can accept that a reanalysis result is truly climate quality. The obvious approach here is to define a set of quasi-orthogonal, societally important metrics to assess against and build a set of independent data for each of these. Assuming that a reduced-input dataset approach is pursued, there should be adequate data from shorter-term records from field campaigns, research satellites, and in situ observations as well as a handful of high-quality academic research/National Observatory/ARM types of records from which to construct such a validation database. A final decision on a set of metrics, acceptance criteria, and validation data is more appropriately decided upon by broad community involvement. However, we would argue that knowing the climate trend within 10% of the expected climate change signal would be an ideal final target.

ACKNOWLEDGMENTS. PWT was supported by the Joint DECC, Defra, and MoD Integrated Climate Programme [(DECC) GA01101, (MoD) CBC/2B/0417_Annex C5]; by the Met Office Government Business PHEATS project; and with travel support from NOAA/NCDC. Leo Haimberger, Roger Saunders, and Peter Stott provided comments on early drafts. Feedback from various Met Office staff and NCDC staff when given as a seminar sharpened the text. Three anonymous reviewers provided in-depth comments, which considerably aided in redrafting the text. David Parker and Roger Saunders provided help in formulating resulting changes. Deb Misch of NOAA/NCDC's graphics team provided considerable help in formulating the figure.

REFERENCES

Bengtsson, L., M. Kanamitsu, P. Kållberg, and S. Uppala, 1982: FGGE 4-dimensional data assimilation at ECMWF. *Bull. Amer. Meteor. Soc.,* **63,** 29–43.

——, S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.,* **109,** D11111, doi:10.1029/2004JD004536.

——, and Coauthors, 2007: The need for a dynamical climate reanalysis. *Bull. Amer. Meteor. Soc.,* **88,** 495–501.

Bitz, C. M., and Q. Fu, 2008: Arctic warming aloft is data set dependent. *Nature,* **455,** E3–E4.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.,* **87,** 175–190.

Danforth, C. M., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev.,* **135,** 281–299.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

——, 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.,* **131**, 3323–3343.

——, and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

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

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.

Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate,* **19,** 53–68.

Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series. *J. Geophys. Res.,* **110,** D22101, doi:10.1029/2005JD006169.

GCOS, 2007: GCOS Reference Upper-Air Network (GRUAN): Justification, requirements, siting and instrumentation options. WMO Tech. Document WMO/TD 1379, GCOS-112, 42 pp.

Gouretski, V., and K. P. Koltermann, 2007: How much is the ocean really warming? *Geophys. Res. Lett.,* **34,** L01610, doi:10.1029/2006GL027834.

Grant, A. N., S. Bronnimann, and L. Haimberger, 2008: Recent Arctic warming vertical structure contested. *Nature,* **455,** E2–E3.

Graversen, R. G., T. Mauritsen, M. Tjernström, E. Källén, and G. Svensson, 2008: Vertical structure of recent Arctic warming. *Nature,* **451,** 53–56, doi:10.1038/nature06502.

Haimberger, L., C. Tavolato, and S. Sperka, 2008: Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data. *J. Climate,* **21,** 4587–4606.

Hollingsworth, A., and Coauthors, 1985: The response of numerical weather prediction systems to FGGE level IIB data. Part I: Analyses. *Quart. J. Roy. Meteor. Soc.,* **111,** 1–66.

John, V. O., and B. J. Soden, 2007: Temperature and humidity biases in global climate models and their impact on climate feedbacks. *Geophys. Res. Lett.,* **34,** L18704, doi:10.1029/2007GL030429.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Karl, T. R., S. J. Hassol, C. D. Miller, and W. L. Murray, Eds., 2006: Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences. U.S. Climate Change Science Program Synthesis and Assessment Product 1.1, 180 pp.

Kistler, R., and Coauthors, 2008: Reanalysis at NCEP: Past, present and future. Extended Abstracts, *Third Int. Conf. on Reanalysis,* Tokyo, Japan, World Climate Research Center, 6 pp. [Available online at http://wcrp.ipsl.jussieu.fr/Workshops/Reanalysis2008/Documents/V1-103_ea.pdf.]

Li, H., E. Kalnay, and T. Miyoshi, 2009a: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 523–533.

——, ——, ——, and C. M. Danforth, 2009b: Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev.,* **137,** 3407–3419.

Lorenc, A. C., and O. Hammon 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *Quart. J. Roy. Meteor. Soc.,* **114,** 515–543.

Matricardi, M., F. Chevallier, G. Kelly, J.-N. Thepaut, 2004: An improved general fast radiative transfer model for the assimilation of radiance observations. *Quart. J. Roy. Meteor. Soc.,* **130,** 153–173.

Mears, C. A., and F. J. Wentz, 2009: Construction of the Remote Sensing Systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders. *J. Atmos. Oceanic Technol.,* **26,** 1040–1056.

Murphy, J. M., and Coauthors, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature,* **430,** 768–772.

Onogi, K., and Coauthors, 2007: The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.

Pawson, S., and M. Fiorino, 1999: A comparison of reanalyses in the tropical stratosphere. Part 3: Inclusion of the pre-satellite data era. *Climate Dyn.,* **15,** 241–250.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108**, 4407, doi:10.1029/2002JD002670.

Schubert, S. D., R. B. Rood, and J. Pfaendtner, 1993: An assimilated dataset for earth science applications. *Bull. Amer. Meteor. Soc.,* **74**, 2331–2342.

Sherwood, S. C., C. L. Meyer, R. J. Allen, and H. A. Titchner, 2008: Robust tropospheric warming revealed by iteratively homogenized radiosonde data. *J. Climate,* **21,** 5336–5350.

Simmons, A. J., and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.,* **109,** D24115, doi:10.1029/2004JD005306.

——, and Coauthors, 2009: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.,* **115,** D01110, doi:10.1029/2009JD012442.

Thorne, P. W., 2008: Arctic tropospheric warming amplification? *Nature,* **455,** E1–E2.

——, D. E. Parker, J. R. Christy, and C. A. Mears, 2005a: Uncertainties in climate—Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.,* **86,** 1437–1442.

——, ——, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005b: Revisiting radiosonde upper air temperatures from 1958 to 2002. *J. Geophys. Res. Atmos.,* **110**(D18), D18105.

Titchner, H. A., and Coauthors, 2009: Critically reassessing tropospheric temperature trends from radiosondes using realistic validation experiments. *J. Climate,* **22,** 465–485.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

——, A. Simmons, D. Dee, P. Kållberg, and J.-N. Thépaut, 2008: Atmospheric reanalyses and climate variations. *Climate Variability and Extremes during the Past 100 Years,* S. Brönnimann et al., Eds., Vol. 33, *Advances in Global Change Research:* Springer, 103–117.

Willis, J. K., J. M. Lyman, G. C. Johnson, and J. Gilson, 2007: Correction to "Recent cooling of the upper ocean." *Geophys. Res. Lett.,* **34,** L16601, doi:10.1029/2007GL030323.