Observing System Simulation Experiments

Michiko Masutani[1], Thomas W. Schlatter[2], Ronald M. Errico[3], Ad Stoffelen[4], Erik Andersson[5], William Lahoz[6], John S. Woollen[7], G. David Emmitt[8], Lars-Peter Riishøjgaard[9], Stephen J. Lord[10]

[1]NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA, and Wyle Information Systems, El Segundo, CA, USA, Michiko.Masutani@noaa.gov
[2]NOAA/Earth System Research Laboratory, Boulder, CO, USA, Tom.Schlatter@noaa.gov
[3]NASA/GSFC, Greenbelt, MD, USA, and Goddard Earth Science and Technology Center, University of Maryland, Baltimore, MD, USA, Ronald.M.Errico@nasa.gov
[4]Royal Dutch Meteorological Institute (KNMI), DeBilt, The Netherlands, Ad.Stoffelen@knmi.nl
[5]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK, erik.andersson@ecmwf.int
[6]Norsk Institutt for Luftforskning (NILU), Norway, wal@nilu.no
[7]NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA and Science Applications International Corporation (SAIC), USA, Jack.Woollen@noaa.gov
[8]Simpson Weather Associates (SWA), Charlottesville, VA, USA, gde@swa.com
[9]NASA/GSFC, Greenbelt, MD, USA; Goddard Earth Science and Technology Center, University of Maryland, Baltimore, MD, USA; and Joint Center for Satellite Data Assimilation,Camp Springs, MD, USA, Lars.P.Riishojgaard@nasa.gov
[10]NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA, Stephen.Lord@noaa.gov

1 Definition and motivation of OSSEs

Observing System Simulation Experiments (OSSEs) are typically designed to use data assimilation ideas (chapter *Mathematical Concepts in Data Assimilation*, Nichols) to investigate the potential impacts of prospective observing systems (observation types and deployments). They may also be used to investigate current observational and data assimilation systems by testing the impact of new observations on them. The information obtained from OSSEs is generally difficult, or in some contexts impossible, to obtain in any other way.

In an OSSE, simulated rather than real observations are the input to a data assimilation system (DAS for short). Simulated observational values are drawn from some appropriate source (several possibilities have been considered; see Section 3). These values are generally augmented by implicitly or explicitly estimating respective values of observational errors to make them more realistic (see Section 4). The resulting values are then ingested into a DAS (that may be as complex as an operational one) just as corresponding real observations would be. Simulations of both analyses and subsequent forecasts are then produced for several experiments, with each considering a distinct envisioned observing system; *i.e.*, a distinct set of observation types and characteristics. The analysis and forecast products are then compared to evaluate the impacts of the various systems considered.

OSSEs are closely related to Observing System Experiments (OSEs). For an observing system in operational use, the OSE methodology consists of:

- A control run in which all observational data currently used for every-day operations are included;
- A perturbation run from which the observation type under evaluation is *excluded* while all other data are kept as for the control;
- A comparison of forecast skill between the control and perturbation runs.

OSEs are effectively Data-Denial Experiments (DDEs, discussed in Section 7.1). They reveal specifically what happens when a DAS is degraded by removing particular subsets of observations and thus measure the impacts of those observations.

The structure of an OSSE is formally similar to that of an OSE with one important difference: OSSEs are assessment tools for *new* data, *i.e.*, data obtained by hypothetical observing systems that do not yet exist. The methodology of an OSSE consists of:

- Generation of reference atmospheric states for the entire OSSE period. This is usually done with a good-quality, realistic atmospheric model in a free-running mode without data assimilation. This is often called the Nature Run (NR for short), providing the proxy "truth," from which observations are simulated and against which subsequent OSSE assimilation experiments are verified;
- The generation of simulated observations, including realistic errors, for all existing observing systems and for the hypothetical future observing system;
- A control run (or experiment) in which all the data representing the current operational observational data stream are included;
- A perturbation run (or experiment) in which the simulated candidate observations under evaluation are added;
- A comparison of forecast skill between the control and perturbation runs.

The most common motivation for OSSEs regards estimating the potential impact of proposed new observation types. Although a new type may be highly accurate and robust, it does not provide complete, instantaneous global coverage with perfect accuracy. All new observation types therefore will be used in conjunction with other, mostly already existing, observation types and a background derived from a short-term model forecast. Since data assimilation is a blending of all such useful information, the impact of a new type can only be estimated by considering it in the context of all the other useful types. It is therefore necessary to investigate potential impacts in a complete and realistic DAS context.

New observation types that do not yet exist cannot provide observational values to be assimilated. If a prototype does exist but is not already deployed as envisioned, impacts that can be currently measured may be unrepresentative of future potential impacts or not statistically significant. The latter is always an issue with data assimilation because the data analysis problem is fundamentally statistical due to unknown aspects of observational and modelling errors. Under these conditions, the only way of estimating the potential impact of new observations is by appropriately simulating them; *i.e.*, performing an OSSE of some kind.

Besides estimating the impact, and therefore the value, of an augmentation to the observing system, an OSSE can be used to compare the effectiveness of competing observation designs or deployment options. What is the cost to benefit ratio, for example, between using a nadir-looking versus a side-scanning instrument on a satellite? Or, for a lidar, what are the relative benefits of using various power settings for the beams? An OSSE can aid the design before putting an instrument in production. Thus, well-conducted OSSEs can be invaluable for deciding trade-offs between competing instrument proposals or designs: the cost of an OSSE is a tiny fraction of the cost of developing and deploying almost any new observing system.

Furthermore, by running OSSEs, current operational data assimilation systems can be tested, and upgraded to handle new data types and volume, thus accelerating use of future instruments and observing systems. Additionally, OSSEs can hasten database development, data processing (including formatting) and quality control software. Recent OSSEs show that some basic tuning strategies can be developed before the actual data become available. All of this accelerates the operational use of new observing systems. Through OSSEs future observing systems can be designed to optimize the use of data assimilation and forecast systems to improve weather forecasts, thus giving maximum societal and economic impact (Arnold and Dey 1986; Lord *et al.* 1997; Atlas 1997).

There is another motivation for OSSEs that has been less often discussed. It exploits the existence of a known "truth" in the context of an OSSE. For a variety of purposes, including validating or improving an existing DAS or designing perturbations for predictability studies or ensemble forecasting, it is useful to characterize critical aspects of analysis errors. Evidence to guide such characterization is generally elusive since the DAS-produced analyses themselves are often the best estimates of the atmospheric state (by design) and, therefore, there is no independent dataset for determining errors. All observations have presumably been used, accounting optimally (to some degree) for their error statistics and accounting for their mutual relationships in time (using a forecast model for extrapolation or interpolation) or in space (*e.g.* quasi-geostrophy and spatial correlations) and thus robust independent datasets for verification are usually absent (although, *e.g.*, research data such as ozonesondes and ozone from some instruments are not commonly assimilated, and thus are available for independent verification). While some information about DAS errors can be derived from existing data sources, it necessarily is incomplete and imperfect. Although any OSSE is necessarily also an imperfect simulation of reality, the analysis and forecast errors can be completely and accurately computed and thus fully characterized within the simulated context.

The fact that they are widely used and relied upon does not mean that OSSEs, or the experimental results created by them, are free of controversy. Because of the wide-ranging consequences of decisions on major Earth Observing Systems, any OSSE results on which these decisions are based will have to withstand intense scrutiny and criticism. One goal of this chapter is to suggest ways in which OSSEs can be made robust and credible.

In this chapter we present the basic guidelines for conducting OSSEs. A historical review is provided, and experiences from OSSEs conducted at the National Centers for Environmental Prediction (NCEP OSSE) are presented; finally, conclusions and the way forward are outlined.

2 Historical summary of OSSEs

The OSSE approach was first adopted in the meteorological community to assess the impact of prospective observations, *i.e.*, not available from current instruments, in order to test potential improvements in numerical weather prediction, NWP (Nitta 1975; Atlas 1997; Lord *et al.* 1997; Atlas *et al.* 2003a). In a review paper, Arnold and Dey (1986) summarize the early history of OSSEs and present a description of the OSSE methodology, its capabilities and limitations, and considerations for the design of future experiments. Meanwhile, OSSEs have been performed to assess trade-offs in the design of observing networks and to test new observing systems (*e.g.* Stoffelen *et al.* 2006).

In early OSSE studies, the same model used to generate the “Nature Run” or truth was used to assimilate the synthetic data, and to run forecasts (Halem and Dlouhy 1984). In these so-called “identical twin” OSSEs the physical parametrizations and discretized dynamical processes in the assimilating model exactly represent those in the surrogate atmosphere. Model errors due to parametrization and numerical implementation are thus neglected and a free model forecast run from given initial conditions would provide identical results for the Nature Run and the DAS model. Consequently, forecast errors arising from deficiencies in the forecast model representation of the real atmosphere are not accounted for; only forecast errors due to errors in the initial conditions are represented. This limitation has been noted to lead to overly optimistic forecast skill in the OSSE DAS.

Another effect of the neglected model errors is that the differences between observations, both existing and future ones, and background (*i.e.*, forecast), O-B, tend to be smaller in case of an identical twin OSSE than in operational practice (Atlas 1997; Stoffelen *et al.* 2006). As a result, both the observation minus analysis (O-A) differences and analysis impact of the observations, A-B (analysis less background), tend to be smaller than expected. Several ways exist to test the reduced observation impact and overly optimistic forecast skill: *e.g.*, by comparing the O-B and O-A distributions, single observing system impacts, and forecast skill metrics in the OSSE and operational practice (calibration). The chapter *Evaluation of Assimilation Algorithms* (Talagrand) provides details of methods used to evaluate the assimilation process.

Since the DAS background model error space in identical twin OSSEs is limited with respect to an operational model’s error space, fewer observations are needed to correct the model state in the analysis step. In fact, the simulated observation set, unlike the real observations, has systematic characteristics consistent with the model formulation (*e.g.* scales of motion, mass-wind balance). Therefore, just a few observations could potentially correct the initial state errors and provide improved forecasts in an identical twin OSSE. On the other hand, as Atlas *et al.* (1985) point out, due to the simplified error space, observation “saturation” in the DAS will tend to occur at lower data volumes in an identical twin OSSE than in the case of assimilation of the real observations. This saturation may lead to underestimation of the impact of observing systems with extensive coverage (*e.g.* satellite systems). Moreover, observing systems that tend to correct errors due to numerical truncation of the dynamics or due to physical parametrization, may be undervalued. This potential non-linear effect of sampling on identical twin OSSE forecast scores, makes the above-mentioned calibration tests (involving, *e.g.*, O-A and O-B distributions) on the OSSE data assimilation system increasingly relevant.

Arnold and Dey (1986) recommend "fraternal twin" OSSEs as a way to address the shortcomings of "identical twin" OSSEs. In fraternal twin OSSEs, the NWP model used to simulate the observations is different from the forecast model in the OSSE data assimilation system, but not as different as the true atmosphere is from an operational forecast model. Examples can be found in Rohaly and Krishnamurti (1993), Keil (2004) and Lahoz *et al.* (2005). It is clear that the problems noted above with identical twin experiments will be reduced, but not absent for fraternal twin experiments. Stoffelen *et al.* (2006) test the absence of unrealistic observation impact in a fraternal twin OSSE. To avoid potential fraternal twin problems, the Nature Run and atmospheric data base may be produced at one NWP centre (Becker *et al.* 1996), while the impact experiments are run by another independent NWP centre (Masutani *et al.* 2006, 2009).

Another reported measure to reduce identical twin effects is to produce the Nature Run at high resolution and run the OSSE data assimilation system at lower spatial resolution. While useful for some studies, a potential disadvantage is that the observing system impact of a prospective system is tested at a resolution which is obsolete by the time the new observing system will be operationally implemented.

Atlas *et al.* (1985) report on the exaggerated OSSE impact of satellite-derived temperature soundings. At that time, the fraternal twin problem was raised as one cause, although these satellite soundings are rather abundant (see above). Other, and with hindsight perhaps more plausible, noted causes are:

- Simplified observation error characteristics. Observing systems can have complicated relationships (geophysical, spatial, and temporal) with the forecast model's atmospheric state and special care is needed to simulate them;
- The simulated observation coverage is over-optimistic. For example, the degree of cloud contamination of the measurements may be underestimated (*e.g.* Masutani *et al.* 1999);
- The simplifying assumption, usually made in OSSEs, that the distribution of observation errors is perfectly known;
- Temperature data are both simulated and assimilated, with no error from the Radiative Transfer Model (RTM) involved.

Again, comparison of observation impact and forecast skill, *e.g.*, by comparing the O-B and O-A distributions; single observing system impacts; and forecast skill metrics in the OSSE and operational practice involving OSSE calibration, should reveal such problems.

Various simulation experiments have been attempted which use real data for existing instruments and only simulate future instruments. These methods do not require a Nature Run and allow experimentation on a specific (extreme) weather event. Observing System Replacement Experiments (OSREs) could, for example, be used to test the impact of existing wind profile observations over Northern Hemisphere land and how these may be replaced by another observing system (Cress and Wergen 2001). Although an OSRE indicates how one could replace existing observing systems, it is, however, not *a priori* clear how to extrapolate these results to faithfully test new observation capabilities, *e.g.*, like the DWL (Doppler Wind Lidar) capability to resolve the incomplete wind profile coverage over the oceans. To test new observation capabilities, Marseille *et al.* (2008a-c) developed a method called the Sensitivity Observing System Experiment (SOSE). In a SOSE, adjoint sensitivity structures are used to define a pseudo-true atmospheric state for the simulation of the prospective observing system. In a SOSE the forecast error is projected back onto the initial state, thereby setting the maximum achievable forecast improvement. An alternative method, the Analysis Ensemble System (AES) (Tan *et al.* 2007) uses the spread in the ensemble as a proxy for the analysis and background uncertainty based on arguments of error growth (Fisher 2003). Since the background, analysis and observation errors are larger in the AES than in the real DAS, it is not clear whether the same set of observations in both systems remain optimal for reducing the background uncertainty. In order to test the realism of the OSRE, SOSE and AES, both the analysis and forecast impacts need to be carefully calibrated, just as in an OSSE.

In this chapter, the term OSSE (sometimes *full* OSSE to distinguish from other simulation experiments) refers to a simulation experiment with a Nature Run model significantly different from the NWP model used for data assimilation. This provides a truth independent of the data assimilation system NWP model and of the Global Observing System (GOS) data coverage and quality. In an OSSE, all observations used for the DAS have to be simulated from the Nature Run. In a SOSE, OSRE or AES, only the future observations are simulated from analysis or forecast fields. These fields used may have limitations in comparison with a Nature Run in terms of biases and temporal consistency due to the GOS, DAS and NWP (adjoint) model involved. It is considered that simulation of all observations is a significant initial investment for an OSSE, but that interpolating observations is part of a DAS. In OSSEs, all the usual analysis and forecast verification metrics can be used to evaluate data impact, and the simulated data can be tested with several different data assimilation systems with minor modification to the operational systems. The data impact for OSSEs (and their variants) often varies with verification metric and DAS used. Note, however, that a truth is available for further verification of the DAS characteristics. Although a SOSE, OSRE or AES allow quick study of real extreme events, the SOSE requires an adjoint model to generate the new observations and the AES requires an established ensemble system. Calibration and interpretation of the results is complicated and needs to be tested carefully for the SOSE, OSRE and AES. Full OSSEs with a long Nature Run allow quantitative assessment of the analysis and forecast impact. Note, however, that there are many OSSEs conducted without calibration. Furthermore, during the early years of OSSEs, identical twin OSSEs or fraternal twin OSSEs were often conducted due to the lack of variety in state-of-the-art NWP models.

To conclude, although initial investment is required for a full OSSE, it is today the most reliable strategy to use full OSSEs for impact assessment of prospective observing systems.

3 The Nature Run

The Nature Run is a long, uninterrupted forecast by a NWP model whose statistical behaviour matches that of the real atmosphere. The ideal Nature Run would be a coupled atmosphere-ocean-cryosphere model with a fully interactive lower boundary. However, it is still customary to supply the lower boundary conditions (sea surface temperature, SST, and ice cover) appropriate for the span of time being simulated. Meteorological science is approaching this ideal, but such coupled systems are not yet mature enough to be used for Nature Runs. Although fully coupled systems are available, their usefulness and accuracy for OSSEs is unknown. Preliminary tests, however, suggest that coupled systems may be good enough for operational NWP in the near future (Saha *et al*. 2006; Kistler *et al.* 2008).

The advantage of using a long, free-running forecast to simulate the Nature Run is that the simulated atmospheric system evolves continuously in a dynamically consistent way. One can extract atmospheric states at any time. Because the real atmosphere is a chaotic system governed mainly by conditions at its lower boundary, it diverges from the real atmosphere a few weeks after the simulation begins. This does not matter *provided that* the climatological statistics of the Nature Run match those of the real atmosphere. A Nature Run should be a separate universe, ultimately independent from but with very similar characteristics to the real atmosphere.

3.1 Characteristics of the Nature Run

One of the challenges for an OSSE is to demonstrate that the Nature Run does have the same statistical behaviour as the real atmosphere in every aspect relevant to the observing system under scrutiny. For example, an OSSE for a wind-finding lidar on board a satellite requires a Nature Run with realistic cloud climatology because lidars operate at wavelengths for which thick clouds are opaque. The cloud distribution thus determines the location and number of observations.

The Nature Run is central to an OSSE. It defines the *true* atmospheric state against which forecasts using simulated observations will be evaluated. This concept deserves more explanation. In 1986, Andrew Lorenc suggested the following definition of the “truth”: the projection of the true state of the atmosphere onto the model basis. As an example, if a spectral model produces a Nature Run, the true atmospheric state might be represented by spectral coefficients corresponding to triangular truncation at total wave number n (Tn) on L vertical levels. Atmospheric features too small to be captured by the model resolution are not incorporated in this truth.

The Nature Run is also the source of simulated observations. For each observing system, existing or future, a set of realistic observing times and locations is developed along with a list of observed parameters. An interpolation algorithm looks at the accumulated output of the Nature Run, goes to the proper time and location and then extracts the value of the observed parameter. If the Nature Run does not explicitly provide an observed parameter, the parameter is estimated from related variables that the model does provide. Because observations extracted from the Nature Run are the same as the defined truth (they are “perfect”), various sources of error must also be simulated and added to form observations with realistic accuracy with respect to the Nature Run itself.

Some OSSEs have used a succession of atmospheric analyses as a substitute for a Nature Run (Keil 2004; Lahoz *et al.* 2005). A succession of analyses is a collection of snapshots of the real atmosphere. For example, in the case of four-dimensional variational assimilation (4D-Var, chapter *Variational Assimilation*, Talagrand), although the analyses may each be a realizable model state, they all lie on different model trajectories. The background (first guess) lies on the same model trajectory as the previous analysis because, in 4D-Var the analysis is a realizable model state (it does not require separate initialization or balancing). Once this background is adjusted by new data in 4D-Var, the model lies on a new trajectory, which may be close to the old one (the one that the background was on) but is nonetheless different. Each analysis marks a discontinuity in model trajectory, determined by the information content extracted by a DAS from the existing global observing systems (chapter *The Global Observing System*, Thépaut and Andersson). Furthermore, residual systematic effects due to the spatially non-uniform and often biased observations, the DAS or the model state, may either favourably or unfavourably affect the potential of new observing systems to improve the forecasts. Thus, considering a succession of analyses as truth seriously compromises the attempt to conduct a "clean" experiment.

3.2 Evaluation and potential adjustment of the Nature Run

No Nature Run is perfect and its shortcomings need to be investigated by comparison with real-world climatology and, if the shortcomings can compromise a particular OSSE, adjustments to the Nature Run may be needed.

Several NCEP OSSEs (Masutani *et al.* 2006, 2009) have used the Nature Run with *T213* horizontal resolution and 31 vertical levels (*T213* NR) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and described in Becker *et al.* (1996). For the *T213* NR, quadratic grids with 60 km horizontal resolution were used to compute the physics. Note the corresponding linear grid space would be 90 km, which is more representative of the scale resolved by the *T213* NR. A one-month model run starting on 5 February 1993 was saved every six hours.

It is important that the Nature Run contain realistic clouds for evaluation of Doppler Wind Lidar (DWL) and cloud motion vector (CMV) data and simulation of radiances. Doppler Wind Lidar data can be retrieved only if the DWL shots hit the target. Clouds are important targets for a DWL but they also interfere with the DWL shots at lower atmospheric levels. Therefore, large differences in the Nature Run cloud amount will affect the sampling of simulated data. Realistic clouds are also necessary for generating realistic cloud track winds from geostationary platforms. Clouds moreover affect the sampling and simulation of radiance data.

The observed estimates for total cloud cover come from three different sources: the USAF Real-Time Nephanalysis (RTNEPH; Hamill 1992; Henderson-Sellers 1986); the International Satellite Cloud Climatology Project (ISCCP); and the NESDIS experimental product, Clouds from the Advanced Very high Resolution Radiometer (CLAVR-phase; http://cimss.ssec.wisc.edu/clavr). The differences between the total cloud cover (TCC) in the three observational sources and the Nature Run are within the variability of the observations. In the *T213* NR, the High-level Cloud Cover (HCC) amount seems larger than the satellite observed estimate across all areas of the globe. The amount of Low-level Cloud Cover (LCC) in the *T213* NR over the ocean is less than observed and the amount of LCC over snow is too high. After careful investigation, it was found that, due to the lack of reliable observations, there is no strong evidence for an over-estimation of HCC and polar cloud by the *T213* NR. However, the under-estimation of low level stratocumulus clouds over the oceans and its over-estimation over snow was clearly evident, and adjustments were consequently applied (see Masutani *et al.* 1999).

Although the OSSE using the *T213* NR produced many valuable results, it also had limitations. First of all, due to advances in model development, it is neither realistic nor suitable to use a Nature Run produced by a NWP model more than 10 years old to test a current DAS. Second, since there is a significant drift from analyses in the tropics during the first several weeks of the Nature Run, the one-month long Nature Run cannot be used to evaluate data impact in the tropics. The *T213* NR employed fixed SSTs; although fixed SSTs were not found to jeopardize the OSSEs, this is still a serious limitation of the *T213* NR. Note that a more recent *T511* Nature Run produced by ECMWF (Reale *et al.* 2007) showed a reduction of tropical convective rainfall during the first few weeks of the Nature Run period. This may mean that the Nature Run has much less convective rainfall compared to the real atmosphere, or that the analysis has too much convective rainfall compared to the real atmosphere. For the Nature Run to be useful, its statistics must lie within the climatological variability in the real analyses.

Producing accurate tropical forcing is a challenge for current NWP models. Nevertheless, the recent *T511* NR produced by ECMWF (see above) faithfully reproduces many aspects of the tropical atmosphere, at least in a statistical sense (Reale *et al.* 2007). For example, it reproduces the African Easterly Jet and African Easterly Waves in good agreement with observations.

There is great interest in OSSEs for studying forecasts of tropical waves and tropical cyclones (TCs). A prerequisite for such studies is a Nature Run that generates realistic tropical disturbances, *e.g.*, hurricanes with well defined warm cores and realistic tracks. However, there are still significant differences between model produced tropical cyclones and observations, and the interaction of TCs with SSTs requires further study (Tsutsui and Kasahara 1996). Finally, the properties of tropical cyclones relevant to the evaluation of a DAS are still to be investigated.

Mid latitude cyclone statistics in the Nature Run must also be realistic. The basic measures commonly used to compile mid latitude cyclone statistics are:

- Distribution of cyclone strength across a wide range of pressures;
- Cyclone lifespan;
- Cyclone deepening;
- Regions of cyclogenesis and cyclolysis;
- Distribution of cyclone speed and direction.

3.3 Requirements for a future Nature Run

The preparation of the Nature Run and the simulation of data from it consume significant resources. It is of practical importance to have one or two good-quality Nature Runs shared by many OSSEs. OSSEs with different Nature Runs are difficult to compare but OSSEs using different data assimilation systems and the same Nature Run can provide valuable cross-validation of data impact results. If Nature Runs are widely accessible, the Nature Runs and simulated data ought to be shared between many of the institutes carrying out the actual OSSEs.

The primary specifications of a Nature Run based on past experience of OSSEs are:

a. Employ a NWP model with demonstrated forecast skill;

b. Simulation span: since the data impact depends on the season, it is important that future Nature Runs cover long periods, preferably a whole year to allow selection of interesting subperiods for closer study;
c. Simulation sample: a temporal resolution higher than the OSSE analysis cycle. If more than one DAS is involved, this would ideally be a resolution higher than that of all participating data assimilation systems;
d. Simulation should resolve scales compatible with the main observing systems;
e. It is desirable that they should be based on an atmosphere-ocean coupled model; or at least, the Nature Run must be forced by an analysis incorporating frequently updated SST and sea ice;
f. Data archiving should be user-friendly and shareable with the community;
g. Simulation should agree with the real analyses in a statistical sense;
h. Chemistry and aerosol information which affect the data should be evaluated;
i. There should be a trade-off between the resolution and the complexity of the model;

The set of archived Nature Run variables should be enhanced to accommodate the need for OSSEs. For example, geopotential height at model levels is very desirable. Archiving of this variable will help the simulation of observations based on height coordinates, such as those from DWL and profilers. Low resolution pressure level data and isentropic level data output on a standard grid are also very useful for OSSEs, as they can be used for verification of the experiments. However, producing these verification datasets can take up significant resources at the initial stages of setting up an OSSE.

A main requirement of full OSSE experiments is to avoid the identical or fraternal twin ("incest") problem, as discussed in Section 2. If the model from which hypothetical observations are extracted is the same as the assimilating model, the OSSE results will show unrealistic observation impact and overly optimistic forecast skill (Arnold and Dey 1986; Stoffelen *et al.* 2006). Thus the forecast model used for the Nature Run should *not* be used later on for DAS experiments in the full OSSE.

4 Assignment of realistic observation errors

The following definitions concern data assimilation in general – see also chapter *Mathematical Concepts of Data Assimilation* (Nichols). In the definitions below, **x**, **y** and $\boldsymbol{\varepsilon}$ are vectors, and $\mathscr{H}$ is a non-linear operator.

a) ***The observation:*** $$\mathbf{y} = \mathbf{y}_t + \boldsymbol{\varepsilon}_m$$

y is the observed value, measured by some instrument, and $\boldsymbol{\varepsilon}_m$ is the observation error. The subscript t refers to the true atmospheric value. We define the true value as the weighted average of the true atmospheric values within the volume sampled by the instrument. Petersen (1968) defined the "true" observation in this way, but quantitatively by means of an integral. Different instruments sample different volumes so that the true value of temperature appropriate for a radiosonde may not match the true value appropriate for the AMDAR (Aircraft Meteorological DAta Relay) system aboard a commercial jet, even if the two observations are assigned to the same location and time. Thus, the observed "truth" is very much scale-dependent, but defining it in this way is consistent with the definition of truth with respect to model resolution as proposed by Lorenc (1986) and discussed immediately below.

$\boldsymbol{\varepsilon}_m$ refers to errors incurred during measurement or subsequent data processing. The errors can be random or systematic (*i.e.*, biased).

b) ***The model state:*** $\mathbf{x} = \mathbf{x}_t + \boldsymbol{\varepsilon}_f$

$\boldsymbol{\varepsilon}_f$ is the model state error. The state of a DAS model is defined by a set of parameters stored at the points of a model grid, or, alternatively, by a set of spectral coefficients. As noted above, we follow Lorenc (1986), in defining the true model state $\mathbf{x}_t$ as the true atmospheric state containing all scales from long waves down to cloud microphysics, but spectrally truncated at the model grid. Scales of motion that cannot be captured by the model grid (or the spectral truncation) are not included in the definition of the true state. The numerical model forecasts the state **x**, but the forecast is subject to error $\boldsymbol{\varepsilon}_f$, which is the result of truncation associated with finite differencing, imperfect dynamics, and errors in the representation of physical processes, whether parametrized or not.

c) ***The forward model:*** $\mathscr{H}(\mathbf{x})$

Forecasts are usually verified against observations (sometimes against an analysis). Because observations hardly ever coincide with model grid points, it is necessary to map the model forecast to the observations in order to make a direct comparison. The forward model $\mathscr{H}$ does this. Another name for $\mathscr{H}$ is the *observation operator,* because $\mathscr{H}$ operates on the model grid to generate a pseudo-observation, a best estimate of the observed value. It relies on the parameters computed by the model on the model grid in order to make a best estimate of the observed value. Sometimes the calculation is as simple as 3-D linear interpolation, but if the observed quantity does not match one of the predicted quantities, then $\mathscr{H}$ will also involve a transformation of variables. For example, the model may predict relative humidity, but the observed quantity is column integrated water vapour. In this case, in addition to interpolation, the forward model has to convert the predicted relative humidity and temperature to a specific humidity and integrate the specific humidity vertically from the surface to the top of the model atmosphere.

d) ***Representativeness:*** $\mathbf{y}_t = \mathscr{H}(\mathbf{x}_t) + \boldsymbol{\varepsilon}_r$

If the forward model $\mathscr{H}$ could be applied to the true values $\mathbf{x}_t$ (unknown in practice) on the model grid, we would have an observation that still lacks a representativeness error $\boldsymbol{\varepsilon}_r$. The representativeness error has two causes:

1) The model grid volume does not match the atmospheric volume that is the object of measurement. If the observed volume is small compared to the model grid volume, the measurement will represent scales of motion that the model grid cannot resolve. From the model's point of view, the observation contains subgrid scale noise, and this will contribute to the value of $\boldsymbol{\varepsilon}_r$. In other words, because the representation of $\mathbf{x}_t$ is spectrally truncated, the projection $\mathscr{H}(\mathbf{x}_t)$ does not capture the subgrid scale atmospheric variance inherent in the observation. If the observed volume is larger than the model grid volume (*e.g.* a measurement of radiance in the microwave portion of the electromagnetic spectrum could involve a volume of atmosphere larger than the model grid volume), then the forward model will be an

averaging operator rather than an interpolation operator. From the model's point of view, the observation is too smooth and $\boldsymbol{\varepsilon}_r$ will relate to how well the model average spatially and temporally represents $\mathbf{y}$.

2) If a transformation of variables is included in $\mathscr{H}$, the relationship is imperfectly known or it is approximated in order to minimize the number of computations, *e.g.*, in case of radiance observations. This also contributes to $\boldsymbol{\varepsilon}_r$. In fact, any operation incorporated in $\mathscr{H}$ may contribute an error component to $\boldsymbol{\varepsilon}_r$.

To summarize, representativeness error arises from the mismatch between the DAS model grid volume and the volume sampled by the instrument, and also from a mismatch between the observed and predicted variables.

Some aspects of the representativeness error are not random but systematic. Even if we exclude subgrid effects that may be small if the model resolution is high enough, a computationally fast radiative transfer model applied to an atmospheric profile will generally yield imperfect radiances compared to the real atmosphere. This error will be almost identical whenever the atmospheric profiles are the same, since the model and physics remain unchanged. If such imperfections are complex functions of the atmospheric state, they may appear as random errors when computed from collections of states although they are in fact systematic. Modelling the representativeness error as though it were random may therefore introduce unrealistic effects if some aspect of the systematic nature of the error is important.

In all aspects of the data assimilation problem, representativeness error appears combined with instrument error. The combined error is called the "observation error", and its covariance (denoted by $\mathbf{R}$) is a key statistic that determines the analysis error covariance. If the instrument and representativeness errors are uncorrelated, then $\mathbf{R}$=$\mathbf{E}$+$\mathbf{F}$, where $\mathbf{E}$ and $\mathbf{F}$ are the covariances of instrument and representativeness errors, respectively:

$$\mathbf{E} = \mathscr{E}[(\boldsymbol{\varepsilon}_m - \langle\boldsymbol{\varepsilon}_m\rangle)(\boldsymbol{\varepsilon}_m - \langle\boldsymbol{\varepsilon}_m\rangle)^T] \text{ and}$$

$$\mathbf{F} = \mathscr{E}[(\boldsymbol{\varepsilon}_r - \langle\boldsymbol{\varepsilon}_r\rangle)(\boldsymbol{\varepsilon}_r - \langle\boldsymbol{\varepsilon}_r\rangle)^T],$$

where $\left\langle \quad \right\rangle$ denotes an average, $\mathscr{E}[.]$ denotes expectation value, and the superscript T means vector transpose. The $\mathbf{R}$, rather than $\mathbf{E}$ or $\mathbf{F}$, is actually specified in the DAS. $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{R}$ are matrices.

Techniques to estimate $\mathbf{R}$ are imperfect. A poor specification will yield a suboptimal system; *i.e.*, one with larger analysis error variance than otherwise. Generally, some further tuning of the error estimates is conducted so that the $\mathbf{R}$ incorporated in the system experiments appears close to optimal. These are, therefore, generally the values that must be duplicated as the observational errors in the OSSE if the responses in the real and simulated systems are to appear similar.

e) ***Application to OSSEs:***

In practice, real observations come with only an instrument error; they are inherently *representative* of the volume of atmosphere sampled. The representativeness error arises from the forward operator and has the two components mentioned above. We account for instrument error and, to be rigorous, also for the representativeness error, when we specify the observation error covariance in the DAS penalty function that is part of the variational analysis. In practice, we compute $\mathscr{H}(\mathbf{x})$, not $\mathscr{H}(\mathbf{x}_t)$.

By contrast, in an OSSE, one uses a forward model to *generate* an observation. After the forward model is applied to the grid point values of the Nature Run, we must add a random contribution $\boldsymbol{\varepsilon}_r$ to the forward model output. The finer the resolution of the Nature Run and the more accurate the forward model, the smaller the representativeness error will be. Finally, we must also add an appropriate instrument error to improve realism. In summary, we must compute:

$$\mathbf{y} = \mathbf{y}_t + \boldsymbol{\varepsilon}_m = \mathscr{H}(\mathbf{x}_t) + \boldsymbol{\varepsilon}_r + \boldsymbol{\varepsilon}_m .$$

The random contribution $\boldsymbol{\varepsilon}_r$ accounts for the missing subgrid scale variance, say $\boldsymbol{\varepsilon}_r^{\mathrm{S}}$ and any error associated with a transformation in the forward model, say $\boldsymbol{\varepsilon}_r^{\mathrm{H}} = \mathscr{H}_t(\mathbf{x}_t) - \mathscr{H}(\mathbf{x}_t)$ (so $\boldsymbol{\varepsilon}_r = \boldsymbol{\varepsilon}_r^{\mathrm{S}} + \boldsymbol{\varepsilon}_r^{\mathrm{H}}$), where $\mathscr{H}_t$ is a hypothetical perfect forward model that operates on the NWP model defined truth. We find that:

$$\mathbf{y} = \mathscr{H}(\mathbf{x}_t) + \boldsymbol{\varepsilon}^S{}_r - \mathscr{H}(\mathbf{x}_t) + \mathscr{H}_t(\mathbf{x}_t) + \boldsymbol{\varepsilon}_m = \mathscr{H}_t(\mathbf{x}_t) + \boldsymbol{\varepsilon}^S{}_r + \boldsymbol{\varepsilon}_m$$

Note that $\mathbf{y}$ represents the Nature Run transformed through the hypothetical perfect forward model. Additionally, one should consider whether the difference $\mathscr{H}_t(\mathbf{x}_t) - \mathscr{H}(\mathbf{x}_t)$ might have a systematic component (*i.e.*, a bias), since a normal random error distribution is assumed above in $\boldsymbol{\varepsilon}_r$.

5 Simulation of observations

5.1 Basic guidelines

Although a particular OSSE may be motivated by evaluation of a single instrument, it is still generally necessary to simulate all observations that are expected to be used along with it. Even a poor observing system will be better than none at all since the atmosphere is chaotic. Irrespective of how close to the real atmosphere a data assimilation experiment begins, without the constraint of further observations, after 15 days or so it will diverge to states expected to be as dissimilar to the atmosphere as two states randomly selected for the same month but different years. Thus, using a single observation type in an OSSE with other observations excluded results in a very large impact compared with no assimilation at all, but a much smaller and more realistic impact if other observations *are* considered.

Current observations quite effectively constrain the atmospheric analysis. In many places, the expected error variance of the analysis is less than that of most observations that have been employed in the analysis of the DAS model state. (Note that many observations contain more information about the local atmosphere than the analysis; however, in the truncated model domain, the errors of these observations are larger, due to the representativeness error.) The analysis is better because it has used all nearby observations, including those implied by the background, accounting for the error statistics of each, at least in a crude but still useful way. The weighting of a new observation within the DAS will be determined by the presence of other observations. The impact of any additional observation essentially competes with that of all others. When the impacts of any single observation type are measured, therefore, the improvements to the analysis or forecasts are generally quite small. Progress occurs when innovative instruments are added to those already used, but by small steps rather than great leaps.

Once the Nature Run is sufficiently validated, observations may be simulated. To do so, it is necessary to understand the relationship between the observations and the atmosphere, both the real atmosphere and the one represented by the Nature Run. Furthermore, at the next step in preparing the OSSE, simulated errors are generated to add to the corresponding simulated observations. The accuracy with which the DAS can reproduce the Nature Run in the OSSE will depend strongly on the characteristics of the errors associated with the observations. Prior to selecting a method for simulating the observations, it is therefore prudent to also understand the nature of all the types of error realistically associated with them.

Various observing instruments are designed to respond to differing atmospheric characteristics. Here, two such instruments will be contrasted: (i) a radiosonde, and (ii) a satellite instrument that measures infrared radiances. Together they represent several of the various aspects that must be considered when simulating observations and their errors.

The radiosonde is a comparatively simple instrument with a thermo-resister used to measure temperature as the balloon ascends. The measurement is made along short segments of the trajectory of the balloon, with their length determined by the response and reporting times of the instrument. Compared with the much coarser resolution represented by the Nature Run, these may be considered as (almost) point values that are affected by all spatial scales. A function must therefore be developed to relate the observed value to the atmosphere as represented by the assimilating model (*i.e*., a function for the spatial representativeness error).

The other instrument is on board a satellite designed to measure infrared radiances coming from the Earth and atmosphere below. The satellite actually measures the energy of photons over some range of electromagnetic wavelengths collected on an antenna (see chapter *Research Satellites*, Lahoz). For the purpose of NWP as opposed to climate monitoring, data assimilation is mainly concerned with atmospheric fields: temperature, wind, pressure and constituents (*e.g.* water vapour, ozone, and perhaps minor species and aerosols). The observed radiances must be related to these fields if they are to be useful. Presumably an appropriate relationship exists; otherwise the observation would not be used for this purpose. The antenna collects radiances emitted from a possibly large volume of the atmosphere and is therefore most accurately related to some kind of average (with spatial weights determined by the viewing characteristics of the antenna and the orbiting satellite). This average will not in general correspond to that defining a grid volume average in a model data representation. Thus, some spatial interpolation or integrating relationship must also be defined.

For any observation types already used within a DAS, a useful relationship between what is observed and the representation of fields being analysed necessarily already exists. In the standard notation used for atmospheric data assimilation, this is the operator $\mathscr{H}$ that acts on the background field during the assimilation cycle (see Section 4). For a new instrument not yet used, this operator needs to be developed. Development can be either empirical or physically based.

In general, $\mathscr{H}$ can be expanded into a sequence of one to several distinct functions acting on the state **x**; *e.g.*, as:

$$\mathscr{H}(\mathbf{x}) = \mathscr{S}(\mathscr{F}(\mathscr{I}(\mathbf{x})))$$

The function $\mathscr{I}$ denotes a possible interpolation from grid-point (or other discrete data representation) to observation locations; $\mathscr{F}$ denotes a possible physical (or other) relationship such as radiative transfer relating temperature and moisture to satellite-observed radiances; $\mathscr{S}$ denotes a possible integration of values, such as along a line of sight or within an antenna footprint. Any of $\mathscr{S}$, $\mathscr{F}$, or $\mathscr{I}$ may be absent for a particular observation type, and some types may be better described by a different sequence of operators or the employment of additional ones. The equation should therefore be considered as schematic, although for some observation types the presentation may be precise.

The more realistic the relationship between values representing the model state and the observed quantity, the more useful the real observation will be to the DAS and, correspondingly, the more realistic the simulated observation will be in the OSSE. A problem is that the time to develop the most accurate relationships may be prohibitive, and the benefits may be tiny compared to other shortcomings in the system. A relationship must be designed to be "good enough" for the intended purpose. Results must be carefully interpreted mindful of these criteria. The way these choices are evaluated will depend on the purpose. Inaccuracies in the results when compared to the "true" physical relationship can be handled to some degree by the statistical approach to representing errors in the DAS.

The $\mathscr{H}$ is designed with speed as well as accuracy in mind, especially if the DAS solves a large variational problem. In that case, a tangent linear version of $\mathscr{H}$ and its adjoint (see chapter *Variational Assimilation*, Talagrand) are generally applied to every iteration of the analysis increments (*i.e.*, the difference between the analysis and the forecast). Thus, some compromises may be made that are not necessary when speed is not an issue. An example of this latter case is the generation of simulated observations from the Nature Run; these need only be produced once to be used in all subsequent relevant OSSEs. Thus, the simulation of observations from the Nature Run need not be done in the same way as the assimilation model. In fact, there are good reasons for selecting a different algorithm. These and other considerations are described in the next section.

5.2 Specific issues related to different observational types

Standard and simple forward models are used for extracting observed quantities from the "true" (*i.e.*, Nature Run) background fields as the basis for the simulation of observations for use in OSSE experiments. This procedure will inevitably omit some fraction of the error (from instrument variability and lack of model representativeness) to be found in real observations. Thus, simulation of observations for OSSE work is usually thought of as the synthesis of a signal from the background truth field (often referred to as a "perfect" observation), and some appropriate amount of noise, or "error." If the noise or error is indeed appropriate, then the impact of simulated observations on an OSSE will be similar to the impact that real-world observations have on operational assimilation. Although the instrument errors are in most cases fairly well defined, the derivation of the total error levels appropriate for application to perfect observations is a complex subject. This section describes some of the issues surrounding the creation of the perfect part of simulated observations.

5.2.1 Simulation of conventional observations

In order to create perfect observations, it is only necessary to locate the observation type to be simulated in the space and time coordinates of the background field. The most straightforward approach to this problem, for the case of simulating existing data sources, is just to use the locations of real observations for any given time and place. In the case of conventional observation sources (for example, TEMP, PILOT, SYNOP, AIREP, SHIP, BUOY, SATOB; see chapter *Assimilation of Operational Data*, Andersson and Thépaut) real world data patterns are readily available, and the specification of realistic simulated data patterns for these data types is simple. For the purposes of many OSSE experiments already conducted, this technique of locating conventional observing patterns is sufficient. However, in the set of simulated observations, the effects of observation circumstances and the expected evolution of the observing system should also be taken into account. Below we discuss several examples.

Radiosonde launch points can be located from existing real world datasets, but the balloon ascent and drift will depend on the atmosphere being sampled. The track of each radiosonde can be calculated using relatively simple transport models. For maximum realism, the calculation should be stepped at intervals sufficiently small to obtain information from the full vertical resolution of the Nature Run true fields. The resulting simulated profiles might be used without change in OSSE experiments, but would more likely be transformed into the more recognizable pattern of mandatory and significant vertical levels as presented to an operational DAS.

Surface land observations (for example, SYNOPs, METAR) present several issues to be considered for achieving realistic simulations. The question of location involves mainly the surface elevation and the measuring height. Although most real-world analogues contain some measure of the observation height, it may be advantageous in some cases to use a very high resolution digital elevation model and tables of particular instrument measuring heights to locate these data. There is also a need to interpolate surface values from the Nature Run background fields on a smoothed topography to a realistic topography of simulated observation points.

Commercial aircraft, the source of most aircraft observations, fly routes which use wind patterns to save fuel cost and avoid turbulence. Ideally, flight tracks for the OSSE should be formulated for simulated aircraft in the same way as they are for real cases. However, the location of jets and turbulence can be very different for the Nature Run and the real world; the flight planning software is complicated, proprietary and even unique to individual airlines. It may be possible and worthwhile to develop a simplified generalized approach to formulating simulated flight track planning based on some general principles, *in lieu* of using the actual software employed by the airlines.

Cloud-tracked wind observations, and their unique observing errors, will depend on the specification and perception of cloud fields from the Nature Run. Satellite-borne instruments and observations of all types have unique relationships with various types of clouds, so this is a very important aspect for realistic simulation of satellite-based observations.

In general, it seems desirable to make use of synoptic features from the background truth fields to determine realistic locations for all simulated observations, at least to the extent this can be accomplished without exerting undue effort, or employing unrealistic assumptions. Many more OSSE experiments will need to be designed, conducted, and carefully examined in order to determine how important a realistic distribution of simulated observation locations is.

5.2.2 Simulation of radiance data

For the NCEP OSSE (see Section 9), the use of different Radiative Transfer Models (RTMs) for simulation and assimilation helps understand the errors associated with RTMs. Radiative transfer models used for simulation have been generally based on the RTTOV-6 (Radiative Transfer for TOVS) algorithm (Saunders *et al.* 1999). At NCEP, the OPTRAN model developed by NESDIS was used in the assimilation (Kleespies *et al.* 2004). Brightness temperatures were simulated and level-1B radiances synthesized with correlated measurement errors; the impact of clouds was also considered (Kleespies and Cosby 2001). Currently, the Community Radiative Transfer Model (CRTM) (Han *et al.* 2006; Weng 2007) and RTTOV are widely used in operational data assimilation systems. The SARTA (Stand-alone AIRS Radiative Transfer Algorithm) model (Strow *et al.* 1998) is also available and has been routinely used to simulate radiance data. These models allow the implementation of OSSEs using different RTMs for simulation and assimilation.

The simulation of radiances involves many procedures: simulation of orbits, evaluation of cloudiness, and assignment of surface conditions. Various properties such as surface emissivity and spectral response function have to be evaluated for each instrument. The characteristics of the instruments can change after launch, requiring a different set of coefficients at each stage. Ideally, the radiance data would be simulated as the Nature Run is produced. However, it is safer to save the Nature Run output frequently and simulate the radiance data afterwards, since radiances have to be simulated repeatedly with various conditions and error assignments.

If only clear-sky radiance data are used, a subgrid-scale sampling algorithm has to be developed when the radiances are simulated. If the footprint sizes are smaller than the Nature Run grid spacing, clear radiance data through small holes within the cloudy grid have to be simulated. Using a probabilistic procedure to simulate cloud porosity is a possible way to produce the correct statistics. A functional relationship between clear sky probability and cloud fraction profile has to be derived to obtain a reasonable distribution (*e.g*. Marseille and Stoffelen 2003). If the cloud cover is used simply as a cut-off criterion for clear sky radiances, much of the clear sky radiance data from the porous areas of cumulus clouds are eliminated and large amounts of radiance data from above the clouds will be eliminated. Note that there are many stratospheric channels which are never affected by cloud.

Although both the OPTRAN and RTTOV models can simulate cloudy radiances, cloudy radiances have not been used in data assimilation systems (McNally *et al.*

2000). Further development of RTMs will include cloudy radiances in data assimilation systems (Liu and Weng 2006a, b). Cloudy radiances allow the simulation of imagery and moisture channels. While most of these channels may not be used for data assimilation, imagery and moisture channels can be used with observations to evaluate the Nature Run as well as the RTM itself. Note that since the Nature Run does not resolve cloud scales, even when radiances are modelled through cloud fraction, subgrid-scale clouds still need to be represented appropriately (*e.g.* in a statistical sense). Modelling the subgrid-scale cloud remains important to simulate cloudy radiances and for assimilation of radiance data. Testing RTMs with clouds is an important area for OSSEs.

Calibration of the radiance data includes a sampling algorithm which produces a similar distribution of observations as the real data. The adjoint technique (Zhu and Gelaro 2008) is especially useful in the calibration of radiance data, as it allows the skill of an individual channel to be assessed. The skill has to be evaluated for various conditions, as real errors are likely to be a function of geography, local atmospheric flow, season, and viewing angle. These errors are also likely to be correlated. The bias, variance, error correlation, and distribution function for the errors have to be modelled to be used by any data assimilation system. Bias correction is now a part of data assimilation systems (chapter *Bias Estimation*, Ménard). As a result, one can bias correct the Nature Run radiances or implement the bias correction in the DAS itself.

5.2.3 Simulation of Doppler Wind Lidar (DWL) data

As noted in the introduction (Section 1), one of the primary uses of OSSEs is to investigate and quantify the potential impact of a new observing system or combination of observing systems not currently being used together. No other instrument has been subjected to OSSE evaluation more than the Doppler Wind Lidar (DWL). With only radiosondes and a few radar wind profilers providing complete vertical profiles of the horizontal wind vector, gaining insight into the impact of a new wind profiler, especially over oceans and sparsely populated land areas, requires simulating the performance of the sounder without the benefit of a heritage instrument. Issues of observation errors including measurement errors and error of representativeness must be addressed. The DWL instrument is critically affected by both clouds and aerosols. While clouds are represented reasonably well by current numerical models, aerosols are not.

In the United States, NASA and the Department of Defense (DoD) have supported the development of a Doppler Lidar Simulation Model, DLSM (Wood *et al.* 2000; Emmitt and Wood 2001). The DLSM was designed specifically to operate with the Nature Runs generated for OSSEs. Much attention has been given to incorporating cloud effects on the scale of the lidar beams (~100 m) and representing subgrid-scale turbulence that would affect the precision of the DWL line-of-sight (LOS) measurement (Emmitt and Wood 1989, 1991a).

A major role for OSSEs in preparing for a space-based DWL mission has been the generation of data requirements and subsequently derived instrument design specifications (Atlas *et al.* 2003b). Instrument designers have used the DLSM to conduct NWP impact trade-off studies related to orbit, instrument wavelengths, laser pulse energies, and signal processing strategies (Emmitt and Wood 1991b). NASA and NOAA have conducted numerous OSSEs using DWL observations simulated by the DLSM (Atlas and Emmitt 1995; Lord *et al.* 2002; Masutani *et al.* 2003; Riishøjgaard *et al.* 2003; Woollen *et al.* 2008).

In Europe, a similar Doppler Lidar In-space Performance Atmospheric Simulator (LIPAS) has been developed (Marseille and Stoffelen 2003) in support of the ADM-Aeolus mission to fly a space-borne DWL in 2011 (Stoffelen *et al.* 2005) - see chapter *Research Satellites* (Lahoz). LIPAS has been used to conduct OSSEs (Stoffelen *et al.* 2006) and simulates aerosol variability, vertical overlap of clouds and all relevant instrument performance characteristics.

The usual OSSE process involves a team composed of representatives of the operational weather forecasting community, instrument specialists and data stakeholders. The availability of models such as the DLSM and LIPAS allows the optimistic perspective of the instrument proposers and the more cautious expectations from the NWP communities to be explored over a range of assumed instrument performance within a realistic model and data assimilation environment. In the case of the DWL, the competition with other sources of wind information (including wind information contained in the background state) leads to an integrated impact which is usually more modest than that expected by the technologists. On the other hand, synergies with other sources of wind information (*e.g.* scatterometers and cloud motion vectors) are illuminated in ways not easily quantified without the OSSE.

6 Initial conditions and spin-up period

6.1 Initial conditions

The initial conditions for an OSSE must be generated carefully to reduce noise due to the difference between the Nature Run and the NWP model used for OSSEs. If an appropriate initial condition is not used, the OSSE will be contaminated by noise from the initial conditions and it will be hard to assess the data impact.

When starting a limited-period OSSE at some point within the Nature Run, initial conditions have to be generated carefully. If the initial conditions are generated from a different model, large biases between the models have to be removed, and some model variables may have to be estimated. Possible strategies to generate initial conditions include:

i) Generate the initial conditions by interpolation from the Nature Run.

It is possible to interpolate the initial conditions from the Nature Run to an OSSE model grid and use this as the initial conditions. As there is a large amount of noise produced from inconsistent initial conditions, it usually takes a few weeks for the OSSE to settle down. This procedure requires careful development of the interpolation procedure. Both the differences between the model variables and the bias between the Nature Run and the OSSE data assimilation system have to be carefully handled.

ii) Take the initial conditions from a precursor analysis.

In this approach one generates a precursor analysis starting from the same time and date as the Nature Run and uses the analysis as the initial conditions with the same DAS used for the OSSEs. The precursor analysis does not have to be of a high resolution but should be provided at the lowest resolution used for the OSSE. Not all operational data have to be included, but there should be enough data over the ocean to provide a reasonable description of large scale features, particularly in the Southern Hemisphere.

If the DAS used for the precursor run is the same as the OSSE data assimilation system, but has a higher resolution than the precursor analysis, the transition from the precursor analysis to the OSSE will be smooth. However, it takes a few time steps for the OSSE system to show the full resolution features.

If the OSSE DAS is different from the DAS used for the precursor run, an interpolation has to be performed. Exchanging analyses between different DAS is routinely done in real operational forecasting. This process can also be evaluated by OSSEs.

6.2 Spin-up period

A real analysis is used for the initial conditions of the Nature Run. During the first 2 to 3 weeks, a drift occurs from the real atmosphere to the model atmosphere, particularly in the tropics. This period (called the spin-up period) should not be used for an OSSE because it lies within the limit of predictability (at least for the largest scales) and still contains traces of the real atmospheric conditions.

The Nature Run NWP model and initial analysis have errors that depend on the real atmospheric state due to data distribution, and DAS and NWP model specification. When the Nature Run state has evolved to one which is unrelated to the real atmosphere, these errors can be assumed to have disappeared. One can use trends in the O-B (observation minus background) and O-A (observation minus analysis) differences to determine whether the Nature Run errors are independent of those of the real atmosphere. Depending on the type of experiment, the time for error independence to occur could be less than 2-3 weeks (see above).

7 Evaluation of OSSE results

The data impact in an analysis and forecast could be very different. For example, if the model is not performing well, large differences between the background (forecast) and observations will create a large analysis impact; however, that improvement will not be maintained in the forecast skill. On the other hand, a small analysis impact may become a large forecast improvement in areas where the model is performing well. The areas showing data impact in the analysis and forecast may not be the same. Improvements can also propagate between regions: *e.g.*, improvements in upper level wind will propagate towards lower levels in the forecast.

Data impact varies with spatial and time scales. For example, the impact in the mass fields could be very different from the impact in the wind fields. Below we discuss various aspects of data impact.

7.1 Data denial (or adding) experiments (DDEs)

The most common method used to test the impact of specific data is to compare the analysis and forecast skill with and without the specific data. Many diagnostic methods used to evaluate the Nature Run can also be used to evaluate the forecast and analysis. With real data the impact is measured as the forecast skill without the specific data compared against the best analysis or fit to observations. Usually, the analysis with the most data is considered to be the best and used as the control (defined in Section 1). Various skill scores for simulated experiments can be evaluated against either the control experiment or the Nature Run itself, while experiments with real data can be evaluated only against the control.

There are many evaluation methods, but it is important to produce a consistent evaluation for all experiments when the results are compared. Many diagnostic techniques used to evaluate the Nature Run can also be used to evaluate the results. Examples are given below.

1) Root Mean Square Error (RMSE). Root mean square error does not require climatology; therefore, this is the easiest evaluation that can be performed, and is often the first evaluation to be implemented. In a real system, RMSE is computed as the departure from the control experiment, which is usually the analysis with the most observations. For simulated experiments, RMSE can be computed from the departure from the Nature Run. The RMSE can be evaluated with the zonal mean or the time mean removed;

2) Anomaly correlation (AC). Anomaly correlation is affected by the climatology used, so it is important to use the same climatology for all skill comparisons. It is better to use a less than perfect climatology than to use different climatologies in skill comparisons. Traditionally, the AC of the 500 hPa geopotential height has been used, but Masutani *et al.* (2006, 2009) showed that other levels and variables need to be evaluated. Calculating ACs for different spatial scales is also crucial;

3) Storm track and intensity. Evaluations are done to determine the improvement in the storm track for selected events;

4) Fit to observations. This requires a forward model (see Section 4). For the NCEP OSSEs, an evaluation against Nature Run will replace this method. It is still important to compare the fit to observations during the calibration process, *i.e.*, test the realism of the O-B and the O-A distributions (see chapter *Evaluation of Assimilation Algorithms*, Talagrand);

5) Evaluation of the realism of a Nature Run by assessing the likelihood of extremes lying outside the normal range of analysed or measured values;

6) Amplitude, wavelength and propagation speed (or phase) of waves;

7) Comparisons which may shed light on the realism of disturbances in the model and identify possibly unrealistic or spurious scales of motion;

8) Evaluating the analysis and forecast of precipitation using, *e.g.*, threat scores TS ($TS=AC/(AF+AO-AC)$), where AC=area correct, AF=area forecast, AO=area observed);

9) The statistics of analysis increments. Errico *et al.* (2007) showed that the spectral decomposition of analysis increments reveals the performance of a DAS.

7.2 Adjoint–based techniques

An adjoint–based technique (ADJ) to estimate the impact of observations on NWP analyses has been developed and is described in detail in Langland and Baker (2004) – see also chapter *Variational Assimilation* (Talagrand). This is a powerful method that describes the contributions from different observations. This technique allows detection of impact, be it positive or negative, from any observation. There are advantages and disadvantages compared with Data Denial Experiments (DDEs) (Zhu and Gelaro 2008; Gelaro and Zhu 2009):

- The ADJ measures the impacts of observations in the context of all other observations present in the assimilation system, while the observing system is modified in the DDE (*i.e.*, gain matrix differs for each DDE member);
- The ADJ measures the impact of observations separately at every analysis cycle versus the background, while the DDE measures the total impact of removing data information accumulated in both the background and analysis;
- The ADJ measures the response of a single forecast metric to all perturbations of the observing system, while the OSE measures the effect of a single perturbation on all forecast metrics;
- The ADJ is restricted by the tangent linear assumption (valid ~1-3 days), while the DDE is not;
- The ADJ and DDE techniques produce a similar qualitative pattern on the short-term forecast with some exceptions;
- The ADJ may help our understanding in the interactions and redundancies among various observing systems.

8 Calibration of OSSEs

Calibration of OSSEs verifies the simulated data impact by comparing it to real data impact. In order to conduct an OSSE calibration, the data impact of existing instruments has to be compared to their impact in the OSSE.

The simulated impact experiments should mimic the equivalent real experiments. In any case, the observation-minus-background (*i.e.*, forecast) difference is the sum of three terms: the measurement error, the representativeness error, and a background error transformed by $\mathscr{H}$. Realistic estimates of the variances and spatial covariance of these errors must be made for an effective OSSE. One way to ensure that measurement errors, representativeness errors, and forecast (background) errors are all properly specified is to compare the statistical properties of $\mathbf{y}$-$\mathscr{H}(\mathbf{x})$ (the innovation) of the OSSE with those of the real world assimilation $\mathbf{y}$-$\mathscr{H}(\mathbf{x})$ for each observing system; they should match. Similarly, the statistical properties of the analysis increments for the OSSE and the real world assimilation should match. Thus, distributions of observation minus background (O-B) differences and observation minus analysis (O-A) differences for each observation type in the simulation should be similar to the statistics in an equivalent experiment with real data. In effect, the simulated observations should force the OSSE model state toward the Nature Run in the same way that real observations force the operational model state toward the projected true atmospheric state.

One way of calibrating an OSSE is to use a DDE (see Section 7.1) to find out whether the assimilation of a specific type of observation has the same statistical effect on a forecast within the simulation as it does in the real world. For example, if automated aircraft reports are withheld from an operational data assimilation system, will the statistical measures of forecast degradation be the same as they would be in a system where all observation types are simulated and the Nature Run provides truth?

An alternative method of calibration is to use the ADJ (see Section 7.2) to adjust the observational error so as to achieve a similar data impact with real observations.

When calibrating the OSSE, similarity in the amount of impact from existing data in the real and simulated atmospheres needs to be achieved. If the impacts are different this needs to be explained. For example, synoptic systems in the Nature Run and the real world are different, and that will cause differences in the data impact. If the differences are caused by the procedure used in simulating the data, the simulation of the data has to be repeated until a satisfactory agreement is achieved.

Ideally, a complete calibration would be performed every time the DAS changes. However, we would spend our entire resources on calibration if we try for perfection. Of course, we will never reach the perfect calibration. Thus, we need to select test sets of experiments to use for calibration and for verification.

9 Experiences from the NCEP OSSE

9.1 Background of the NCEP OSSE

Various types of OSSEs have been performed (see Section 2); however, to our knowledge, the OSSE performed at the National Centers for Environmental Prediction (NCEP) is the most extensive one so far, and one where calibrations have been performed and presented in a regular manner. The calibration of data impact has been performed by comparing the data impact with both real and simulated data. Without calibration, the simulated data impact cannot be related to the real data impact. The NCEP OSSE is also the first OSSE where radiance data from satellites were simulated and assimilated. A forecast run with a version of the ECMWF model was used to produce the Nature Run, instead of using an analysis or using the same NWP model used for the assimilation (see Sections 1-3).

Since the DWL is one of the most costly instruments, various simulation experiments have been funded and performed. In the NCEP OSSE, instead of evaluating a specific instrument, four representative types of DWL were evaluated (see Section 9.3 below for details). The results show a potentially powerful impact from DWL, but also show that without a careful design of the observing system and a significant effort in developing the data assimilation system, DWL will not be utilized to its best potential.

9.2 Calibration performed for NCEP OSSE

The calibrations were performed on existing instruments, such as the denial of RAOB (radiosonde observations) wind, RAOB temperature, and TIROS Operational Vertical Sounder (TOVS) radiances in various combinations. The geographical distribution of time-averaged Root Mean Square Error (RMSE) shows generally satisfactory agreement between real and simulated impacts. In both the real and simulated analysis, a large analysis impact in the tropics is seen to decrease in the forecast fields. In the Northern Hemisphere mid latitudes, the RMSE distribution of forecasts shows similar spatial patterns in the real and simulated analyses.

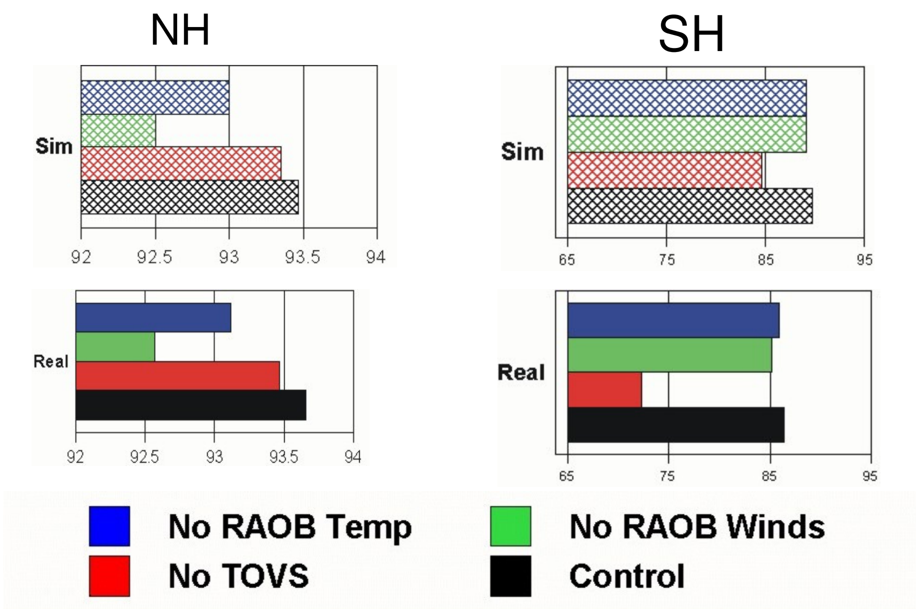

Fig. 1. 500 hPa height anomaly correlation, time averaged between February 13 and 28. 72-hour forecast fields are verified against the control analysis. Control runs include all conventional data and TOVS radiances. For each run the RAOB winds, RAOB temperatures and TOVS radiance are withdrawn in turn (experiments NWIN, NTMP, NTV, respectively). The left two panels are for the Northern Hemisphere and the right two panels for the Southern Hemisphere. The top two panels are for simulation experiments and the bottom two are for real experiments. With permission from Masutani *et al.* (2009).

Figure 1 shows anomaly correlation (AC) skill in the 72-hour 500 hPa geopotential height forecasts verified against the analysis from control experiments. The analysis of the control experiments (CTL) includes conventional observations and TOVS. TOVS (NTV experiment), RAOB wind (NWIN experiment), and RAOB temperature (NTMP experiment) are withdrawn, and the real and simulated data impacts are compared.

In both real and simulated experiments, the RAOB wind has the most impact: overall for the Northern Hemisphere; very slightly more than RAOB temperature for the Southern Hemisphere. Its impact and the magnitude and spatial pattern of the impact are in good agreement for real and simulated experiments. However, the effect of withholding TOVS data in the Southern Hemisphere is much greater in reality than in simulation. Note that a time-varying real SST was used in the assimilation and a constant SST in the simulation. In order to investigate the cause of this inconsistent result, eight experiments were compared: real or simulated analysis, constant or real SSTs, and with or without TOVS data. The consistency in response between the simulated and real atmosphere to the two different SSTs was confirmed. These results suggest that if the SST has a large temporal variability, the impact of TOVS data becomes more important. When TOVS data are used, the analyses with the two different SSTs become closer because TOVS data contain information about SSTs. Although using constant SSTs to generate the Nature Run is not desirable, we conclude that the data impact of slowly varying SSTs in the Southern Hemisphere can be tested with the *T213* NR (see Section 3.2).

These results suggest that a realistically variable SST is required for a more reliable OSSE. Ideally, an ocean-atmosphere coupled model is desirable for producing a better Nature Run, but this may require further development of current coupled models.

9.3 Evaluation of DWL impact using the NCEP OSSE

In the NCEP OSSE, instead of evaluating a specific type of the DWL instrument, four representative types of DWL are evaluated. The data impact from a specific type of DWL is expected to be estimated from the data impact of these four types of DWL. After these idealized experiments, a more realistic DWL will be simulated and evaluated. The four types of DWL are as follows:

- DWL with scanning, while sampling is from all vertical levels;
- DWL without scanning, while sampling is from all vertical levels and in only one direction;
- DWL with scanning, while sampling is from upper levels;
- DWL with scanning, while sampling is from lower levels and clouds.

Upper and lower level sampling represent DWL measurements of molecular and aerosol particle returns, respectively.

First, many experiments were done to illustrate the impact of conventional and DWL data over the first few days of the period under investigation. Then, selected sets of experiments, including model forecasts, were extended to the whole Nature Run period. The impact of DWL was assessed using AC (anomaly correlation) for 500 hPa geopotential heights; then the results of time-averaged geographical distributions and a time series of RMSE were also studied. Traditionally, the AC for 500 hPa geopotential height is used to evaluate the data impact, but it was soon evident in this study that the impact on 500 hPa geopotential height is very limited.

The meridional wind (v) is mainly used to assess the performance of the DWL. Note that the evolution of atmospheric phenomena at the shorter time and smaller spatial scales is dominated by the wind field, while for longer time and larger spatial scales the mass (temperature) field is dominant (Stoffelen *et al.* 2005; Kalnay 1985). In the Northern Hemisphere, excellent skill at the global scale is mostly achieved by existing data (conventional and TOVS). Therefore, the impact of DWL is expected at the synoptic scales. The skill to predict temperature (T) comes mainly from planetary scale events, while the skill to predict v comes mainly from the synoptic scale. The zonal wind (u) and meridional wind (v) contain the information about relative vorticity at the synoptic scale, while u and T contain information about the wave guide (Hoskins and Ambrizzi 1993). Therefore, v depicts information about relative vorticity. The large scale u component can be inferred from temperature, T, observations in the extratropics, while DWL wind observations mainly define the synoptic scale wave which is represented in relative vorticity and the meridional wind, v.

The data impact further depends on the resolution of the DAS. There are many reasons to expect that the data impact might be reduced with higher resolution models (or better forecast models), because they can provide much better background (forecast) fields and there is less room for data to improve the analysis. On the other hand, a higher resolution model will be able to effectively utilize data in finer detail, and that may lead to a higher data impact. Moreover, the smaller scales evolve faster than the larger scales, so their evolution needs to be analysed more often with new observations.

Masutani *et al.* (2006, 2009) showed that improvements in AC (anomaly correlation) scores caused by the insertion of DWL winds are less in a higher-resolution DAS than for a lower-resolution DAS because the first guess field from a higher resolution forecast model is more accurate and leaves less room for improvement. At the larger spatial scales, improvements in the model are more important, but wind data clearly improve the analysis and forecast at the smaller spatial scales. The results very much depend on the spatial scale considered. The NCEP OSSE also showed that the data impact depends on the DAS. Therefore, OSSEs performed using various DAS will be needed to establish confidence in the evaluation of future instruments.

Finally, data impact is also tested using various thinning strategies. For example, data are thinned to 10% in various ways:

- Uniformly;
- 10 minutes on followed by 90 minutes off;
- Targeted to areas with large analysis error;
- Targeted to data void areas;
- Comparison between thinning and increase in observational error.

The NCEP OSSE results show that OSSEs are a very powerful tool for assessing the effect of data distribution.

10 Summary and concluding remarks for OSSEs

Credible OSSEs may be performed that realistically evaluate the impact of prospective observations. The challenges of OSSEs, such as differences in character between the Nature Run and real atmosphere, the process of simulating data and the estimation of observational errors all affect the results. Evaluation metrics moreover affect the conclusions. Thus, consistency in results is important. Some results may be optimistic and some pessimistic. However, it is important to be able to evaluate the sources of errors and uncertainties. As more information is gathered, we can perform more credible OSSEs. If the results are inconsistent, the cause of the inconsistency needs to be investigated carefully. Only when the inconsistencies are explained, interpretation of the results becomes credible.

The NCEP OSSEs (Masutani *et al.* 2006, 2009) have demonstrated that carefully conducted OSSEs are able to provide useful recommendations which influence the design of future observing systems. Based on this work, OSSEs can be used to investigate:

- The effective design of orbit and configuration of an observing system;
- The effective horizontal and vertical data density;
- The evolution of data impact with forecasts;
- The balance between model improvement and improvements in data density and quality;

- The combined impacts of mass (temperature) data and wind data;
- The development of bias correction strategies.

As models improve, there is less improvement in the forecast due to the observations. Sometimes the improvement in forecasts due to model improvements can be larger than the improvement due to observations. However, even in the Northern Hemisphere, forecasts at the subsynoptic scales require much better observations. In the tropics, models need to be improved to retain the analysis improvement for more than a few days of the forecast (Žagar *et al.* 2008). OSSEs will be a powerful tool for providing guidelines for future development in these areas.

(i) Value of OSSEs:

Operational centres are busy getting the best possible value out of existing instruments. We expect that carefully designed OSSEs will enable scientists to make strong and important contributions to the decision making process for future observing systems. Time will be saved in using the new data when compared to the work required to use observing systems that were built without any guidance from OSSEs. However, there is a serious dilemma in spending resources on OSSEs. If a NWP centre devotes resources to getting the greatest benefit out of existing data sources, it misses the opportunity to assess critical future observing systems, with the result that it must live with whatever new observing systems appear in the future rather than influence their development. If it devotes its resources entirely to OSSEs, it may not be paying enough attention to today's valuable data.

(ii) Challenges of OSSEs:

OSSEs are a challenge to weather services. OSSEs require strong leaders with a clear vision, because many of the efforts offer long-term rather than short-term benefits. Although operational systems should benefit from carefully executed OSSEs through lower cost of implementation, there are immediate costs to OSSEs.

OSSEs are very labour intensive. The Nature Run has to be produced using state-of-the-art NWP models at the highest resolution. Simulating data from a Nature Run requires large computational resources, and simulations and assimilations have to be repeated with various configurations. OSSEs also require extensive knowledge of many aspects of the NWP system. Expert knowledge is also required for each instrument. Efficient collaborations are thus essential for producing timely and reliable results.

(iii) Role of stakeholders:

OSSEs will be conducted by various scientists with different interests. Some will want to promote particular instruments. Others may want to aid in the design of the global observing system. Specific interests may introduce bias into OSSEs but they may also introduce strong motivations. Operational centres will perform the role of finding a balance among conflicting interests to seek an actual improvement in weather predictions. They may be regarded as unbiased and thus be best placed for this role; on the other hand, difficulties in finding resources may hamper their effort.

(iv) Recommendations:

Ideally, all new instruments should be tested by OSSEs before they are selected for construction and deployment. OSSEs will also be important in influencing the design of the instruments and the configuration of the global observing system (chapter *The Global Observing System*, Thépaut and Andersson). While the instruments are being built, OSSEs will help prepare the DAS for the new instruments. Developing a DAS to assimilate a new type of data is a significant task. However, this effort has traditionally been made only after the data became available. The OSSE effort demands that this same work be completed earlier; this will speed up the actual use of the new data and proper testing, increasing the exploitation lifetime of an innovative satellite mission.

From the experience of performing OSSEs during recent decades, we realize that using the same Nature Run is essential for conducting OSSEs to deliver reliable results in a timely manner. The simulation of observations requires access to the complete model data and a large amount of resources; thus it is important that the simulated data from many institutes be shared among all the OSSEs. By sharing the Nature Run and simulated data, multiple participants in OSSEs will be able to produce results which can be compared; this will enhance the credibility of the results.

(v) Final word:

NCEP's experience with OSSEs demonstrates that they often produce unexpected results. Theoretical predictions of the data impact and theoretical backup of the OSSE results are very important as they provide guidance on what to expect. On the other hand, unexpected OSSE results will stimulate further theoretical investigations. When all efforts come together, OSSEs will help with timely and reliable recommendations for future observing systems.

References

Arnold, C. P., Jr.and C.H. Dey, 1986. Observing-systems simulation experiments: Past, present, and future. *Bull. Amer. Meteorol. Soc.,* **67**, 687-695.

Atlas, R. 1997. Atmospheric observation and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Jpn,* **75**,111-130.

Atlas, R. and G. D. Emmitt, 1995. Simulation studies of the impact of space-based wind profiles on global climate studies. *Proceedings American Meteorological Society's Sixth Symposium on Global Change Studies*, January, Dallas, TX.

Atlas, R., G.D. Emmitt, J. Terry, E. Brin, J. Ardizzone, J.C. Jusem and D. Bungato, 2003a. Recent observing system simulation experiments at the NASA DAO. Preprints, *Seventh Symposium on Integrated Observing Systems*, 9-13 February 2003, Long Beach, California, American Meteorological Society.

Atlas, R., G.D. Emmitt, J. Terry, E. Brin, J. Ardizzone, J C. Jusem and D. Bungato, 2003b. OSSEs to determine the requirements for space-based lidar winds for weather prediction. *SPIE's Laser Radar Technology and Applications VIII Conference*, April, Orlando, FL.

Atlas, R., E. Kalnay, J. Susskind, W.E. Baker and M. Halem, 1985. Simulation studies of the impact of future observing systems on weather prediction. *Proc. Seventh Conf. on* NWP, 145-151.

Becker, B.D., H. Roquet and A. Stoffelen 1996. A simulated future atmospheric observation database including ATOVS, ASCAT, and DWL. *Bull. Amer. Meteorol. Soc.*, **10**, 2279-2294.

Cress, A. and Wergen, W. 2001. Impact of profile observations on the German Weather Service's NWP system. *Meteor. Zeitschrift,* **10**, 91–101.

Emmitt, G.D. and S.A. Wood, 1989. Simulation of a space-based Doppler lidar wind sounder – sampling errors in the vicinity of wind and aerosol inhomogeneities. *Fifth Conference on Coherent Laser Radar*, June, Munich, Federal Republic of Germany.

Emmitt, G.D. and S.A. Wood, 1991a. Simulating thin cirrus clouds in observing system simulations experiments (OSSE) for LAWS. *Proceedings American Meteorological*

Society's Seventh Symposium on Meteorological Observations and Instrumentation, Special Session on Laser Atmospheric Studies, January 14-18, New Orleans, LA, pp 460-462.

Emmitt, G. D. and S.A. Wood, 1991b. Simulated wind measurements with a low power/high PRF space-based Doppler lidar. *Optical Remote Sensing of the Atmosphere, Fifth Topical Meeting*, November 18-21, Williamsburg, VA.

Emmitt, G.D. and S.A. Wood, 2001. Simulating space-based lidar performance using global and regional scale atmospheric numerical models. *Optical Remote Sensing Topical Meeting*, February, Coeur d'Alene, ID.

Errico, R.M., R. Yang, M. Masutani and J. Woollen, 2007. Estimation of some characteristics of analysis error inferred from an observation system simulation experiment. *Meteor. Zeitschrift*, **16**, 695-708.

Fisher, M., 2003. Background error covariance modelling. Proceedings of ECMWF Seminar, *Recent developments in data assimilation for atmosphere and ocean*, 8-12 September 2003, Reading, UK, pp 45-64.

Gelaro R. and Y. Zhu, 2009. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179-193.

Halem, M. and R. Dlouhy, 1984. Observing system simulation experiments related to space-borne lidar wind profiling. Part 1: Forecast impact of highly idealized observing systems. Preprints, *Conference on Satellite Meteorology/Remote Sensing and Applications*, June 25-29, 1984, Clearwater, Florida, American Meteorological Society, pp 272-279.

Hamill, T.M., R.P. d'Entrement and J.T. Bunting, 1992. A description of the Air Force real-time, nephanalysis model. *Weather and Forecasting*, **7**, 288-306.

Han, Y, P. van Delst, Q. Liu, F. Weng, B.Yan, R. Treadon and J. Derber, 2006. JCSDA Community Radiative Transfer Model (CRTM) - Version 1, *NOAA Tech Report 122*.

Henderson-Sellers, A. 1986. Layer cloud amount for January and July 1979 from 3D-Nephanalysis. *J. Climate Appl. Meteor.*, **24**, 118-132.

Hoskins, B.J. and T. Ambrizzi, 1993. Rossby Wave Propagation on a Realistic Longitudinally Varying Flow. *J. Atmos. Sci.,* **50,** 1661–1671.

Kalnay, E, J.C. Jusem and J. Pfaendtner, 1985. The relative importance of mass and wind data in the FGGE observing system. *Proceedings of the NASA Symposium on Global Wind Measurements, Columbia, MD, NASA,* 1–5.

Keil, M., 2004. Assimilating data from a simulated global constellation of stratospheric balloons. *Q. J. R. Meteorol. Soc*., **130**, 2475-2493**.**

Kistler, R., NCEP Staff, Contractors, and Visiting Scientists, Past and Present, 2008. Reanalysis at NCEP: Past, Present and Future. *Third WCRP international conference on reanalysis*. February, 2008, Tokyo, Japan.

Kleespies, T.J. and D. Crosby 2001. Correlated noise modelling for satellite radiance simulation. *AMS preprint volume for the11th Conference on Satellite Meteorology and Oceanography*, October 2001, Madison Wisconsin, pp 604-605.

Kleespies, T.J., P. van Delst, L.M. McMillin and J. Derber, 2004. Atmospheric Transmittance of an Absorbing Gas. 6. An OPTRAN Status Report and Introduction to the NESDIS/NCEP Community Radiative Transfer Model. *Applied Optics*, **43**, 3103-3109.

Lahoz, W.A., R. Brugge, D.R. Jackson, S. Migliorini, R. Swinbank, D. Lary and A. Lee 2005. An observing system simulation experiment to evaluate the scientific merit of wind and ozone measurements from the future SWIFT instrument. *Q. J. R. Meteorol. Soc*., **131**, 503-523.

Langland, R.H. and N.L. Baker, 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189-203.

Liu, Q. and F. Weng, 2006a. Radiance Assimilation in Studying Hurricane Katrina. *Geophys. Res. Lett.*, **33**, L22811, doi:10.1029/2006GL027543.

Liu, Q. and F. Weng, 2006b. Detecting Warm Core of Hurricane from the Special Sensor Microwave Imager Sounder. *Geophys. Res. Lett.,* **33**, L06817.

Lord, S.J., E. Kalnay, R. Daley, G.D. Emmitt and R. Atlas 1997. Using OSSEs in the design of the future generation of integrated observing systems. Preprints, *First Symposium on Integrated Observing Systems*, Long Beach, California, 2-7 February 1997, American Meteorological Society.

Lord, S.J., M. Masutani, J.S. Woollen, J.C. Derber, G.D. Emmitt, S.A. Wood, S. Greco, R. Atlas, J. Terry and T.J. Kleespies, 2002. Impact assessment of a Doppler wind lidar for NPOESS/OSSE. *American Meteorological Society's Sixth Symposium on Integrated Observing Systems*, January, Orlando, FL.

Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

McNally, A.P., J.C. Derber, W.-S. Wu and B.B. Katz, 2000. The use of TOVS level-1 radiances in the NCEP SSI analysis system. *Q. J. R. Meteorol. Soc*., **129**, 689-724.

Marseille, G.J. and A. Stoffelen, 2003. Simulation of wind profiles from a space-borne Doppler wind lidar. *Q. J. R. Meteorol. Soc.*, **129**, 3079-3098.

Marseille, G.J., A. Stoffelen and J. Barkmeijer, 2008a. Sensitivity Observing System Experiment (SOSE) - A new effective NWP-based tool in designing the global observing system. *Tellus A*, **60**, 216-233. doi: 10.1111/j.1600-0870.2007.00288.x.

Marseille, G.J., A. Stoffelen and J. Barkmeijer, 2008b. Impact assessment of prospective space-borne Doppler wind lidar observation scenarios. *Tellus A*, **60**, 234-248. doi: 10.1111/j.1600-0870.2007.00289.x.

Marseille, G.J., A. Stoffelen and J. Barkmeijer, 2008c. A cycled sensitivity observing system experiment on simulated Doppler wind lidar data during the 1999 Christmas storm "Martin". *Tellus A*, **60**, 249-260. doi: 10.1111/j.1600-0870.2007.00290.x.

Masutani, M., K. Campana, S. Lord and S.-K. Yang, 1999. Note on cloud cover of the ECMWF nature run used for OSSE/NPOESS project. *NCEP Office Note* No. 427.

Masutani, M., J. S. Woollen, S. J. Lord, G. D. Emmitt, T. J. Kleespies, S. A. Wood, S. Greco, H. Sun, J. Terry, V. Kapoor, R. Treadon, and K. A. Campana, 2009. Observing System Simulation Experiments at the National Centers for Environmental Prediction. *J. Geophys. Res.,* **114**, doi:10.1029/2009JD012528.

Masutani, M., J.S. Woollen, S.J. Lord, G.D. Emmitt, S. Wood, S. Greco, T.J. Kleespies, H. Sun, J. Terry, J. C. Derber, R.E. Kistler, R.M. Atlas, M.D. Goldberg and W. Wolf, 2003. Observing system simulation experiments for NPOESS – assessment of Doppler wind lidar and AIRS. *American Meteorological Society's The Simpson Symposium*, February, Long Beach, CA.

Masutani, M., J.S. Woollen, S.J. Lord, T.J. Kleespies, G.D. Emmitt, H. Sun, S. A. Wood, S. Greco, J. Terry, and K. Campana, 2006. Observing System Simulation Experiments at NCEP. *NCEP Office Note* No. 451.

Nitta, T., 1975. Some analyses of observing systems simulation experiments in relation to First GARP Global Experiment. *GARP Working Group on Numerical Experimentation, Report No. 10*, US GARP Plan, pp 1-35. [Available from the National Academy of Sciences, 2101 Constitution Ave. N.W., Washington, D.C. 20418.]

Petersen, D.P., 1968. On the concept and implementation of sequential analysis for linear random fields. *Tellus*, **20**, 673-686.

Reale, O., J. Terry, M. Masutani, E. Andersson, L.P. Riishøjgaard and J.C. Jusem, 2007. Preliminary evaluation of the European Centre for Medium-Range Weather Forecast (ECMWF) nature run over the tropical Atlantic and African monsoon region. *Geophys. Res. Lett*., **34**, L22810, doi:10.1029/2007GL031640.

Riishøjgaard, L.P., R. Atlas and G.D. Emmitt, 2003. Analysis of simulated observations from a Doppler wind lidar. *American Meteorological Society's 12th Conference on Satellite Meteorology*, February, Long Beach, CA.

Rohaly, G.D. and T.N. Krishnamurti, 1993. An observing system simulation experiment for the laser atmospheric wind sounder (LAWS). *J. Applied. Meteor.,* **32**, 1453-1471.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H.M. Van den Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng and P. Xie, 2006. The NCEP Climate Forecast System. *J. Climate*, **16**, 3483–3517.

Saunders R.W., M. Matricardi and P. Brunel, 1999. An Improved Fast Radiative Transfer Model for Assimilation of Satellite Radiance Observations. *Q. J. R. Meteorol. Soc.,* **125**, 1407-1425.

Stoffelen, A., G.J. Marseille, F. Bouttier, D. Vasiljevic, S. De Haan and C. Cardinali, 2006. ADM-Aeolus Doppler wind lidar Observing System Simulation Experiment. *Q .J. R. Meteorol. Soc*., **619**, 1927-1948.

Stoffelen, A., J. Pailleux, E. Källén, J.M. Vaughan, L. Isaksen, P. Flamant, W. Wergen, E. Andersson, H. Schyberg, A. Culoma, R. Meynart, M. Endemann and P. Ingmann, 2005. The Atmospheric Dynamic Mission for Global Wind Fields Measurement. *Bull. Amer. Meteorol. Soc.*, **86**, 73-87.

Strow, L.L., H.E. Motteler, R.G. Benson, S.E. Hannon and S. De Souza-Machado, 1998. Fast computation of monochromatic infrared atmospheric transmittances using compressed lookup tables. *J. Quant. Spectrosc. Radiat. Transfer,* **59**, 481-493.

Tan, D.G.H., E. Andersson, M. Fisher and L. Isaksen 2007. Observing system impact assessment using a data assimilation ensemble technique: Application to the ADM-Aeolus wind profiling mission. *Q. J. R. Meteorol. Soc*, **133**, 381-390.

Tsutsui, J., and A. Kasahara 1996. Simulated tropical cyclones using the National Center for Atmospheric Research community climate model. *J. Geophys. Res.*, **101**, 15,013-15,032.

Weng, F., 2007. Advances in radiative transfer modelling in support of satellite data assimilation. *J. Atmos. Sci*., **64**, 3803-3811.

Wood, S.A., G.D. Emmitt and S. Greco, 2000. DLSM. A coherent and direct detection lidar simulation model for simulating space-based and aircraft-based lidar winds. *AeroSense 2000*, April, Orlando, FL.

Woollen, J.S., M. Masutani, H. Sun, Y. Song, G.D. Emmitt, Z. Toth, S.J. Lord and Y. Xie 2008. Observing Systems Simulation Experiments at NCEP OSSEs for realistic adaptive targeted DWL Uniform observation and AIRS. *AMS preprint, Symposium on Recent Developments in Atmospheric Applications of Radar and Lidar*, New Orleans, LA, pp 20-24, January 2008.

Žagar, N., A. Stoffelen, G.J. Marseille, C. Accadia and P. Schlüssel, 2008. Impact Assessment of Simulated Doppler Wind Lidars with a Multivariate Variational Assimilation in the Tropics. *Mon. Weather Rev.,* **136**, 2443-2460, doi:10.1175/2007MWR2335.1.

Zhu, Y. and R. Gelaro, 2008. Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Weather Rev.*, **136**, 335-351.