P741 # A LIDAR DATA ASSIMILATION SYSTEM FOR THE UPPER TROPOSPHERE AND STRATOSPHERE

George Modica*, Thomas Nehrkorn[1], Mark Leidner[1], Jeff Picton[1], John Henderson[1], and David Johnson[2]

[1]Atmospheric and Environmental Research, Inc., Lexington, Massachusetts
[2]Michigan Aerospace Corp., Ann Arbor, Michigan

1. INTRODUCTION

This paper presents the results of an effort to develop a data assimilation capability for measurements of environmental parameters at altitudes up to the mesopause (~80 km). Some background information is provided in Sections 1.1 and 1.2; descriptions of the data assimilation system and the data are presented in Sections 2 and 3, respectively; Results are provided in Section 4.

1.1. The High Speed Systems Test Program

The High Speed Systems Test (HSST) area of the Department of Defense Test Resource Management Center (TRMC) Test/Evaluation and Science/Technology (T&E/S&T) program has articulated a requirement for real-time meteorological assessment in support of flight testing hypersonic vehicles, particularly, "detailed atmospheric assessments for a given volume along the flight path close to the real-time location of the vehicle" (U.S. Army BAA 2008). Furthermore, "…prediction of atmospheric flight conditions on the flight paths of high speed, hypersonic vehicles is highly desired to better assess flight test variables and also to better understand flight vehicle performance as it accomplishes its mission." The altitudes of interest are from the surface to 80 km. Situational awareness of the meteorological conditions will result in a reduction in the vehicle performance uncertainty that is a large contributor to the costs of flight testing as well as providing numerous other benefits. Michigan Aerospace Corporation, Inc. (MAC) is leading the development of a LIDAR instrument that will acquire meteorological conditions at the altitudes of interest. In a parallel effort, Atmospheric and Environmental Research, Inc. (AER) is providing a data assimilation system (DAS) intended to combine the LIDAR measurements with other conventional sources of data as a means of producing useful 3- and 4-dimensional datasets of state variables, thereby directly addressing the requirements described above.

1.2. The High-Altitude LIDAR Atmospheric Sensing

The effort to acquire a high-altitude atmospheric sensing capability for the HSST program is being led by Michigan Aerospace Corporation (MAC). They have developed the High-Altitude LIDAR Atmospheric Sensing System capable of providing measurements of density, temperature, and wind speed and direction within the troposphere and stratosphere, up to approximately 80 km. The LIDAR is based on the principle of ultraviolet-based direct detection of constituent gases and is combined with a Raman channel to measure nitrogen and oxygen concentration. Results from a demonstration and test phase were provided for use in the DAS acquired as part of this project.

2. DATA ASSISMILATION SYSTEM

The purpose of the DAS is to optimally combine meteorological information distributed irregularly in space and time with a background, thereby creating a new, dynamically consistent dataset having better quality and more information content than what would otherwise be provided by either source alone. Like most DAS, the one described here for HALAS can utilize data originating from a wide variety of atmospheric observational platforms and backgrounds; however, it is also well-suited to non-conventional sources of data such as the high-altitude observations from HALAS. The HALAS DAS utilizes the ensemble method. The main components of that system are the Community Earth System Model (CESM) and an algorithm provided by the Data Assimilation Research Testbed (DART). The specific details of the DAS configuration adopted for the HALAS project are similar to that described in Raeder et al. (2012) and Pedatella et al. (2013) and are detailed in the next sections.

2.1. CESM

CESM is the Community Earth System Model designed and created by the National Center for Atmospheric Research (NCAR) as a community tool capable of representing many hydrodynamic physical processes over a wide range of time and space scales by coupling atmospheric, land, ocean, and other components (Figure 1). For the HALAS project the DAS was run as single-component assimilation for the atmosphere only; we assume the physical processes relevant to the time and space scales of the stratosphere and lower mesosphere (SLM) that are

Corresponding author address: George D. Modica, AER, Inc., 131 Hartwell Ave., Lexington, MA, 02421-3126; email: gmodica@aer.com

Approved for Public Release, AEDC PA 2014-292 Distribution is Unlimited.

important for range support are not significantly affected by the atmospheric interfaces at the ocean and land surfaces. Therefore, to a good approximation the ocean and land surface can be treated with prescribed data, obviating the need to assimilate these data at the surface for the HALAS project.

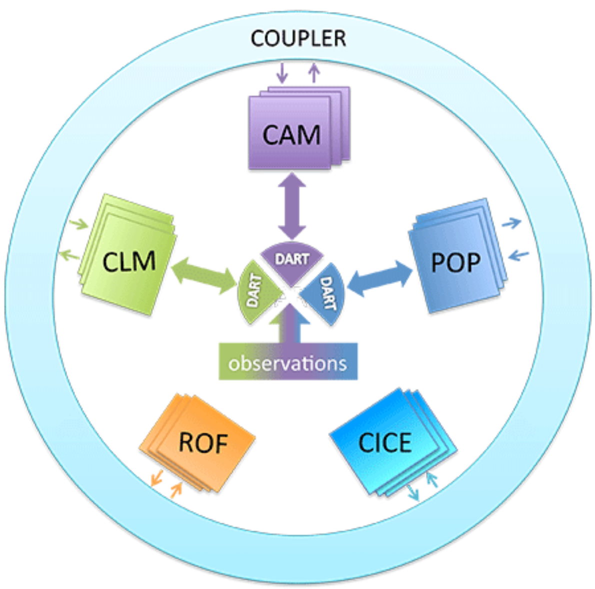

Figure 1. Illustration of the CESM coupled climate system. CAM = Community Atmospheric Model; CLM = Community Land Model; ROF = River Runoff Model; CICE = Community Ice Code; POP = Parallel Ocean Program (From climatesight.org)

The version of CESM used here was v1.1.1 configured with the F_2000_WACCM compset consisting of a 'present-day' configuration of the Community Atmospheric Model (CAM v5.2; Park et al., 2014), the Community Land Model (CLM), and prescribed ice/ocean data. The physical processes of the stratosphere and lower mesosphere were represented with the Whole Earth Atmospheric Community Climate Model (WACCM; Garcia et al. 2007, Marsh et al. 2007, Richter et al. 2010) version 4. This configuration provides global coverage at a horizontal spatial resolution of 1.9° × 2.5° in latitude and longitude, and 66 levels between the surface and 5 × 10-6 hPa (~145 km), varying from roughly 1.1 to 1.7 km in the troposphere and stratosphere to about 3.5 km in the mesosphere.

2.2. DART

DART is a community-based hydrodynamic analysis application designed around the ensemble adjustment Kalman filter (EAKF; Anderson 2009). The EAKF adjusts model values from an NWP model toward a state that is more consistent with information from a set of observations. As with CESM described in Section 2.1, DART was designed and created at NCAR. The main features of the DART configuration relevant to this report are presented in Table 2. DART has been tested specifically to work with CESM—as well as several other geophysical hydrodynamic models—and is informally known in this configuration as CESM-DART (Hurrell et al. 2013, Anderson 2009), or WACCM-DART if CESM is configured with the WACCM model as is the case here. Ensemble data assimilation is a practical methodology for performing data assimilation in whole atmosphere models. Nearly all data assimilation methods require error information for both observations and the (model) background in the form of an error covariance. Ensemble data assimilation methods obtain the background error covariance directly from an ensemble of forecast state vectors. This avoids the need to specify the error covariance, which can introduce spurious correlations between observations and the background.

The 'Lanai' version of DART was used in this project. We developed the WACCM-DART system on a 2-node, 24-CPU facility that was available in-house. The sometimes limited availability of computational resources can restrict the number of ensemble members and, therefore, the representativeness of the statistical draw of the modeled atmospheric state. DART applies techniques to compensate for the resulting sub-optimal sample covariance that can sometimes lead to a problem known as "filter divergence," in which the prior is too confident, i.e., the model solutions comprising the prior tend to diverge too far from the observations. One of the undesirable consequences of filter divergence is that too many observations are ignored during assimilation. Inflating the variance of the model error PDF can improve filter performance. Anderson (2012) describes algorithms to address these deficiencies. They are inflation and covariance localization. The vertical localization was specified with a Gaspari-Cohn function (Gaspari and Cohn, 1999) half-width of 0.5 in ln(p/p0) coordinates; see also Anderson 2012. The horizontal localization was specified with a half-width of 0.1 rad. Adaptive prior inflation was implemented during the data assimilation, which permitted the inflation factors to vary in both time and space.

3. DATA

As a demonstration of the HALAS DAS capability, we conducted an experiment with real data collected during the period 1-30 November, 2008. The baseline experiment assimilates atmospheric observations from the lowest available reporting levels (DART does not yet assimilate surface observations) up to the height of the highest reported observations (~5 hPa for radiosondes). The observations were acquired in the NCEP BUFR format (NCEP 2014) and are the same observations used to construct the NCEP/NCAR Global Reanalysis (Kalnay et al. 1996). The observation data was quality-checked by tools provided within DART. The NCEP BUFR observation types used in our baseline experiment are listed in Table 1. Among the more conventional observation types, ACARS is the Communications Addressing and Reporting System and is a digital communication system for the transmission of data—including weather parameters—

between aircraft and ground station. These are reports obtained during takeoff and landing, whereas AIRCRAFT report are generally flight-level reports from commercial, some military and reconnaissance aircraft. GPSRO is GPS Radio Occultation is a technique for measuring temperature and moisture profiles in the atmosphere from the GPS network of navigation satellites. SAT_U(V)_WIND_COMPONENT are wind measurements derived from geostationary satellite imagery. The gridded initial conditions were provided by GEOS5 as described in Lamarque et al. 2012, and used for a WACCM-DART cold start to begin the data assimilation cycle (described in Section 4).

Table 1. DART observation types used in the baseline assimilation.

No.	***Observation Type***
1	RADIOSONDE_TEMPERATURE
2	ACARS_TEMPERATURE
3	AIRCRAFT_TEMPERATURE
4	RADIOSONDE_U_WIND_COMPONENT
5	RADIOSONDE_V_WIND_COMPONENT
6	AIRCRAFT_U_WIND_COMPONENT
7	AIRCRAFT_V_WIND_COMPONENT
8	ACARS_U_WIND_COMPONENT
9	ACARS_V_WIND_COMPONENT
10	SAT_U_WIND_COMPONENT
11	SAT_V_WIND_COMPONENT
12	GPSRO_REFRACTIVITY

4. RESULTS

During the development of the HALAS baseline DAS we confined our initial efforts to the assimilation of tropospheric observation data. However, verification of the lower atmospheric results can still provide beneficial insight into the overall effectiveness of the DAS. It is also important because tropospheric variability has been shown to influence large-scale variability at mesosphere/lower thermosphere (MLT) altitudes (Liu et al. 2009). As a baseline we chose to assimilate tropospheric observations only; a second WACCM-DART experiment with higher-altitude observations will be conducted in subsequent phases of the project that will address the LIDAR data.

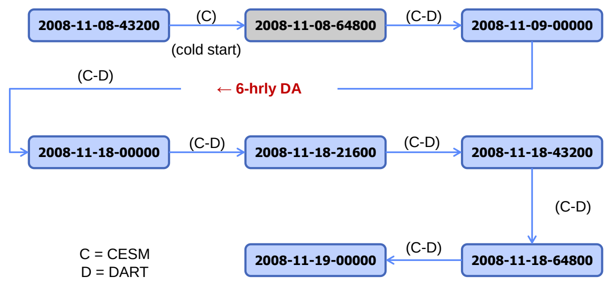

Figure 2. Data flow to spin-up DART analysis for 00 UTC 19 November 2008 verification date. DART added randomly-generated perturbations to the analysis at the second (2008-11-08-64800) time period in order to 'seed' the ensemble spread. See text for additional details.

We verified the CESM-DART output products against data from the NCEP/NCAR Reanalysis (Kalnay et al. 1996). This dataset has evolved into a de facto ground truth. The 40-year reanalysis uses a frozen DAS and a best set of atmospheric data to produce a robust dataset of the atmospheric state. The core of the NCEP/NCAR Reanalysis is the 3d-variational method as implemented in the NCEP Spectral Statistical Interpolation (Parrish and Derber 1992). We chose a meteorological case that spanned the month of November 2008, affording the added opportunity to compare the baseline results with those in the literature (Pedatela et al. 2014).

The DART assimilation tool typically requires a spin-up period to develop a desirable amount of spread among the individual ensemble members. The spin-up process used for the baseline meteorological case is illustrated in Figure 2. The CESM NWP model was started in a "cold start" mode at 12 UTC on November 8, 2008 (from this point forward we will use the notation 'YYYY-MM-DD-SSSSS' or 2008-11-08-43200), i.e., the model started from previously prepared GEOS5 initial conditions (Lamarque et al. 2012) only and no observations were assimilated with DART. Note that at this point all ensemble members were identical since no attempt had been made to cultivate an ensemble spread. The CESM model was next run for 6 hours to 2008-11-09-00000, at which time the forecast was updated with observations. DART provides an option to randomly perturb the state variables to effectively create an ensemble with a minimal amount of spread. The intent is for these small, random perturbations to grow into larger, more meaningful departures from the mean during subsequent forecasts. These perturbations were added to the DART analysis at 2008-11-09-00000; we chose to perturb the u- and v- wind components and temperature thereby creating an ensemble of unique initial conditions for the subsequent run. It is usually recommended to have many ensemble members (20-80) in order to provide a more reliable statistical draw of the state variables; however, in this study the ensemble size was limited to 6 in order to reduce computational demands. DART has the capability to correct for such limited ensemble sizes through a sampling error correction (Anderson 2012), and this feature was activated in this experiment. A series of 6-hourly CESM forecasts and DART updates resulted in an analysis valid at 2008-11-19-00000. Covariance inflation was applied to the prior (Anderson 2012) during all applications of DART in Figure 2.

We discovered that the spin-up process mentioned earlier indeed took quite a few WACCM-DART iterations before an acceptable number of observation were included in the analysis. Figure 3 includes two plots: the first is the percentage of the total number of observations accepted into each DART analysis at 00 UTC during the assimilation period. (Data for the 06, 12, and 18 UTC times were omitted for clarity.) The second plot shows the percentage of the total number of observations that were rejected by the DART outlier test. The outlier test computes a measure of the difference between the ensemble mean and the observation value; an observation will be rejected if it exceeds a predetermined threshold. The two curves in Figure 3 together suggest that fewer observations are rejected with time as the ensemble mean evolves toward the observations.

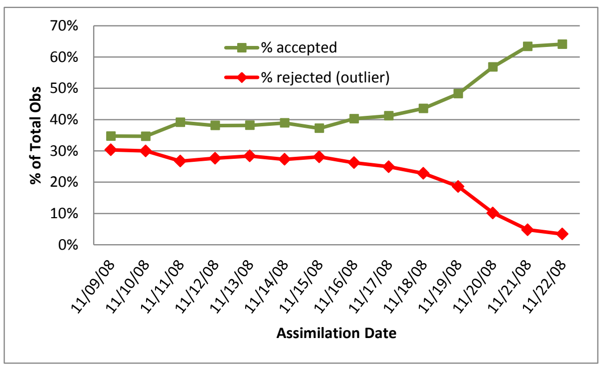

Figure 3. Observations accepted (green curve) and rejected (red curve) by DART as a percentage of the total number of available observations at each 00 UTC assimilation time for the baseline experiment.

A comparison of some representative output from WACCM-DART and the NCEP/NCAR reanalysis is presented in Figure 4. The 500 hPa reanalysis temperature field is shown in the top figure with the WACCM-DART ensemble mean temperature field at the bottom. Note that the WACCM-DART output is for the nearest model computational surface, or ~510 hPa. A similar comparison of the zonal wind at 300 hPa is provided in Figure 5. Subjectively, the pairs of data compare fairly well; a quantitative assessment will be included in a later effort in the project.

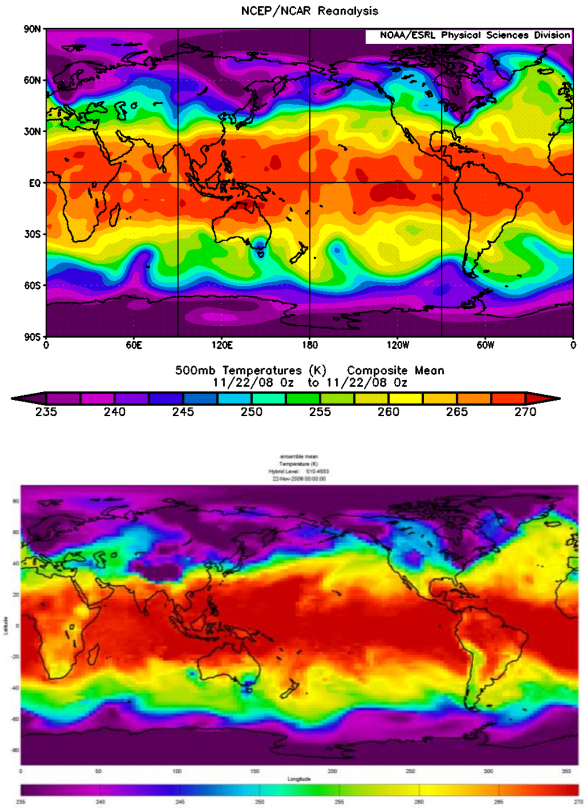

Figure 4. The 500 hPa temperature field (K) valid 22 November 2008 at 0000 UTC from (TOP) the NCEP/NCAR Reanalysis (Kalnay et al. 1996) and (BOTTOM) the WACCM-DART DAS posterior ensemble mean from the nearest model computational surface (~510 hPa). Reanalysis image provided by the NOAA/ESRL Physical Sciences Division, Boulder Colorado from their Web site at http://www.esrl.noaa.gov/psd.

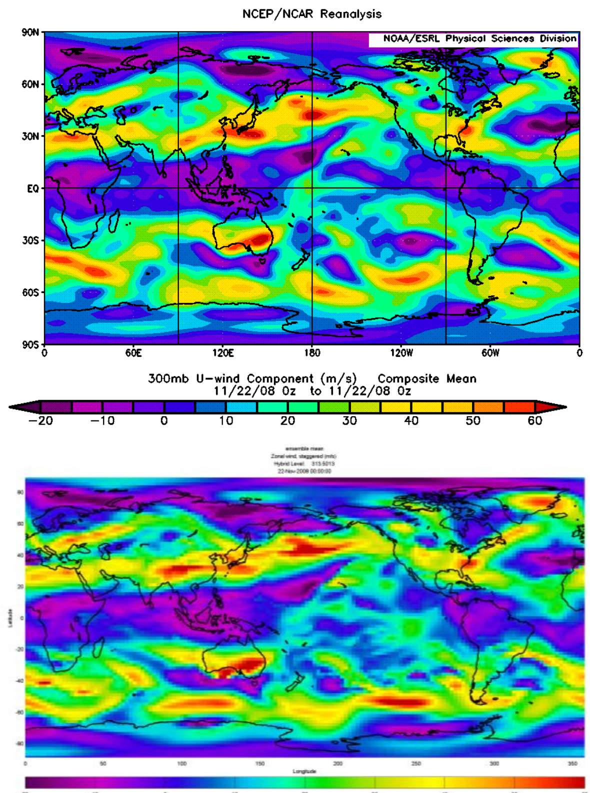

Figure 5. Same as Figure 4 but for 300 hPa zonal wind (ms^{-1}).The WACCM-DART output in this case is for a model computational surface at ~313 hPa.

5. SUMMARY AND CONCLUSIONS

The High Speed Systems Test program requires best estimates of the parameters of state at altitudes extending to about 80 km. We adapted a version of the Whole Earth Community Climate Model coupled with the DART data assimilation tool to combine HALAS Lidar observations with conventionally acquired observational data and a dynamically consistent background in order to provide improved meteorological analyses for selected field experiments.

We first conducted a baseline test of the DA system using conventional 6-hourly prepBUFR observations. Results from the baseline test helped to identify the amount of time that will be required to spin-up the modeled atmosphere to a point where the ensemble spread is sufficiently large. Note that up until this point is reached, the DART outlier rejection test rejects too many observations due to insufficient ensemble spread, even though the ensemble mean was far from the observations. In the baseline test this point was reached after about 32 6-hourly DA cycles, or about 8 days of simulated time. After this point, DART began to accept a larger proportion of the observations, rising from 40% at the beginning of the DA period to about 65% before leveling off. During this time, the analysis increments were observed to grow both in terms of spatial extent and magnitude.

Acknowledgements: The authors would like to thank the Test Resource Management Center (TRMC) Test/Evaluation and Science/Technology (T&E/S&T) program for their support. This work is funded by T&E/S&T program through the High Speed System Test Technology (HSST) area. The CESM project is supported by the National Science Foundation and the Office of Science (BER) of the U.S. Department of Energy. We gratefully acknowledge Nick Pedatella from the NCAR High Altitude Observatory for valuable suggestions on configuring a suitable WACCM-DART configuration for this project. We are also indebted to Nancy Collins and Kevin Raeder from NCAR's Data Assimilation Research Section for their very helpful and responsive DART guidance. The NCEP 6-hourly reanalysis plots were acquired from the NOAA Earth System Laboratory data server at http://www.esrl.noaa.gov/psd/data/composites/hour.

REFERENCES

Anderson, Jeffrey, 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371. doi: http://dx.doi.org/10.1175/MWR-D-11-00013.1.

Anderson, Jeffrey, Tim Hoar, Kevin Raeder, Hui Liu, Nancy Collins, Ryan Torn, Avelino Avellano, 2009: The Data Assimilation Research Testbed: A Community Facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi: http://dx.doi.org/10.1175/2009BAMS2618.1.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation, *Mon. Wea. Rev.*, **129**, 2884–2903.

Garcia, R. R., D. R. Marsh, D. E. Kinnison, B. A. Boville, and F. Sassi, 2007: Simulation of secular trends in the middle atmosphere, 1950-2003, *J. Geophys. Res.*, **112**, D09301. doi:10.1029/2006JD007485.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions, *Q. J. R. Meteor. Soc.*, **125**(554), 723–757.

Hurrell, J. W. and coauthors, 2013: The Community Earth System Model: A Framework for Collaborative Research, *Bull. Amer. Met. Soc.*, **94**, 1339-1360, doi: http://dx.doi.org/10.1175/BAMS-D-12-00121.1.

Kalnay, E. and others, 1996: The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Amer. Met. Soc.*, **77**, 437-471.

Lamarque, J.-F. and coauthors, 2012: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. *Geosci. Model Dev.*, **5**, 369-411. doi:10.5194/gmd-5-369-2012.

Liu, H.-L., F. Sassi, and R. R. Garcia, 2009: Error growth in a whole atmosphere climate model, *J. Atmos. Sci.*, **66**, 173–186.

Marsh, D. R., R. R. Garcia, D. E. Kinnison, B. A. Boville, F. Sassi, S. C. Solomon, and K. Matthes, 2007: Modeling the whole atmosphere response to solar cycle changes in radiative and geomagnetic forcing, *J. Geophys. Res.*, **112**, D23306, doi:10.1029/2006JD008306.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 1994: NCEP/NCAR Global Reanalysis Products, 1948-continuing. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at http://rda.ucar.edu/datasets/ds090.0/.] Accessed 24 Jul 2014.

Park, S., C.S. Bretherton, and P.J. Rasch, 2014: Integrating Cloud Processes in the Community Atmosphere Model, Version 5. *J. Climate*, **27**, 6821–6856, doi: http://dx.doi.org/10.1175/JCLI-D-14-00087.1.

Parrish, D. F. and J.C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system, *Mon. Wea. Rev.*, **120**, 1747-1763.

Pedatella, N.M., K. Raeder, J. L. Anderson, and H.-L. Liu, 2014: Ensemble data assimilation in the Whole Atmosphere Community Climate Model, *J. Geophys. Res.: Atmos.*, **119**, doi:10.1002/2014JD021776.

Pedatella, N. M., K. Raeder, J. L. Anderson, and H.-L. Liu, 2013: Application of data assimilation in the Whole Atmosphere Community Climate Model to the study of day-to-day variability in the middle and upper atmosphere, *Geophys. Res. Lett.*, **40**, 4469–4474, doi:10.1002/grl.50884.

Richter, J. H., F. Sassi, and R. R. Garcia, 2010: Towards a physically based gravity wave source parameterization in a general circulation model, *J. Atmos. Sci.*, **67**, 136–156.

U.S. Army Broad Agency Announcement W900KK-08-R-0017 AMENDMENT 0004 (2008), High Speed Systems Test Science & Technology (HSST S&T) Test Technology Area (TTA), 80 pp.