ASSIMILATION OF AIRS VERSION 6 DATA IN AMPS

Jordan G. Powers, Priscilla A. Mooney, and Kevin W. Manning
Mesoscale and Microscale Meteorology Laboratory
National Center for Atmospheric Research
Boulder, Colorado, USA

1. INTRODUCTION

The Antarctic Mesoscale Prediction System (AMPS) is a real-time numerical weather prediction capability that provides model guidance for the forecasters of the U.S. Antarctic Program (Powers et al. 2012). AMPS also supports researchers and students, international Antarctic efforts, and field campaigns. AMPS uses the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008), and the forecast initialization involves data assimilation using WRFDA (WRF data assimilation system; Barker et al. 2004; Barker et al 2012). For this a three-dimensional variational (3DVAR) assimilation approach is employed.

Atmospheric observations from satellites are particularly important for data assimilation for model domains over the high southern latitudes given the expanses that lack in-situ measurements. One such observation source is the Atmospheric Infrared Sounder (AIRS), an instrument on NASA's polar-orbiting Aqua satellite that provides vertical profiles of temperature and moisture. Over time, new versions of AIRS data come out, and the most recent upgrade is AIRS Version 6 (AIRS V6). Studies using the previous version of the data, AIRS V5, for forecast initialization using WRFDA have shown it to improve WRF simulations (Singh et al. 2011; Chou et al. 2010). In the context of AMPS, however, a prior investigation of the impact of AIRS V5 returned mixed results (Powers and Manning 2012), and thus AIRS V5 has not been data assimilated in AMPS. With the advent of AIRS V6, this study investigates the data's impact on WRF Antarctic forecasts in order to decide upon its use in AMPS.

2. BACKGROUND AND METHODOLOGY

The Atmospheric Infrared Sounder (*see* Chahine et al. (2006), LeMarshall et al. (2006)) measures upwelling infrared energy (the brightness from the surface and the atmosphere) and estimates atmospheric water vapor and temperature. Carried on NASA's Aqua satellite, it scans across a ground-projected track with an 800-km swath. AIRS has over 2300 spectral channels and uses these to develop temperature and water vapor profiles. AIRS's resolution at nadir is 13.5 km, and its temperature and humidity accuracies in the troposphere are 1K and 15%. This study uses the AIRS Level 2 retrieval products of temperature and water vapor, with data thinned to 45 km for the data assimilation experiments.

AIRS data are periodically updated, and the most recent revision is Version 6 (Olsen 2013). V6 first differs from the previous V5 in its use of neural networks to generate a required first guess, compared the linear regression of V5. Second, unlike V5, surface temperature is determined from shortwave channels only, using measurements of emissivity and reflectance. Third, in contrast to the 17% rejection rate for V5, less than 1% of the retrievals are rejected in V6. The processing differences yield more upper atmosphere data for V6. Fourth, V6 and V5 differ in the treatment of quality control (QC) flags. In contrast to V5, in V6 there is a quality control flag for each variable and level. This information allows a more targeted selection within the dataset.

Knowledge of the surface emissivity is essential for deriving accurate temperature and moisture profiles from radiances. As acquiring the surface emissivity over land surfaces is generally more difficult than over ocean surfaces, the quality of the retrievals over the ocean is, in general, higher. This prompts experiments to investigate the differences from assimilating ocean or land AIRS data. Note that AIRS discriminates among eight surface types: coastline, land, ocean, sea ice, sea ice MW (derived from microwave sensor), snow, glacier/snow, and snow MW (derived from microwave sensor).

This study consists of AIRS assimilation experiments for two periods of forecasts: July–August 2014 (winter) and November 2014–January 2015 (summer). For all experiments, the 30-km and 10-km AMPS grids are used (Fig. 1), and WRF is run out to five days from 0000 UTC and 1200 UTC initializations.

In the experiments here, WRF primarily uses output from the NCEP Global Forecast System (GFS) for its first-guess field and boundary conditions. However, for one setup for the summer period, WRF is cycled (i.e., WRF forecasts are used for the first-guess). The WRFDA system using 3DVAR is used for the AIRS assimilation. The base collection of observations assimilated in all of the runs is the standard set used in AMPS: surface data (AWS, SYNOP, METAR); upper-air soundings; aircraft observations; ship and buoy observations; geostationary and polar-orbiting satellite AMVs (atmospheric motion vectors); GPS radio occultations; and AMSU (Advanced Microwave

Sounding Unit) radiances. On top of that, AIRS data are ingested.

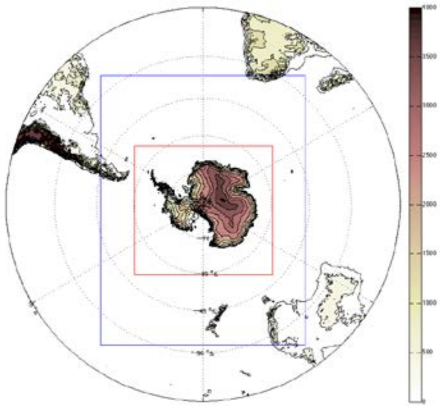


Fig. 1: AMPS domains used for AIRS impact testing. Outer frame (blue) is the 30-km domain, while inner frame (red) is the 10-km domain. Topography shaded; scale (m) to right.

3. RESULTS

a. ***Winter Experiments***

There are five experiments in total for the winter period (29 July–31 August 2014). Experiment 1 is a control run assimilating only the base observation set. Experiment 2 assimilates AIRS data over all surfaces where the data has quality control (QC) levels of "best" and "good" (i.e., QC flag= 0 or 1). Experiment 3 includes both categories of QC, but only AIRS data over ocean surfaces. Experiments 4 and 5 use the surface classifications of 2 (all surfaces) and 3 (ocean), respectively, but use QC level "best" (QC flag= 0).

For the winter experiments Taylor diagrams (Taylor 2001) are used to determine the influence of the AIRS V6 data. Taylor diagrams display three statistics together: the correlation coefficient (*r*) for the experiment and observations, the standard deviation (SD) of the experiment or observed values, and a centered root mean squared difference (RMSD) from the observations. The statistical significance of the difference of the standard deviation values between the control and an AIRS run are analyzed using an F-ratio test, while the significance for correlation coefficient differences are analyzed using the Fisher Z-transformation.

The forecast parameters reviewed are temperature (T), relative humidity (RH), and pressure at the surface and T and RH at upper levels. The verification uses observations from surface stations and upper-air sites across Antarctica. While a full set of verification diagrams were produced, only a sample is reviewed here to illustrate the overall results.

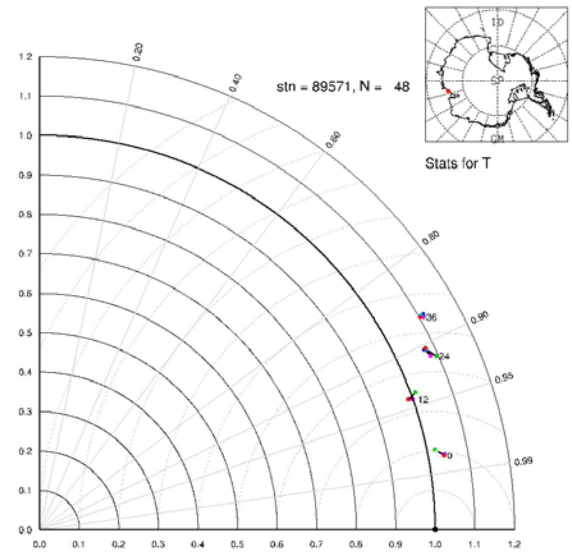


Fig. 2: Taylor diagram verification of surface temperature forecasts for Davis Station, winter experiments. Location of station shown in inset (red dot). Experiment values colored as follows: Expt 1: No AIRS; Expt 2: AIRS all QC, all surfaces; Expt 3: AIRS all QC, ocean only; Exp 4: AIRS best QC, all surfaces; Expt 5: AIRS best QC, ocean only. Clusters of results from bottom to top: Hr 0, Hr 12, Hr 24, Hr 36.

Figure 2 shows results for surface temperature at Davis Station. The location of Davis is shown in the inset map. Hours 0, 12, 24, and 36 are presented, with the experiments color-coded (see Fig. 2 caption). The diagram first reveals that, predictably, the errors for all of the experiment configurations increase over time. More importantly, however, for each verification time the error values are clustered, and the error differences are small. There is no consistently better-performing run: the experiment with the best correlation, RMSD, and standard deviation varies with forecast hour.

Figure 3 shows the results for temperatures at 500 mb for Casey Station, as averaged for all Antarctic upper-air sites. Again, the experiment differences are small and the errors are clustered. Figures 2 and 3 are representative of the diagrams for the temperature and RH at surface stations as well as upper levels across the continent. All show little difference between the simulations for a given forecast hour. And, among the variations in QC level or underlying surface, no AIRS experiment emerges with consistently or statistically significantly better results. Conversely however, there is no degradation seen from the assimilation of AIRS.

The differences between the experiments and control run (no AIRS) have been compared and significance testing performed. This reveals no statistically significant differences from the inclusion of AIRS, whether stratified for underlying surface type or for QC level. Figure 4 provides an example in a comparison of correlation coefficients of surface temperature for the AIRS runs and the control run. The 95% and 90% confidence intervals are plotted with the z-score. If the score exceeds the range, then the difference in correlation coefficient of the given AIRS run from the control is statistically significant. Here, these values at any station do not exceed the confidence limits. The results for surface RH and pressure are comparable. Similarly, differences in the standard deviations of forecast values between the simulations are generally not statistically significant.

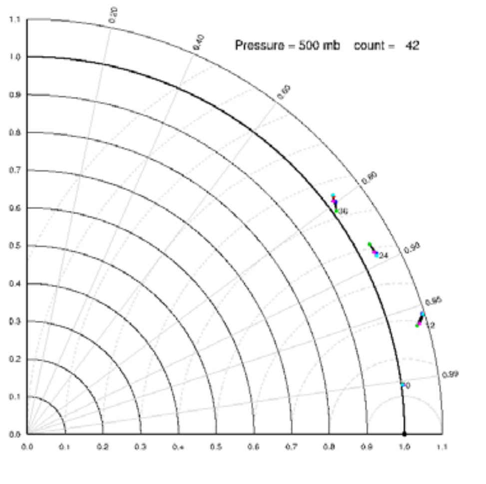


Fig. 3: Taylor diagram verification of 500 mb temperature forecasts for Casey Station, winter experiments. Experiment values colored as in Fig. 2. Clusters of results for varying forecast hours from bottom to top: Hr 0, Hr 12, Hr 24, Hr 36.

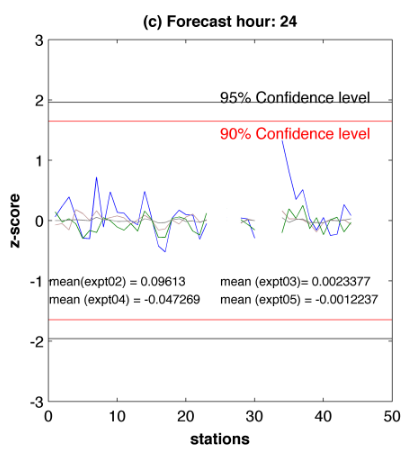


Fig. 4: Z-scores for comparison of correlation coefficients for surface temperatures across all Antarctic surface stations verified for hour 24 of winter experiment runs. 95% and 90% confidence intervals plotted and averaged z-scores for the AIRS experiments shown. Stations numbered randomly on x-axis. Gaps indicate lack of a value of the statistic for the numbered station.

In summary, the winter test period does not reveal significant impacts from the assimilation of AIRS V6 data. In addition, differences from discriminating between QC levels and underlying surface types do not emerge. In contrast to the previous AIRS V5 testing in AMPS, however, there are no degradations seen from AIRS V6. The results may in part reflect the use of the GFS as a first-guess, a background which has assimilated AIRS data already. While there is a redundancy in assimilation, here, this is part of the AMPS WRF DA approach and has been found to benefit the forecasts overall.

These analyses for the winter period adopted a less-traditional Taylor diagram focus, which may not capture all impacts. Given this, and given that it would be worthwhile to examine other forecast quantities (e.g., winds, heights) during the field season (Austral summer), further experiments have been performed. Furthermore, the issue of the background field used suggests that the assimilation of AIRS V6 in a cycling context may better illuminate the data's potential impact. Described in the next section, the summer experiments look at a different season, take a different verification approach, and consider different first-guess fields.

b. *Summer Experiments*

The summer test period covers late November 2014–January 2015. Two experiments are performed in which forecasts with the standard AMPS data are compared to runs adding all of the AIRS data (i.e., over all surfaces and using good and best QC levels). The first experiment uses the regular AMPS first-guess (GFS), while the second adopts cycling.

For the first experiment, Fig. 5 presents the results for upper-air temperatures. Multiple pressure levels and forecast hours are shown, and the run pairs are "AIRS" (AIRS V6 assimilated) and "No AIRS" (standard data only). Dots for any hour indicate that the differences are statistically significant. On the bias and RMSE panels, the bias curves for the runs are the pair closer to the zero line. The AIRS run mostly has lower T biases, and the differences are statistically significant. While there are statistically significant differences in the RMSEs, the differences are too small to be judged practically significant.

In contrast to the better biases for AIRS seen at upper levels, the aggregate surface temperature results actually show a cold bias increase for the AIRS runs (Fig. 6). Note that these statistics reflect an average across all surface stations used, with equal weight given to each site. However, it is found that the error differences for individual stations vary, and certain regions show improvements from AIRS.

For example, Fig. 7 shows the results for surface temperature at South Pole. The top panel presents the AIRS (maroon) and No AIRS (blue) forecasts, along with the observations (green). The AIRS forecasts are, on the whole, not as warm as the No AIRS runs, which is an improvement given the warm bias of the standard (No AIRS) AMPS setup. This AMPS warm bias at Pole has been a consistent issue over the years.

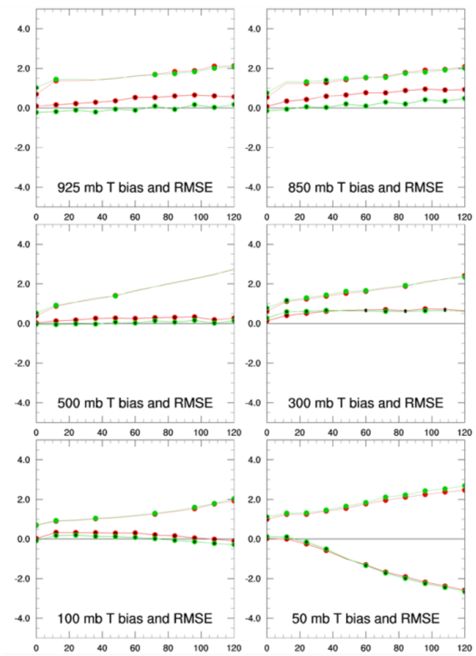


Fig. 5: Summer experiment AIRS (green) and No Airs (red) temperature verification statistics for RMSE (upper pairs of curves) and bias (lower pairs of curves) at upper levels. Errors in °C. Forecast hours (0–120) along abscissa. Dots indicate where error differences are statistically significant.

The lower left panel shows bias (blue), RMSE (red), bias-corrected RMSE ("Stdv"; pink), and correlation coefficient (black). AIRS results are in the thick curves, with No Airs results in the thin curves. The lower right panel presents the average forecast and observed temperatures over the diurnal cycle. At South Pole, and at some other Plateau sites, the warm bias in AMPS is reduced with the AIRS data. The bias decreases from 2.0C to 1.3C, and the RMSE decreases from 2.6C to 2.2C. The lower right panel shows the overall warm bias reduction clearly.

Errors in the winds have also been compared. For *u*, *v*, and wind speed, both at the surface and upper levels, error differences (not shown) are negligible and not statistically significant. This is perhaps not surprising, given that the data assimilated directly address the mass field, rather than the momentum field.

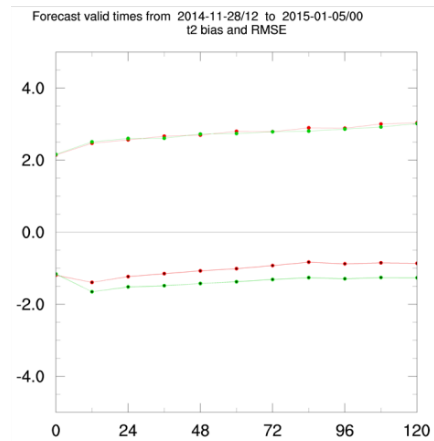


Fig. 6: Summer experiment AIRS (green) and No Airs (red) surface temperature RMSE (upper pairs of curves) and bias averaged over all surface stations examined. Errors in °C. Forecast hours (0–120) along abscissa. Dots indicate that error differences for the given forecast hour are statistically significant.

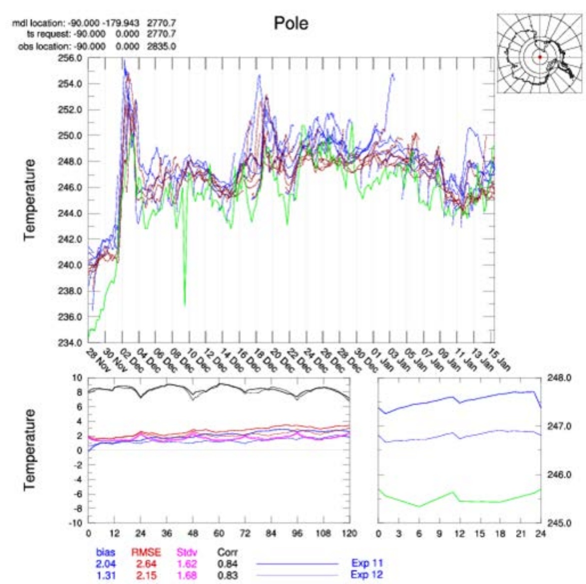
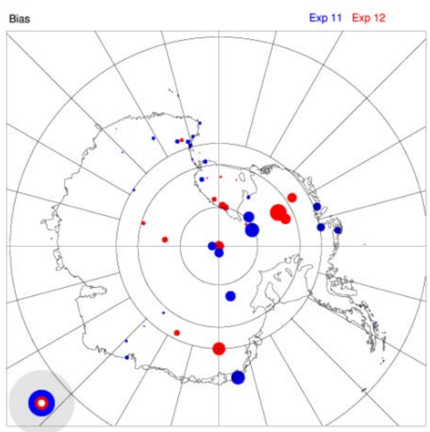
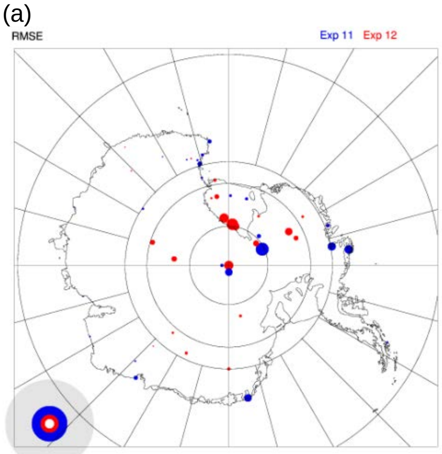


Fig. 7: Summer experiment surface temperature results for AIRS and No AIRS experiments at South Pole. Top panel: Observations (green), AIRS forecast (maroon) temperatures, and No AIRS forecast (blue) temperatures. Bottom left: Average errors per forecast hour (hrs 0–120)— AIRS thick solid, No AIRS thin solid. Blue= bias; red= RMSE; pink= bias-corrected RMSE; black= correlation. Bottom right: Average biases (°C) for a 24-hr diurnal period.

Figures 8(a)–(d) show the distribution of the results for surface temperature errors. Red circles mark the AIRS run as better, while blue circles mark the No AIRS run as better, with the size of the circle proportional to the magnitude of the improvement. Figure 8(a) presents the bias comparisons. Lower biases from AIRS are seen at Pole, in West Antarctica, and in the East Antarctic plateau. At other sites, the biases are lower in No AIRS, and the results are consistent with the overall surface temperature bias result in Fig. 6. As seen in Figs. 8(b) and (c), however, the AIRS V6 data improve the RMSEs and bias-corrected RMSEs at the majority of sites. Lastly, Fig. 8(d) shows the predominant improvements in correlation coefficients from assimilation of AIRS V6 data.

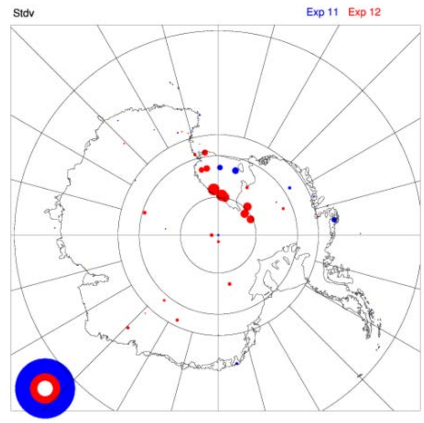


(a)

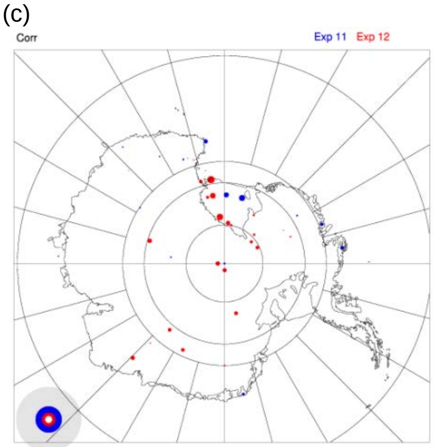


(b)

Fig. 8: Comparison of surface temperature errors for summer AIRS and No AIRS experiments. Red= AIRS better; blue= No AIRS better. Circle size is proportional to magnitude of difference in improvement. (a) Bias. (b) RMSE. (c) Bias-corrected RMSE. (d) Correlation coefficient.



(c)



(d)

The second summer experiment examines AIRS V6 assimilation in WRF cycling. This may provide a cleaner test of the impact of AIRS on a WRF forecast in AMPS. It is found that for temperature and heights aloft, there are improvements in each statistic: biases and RMSEs decrease, while correlations increase. Figure 9 shows statistically significant reductions in temperature bias aloft for most forecast hours. The RMSE decreases are statistically significant as well, but being small, they are not practically significant.

While upper-level height and temperature statistics are improved with AIRS in cycling mode, the surface temperature biases and RMSEs (aggregated across all stations; not shown) still are not. This result mirrors that of the first summer experiment (GFS first-guess). Lastly, wind speed and wind component statistics are not significantly different in the AIRS and No AIRS cycling runs.

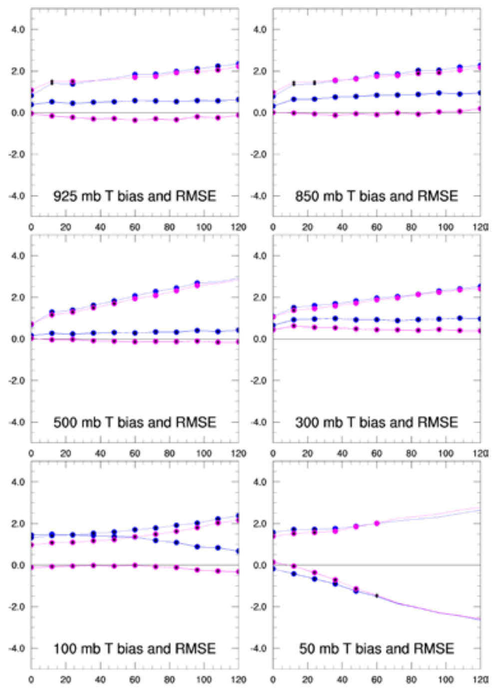


Fig. 9: Summer cycling experiment AIRS (magenta) and No Airs (blue) temperature verification statistics for RMSE (upper pairs of curves) and bias at upper levels. Errors in °C. Forecast hours (0–120) along abscissa. Dots indicate where error differences are statistically significant.

4. SUMMARY AND CONCLUSIONS

Given its high-latitude coverage, and improvements over previous the version, AIRS Version 6 (V6) data are tested for their impact on AMPS forecasts. Here the V6 data used are the retrievals of temperature and moisture, and the tests assimilate the AIRS data on top of the standard observation suite. Two periods, summer and winter, are examined. For the summer period, the standard AMPS first-guess field is used, with tests addressing differing underlying surface types and QC levels. For the winter period, the impact of all AIRS data is investigated using both the standard cold-start first-guess approach and a cycling approach.

For the summer experiments, there is no consistent signal seen from the assimilation of AIRS V6 data. Using the data over all surface types and for all acceptable QC levels, versus limiting the data to the highest QC level or over ocean surfaces, does not make a significant difference. However, in contrast to previous AMPS testing with AIRS V5 data, there are no negative impacts seen from AIRS V6.

The winter experiments take a slightly different verification approach and also explore the effect of first-guess mode. The first experiment (standard AMPS first-guess) shows overall positive impacts from AIRS V6 data. At the surface, while the temperature cold bias at the margins of continent increases, the East Antarctic plateau warm bias decreases and West Antarctica results improve. Temperature correlations and RMSEs are somewhat better as well. Aloft, the temperature bias was substantially reduced, especially in the lower troposphere. Height biases are also generally improved.

In the cycling experiment, the benefits of AIRS V6 are more pronounced. Although averaged over all surface stations a temperature cold bias increases, the surface RMSEs are improved, and aloft both temperature biases and RMSEs are improved. Height and wind verifications improve overall with AIRS V6 in this mode. Variable correlations are also better with the assimilation of the V6 data.

Taken as a whole the results were deemed sufficient to support the real-time assimilation of AIRS V6 data in AMPS WRF forecasts. AIRS V6 have been used in AMPS since February 2015.

ACKNOWLEDGEMENTS

The authors thank the NSF Division of Polar Programs for its support of AMPS.

REFERENCES

Barker, D.M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional (3DVAR) data assimilation system for use with MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.

Barker, D.M., and Co-Authors, 2012: The Weather Research and Forecasting Model's community variational/ensemble data assimilation system: WRFDA. *Bull Amer. Meteor. Soc.*, **93**, 831–843.

Chahine, M.T., and Co-authors, 2006: Improving weather forecasting and providing new data on greenhouse gases. *Bull. Amer. Meteor. Soc.*, **87**, 911– 926.

Chou, S.-H., B. Zavodsky, and G. Jedlovec, 2010: Regional precipitation forecast with Atmospheric Infrared Sounder (AIS) profile assimilation. 11th WRF Users' Workshop, Nat'l Center for Atmos. Res., Boulder, CO.

LeMarshall, J., and Co-authors, 2006: Improving global analysis and forecasting with AIRS. *Bull. Amer. Meteor. Soc.*, **87**, 891–894.

Olsen ET (2013) AIRS/AMSU/HSB Version 6 changes from Version 5. *NASA-JPL Tech. Rep.*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California [available online at http://disc.sci.gsfc.nasa.gov/AIRS/documentation/v6_docs/v6releasedocs-1/V6_Changes_from_V5.pdf].

Powers, J.G., and K.W. Manning, 2012: Assimilation of AIRS retrievals in Antarctic WRF forecasts. 13th WRF Users' Workshop, Nat'l Center for Atmos. Res., Boulder, CO.

Powers, J.G., K.W. Manning, D.H. Bromwich, J.J. Cassano, and A.M. Cayette, 2012: A decade of Antarctic science support through AMPS. *Bull. Amer. Meteor. Soc.*, **93**, 1699–1712.

Singh R., C. M. Kishtawal, and P. K. Pal, 2011: Use of Atmospheric Infrared Sounder clear-sky and cloud-cleared radiances in the Weather Research and Forecasting 3DVAR assimilation system for mesoscale weather prediction over the Indian region. *J. Geophys. Res.*, **116** (D22116).

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X.-Y. Huang, W. Wang, and J.G. Powers, 2008: A description of the Advanced Research WRF Version 3. NCAR Tech. Note, NCAR/TN-475+STR, 113 pp.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106** (D7), 7183–7192.