TOWARD A CONSISTENT REANALYSIS OF THE CLIMATE SYSTEM

BY D. P. DEE, M. BALMASEDA, G. BALSAMO, R. ENGELEN, A. J. SIMMONS, AND J.-N. THÉPAUT

Past and current reanalysis activities at ECMWF are reviewed and plans for developing consistent climate reanalyses of the coupled Earth system are discussed.

Since the early 1980s, steady progress in numerical weather prediction has led to better and better descriptions of the global atmospheric circulation as observed during the recent past. This is achieved by "reanalysis," which is a consistent reprocessing of archived weather observations using a modern forecasting system. Reanalysis produces multidecadal, gridded datasets that estimate a large variety of atmospheric, sea-state, and land surface parameters, including many that are not directly observed. Such datasets have become fundamental to research and education in the Earth sciences.

Reanalysis differs from traditional methods for processing observations into useful data products. It relies on models to interpret, relate, and combine many different observations from multiple sources. The types of observation that can be assimilated are limited only by the condition that they can be accurately modeled. The data assimilation uses prior information about uncertainties in models and observations for quality checks, to derive bias adjustments, and to assign proportional weights to the data. The equations of motion and physical processes as represented in a forecast model are used to generate data products that are spatially complete and physically consistent. In essence, the aim in reanalysis is to derive a comprehensive description of the observed atmospheric circulation by using as much information as possible.

Several generations of atmospheric reanalyses have been produced at the National Centers for Environmental Prediction (NCEP; Kalnay et al.1996), the National Aeronautic and Space Agency (NASA; Schubert et al. 1993; Rienecker et al. 2011), the European Centre for Medium-Range Weather Forecasts (ECMWF; Gibson et al. 1997; Uppala et al. 2005; Dee et al. 2011a), and at the Japanese Meteorological Agency (JMA; Onogi et al. 2007). The most recent global atmospheric reanalysis is the 55-yr Japanese Reanalysis (JRA-55; Ebita et al. 2011), completed at the JMA in 2013, but not yet fully documented.

A coupled reanalysis of the global atmosphere, ocean, land surface, and cryosphere was recently

AFFILIATIONS: DEE, BALMASEDA, BALSAMO, ENGELEN, SIMMONS, AND THÉPAUT—European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
CORRESPONDING AUTHOR: D. P. Dee, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom
E-mail: dick.dee@ecmwf.int

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-13-00043.1

In final form 25 November 2013

created by NCEP to support the development of their Climate Forecast System (Saha et al. 2010, 2014). The University of Colorado's Cooperative Institute for Research in Environmental Sciences (CIRES) together with the National Oceanic and Atmospheric Agency (NOAA) has developed for the first time an atmospheric reanalysis that extends back to the late nineteenth century, using only surface pressure observations and prior estimates of sea surface temperature and sea ice concentration (Compo et al. 2011).

To help users find their way among these various datasets, a website (reanalyses.org) has been created with an up-to-date overview maintained by the reanalysis producers. The site includes details about temporal coverage, spatial resolution, data access, etc., and also contains space for comments and feedback from users.

IMPROVING THE MEDIUM-RANGE FORECASTS. Reanalysis activities at ECMWF have always been closely connected with the development of its operational forecasting system. A reanalysis of observations collected for the First Global Experiment of the Global Atmospheric Research Programme (FGGE) started only months after the first operational forecast was issued in August 1979 (Bengtsson et al. 1982a,b). This first FGGE reanalysis was completed by summer 1981. Based on its results and feedback to the data providers, various corrections and additions were made to the original FGGE input dataset. A second FGGE reanalysis covering the two special observing periods (January–February and June–July 1979) was produced in 1986, using the improved FGGE data and an updated version of the forecasting system. These pioneering reanalyses, together with a complementary reanalysis by Ploshay et al. (1992), provided the first global atmospheric datasets available for scientific research, widely utilized in predictability studies and for diagnostic purposes (e.g., ECMWF 1985).

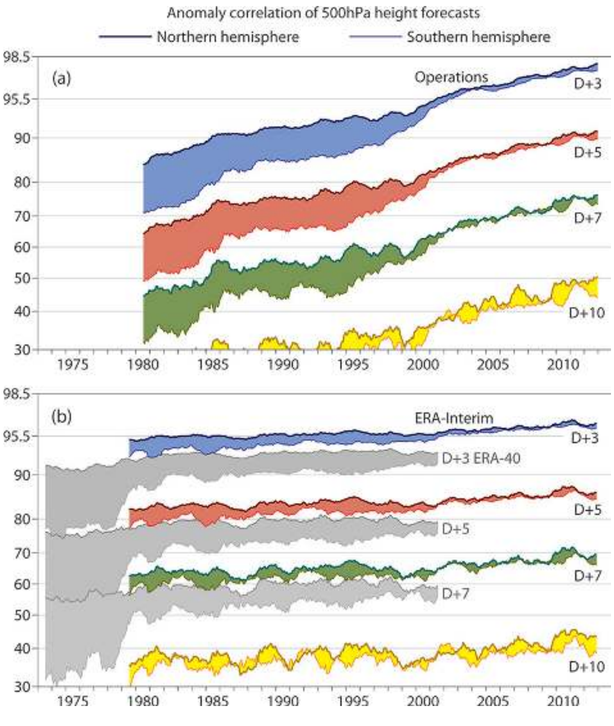

FIG. 1. Twelve-month running mean anomaly correlations (%) of 3-, 5-, 7- and 10-day 1200 UTC forecasts of 500-hPa height for the extratropical Northern and Southern Hemispheres from (a) ECMWF operations from Jan 1980 to May 2013 and (b) ERA-Interim from Jan 1979 to Apr 2013 and ERA-40 from Jan 1973 to Dec 2001. The shading shows the difference in scores between the two hemispheres at the forecast ranges indicated.

ECMWF's role in the FGGE project set in motion a strong feedback loop between improvements in the global observing system, advances in data assimilation methodology, and development of better forecast models through reanalysis. Between 1993 and 1996 an early version of the newly developed Integrated Forecast System (IFS) was used to complete a reanalysis of the period 1979–93 [the 15-yr ECMWF Re-Analysis (ERA-15); Gibson et al. 1997]. Building on experience gained with ERA-15, a reanalysis of the period September 1957–August 2002 was produced (ERA-40; Uppala et al. 2005). ERA-40 used a version of the IFS that was operational in 2001, but at a lower resolution (125 km) than used for medium-range forecasting and with a three-dimensional variational (3D-Var) analysis scheme. ECMWF's most recent atmospheric reanalysis is ERA-Interim (Dee et al. 2011a), covering the modern satellite era from January 1979 to the present at a spatial resolution of 80 km. ERA-Interim is based on a 2006 version of the IFS and uses a four-dimensional variational (4D-Var) analysis for data assimilation.

The IFS model has fully interactive components for the atmosphere, the land surface (since 1991), and the sea state (since 1998). The ERA datasets therefore include estimates of land surface parameters (e.g., soil temperature, soil moisture, snow) and, beginning with ERA-40, parameters that describe the sea state (e.g., wave spectra, significant wave height). These estimates are consistent with the meteorological parameters, in the sense that they are constrained by the coupled model. However, the analysis schemes for the different components are separate and use different methodologies (see Dee et al. 2011a).

Reanalysis data are routinely used to assess the performance of ECMWF's operational forecast

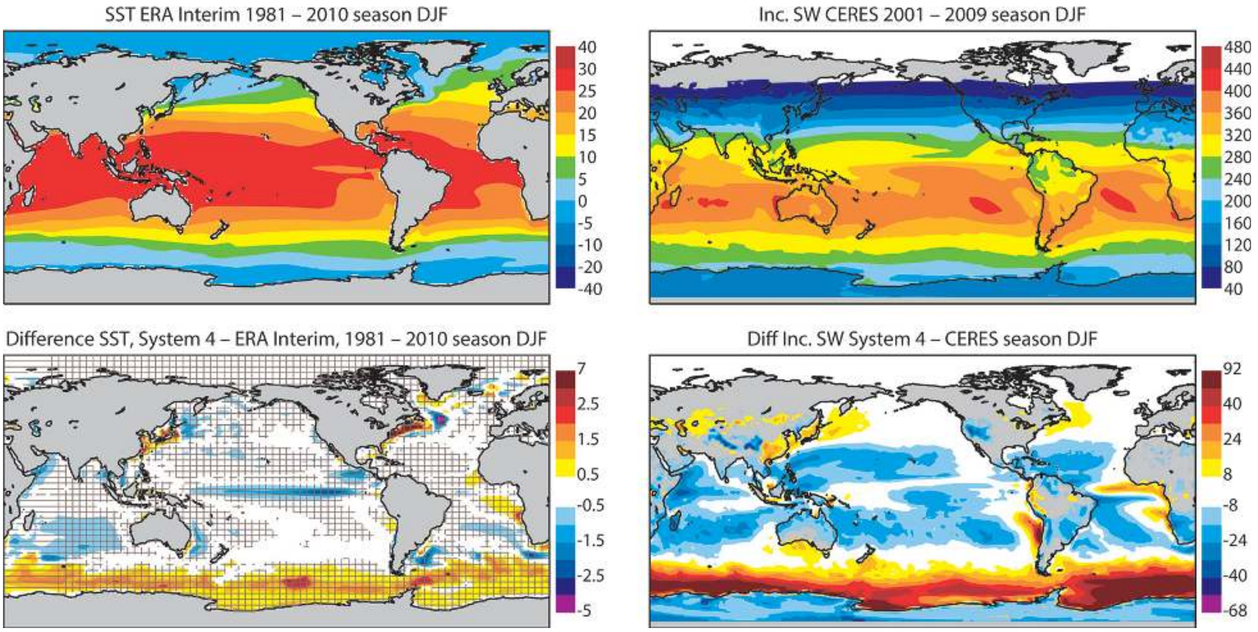

Fig. 2. Model drifts in the current ECMWF seasonal forecasting system (System 4). (left) Mean forecast errors in (bottom) SST relative to (top) ERA-Interim data, averaged over 1981–2010 Northern Hemisphere winter seasons. (right) Mean forecast errors in (bottom) incoming shortwave radiation relative to (top) CERES observations, averaged over 2001–09 Northern Hemisphere winter seasons.

system and to evaluate the effect of new model developments and other changes in the IFS (Simmons and Hollingsworth 2002). Since reanalyses are produced with a fixed configuration of the IFS, it can be very useful to compare the evolution over time of medium-range forecast skill with that of reforecasts produced with the reanalysis system. This is illustrated in Fig. 1 for anomaly correlations of 500-hPa geopotential height forecasts averaged over the hemispheres, obtained from operations and from the ERA-40 and ERA-Interim reanalyses. The figure shows, for example, the effects of atmospheric predictability (common to all curves); the role of an improving observing system (visible in the reanalyses), including better satellite data (convergence of hemispheric scores for ERA-Interim); and the role of improvements in satellite data assimilation (convergence of hemispheric scores for operations). Comparing the slopes in the top and bottom panels suggest that, on average, at most 15% of medium-range forecast skill improvement achieved during the last three decades can be attributed to the evolution of the observing system—the lion's share is due to advances in modeling and data assimilation. The figure does not show, of course, that the research and development conducted at ECMWF that led to those advances benefited greatly from the improved observations–a good example of the feedback loop mentioned earlier.

CALIBRATING MONTHLY AND SEASONAL FORECASTS. Since 2006, ECMWF has produced two major reanalyses of the global oceans [the Ocean Re-Analysis System 3 (ORA-S3; Balmaseda et al. 2008) and System 4 (ORA-S4; Balmaseda et al. 2013)] to support its monthly and seasonal forecasting capability. Prediction of large-scale atmospheric anomalies beyond the medium range requires an accurate representation of slow interactions between the atmosphere and its surface boundaries over land and ocean. ECMWF's monthly and seasonal forecast system therefore uses a coupled atmosphere–ocean model, based on the IFS extended with the Nucleus for European Modeling of the Ocean (NEMO) ocean model.

Coupled atmosphere–ocean models tend to develop biases and drifts at the interface (e.g., see Fig. 2), so seasonal forecasts must be corrected a posteriori for them to be useful. In current practice this is accomplished by generating an extensive set of coupled reforecasts (hindcasts) initialized from a combined ocean–atmosphere reanalysis, then computing the climatology of the model errors relative to the reanalysis, and correcting the model output accordingly. Ideally the reanalysis used as a reference for the calibration should be fully consistent with the coupled prediction system. Instead, the datasets used to calibrate ECMWF's monthly and seasonal forecast systems have been constructed from separately

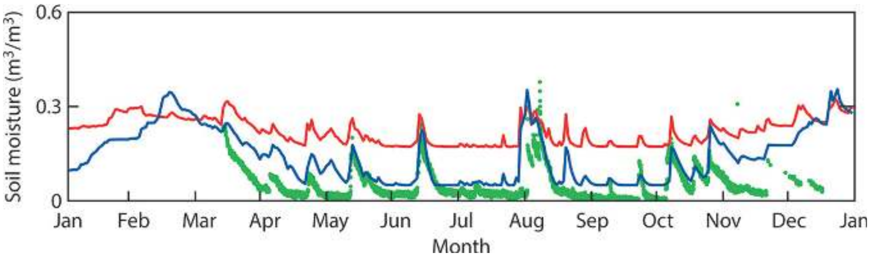

Fig. 3. Evolution of volumetric soil moisture at a site in Utah for the year 2010. In situ observations in green, ERA-Interim estimates in red, and ERA-Interim/Land estimates in blue.

produced (hence not fully consistent) reanalyses of the atmosphere and ocean.

The ocean reanalyses use the NEMO ocean model with prescribed atmospheric forcing (wind, temperature, precipitation) provided by the most recent ERA data. The lack of feedback between ocean and atmosphere in the reanalyses can create physical inconsistencies in the combined dataset. Nevertheless, the reanalysis of the ocean circulation is quite sensitive to the quality of the atmospheric fluxes, which essentially drive the circulation in the upper layers of the ocean. The impact can propagate to the deeper ocean, which is not well constrained by observations, especially prior to the deployment of the Argo floats about a decade ago.

UPGRADING THE LAND SURFACE PARAMETERS. Discrepancies in model versions used for calibrating and initializing seasonal forecasts also concern the representation of the land surface. Koster et al. (2011) have shown that improved initialization of the land surface can enhance the predictability of near-surface temperature by up to 6 weeks. The land surface component of the IFS configuration used for seasonal forecasting at ECMWF has improved a great deal in recent years, especially in the surface hydrology, the treatment of snow, and the representation of carbon processes. The need to include these improvements also in the calibration datasets has motivated the development of an offline land surface modeling system, similar to that used by Reichle et al. (2011). The offline system uses meteorological forcing from an existing atmospheric reanalysis to create an upgraded land surface dataset compatible with a new land surface model. Observational constraints are implicit in the meteorological forcing, and additional information can optionally be included to correct biases (e.g., in precipitation). The land surface parameters can be produced at multiple spatial resolutions, and the offline tool is sufficiently efficient to support testing and development of new land surface model components.

An upgraded land surface reanalysis for 1979–2010, ERA-Interim/Land, has been derived from ERA-Interim in this fashion and is now available for general use. It is based on a recent version of the IFS land surface model [the Tiled ECMWF Scheme for Surface Exchanges over Land with revised Hydrology (HTESSEL); Balsamo et al. 2011] and includes corrections derived from estimates of monthly averaged precipitation provided by the Global Precipitation Climatology Project (GPCP), largely based on rain gauge data over land. Figure 3 illustrates the impact of the new model on the reanalysis of surface hydrology. Balsamo et al. (2012) provide additional information about ERA-Interim/Land with assessments for various parameters of the upgraded land surface.

The offline land surface system can be regarded as a sophisticated tool for downscaling and enhancing the description of the land surface as given by an existing global atmospheric reanalysis. This capability presents interesting possibilities for specialized applications and services, as the downscaling to higher resolution offers support for local generation of products tailored to local needs. Additional developments in the offline system planned at ECMWF include the ability to directly assimilate (or reassimilate) terrestrial and screen-level observations. Improvements in the land surface model, especially when combined with enhanced spatial resolution, should allow better use of near-surface observations and can potentially accommodate many observations that are not currently useable in atmospheric reanalyses.

The use of near-surface observations in the IFS, even at the spatial resolutions used for operational forecasts, is still limited both by poor model representativity near the surface and by shortcomings in the analysis method used. In its current configuration the 4D-Var analysis of upper-air prognostic variables is performed separately from simpler analyses of screen-level parameters (temperature, humidity) and land surface parameters (soil moisture, soil temperature, snow depth). As a result, incremental 4D-Var updates of the atmospheric state are constricted by fixed (but uncertain) conditions at the surface, even if the available observations indicate that these conditions have changed. This lack of dynamic coupling in the analysis between the land surface and the atmospheric boundary layer limits the ability of the IFS to represent fast surface interactions (e.g., associated with precipitation), which may affect forecast skill.

COUPLED REANALYSIS OF ATMOSPHERIC COMPOSITION. Two successive reanalyses of global atmospheric composition have been produced within the framework of the collaborative European projects Global and Regional Earth-System (Atmosphere) Monitoring using Satellite and In Situ Data (GEMS) and Monitoring Atmospheric Composition and Climate (MACC), as part of the development of an operational capability for global air-quality monitoring and forecasting at ECMWF (Hollingsworth et al. 2008; Inness et al. 2013). Both reanalyses used an extended version of the IFS that includes chemically reactive gases, aerosols, and greenhouse gases. Because of the limited availability of global satellite observations of atmospheric composition, these reanalyses extend back only as far as 2003.

The extensions to the IFS developed in the GEMS and MACC projects allow integrated modeling of meteorological, chemical, and aerosol variables and combined use of observations of trace species and meteorology in the 4D-Var analysis. These are the basic elements needed for a fully coupled data assimilation system, in which observations of atmospheric constituents lead to physically consistent adjustments to the meteorological variables, and conversely, meteorological observations can have an immediate impact on estimates of the constituent concentrations. In principle, such a system can produce coherent global analyses with all estimated variables constrained by a unified set of model equations. Coupled data assimilation potentially allows for better use of observations with information about both meteorology and aerosols or chemistry. These are important advantages over uncoupled or weakly coupled systems, in which either the model integration or the analysis of observations (or both) is performed in separate steps.

The actual impact of any single observation in a fully coupled data assimilation system depends on many factors, including the choice of control variables in the analysis, the background error covariances, and details of the forecast model itself. The increased complexity in the system inevitably leads to additional assumptions and pragmatic decisions required for implementation. Fundamentally, a realistic analysis (as in true to nature) is possible only if the additional degrees of freedom in the modeling system can be adequately constrained by observations. This has important implications for climate reanalysis, since the instrumental record available for a reanalysis of atmospheric composition is limited, both in quality and quantity.

Some pitfalls associated with coupled data assimilation are evident even in the ECMWF's medium-range forecast system, which contains ozone as a prognostic model variable. In theory it should be possible to extract useful information about advection from stratospheric tracer observations in the 4D-Var analysis (Riishøjgaard 1996). However, during production of ERA-Interim it became evident that the ozone assimilation caused large and unrealistic changes in the upper-stratospheric circulation, where the model

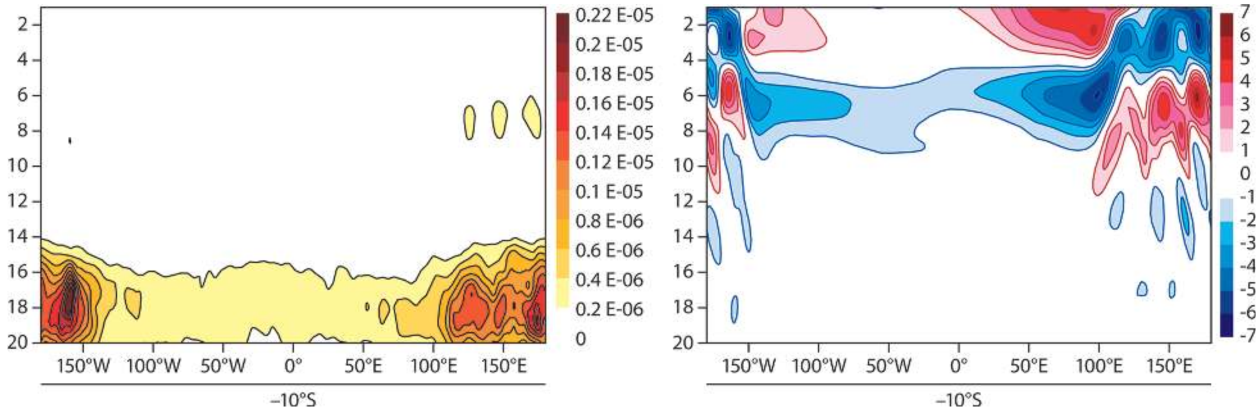

Fig. 4. Impact of the Global Ozone Monitoring Experiment (GOME) ozone profile observations in a single 12-h 4D-Var analysis (0000 UTC 4 Jul 1995), along the latitude circle 10°S for the top 20 model levels [of a 60-level model (i.e., from 40 hPa up to 0.1 hPa)]. Vertical axes indicate IFS model level numbers. (left) Ozone increments with maximum values of about 2 g kg^{-1} are concentrated in locations where the satellite track crosses 10°S. They are everywhere positive in this vertical plane, because the model ozone concentrations are biased low. (right) Unrealistic temperature increments ranging from –6.6 to +6.3 K occur at much higher levels.

background is not well constrained by observations (see Fig. 4). In the 4D-Var framework, these upper-level increments represent the least costly way to accommodate large local changes in observed ozone concentration further below. More realistic increments can be achieved only if both the model background and the observations are sufficiently accurate and largely unbiased, but this is currently not the case.

For similar reasons, the MACC assimilation system does not yet allow direct adjustments to the meteorological parameters based on trace-gas observations. Variational bias corrections for most of the constituents used in the MACC system are being implemented as a necessary step before the 4D-Var analysis can be fully coupled. Additional developments will couple the atmospheric data assimilation with the carbon component of the land surface model (C-TESSEL). Potentially this will generate estimates of surface fluxes of CO_2 for the land biosphere that are fully consistent with the meteorology, provided they can be effectively constrained with available observations.

REANALYSIS FOR CLIMATE APPLICATIONS. The use of reanalysis for climate change assessment is not uncontroversial (e.g., see Thorne and Vose 2010; Dee et al. 2011b). There are well-known difficulties with the representation of low-frequency variability and trends in reanalysis data. Early generations of reanalyses, and to varying degrees more recent ones, show spurious shifts and other artifacts that can be attributed to changes in the observing system in the presence of model error, improper use of observations, transitions between multiple production streams, or various mistakes that can occur in a complex reanalysis production. Many of these issues are technical in nature, but clearly there are also fundamental limits to what is achievable with incomplete observations and imperfect models.

Between ERA-40 and ERA-Interim considerable progress was made at ECMWF in addressing many of the technical issues just mentioned. The ERA-Interim reanalysis uses a more sophisticated data assimilation system, based on a 4D-Var analysis that includes variational adjustment of bias parameters for satellite observations (Dee and Uppala 2009). Technical tools and computing services, including facilities for observation handling, monitoring, and diagnostics, have greatly improved as well. It is now relatively straightforward to maintain a reanalysis production in near–real time and to provide regular monthly updates of the dataset to a large number of users.

An important technical breakthrough was achieved with the 10-yr backward extension of ERA-Interim, which was produced without introducing major discontinuities in the combined dataset (see Fig. 5). Close inspection of a 1-yr overlap between the two

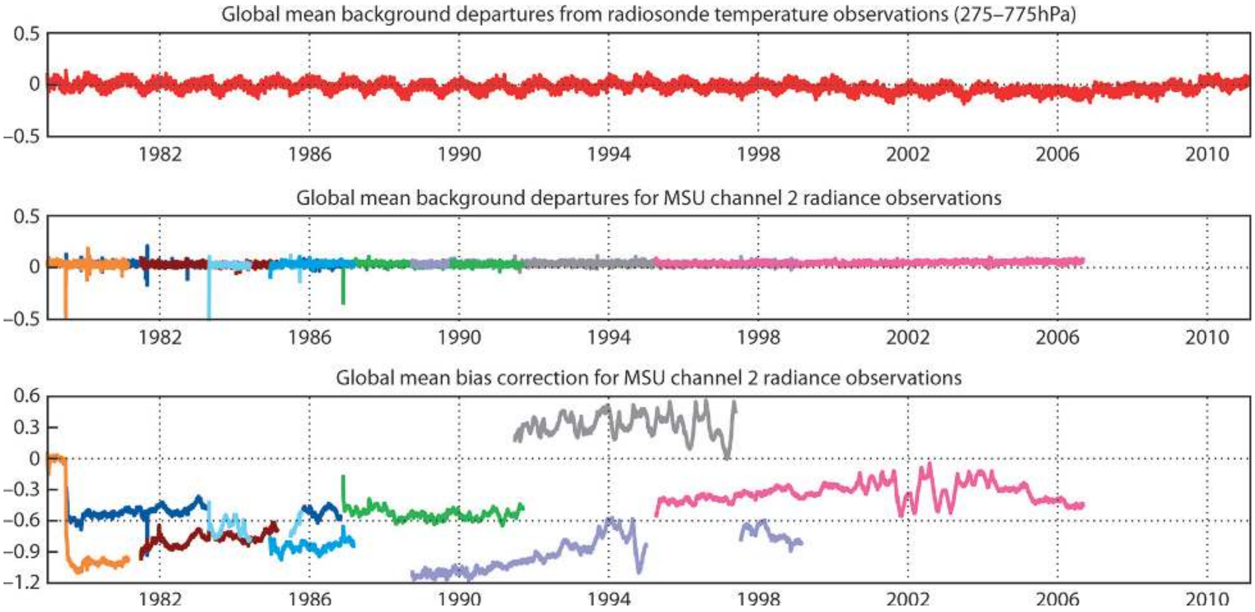

Fig. 5. The three panels illustrate the stability and temporal consistency of the extended ERA-Interim reanalysis and the nearly seamless transition between the two production streams on 1 Jan 1989. Reanalyzed temperatures in the midtroposphere are largely consistent with (top) radiosonde observations and with (middle) bias-corrected radiance measurements from Microwave Sounding Units flown on successive NOAA satellites (colors indicate different satellites). (bottom) The bias corrections for the MSU data, produced by the variational analysis in ERA-Interim, account for calibration differences, orbital drifts, and various other instrument errors.

production streams showed an excellent match for temperature and wind fields, except in the tropical upper stratosphere, and adequate agreement of humidity and ozone throughout the troposphere and lower stratosphere. Convergence of streams depends on the strength of the observational constraint, and requires adequate methods for correcting biases in observations. Nevertheless, it appears feasible to compute a multidecadal atmospheric reanalysis in separate segments, with practical implications for the planning of future high-resolution reanalysis productions.

Ultimately, the achievable temporal consistency in reanalyses depends on the availability of high-quality input observations. The variational bias adjustments applied to the satellite radiances assimilated in ERA-Interim (Fig. 5) are implicitly derived from all other observations available to the reanalysis, most notably from radiosondes, aircraft, and (after 2006) radio occultation data from the Global Positioning System (Poli et al. 2010). Conversely, efforts to improve the existing instrumental record (e.g., by reprocessing, intercalibration, and homogenization) can benefit greatly from the bias estimates and other information about data quality generated by reanalysis. The radiosonde temperature record used in ERA-Interim, for example, incorporates adjustments for changes in equipment that were identified partly on the basis of output from the previous ERA-40 reanalysis (Haimberger 2007).

Reanalysis arguably offers the best potential for extracting maximum information about the recent climate from the total instrumental record by using models to relate and combine information from otherwise disparate observations. Even though meteorological soundings from space have been available since the 1970s, today's benchmark datasets used for monitoring the climate still do not cover critical areas such as the polar regions and parts of the tropics. Changes in near-surface temperature and humidity estimated from modern reanalyses have been shown to closely match those from in situ station records where they exist (Simmons et al. 2004, 2010; see also Fig. 6). For well-observed variables the global reanalyses are now routinely used, along with other observational datasets, in the BAMS annual assessments of climate change (e.g., Blunden and Arndt 2013).

Reanalysis produces useful estimates for model variables that are not well observed, such as stratospheric winds, radiative fluxes, root-zone soil moisture, etc. (see Fig. 7), because these variables are indirectly constrained by the observations used to initialize the assimilating forecast model. In the absence of direct observations, however, it is difficult to quantify the uncertainties in estimates of model-generated variables, as they depend on errors in the model as well as on the strength of the (indirect) observational constraint. An indication of uncertainties can be obtained by using ensemble techniques, with the important caveat that it is not practical to sample more than a few selected sources of uncertainty in a reanalysis.

Nevertheless, the complete description of a physically plausible atmosphere consistent with observations, as provided by reanalysis, makes it possible to do many things that simply cannot be done otherwise (Bengtsson et al. 2007). It permits, for example, detailed diagnostics of the global energy budget and the hydrological cycle (Trenberth et al. 2011). Such diagnostics are especially useful if they involve known time-invariant properties of the climate system. These are usually quite well conserved by the assimilating model in a reanalysis, but tend to be perturbed by the assimilation increments, depending on the nature of the observational constraints and on the method of assimilation. Budget diagnostics are useful for demonstrating shortcomings as well as progress in climate reanalysis (Berrisford et al. 2011), and examination of the increments can be highly informative about shortcomings in the assimilating model (Mapes and Bacmeister 2012). Ironically, inconsistencies in the mass and energy budgets are often used to question the usefulness of reanalysis data for climate applications, even though it is clearly not possible to estimate most of the quantities involved from observations alone.

There is still a great deal of room for improvement in reanalysis; various issues with the quality of ERA-Interim have been identified that remain to be addressed (e.g., Dee et al. 2011a). These include the temporal consistency of mean precipitation over the tropical oceans, which, although much improved over ERA-40, is affected by incorrect use of rain-affected satellite radiances in the ERA-Interim system. Various aspects of the land surface representation in ERA-Interim, in particular snow cover, have been affected by problems with the input data as well as shortcomings in the surface analysis scheme. The energy balance at the surface boundary in ERA-Interim is poor, especially over tropical oceans owing to excessive solar radiation associated with a bias in cloud cover.

Recent IFS upgrades aimed at improving forecast skill will address some of the shortcomings. However, additional development is needed specifically to address the key requirements for climate applications: 1) reanalyses need to extend further back in time to provide a longer record for climate studies and climate model validation, and 2) interactions and feedbacks between the atmosphere and other components of the climate system need to be better represented. In addition, it is necessary to provide users with much better information about uncertainties relevant to the assessment of low-frequency variability and trends.

GOING FURTHER BACK IN TIME. The production of a model-based reanalysis that extends as far back as the instrumental record allows was first pursued in the 20th-Century Reanalysis Project (Compo et al. 2006) at NOAA's Earth Systems Research Laboratory. The project was based on the idea that a reanalysis assimilating only surface pressure observations is relatively straightforward to compute and avoids many of the problems associated with observing system changes.[1] Surface weather observations are available in reasonably large

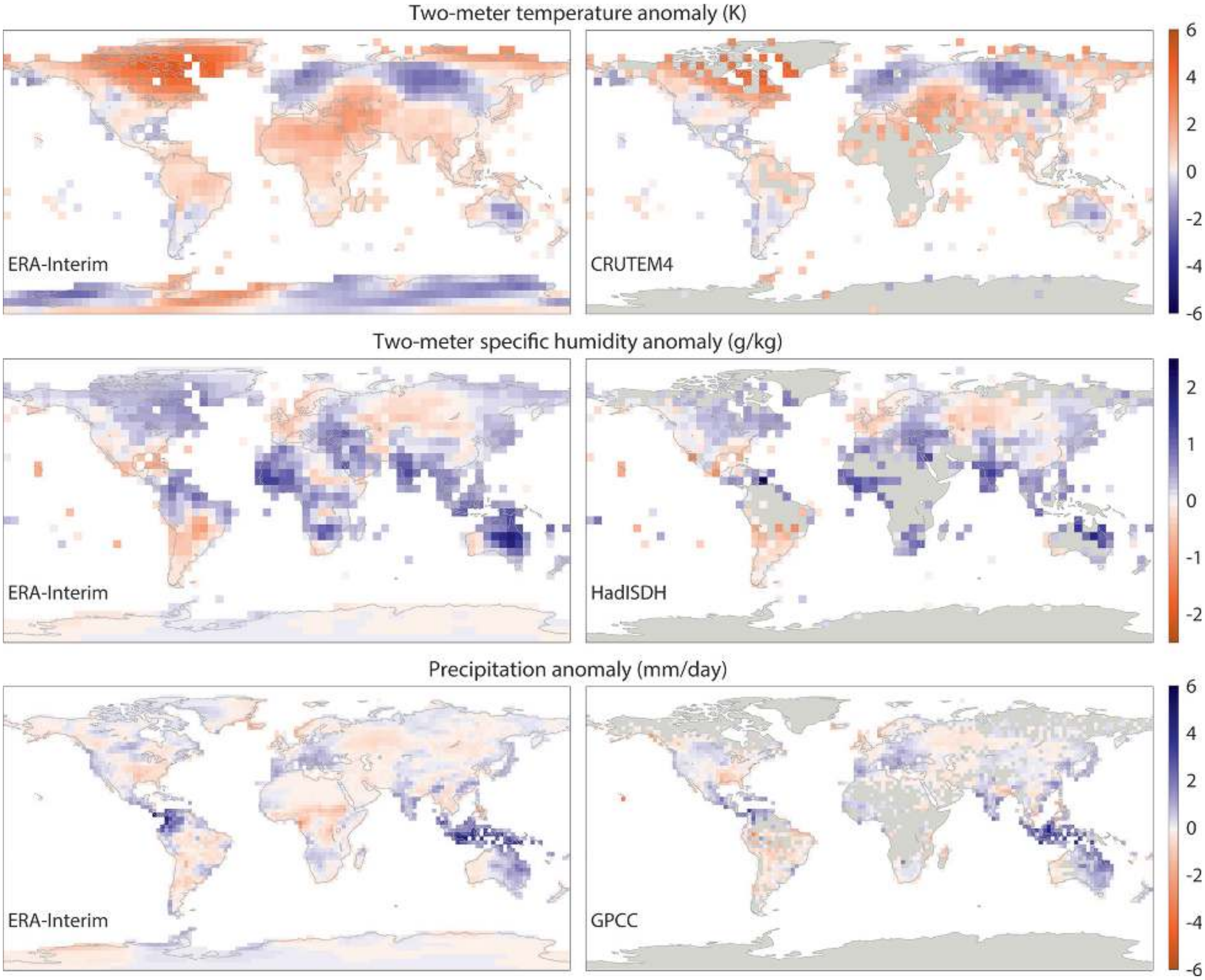

Fig. 6. Anomalies for 2010 relative to 1981–2010 in (top) surface air temperature (K) and (middle) specific humidity (g kg^{-1}), and in (bottom) precipitation rate (mm day^{-1}), from (left) ERA-Interim and from (top right) the Climatic Research Unit temperature dataset, version 4 (CRUTEM4; Jones et al. 2012), (middle right) the Hadley Centre integrated surface humidity dataset (HadISDH; Willett et al. 2013), and (bottom right) Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis version 6 (Becker et al. 2013). ERA-Interim values for a particular variable are averaged over the 5° or 2.5° grid boxes of the dataset with which comparison is made, and are plotted for grid boxes that are at least 10% land or where there are otherwise data values for comparison. Values from CRUTEM4, HadISDH, and GPCC are plotted only for grid boxes with a complete monthly data record for 2010. For GPCC it is also required that there be at least one station per grid box. 2010 was a relatively warm and moist El Niño year.

---

[1] It must be noted that any such reanalysis requires model boundary conditions (e.g., for sea surface temperature, which involve observations from multiple sources and can therefore also be affected by changes in the observing system).

numbers throughout the twentieth century, initially concentrated in the Northern Hemisphere but with global coverage increasing with time. Modern data assimilation systems are remarkably well able to reconstruct a large-scale tropospheric circulation from surface pressure observations alone (Whitaker et al. 2009), although the quality of the reconstruction strongly depends on the quality of the assimilating model (including its boundary conditions), especially where observations are sparse.

A reanalysis of the 140-yr period 1871–2010 was produced using NCEP's Global Forecast System (GFS) with an ensemble Kalman filter especially developed for the purpose (Compo et al. 2011). Future versions have been proposed that will extend even further back in time. Many of the earlier observations needed for this ground-breaking project were obtained from analog sources, then digitized and collected in a global database [the International Surface Pressure Databank (ISPD)] and are now available for general use.

At ECMWF a similar reanalysis for the period 1900–2010 is now taking place, within the framework of the EU-funded ERA-CLIM project (see sidebar). The ERA-20C reanalysis is being produced with the IFS at 125-km global resolution, as an ensemble of 4D-Var data assimilations (EDA), using surface pressure and marine wind observations from the ISPD and the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Poli et al. 2013). The assimilating model has been supplied with atmospheric forcing related to model radiation and land surface processes (Hersbach et al. 2013), using datasets prepared for the Coupled Model Intercomparison Project phase 5 (CMIP5). Boundary conditions for the model make use of new estimates of the evolution of sea surface temperature (SST) and sea ice concentrations (SIC) during the twentieth century, developed for the ERA-CLIM project by the Met Office Hadley Centre. These are provided as an ensemble of equally likely, physically plausible realizations, which take into account some of the fundamental uncertainties in the available observational sources (see Fig. 8). The assimilating model in ERA-20C uses an ensemble of 10 such realizations to represent at least one key source of uncertainty in the reanalysis.

The full set of ERA-20C data products, including a model-only simulation (ERA-20CM) and a 25-km global land surface product (ERA-20CL), will become available during the first half of 2014.

ERA-20C will be ECMWF's first extended climate reanalysis, based on a restricted set of observations and other input data for the model specifically prepared for climate applications. In contrast, previous ECMWF reanalyses have assimilated the majority of observations used in numerical weather prediction (NWP) in an attempt to produce the best possible estimate of the

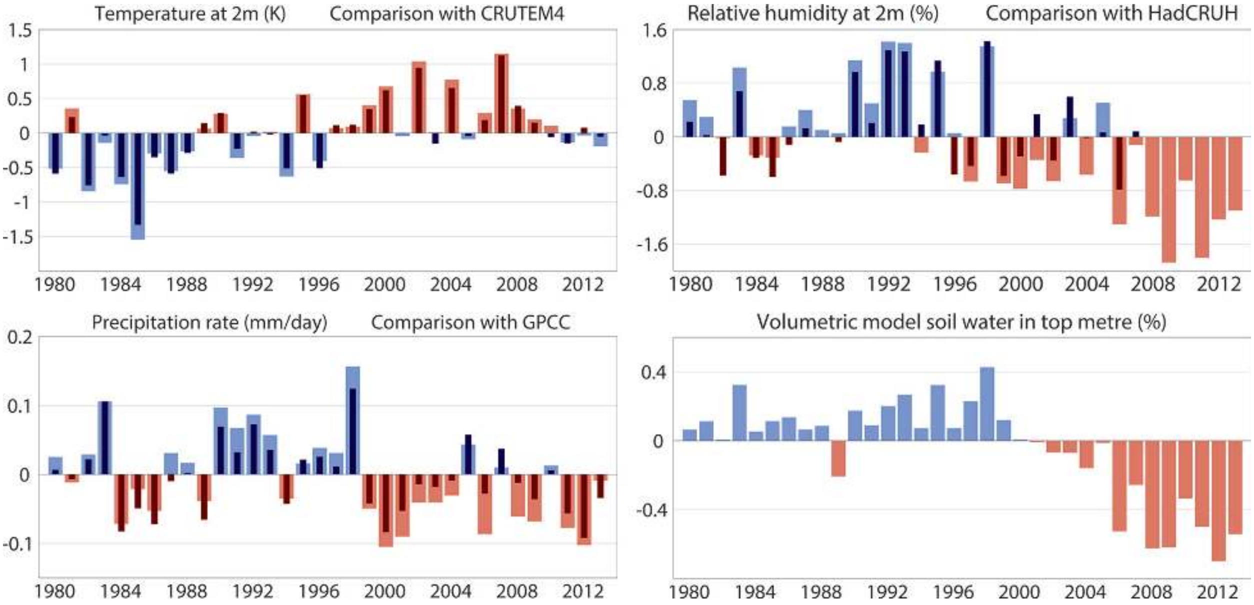

Fig. 7. Dec–Mar anomalies averaged over Northern Hemisphere land locations computed from ERA-Interim data and independently from in situ observations. Light-colored bars show ERA-Interim estimates; vertical darker lines show corresponding estimates from (top left) CRUTEM4 (for temperature), (top right) HadCRUH (for relative humidity), and (bottom left) GPCC (for precipitation). ERA-Interim estimates of changes in soil moisture (bottom right) correlate well with observed changes in humidity.

atmospheric state at any given time. Clearly both types of reanalysis have a role in climate studies and climate change monitoring. Extended reanalyses will provide the longest possible record of low-frequency variability and change consistent with observations, which is needed to put more recent large-scale anomalies in perspective. The spatial and temporal resolution that can be achieved in such a reanalysis will be limited mainly by the available observations. A shorter reanalysis of the satellite era, such as ERA-Interim, can provide a more detailed and complete view of recent changes taking place in the climate system, which can be continuously updated by making use of observations used for operational forecasting.

The two kinds of reanalyses have different requirements in terms of data usage. A traditional NWP-like reanalysis of the satellite era tends to assimilate all available observations unless they are known to be of poor quality or unusable for other reasons; this is the familiar blacklisting approach. In contrast, an extended climate reanalysis ideally follows a whitelisting approach to data selection, where observations are used only if they are known to be suitable for climate applications. In practice the distinction may not be quite as strict, since judgments on data quality are often difficult to make, but the contrasting objectives can provide useful guidance. It implies, for example, that a climate reanalysis requires extra effort in selecting and preparing the input data prior to assimilation, with preference given to observations that have been reprocessed, homogenized, or otherwise prepared specifically for climate applications.

COUPLING THE ATMOSPHERE AND OCEAN. Coupled data assimilation is the second major development thrust needed to better address climate requirements, since only coupled models can provide consistent global transport of mass, water, and energy on the relevant time scales. In addition, the use of a coupled atmosphere–ocean model potentially allows better use of near-surface observations, and this can lead to reduced uncertainties in estimates of surface conditions.

To illustrate the latter point, consider the best available observational estimates of global sea surface temperature, which are essential input for any atmosphere-only reanalysis. These estimates cannot be considered reliable on time scales less than a month or so in the presatellite era. Daily boundary conditions for the atmospheric model are then typically obtained by time interpolation from monthly averages, which introduces additional errors and creates unphysical (and usually damped) variability. On the other hand, a reanalysis produced with a coupled atmosphere–ocean model generates daily sea surface temperature evolutions that are at least physically plausible and, at best, constrained by all available observations—including any atmospheric observations affected by local properties of the sea surface. This scenario presumes, of course, a sufficiently accurate model whose drift at the surface can be adequately constrained.

As mentioned earlier, NCEP has recently produced a Climate Forecast System Reanalysis (CFSR), a first global reanalysis based on a coupled model of the

THE ERA-CLIM AND ERA-CLIM2 PROJECTS

The European Commission has provided research funding for two consecutive 3-yr projects aimed at developing global climate reanalyses extending back to the early twentieth century to support the implementation of an operational Climate Change Service in Europe. Both projects are led by ECMWF but involve participants from various EU member states as well as Russia and Chile. The first ERA-CLIM project, which started in 2011, focused on data rescue of early in situ upper-air observations, preparation and reprocessing of satellite datasets for reanalysis, and production of several pilot reanalyses at ECMWF including ERA-20C and ERA-20C/Land.

The follow-up ERA-CLIM2 project will add a new focus on coupled data assimilation research. The project will develop a first coupled ocean–atmosphere reanalysis of the twentieth century, together with consistent estimates of carbon fluxes and stocks. A longer-term aim of the project is to develop new climate monitoring products based on high-resolution coupled reanalyses.

An important activity initiated in ERA-CLIM is the development of an Observation Feedback Archive. This facility will provide open access to all input observations used in ECMWF reanalyses, supplemented with information about data quality generated by the reanalyses such as departures, bias adjustments, and error estimates. The goal is to make it easier for users to assess the observational constraint in reanalysis products and to allow traceability of observations to their source.

atmosphere, land surface, ocean, and sea ice (Saha et al. 2010). The data assimilation scheme developed for CFSR is weakly coupled in the sense that it uses the coupled model only for generating background estimates for each analysis cycle. The analysis itself is uncoupled (i.e., performed in separate and independent steps for each of the model components). The coupling is described as weak because information from observations in any one component can affect the state of the other components only indirectly by propagating the information to the next analysis cycle. Another important feature of the CFSR scheme is that it makes use of external datasets for prior estimates of sea surface temperature, sea ice concentration, snow depth, and precipitation to constrain model drift.

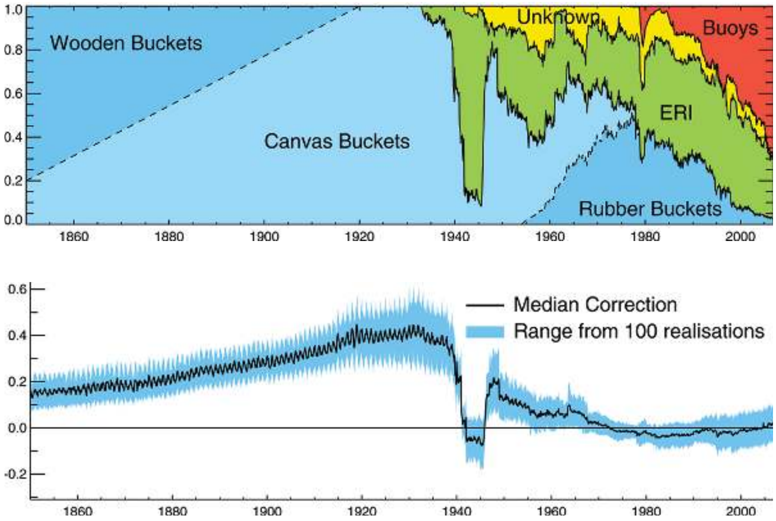

Fig. 8. Observational uncertainty in estimates of sea surface temperature. (top) The evolution of different measurement methods used for in situ observation of sea surface temperature, as a fraction of total. The exact proportion between wooden and canvas buckets used in the early part of the record is not known. (bottom) The associated uncertainties in bias corrections (K) applied when constructing global SST fields. Figure courtesy of Met Office Hadley Centre; see also Kennedy et al. (2011a,b).

At ECMWF work is well underway to expand the technical framework of the IFS to allow coupled data assimilation for the atmosphere and ocean. Ultimately this will provide the IFS with the ability to initialize coupled forecasts, offering prospects for skill improvements in the medium as well as the monthly to seasonal range. The first major implementation target for this development is the production of a coupled atmosphere–ocean reanalysis of the twentieth century by the end of 2015.

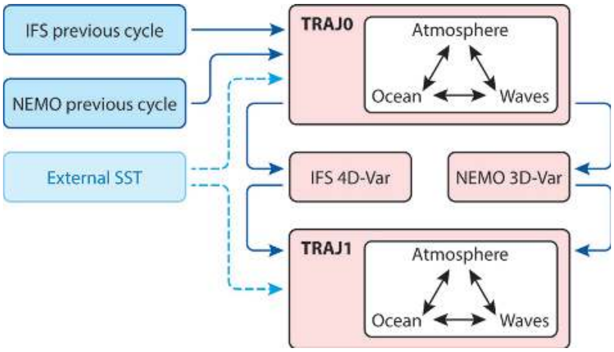

Fig. 9. Schematic of a single analysis cycle for coupled data assimilation based on the incremental variational approach. TRAJ0 represents a first-guess four-dimensional coupled model trajectory initialized with output from the previous analysis cycle. IFS 4D-Var and NEMO 3D-Var provide linearized analysis increments separately for the ocean and atmosphere, which are then used to create an updated model trajectory in TRAJ1. An external SST product is used to constrain model drift in the coupled model integrations.

The new system will use a fully coupled model that combines the IFS atmospheric forecast model (including land surface and sea-state components) with the NEMO ocean model (optionally including dynamic sea ice and biogeochemistry components), as used for seasonal forecasting at ECMWF. The design of the coupled data assimilation scheme follows the incremental variational approach as currently implemented in the IFS (Courtier et al. 1994), in which the nonlinear 4D-Var analysis problem is solved iteratively using successive linearizations of the model and observation operators. Each iteration begins with a four-dimensional trajectory obtained by integrating the coupled model forward in time, over the length of the analysis window. The variational problem is linearized about this reference trajectory and then solved separately for the atmosphere and ocean components of the system. From the updated state the coupled model is again integrated to generate a new nonlinear reference trajectory, which is then used for the next iteration of the scheme. These steps are schematically represented in Fig. 9.

This incremental approach can accommodate strong coupling because it allows dynamic two-way exchange of information between ocean and atmosphere within a single analysis cycle. The exchange takes place during the coupled model integrations used to generate the reference trajectories for the linearized analysis updates. Thus, assimilation of an ocean observation can change the atmospheric state, and conversely, atmospheric observations can directly affect the ocean. In principle the design can accommodate observations that are sensitive to both oceanic and atmospheric variables. By construction, the final four-dimensional state estimates produced by the analysis will be consistent at the ocean–atmosphere interface.

To address the issue of model drift and its possible impact on a reanalysis of the presatellite era, the data assimilation will include a weak observational constraint on the atmosphere–ocean interface. The initial implementation of this bias constraint will rely on prior estimates of global monthly averaged sea surface temperature and sea ice concentrations to selectively constrain the well-observed time scales of variability.

A ROLE FOR REANALYSIS IN CLIMATE SERVICES. Climate services encompass a wide range of activities that deal with generating, processing, and delivering information about past, present, and future climate and its effect on society and the environment (WMO 2011). Applications include, for example, monitoring and seasonal prediction of droughts; policy development in agriculture, water use, health, and urban planning; climate-related risk assessments for the reinsurance industry; optimal design of sustainable energy projects; and others. Global reanalyses can provide the background information about the observed climate that is needed, together with other specialized datasets, to produce reliable climate indicators relevant to such applications. Indeed, a pivotal role for reanalysis in the information chain for climate services is foreseen in current plans of the European Commission for the establishment of an operational Climate Change Service (Uppala et al. 2011).

Different types of reanalyses are needed to serve both atmospheric science and climate services in future. These include the extended climate reanalyses reaching back a century or more and the comprehensive reanalyses of the recent observing period discussed in this article. In addition, regional reanalyses can add valuable information from high-resolution observations, and can be afforded at higher spatial resolutions than the global reanalyses. Comprehensive ocean reanalyses, high-resolution reanalyses of the land surface, reanalyses of atmospheric composition, and other specialized model-based reanalyses all have important roles to play in advancing our understanding of the climate system. The common goal is to make the best possible use of observations. In all cases, of course, there are fundamental limitations on the ability to represent what is essentially unobserved.

Research funding from both the European Commission and the European Space Agency is currently supporting the implementation of a coupled data assimilation capability targeted for the production of climate reanalyses, following the approach outlined in this article. The work is coordinated by ECMWF but involves collaboration with many research institutions and data providers around the world. Further investments are needed to transform reanalysis from the valuable research activity it currently is into a dependable operational service. Resources for this purpose will likely become available through the Copernicus Climate Change Services being established by the European Commission. The technical expertise and resources available at ECMWF, resulting from many years of investment in numerical weather prediction in Europe, can then be enlisted in the effort to address one of the most pressing problems of our times—namely, adaptation to and mitigation of climate change.

REFERENCES

Balmaseda, M., A. Vidard, and D. L. T. Anderson, 2008: The ECMWF Ocean Analysis System: ORA-S3. *Mon. Wea. Rev.,* **136,** 3018–3034, doi:10.1175/2008MWR2433.1.

——, K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF Ocean Reanalysis ORAS4. *Quart. J. Roy. Meteor. Soc.,* **139,** 1132–1161, doi:10.1002/qj.2063.

Balsamo, G., S. Boussetta, E. Dutra, A. Beljaars, P. Viterbo, and B. Van den Hurk, 2011: Evolution of land surface processes in the IFS. *ECMWF Newsletter,* No. 127, ECMWF, Reading, United Kingdom,17–22. [Available online at www.ecmwf.int/staff/gianpaolo_balsamo/DOC/NL127-Balsamo-lr.pdf.]

——, and Coauthors, 2012: ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing. ERA Rep. Series 13, 25 pp. [Available online at www.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_13.pdf.]

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data,* **5,** 71–99, doi:10.5194/essd-5-71-2013.

Bengtsson, L., M. Kanamitsu, P. Kållberg, and S. Uppala, 1982a: FGGE research activities at ECMWF. *Bull. Amer. Meteor. Soc.,* **63,** 277–303.

——, ——, ——, and ——, 1982b: FGGE four-dimensional data assimilation at ECMWF. *Bull. Amer. Meteor. Soc.,* **63,** 29–43.

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

Berrisford, P., P. Kållberg, S. Kobayashi, D. Dee, S. Uppala, A. J. Simmons, P. Poli, and H. Sato, 2011: Atmospheric conservation properties in ERA-Interim. *Quart. J. Roy. Meteor. Soc.,* **137,** 1381–1399, doi:10.1002/qj.864.

Blunden, J., and D. S. Arndt, 2013: State of the Climate in 2012. *Bull. Amer. Meteor. Soc.,* **94,** S1–S258.

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, doi:10.1175/BAMS-87-2-175.

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

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387, doi:10.1002/qj.49712051912.

Dee, D. P., and S. M. 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, 2011a: 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.

——, E. Källén, A. J. Simmons, and L. Haimberger, 2011b: Comments on "Reanalyses suitable for characterizing long-term trends. " *Bull. Amer. Meteor. Soc.,* **92,** 65–70, doi:10.1175/2010BAMS3070.1.

Ebita, A., and Coauthors, 2011: The Japanese 55-year Reanalysis "JRA-55": An interim report. *SOLA,* **7,** 149–152, doi:10.2151/sola.2011-038.

ECMWF, 1985: *Proc. 1984 Seminar/Workshop on Data Assimilation Systems and Observing System Experiments with Particular Emphasis on FGGE.* Reading, United Kingdom, ECMWF.

Gibson, J. K., P. Kållberg, S. M. Uppala, A. Nomura, A. Hernandez, and E. Serrano, 1997: ERA description. ERA-15 Rep. Series 1, 72 pp. [Available online at www.ecmwf.int/publications/library/do/references/show?id=83253.]

Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate,* **20,** 1377–1403, doi:10.1175/JCLI4050.1.

Hersbach, H., and Coauthors, 2013: ERA-20CM: A twentieth century atmospheric model ensemble. ERA Rep. Series 16, 44 pp. [Available online at www.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_16.pdf.]

Hollingsworth, A., and Coauthors, 2008: Toward a monitoring and forecasting system for atmospheric composition: The GEMS Project. *Bull. Amer. Meteor. Soc.,* **89,** 1147–1164, doi:10.1175/2008BAMS2355.1.

Inness, A., and Coauthors, 2013: The MACC reanalysis: An 8 yr data set of atmospheric composition. *Atmos. Chem. Phys.,* **13,** 4073–4109, doi:10.5194/acp-13-4073-2013.

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., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471, doi:10.1175/1520-0477(1996)0772.0.CO;2.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.,* **116,** D14103, doi:10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850: 2. Biases and homogenisation. *J. Geophys. Res.,* **116,** D14104, doi:10.1029/2010JD015220.

Koster, R. D., and Coauthors, 2011: The second phase of the Global Land–Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeor.,* **12,** 805–822, doi:10.1175/2011JHM1365.1.

Mapes, B. E., and J. T. Bacmeister, 2012: Diagnosis of tropical biases and the MJO from patterns in the MERRA analysis tendency fields. *J. Climate,* **25,** 6202–6214, doi:10.1175/JCLI-D-11-00424.1.

Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan,* **85,** 369–432, doi:10.2151/jmsj.85.369.

Ploshay, J. J., R. White, and K. Miyakoda, 1992: FGGE reanalysis at GFDL. *Mon. Wea. Rev.,* **120,** 2083–2108, doi:10.1175/1520-0493(1992)1202.0.CO;2.

Poli, P., S. B. Healy, and D. P. Dee, 2010: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.,* **136,** 1972–1990, doi:10.1002/qj.722.

——, and Coauthors, 2013: The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). ERA Rep. 14, 59 pp. [Available from www.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_14.pdf.]

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate,* **24,** 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

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.

Riishøjgaard, L. P., 1996: On four-dimensional variational assimilation of ozone data in weather-prediction models. *Quart. J. Roy. Meteor. Soc.,* **122,** 1545–1571, doi:10.1002/qj.49712253505.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.,* **91,** 1010–1057, doi:10.1175/2010BAMS3001.1.

——, and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate,* **27,** 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Schubert, S. D., R. Rood, and J. Pfaendtner, 1993: An assimilated dataset for earth science applications. *Bull. Amer. Meteor. Soc.,* **74,** 2331–2342, doi:10.1175/1520-0477(1993)0742.0.CO;2.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **128,** 647–677, doi:10.1256/003590002321042135.

——, 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.

——, K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee, 2010: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational datasets. *J. Geophys. Res.,* **115,** D01110, doi:10.1029/2009JD012442.

Thorne, P. W., 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., J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Climate,* **24,** 4907–4924, doi:10.1175/2011JCLI4171.1.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012, doi:10.1256/qj.04.176.

——, and Coauthors, 2011: GMES climate service: Towards a European knowledge base in support of mitigation and adaptation. Draft Report European Commission DG ENTR/H.4/CC (2011), 22 pp. [Available online at http://ec.europa.eu/enterprise/newsroom/cf/_getdocument.cfm?doc_id=6611.]

Whitaker, J. S., G. P. Compo, and J.-N. Thépaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Wea. Rev.,* **137,** 1991–1999, doi:10.1175/2008MWR2781.1.

Willett, K. M., C. N. Williams Jr., R. J. H. Dunn, P. W. Thorne, S. Bell, M. de Podesta, P. D. Jones, and D. E. Parker, 2013: HadISDH: An updateable land surface specific humidity product for climate monitoring. *Climate Past,* **9,** 657–677, doi:10.5194/cp-9-657-2013.

WMO, 2011: Climate knowledge for action: A global framework for climate services—Empowering the most vulnerable. WMO-1065, 240 pp. [Available online at ftp://ftp.wmo.int/Documents/SESSIONS/Cg-XVI/English/DOCs/d11-1%281%29_1065_HLT_report_en.pdf.]