Observation system simulation experiments for a global wind observing sounder

Lars Peter Riishojgaard,1,2 Zaizhong Ma,1,2 Michiko Masutani,1,3,4 John S. Woollen,3,5 George D. Emmitt,6 Sidney A. Wood,6 and Steven Greco6

Received 6 July 2012; accepted 3 August 2012; published 13 September 2012.

[1] Observing System Simulation Experiments (OSSEs) are a powerful tool used to assess the potential impact on numerical weather prediction skill of planned or hypothetical future observing systems. Over the past several years an international Joint OSSE collaboration has emerged centered on the use of NASA’s and NOAA’s data assimilation systems. A Nature Run provided by the European Centre for Medium Range Weather Forecasts (ECMWF) has undergone extensive validation, and a set of simulated reference observations have been subjected to a set of calibration experiments. One of the first candidate observing systems assessed by this system is a wind lidar based on the Global Wind Observing Sounder (GWOS) concept developed by NASA in response to the National Research Council (NRC) Decadal Survey. A set of 45-day assimilation and forecast experiments has been set up and executed. The experiments were done using the NCEP operational forecast system, and the results show that wind lidar observations are likely to provide a significant increase in forecast skill even when used in a modern, radiance-based assimilation system.


1. Introduction

[2] Out to a range of a few days, numerical weather prediction is largely an initial value problem, and much of the progress realized over the course of the last couple of decades can be directly traced to improvements in the specification of the initial conditions. The initial conditions are established based on observational input coming from the World Meteorological Organization (WMO) Global Observing System (GOS), and one of the main limitations of this system today (2012) is the lack of vertically resolved wind measurements over regions for which little or no radiosonde coverage is provided [e.g., World Meteorological Organization (WMO), 2008]. A space-borne Doppler Wind Lidar (DWL) [Baker et al., 1995] is arguably the most promising way to address this limitation, and the purpose of this article is to report the initial results from a series of impact assessment experiments (Observing System Simulation Experiments, or OSSEs) for a hypothetical DWL mission [Stoffelen et al., 2006; Masutani et al., 2010b]. The experiments were carried out with the new Joint OSSE system developed by the NASA’S Goddard Space Flight Center, NOAA/NCEP, NOAA/NESDIS and the Joint Center for Satellite Data Assimilation [Reale et al., 2007; Masutani et al., 2009; Andersson and Masutani, 2010].

[3] The DWL system studied here is based on the Global Wind Observing Sounder (GWOS) concept developed by NASA in response to the National Research Council (NRC) Decadal Survey [Committee on Earth Science and Applications from Space, 2007]. GWOS is a so-called hybrid system, being based on both coherent and direct detection lidar techniques. The coherent subsystem [Frehlich and Kavaya, 1991] provides wind observations by measuring the backscatter from atmospheric aerosol (primarily in boundary layer) and cloud particles, while the direct detection subsystem [Korb et al., 1992] makes its observations from molecular returns in regions of clear air, primarily in the middle and upper troposphere. Due to this synergy between the two techniques, a hybrid wind lidar is expected to fill a critical gap in the GOS, especially in terms of its impact on NWP [WMO, 2008].

[4] Compared to, for instance, the European ADM/Aeolus DWL mission [Stoffelen et al., 2005], GWOS differs by providing measurements of the motion of a given airmass along more than one direction. In its basic configuration, a wind lidar measures the average atmospheric motion (“wind”) along a line of sight (LOS) between the instrument and the air volume being measured. To provide more than one vector component of the wind, more than one line of sight must be used. In order for the data to have as much impact as possible, it is desirable to have the lidar measure at least the two horizontal components of the wind [Riishojgaard et al., 2004]. Earlier DWL concepts achieved this through the use of a scanning mechanism [Emmitt and Wood, 1991; Wood et al., 1991], whereas GWOS uses four fixed telescopes mounted at horizontal angles of 45, 135, 225 and 315 degrees with respect to the velocity vector of the spacecraft and tilted downward at an angle of 45 degrees off nadir. The data acquisition is such that the spacecraft first measures a given volume of air for instance through the fo right telescope. Approximately 80 seconds later, for a 400 km orbit, that same volume is measured through the aft right
telescope, and the two measurements are then combined to derive the horizontal wind under the assumption of zero mean vertical wind in the volume. The left pair of telescopes takes a separate set of similar measurements on the opposite side of the orbital track, and is included to improve coverage in the zonal direction. Wind observations are provided every 350 km along the two tracks, and the maximum horizontal averaging distance is 80 km. Additional information about GWOS can be found in Text S1 and Figure S1 in the auxiliary material.

2. Experimental Setup

[5] The methodology of the experiments follows the classical OSSE design described, e.g., by Atlas et al. [1984], Arnold and Dey [1986] and Masutani et al. [2010a]. Additional information about the method and set up can be found in Text S2. A 13-month long Nature Run was provided by ECMWF using cycle 30r1 of their Integrated Forecast System (IFS) model with triangular truncation T511 (T511NR). This horizontal resolution corresponds to a grid spacing of roughly 40 km. There are 91 levels in the vertical, and the output was saved every 3 hours. The initial condition was the IFS operational analysis valid at 1200 UTC on 1 May 2005, and the T511NR simulation ended at 0000 UTC on 1 June 2006. It should be noted that apart from the initial conditions, the only link to the “real” 2005–2006 period was through the use of actual, observed sea surface temperature (SST) and sea ice data from this period as lower boundary conditions. In all other respects, the atmospheric state was free to evolve in a manner that was quite different from the real atmosphere, but nonetheless meteorologically plausible. The T511NR has been extensively validated and has been shown to have realistic hurricanes and mid-latitude cyclone statistics [Reale et al., 2007; Masutani et al., 2009; Andersson and Masutani, 2010].

[6] Since the sequence of atmospheric states represented by T511NR is different from those of the real atmosphere, all types of observations must be simulated, irrespective of whether or not real-world observations of that type were available. The approach taken for these initial experiments was simple: For datatypes for which existing parallels existed (radiosondes, surface observations, aircraft data, existing satellite systems) observations were simulated at the times and locations for which actual observations were available in the real 2005–06 period, as recorded by the operational NCEP Global Forecast System. It should be noted that this methodology leads to a plausible spatial distribution as well as a realistic number of observations. However, it does not take into account the actual cloud field of the Nature Run.

[7] The process of simulating the so-called conventional observations — i.e., wind (including satellite AMVS from geostationary and polar orbiters), temperature, surface pressure, humidity - thus basically amounts to sampling T511NR at the appropriate time and location. A set of real observation types, times and locations obtained from NCEP operations for the real 2005–06 time period was used to generate observation templates, which were then populated with information from T511NR to generate the simulated observations. Simulated satellite radiances were obtained by processing T511NR profiles of temperature and humidity with the JCSDA Community Radiative Transfer model version 1.2.2. Land surface types and vegetation types provided by the Nature Run were used for the simulation, and data were simulated at the footprints and data densities used by the NCEP operational data assimilation. The simulated radiances were compared with real data in the early part of the T511NR period (before the simulated state had time to diverge strongly from the actual atmospheric state) and were found to be in general agreement. Satellite radiances (clear only) from AIRS (Aqua), AMSU-A (Aqua, NOAA-15, 16, 18), AMSU-B (NOAA-15, 16, 17), HIRS2 (NOAA 14), HIRS-3 (NOAA 15,16,17), HIRS-4 (NOAA-18), MSU (NOAA-14), MHS (NOAA-18) and the GOES sounder (GOES-10, 12) were simulated and included in the experiments.

[8] The simulation of DWL observations was conducted by Simpson Weather Associates (SWA) using the multi-agency funded Doppler Lidar Simulation Model (DLSM) described by Wood et al. [2000]. The DLSM was designed to work exclusively with numerical weather models, in this case the ECMWF T511 Nature Run. Critical to the realism of the lidar simulations were the respective representations of clouds, subgrid scale turbulence and aerosol distributions, all of which were adjusted in or added to the nature run fields. The simulations resulted in LOS observations containing both random and systematic errors as well as measurement accuracy estimates based upon the number of reflected photons detected by the instruments. Both direct and coherent detection returns were simulated.

[9] The experiments were conducted using the December 2009 version of the NCEP Global Data Assimilation System [Kleist et al., 2009]. The experiment setting is consistent with the way the system is used in NCEP operations, except that model resolutions of T126L64 and T382L64 (i.e., spectral triangular truncation of either 126 or 382 wavenumbers with 64 layers) were used rather than the current (April 2012) operational resolution of T574. Only results from the higher resolution (T382L64) experiments are shown here.

[10] It is important to point out that with the exception of the DWL observations for which the simulation approach is described above, no attempt was made to add random or systematic observation errors to the simulated data. However, the fact that the simulated observations were obtained by sampling a free-running forecast produced by a model that was different from the assimilating model and running at a higher horizontal resolution does add an element of “error” to the observations, since they would be extremely unlikely to perfectly represent an achievable state of the assimilating model. This relationship between the high resolution T511NR and the lower resolution assimilating model mimics the real-world relationship between the true atmospheric state and its imperfect representation by the forecast model, and the simulated data were in fact assimilated with assumed observation errors identical to what is used in NCEP operations.

[11] We recognize that one potential consequence of the use of perfect observations is that the forecast skill of the system may be overestimated. In contrast to a realistic setting in which the true state of the atmosphere is only imperfectly represented by the observing system, our simulated observations represent the projection of the T511NR states onto
the observation space. Since the metric of skill in the OSSE context is a measure of similarity between the forecast and the Nature Run, perfect observations may draw the forecast unrealistically close to the Nature Run. This would have two effects: On one hand it would translate into excessively optimistic estimates of the forecast skill provided by the reference observing system. On the other hand, the fact that both measures of skill used here are bounded means that it would be more difficult for the perturbation observations (in this case the wind lidar data) to further improve on this skill and the assessment of the lidar impact is therefore likely to be conservative. In two calibration experiments carried out specifically to assess the severity of this issue, we did see signs of the former of these two effects (sections 3.4 below). However, the drop in forecast skill observed when the wind observations were taken out was similar enough between OSSE and real world experiments that we have some confidence in the qualitative conclusions regarding the impact of the lidar observations.

The observation error standard deviation for the DWL data – for which no operational parallel exists - was set separately according to estimates provided by the instrument simulator and varies from 0.2 to more than 10 m/s, depending on detection technology, the number of individual laser returns used for a specific retrieval and the number of photons detected in the simulated returns. In order to avoid overfitting the wind observations, a minimum error threshold of 1.8 m/s was used, and the assumed observation error standard deviation was thus constrained to fall within the range from 1.8 m/s to 10 m/s.

3. Experimental Results

3.1. Calibration Experiments

An initial series of data assimilation experiments was conducted for the period from 1200 UTC on 1 July 2005 to 0000 UTC on 15 August 2005. Each day a 5-day forecast was issued at 0000 UTC. Data from the first week of each run were removed from these results to allow the system to adjust to the new data (“spin-up”). This left a total of 40 five-day forecasts (initial times from 7 July to 15 August 2005) for verification, which was carried out using the operational NCEP verification package. Due to space limitations only sample diagnostics are included. Figures S2–S4 can be found in the auxiliary material.

In preparation for the actual DWL OSSEs, an assessment was performed of the following basic characteristics of the OSSE system: (i) the overall skill of the OSSE system using perfect data, (ii) the impact of some of the existing wind observations, and (iii), the degree of similarity of (i) and (ii) when using simulated and real-world data, respectively.

This assessment was done based on the following two basic experiments: A control run (“CRTL”) in which all the current operational observational data were included, and a perturbation run (“NOUV”) from which all radiosonde, pilot balloons and dropsonde observations were removed. Both CRTL and NOUV were run twice: Once in “OSSE mode” on the simulated T511NR data, and once in “hindcast mode” on real observations from the corresponding period.

Figures 1 and 2 show forecast skill (extratropical 500 hPa height anomaly correlations, and tropical RMS errors of 200 and 850 hPa winds) for real and simulated experiments. For both the real and the simulated observations, the two experiments are verified against their own analyses. As one might expect due to the use of perfect observations, the skill in the T511NR experiments (Figure 2) is slightly higher than in the real case (Figure 1). This is true in both hemispheres as well as in the tropics. However, the decrease in skill (red versus black curves in Figures 1 and 2) when the wind observations are withheld is very similar in both cases. Both give rise to roughly a one-point drop in the Anomaly Correlation (AC) score at day 5, which roughly corresponds to a decrease of the useful forecast range of 8 hours. In both cases the difference is statistically significant at the 95% level. This gives us additional confidence in the ability of the system to assess the impact of wind observations, even in the case of perfect observations.

3.2. Wind Lidar Perturbation Experiments

Next, the following two perturbation experiments were done in the simulated T511NR domain. In one experiment (“NONW”), all wind observations were withheld. In the other (“DWL”), simulated GWOS observations were added to the set of observations used for the control run. Figure 3 shows basic skill measures for these two experiments together with the CRTL and NOUV experiments described above. Since the measure of skill for the OSSEs is the degree to which the Nature Run states can be simulated by the forecast, all verification is done against T511NR, and the lower plot in all four panels shows differences with respect to control run. Elimination of all wind observations from the assimilation leads to a very significant decrease in skill by the measures shown here. We also see that the addition of the simulated lidar wind observations leads to a statistically significant increase in AC score at day five (120 h) of approximately 1.5 points in the northern hemisphere and approximately 2 points in the southern hemisphere.

The impact of the simulated lidar wind observations in the tropics is initially very large, especially at the 200 hPa level. The 850 hPa level is more strongly influenced by the lower boundary conditions and due to the nature of the observing technique, fewer wind measurements are available at this level. However, as also seen in other data impact experiments, the impact tends to decrease rapidly over time at either level. This behavior is typical for the tropics and it illustrates the general problem of information retention and hence limited forecast skill in the tropics. It should not be taken as indicative of any real problem with the simulated wind observations.

Finally, Figure 4 shows time series of daily day-5 AC scores in the northern and southern hemispheres over the 40-day verification period. They exhibit the noisy behavior that is typical for AC time series. However, the day to day consistency confirms the statistical significance of the results: the skill in the DWL experiment only rarely drops below that of the control, while the skill of NONW is well below that of the control on most days in both hemispheres. In addition to average skill, temporal consistency is a highly desired attribute of an operational forecast system, where not only the forecast quality itself but also the measure of confidence associated with it is of interest. Observations that improve both the former and the latter are therefore considered particularly valuable, and it is encouraging to see that in both hemispheres the DWL experiments had not only the
Figure 1. Average 500-hPa geopotential height anomaly correlation as a function of forecast range for the (a) Northern and (b) Southern Hemisphere. Tropical wind vector RMS errors (m/s) at (c) 200 hPa and (d) 850 hPa as a function of forecast range. CTRL shown in black, NOUV in red. All observations used were real. Lower plot of each panel shows difference between NOUV and CTRL with, error bars indicating differences that are significant at the 95% confidence level.
Figure 2. As Figure 1, except that all observations were simulated based on T511NR.
highest average skill but also the lowest variance in skill, corresponding to the best temporal consistency.

4. Discussion

[20] The continued progress in data assimilation and numerical modeling, in the speed and memory of modern supercomputers and in the quantity and quality of meteorological observations poses an inherent difficulty for any OSSE-based assessment of the impact of future observations. The difficulty – particularly for space-based observing systems that have very long development lead times – stems from the fact that we are attempting to assess the impact of new data on future forecast systems that may be far more capable than today’s systems, in the context of a future GOS that is likely to be quite different from what we have today. However, the impact of adding wind observations in particular is relatively straightforward to assess, given the current scarcity of such observations over large areas of the globe. The general case for adding more wind observations to improve forecast skill appears to be well supported by our experiments. The NOUV and NONW experiments both show a significant negative impact of withholding wind experiments, and the additional skill shown in the DWL experiment indicates that the GOS is far from saturated with wind measurements.

[21] The calibration experiments discussed in Section 3.1 show that while the OSSE system slightly overstates the overall level of skill compared to real-world forecasts executed with the same system, the contribution to skill from the wind data in particular is very similar between the two systems. While no firm conclusions concerning the actual size of the impact of a specific new observing systems should be drawn from these experiments alone for the reasons just stated, it is nonetheless encouraging that the perturbation experiments shown in Section 3.2 indicate that wind lidar...
The T511NR was produced by Erik Andersson of the ECMWF. Initial resources for the experiments were made available by NOAA/NCEP. NASA (R. Kakar) through ROSES (grant NNX08AQ44G). Computational sing specific configuration issues for the proposed wind lidar the impact remains strong throughout the range. In the extra- the impact is very large at upper levels in the tropics but tends hemispheres as well as in the tropics. As one would expect, significant at the 95% level in the extratropics in both

5. Summary and Conclusion

A new OSSE system based on a nature run provided by ECMWF has been tested and validated. The first set of experiments done with the new system demonstrate the potential for a very significant positive impact on forecast skill of wind observations obtained by a hypothetical space-based hybrid wind lidar system. The impact is statistically significant at the 95% level in the extratropics as well as in the tropics. As one would expect, the impact is very large at upper levels in the tropics but tends to dissipate rapidly during the forecast range. In the extratropics the information retention is generally better and here the impact remains strong throughout the range.

Additional experiments underway are aimed at assessing specific configuration issues for the proposed wind lidar system and will be reported on separately at a later stage.

Acknowledgments. Support for this work was provided by NASA (R. Kakar) through ROSES (grant NNX08AO44G). Computational resources for the experiments were made available by NOAA/NCEP. The T511NR was produced by Erik Andersson of the ECMWF. Initial simulation of GOES radiance data was conducted by Tong Zhu of NESDIS. The authors thank Fanglin Yang for help with the use of the NCEP Verification System.

References

Andersson, E., and M. Masutani (2010), Collaborating on Observing System Simulation Experiments (Joint OSSE), ECMWF News, 123, 14–16.


