Full OSSEs at NCEP and International Joint OSSE effort

Michiko Masutani[1], John S. Woollen[2], Erik Andersson[3], Lars-Peter Riishøjgaard[4], Stephen J. Lord[5]

[1]NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA, and Wyle Information Systems, McLean, VA, USA, Michiko.Masutani@noaa.gov
[2]NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA and Science Applications International Corporation (SAIC), USA, Jack.Woollen@noaa.gov
[3]European Centre for Medium-Range Weather Forecasts (ECMWF) Reading, UK, erik.andersson@ecmwf.int
[4]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
[5]NOAA/NWS/NCEP/EMC Camp Springs, MD, USA, Stephen.Lord@noaa.gov

1. Introduction

Building and maintaining observing systems with new instruments is extremely costly, particularly when satellites are involved. Any objective method that can evaluate the expected improvement in forecast skill due to the selection of future instruments and help determine their configurations will be quite valuable. The forecast skill evaluations performed using simulation experiments are known as Observing System Simulation Experiments (OSSEs). Since all the details of operational Data Assimilation Systems (DAS) have to be reconstructed for a simulated world, OSSEs are quite labor intensive. However, the OSSE cost is only small fraction of the cost of an actual observing system and, therefore, is a relevant aid in its design. By running OSSEs, current operational DAS are upgraded and pre-prepared to handle future data types and larger data volumes. OSSEs can thus speed up the development of databases, data processing (including formatting) and quality control software.

The general guideline for an OSSE was reviewed in Arnold and Dey (1986) and Masutani et al. (2010). In this paper, this type of OSSE is called a "Full OSSE" to distinguish it from other types of OSSEs. In a Full OSSE, all existing observations are simulated and a calibration is performed.

In Section 2 Full OSSEs are defined as well as calibration in Full OSSEs, and requirements for the Nature Run and calibration are discussed. In Section 3, Full OSSEs conducted at NCEP are described, based on the experience in international collaborative OSSEs that has emerged in the past few years. In Section 4, the Nature Run for Joint OSSE and Joint OSSE strategies are discussed. Finally, in Section 5 further guidelines for OSSEs are described.

2. Full OSSEs

2.1. Full OSSEs and other types of simulation experiments

The full OSSE refers to a simulation experiment with a Nature Run model that is significantly different from the NWP model used for data assimilation. The Nature Run provides a simulated substitute for the true atmosphere. 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. In a full OSSE, all significant observations used for the DAS have to be simulated from the Nature Run. They are used as the control run for OSSEs and also used for the calibration experiments explained in Section 2.3.

The simulation of all observations is a necessary significant initial investment for an OSSE,. State-of-the-art observation operators (or forward models) are used. In Full OSSEs, all the usual analysis and forecast verification metrics can be used to evaluate data impact. The simulated data can be tested with several different DAS with minor modification to the operational systems. The data impact for OSSEs often varies with the verification metric and DAS used. The reference (truth) for the verification is provided by the Nature Run,

2.2. The Nature Run for a full OSSE

The Nature Run is a long, uninterrupted forecast by a NWP 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. However, it is still customary to prescribe the lower boundary conditions (sea surface temperature, SST, and ice cover) using analysed values for the entire simulation period. NWP 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 in high resolution Nature Run is unknown.

The advantage in 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 into the simulation. This does not matter in the context of OSSEs, provided that the climatological statistics of the Nature Run match those of the real atmosphere. A Nature Run could be seen as a separate universe, ultimately independent from but with very similar characteristics to the real atmosphere.

2.3. Calibration

Calibration of OSSEs verifies the simulated data impact for existing observing systems by comparing it to the real data impact. The calibration experiments provide invaluable guidance for the interpretation of OSSE results. Without calibration, the quantitative evaluation of data impact using OSSE could mislead the meteorological community. Calibration evaluates the quality of the Nature Run for OSSEs. If a Nature Run is not suitable for OSSE it will exhibit an unrealistic data impact which cannot be adjusted by simulated observational errors.

3. Full OSSE at NCEP with T213 Nature run

3.1. Background of the Full OSSE at NCEP

Full OSSEs were conducted by NCEP using a Nature Run provided by ECMWF (Becker et al. 1996). 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 that was used for the assimilation (see Sections 1-3).

Since the Doppler wind lidar (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.

Note that the NCEP OSSE was conducted using the old Nature Run and old DAS at NCEP. Therefore, all results must be confirmed using the new Nature Run and DAS. The detail results from NCEP OSSEs are presented in Masutani et al. (2006) and Woollen et al. (2008).

3.2. Calibration performed for NCEP OSSE

Calibration of NCEP OSSE was conducted using 1993 observing systems. The calibrations were performed on existing instruments using data denial experiments, 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 a generally satisfactory agreement between real and simulated impacts. In both the real and simulated analysis, a large analysis impact in the tropics was 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. (See presentation at the work shop)

In both real and simulated experiments, the RAOB wind has the most impact. 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 the 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 atmospheres 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. Further detailed evaluation of the data impact in the simulation experiments is discussed by Errico et al. (2007) and the unrealistic analysis increment in the SH was pointed out. Although the data impact with a slowly varying SST could be tested in the SH with this Nature Run, because of this problem in SST the results are mainly presented for the NH.

3.3. Evaluation of DWL impact using the NCEP OSSE

In the NCEP OSSE, instead of evaluating a specific type of 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.

Mainly the meridional wind (V) is used to assess the performance of DWL. Note that the evolution of atmospheric phenomena at 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. U and V contain the information about relative vorticity at the synoptic scale while U and T contain information about wave guides (Hoskins et al. 1993). Therefore, V depicts the information about relative vorticity. The large scale U component can be inferred from T observations in the extratropics, while DWL wind observations mainly define the synoptic scale wave which is represented in relative vorticity and V.

DWL data impacts are evaluated within various contexts. Since the DAS used for this OSSE is old and the Nature Run is also old, detailed results have to be confirmed using the newer Nature Run described in Section 4 with newer DAS.

Radiance data: Without radiance data major impacts of DWL are over SH. However, with radiance data the center of the impact moved to the tropics and the magnitude of impact over NH was reduced by half. However, the large impact over the tropics was quickly reduced in forecast fields.

Scanning: The experiments showed the definite advantage of scanning. With the same amount of data uniformly distributed, scanned DWL can produce about four times more impact compared to non-scanned DWL. Note there is an overwhelming technical difficulty in building space based DWLs which can continuously scan in all directions.

Different DAS systems and model resolutions: Analysis impact does not change greatly with improved DAS system and model resolution. However, with an improved DAS and model, improvement in the analysis is maintained in the forecast. This is made clear at large scales.

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 effects of data distribution Masutani et al. (2006) and Woollen et al. (2008).

4. International collaborative Joint OSSE

4.1. Background

OSSEs also require the best knowledge of many areas of NWP systems. The Nature Run has to be produced using a state of art NWP model at the highest possible resolution. Simulating data from a Nature Run requires experts for each instrument. Simulations and assimilations have to be repeated with various configurations. Efficient collaborations are essential for producing timely and reliable results.

From the experience of the OSSEs performed during recent decades, we realize that introducing a new Nature Run consumes a significant amount of resources. For a Full OSSE all major existing observations have to be simulated. The simulation of observations requires access to the complete model level data and a large amount of resources, and it is important that the simulated data from many institutes be shared among all the OSSEs. By sharing the Nature Run and simulated observations, OSSEs will be able to produce results which can be compared, which will enhance the credibility of the results. Based on these thoughts a broad group of US and international partners formed the "Joint OSSEs" (Masutani et al. 2007, Masutani et al. 2008).

4.2. 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 and the same Nature Run can provide a valuable crosscheck of data impact results.

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 (25 km) 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. A period from April 10 to May 15 was selected to study mid latitude storms. The version of the model used was the same as the ERA-Interim reanalysis at ECMWF (labeled 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 at ECMWF) which is used throughout the experiments.

Mid latitude systems in the T511NR were found to be remarkably realistic (Masutani et al. 2007). Mid latitude cyclone statistics were produced using Goddard’s objective cyclone tracker. Distribution of cyclone strength across the pressure spectrum, cyclone lifespan, cyclone deepening, regions of cyclogenesis and cyclolysis, and distribution of cyclone speed and direction were studied. All statistics showed the Nature Run is within inter-annual variability. Location and intensity of the jet were found to be realistic. The cloud cover was also evaluated and found to be much improved compared to earlier nature runs.

Once over the Atlantic Ocean, signs of the development and organization of some waves into smaller-scale circulations are observed. In particular, the ECMWF Nature Run seems to also show the capability of spontaneously producing realistic Atlantic hurricanes. These findings, albeit preliminary, are suggestive that the ECMWF Nature Run simulates a realistic meteorology over tropical Africa and the nearby Atlantic and may prove itself beneficial to OSSE research focused over the AMMA or the Atlantic Hurricane regions (Reale et al. 2007). However Jung et al. (2005) reported the MJO in cy30r1 is weak compared to current ECMWF operational model.

There have been significant developments in high resolution forecast models at ECMWF since 2006 and a more realistic tropics in the T799 Nature Run is expected with a newer version of the ECMWF model. ECMWF agreed to generate a new T799 NR, when the Joint OSSE team has gained enough experience in OSSEs with the T511NR and is ready to make the best use of the high resolution Nature Run. Therefore, most of the work by the Joint OSSE team will concentrate on OSSEs using T511 Nature Run.

4.3. Calibration and simulation of observations for calibration

An extensive effort to simulate observations was conducted at NASA/GSFC/Global Modeling and Assimilation Office (GMAO) and NCEP-NESDIS. GMAO provided software and simulator have been set up to simulate existing radiance observations as well as conventional data. Calibration experiments were also conducted at GMAO using an adjoint technique. A GMAO program to add random errors to observations will also be available to Joint OSSE partners. They are planning to provide calibrated observations.

NCEP is processing a 91 level Nature Run profile, including complete surface data and climatology at observation locations. This data is intermediate data which does not have to be saved but is posted for the development of test simulations of radiance data. Simulations of radiance have to be repeated many times with different radiative transfer models and various sampling methods. NCEP and NESDIS will provide preliminary simulated observations without calibration. Users are expected to perform their own calibration. NOAA/ESRL and NCEP are working on a calibration using traditional data denial experiments.

4.4. Data distribution

The complete data for the T511NR and T799NR are saved at ECMWF, NCEP, the NASA/GSFC portal, and ESRL. Verification data at pressure levels are distributed to several institutes. This data must not be used for commercial purposes and re-distribution rights are not given. ECMWF and Joint OSSE must be given credit in any publications in which this data set is used. NCAR will track users and send the information to ECMWF and the Joint OSSE. If anyone is interested in using the data, please contact Michiko Masutani (michiko.masutani@noaa.gov) for the required User Agreement.

Simulated observations will be posted at the NASA/GSFC data portal. Separate directories will be set up and maintained for each institute.

5. Summary and concluding remarks for OSSEs

It is clear that "Full OSSEs" are a cost-effective way to optimize investment in future observing systems. Using a Full OSSE, various experiments can be performed and various verification metrics can be tested to evaluate data impact from future instruments and data distributions. An OSSE capability should be broadly based (multi-agency) to enhance credibility and to save costs.

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 very positive whereas others will be negative. 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 the interpretation of the results become credible.

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, full 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 that deliver reliable results in a timely manner. The simulation of observations for control experiments has to be repeated by each Nature Run and each OSSE. 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.

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 from simplified theoretical expectations will stimulate further theoretical investigations. When all efforts come together, OSSEs will help with timely and reliable recommendations for future observing systems.

Acknowledgements

Throughout this project many staff members at NOAA/NWS/NCEP, NASA/GSFC/GMAO, NASA/GSFC/NCCS, NOAA/NESDIS, NOAA/ESRL and ECMWF contributed. Simulation of observation at GMAO is lead by Ronald Errico and Runhua Yang. Calibration at ESRL is conducted by Yuanfu Xsi and Nikki Prive.

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.

Becker, B. D., H. Roquet, and A. Stoffelen 1996: A simulated future atmospheric observation database including ATOVS, ASCAT, and DWL. *Bull. Amer., Meteor. Soc*., **77**, 2279-2294.

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.

Hoskins, B.J. and T. Ambrizzi, 1993. Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci*., 50, 1661–1671.

Jung, T. , A.M. Tompkins and M.J. Rodwell 2005: Some aspects of systematic error in the ECMWF model. *ECMWF Technical Memo* No.471

Kalnay, E, Jusem, J C 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*, Clumbia, MD, NASA, 1-5.

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

Masutani, Michiko, Erik Andersson, Joe Terr, Oreste Reale, Juan Carlos Jusem, Lars Peter Riishojgaard, Tom Schlatter, Ad Stoffelen, Jack Woollen, Steve Lord, Zoltan Toth, Yucheng Song, Daryl Kleist, Yuanfu Xie, Nikki Priv5, Emily Liu, Haibing Sun, Dave Emmit6, Steve Greco, Sid A. Wood, Gert-Jan Marseille, Ron Errico, Runhua Yang, Gail McConaughy, Dezso Devenyi, Steve Weygandt, Adrian Tompkins, Thomas Jung, Valentine Anantharaj, Chris Hill, Pat Fitzpatrick, Fuzhong Weng, Tong Zhu, Sid Boukabara 2007: Progress in Joint OSSEs, *AMS preprint volume for 18th conference on Numerical Weather Prediction,* Parkcity UT. 25-29 June, 2007.

Masutani, Michiko, J. S. Woollen, R. Errico, Y. Xie, T. Zhu, H. Sun, J. Terry, R. Yang, S. Greco, N. Prive, E. Andersson, T. W. Schlatter, A. Stoffelen, F. Weng, O. Reale, L. Riishojgaard, G. D. Emmitt, S. Lord, Z. Toth, G.J. Marseille, V. Anantharaj, K. Fielding, G. McConaughy, S.Worley, C.-F. Shih, M. Yamaguchi ,J. C. Jusem, C. Hill, P. J. Fitzpatrick, D. Devenyi, S. Weygandt, S. A. Wood, Y. Song, E. Liu, D. Groff, M. Hart, G.Gayno, A. da Silva, M. J. McGill, D. Kleist, Y. Sato, S. Boukabara, 2008: Progress in Joint OSSEs. Three Joint OSSE nature runs and simulation of observation . *AMS preprint, 12th Conference on IOAS-AOLS* New Orleans, LA, 20-24 January 2008

Masutani, M., T. W. Schlatter, R. M. Errico, A. Stoffelen, E. Andersson, W. Lahoz, J. S.. Woollen, G. D. Emmitt,L.-P. Riishøjgaard, S. J. Lord 2010: *Observing System Simulation Experiments. Data Assimilation: Making sense of observations*, to be published by Springer in 2010.

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

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, *Quart. J. Roy. Meteorl. Soc*., **619**, 1927-1948.

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