The Impact of Dropsonde Data on the Performance of the NCEP Global Forecast System during the 2020 Atmospheric Rivers Observing Campaign. Part II: Dynamic Variables and Humidity

STEPHEN J. LORD,[a,b] XINGREN WU,[c,b] VIJAY TALLAPRAGADA,[b] AND F. M. RALPH[d]

[a] *University Corporation for Atmospheric Research/CPAESS, Boulder, Colorado*
[b] *NOAA/NCEP Environmental Modeling Center, College Park, Maryland*
[c] *I. M. Systems Group, Inc., Rockville, Maryland*
[d] *CW3E, Scripps Institution of Oceanography, University of California, San Diego, San Diego, California*

(Manuscript received 15 April 2022, in final form 6 January 2023)

ABSTRACT: The impact of assimilating dropsonde data from the 2020 Atmospheric River (AR) Reconnaissance (ARR) field campaign on operational numerical weather forecasts was assessed. Two experiments were executed for the period from 24 January to 18 March 2020 using the National Centers for Environmental Prediction (NCEP) Global Forecast System version 15 (GFSv15) with a four-dimensional hybrid ensemble–variational (4DEnVar) data assimilation system. The control run (CTRL) used all of the routinely assimilated data and included data from 628 ARR dropsondes, whereas the denial run (DENY) excluded the dropsonde data. Results from 17 intensive observing periods (IOPs) indicate a mixed impact for mean sea level pressure and geopotential height over the Pacific–North American (PNA) region in CTRL compared to DENY. The overall local impact over the U.S. West Coast and Gulf of Alaska for the 17 IOPs is neutral (−0.45%) for integrated vapor transport (IVT), but positive for wind and moisture profiles (0.5%–1.0%), with a spectrum of statistically significant positive and negative impacts for various IOPs. The positive dropsonde data impact on precipitation forecasts over U.S. West Coast domains appears driven, in part, by improved low-level moisture and wind fields at short-forecast lead times. Indeed, data gaps, especially for accurate and unbiased moisture profiles and wind fields, can be at least partially mitigated to improve U.S. West Coast precipitation forecasts.

KEYWORDS: Numerical weather prediction/forecasting; Operational forecasting; Data assimilation; Atmospheric river; Forecast verification/skill

1. Introduction

In Lord et al. (2022b, hereafter Part I), it was demonstrated that forecasts of precipitation associated with landfalling atmospheric river (AR) phenomena were improved by deploying dropsonde soundings into AR regions at least several days before landfall. ARs have an important impact on weather over the U.S. West Coast and Canada and are associated with 84% of flood damage in the western United States (Ralph et al. 2020). However, forecasting AR impact on typical watersheds, with an extent of 100 km or less, has many difficulties, among which are the comparable horizontal extents of both the AR and the target watershed areas, coastal topographic forcing on the U.S. and Canadian west coasts, and the inherent challenges of accurate precipitation prediction from operational forecast and data assimilation systems.

Real-time AR reconnaissance (ARR) observing campaigns (OCs; Ralph et al. 2020) have been designed and executed in a similar manner to those initiated for tropical cyclones (e.g., Aberson et al. 2010; Majumdar et al. 2013; Brennan et al. 2015). Moreover, dropsondes deployed in the vicinity of extratropical and tropical cyclones in midlatitudes have shown positive downstream impact (Schäfler et al. 2018; Schindler et al. 2020).

In the case of ARs, aircraft are deployed to sample the jet-like features throughout the troposphere and regions with enhanced moisture emanating primarily from the tropics. Ensemble-based (Torn and Hakim 2008; Elless et al. 2021) and adjoint-based (Reynolds et al. 2019; Doyle et al. 2019) sensitivity analyses are used for each AR IOP to direct targeted AR dropsonde observations into regions with potential for improving predictions of U.S. West Coast precipitation. The ensemble sensitivity tools use data from the European Centre for Medium-Range Weather Forecasts (ECMWF), the Canadian Meteorological Center (CMC), and the NCEP Global Ensemble Forecast System (GEFS). In the majority of cases, both the adjoint and ensemble sensitivity analyses indicated forecast sensitivity along the AR core, its edges, and in the warm conveyor belt, all of which provided complementary information where additional observations may improve forecasts (Ralph et al. 2020). More details are given in Part I.

Dropsonde soundings from each aircraft provided temperature, wind, pressure and moisture data that were transmitted to international weather prediction centers in real time for ingest into their respective operational data assimilation and forecast systems. The resulting forecasts have the benefit of these data as well as previous soundings (through the cycled data assimilation) and would be expected to be of improved predictive quality for users concerned with ARs and their regional impacts along the U.S. West Coast.

This paper examines the impact of AR dropsonde observations taken during 17 AR IOPs over January–March 2020 (AR2020), and focuses on impacts to wind, geopotential height and moisture fields as well as the integrated vapor transport (IVT; Ralph et al. 2018). The NCEP operational Global Forecast System (GFS),

Corresponding author: Vijay Tallapragada, Vijay.Tallapragada@noaa.gov

DOI: 10.1175/WAF-D-22-0072.1

TABLE 1. Summary of the time and space domains used in calculating verification statistics. Analyzed variables (and their abbreviations), vertical levels, and a short description are also listed. Time-mean statistics derived from large domains 1 and 2 are comparable, even though the domains are not identical. Appendix A contains the valid dates for each IOP. Descriptions are abbreviated (TMS: time-mean statistics for the OC; CS: case study; FS: forecast statistics; and QCD: qualitative case description). Spatial domains are shown in Fig. 1.

Domain No.	Time domain	Spatial domain and projection	Analyzed variables (abbreviation)	Levels (hPa)	Description
1	0000 UTC 24 Jan 2020–0000 UTC 18 Mar 2020	Large domain (40°W–180°, 20°–75°N, lat–lon projection)	Geopotential height (Z) Wind speed (WSPD)	500 500	TMS
2	0000 UTC 24 Jan 2020–0000 UTC 18 Mar 2020	Large domain (83°–197°W, 14°–75°N, polar stereographic projection)	Geopotential height (Z) Mean sea level pressure (PMSL)	500 Surface	TMS
1	0000 UTC 29 Jan	Domain 1	Geopotential height (Z)	500	CS
3	17 IOP dates	Small domain (115°–160°W, 33°–65°N, lat–lon projection)	Geopotential height (Z) Mean sea level pressure (PMSL) Wind speed (WSPD) Specific humidity (SPCH)	200, 500, 700, 850 Surface 200, 300, 400, 500 700, 850, 925 700, 850, 925	FS for IOP cases
4	As in domain 3	Small domain (110°W–180°, 18°–60°N, lat–lon projection) traversed by all IVT maxima	Integrated vapor transport (IVT)	250	FS for all IOP cases
5	IOP-13 (0000 UTC 7 Mar)	Baja California (60°–130°W, 15°–35°N, lat–lon projection)	IVT	Surface	QCD

consisting of a forecast model and a cycled data assimilation system, is executed with and without the observations and provides the raw impact data. The impact is measured in terms of reduced error for AR-important weather predictands such as wind and water vapor comprising the IVT, as well as other standard weather prediction variables such as geopotential height and sea level pressure.

Section 2 describes the methodology of the data impact experiments and defines fundamental diagnostic quantities such as the IVT and section 3 briefly reviews the 17 AR2020 IOPs. Section 4 gives the impact results for IVT and the basic model variables (wind, geopotential height, and specific humidity). Section 5 provides the summary and discussion. Appendix A provides details on the AR2020 IOPs, and appendix B details a case study of large positive remote impact resulting from dropsondes deployed during AR2020 IOPs.

2. Methodology

a. Data assimilation and forecast model

The experimental methodology is that of a standard NWP data impact experiment as described in Part I. One cycled data assimilation and forecast run is executed with the additional AR observations (the CTRL) and a second run (DENY) is made without these observations. All other aspects of the analysis and forecast system and generated data are the same. A set of diagnostic error post processing software is applied to each run and differences in verification statistics are generated and evaluated. The forecast model and data assimilation system used to generate the CTRL and DENY impact experiments is the GFS version 15 (GFSv15), which was an upgrade over the version used in NCEP's operations during most of the AR2020 OC.

The GFSv15 model (Yang and Tallapragada 2018) has a horizontal resolution of 13 km, and 64 levels in the vertical extending up to 0.2 hPa. The GFDL finite-volume cubed-sphere (FV3) dynamical core (Lin and Rood 1997; Lin 2004; Putman and Lin 2007; Harris and Lin 2013; Harris et al. 2020a,b) is the basis of GFSv15, and a suite of physical parameterizations comprise the GFS model. The Global Data Assimilation System (GDAS) is a 4D hybrid ensemble–variational data assimilation system (Kleist and Ide 2015). The ensemble system has 80 members at a resolution of 25 km. The hybrid algorithm combines uncertainty estimates from the ensemble and a climatological, spatially varying but fixed in time, variance. The operational GDAS observations and other data assimilation information are described in Part I.

b. Evaluation

The AR2020 experimental period began at 0000 UTC 24 January 2020 and continued until 0000 UTC 18 March, covering 55 days, of which 17 days have IOP dropsonde observations centered on 0000 UTC. Further details on the IOPs are in appendix A.

The two GFS experiments (CTRL and DENY) were executed with a 6-hourly GDAS cycle over the experimental period; forecasts out to 168 h generated from initial conditions

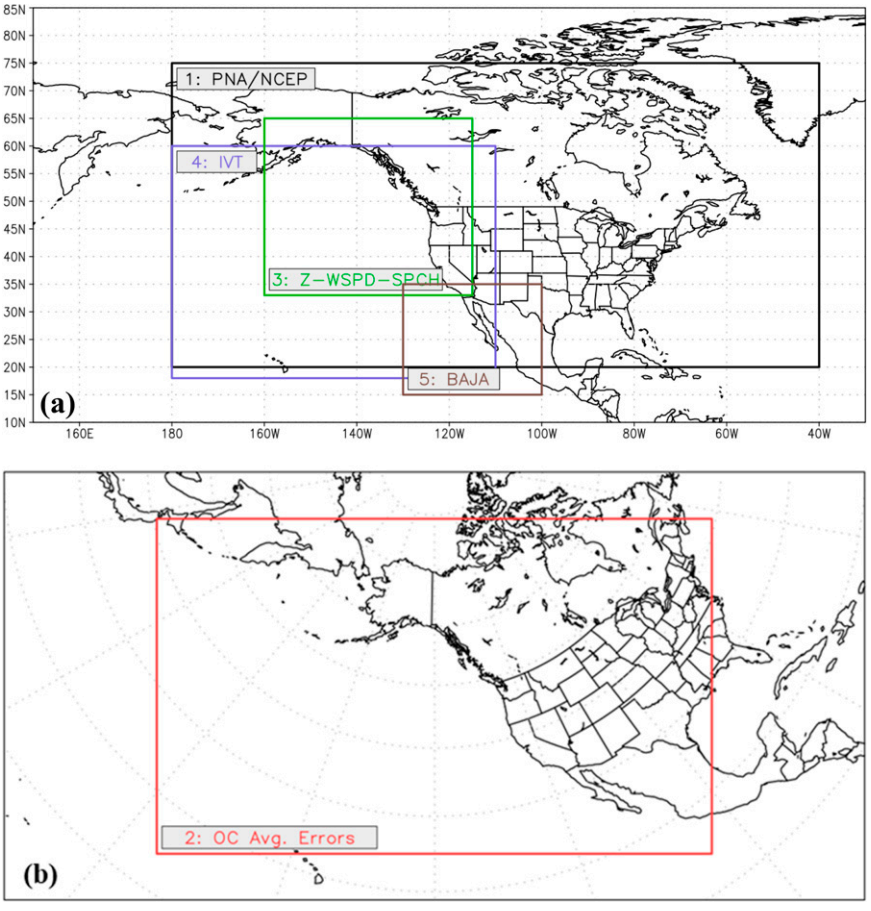

FIG. 1. (a) Spatial domains 1, 3, 4, and 5 for the NCEP statistical, model variable (Z, WSPD, SPCH) and IVT verifications, respectively (see Table 1), and (b) domain 2 for the large-scale verification for PMSL and Z over the AR OC.

at 0000 and 1200 UTC daily through 0000 UTC 11 March 2020, the last IOP date. Output data consisted of global analyses and forecast fields on a 1° × 1° latitude–longitude grid. Similar 1° × 1° gridded data from the operational ECMWF ("ECMO") system was also used to provide independent verifying analyses for calculating forecast errors. Choice of the ECMO verification is discussed in Lord et al. (2022a).

As discussed in Part I, the 17 IOP cases represent approximately 30% of the total cases (55) at 0000 UTC over the experimental period. Thus, statistical mean values over all verifications can dilute the impact of the IOP cases alone. The approach taken here is to examine mean statistics over all 55 initial conditions, but to also examine impacts over the 17 IOP cases separately and in more detail. Similarly, for geographical extent, the observation impacts over North America are likely to be small, having been diluted by a large areal extent compared to the area covered by and impacted by the observations. Again, the approach taken here is to describe impact over North America briefly, but to focus on the impact over a much smaller domain, the northeast Pacific and U.S. West Coast, which is also most consistent with the goals of the AR Program.

CTRL and DENY were evaluated graphically and statistically by their error patterns with respect to ECMO verifying analyses at 0000 and 1200 UTC, as appropriate, throughout the experimental period. Statistical evaluations consisted of a subset of NCEP's standard performance scores and mean absolute error (MAE) statistics for selected model variables such as mean sea level pressure (abbreviated as PMSL), geopotential height (Z), wind speed (WSPD), and specific humidity (SPCH) as well as the IVT, a fundamental derived quantity characterizing the ARs as the primary atmospheric feature leading to major precipitation events. The IVT (kg m^{-1} s^{-1}) is defined as

$$\mathrm{IVT} = \mathrm{Mag}\left\{\left[\int_{\mathrm{sfc}}^{225}(uq)dp/g\right], \left[\int_{\mathrm{sfc}}^{225}(vq)dp/g\right]\right\}, \qquad (1)$$

where sfc is the surface pressure; 225 hPa is the upper integration limit; Mag is the magnitude operator; u and v are the horizontal zonal and meridional velocities (m s^{-1}), respectively; q is the specific humidity (g g^{-1}); and g is the

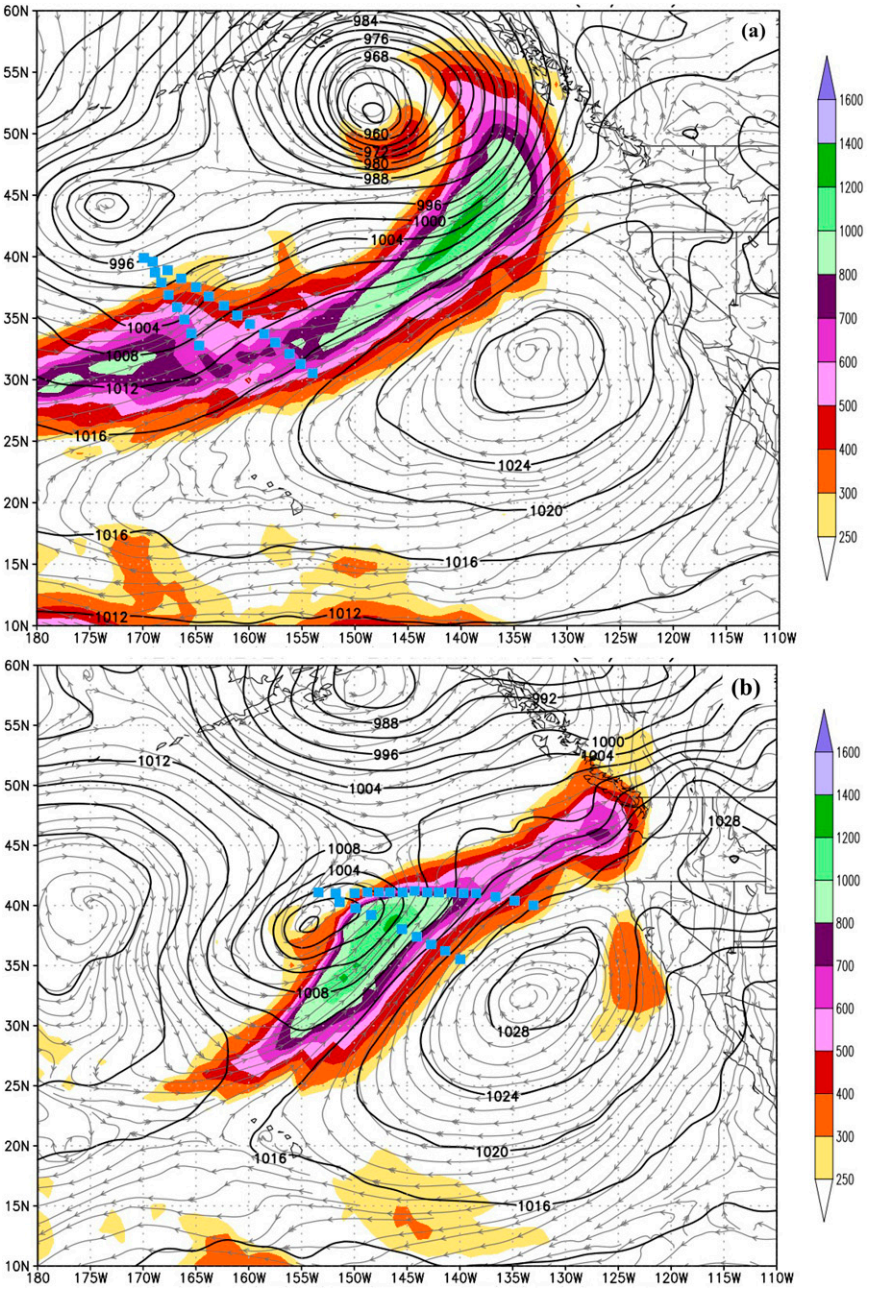

FIG. 2. Magnitude of the CTRL vertically integrated specific humidity flux (kg m^{-1} s^{-1}) for (a) 0000 UTC 29 Jan and (b) 0000 UTC 31 Jan 2020. Contours of mean sea level pressure (hPa) and 850-hPa streamlines are included. The sounding locations for 27 (24) dropsondes from 2 (1) AFRES aircraft for the 0000 UTC data assimilation cycle on 29 (31) Jan are shown in blue.

acceleration of gravity. Since the IVT depends on integrated wind and q values, but is episodic in nature, separate evaluations of the forecast model variables in regions of large IVT can be useful in diagnosing the source(s) of IVT errors. Table 1 summarizes the variety of statistical results presented in this paper that are calculated over appropriate verifying domains (see also Fig. 1) and representative vertical levels.

Since dropsonde data are ingested into a continuously cycled data assimilation system, typically with 4 cycles per day, impact can occur at the cycle with assimilated AR observations due to an improved initial atmospheric state at the assimilation time, or it may occur at a later data assimilation cycle due to spread of observed information throughout the model domain as the continuous cycling operates over time, even though there are no new AR observations assimilated.

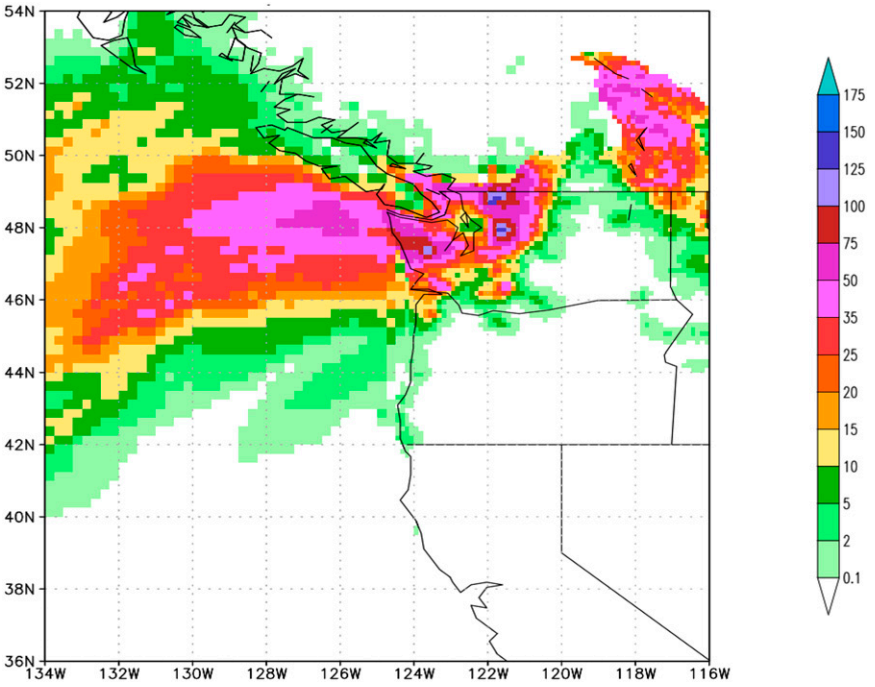

FIG. 3. The 24-h accumulated observed precipitation (mm) ending at 1200 UTC 1 Feb. Over land precipitation is gauge based, using the Climatologically Calibrated Precipitation Analysis (CCPA) version 4 technique (Hou et al. 2014). Over the ocean, it is satellite estimated, using the Climate Prediction Center morphing technique (CMORPH; Joyce et al. 2004). See Part I for details.

Due to the continuous nature of a cycled data assimilation system, it is likely that both impacts may be realized, but "direct impact" from dropsonde data at the initialization (analysis) time is more easily identifiable than "indirect impact" from previously assimilated data.

3. The AR2020 observing campaign and IOPs

Each IOP dropsonde deployment was focused on providing supplementary observations for potential positive impact on forecasts of major precipitation events and their associated IVTs as they made landfall over the U.S. West Coast and Canada. The IOPs were initiated based on satellite-based evidence of an existing AR, real-time, operational forecasts predicting their future impacts, and sensitivity analyses as described in section 1. Part I provides a general description of the AR2020 OC deployment strategy and resources, and includes time series of average and maximum precipitation over selected, impacted domains; Table A1 from Part I is copied in appendix A here for convenience. All IOPs begin on the 0000 UTC cycle in year 2020 so that dates are identified by their calendar day and month (e.g., 5 February).

IOP-1 begins at 0000 UTC 24 January as described in Part I. At this time, a precipitation event is already impacting the Oregon–Washington coast with an associated IVT maximum and southwest flow along the coast. IOP-2 begins on 29 January with an IVT maximum stretching from the southeast Alaska coast to the west of the date line, two low pressure centers at 52°N, 150°W and 44°N, 174°W to the north, and a high pressure center to the southeast at 32°N, 134°W (Fig. 2a). Deployed dropsondes sampled cross sections normal to the lower tropospheric flow. Two days later, on 31 January (IOP-3), AFRES flights again sampled this AR (Fig. 2b), which continues to propagate northeast; the leading edge influences the Oregon, Washington, and northern California area and produces peak precipitation at 1200 UTC 1 February (Fig. 3). Further details on these and remaining IOPs are given in Lord et al. (2022a).

4. Results

a. Large-scale mean atmospheric conditions and errors

The time-mean ECMO PMSL analysis over the experimental period and an area encompassing the North Pacific Ocean and most of North America (Fig. 4a) shows a low pressure area centered west of the Aleutians and a Z500 minimum centered over Kamchatka (Fig. 5a). The midlevel trough extends eastward over the northeast Pacific. In the Gulf of Alaska, a secondary PMSL low and collocated Z500 trough are just south of the Alaskan coast. Further downstream are a weak ridge over the Canadian Yukon and British Columbia (BC), and a trough over Hudson's Bay extending south through the U.S. Great Lakes.

An important feature of the mean analyses is the low pressure over Kamchatka and the Bering Sea. Climatologically, in Northern Hemisphere winter, this area is a major growth area

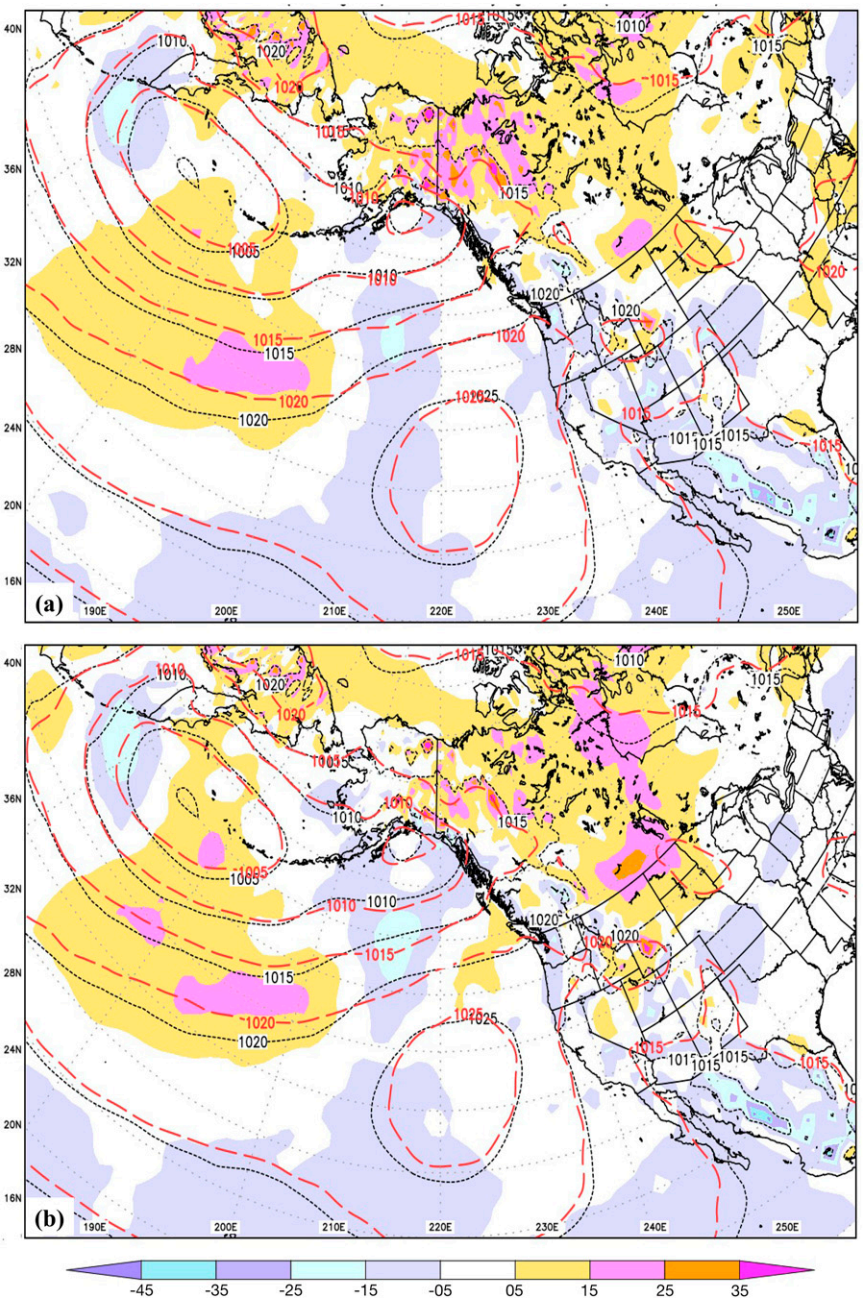

FIG. 4. Time-mean mean sea level pressure (hPa) over the period 0000 UTC 24 Jan 2020–0000 UTC 18 Mar 2020 for the ECMO verifying analysis (black contours) and the (a) CTRL and (b) DENY experiments (red contours). The mean 120-h forecast errors (CTRL-ECMO, DENY-ECMO; m) are in solid colors.

for baroclinic storms. The westward tilt with height of the Bering Sea low pressure center is consistent with the area's frequent cyclogenesis. Cyclones born over the AR2020 experimental period are critical to the formation of ARs impinging on the U.S. and Canadian west coast. This winter is, therefore, a highly representative period for investigating ARs and for examining the impact of additional observations to reduce analysis forecast errors in downstream AR landfalls over the United States.

The 120-h CTRL MSLP mean forecast error with respect to the ECMO verifying analysis (Fig. 4a) shows a maximum southeast of the surface low; for Z500 (Fig. 5a) the prominent minimum error is south of the Aleutians at the eastern extent of the mean trough. The DENY experiment (Figs. 4b and 5b) shows very similar patterns over the oceanic domain, with the largest errors shifted slightly poleward compared to CTRL. DENY also has a somewhat larger Z500 error over

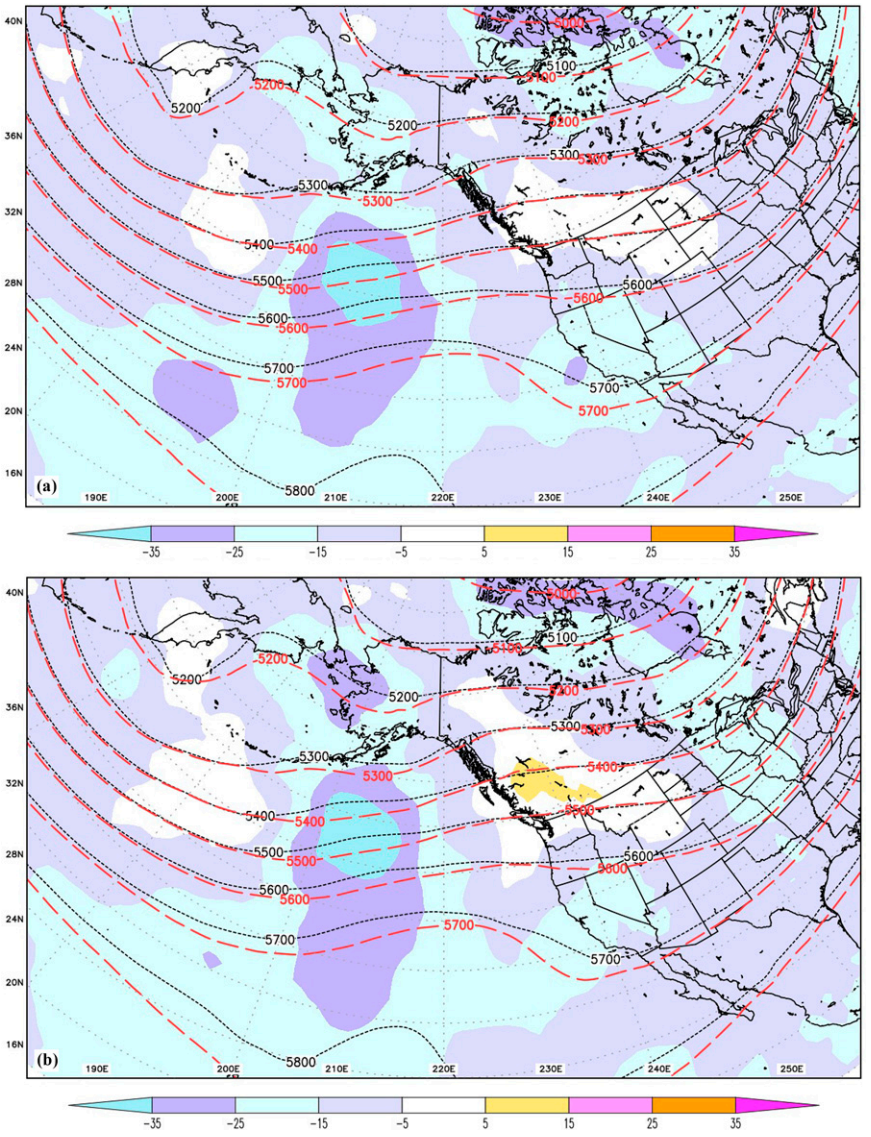

FIG. 5. Time-mean 500-hPa geopotential height fields over the period 0000 UTC 24 Jan 2020–0000 UTC 18 Mar 2020 for the ECMO verifying analysis (black contours) and the (a) CTRL and (b) DENY experiments (red contours). The mean 120-h forecast errors (CTRL-ECMO, DENY-ECMO) are in solid colors.

southwestern Canada. The GFS, therefore, exhibits a notable tendency to propagate troughs too fast to the east with respect to the verification; AR2020 observations do not mitigate this error significantly.

The majority of large-scale, time-mean statistics confirm the lack of significant impact over the North American continent and surrounding oceanic areas (Fig. 6). Mean anomaly correlations for model winds, geopotential height and temperature improved very slightly in CTRL, except for meridional wind, which is 1.2% improved in CTRL. The largest 120-h CTRL improvement is valid at 0000 UTC 5 March and initialized at 0000 UTC 29 February (Fig. 6), which was 5 days after the nearest IOP. Statistical significance (95% confidence level) was tested using the Student's *t* test (Wilks 2006; Lindgren and McElrath 1959) for each of the variables in Fig. 6; all differences were statistically insignificant over the 48 experimental cases.

b. *The 0000 UTC 29 February case*

For reasons discussed above (and in Part I), the 29 February case is clearly an indirect improvement attributable to earlier assimilation of dropsonde data and the nonlinear propagation of observational information through the data assimilation system to future initial atmospheric states. The CTRL forecast error growth (Fig. 7) is close to that of ECMO out to 72 h, but increases at a higher rate than ECMO beyond 72 h. CTRL error patterns closely resemble DENY out to 96 h, after which the DENY errors increase much more rapidly and

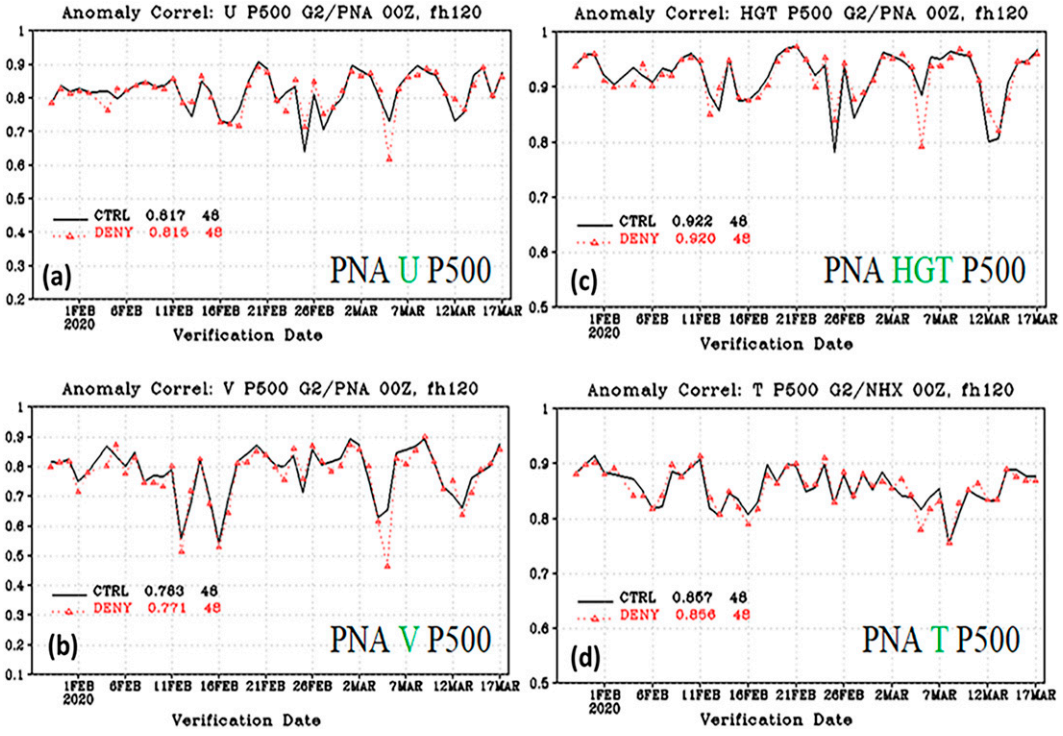

FIG. 6. Anomaly correlation time series over the experimental period and area 20°–75°N, 180°–320°E for 120-h CTRL (black) and DENY (red) forecasts at 500 hPa: (a) zonal wind, (b) meridional wind, (c) 500-hPa geopotential height, and (d) temperature. There were 48 cases in the time series verifying at 0000 UTC.

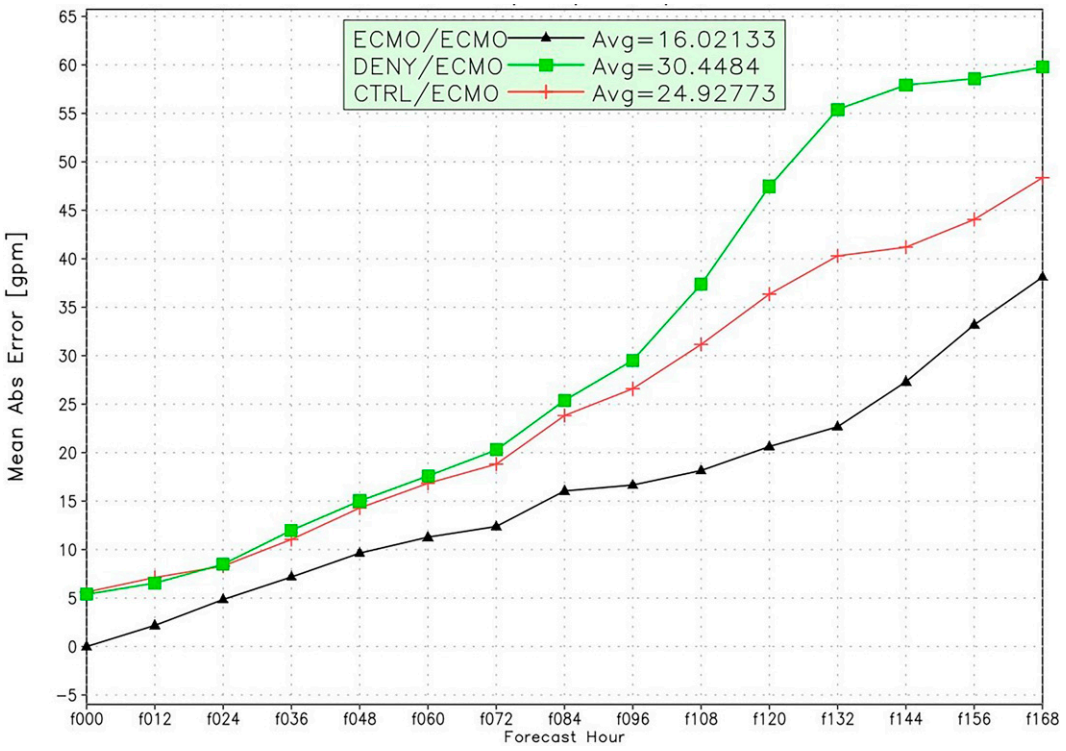

FIG. 7. Z500 MAE for the analysis and 12–168-h CTRL, DENY, and ECMO forecasts initialized at 0000 UTC 29 Feb 2020. All MAE are calculated with respect to the verifying ECMO analysis. Verification is over domain 1 (Fig. 1 and Table 1). The CTRL and DENY analysis MAE results from a Z500 bias between the GFS and ECMO geopotential height fields. The average MAE over the analysis and all forecasts is shown in the legend.

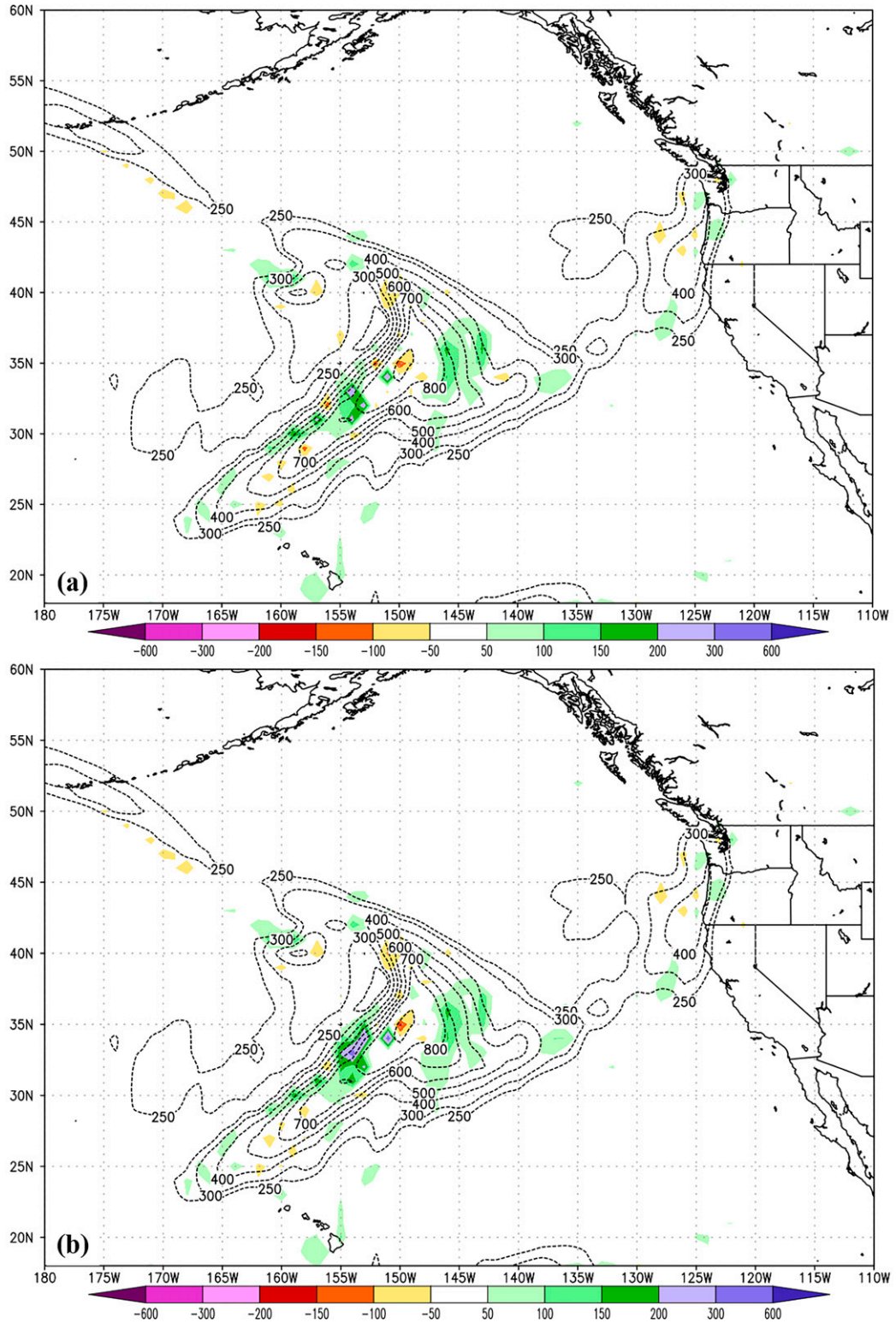

FIG. 8. (a) DENY-ECMO and (b) CTRL-ECMO IVT analysis differences (shaded) and ECMO IVT analysis contours (black) for 0000 UTC 24 Jan 2020.

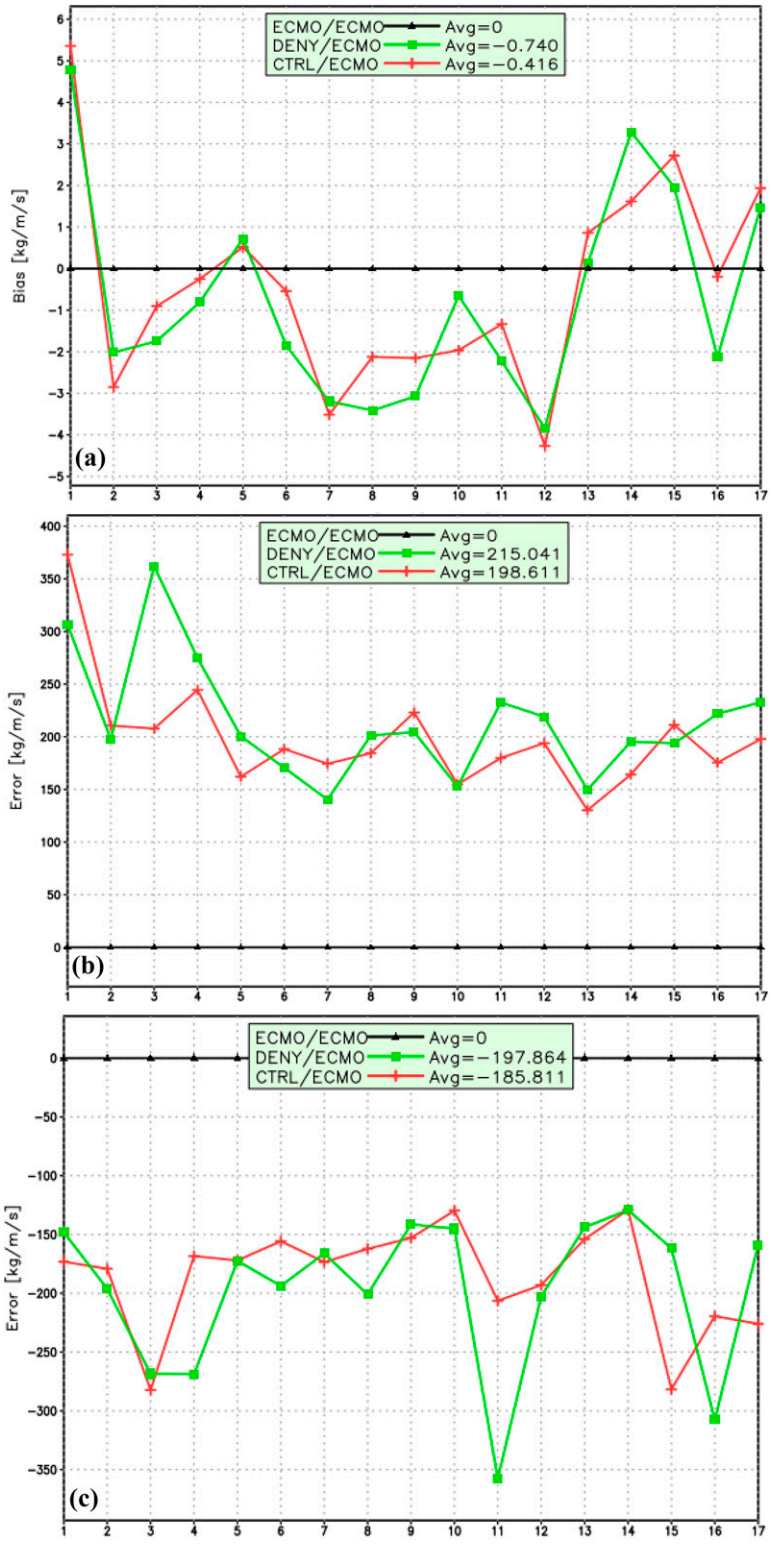

FIG. 9. (a) IVT area-mean and domain-wide (b) maximum and (c) minimum IVT analysis differences (kg m^{-1} s^{-1}) between DENY (green) and CTRL (red) and the verifying ECMO analysis (D-E and C-E, respectively) over the verification domain 4 (Table 1, Fig. 1) for each IOP. The index along the abscissa corresponds to the IOP number (Table A1). Average D-E and C-E values over the 17 IOPs are given in the legend. Neither of these analysis differences is statistically significant at the 80% level.

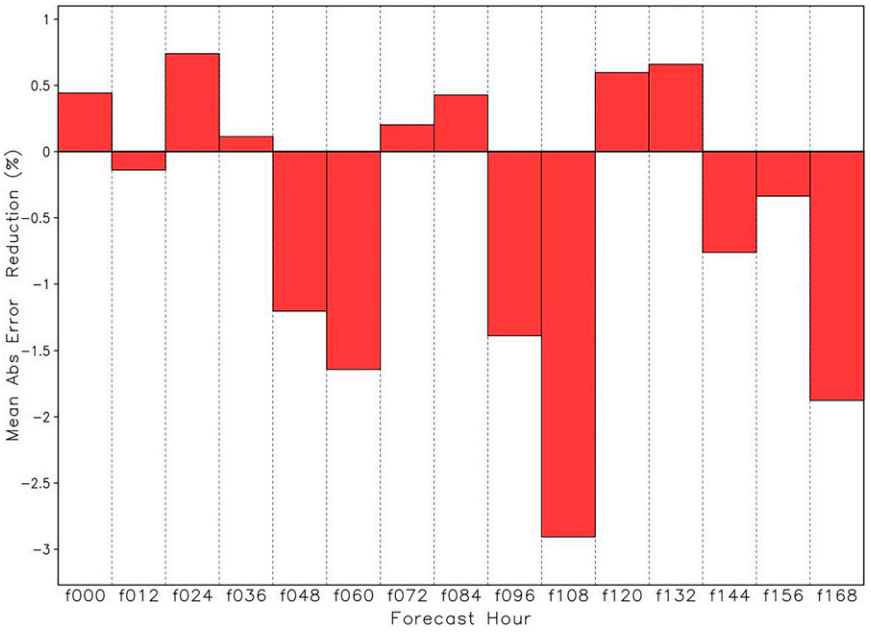

FIG. 10. Average IVT MAE reduction (%) over 17 IOP cases for analyses and 12–168-h forecasts over domain 4 (Table 1). The mean reduction over all forecasts and IOPs is −0.47%.

error patterns deviate substantially from those of both CTRL and ECMO over the central and northeastern United States and eastern Canada (appendix B). At 120 h and beyond, DENY forecasts a ridge development over the northeastern United States and Nova Scotia instead of a closed low pressure system as indicated in the verifying analysis, resulting in a large anomaly correlation decrease for DENY. More details of the error evolution in this case are presented in appendix B.

c. *Forecast impacts*

The direct positive impact on forecast variables is, of course, the purpose of the AR2020 OC. Therefore, to focus on these potential impacts, statistics have been generated for each of the 17 forecasts initialized by AR2020 OC IOP soundings over an appropriate domain (Table 1). For each forecast hour, and for both CTRL and DENY experiments, the statistics are averaged over all 17 IOP cases. Differences between averages are calculated as a percent impact:

$$I(v) = 100 \times (|S_{\mathrm{DENY}}| - |S_{\mathrm{CTRL}}|)/|S_{\mathrm{DENY}}|, \qquad (2)$$

where v is a variable (IVT, PMSL, Z, WSPD, or SPCH), S_X is a statistic [bias, MAE, min(v), max(v), etc], and X = CTRL/DENY are the experimental runs. An $I > 0$ reflects an improvement of CTRL over DENY while $I < 0$ is a degradation due to assimilating dropsonde soundings. Averaging I over all forecast hours provides an estimate of the overall impact for each IOP.

1) IVT IMPACTS

Since the IVT is a strong meteorological feature amid a lower amplitude environment, domain-wide statistics over the area encompassing one or more IVT maxima can evaluate both analysis differences and forecast quality. IVTs propagate through the verification area during the entire OC so that both direct and indirect impacts are measured by evaluating the IVT errors in CTRL and DENY analyses and forecasts over all IOPs. Since the IVT is an isolated feature, verifications should include only cases in which IVTs are the prominent feature in the verification area, and hence are performed only for IOP dates.

IVT verification statistics [Eq. (2)] were generated for 12–168-h forecasts from initial conditions on each IOP date (Table A1). Impacts are measured by the percent improvement (I) of the CTRL MAE relative to the DENY MAE over domain 4 (Table 1). For analyses, I measures the difference between the CTRL/DENY analysis and the ECMO analysis. To measure overall impact, I is averaged over all IOPs for the analysis and each forecast hour. To measure the impact of each IOP, I is averaged over the analysis and all forecasts initialized from that IOP.

(i) IVT analyses

DENY and CTRL IVT analyses are compared to ECMO IVTs for all 17 IOPs. Both CTRL and ECMO analyses assimilated OC dropsonde data while DENY analyses did not. For the first IOP, 0000 UTC 24 January (Fig. 8a), the DENY analysis has the general shape and magnitude of the ECMO IVT; the DENY-ECMO (D-E) differences are scattered and generally <20% of the IVT magnitude on the leading edge (35°N, 145°W) and west-central (W-C, 33°N, 155°W) sectors. The CTRL IVT (Fig. 8b) shows the same pattern of C-E differences with the ECMO but the W-C sector has larger differences. Isolated maxima ("bull's-eyes") are present in both D-E and C-E differences and result from a more variable GFS moisture distribution than for the ECMO (Lord et al. 2022a).

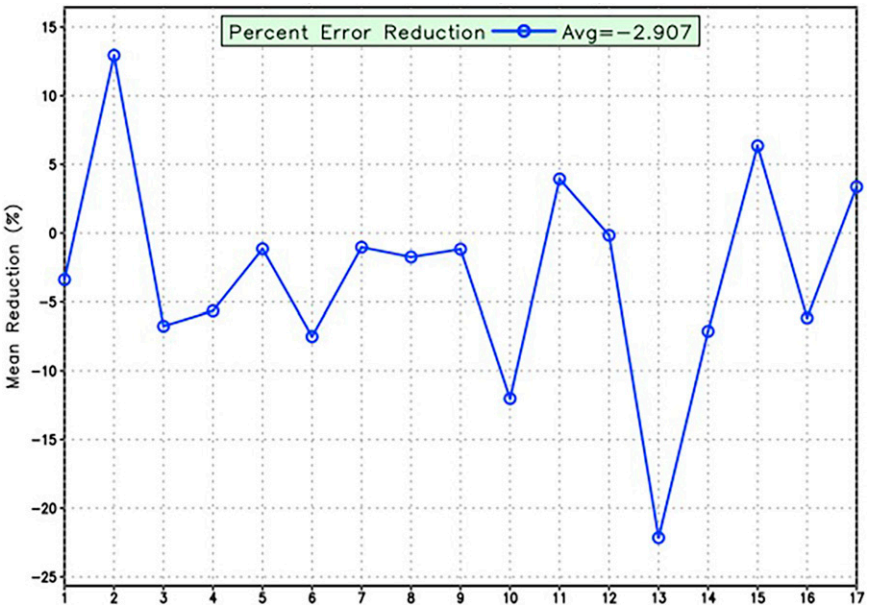

FIG. 11. IVT MAE reduction for each IOP and 108-h forecasts. Abscissa labels are as in Fig. 9.

IVT analysis differences at some other times also contain bull's-eyes (not shown).

Overall, there are small, but positive, impacts from the assimilated dropsonde observations on the CTRL IVT structures. The domain-wide mean analysis difference (Fig. 9a) is reduced by 43.8% (from −0.740 to −0.416 kg m^{-1} s^{-1}, i.e., closer to the ECMO IVT mean structure). Maximum C-E values are reduced compared to D-E in 10 IOPs and the IOP average maximum value is decreased by 7.6% (Fig. 9b), while the domain-wide mean absolute difference is improved by 0.5% (not shown). Last, domain-wide minimum C-E values (compared to D-E) are reduced in a minority of the IOPs (7 of the 17) and the IOP-averaged (C-E) is 6% smaller (Fig. 9c).

(ii) IVT forecasts

The IOP-averaged IVT MAE reduction (Fig. 10) is positive for the analyses (0.44%) and 24-h forecasts (0.74%), while impacts for analyses, 12- and 36-h forecasts are less than 0.15% magnitude. Longer forecasts (48–168 h) are mostly degraded, with the largest values at 108 and 168 h. The major

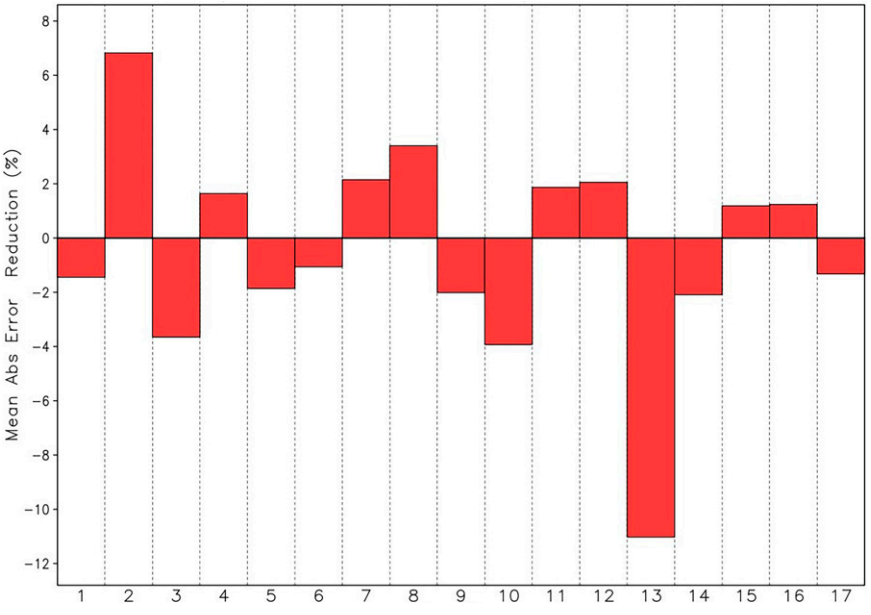

FIG. 12. Average IVT MAE reduction (%) over 0–168-h forecasts for each of the 17 IOP cases.

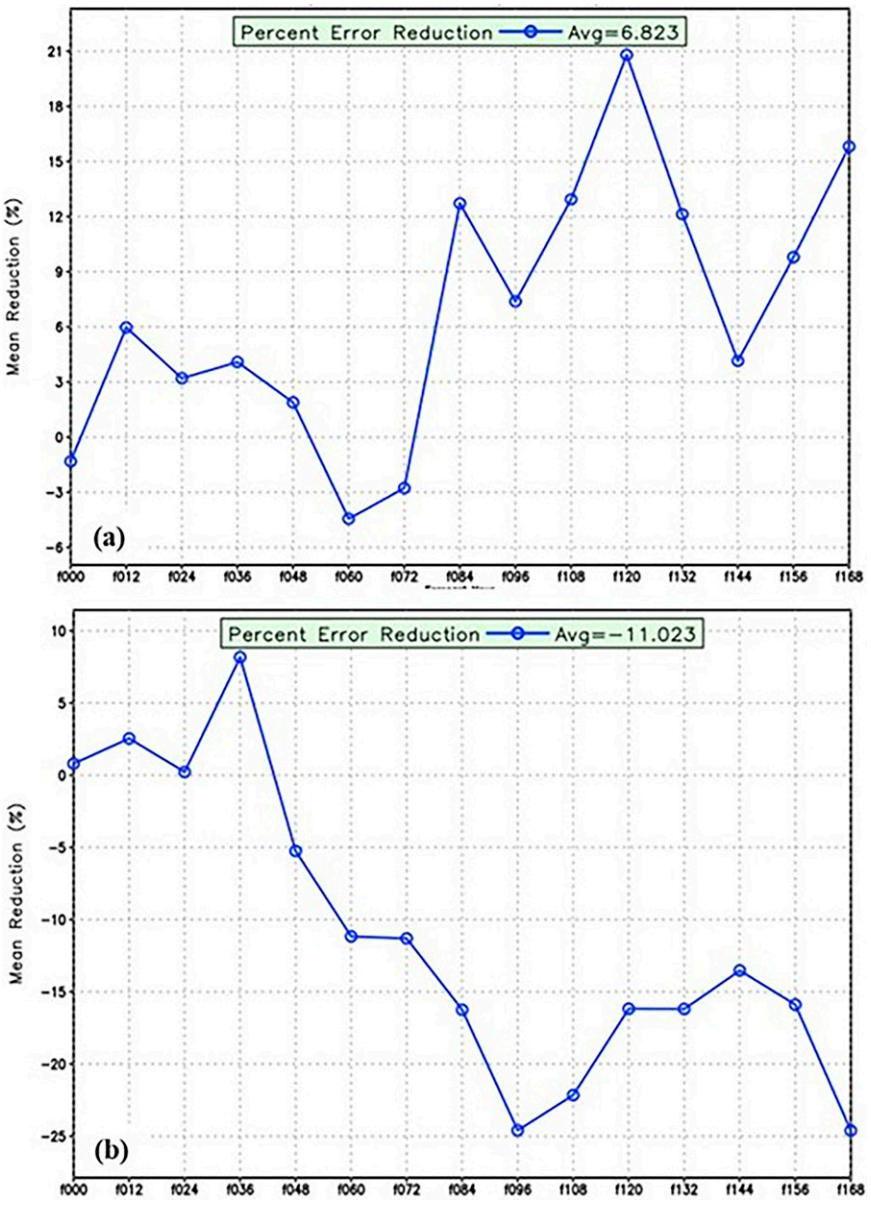

FIG. 13. IVT MAE percent reduction for 0–168-h forecasts and (a) IOP-2, initialized 0000 UTC 29 Jan and (b) IOP-13, initialized 0000 UTC 7 Mar.

contributor to degraded forecasts at 108 h is IOP-13 (0000 UTC 7 March, Fig. 11) with a mean error of −2.9% which is statistically significant, but at a low confidence level (80%) due to the large error standard deviation across the IOPs for longer forecasts. For each IOP, the average percent error reductions over all analysis and forecast hours are positive for 8 IOPS and negative for 9 IOPs (Fig. 12); the mean impact over all IOPs and all forecast hours is −0.47%. The largest improvement is 6.8% for IOP-2 (0000 UTC 29 January) and the largest degradation is −11.0% for IOP-13 (0000 UTC 7 March), both of which are statistically significant at the 95% level as are the mean forecast improvements for IOPs 1 and 8. For IOP-2 (Fig. 13a), there are consistent positive impacts both early in the forecast (12–48 h) and later (84–168 h), mostly exceeding 6%. The early impacts are for the observed IVT making landfall while the later impacts are for a new IVT entering the verification domain from the west. For IOP-13 (Fig. 13b), the average reduction is dominated by degradations exceeding 13% from forecasts longer than 72 h, even though there are consistent improvements through 36 h. More details on WSPD, *Z*, and SPCH errors that are forcing IVT errors, including statistical significance of these results, are given in section 2 below and in Lord et al. (2022a).

(iii) IVT impact summary

To summarize all analysis and forecast impacts, *I* is categorized by assigning an integer to a range of impact magnitude

(a)

Impact (%)	Category	Color
I < -17.5	-4	
-17.5 <= I < -10	-3	
-10.0 <= I < -5.0	-2	
-5.0 <= I < -0.5	-1	
-0.5 <= I <= 0.5	0	
0.5 < I <= 5.0	1	
5.0 < I <= 10.0	2	
10.0 < I <= 17.5	3	
I > 17.5	4	

(b)

IOP	Initial Date	anal	12	24	36	48	60	72	84	96	108	120	132	144	156	168
1	00Z24JAN2020	-1	1	-1	1	-1	0	0	1	-1	-1	1	0	-2	-1	-2
2	00Z29JAN2020	-1	2	1	1	1	-1	-1	3	2	3	4	3	1	2	3
3	00Z31JAN2020	2	1	-1	-2	-1	-1	2	2	0	-2	-4	-4	-4	1	2
4	00Z04FEB2020	1	1	1	1	0	-2	1	1	-2	-2	-1	1	3	3	2
5	00Z05FEB2020	-1	-1	-1	-2	-3	-2	-2	-2	-1	-1	2	3	2	2	-1
6	00Z06FEB2020	-1	0	-1	-1	-2	-2	1	-1	-1	-2	-1	0	1	2	2
7	00Z14FEB2020	-1	0	1	-1	-1	-2	-1	2	2	-1	2	2	2	2	1
8	00Z15FEB2020	1	1	1	1	1	2	3	3	2	-1	2	2	-2	-1	-1
9	00Z16FEB2020	-1	0	1	1	2	2	1	1	2	-1	-2	-2	-2	-3	-4
10	00Z21FEB2020	0	-1	1	2	2	2	1	-1	-1	-3	-3	-2	-3	-4	-2
11	00Z24FEB2020	1	1	2	3	2	1	2	2	2	1	-1	-2	-2	-1	-2
12	00Z02MAR2020	0	-1	-2	-2	0	-1	-1	-1	-1	0	2	3	4	2	1
13	00Z07MAR2020	1	1	0	2	-2	-3	-3	-3	-4	-4	-3	-3	-3	-3	-4
14	00Z08MAR2020	1	0	1	-3	-3	-3	-3	-2	-2	-2	2	3	2	2	1
15	00Z09MAR2020	0	-2	-1	-2	-1	-1	0	0	1	2	2	2	1	2	2
16	00Z10MAR2020	2	-1	-1	2	2	3	2	1	-1	-2	0	0	1	-2	-2
17	00Z11MAR2020	-2	-1	2	1	0	1	0	-1	-2	1	2	1	-2	-3	-1

(c)

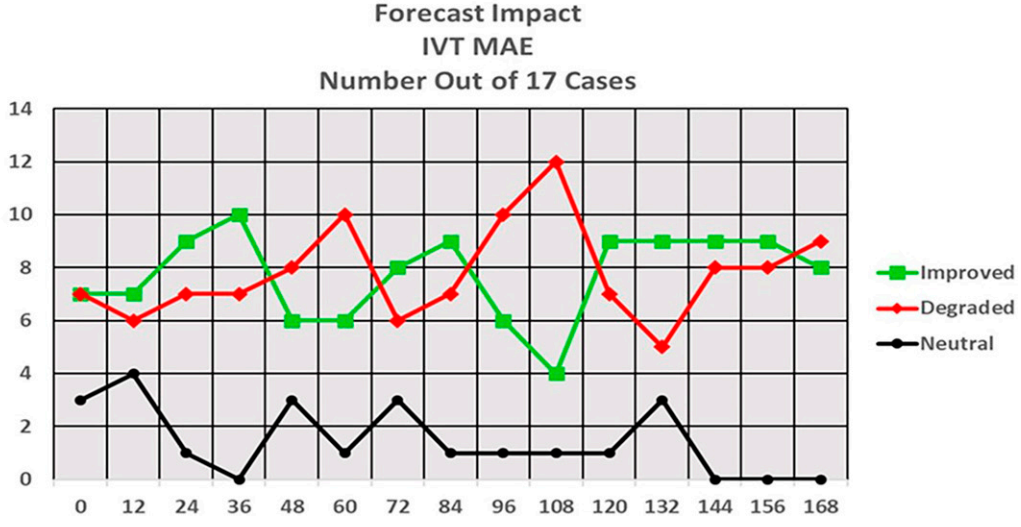

FIG. 14. (a) IVT MAE impact (*I*; %) categories for CTRL error improvement relative to DENY, (b) categorical summary for each IOP and forecast hour (analysis and 12–168 h), and (c) number of IOPs, out of 17, improved (green), degraded (red), or with neutral impact (black). Positive categories indicate improvement, negative are degradations, and zero (magnitude of 0.5% or less) is neutral.

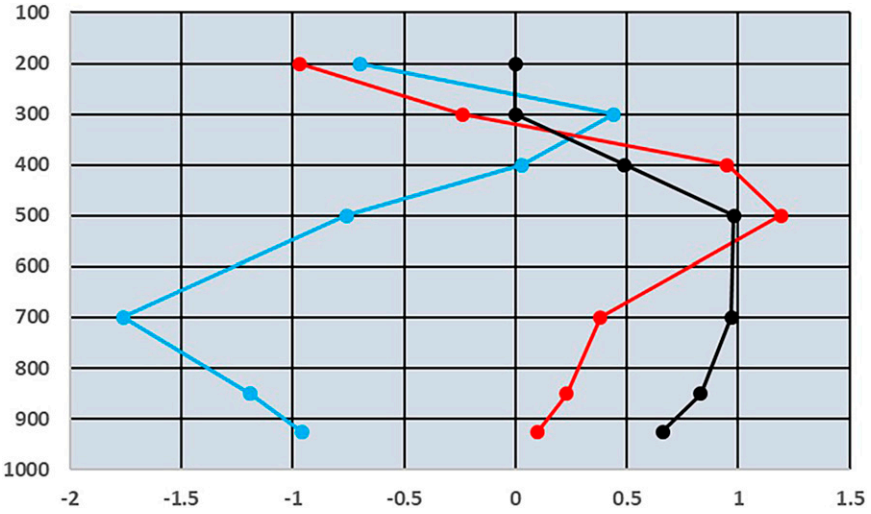

FIG. 15. Vertical profile of MAE improvement (%) averaged over all analyses, forecast hours and IOPs for geopotential height (blue), wind speed (red), and specific humidity (black). Specific humidity improvements are zeroed out at 200–300 hPa due to insufficient precision in their calculation. Results for Z700, SPCH500, SPCH700, and SPCH850 are statistically significant at the 90% level and WSPD500 is significant at the 95% level.

(Fig. 14a). Over the complete set of IOPs and forecast ranges, the categorized I (Fig. 14b) varies over the full range ($-4 \le I \le 4$), generally without strong coherence from one IOP to the next or from one forecast projection to smaller or larger projections. For 36–60-h forecasts, there is a weak trend from small negative impact to moderate positive impact for consecutive IOPs (14–16 February) and from negative impact to moderate positive impact (7–10 March). However, for consecutive IOPs on 4–6 February, almost all impacts are negative for 12–108 h. Further details, together with statistical significance determinations,

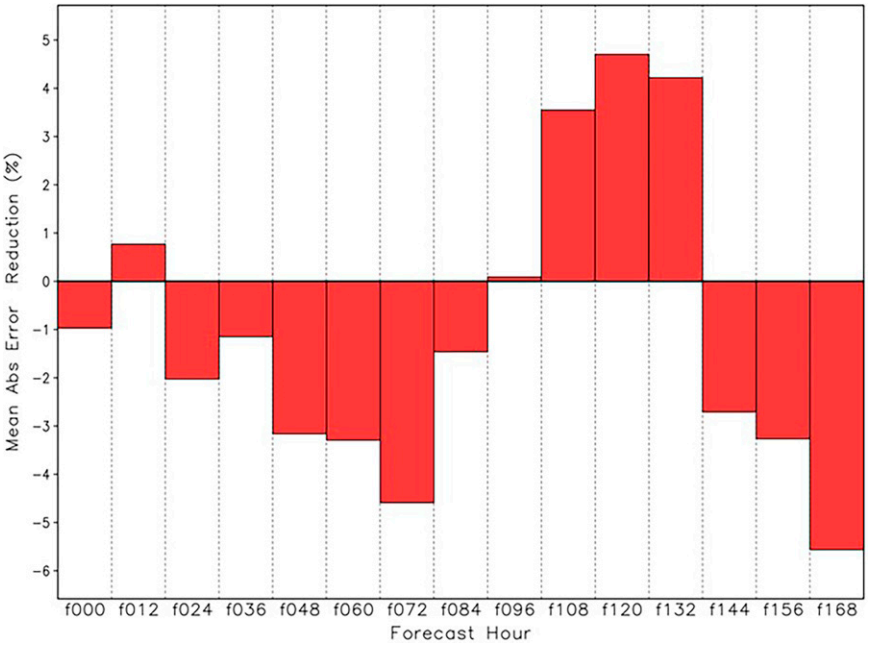

FIG. 16. Average MAE reduction (%) over 17 IOP cases for the analysis and 12–168-h PMSL forecasts, verified against ECMO analyses. Positive reductions indicate improvement for CTRL relative to DENY. The verification is over domain 3 (Table 1). Improvements/degradations have low statistical significance (80%) level for 24, 60, 72, and 120 forecast hours. Figure 17 summarizes statistical significance for all forecast variables and forecast hours.

	Level	f000	f012	f024	f036	f048	f060	f072	f084	f096	f108	f120	f132	f144	f156	f168	NS+	NS-
Z	200			+	+			-									2	1
	500	+					-	-									1	2
	700			-		-	-										0	3
	850			-			-	-				+					1	3
WSPD	200			+			-		-	-							1	3
	300	+					-				+	-					2	2
	400	+	+									+					3	0
	500	+		+													2	0
	700	+	+	+	+												4	0
	850		+		+												2	0
	925																	
SPCH	700	+		+													2	0
	850																	
	925		-											+			1	1
PMSL	Sfc			-			-	-				+					1	3

FIG. 17. Summarized statistical significance for CTRL-DENY, IOP-averaged, analysis and 12–168-h forecast differences (improvements/degradations) for *Z*, WSPD, and SPCH at various levels and PMSL (Figs. 16 and 18–20). Significance at the 95% (dark green), 90% (light green), and 80% (yellow) levels are indicated for each variable, level, and forecast length. The plus and minus signs in each box indicate statistically significant improvement and degradation, respectively, and the numbers of each (NS+ and NS−) are totaled in the two right-hand columns.

are presented below in the context of forecast impacts from all variables (IVT, *Z*, WSPD, and SPCH).

From the analysis to 108-h forecasts, there is an expected downward trend in the number of improved IOP cases and a corresponding positive trend in the number of degraded cases (Fig. 14c). Also expected is the number of neutral cases dwindling to zero beyond 132 h. The number of improved IVT cases beyond 108 h is unexpected, however, as is the corresponding improvement to both wind speed and moisture forecasts over the same forecast period as shown later.

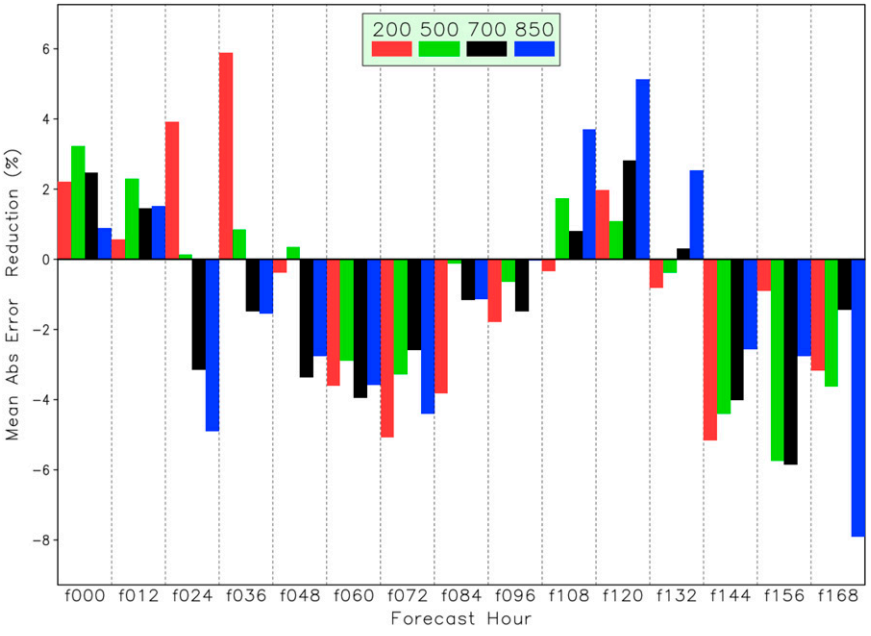

FIG. 18. As in Fig. 16, but for geopotential height at 200, 500, 700, and 850 hPa. The statistical significance levels for Z200 at 24–36 h, Z850 at 24 h, and Z700 at 60 h is 90% (Fig. 17).

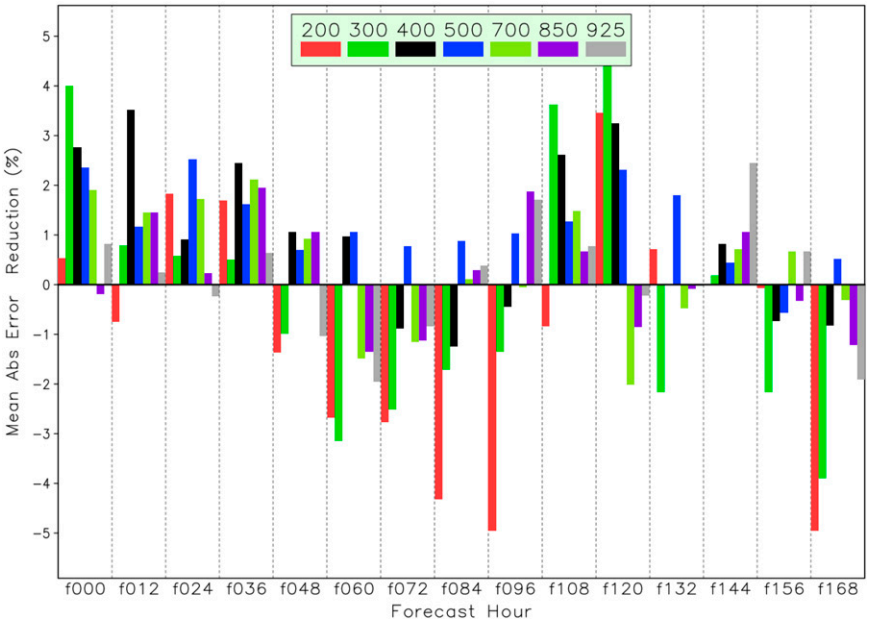

FIG. 19. As in Fig. 16, but for wind speed at 200, 300, 400, 500, 700, 850, and 925 hPa. Missing data (e.g., large values removed due to division by near-zero averages) are represented as a zero value. Analysis improvements are statistically significant at the 95% level for 300–700 hPa as are 12–24-h forecasts in the midtroposphere (Fig. 17).

For a particular IOP, as noted for IOP-2 above, a common, major meteorological feature over the AR OC period is a subsequent IVT system entering the verification domain, typically from the west or southwest. Of course, this system is not observed directly, but its performance can be impacted indirectly (positively or negatively), by previous observations through the GDAS analysis and cycling. Ideally, impacts would be positive, but practically, limited-area observations have the most chance of improving downstream, rather than upstream, forecasts. However, some of the impact at longer

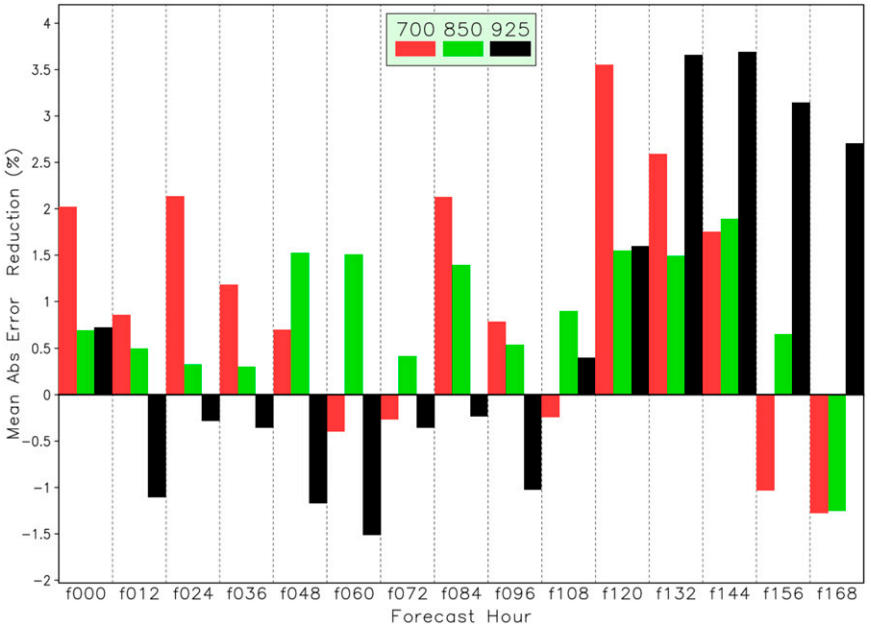

FIG. 20. As in Fig. 16, but for specific humidity at 700, 850, and 925 hPa. Most impacts lack statistical significance (Fig. 17).

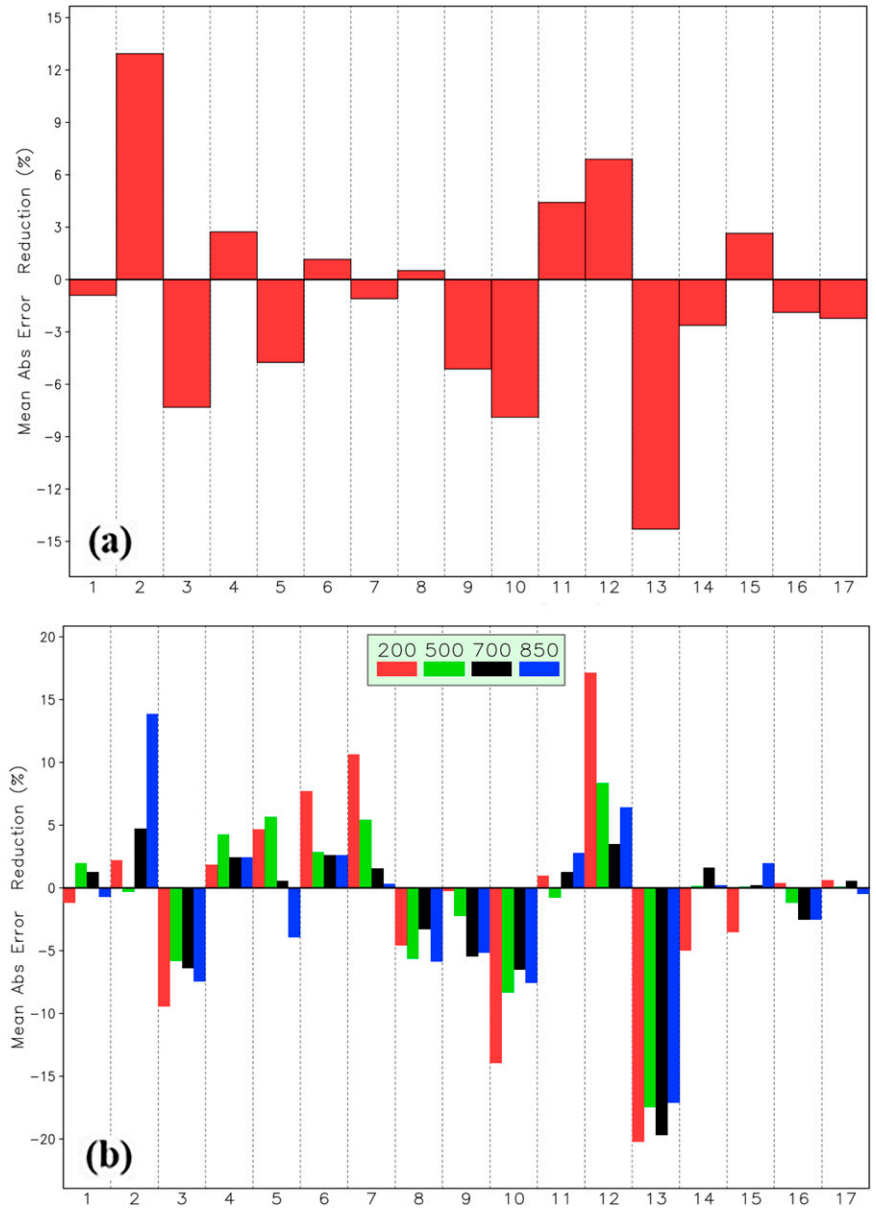

FIG. 21. As in Fig. 12, but for (a) mean sea level pressure and (b) geopotential height at 200, 500, 700, and 850 hPa. PMSL impacts are statistically significant at the 95% level for IOPs 2, 11, 13, and 17, and Z impacts are scattered among the IOPs but strongly negative for IOP-13.

forecast projections may be positive overall, such as the IVT MAE at 120–132 h (Fig. 10). This may also be the case for some or all of the model forecast variables themselves as noted below.

2) PMSL, Z, WSPD, AND SPCH IMPACTS

(i) IOP-averaged impacts

The overall MAE percent improvements for PMSL, averaged over all analyses, forecast hours and IOPs, and for Z, WSPD, and SPCH (additionally averaged over 200–925 hPa for Z and WSPD and over 400–925 hPa for SPCH) are −0.99%, −0.70%, 0.23%, and 0.79%, respectively. The degradations for PMSL and Z are consistent as expected, and the Z degradation is statistically significant from zero at the 90% level. The overall improvements for WSPD and SPCH are encouraging, but only SPCH is significantly different from zero at the 95% confidence level. These average percent improvements, however, show varied impact with height, with some levels showing statistical significance (Fig. 15). Degradations in Z dominate the mid- and lower troposphere, while WSPD is positive (0.1–1.2) at all levels up to 400 hPa and SPCH improvements are largest (~1%) up through the lower and midtroposphere before decreasing in the upper troposphere.

	Level	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Z	200			-			+	+			-		+	-		-		
	500	+		-		+	+	+					+	-				
	700	+	+	-			+							-				
	850		+	-			+					+	+	-				
WSPD	200		+	-			-						+	-				
	300	+	+		+			+						-			+	
	400	+	+	-	+				+				+	-			+	
	500		+		+				+				+	-			+	
	700		+	-			+	+	+		-		+	-		+	+	
	850	+	+	-	+	-	+		+	-			+	-		+		
	925		+		+	-	+		+	-	-		+	-		+	-	-
SPCH	700		+		+		+	-	+		-		+	-		-	+	
	850		+	-			+		+		-		+	-			+	+
	925	+	+		+			-	+		-			-				+
PMSL	Sfc		+	-						-	-	+	+	-		+		-
	NS+	6	13	0	7	1	9	4	8	0	0	2	12	15	0	4	6	2
	NS-	0	0	10	0	2	1	2	0	3	7	0	0	0	0	2	1	2

FIG. 22. As in Fig. 17, but for forecast-averaged impacts for each IOP, with data from Figs. 21 and 23. In the bottom rows, NS+ and NS− are the number of statistically improved and degraded forecasts for each IOP.

The IOP-averaged PMSL MAE for the analysis and 12–168-h forecasts, verified against ECMO fields (Fig. 16), shows less than 1% improvement at 12 h, 2%–5% CTRL degradation from 24 to 72 h, 3%–5% improvement at 108–132 h, and 3%–6% degradation from 144 to 168 h. While the forecast- and IOP-averaged PMSL degradation (−0.99%) itself is not statistically significant at the 80% level, impacts at some forecast hours (Fig. 17), discussed below, can still be significant.

All CTRL geopotential height analyses (Fig. 18) are closer to the verification than the DENY, and a similar pattern to PMSL forecasts is shown for all levels over the entire forecast range: CTRL MAE improvements for 12–36 h, degradations for 48–96 h, 1%–4% improvement at 120 h, and 1%–8% degradation at and beyond 132 h. Impacts at 200 hPa are generally the largest magnitude but lower and midtropospheric impact magnitudes are larger at 156–168 h. The magnitudes of IOP-averaged PMSL and mid-to-upper tropospheric Z error improvements over the entire forecast range are not highly correlated visually with IVT errors (Figs. 16, 17, and 10). Lower tropospheric Z error improvements, however, do have a positive correlation with IVT improvements for 24–120-h forecasts as noted in Part I.

Wind speed impacts (Fig. 19) are generally 1%–4% improvements of CTRL over DENY across all pressure levels out to 48 h. Beyond 48 h, impacts at 200–300 hPa have the largest magnitudes and dominate the negative impacts from 48 to 96 h. Average negative wind impacts for almost all levels at 60–96 h appear to precede negative impacts on IVT forecasts at 96–108 h, while positive wind impacts at 108–120 h tend to precede positive IVT forecast impacts at 120–132 h (Fig. 10).

Specific humidity impacts (Fig. 20) for 700, 850, and 925 hPa are also generally positive, except at 925 hPa for 12–96-h forecasts. These impacts are widespread: for all analyses over the 700–925-hPa layer, improved moisture MAE for CTRL versus DENY analyses occurs in 65% of the 17 IOPs, and positive forecast impacts occur in 53% of all 12–168-h forecasts in these layers. These positive forecast impacts range from 47% of all IOPs at 12–24 h to 63% at 120 and 144 h and are more consistent across all forecast ranges than for geopotential height and wind speed. Overall, and despite a lack of widespread statistical significance, they confirm that the GDAS responds positively throughout the AR2020 OC to the unique, accurate and high vertical resolution dropsonde specific humidity observations by improving the initial CTRL moisture analyses, which results in consistent positive value to the majority of CTRL moisture forecasts relative to the DENY experiment.

Correlations between IVT improvements and those for lower tropospheric Z, WSPD, and SPCH are presented and discussed further in Part I.

In terms of statistical significance, IOP-averaged improvements are scattered among the variables PMSL, Z, WSPD, and SPCH and various forecast hours, mostly for 0–36 h (Fig. 17).

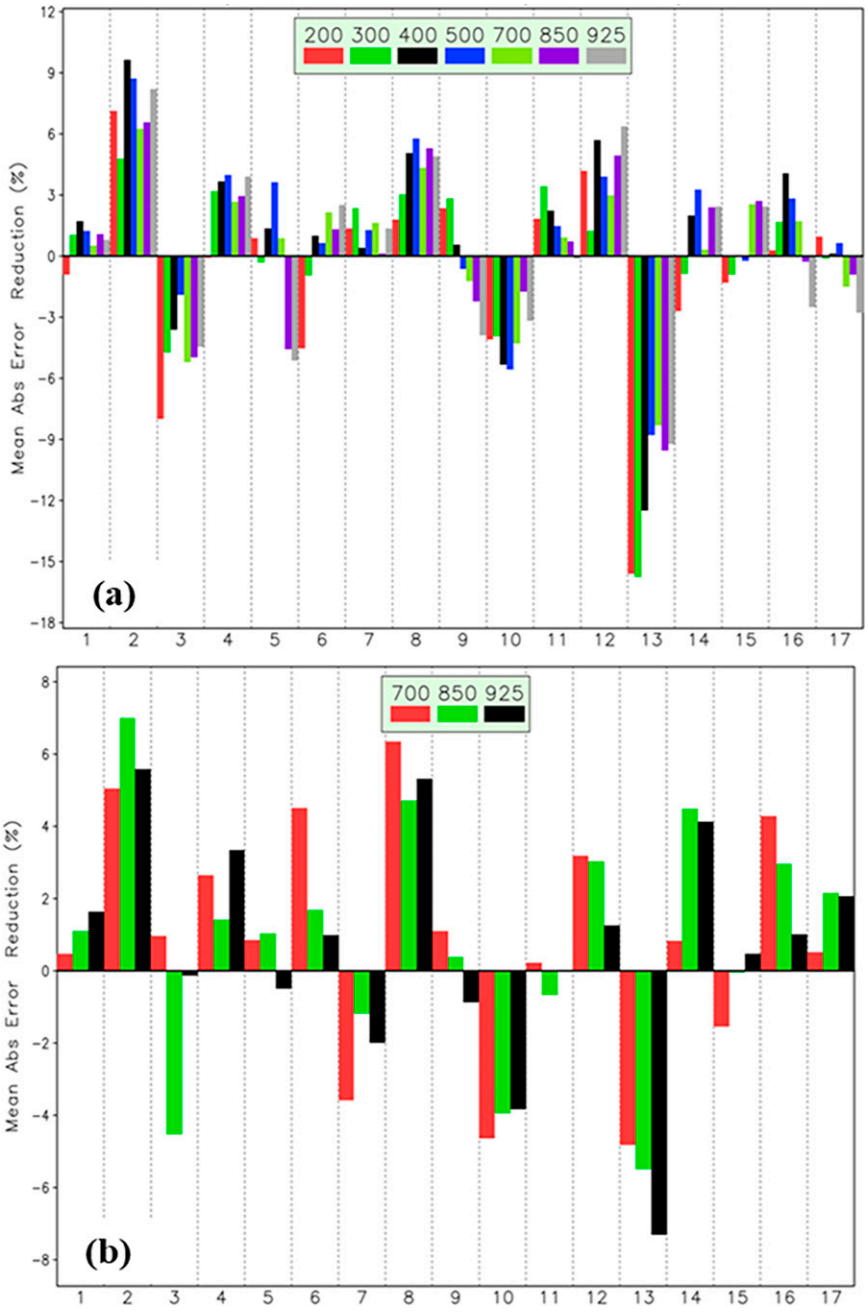

FIG. 23. As in Fig. 12, but for (a) tropospheric wind speeds and (b) specific humidity at 700, 850, and 925 hPa. IOPs 2, 12, and 13 have statistically significant results at the 95% level for WSPD impacts throughout the troposphere (Fig. 22). SPCH impacts at the 95% significance occur for IOP-2 and IOP-8.

Some PMSL degradations for 24, 60, and 72 h are significant at a low (80%) level and the only significant improvement, at 80% confidence, is for forecast hour 120. CTRL 24–36-h geopotential height forecast improvements are statistically significant at 90% for 200 hPa only; they are predominantly degradations and marginally significant (at 80% level) for 24 to 72 h at 700–850 hPa. Longer-range forecasts beyond 72 h lack significance, except for an isolated value for 850 hPa at 120 h. Impacts at 144 to 168 h, while mostly larger in magnitude than for shorter forecast ranges, are not statistically significant due to the larger forecast standard deviations at long ranges.

WSPD analyses and forecasts to 36 h are notably and uniformly positive, and statistically significant (95% for analyses and at least 80% for forecasts). Upper-tropospheric WSPD

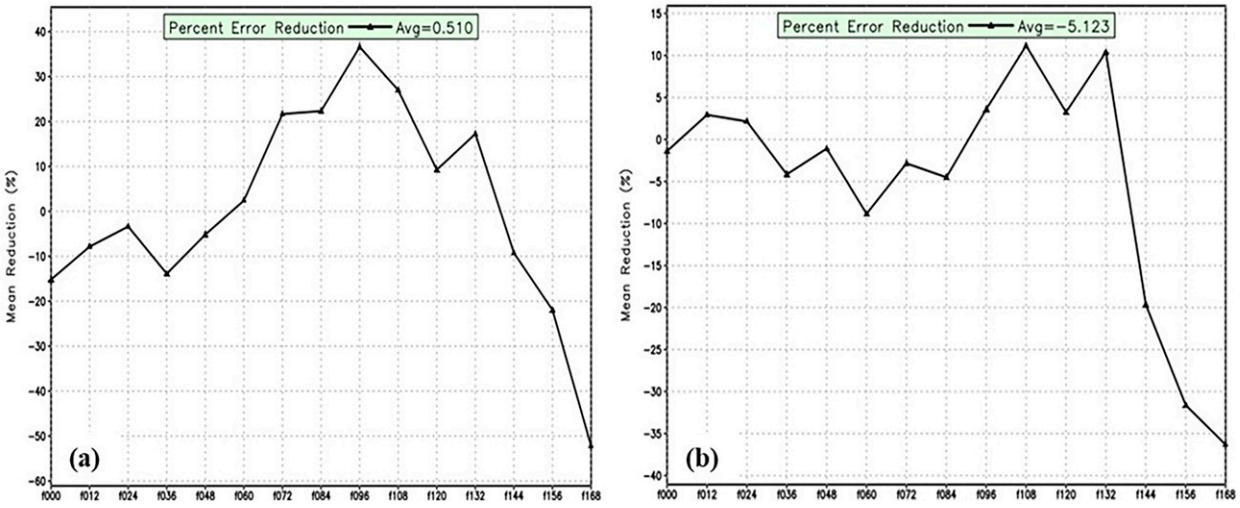

FIG. 24. Mean sea level pressure MAE reduction for 0–168-h forecasts and (a) IOP-8, initialized at 0000 UTC 15 Feb, and (b) IOP-9, initialized at 0000 UTC 16 Feb.

forecasts are dominantly degradations and 80% significant at 60 to 96 h, but with some scattered improvements at 108 and 120 h. While SPCH improvements are comparatively large (Fig. 20) they are not statistically significant beyond 24 h (except for 144 h at 925 hPa), due to the large variability in SPCH impacts at these time ranges.

(ii) Forecast-averaged impacts

By averaging the percent impact for the basic model forecast variables over all forecast hours, as done for IVT (Fig. 12), the cumulative impact of added observations from each IOP can be assessed. For PMSL (Fig. 21a), the largest positive (+12.9%) and negative (−14.3%) impacts are for IOP-2 (29 January) and IOP-13 (7 March), respectively, which are statistically significant at the 95% level (Fig. 22). PMSL impacts for other IOPs range from +6.9 to −7.9%, with 9 of the remaining 15 IOPs with less than 3% impact magnitude, most of which have either lower or no statistical significance (Fig. 22). Across all IOPs, PMSL impacts do not always appear correlated with those for Z200 and Z500 (e.g., IOP-2, Figs. 21a,b), but there is somewhat higher correlation with Z700 and Z850. The large PMSL degradation for IOP-13, followed by more neutral impacts in subsequent IOPs 14–17, is also reflected in all geopotential height impacts and implies successive forecast improvement in consecutive IOPs through the GDAS cycle. Tropospheric wind speed impacts across all IOPs (Fig. 23a) appear well correlated with PMSL impacts. Lower tropospheric moisture impacts, while generally positive, are less uncorrelated with the other model variables

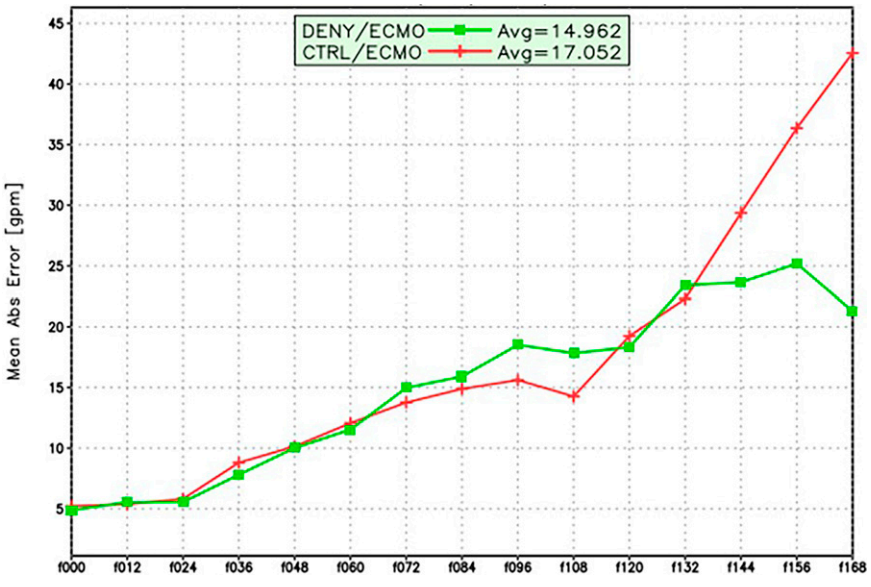

FIG. 25. Z700 forecast MAE for IOP-8, initialized at 0000 UTC 15 Feb.

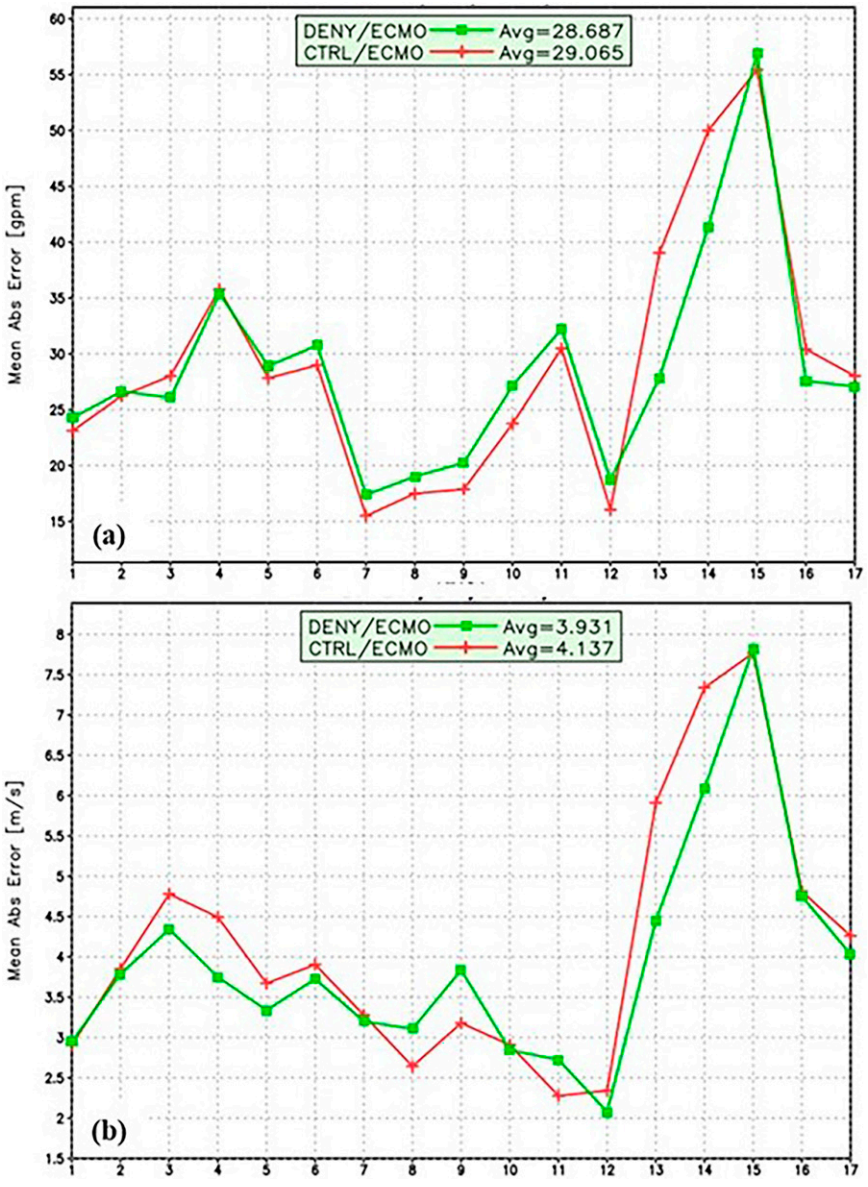

FIG. 26. MAE for (a) 500-hPa geopotential height and (b) 200-hPa wind speed for 84-h forecasts and each of the 17 AR2020 IOPs.

(as expected) but nevertheless show strong correlation with winds and PMSL for IOP-2 and IOP-13.

Forecast-averaged statistical significance levels for each IOP (Fig. 22) are highly consistent across both the pressure level ranges and multiple variables for each IOP (WSPD and SPCH, for example), whether for improvement or degradation. In this respect, IOP-2 is notable for its strong and significant improvements and IOP-13 is notable for its large and significant degradation. IOP-12 has strong and significant (95%) improvements for Z and WSPD, but weak SPCH impact and a surprisingly small IVT impact (Fig. 12). IOP-6 has small, but highly significant, improvements for Z, lower-tropospheric WSPD, and 700–850-hPa SPCH, but the overall IVT impact is weakly negative. IOP-14 is notable for its lack of significant impact, even though 55 dropsondes were deployed. Statistically significant (90%–95%), but relatively small impacts for model forecast variables do not always translate into significant IVT impacts of the same sign.

While the forecast-averaged PMSL impact for IOP-8 (15 February) is close to zero (Fig. 21a), impact over the full forecast range (Fig. 24a) varies from negative (12–48 h), to positive (60–132 h), and finally large negative (144–168 h) as the forecast proceeds. None of these excursions is statistically significant (Fig. 22) due to the large impact variability over the entire forecast. IOP-9 (16 February) also has a similar pattern (Fig. 24b) with small positive impact (12–24 h),

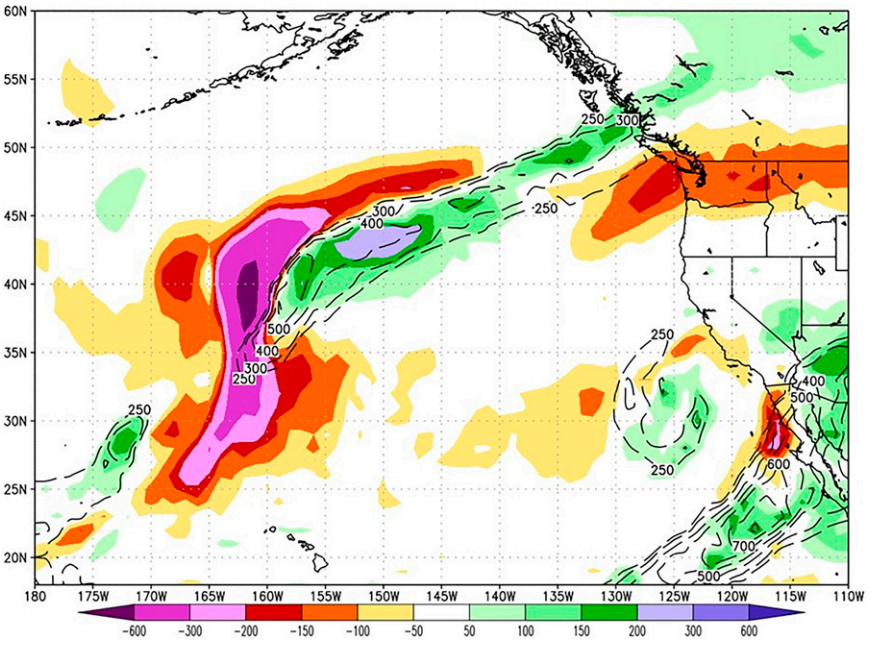

FIG. 27. IOP-13 (7 Mar) CTRL IVT 96-h forecast error (shaded) and forecast contours (black) valid for 0000 UTC 11 Mar 2020.

followed by moderate negative impact (36–84 h), positive impact (96–132 h) and large negative impacts at 144–168 h, which are statistically significant at the 80% level. Impacts for these IOPs are discussed further below, with further details in Lord et al. (2022a).

(iii) Impacts for IOPs 2, 8–9, and 13

The positive impact for IOP-2 occurs for wind speed and moisture at the selected levels (Figs. 23a,b), and also for lower tropospheric geopotential height, but to a lesser extent for Z200 and not for Z500 (Fig. 21b). Forecast-averaged geopotential height impacts over the deep 200–925-hPa layer for IOP-2 average 6.1%, with a range of −0.3% at 500 hPa to 15.2% at 925 hPa (not shown), with improvements of 2%–4% above 500 hPa. This vertical structure suggests that the AR OC dropsondes and the GDAS provide an improved analyzed CTRL baroclinic structure for this IOP and that both play a positive role in forecast improvement. From the analysis time (0000 UTC 29 January) to the end of the major rainfall event over Washington state at 0000 UTC 2 February (appendix B), the forecast CTRL wind speed, moisture and IVT verifications are generally improved over the DENY experiment.

While IOP-8 (15 February) is characterized by a small positive (3.4%) impact on IVT (Fig. 12), and overall positive impact for both wind speed (~4%, Fig. 23a) and specific humidity (~4%, Fig. 23b), the impact for mean sea level pressure is near zero (as noted previously) and negative (~−4%) for geopotential height (Fig. 21b). This neutral/negative impact on the mass field is exemplified by the Z700 errors (Fig. 25), and is driven by a rapid increase of the CTRL IOP-8 forecast error from 108 h, when the CTRL was improved, onward to 168 h, when there is a major (but not statistically significant) CTRL degradation. At 168 h the CTRL error is double that of the DENY experiment. A similar negative forecast impact at 132–168 h for geopotential height is present throughout the 200–925-hPa layer for IOP-8 and also for IOP-9 (16 February).

Enhanced IVT degradation at longer forecast times often reflects errors for IVTs that enter the verification domain during the forecast, typically from the western boundary (see Lord et al. 2022a for examples). However, the large geopotential height CTRL degradations at 144 h onward for IOPs 8–9 derive primarily from enhanced errors in the CTRL large-scale mass field due to increased growth of a low pressure system over the northern Gulf of Alaska and Prince William Sound and progressive southeast propagation of this system to west of Vancouver Island subsequent to IOP-9. While the DENY experiment has errors, they are less impactful on the day-6–7 forecast accuracy so that the CTRL error growth unexpectedly produces an inferior forecast.

The IOP-13 forecast includes two major midlatitude IVT systems and a long-lasting sequence of IVTs landfalling over Baja California and southern California. This complex event sequence is discussed in more detail in Lord et al. (2022a). The first midlatitude IVT enters the verification domain at 50°N at the analysis time (0000 UTC 7 March) and the second develops on the southwestern end of the previous IVT at 0600 UTC 9 March in association with a developing cyclone at 30°N, 178°W. This cyclone continues to strengthen as it moves north northeast to 42°N, 165°W by 1200 UTC 12 March.

As noted earlier (Fig. 13b), in the first 36 h of the IOP-13 forecast, the CTRL has neutral to positive IVT MAE reductions of 2% at 12 h and 8% at 36 h. However, the forecast rapidly degrades thenceforth to a major degradation of 25% by

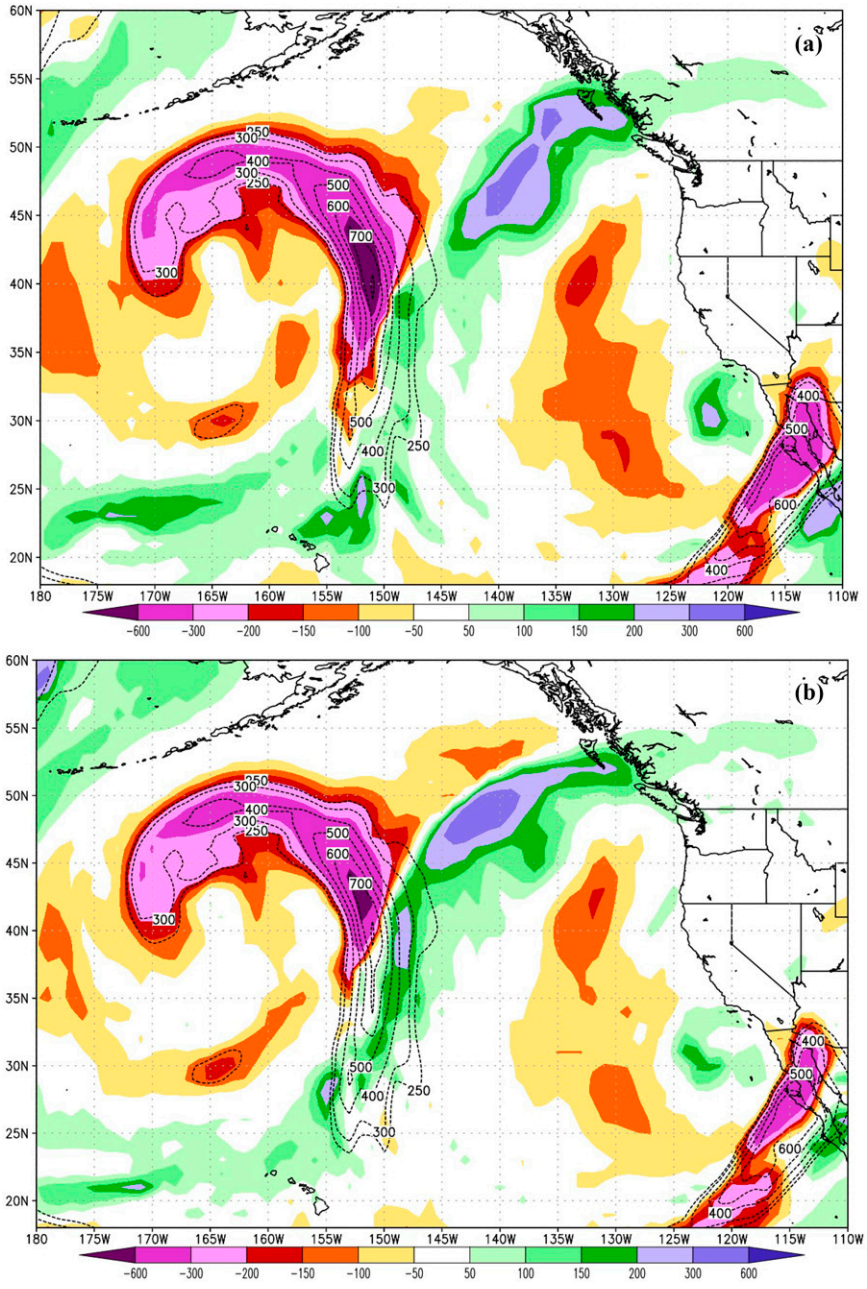

FIG. 28. IOP-13 (7 Mar) (a) CTRL and (b) DENY IVT 132-h forecast error (shaded) and ECMO IVT verification contours (black) for 1200 UTC 12 Mar 2020.

96 h, such that the forecast-averaged IVT degradation is statistically significant at the 95% level. Similar impacts occur for PMSL, geopotential height and wind speed, resulting in large negative and statistically significant (95%) forecast-averaged impacts for all forecast variables in IOP-13 (Figs. 21–23). For example, at 84 h, IOP-13 CTRL and DENY forecast errors for Z500 and WSPD200 are almost twice as large as for other IOPs (Figs. 26a,b) and they, along with corresponding errors at other levels, account for the degradation in the IVT landfall errors at 1200 UTC 10 March (Lord et al. 2022a). Relative errors initialized 24 h later for IOP-14 (and for the 72–84-h forecasts) have the same pattern but are smaller and not statistically significant (Fig. 22).

Forecast errors for the second IOP-13 IVT are associated with a developing trough moving northeast from the date line (not shown here, but see Lord et al. 2022a, appendix D for further details) which is underpredicted along with its accompanying IVT. The errors grow from 60 h onward in both the CTRL and DENY forecasts, with the CTRL errors growing faster and dominating the MAE statistic from 72 to 156 h.

TABLE 2. Domain-wide IVT MAE for CTRL and DENY forecasts from consecutive IOPs 13–17, all valid at 1200 UTC 12 Mar. The forecast hour and initial dates are also tabulated. The verification domain is 4 (Table 1).

IOP	Forecast hour	Initial date (0000 UTC)	CTRL MAE (kg m^{-1} s^{-1})	DENY MAE (kg m^{-1} s^{-1})
13	132	7 Mar	88.6	76.3
14	108	8 Mar	81.0	75.6
15	84	9 Mar	60.3	60.1
16	60	10 Mar	30.4	34.1
17	36	11 Mar	25.2	26.0

IVT-related MAEs for this forecast period, and MAEs for all model forecast variables, are a consequence of the GFS IVT forecast persisting the eastward IVT path (Figs. 27 and 28), i.e., following its predecessor, until landfall at 132 h (1200 UTC 12 March) vice the IVT verification path that continues to follow the strengthening and northward moving cyclone (Figs. 28a,b), resulting in an eventual landfall over the Aleutians (see Lord et al. 2022a for additional details). Consecutive CTRL and DENY forecasts valid at 1200 UTC 12 March (12 h before landfall) from IOPs 13–17 (Table 2) show that the above forecast scenario continues, but to a lesser extent, for IOPs 14–15 (8–9 March) until the error growth is reduced by the assimilation of later IOP data and the shortened lead time to landfall for IOPs 16–17, at which time the CTRL has reduced error compared to DENY.

During IOP-13, observations were not deployed for prospective landfalling IVTs over Baja and southern California. However, the IVT-13 forecast includes multiple impacting events from the merged cyclone–IVT system over this area (Table 3). Over 12–48 h, an IVT makes landfall over south Baja and is well forecast by both CTRL and DENY experiments through 48 h. Over the forecast range 60–84 h and verifying over 1200 UTC 9 March–1200 UTC 10 March, the system begins to be influenced by the offshore cyclone and the IVT progresses northward and reaches the U.S.–Mexico border. At this time, both CTRL and DENY forecasts are too strong and displaced to the north, but the CTRL MAE is reduced. During the 96–108-h forecast period, a second IVT maximum approaches Baja: both the CTRL and DENY IVTs are displaced further to the south, with the CTRL more so. In the last event, from 132 to 144 h of the forecast, the IVT splits into two branches, with the western branch extending northward to eastern Arizona and the second branch heading northeast into Texas and Oklahoma. Over this 12-h period, the CTRL IVT MAE is much larger than the DENY (e.g., Figs. 28a,b) and reflects the southern displacement of the IVT and the weak extension into the southern United States. Later on, as captured in forecasts from IOPs 16–17, major precipitation occurs in both of these areas over 13–14 March as shown in Part I.

(iv) Impact summary

Forecast-averaged IVT impacts for all IOPs (Fig. 12) were sorted into five categories and ranked from 1 to 5, where category 1 contains the largest (three) negative IOP impacts (Fig. 29a), category 5 the largest (three) positive impacts, category 3 contains the five middle/neutral IOP forecast impacts and categories 2 and 4 are intermediate impacts (three each). Likewise, forecast-averaged impacts for the model forecast variables were similarly sorted and categorized after vertical averaging for each IOP: geopotential height and wind speed were averaged for mandatory pressure levels over the 200–925-hPa layer and moisture was averaged over the 700–925-hPa layer.

As noted previously (Fig. 12), IVT impacts number 8 IOPs being improved and 9 degraded, with 3 improvements and 4 degradations being statistically significant at 80% or above (90% and 95%). Overall, IOP impact improvements for vertically averaged model variables (Z, WSPD, and SPCH) are more prevalent than for IVT (Fig. 29b),[1] with 11 improvements and 6 degradations for both WSPD and SPCH; of these, 9 improvements and 3 degradations are statistically significant. For Z, 6 of 9 improvements are significant while 5 of 8 degradations are also significant. This complex mixture of improvements/degradations and significant/insignificant results among three basic model variables does not uniformly translate into a clear and statistically significant positive result for IVT impacts although impact correlations between Z, WSPD, SPCH, and IVT are positive overall as shown in Part I.

Forecast-averaged IVT impacts for the two IOPs with the largest magnitudes (IOP-2, improved and IOP-13, degraded, Figs. 12 and 29b) are each significant at the 95% level; furthermore, vertically averaged Z, WSPD, and SPCH impacts for each of these IOPs are also of the same sign and significant at 95%. For the second ranked IOPs for IVT improvement/degradation (8 and 10, respectively), this consistency breaks down: there is a degraded Z result for IOP-8 (significant at the 90% level) and the IVT degradation for IOP-10 is less significant (90%) even though impacts for Z, WSPD, and SPCH are all at 95% significance. Throughout this impact matrix, the highest/lowest impacts for the model variables themselves are not necessarily for those IOPs with the largest IVT impacts. The third-ranked IOPs in the highest/lowest sorted list for IVT (IOP-7 and IOP-3, respectively) are marginally significant at the 80% level despite a significant degradation for SPCH (IOP-7) and a weaker, nonsignificant SPCH negative impact for IOP-3.

Generally, moderate and neutral forecast-averaged IVT impacts (sort categories 2–4, Fig. 29a) are not statistically significant, even though impacts for corresponding model variables may be significant at the 95% level (e.g., IOPs 4, 12), but of small magnitude. In most cases of small and mixed positive/negative impact across the forecast variables or (e.g., IOPs 14 and 16) IVT impacts are also small and insignificant. The exception is IOP-1, the first AR2020 OC forecast, which has neutral, but statistically significant (95%), improvements for WSPD and SPCH and a small but significant negative impact for IVT. The fact that IOP-1 is the first dropsonde

[1] As expected since an improved/degraded IVT depends on like contributions from multiple, independent model variables such as wind and moisture.

TABLE 3. Summary of significant system events for the Baja California IVT as covered by the IOP-13 12–168-h forecast. A brief description of the event, the covering forecast range, verification date, and CTRL and DENY MAEs are tabulated. The verification domain is 5 (Table 1, Fig. 1).

Event	Description	Forecast event range (h)	Forecast hour	Verification date	CTRL MAE (kg m^{-1} s^{-1})	DENY MAE (kg m^{-1} s^{-1})
1	Initial landfall	12–48	48	0000 UTC 9 Mar	48.4	45.7
2	Cyclone influence	60–84	84	1200 UTC 10 Mar	57.6	68.5
3	Second IVT maximum landfall	96–108	108	1200 UTC 11 Mar	66.4	53.7
4	Baja IVT splits into two streams	132–144	144	0000 UTC 13 Mar	94.2	71.4

deployment, and remote errors have not yet had an impact through the GDAS, may be contributing to this exception.

5. Summary and discussion

Aircraft-deployed wind and thermodynamic soundings, ingested into the NCEP GDAS during the 17 AR2020 IOPs, produced IOP-averaged analysis and forecast error reductions of basic forecast variables, from 0 to 36 h, as follows: wind speed (300–850 hPa), moisture (700 hPa), and geopotential height (200 and 500 hPa). These reductions were statistically significant at 95%–80% confidence levels (Fig. 17). Beyond 36 h and for other pressure levels, however, statistical significance rarely meets or exceeds the 80% level and, when present, is mostly for forecast degradations, not improvements.

IOP-averaged IVT impacts for analyses are positive for domain-wide mean, minimum, and maximum values in the sense that the CTRL analyses are closer to the ECMO verification, but they are not statistically significant. Forecast-averaged IVT impacts over all IOPs has a plurality of degradations (Fig. 12) and the mean impact over all forecast hours and IOPs was marginally negative, but not statistically different from zero. Nevertheless, some IOPs (2, 8) had statistically significant (95%) improvements (Fig. 29) in which basic model

(a)

Rank	Key	#IOPs
1		3
2		3
3		5
4		3
5		3

(b)

IOP	Date	IVT	Sig.	Z	Sig.	WSPD	Sig.	SPCH	Sig.
1	00Z24JAN2020	-1.45	***	0.26		0.78	***	1.07	***
2	00Z29JAN2020	6.82	***	6.12	***	7.33	***	5.88	***
3	00Z31JAN2020	-3.65	*	-7.05	***	-4.7	***	-1.23	
4	00Z04FEB2020	1.64		3.24	***	2.91	***	2.47	***
5	00Z05FEB2020	-1.86		1.86		-0.47		0.47	
6	00Z06FEB2020	-1.06		3.90	***	0.31		2.39	***
7	00Z14FEB2020	2.15	*	5.05	***	1.22	***	-2.25	***
8	00Z15FEB2020	3.40	***	-4.12	**	4.33	***	5.47	***
9	00Z16FEB2020	-2.01		-2.36	*	-0.32		0.20	
10	00Z21FEB2020	-3.93	**	-9.23	***	-4.03	***	-4.14	***
11	00Z24FEB2020	1.87		1.33	***	1.51	*	-0.22	
12	00Z02MAR2020	2.05		9.82	***	4.19	***	2.50	***
13	00Z07MAR2020	-11.02	***	-18.16	***	-11.39	***	-5.87	***
14	00Z08MAR2020	-2.09		-1.29		0.98		3.15	*
15	00Z09MAR2020	1.19		0.08		0.75	*	-0.37	
16	00Z10MAR2020	1.24		-1.13	*	1.11	**	2.75	***
17	00Z11MAR2020	-1.32		-0.20		-0.51		1.58	***
	N+/NS+	8	3	9	6	11	9	11	9
	N-/NS-	9	4	8	5	6	3	6	3

FIG. 29. Ranked impact (I; %), colorized categories for (a) forecast-averaged IVT and (b) forecast- and vertically averaged geopotential height, wind speed, and specific humidity impacts. Statistical significance at 95% (***), 90% (**), and 80% (*) is given in adjacent columns to the right of the ranked impacts. Here, N+ (N−) and NS+ (NS−) are the numbers of IOPs with positive (negative) and statistically significant positive (negative) impacts, respectively. See text for processing details.

forecast variables (Z, WSPD, and SPCH) supported the IVT improvements with significant (90%–95%) error reductions of their own throughout the forecast. IOP-13 (7 March) is notable for an extremely negative, statistically significant (95%) impact (Figs. 27 and 28) which resulted from complex nonlinear interactions governing evolution of an AR system entering the AR2020 verification domain late in the forecast. Within the AR phenomenon, the IVT geographical extent and strength are more difficult features to verify since they are localized and intermittent; a larger sample size is most likely required to demonstrate significant impact overall. IVT features are also challenging to improve as they depend on impacts to all of the multilevel model forecast variables (moisture, wind, and geopotential height/temperature).

In the AR2020 OC, additional in situ moisture information provided by the dropsondes appears to have a widespread positive forecast impact as it is a unique and accurate measurement over most oceanic areas and of sufficient density to characterize the ARs. Improved IOP-averaged CTRL moisture verification at 120 h and beyond (Fig. 19) is notable, but not statistically significant at the 80% level due to large moisture variability at these longer forecast hours. Over all IOPs, however, moisture changes are positively correlated with IVT improvements for 24–120-h forecasts (as are impacts for geopotential height and wind speed) and likely play a positive role in precipitation impacts as well (see Part I).

Episodic dropsonde deployments impact larger areas both upstream and downstream of the verification area through the GDAS cycling process as time proceeds. Observations for a particular AR event may impact subsequent AR events as they enter the verification area beyond 36–48 h. The sign of this impact is generally unpredictable, but for very limited observation coverage it is more likely to be negative or neutral than positive. Some positive regional impacts for 24–120-h precipitation forecasts (see Part I) and for 108–168 h (see Part I) and elsewhere (Wu et al. 2021) have been described; nevertheless, a more detailed comparison between precipitation and IVT forecast improvements for each IOP is warranted, together with examination of associated wind speed and moisture statistics over the entire OC.

Due to small sample sizes, statistics averaged over the 17 IOPs can be dominated (in either direction) by a single case. For example, the improvement at 144 h for wind speed and moisture (Figs. 18 and 19) comes primarily from IOP-14 (0000 UTC 8 March deployment, Figs. 22a,b), which is large compared to all other impacts (not shown). Indeed, there is considerable impact variability in the complete forecast sample for forecast length, IOP and model variable. Furthermore, small forecast differences in the short range can occasionally lead to major subsequent forecast errors, as noted earlier for IOP-13 (Lord et al. 2022a). While the mechanisms described by Hodyss and Majumdar (2007) and Ancell et al. (2018), related to spreading of model-generated noise cannot be ruled out, tracking of the forecast differences between CTRL, DENY, and the verifying analyses has revealed error mechanisms that appear physically realistic. Given that the cycled data assimilation spreads the inserted dropsonde information globally over time, occasional large forecast differences are bound to occur by virtue of the intrinsic nonlinearity of the atmosphere.

This study also documents some heightened PBL moisture variability in the GFS/GDAS system relative to that in the ECMWF forecast system. A more substantial study of the low-level moisture distribution, relative to the dropsonde data, may prove useful in diagnosing this issue.

The ARR program has evolved from a field demonstration to a real-time operational capability as documented in the National Winter Season Operations Plan by the Office of Federal Coordinator of Meteorology (OFCM 2020), thereby providing more opportunities to thoroughly investigate the impact of aircraft observations on NCEP's operational global model analysis and forecasts. The ARR program in 2021 consisted of multiple aircraft and several sequential (multiday) IOPs, giving unprecedented temporal coverage of ARs. In addition, NCEP made significant advancements to the operational GFS and GDAS in 2021 (GFSv16; Yang et al. 2021; Kleist et al. 2021). For the first time, real-time data denial experiments were conducted using GFSv16. Results from those experiments will be documented in a sequel to this manuscript, along with a focused examination of precipitation forecast improvements at local watershed levels where the impacts are found to be more pronounced.

Acknowledgments. This work is supported by generous funding made available from NOAA/Office of Marine and Aviation Observations (OMAO). The authors are grateful for feedback from the Modeling and Data Assimilation Steering Committee of the AR Recon Program and for comments from the internal and external reviewers of this manuscript, especially from external reviewer 2 on the topic of statistical significance.

Data availability statement. All model data, observations, and statistical results are available through the corresponding author at the NOAA/National Centers for Environmental Prediction, Environmental Modeling Center.

APPENDIX A

AR2020 Intensive Observing Periods

Table A1 provides some details on the 17 AR2020 IOPs, including dates, participating aircraft, number of dropsondes (and failures) from each of the G-IV and C-130 aircraft, and the number of observations assimilated in real time by the GDAS for the 0000 UTC cycle and the surrounding cycles (1800 and 0600 UTC).

APPENDIX B

A Case Study of Forecasts Initialized at 0000 UTC 29 February 2020

The 120-h forecast impact from 0000 UTC 29 February 2020 is a rare case of substantial indirect impact as defined

TABLE A1. Summary of the AR2020 IOPs. Columns identify the serial number of each IOP, the central time for the aircraft observations, the number of dropsondes (failed number), and the number of sondes included in the 1800, 0000, and 0600 UTC GDAS cycles.

IOP No.	Date 2020	No. of dropsondes (failed drops) AF C-130	AF C-130	NOAA G-IV	GDAS assimilated IOP dropsondes	GDAS cycle (1800/0000/0600 UTC)
1	0000 UTC 24 Jan	25 (0)	14 (1)		38	1/37/0
2	0000 UTC 29 Jan	26 (1)	2 (0)		27	(2/21/4)
3	0000 UTC 31 Jan		25 (1)		24	0/24/0
4	0000 UTC 4 Feb	19 (4)	25 (0)	35 (5)	70	0/70/0
5	0000 UTC 5 Feb			34 (4)	30	0/30/0
6	0000 UTC 6 Feb	25 (3)	26 (1)	35 (5)	59	0/59/0
7	0000 UTC 14 Feb			30 (0)	27	0/27/0
8	0000 UTC 15 Feb	26 (1)	31 (6)	35 (5)	80	0/78/2
9	0000 UTC 16 Feb			30 (0)	30	0/30/0
10	0000 UTC 21 Feb			30 (0)	30	1/29/0
11	0000 UTC 24 Feb	29 (3)		32 (2)	56	1/55/0
12	0000 UTC 2 Mar	27 (5)	25 (1)	30 (0)	76	0/74/2
13	0000 UTC 7 Mar			31 (1)	30	0/30/0
14	0000 UTC 8 Mar		28 (3)	31 (1)	55	0/54/1
15	0000 UTC 9 Mar		23 (3)	29 (0)	29	0/29/0
16	0000 UTC 10 Mar		30 (0)	27 (0)	43	0/39/4
17	0000 UTC 11 Mar		2 (0)		2	0/2/0

in section 2 and in Part I. The CTRL initialization contains no dropsonde data and the closest previous IOP was at 0000 UTC 24 February, 5 days earlier. Analysis differences at 0000 UTC 29 February (Fig. B1) are of magnitude 2–4 m throughout most of the domain, with isolated minimum/maximum differences of −13 and 9 m, respectively. The impact at forecast hour 120 is the largest Z500 improvement over the experimental period (Fig. 6). Forecast MAE (Fig. 7) increases linearly and almost equally with forecast hour for both CTRL and DENY for the first 96 h. After 96 h, the DENY MAE increases much faster before leveling off at 144 h.

At 96 h, in comparison with the ECMO verifying analysis (Fig. B2a), both CTRL and DENY forecasts exhibit much the same error patterns (Figs. B2b,c). The CTRL forecast exhibits maximum positive forecast errors exceeding 100 m due to an overdeveloped ridge in southwest Canada, a progressive trough over Texas instead of a cutoff low over Arizona, New Mexico, and northern Mexico, and a developing ridge with errors exceeding 150 m over the eastern United States instead of a shallow trough over the Great Lakes. These error patterns are accentuated in the DENY forecast, especially as it develops a deeper trough from Texas to Minnesota (errors less than −160 m) and amplifies the

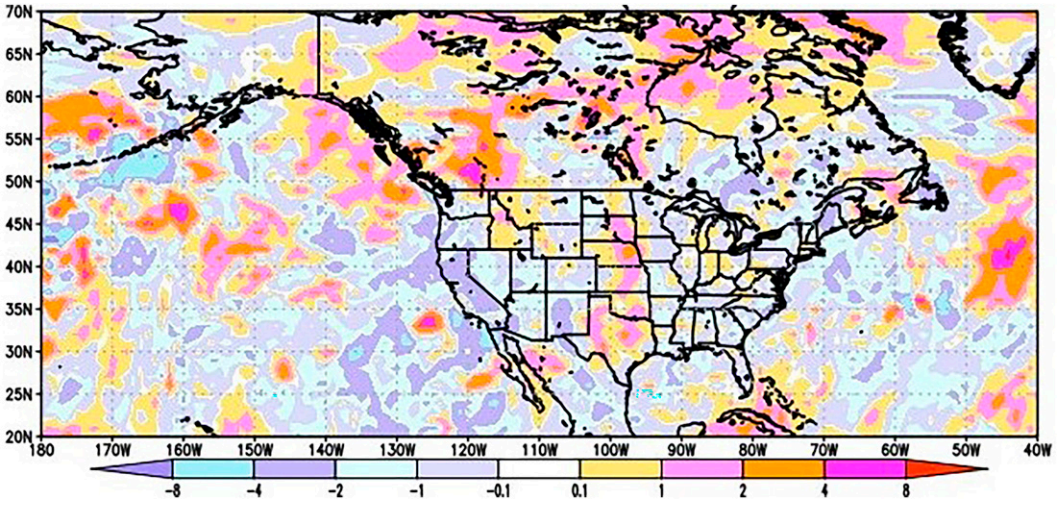

FIG. B1. Z500 analysis difference between CTRL and DENY experiments at 0000 UTC 29 Feb. Minimum (maximum) differences in the domain 20°–75°N, 40°W–180° are −13.4 (9.1) m, the mean difference is −0.3 m, and the mean absolute difference is 1.2 m.

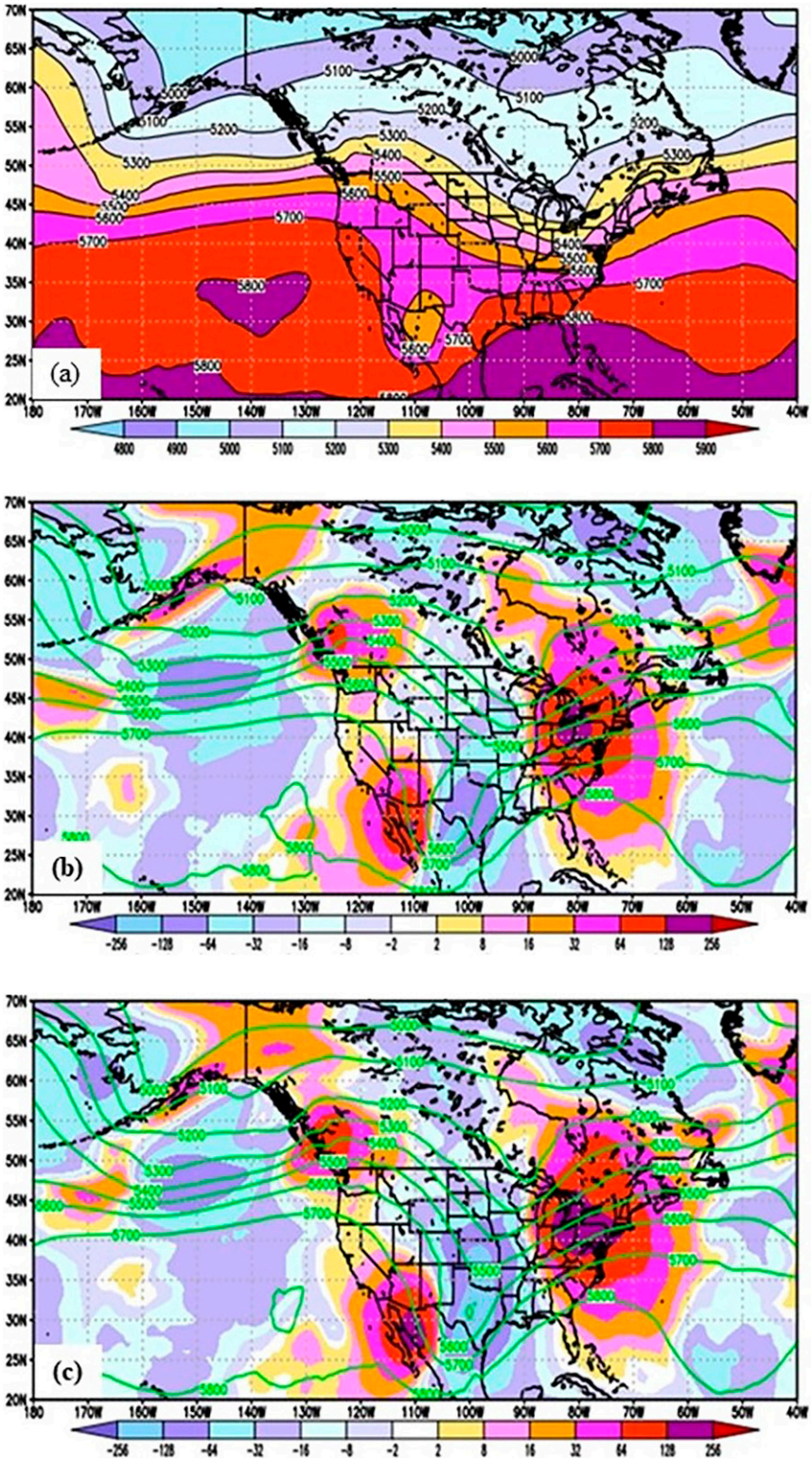

FIG. B2. (a) Z500 ECMO verifying analysis at 0000 UTC 4 Mar 2020 and (b) Z500 CTRL and (c) DENY forecast errors at 96 h. Solid colors depict the sign and magnitude of the forecast error, and green contours depict the CTRL forecast.

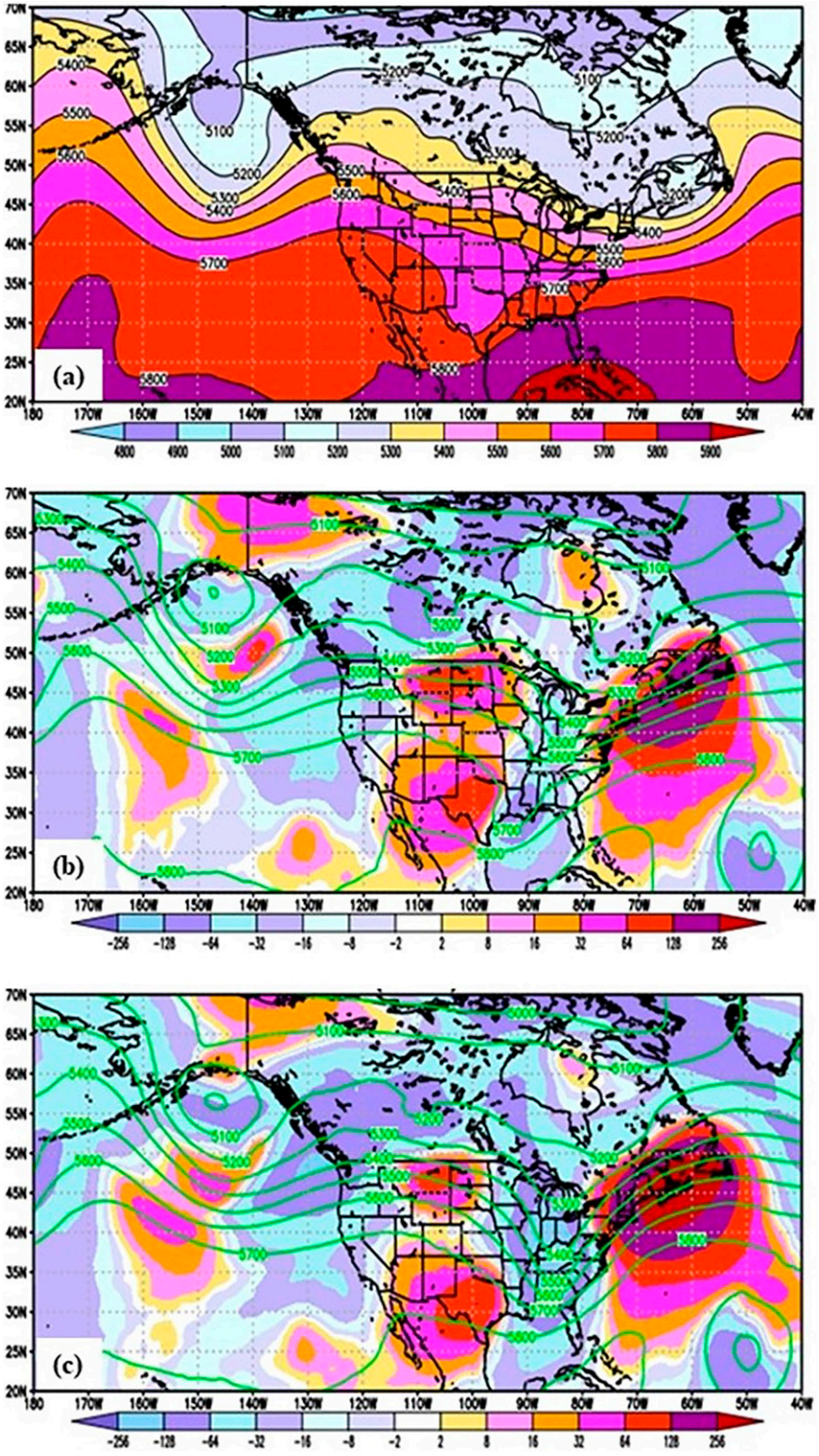

FIG. B3. As in Fig. B2, but for the valid date at 0000 UTC 5 Mar 2020 and 120-h forecasts.

ridge (errors exceeding 180 m) from the Great Lakes to New England.

At 120 h, the ECMO verifying analysis (Fig. B3a) shows a ridge over the western United States and southwestern Canada and a trough extending southeastward from Hudson Bay to the Canadian Maritime Provinces. The CTRL exhibits a weak trough southeast of Hudson Bay but does not extend it sufficiently eastward as in the verification. CTRL errors exceed 270 m in this region. DENY errors exceed 390 m as it forecasts a ridge over the Maritime Provinces down through New England. Over the Great Lakes region, CTRL predicts a weaker trough than DENY, but is still too progressive as the storm system actually develops 12–24 h later (not shown). Even further out in the forecast, at forecast hour 168 (not shown), DENY has the largest forecast errors over the mid-Atlantic region due to the Great Lakes trough having moved through that area in advance of the verification. This is a case of fast-growing error development that produced an indirect impact on AR2020 forecasts, but did not impact the important AR region of the western United States and Canada.

REFERENCES

Aberson, S. D., J. Cione, C.-C. Wu, M. Bell, J. Halverson, C. Fogarty, and M. Weissmann, 2010: Aircraft observations of tropical cyclones. *Global Perspectives on Tropical Cyclones: From Science to Mitigation*, J. C. L. Chan and J. D. Kepert, Eds., World Scientific, 227–240, https://doi.org/10.1142/9789814293488_0008.

Ancell, B. C., A. Bogusz, M. J. Lauridsen, and C. J. Nauert, 2018: Seeding chaos: The dire consequences of numerical noise in NWP perturbation experiments. *Bull. Amer. Meteor. Soc.*, **99**, 615–628, https://doi.org/10.1175/BAMS-D-17-0129.1.

Brennan, M. J., D. Kleist, K. Howard, and S. J. Majumdar, 2015: The impact of supplemental dropwindsonde data on the structure and intensity of Tropical Storm Karen (2013) in the NCEP global forecast system. *Wea. Forecasting*, **30**, 683–691, https://doi.org/10.1175/WAF-D-15-0002.1.

Doyle, J. D., C. A. Reynolds, and C. Amerault, 2019: Adjoint sensitivity analysis of high-impact extratropical cyclones. *Mon. Wea. Rev.*, **147**, 4511–4532, https://doi.org/10.1175/MWR-D-19-0055.1.

Elless, T. J., X. Wu, and V. Tallapragada, 2021: Identifying atmospheric river reconnaissance targets using ensemble forecasts. The 2021 Blue Book 1-07, 2 pp., https://wgne.net/bluebook/uploads/2021/docs/01_Elless_Travis_EnsembleForecasts,_AtmosphericRivers.pdf.

Harris, L., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, https://doi.org/10.1175/MWR-D-11-00201.1.

——, L. Zhou, X. Chen, and J.-H. Chen, 2020a: The GFDL finite-volume cubed-sphere dynamical core. NOAA Tech. Memo. OAR GFDL 2020-001, 10 pp., https://doi.org/10.25923/7h88-c534.

——, and Coauthors, 2020b: GFDL SHiELD: A unified system for weather-to-seasonal prediction. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002223, https://doi.org/10.1029/2020MS002223.

Hodyss, D., and S. J. Majumdar, 2007: The contamination of 'data impact' in global models by rapidly growing mesoscale instabilities. *Quart. J. Roy. Meteor. Soc.*, **133**, 1865–1875, https://doi.org/10.1002/qj.157.

Hou, D., and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeor.*, **15**, 2542–2557, https://doi.org/10.1175/JHM-D-11-0140.1.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, https://doi.org/10.1175/MWR-D-13-00350.1.

——, and Coauthors, 2021: NCEP operational global data assimilation upgrades: From versions 15 through 16. *WAF Symp. General Session*, Online, Amer. Meteor. Soc., 12.3, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378554.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

——, and R. B. Rood, 1997: An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quart. J. Roy. Meteor. Soc.*, **123**, 2477–2498, https://doi.org/10.1002/qj.49712354416.

Lindgren, B. W., and G. W. McElrath, 1959: *Introduction to Probability and Statistics*. McMillan, 277 pp.

Lord, S. J., X. Wu, and V. Tallapragada, 2022a: Overview of the 2020 atmospheric rivers field campaign. NOAA/NCEP Office Note 508, 168 pp., https://doi.org/10.25923/pjwn-p075.

——, ——, ——, and F. M. Ralph, 2022b: The impact of dropsonde data on the performance of the NCEP global forecast system during the 2020 atmospheric rivers observing campaign. Part I: Precipitation. *Wea. Forecasting*, **38**, 17–45, https://doi.org/10.1175/WAF-D-22-0036.1.

Majumdar, S. J., M. J. Brennan, and K. Howard, 2013: The impact of dropsonde and supplemental rawinsonde observations on track forecasts for Hurricane Irene (2011). *Wea. Forecasting*, **28**, 1385–1403, https://doi.org/10.1175/WAF-D-13-00018.1.

OFCM, 2020: National winter season operations plan. Rep. FCM-P13-2020, OFCM, 126 pp., https://www.icams-portal.gov/resources/ofcm/nwsop/2020_nwsop.pdf.

Putman, M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Ralph, F. M., M. D. Dettinger, M. M. Cairns, T. J. Galarneau, and J. Eylander, 2018: Defining "atmospheric river": How the Glossary of Meteorology helped resolve a debate. *Bull. Amer. Meteor. Soc.*, **99**, 837–839, https://doi.org/10.1175/BAMS-D-17-0157.1.

——, and Coauthors, 2020: West Coast forecast challenges and development of atmospheric river reconnaissance. *Bull. Amer. Meteor. Soc.*, **101**, E1357–E1377, https://doi.org/10.1175/BAMS-D-19-0183.1.

Reynolds, C. A., J. D. Doyle, F. M. Ralph, and R. Demirdjian, 2019: Adjoint sensitivity of North Pacific atmospheric river forecasts. *Mon. Wea. Rev.*, **147**, 1871–1897, https://doi.org/10.1175/MWR-D-18-0347.1.

Schäfler, A., and Coauthors, 2018: The North Atlantic waveguide and downstream impact experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1607–1637, https://doi.org/10.1175/BAMS-D-17-0003.1.

Schindler, M., M. Weissmann, A. Schäfler, and G. Radnoti, 2020: The impact of dropsonde and extra radiosonde observations during NAWDEX in autumn 2016. *Mon. Wea. Rev.*, **148**, 809–824, https://doi.org/10.1175/MWR-D-19-0126.1.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Wu, X., V. Tallapragada, S. Lord, K. Wu, and M. Ralph, 2021: Impact of atmospheric river reconnaissance dropsonde data on NCEP GFS forecast: A case study. The 2021 Blue Book, 23–24, http://bluebook.meteoinfo.ru/uploads/2021/sections/BB_21_S1.pdf.

Yang, F., and V. Tallapragada, 2018: Evaluation of retrospective and real-time NGGPS FV3GFS experiments for Q3FY18 beta implementation. *25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 5B.3, https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345231.html.

——, V. S. Tallapragada, D. T. Kleist, A. Chawla, J. Wang, R. Treadon, and J. Whitaker, 2021: On the development and evaluation of NWS global forecast systems version 16. *12th WAF Symp. General Session*, Online, Amer. Meteor. Soc., 12.2, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378135.