P 1.2

INTERNATIONAL COLLABORATIVE JOINT OSSES

- Toward reliable and timely assessment of future observing systems -

M. Masutani[1,10,#,*], L.-P. Riishojgaard[10,2,$], E. Andersson[4],
J. S. Woollen[1+], A. Stoffelen[7], G.-J. Marseille[7], Y. Xie[5], N. Prive[5,@],
R. Errico[2$], R. Yang[2,&], T. Zhu[3,@], F. Weng[3],
G. D. Emmitt[6], S. Greco[6], S. A. Wood[6],
O. Reale[2,$], T. W. Schlatter[5], S. J. Lord[1],
H. Sun[3,##], T. J. Kleespies[3,10], Y. Song[1%]

[1]NOAA/NWS/NCEP/EMC, Camp Springs, MD
[2]NASA/GSFC, Greenbelt, MD
[3]NOAA/NESDIS/STAR, Camp Springs, MD
[4]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
[5]NOAA/Earth System Research Laboratory, Boulder, CO
[6]Simpson Weather Associates (SWA), Charlottesville, VA
[7]Royal Dutch Meteorological Institute (KNMI), DeBilt, Netherlands
[8]Joint Center for Satellite Data Assimilation, MD
[#]Wyle Information Systems Inc., El Segundo CA
[+]Science Applications International Corporation (SAIC)
[$]Goddard Earth Science and Technology Center, University of Maryland, Baltimore, MD
[%]I. M. Systems Group, Inc. (IMSG), MD
[&]Science Systems and Applications Inc (SSAI). MD
[@]Cooperative Institute for Research in the Atmosphere (CIRA)/CSU, CO
[##]QSS Group, Inc., MD

1. Definition and motivation of OSSEs

Observing System Simulation Experiments (OSSEs) are typically designed 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.

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. 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 but 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 is in regard to 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 be 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 modeling 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 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 in the design before putting an instrument into production. Thus, well-conducted OSSEs can be invaluable in 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 the use of future instruments and observing systems. Additionally, OSSEs can hasten database development and the development of data processing techniques (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 the maximum societal and economic impact (Arnold and Dey 1986; Lord *et al.* 1997; Atlas 1997).

There is another motivation for OSSEs that has been discussed less often. 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 the analysis errors. Evidence to guide such a 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 data set 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), Thus, robust independent data sets for verification are usually absent (although, *e.g.*, research data such as ozonesondes and ozone from some instruments are not commonly assimilated, and are thus 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 also necessarily 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 made on major Earth Observing Systems, any OSSE results upon which these decisions are based will have to withstand intense scrutiny and criticism. One goal of this manuscript is to suggest ways in which OSSEs can be made robust and credible.

OSSEs are very labor intensive projects. It has been realized that the preparation of a NR including the evaluation, simulation of observations, and distribution of data consumes a significant amount of effort. An internationally collaborative effort for full OSSEs, called Joint OSSEs, has been formed over the past three years. In Joint OSSEs a common NR will by used by the various DAS at many institutes. The first Joint OSSE NRs have been produced by the European Center for Medium-Range Weather Forecasts (ECMWF).

2. Various type of OSSEs and full OSSEs

In the Joint OSSE, the term OSSE (sometimes *full OSSE* to distinguish it from other simulation experiments) refers to a simulation experiment with a Nature Run model that is significantly different from the Numerical Weather Prediction (NWP) model used for data assimilation. This provides a truth independent of the data assimilation system NWP model and the Global Observing System (GOS) data coverage and quality. Simulation of all observations is considered to be a significant initial investment for an OSSE, but interpolating the 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 using several different data assimilation systems with minor modifications to the operational systems. The data impact for OSSEs (and their variants) often varies with the verification metric and DAS used. Note, however, that a truth is available for further verification of the DAS characteristics.

Various simulation experiments have been attempted which use real data for existing instruments and only simulate the 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 the Northern Hemisphere land surface and how these might be replaced by another observing system (Cress and Wergen 2001). 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. An alternative method, the Analysis Ensemble System (AES) (Tan *et al.* 2007) uses the spread in an ensemble as a proxy for the analysis and background uncertainty based on arguments of error growth (Fisher 2003). 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.

Although a SOSE, OSRE or AES allows the quick study of real extreme events, the SOSE requires an adjoint model to generate new observations and the AES requires an established ensemble system. Calibration and interpretation of the results are complicated and need 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. Therefore, although an initial investment is required for a full OSSE, today the most reliable strategy is to use full OSSEs for impact assessment of prospective observing systems. There are many OSSEs conducted without calibrations. During the early years of OSSEs, identical twin OSSEs or fraternal twin OSSEs (see Section 3) were often conducted due to the lack of variety in high-fidelity NWP models.

3. The Nature Run

3.1 Requirement for the Nature Run

The Nature Run is a long, uninterrupted forecast by a model whose statistical behavior matches that of the real atmosphere. The ideal Nature Run would be a coupled atmosphere-ocean-cryosphere model with a fully interactive lower boundary. Meteorological science is approaching this ideal but has not yet reached it. For example, it is still customary to supply the lower boundary conditions (SST and ice cover) appropriate for the span of time being simulated. Such coupled systems are not quite mature enough yet 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).

In Joint OSSEs a succession of analyses are not being used for the Nature Run. In the case of four-dimensional variational assimilation (4D-VAR), although the analyses may each be a realizable model state, they all lie on different model trajectories. Each analysis marks a discontinuity in the model trajectory, determined by the information content extracted by a DAS from the existing global observing systems and forced by observations. Furthermore, residual systematic effects due to the spatially non-uniform and often biased observations, the DAS, or the model state may either favorably or unfavorably affect the potential of new observing systems to improve the forecast. Thus, considering a succession of analyses as truth seriously compromises the attempt to conduct a “clean” experiment.

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

3.2 Joint OSSE Nature Run

The Nature Runs and simulated data ought to be shared between many institutes carrying out the actual OSSEs. OSSEs with different Nature Runs are difficult to compare but OSSEs using a different DAS but the same Nature Run can provide a valuable crosscheck of data impact results.
The primary specifications for a new Nature Run are:

- To cover a long enough period to span all seasons and to allow the selection of interesting sub-periods for closer study;
- To provide data at a temporal resolution higher than the OSSE analysis cycle;
- To simulate the atmosphere at scales compatible with the main OS;
- To use daily SSTs;
- To have user-friendly archiving.

Based on the recommendations from NOAA and NASA, ECMWF produced a new Nature Run in July 2006 at T511 (40 km) spectral truncation and 91 vertical levels, with the output saved every 3 hours. Two high resolution Nature Runs at T799 (25km) horizontal resolution and 91 vertical levels have been generated to study data impacts when forecasting hurricanes and midlatitude storms. The output is saved every hour. A hurricane period from September 27 to November 1 was selected and a period from April 10 to May 15 was selected to study midlatitude storms. The version of the model used was the same as for the interim reanalysis at ECMWF (cy30r1). The initial condition is the operational analysis on 12Z May 1st, 2005 and the Nature Run ends at 00Z June 1st, 2006. The model was forced by daily SST and ice provided by NCEP (also used in the operational forecasts) which is used throughout the experiments.

The complete data for the T511NR and T799NR are saved at ECMWF, NCEP, NASA/GSFC, and ESRL. The complete Nature Runs are accessible from the NASA/GSFC/NCCS portal system. Access to the data from this site requires an account, which is available to the research community. The complete Nature

Runs will also be available from ECMWF. Verification data (1degx1deg data) for the T511NR are also available from JMA and the NCAR/ CISL Research Data Archive as data set ID ds621.0. Complete verification data for the T511NR and T799 NR are also available from NRL/Monterey, University of Utah, and Mississippi State University. (Masutani et al 2008).

3.3 Evaluation of the Joint OSSE Nature Run

Tropical rainfall showed some dumping during the first few weeks of the T511NR, and it takes about 10 days for the convective rainfall to settle down and 20 days for the large scale rainfall. The amount of area average rainfall does not show any apparent drift in the midlatitudes.

The large scale structure of the T511 NR is very realistic. At some times smaller scale structures in the NR are more realistic than in the reanalysis, which is processed by a much lower resolution model. Some results presented in this section are published in Masutani et al. (2007), Reale et al. (2007), and Masutani et al. (2009).

Midlatitude cyclone statistics were produced using Goddard's objective cyclone tracker. The cyclone tracker produces a:

- Distribution of cyclone strength across the pressure spectrum;
- Cyclone lifespan;
- Cyclone deepening;
- Regions of cyclogenesis and cyclolysis;
- Distribution of cyclone speed and direction.

The location and height of midlatitude jets in the nature run were within the interannual variability of the ECMWF analysis. (Fig.1)

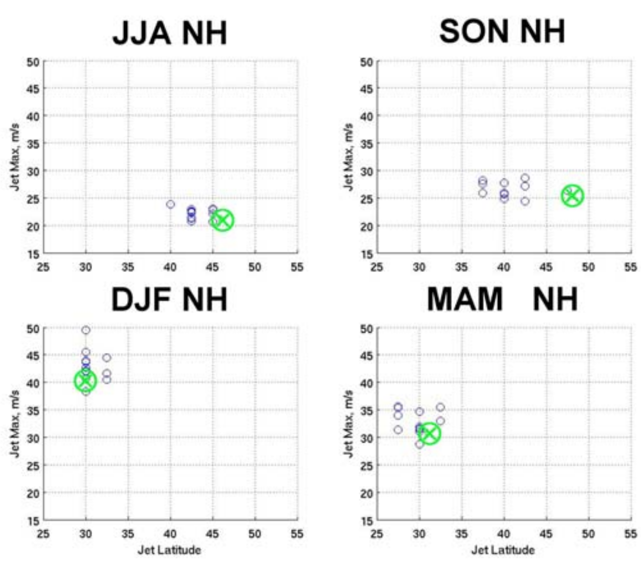

Fig.1 Seasonal mean, zonal mean, zonal wind, jet maximum strength and latitude of the jet maxima for the ECMWF reanalysis (1989-2001, blue circles) and the Nature Run (green circle with a cross), northern hemisphere. (by N. Prive.)

Some preliminary analyses performed over the first four months of the ECMWF NR for the African Monsoon and tropical Atlantic regions are presented. The analyzed data are the 1x1 degree resolution pressure-level fields.

The overall representation of the African Easterly Jet (AEJ) is realistic and a number of important well-known observational features are observed, such as the axis of the AEJ core slightly tilted northward and westward, a clear separation from the low-level Harmatthan flow, and a stronger low-level monsoonal flow on the western side.

Disturbances resembling African Easterly Waves are being produced in the NR. The propagation speed (about 5-9 degrees/day) and the amplitude appear realistic, as the evident modifications occurring at or about the transition (approximately 15W) when some waves intensify and most accelerate.

Once over the Atlantic Ocean, signs of the development and organization of some waves into smaller-scale circulations are observed. In particular, the ECMWF NR seems to also show the capability of spontaneously (without any form of vortex bogus using, relocation or ad-hoc data assimilation) producing realistic Atlantic hurricanes. In spite of the 1x1 degree interpolation, the vortex looks very realistic, with a prominent warm core and vertically aligned isotachs indicating an eye-like feature.

Monthly, seasonal, and annual mean cloud cover and radiation budgets evaluated at ECMWF using routine diagnostic software and time mean fields show remarkable agreement with the observed (Fig.2 and Fig.3). Further evaluation of variance is yet to be performed.

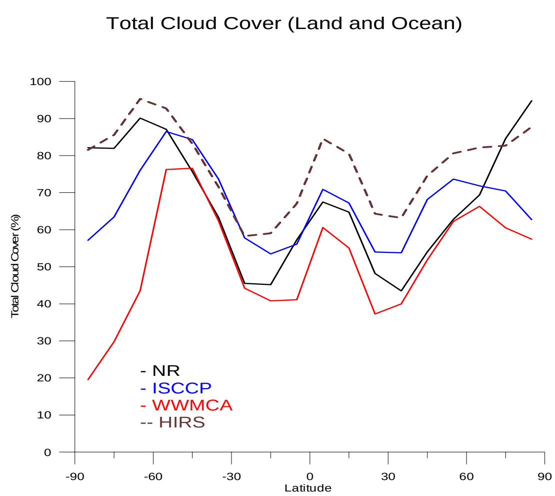

Fig.2 Zonal mean total cloud cover for August 2005. IT511 NR is compared with ISCCP monthly cloud climatologies, MODIS based cloud climatology, UW/HIRS based climatology, and WWMCA (Nephanalyses). (by S. Greco.)

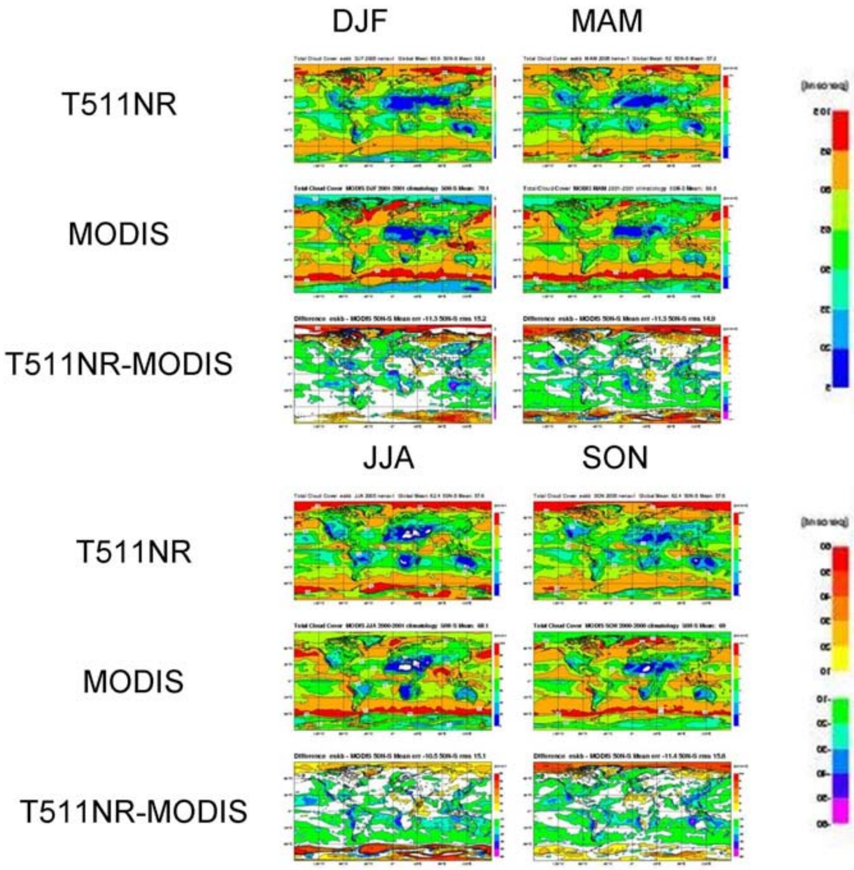

Fig. 3 Seasonal mean total cloud cover. T511NR, observed estimate from MODIS data and the difference.. (by A. Tompkins)

3.4 Possibilities for future Nature Runs

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. The Nature Runs should be widely accessible, and 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 our past experience of OSSEs are:

a. Employ a NWP model with demonstrated forecast skill;
b. Simulation span should include all seasons to allow selection of interesting sub-periods for closer study;
c. Simulation sample should have 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. Simulations should resolve scales compatible with the main observing systems;
e. Base the NR 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. Simulations 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;
j. Since the data impact depends on the season, it is important that future Nature Runs cover long periods, preferably a whole year.

The set of archived Nature Run variables should be expanded to accommodate the need for OSSEs. For example, geopotential height at model levels is very desirable. The archiving of this variable will help with 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.

4. Simulation of observations

4.1 Basic guidelines

Although a particular OSSE may be motivated by the 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 initially is, 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 but with other observations excluded results in a very large impact compared with no assimilation at all, but has a much smaller and more realistic impact if other observations *are* considered.

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, therefore, it is prudent to also understand the nature of all the types of error realistically associated with them.

Simulation of observations requires experts for each instrument. Since this process requires access to the full resolution Nature Run, computing facilities with large memory capacity are required. If the observational errors, added to the true values extracted from the nature run, are properly specified, then the statistical behavior of the assimilation system will be similar in the simulated and real worlds and the OSSE will be properly calibrated. The calibration process is time consuming and calibration was not often performed in most of OSSEs, except for the OSSE at NCEP (Masutani et al. 2006).

4.2 Observation simulator

An initial preliminary simulation of conventional data was conducted by NCEP and NESDIS. The data were made available to Joint OSSE for calibration purposes. In order to simulate radiance data, vertical profiles were generated based on actual operational usage to keep some statistics similar to a real assimilation. Initial simulation of GOES, AMSUA and AMSUB radiance data are also completed for the whole period of the T511 NR.

An extensive effort for simulation of observations was conducted at NASA/GSFC/Global Modeling and Assimilation Office (GMAO). GMAO simulator has been set up to simulate HIRS2, HIRS3, AIRS, AMSUA, AMSUB, MSU radiance data as well as conventional data. Calibration experiments were also conducted at GMAO using adjoint technique.

GMAO simulation software includes:

- Software for generating conventional obs (Observation type included in NCEP .prepbufr file). The codes are set up for raobs, aircraft, ships, vad winds, wind profilers, surface station data, SSMI and Quick scat surface winds, Cloud Motion Vector (CMV).
- Software for simulating radiances Code to simulate HIRS2/3, AMSUA/B, AIRS, MSU has been set up. Community Radiative Transfer Model (CRTM) is used for forward model.
- Software for generating random obs. error. Observations are generated without errors but software to simulate errors is provided.

The output of the data are saved BUFR format which can be read by Gridpoint Statistical Interpolation (GSI). GSI is a DAS used at NCEP, GMAO and ESRL. The codes are flexible and include many tunable parameters. The codes will be available to Joint OSSE and software is well documented. Since the software will continued to be developed, all interested people are expected to contact GMAO (Ronald Errico: ronald.m.errico@nasa.gov) or Joint OSSE (Michiko Masutani: michiko.masutani@noaa.gov) for the current version. The GMAO simulation software was successfully installed at NCEP and initial simulations of AIRS, HIRS2 and HIRS3 radiance data were completed for the entire period of the T511 NR. It is also versatile enough to simulate other observing systems.

Calibration using an adjoint technique was conducted at GMAO and remarkable similarity between simulated data impact and real impact was achieved. Further detailed adjustments are being conducted. ESRL is working on calibration experiments including GOES. Some initial results are reported by Privé et al. (2009). Significant inconsistent results are observed in the data impact of CMV and SSMI wind. This is possibly due to the preliminary sampling strategies. In the initial simulation, CMV was simulated using actual observation locations. SWA has developed strategies for a realistic sampling of CMV from the Nature Run and a coordinated effort will be conducted to simulate a more realistic CMV.

Alternative software to simulate radiance data using the Stand-alone AIRS Radiative Transfer Algorithm (SARTA) as well as CRTM is also being developed at NESDIS. This will be important in evaluating CRTM in Joint OSSEs.

The preliminary data used for ongoing calibration require further tuning and evaluations and should be used with caution. These data are useful for building and testing scripts and are available to participating scientists who are expected to share their results.

4.3 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, the 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 an operational assimilation. Although the instrument errors are in most cases fairly well defined, the derivation of total error levels that are 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.

In order to create perfect observations, it is only necessary to locate the observation type to be simulated within 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 to just 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, and SATOB) 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 was 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. We discuss several examples below.

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 (SYNOPS and METAR for example) present several issues to be considered in achieving realistic simulations. The question of location mainly involves 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 a particular instrument's measuring heights to locate these data. There is also a need to interpolate surface values from the Nature Run background fields 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 costs and avoid turbulence. Ideally, flight tracks for the OSSE should be formulated for the simulated aircraft in the same way as they are for real cases. However, the location of jets and turbulence can be very different between 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 in 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 in a 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.

In the NCEP OSSE (see Section 9), the use of different radiative transfer models (RTMs) for simulation and assimilation helps in understanding 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 from 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). Modeling subgrid-scale cloud remains important in simulating cloudy radiances and for the assimilation of radiance data. Testing RTMs with clouds is an important area for OSSEs. 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 modeled through cloud fractions, subgrid-scale clouds still need to be represented appropriately (*e.g.,* in a statistical sense).

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 modeled to be used by any data assimilation system. Bias correction is now a part of data assimilation systems. As a result, one can bias correct the Nature Run radiances or implement the bias correction in the DAS itself.

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 a combination of observing systems not currently being used together. No other instrument has been subjected to more OSSE evaluations 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 the 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 specifically designed 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 sub-grid 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 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 2010 (Stoffelen et al. 2005). 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 from 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.

5. Assignment of realistic observation errors

Assignment of realistic observational errors is the most challenging task in OSSEs. There are various sources of observational error. Initially, Gaussian random errors with the amplitude from an operational error table were used. However, many errors are not random and specially (sp?) correlated. Estimates of instrument error have to be provided by instrument communities. The Magnitude of the errors has to be tuned throughout the calibrations.

The most challenging task is to simulate representativeness errors. The model-resolved volume (Skamarock, 2004) does not match the atmospheric volume that is the object of measurement. If the observed volume is small compared to the model-resolved atmospheric volume, the measurement will represent scales of motion that the model cannot resolve. From the model's point of view, the observation contains sub-resolution scale noise, and this will contribute to the value of representativeness error *(ε_r)*. In other words, because the representation of model values at an observation location ($x_{t)}$) is spectrally truncated the projection does not capture the small-scale atmospheric variance inherent in the observation. If the observed volume is larger than the model resolution volume (*e.g.*, a measurement of phase in the microwave portion of the electromagnetic spectrum by GPS could involve a dimension in the volume of atmosphere larger than the equivalent model resolution dimension) then the forward model will be an averaging operator in this direction rather than an interpolation operator. From the model's point of view ε_r will relate to how well the observation spatially and temporally (*i.e.*, in 4D) represents the equivalent model-average value (*y)*.

Some aspects of the representativeness error are not random but systematic. Even if we exclude sub-grid 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. Modeling 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.

More details of the observational errors are discussed in the OSSE chapter in "Data assimilation: Making sense of Observations" which is edited by Menard and Lahoz, to be published from Springer.

6. Evaluation of OSSE results

The data impacts on 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 on the mass fields could be very different from the impact on the wind fields. Below we discuss various aspects of data impact.

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 is 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.

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). This is a powerful method that describes the contributions from different observations. This technique allows the detection of impact, be it positive or negative, from any observation. There are advantages and disadvantages as compared with Data Denial Experiments (DDEs) (Gelaro and Zhu 2008; Zhu and Gelaro 2008).

7. Calibration of OSSEs

The 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 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 y-$H(x)$ of the OSSE with those of the real world assimilation y-$H(x)$ 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 (Stoffelen et al., 2006).

One way of calibrating an OSSE is to use a DDE 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 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 data simulation 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 all our 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. 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 also 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 does interpretation of the results become credible.

The NCEP OSSEs (Masutani et al. 2006) 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 sub-synoptic 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.

9.1 Value of OSSEs:

Operational centers 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 an NWP center 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 having influence over their development. If it devotes its resources entirely to OSSEs, it may not be paying enough attention to today’s valuable data.

9.2 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 labor 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.

9.3 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 centers 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.

9.4 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. 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.

9.5 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. Meteor. Soc.,* **67**, 687-695.

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.

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

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.

Becker, B.D., H. Roquet and A. Stoffelen 1996. A simulated future atmospheric observation database including ATOVS, ASCAT, and DWL. *Bull. Amer. Meteor. 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, 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, 45-64.

Gelaro R. and Y. Zhu 2008. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus* (accepted).

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, 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.

Keil, M., 2004. Assimilating data from a simulated global constellaion 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. 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, 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, 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.

Skamarock, W.C., Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, *M.W.R.,* **132**, 3019-3032.

Stoffelen, A., J. Pailleux, E. Källén, J.M. Vaugham, I. 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. Meteor. Soc.*, **86**, 73-87.

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.

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**, 15013-15032.

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, 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.