BAMS Article

Observing System Simulation Experiments (OSSEs) in Support of Next-Generation NOAA Satellite Constellation

Lidia Cucurull,[a] Richard A. Anthes,[b] Sean P. F. Casey,[b,c] Michael J. Mueller,[b,c] and Andres Vidal[b,c]

KEYWORDS: Remote sensing; Satellite observations; Sensitivity studies; Data assimilation; Numerical weather prediction/ forecasting

ABSTRACT: Between 2014 and 2018, the National Oceanic and Atmospheric Administration conducted the NOAA Satellite Observing System Architecture (NSOSA) study to plan for the next generation of operational environmental satellites. The study generated some important questions that could be addressed by observing system simulation experiments (OSSEs). This paper describes a series of OSSEs in which benefits to numerical weather prediction from existing observing systems are combined with enhancements from potential future capabilities. Assessments include the relative value of the quantity of different types of thermodynamic soundings for global numerical weather applications. We compare the relative impact of several sounding configuration scenarios for infrared (IR), microwave (MW), and radio occultation (RO) observing capabilities. The main results are 1) increasing the revisit rate for satellite radiance soundings produces the largest benefits but at a significant cost by requiring an increase in the number of polar-orbiting satellites from 2 to 12; 2) a large positive impact is found when the number of RO soundings per day is increased well beyond current values and other observations are held at current levels of performance; 3) RO can be used as a mitigation strategy for lower MW/IR sounding revisit rates, particularly in the tropics; and 4) smaller benefits result from increasing the horizontal resolution along the track of the satellites of MW/IR satellite radiances. Furthermore, disaggregating IR and MW instruments into six evenly distributed sun-synchronous orbits is slightly more beneficial than when the same instruments are combined and collocated on three separate orbits.

SIGNIFICANCE STATEMENT: The results of this paper are significant because they inform decision-makers about the future configuration of the NOAA's environmental satellite constellation, which serves millions of diverse users.

DOI: 10.1175/BAMS-D-23-0060.1
Corresponding author: Lidia Cucurull, lidia.cucurull@noaa.gov
Manuscript received 17 March 2023, in final form 26 February 2024, accepted 1 March 2024

AFFILIATIONS: [a] NOAA/Quantitative Observing System Assessment Program, Boulder, Colorado; [b] University Corporation for Atmospheric Research, Boulder, Colorado; [c] CPAESS/QOSAP, Boulder, Colorado

1. Introduction

As planning for future satellite systems takes a long time, in 2014, the National Oceanic and Atmospheric Administration (NOAA) began evaluating existing and new opportunities for the design of the next-generation operational environmental satellite constellation beyond 2030. Costs of developing, deploying, and maintaining new space-based observing systems typically exceed $100–500 million per instrument; thus, it is critical to determine the most cost-effective system. The NOAA Satellite Observing System Architecture (NSOSA) study (Volz et al. 2016; NOAA 2018a,b; Maier et al. 2021) investigated alternative satellite constellations and new capabilities to ensure the readiness of new satellites when current operating satellites are expected to retire. The study considered a more flexible approach beyond the current legacy systems, which typically take a decade or more when combining design, acquisition, and launch.

A study by Maier et al. (2021) was a comprehensive consideration of the future NOAA constellation of operational satellites. It investigated over 180 possible constellations with no preconceived notions of instruments, platforms, orbits, or other factors while considering a wide range of user needs and was required to fit within the expected NESDIS budget. The cost-effectiveness of each constellation was considered by plotting a metric of its value to users versus the estimated cost on an efficient frontier plot. Key to this analysis was the environmental data record (EDR) value model (EVM), which was developed by the Space Platform Requirements Working Group (SPRWG), a group of subject matter experts knowledgeable about satellite capabilities and user needs (Anthes et al. 2019). The EVM ranked 44 objectives in priority order for improvements of capabilities over the Program of Record (POR) in 2025, which was the expected constellation in 2025 (Maier et al. 2021, their Table 1). The 44 objectives included 19 terrestrial weather- and climate-related observations, 19 space weather observations, and six strategic objectives (such as compatibility with fixed budgets and assurance of all capabilities). The highest priorities for improvement over the POR in 2025 in weather- and climate-related observations for numerical weather prediction (NWP) were 1) three-dimensional (3D) winds, 2) radio occultation (RO) soundings, 3) microwave (MW) soundings, and 4) infrared (IR) soundings. In this paper, we extend this study by considering the relative merits of various alternatives for the thermodynamic soundings, RO, IR, and MW, using observing system simulation experiments (OSSEs). Though 3D winds were the highest priority for improvement from the SPRWG study, we were asked and funded by the NOAA to investigate the impact of variations in the IR, MW, and RO mix. Previous OSSEs have shown a high value of wind observations, but comparisons of various configurations of RO have not been as well studied. Therefore, the first experiments we conducted were for IR, MW, and RO.

Several complementary methods have been developed to determine quantitatively the value of different types of observations, including observing system experiments (OSEs) and model adjoint methods such as forecast sensitivity to observation impact (FSOI), OSSEs, and ensemble of data assimilations (EDA).

OSEs address modeling and observing systems capabilities as of a given time (Bouttier and Kelley 2001; Boukabara et al. 2016). They are data denial studies using real data that evaluate the impact of these data. For example, an OSE may compare numerical weather forecasts using all the current observations with forecasts in which one type of data is removed. The differences in forecast accuracy between both experiments indicate the impact of the data that were removed. OSEs also help optimize the use of observations in current modeling systems, i.e., enhanced data assimilation strategies, more realistic characterization of observations, and incorporation of existing observations not currently used.

Another valuable tool for assessing the value of current observations is the FSOI, which is a method of quantifying the relative impact of all assimilated observations in a model (Langland and Baker 2004). The technique computes the variation in the short-range (24 h) forecast errors due to the assimilated data. FSOI studies consistently show that MW, IR, and RO soundings are among the top five observational systems contributing to short-range NWP forecast accuracy (Cardinali and Healy 2014).

A technique used to evaluate one component of the impact of future observations is the EDA. This method combines real and additional simulated observations, and the impact of an observing system is quantified by the reduction of uncertainty ("spread" of the forecasts) in the ensemble after adding the new observing system (Tan et al. 2007). A positive impact of the new observations is indicated by a reduction in spread. For example, Harnisch et al. (2013) used the EDA method to investigate the impact of increasing the number of RO profiles from the current number of around 6000–10 000 to 128 000 profiles per day and found no saturation of impact up to this very large number.

Complementing OSEs in which the impact of different combinations of real observations is evaluated, OSSEs provide information on the impact of not-yet-existing observing capabilities (Atlas et al. 1985; Hoffman and Atlas 2016). OSSEs are modeling experiments used to quantitatively evaluate the potential benefits of additional or new observations. They differ from OSEs in that all the observations (current and proposed) are first simulated and then assimilated with realistic errors. The impact of one data type is investigated by comparing the difference in skill between a forecast experiment with all the observations and one without the proposed data type. A simulated truth using a realistic high-resolution model, called a nature run, is used to simulate all the observations and to quantify the skill of the experimental forecasts. OSSEs, though relatively expensive computationally, have the advantage that all aspects of forecasts can be evaluated, including various statistical quantitative measures of accuracy of the forecast temperature, pressure, winds, water vapor, and precipitation as well as the quality and frequency of forecasts of extreme events such as tropical cyclones. They also provide complete four-dimensional datasets that can be analyzed to determine the scientific reasons for the differing impact of various observations. Finally, OSSEs can also be used to optimize the assimilation of observational data in numerical weather prediction models.

OSSEs can also be used to quantitatively investigate the optimum deployment of observing systems. While the NSOSA study has provided high-level guidance into the relative merits of different constellations, OSSEs may be used to analyze trade-offs among details such as spatial coverage, horizontal resolution, frequency, latency, accuracy, and data redundancy, as well as optimize data assimilation and modeling strategies. Privé et al. (2022) investigated the impact of the increased number of RO observations from a control value based on 2015 data availability (approximately 3300 profiles per day) to 100 000 (100K) profiles per day, in experiments of 25K, 50K, 75K, and 100K profiles per day. They found the greatest increase in forecast skill with the first increment (the 25K experiment) but a continued increase in positive impact up to 100K profiles per day, indicating no saturation in improvement at 100K per day, in agreement with Harnisch et al. (2013).

The Weather Research and Forecasting Innovation Act of 2017 (H.R. 353, Public Law 115–25) mandates the NOAA to undertake OSSEs to assess the value and benefits of observing capabilities and systems (Zeng et al. 2020). OSSEs must be conducted before 1) the acquisition of major government-owned or government-leased operational observing systems with a life cycle cost of more than $500 million and 2) the purchase of any major new commercially provided data with a life cycle cost of more than $500 million. As an example, the United States mandated the NOAA to conduct OSSEs for RO observations in support of the originally planned follow-on mission to Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC; Anthes et al. 2008; Cucurull et al. 2018; Mueller et al. 2020).

In this article, we describe some of the OSSEs conducted at the NOAA under the Quantitative Observing System Assessment Program (QOSAP; Zheng et al. 2020) to provide additional information on the design of proposed observing systems in support of the NSOSA study. Because of their importance to NWP and high priorities in the NSOSA study, we investigate different numbers and configurations of MW, IR, and RO atmospheric thermodynamic soundings. For MW and IR, we simulate and assimilate radiances, while for RO, we simulate and assimilate bending angles. These configurations are based on augmentations of the current operational satellite constellation, i.e., changes in the number of polar orbiters for MW and IR observations and the number of RO observations, globally distributed. The OSSEs undertaken in this study correspond approximately to several levels of performance identified by the SPRWG (Anthes et al. 2019). These levels of performance include the satellite configuration and expected capabilities in 2025 (the Program of Record 2025), a study threshold level below which the observations have no value (ST), an expected level (EXP), and a level of maximum effective impact (ME).

This study contributes to the understanding of how future operational satellites should be designed and configured in space to give maximum benefit to users of NWP, which includes many applications to operational decision-making and scientific research. This work complements the NSOSA, SPRWG, and other studies. However, like all studies, it has limitations. Common to all OSSEs, there is a limitation to the number of combinations of observational variables that can be studied and the variation of their performance levels (horizontal and vertical resolution, accuracy, update rate, coverage of Earth, etc.). Thus, it would be prohibitively expensive to repeat the studies of the NSOSA and SPRWG in their entirety, which depended on the collective judgment of many subject matter experts, in addition to OSEs and OSSEs, and covered many applications other than NWP. OSSE results depend on the nature run (NR), which is not a perfect replica of nature, the model being used for the OSSEs, including its physics and data assimilation system, the generation of the synthetic observations and their errors, and many other factors. Because of these limitations, the relative differences in NWP accuracy between two different experiments are more likely to be more relevant than their absolute differences. Furthermore, OSSEs conducted in this study pertain to NWP and therefore have the greatest relevance to applications that depend on NWP, such as forecasting beyond approximately one day and the myriad of applications that depend on these forecasts. They are also relevant to reanalyses, which are extremely useful for climate and research studies. At the present time, they are not as relevant to the many important short-range applications (minutes to hours) that depend on nowcasting, such as warnings of severe weather, and are not applicable at all to space weather applications. For all these reasons, there are limitations and uncertainties in the results and conclusions. In other words, OSSEs are powerful tools that can be used to provide guidance to the design of future space observing systems, but they should be considered cautiously and in conjunction with other studies and the judgment of subject matter experts, including users, scientists, and engineers.

2. Global OSSE system

The global OSSE used in this study begins with an NR generated by a state-of-the-art atmospheric model, the 9-km ECMWF forecast model (ECO1280; Hoffman et al. 2019). The ECO1280 was based on the TCo1279L configuration of the ECMWF Integrated Forecast System (IFS) model cycle 43r1, which was the operational configuration at ECMWF between November 2016 and July 2017. The model is a hydrostatic hybrid vertical sigma-pressure coordinate system with 137 levels and a top of ~80 km. The NR simulation extended over 14 months, from 30 September 2015 to 30 November 2016. Model variables were output on an hourly basis for the first month of the NR and every 3 h afterward. The NR was validated against a variety of variables and phenomena using the fifth major global reanalysis produced by ECMWF (ERA5) from 1979 to 2015. As an example, Fig. 1 shows the time average of the zonal mean of the wind and temperature fields for both the ECO1280 and ERA5 for July 2016. Similar results were found for water vapor and for horizontal maps at different levels (not shown).

The QOSAP consolidated observing system simulator (COSS) package was used to generate synthetic observations from the NR. This simulator has the capability to simulate conventional observations, RO bending angle and refractivity profiles, and MW and IR satellite radiances. The Community Radiative Transfer Model (CRTM; Han et al. 2006) code was slightly modified to enable the simulation of radiances under cloudy conditions. COSS has been recently upgraded to also simulate Doppler wind lidar observations, and the generation of three-dimensional passive atmospheric vector winds by tracking moisture features is an ongoing work. Explicit errors, including both standard deviation and systematic (bias) errors when needed, were added to the perfect simulated observations according to the differences in statistics between the real world and the simulated OSSE world (Errico et al. 2013). For example, the RO observation error adjustment for *MetOp-B* is shown in Fig. 2. In the figure, the incremental bending angles [defined as $(O - B)/B$, where O is the observation and B is the model value computed from the forward model] are too low with the use of perfect observations. When random errors are added to the perfect observations according to the error variance values observed in forecasts with real observations, these incremental bending angle differences match those of the real world much more closely. Following this procedure, error-added observations were simulated for 2 months, June and July 2016, for the observing configuration that existed in June–July 2020. The NR field variables were spatially and temporally interpolated to the location of the observations for these observation simulations.

The NOAA Finite Volume Cubed-Sphere (FV3) Global Forecast System (GFS) operational in May 2021 (GFSv16.1.1) was used as the global data assimilation and forecast system in the OSSEs. This code configuration was slightly modified to include RO bug fixes that were implemented operationally soon after (September 2021, GFSv16.1.4). The OSSE proceeds by running data assimilation and forecast experiments with and without the new observations, and the analyses and forecasts are verified against the NR. Experiments ran from 1 June to 15 July, with the verification time period conducted for 15 June–15 July.

We validated the OSSE system for a full control configuration (i.e., with all the observations). The real-world predictability in June–July 2020 and the OSSE system predictability for ECO1280 June–July 2016 were verified against self-analyses and compared. All of the metrics shown in this paper were checked. Overall, the OSSE system generally has lower predictability than the real world in the Southern Hemisphere and the tropics. In the Northern Hemisphere, the OSSE system generally has comparable predictability. While the predictability may be slightly lower than the real world, the overall OSSE system performed well, and the difference in impacts of the different observations tested in the experiments should be realistic.

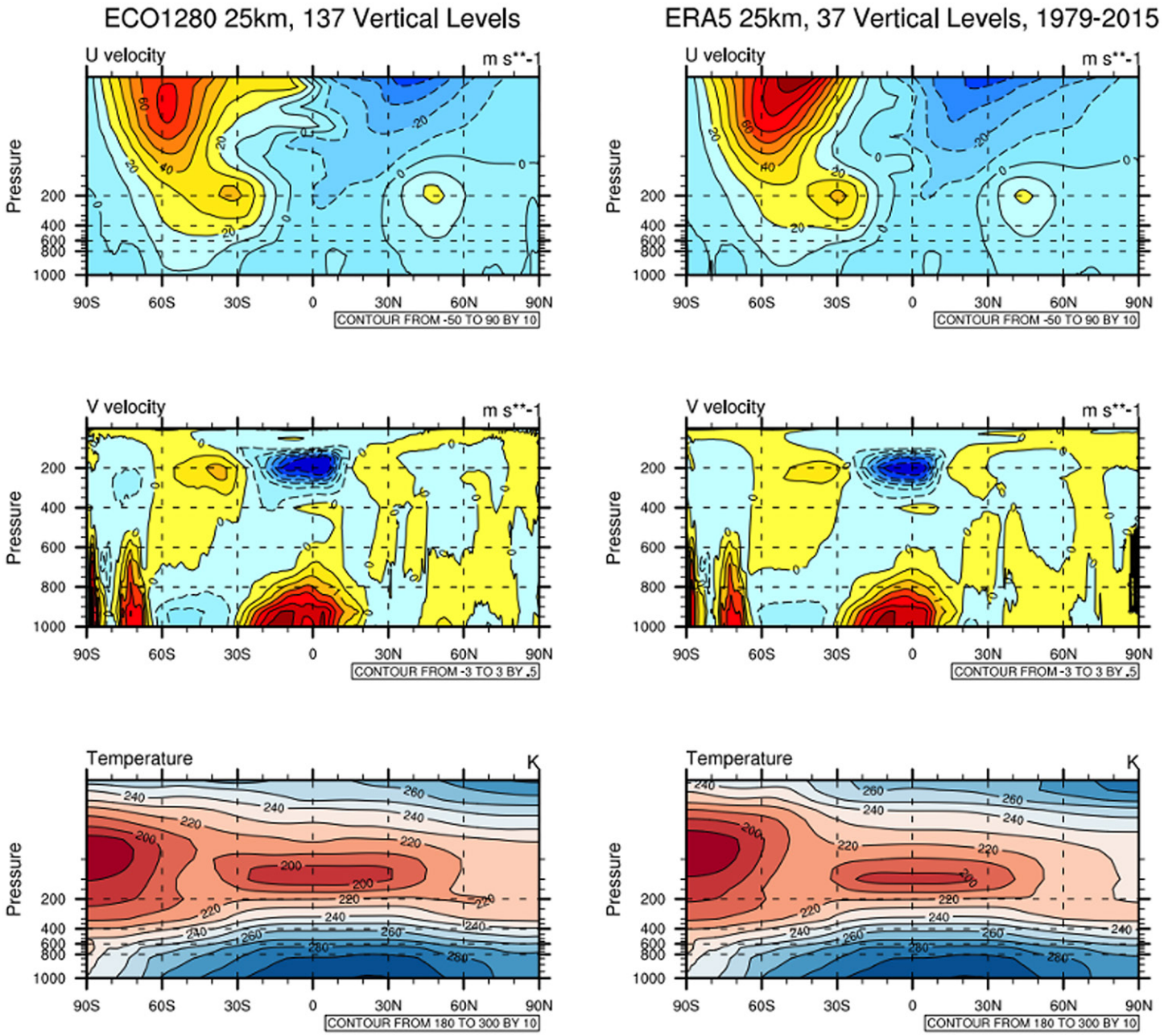

Fig. 1. Comparison of the (left) NR (ECO1280) with (right) ERA5 for Jul 2016. Monthly averages of the zonal mean. (top) The *u* component of the wind (m s^{-1}), (middle) v component of the wind (m s^{-1}), and (bottom) temperature (K) are shown.

3. Performance levels of simulated observations

As summarized earlier, the SPRWG study identified three performance levels (ST, EXP, and ME) of each observational system as well as the Program of Record for 2025 (POR2025). Additional details for each of these levels can be found in Anthes et al. (2019) and Maier et al. (2021). When assessing the impact in NWP forecasts of using observations with different performance levels, our experiments roughly correspond to several of the SPRWG performance levels with respect to the number of observations. However, the experiments consisted of relatively simple enhancements of existing sounding configurations, such as increasing the number of IR/MW orbits, increasing the IR/MW horizontal resolutions along the track of each satellite, and increasing the number of global RO profiles. We conducted a few tests where

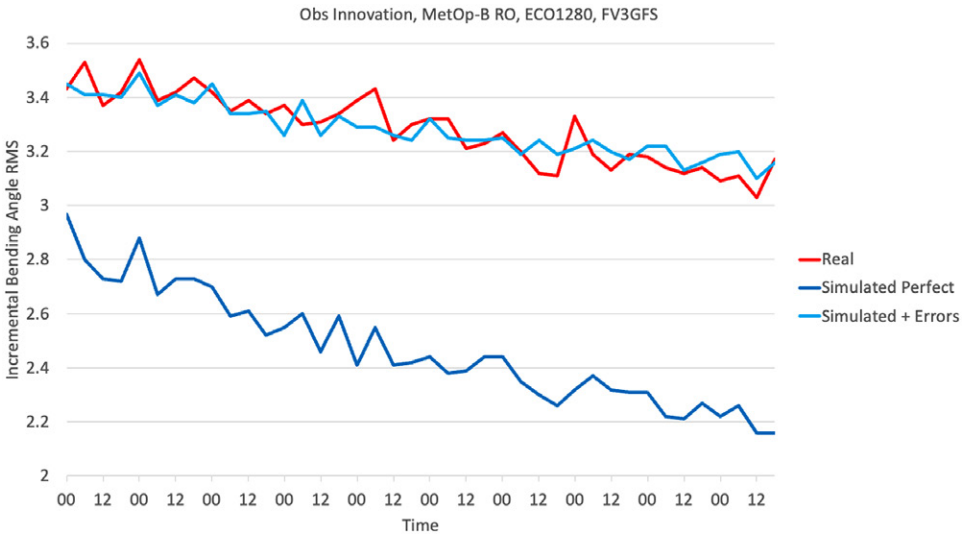

Fig. 2. RMS values of the normalized differences between *MetOp-B* observations (*O*) and their model counterpart (*B*) in incremental bending angle, (*O* – *B*)/*B*, from 0000 UTC 30 Sep 2020 to 1200 UTC 9 Oct 2020. Values are averaged over the entire model domain and plotted every 6 h.

the IR/MW observation errors were modified according to some of the SPRWG performance levels. However, the sensitivity of NWP forecasts to these modified errors was negligible when the representativeness error was added in the data assimilation system, and the results are not discussed here.

4. Minimum (threshold) and POR sounding capabilities

First, we evaluated the differences in NWP skill between the ST and POR2025 levels of performance. This was done to confirm and also quantify the expected higher skill of POR2025 versus a satellite configuration with limited sounding capabilities. For this purpose, we ran two experiments. A revisit rate of 12 h for MW and IR soundings was used in the ST experiment (EXP1). This corresponds to the time a single polar-orbiting satellite in a sun-synchronous orbit (such as *JPSS-1*) passes over the same point on the equator. Simulations based on IASI on *MetOp-B* quality were used as a proxy for the IR observations and the Microwave Humidity Sounder (MHS) and Advanced Microwave Sounding Unit-A (AMSU-A) on *MetOp-B* for the MW observations. The horizontal spatial resolution along the track was set to 36 km for both IR and MW observations (we kept one of every four observations along the track of the original IR and MW simulated observations at the finest 9-km spatial resolution), while the vertical resolution was driven by the channels currently being assimilated. Note that in this paper, we define the horizontal resolution of the IR and MW observations as the average horizontal distance between observations along the track of the satellite, rather than the average distance between observations in a uniformly distributed global network of observations as was done in the SPRWG study. The ST performance level also includes a total of 5000 RO globally distributed profiles per day, based on the geographical location of COSMIC-1, *TanDEM-X*, *TerraSAR-X*, and MetOp real observations. Finally, assumed errors in the variational assimilation were not changed for any of the observations from those prescribed in the OSSE system.

The revisit rate was increased from 12 to 6 h in EXP2 (the POR2025 experiment), which required a second polar orbiter. This was done by adding simulated Cross-Track Infrared Sounder (CrIS) observations from *NOAA-20* as the second IR sounder, and Advanced Technology Microwave Sounder (ATMS) observations on NPP as the second MW sounder. The number of RO profiles per day in EXP2 was increased from 5000 to 8000. The 8000 RO profiles per day included simulations of COSMIC-2 and *MetOp-A*/*B* observations only. No additional

Table 1. Summary of the experiments conducted in this study to investigate sounding quantity and orbital trade-offs. All experiments ran from 1 Jun to 15 Jul, with the verification time period conducted for 15 Jun–15 Jul.

Experiment	MW/IR revisit rate (No. of polar orbiters)	MW/IR horizontal resolution along the track of the satellite	RO profiles per day distributed globally	Key results from OSSEs
1 (ST level)	12 h (1)	36 km	5K	Lowest performance by all metrics
2 (POR2025)	6 h (2)	36 km	8K	Significantly higher performance than ST
3	12 h (1)	36 km	50K (ME level)	Significantly higher performance than EXP1 (ST) by all metrics. Similar to POR2025 in extratropics, better than POR2025 in tropics
4	12 h (1)	9 km MW 36 km IR	5K	No significant change from EXP1 (ST)
5	1 h (12); ME level	36 km	5K	Highest performance by all metrics, likely the highest cost because it requires 12 polar orbiters
6	3 h (4)	18 km	20K (EXP level)	Similar to EXP3 in the tropics, closer to EXP5 in the extratropics. Overall, perhaps the most cost-effective
7 (MWIR 3 orbits) IR and MW instruments on each satellite)	4 h (3)	36 km	5K	Significant improvements over ST level
8 (MWIR 6 orbits) IR on 3 orbits and MW on other 3 orbits	2 h (6)	36 km	5K	Significant improvements over ST level. A minor improvement over EXP7 from disaggregating IR and MW sounders

changes were made to ST in configuring the POR2025 experiment. Table 1 summarizes the experiments conducted in this study. The number of conventional (nonsatellite) and satellite data for each experiment is shown in Table 2.

The version of GFS available in 2021 at a research resolution was used to run the OSSEs in both experiments. Deterministic forecasts were run at ~25-km horizontal resolution (~13 km in the operational suite), while the analyses and ensemble members ran at ~50-km resolution (~25 km in the operational suite). As in the operational configuration, 127 vertical levels were used. Twenty ensemble members (80 in the operational configuration) were used in each experiment. Eight-day forecasts were initialized at 0000 UTC.

Figures 3a and 3b show the anomaly correlation skill for EXP1 (ST) and EXP2 (POR2025) for the 500-hPa geopotential heights for the Northern Hemisphere (NH, 20°–80°N) and the Southern Hemisphere (SH, 20°–80°S), respectively. As expected, POR2025 outperforms ST in both hemispheres, and differences are statistically significant for most lead times in the NH, and all lead times in the SH. Differences between POR2025 and ST are larger in the SH, probably due to the lower number of available conventional observations to constrain the system and thus increasing the benefits of using a larger number of MW, IR, and RO observations.

The root-mean-square (RMS) wind errors also decrease with the use of a larger number of satellite radiances and RO observations. The benefits extend to upper- and lower-level winds,

Table 2. Observation counts in thousands for the experiments in this study. The counts for the conventional observations are the same in all experiments.

Observation type	EXP1	Observation type	EXP1	EXP2	EXP3	EXP4	EXP5	EXP6	EXP7	EXP8
Surface pressure	51.9	**Atmospheric motion vector winds**	300.1	305.4	303.9	301.0	308.3	306.8	305.3	307.3
In situ wind	85.0	**MW radiances**	36.1	197.4	39.1	74.9	1839.2	601.6	367.1	366.9
Temperature	73.7	**IR radiances**	351.4	730.9	359.6	351.5	4425.1	1951.4	1094.7	1109.8
Specific humidity	20.5	**RO bending angles**	288.5	296.2	2779.2	288.5	289.9	1111.7	289.6	289.7

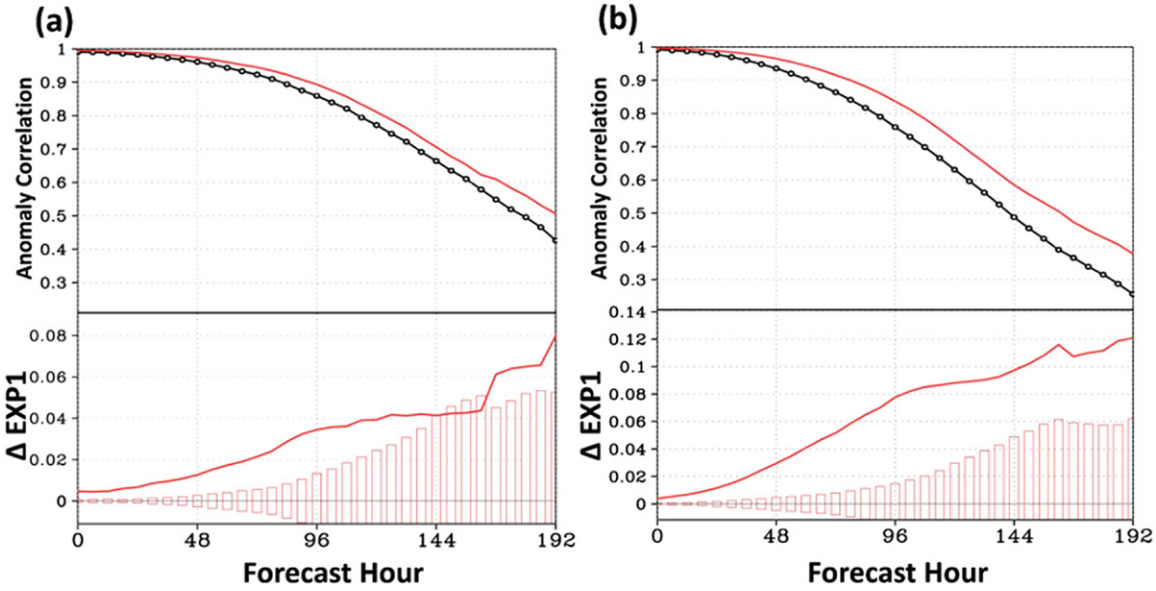

Fig. 3. Anomaly correlation score for the 500-hPa geopotential heights for EXP1 (ST; black) and EXP2 (POR2025; red) for (a) NH and (b) SH. The lower parts of each panel show differences between EXP2 and EXP1, with positive being an improvement. Bars show limits of statistical significance at the 95% confidence level; values above bars are statistically significant.

and the reduction in wind errors is larger in the SH (Figs. 4e,f) and tropics (TR, 20°S–20°N; Figs. 4c,d) than in the NH (Figs. 4a,b).

5. Evaluation of different sounding quantities

We ran four experiments to analyze some trade-offs on the number of IR, MW, and RO profiles and several IR and MW orbits. The skill of each experiment was compared to the performance of EXP1 (ST). We compared these experiments to the ST level of performance rather than the POR2025 because a ground rule of the NSOSA study was that all future NOAA architectures would meet at least the ST level of performance (Anthes et al. 2019). Experiment 3 (EXP3) increased the number of RO daily profiles from 5000 (ST) to 50 000 globally distributed profiles per day (the SPRWG ME level). Experiment 4 (EXP4) increased the resolution of MW soundings along the track of the satellite to 9 km, the maximum resolution available from the NR. Experiment 5 (EXP5) increased the number of IR and MW soundings to a 1-h revisit rate (ME level) while maintaining the number of RO profiles at ST values (5000). The higher revisit rate was achieved by adding 10 additional ATMS and 10 additional CrIS instruments to the POR2025 configuration (12 orbits total for each instrument type). Finally, experiment 6 (EXP6) increased the number and along-track horizontal resolution of IR and MW soundings to a 3-h revisit rate and 18-km resolution, and the number of RO daily profiles was set to 20 000. EXP6 followed the SPRWG expected (EXP) performance level requirements for RO and is the minimum number of RO observations recommended by the WMO International Working Group on Radio Occultation (IROWG 2022).

Figures 5a and 5b show the anomaly correlation score for the 500-hPa geopotential heights for the NH and SH as a function of the forecast hour. A notable result is the statistically significant increase in forecast skill by increasing the number of RO profiles from the ST level of 5000 to the ME level of 50 000 profiles per day while leaving all other observation types at the ST level (EXP3). This result is in general agreement with those of Harnisch et al. (2013) and Privé et al. (2022) and supports the IROWG recommendation of at least 20 000 RO profiles per day.

In both the NH and SH, little benefits are found by increasing the along-track horizontal resolution of MW from 36 km in the ST experiment to 9 km in EXP4, and the differences are not statistically significant. This small change indicates that the horizontal resolution along

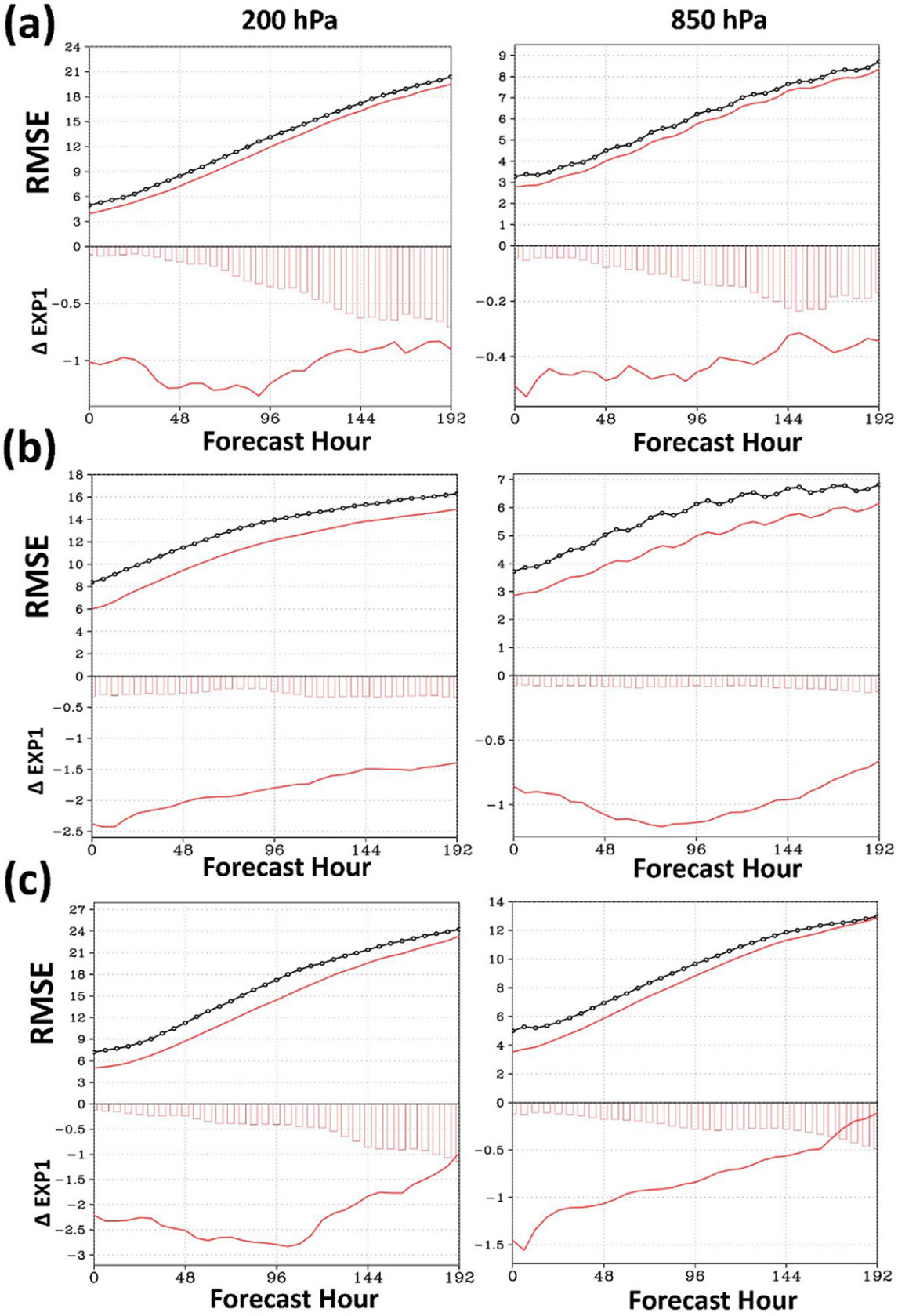

FIG. 4. RMS wind vector error (m s^{-1}) as a function of forecast lead time for EXP1 (ST; black) and EXP2 (POR2025; red) at (left) 200 and (right) 850 hPa for (a) NH, (b) TR, and (c) SH. The lower parts of each panel show differences with respect to EXP1 (ST). Bars show limits of statistical significance at the 95% confidence level; differences outside bars (solid red lines) are statistically significant.

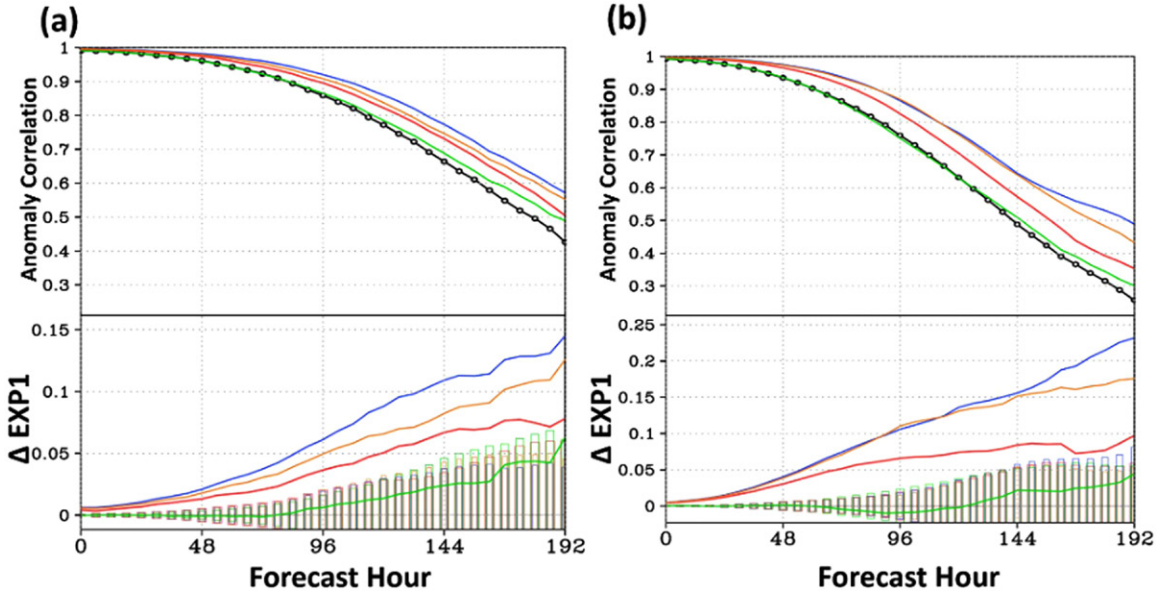

Fig. 5. Anomaly correlation score for the 500-hPa geopotential heights for EXP1 (ST; black), EXP3 (red), EXP4 (green), EXP5 (blue), and EXP6 (orange) for (a) NH and (b) SH. The lower parts of each panel show differences between each experiment and EXP1 (ST), with positive being an improvement. Bars show limits of statistical significance at the 95% confidence level; values above the bars are statistically significant.

the track of one satellite is not a large factor in the forecast skill. Adding more observations of the same type along the same track at the same time does not have much impact. Adding more independent observations with relatively uniform horizontal resolution globally would have a much larger impact because the medium- and large-scale atmospheric features would be better resolved.

Larger and statistically significant improvements over ST are found in the other experiments, with EXP5 providing the largest forecast skill, at the expense of an increase in the number of polar orbiters from 1 to 12. The large impact of a higher revisit rate is consistent with previous results, which indicate the importance of low latency (high refresh rate) of observations in data assimilation (McNally 2019; Noh et al. 2020; Wang et al. 2020; Casey and Cucurull 2022). EXP5 is followed in terms of overall improvement by EXP6 (differences between EXP5 and EXP6 in the SH are small) and EXP3. The differences between ST and any of the experiments are larger in the SH than in the NH, which is consistent with earlier studies showing that the benefits of assimilating satellite radiances and RO observations are larger in the SH. Similar results exist at other pressure levels (not shown).

EXP5 also has the smallest RMS temperature errors (Figs. 6a–c). At the 500-hPa pressure level, EXP4 shows a neutral impact over EXP1 (ST), as seen for the geopotential heights anomaly correlation score. A large reduction in RMS error is achieved with the increased number of RO daily profiles (50 000) in EXP3, and the reduction is greater in the TR and SH. This increase in forecast accuracy with an increasing number of RO observations is consistent with previous results (Harnisch et al. 2013; Lonitz et al. 2021; Privé et al. 2022.) In the NH and SH, a further decrease in temperature error is obtained in EXP6, while little difference between EXP6 and EXP3 is seen in the TR. For all latitudinal ranges, the lowest RMS temperature errors are achieved with EXP5, which has the largest revisit rate (1 h) for MW and IR. Similar results are seen at 200 hPa (Fig. 6a) and at lower pressure levels (850 hPa; Fig. 6c) although EXP6 shows an increased benefit over EXP3 in the TR. At this pressure level, the largest reduction in temperature error in EXP5 is found in the TR. The difference between EXP3 and EXP6 is larger than the difference between EXP6 and EXP5,

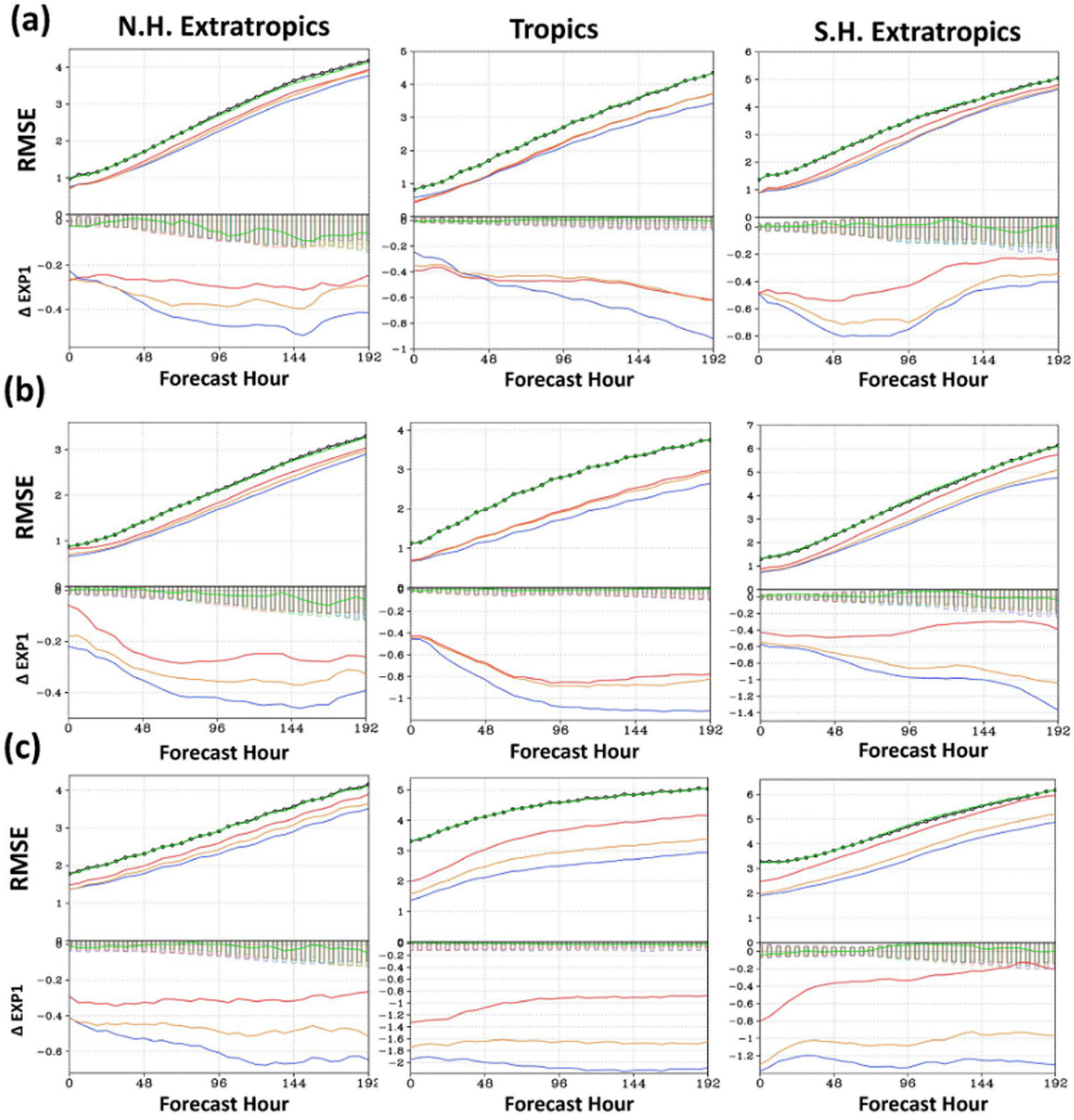

Fig. 6. RMS temperature errors as a function of the forecast lead time for EXP1 (ST black), EXP3 (red), EXP4 (green), EXP5 (blue), and EXP6 (orange) at (a) 200, (b) 500, and (c) 850 hPa. (left) NH, (middle) TR, and (right) SH. The lower parts of each panel show differences with respect to EXP1 (ST). Bars show limits of statistical significance at the 95% confidence level; values outside bars are statistically significant.

in particular in the SH, and this finding extends to all pressure vertical levels. This suggests that increasing the number of RO profiles and a modest increase in the revisit rate for satellite radiances beyond ST levels might be close to an optimum strategy among the options considered here.

The RMS wind errors at 200 and 850 hPa for the different experiments are shown in Figs. 7a–c. Overall, differences among the experiments for the reduction in wind errors are

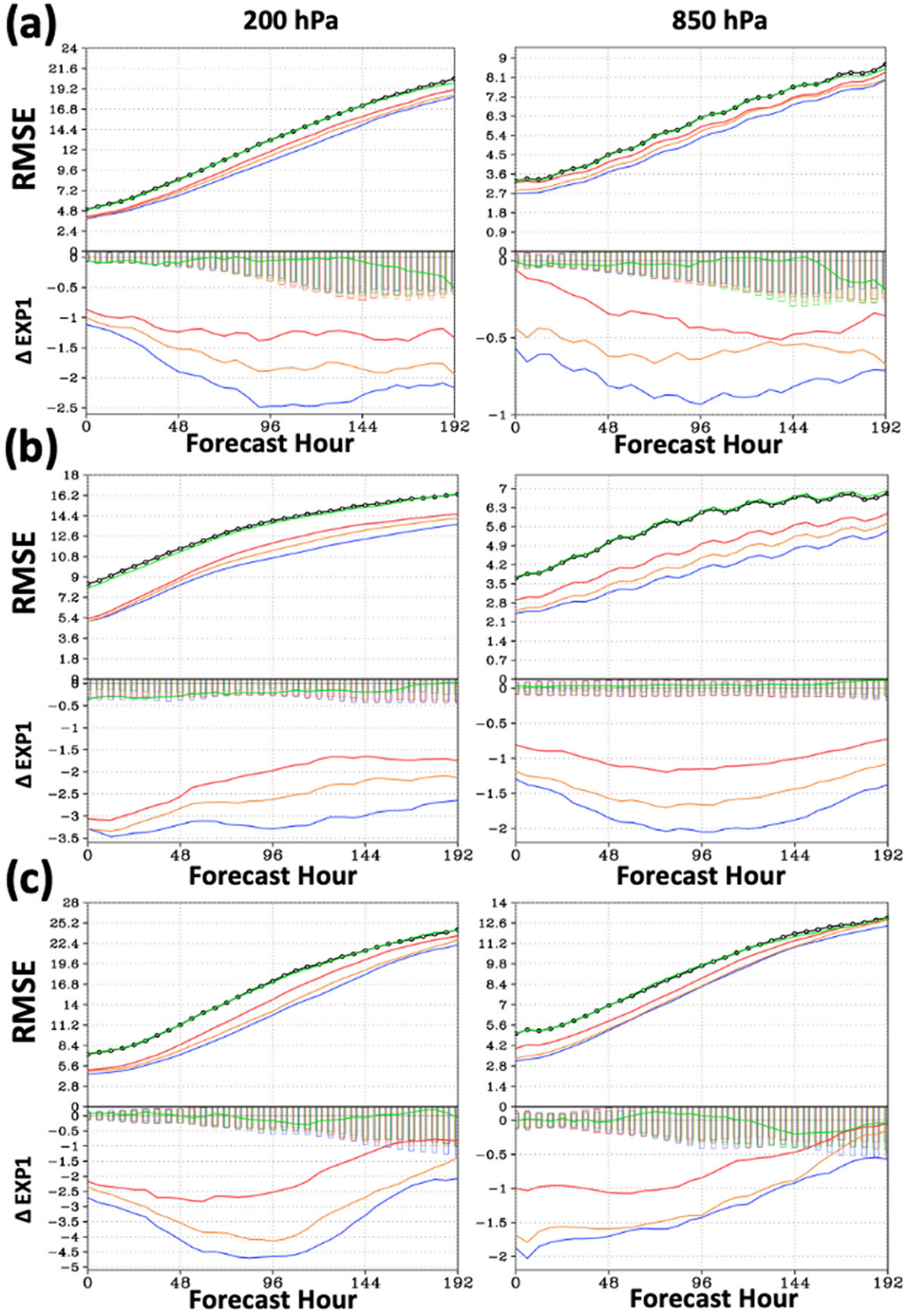

FIG. 7. Upper-level (200 hPa) and lower-level (850 hPa) RMS wind vector error (m s^{-1}) as a function of the forecast lead time for EXP1 (ST; black), EXP3 (red), EXP4 (green), EXP5 (blue), and EXP6 (orange) for (a) NH, (b) TR, and (c) SH. The lower parts of each panel show differences with respect to EXP1 (ST). Bars show limits of statistical significance at the 95% confidence level; values outside the bars are statistically significant.

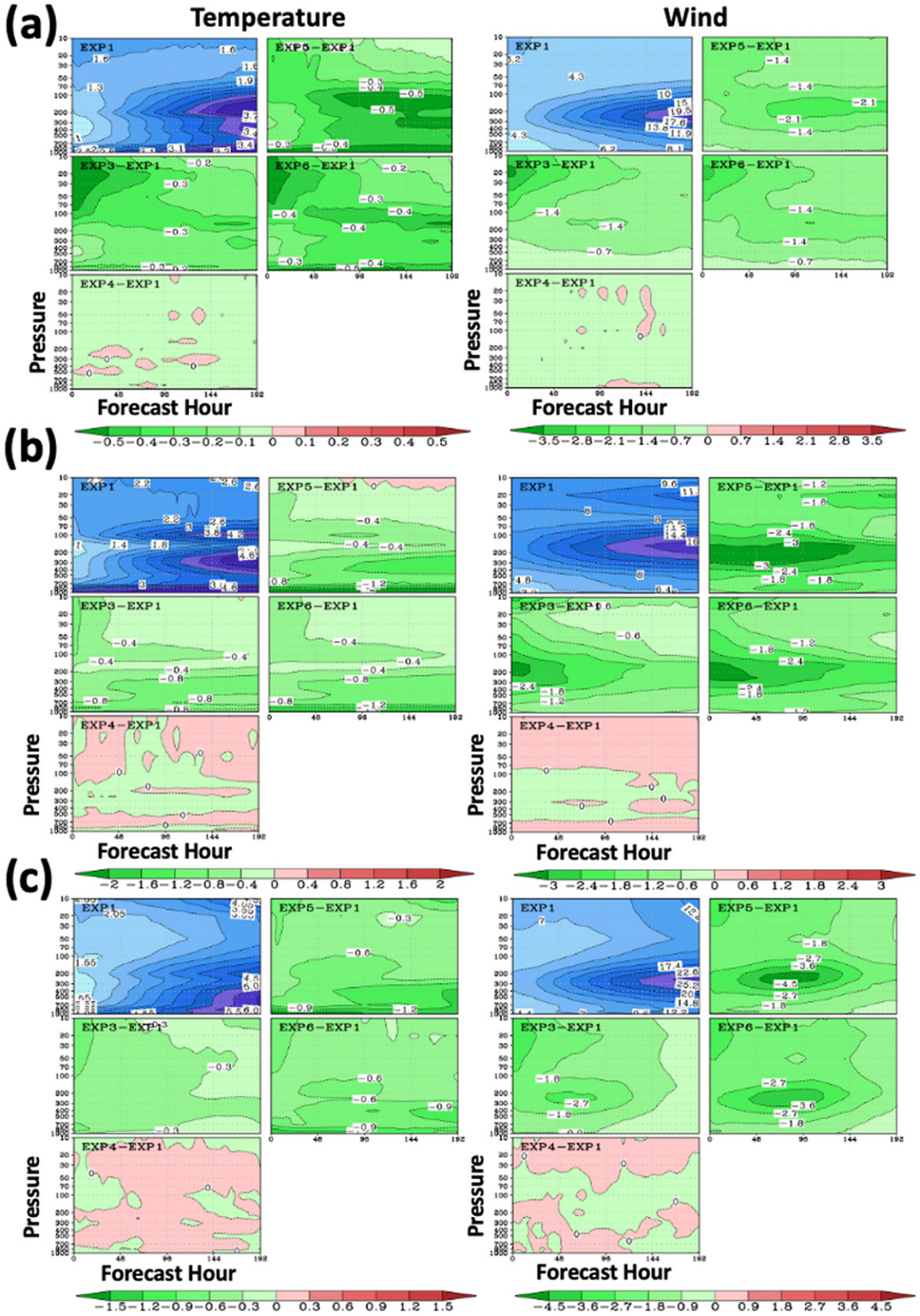

FIG. 8. Vertical cross section of the RMS temperature errors (K) and wind errors (m s^{-1}) as a function of the forecast lead time for (a) NH, (b) TR, and (c) SH. In each plot, the top left panel shows the RMS error of EXP1 (ST), while the other panels show the difference in RMS error between each of the experiments and EXP1 (ST): EXP3 (middle left), EXP4 (lower left), EXP5 (upper right), and EXP6 (middle right). Green shading indicates an improvement over EXP1 (ST). Note the different ranges of the color bar for each latitudinal region.

similar to those in the temperatures. No significant improvements beyond ST are found for EXP4, while significant improvements are found in EXP3, where a larger number of RO daily profiles (50 000) are assimilated. Further improvements exist in EXP6 when the number of RO profiles is increased to 20 000 profiles per day along with a higher revisit rate of 3 h and a finer along-track horizontal resolution of 18 km for satellite radiances. The largest reduction in RMS wind error is found in EXP5, with the larger revisit rate for satellite radiances (1 h). In the SH, lower-level wind errors are similar for EXP5 and EXP6.

Finally, vertical cross sections for the RMS temperature and wind vector errors are depicted in Figs. 8a–c. Overall, when compared to the skill of experiment ST, EXP4 shows a neutral impact, while EXP5 has the largest reduction in RMS error.

6. Aggregated versus disaggregated MW and IR instruments

QOSAP was also asked to conduct a trade-off analysis to quantify the impact of having three pairs of MW/IR sensors collocated in three evenly distributed orbits versus having the same six instruments disaggregated in six evenly distributed sun-synchronous orbits. In both scenarios, one of the orbits had to have a 1330 local equator crossing time (LECT). In EXP7 (MWIR 3 orbits), the locations of CrIS/ATMS on *NOAA-20* were used as a starting point and modified to generate orbits at 0530, 0930, and 1330 LECT. Additional orbits were added in EXP8 (MWIR 6 orbits) to alternate IR and MW instruments. The 1330 LECT was chosen to have an IR sensor. As an example, the spatial coverage for EXP7 and EXP8 at 0000 UTC 15 June 2016 is shown in Figs. 9a and 9b.

Both experiments were validated against EXP1 (ST) for consistency with previous experiments. However, we note that EXP1 (ST) had an IR/MW pair at 0930 LECT, so it is the relative performance between three orbits and six orbits that is more relevant here.

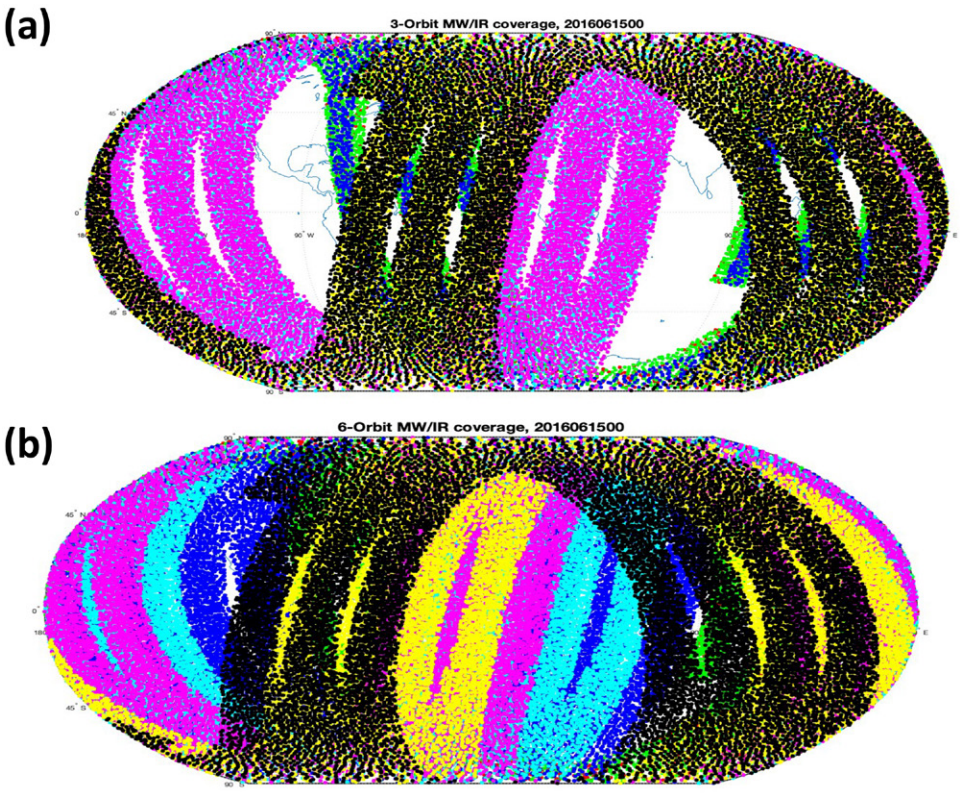

Fig. 9. Spatial coverage for IR and MW satellite radiances for the 3-orbit configuration (EXP7) and 6-orbit configuration (EXP8). Sensors plotted are as follows: AMSU-A (red), MHS (green), ATMS (cyan and yellow), CrIS (magenta and black), IASI (blue in EXP7), and CrIS (blue in EXP8).

RMS temperature errors as a function of the forecast lead time are shown in Figs. 10a–c for 200-, 500-, and 850-hPa vertical pressure levels. As expected, both configurations with enhanced satellite radiance coverage overperform that of ST, since the number of IR and MW instruments has increased from one each in EXP1 (ST) to six in EXP7 (MWIR 3 orbits) and EXP8 (MWIR 6 orbits). The benefits are larger in the SH at 200 hPa and in the TR at 850 hPa. As found with other satellite configurations investigated in this study, the impact is overall

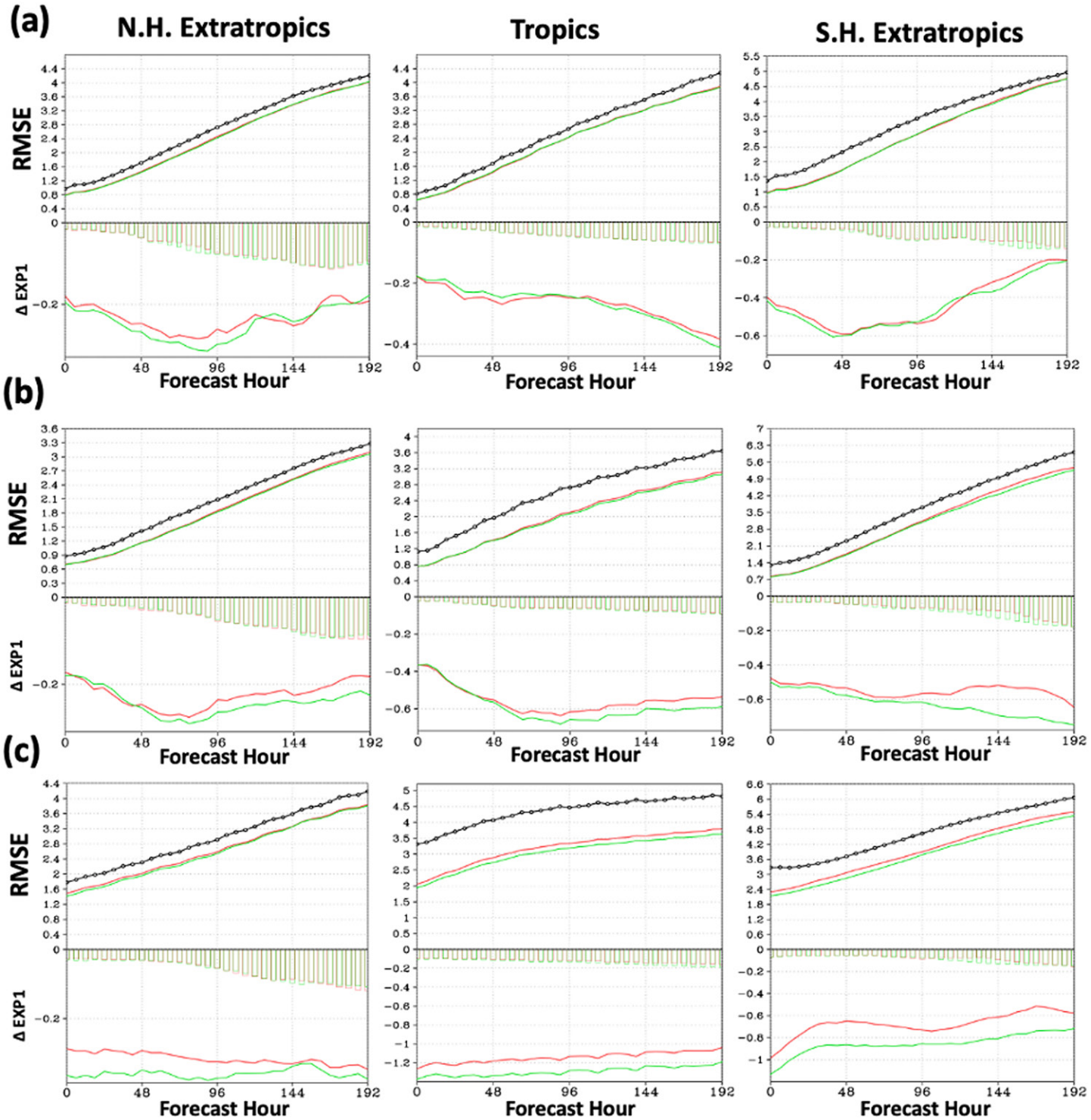

Fig. 10. RMS temperature error as a function of the forecast lead time for EXP1 (ST; black), EXP7 (MWIR 3 orbits, red), and EXP8 (MWIR 6 orbits, green) at (a) 200, (b) 500, and (c) 850 hPa. The lower parts of each panel show differences with respect to EXP1 (ST). Bars show limits of statistical significance at the 95% confidence level; values outside bars are statistically significant.

lower in the NH. Small differences are found when comparing three orbits and six orbits though some benefits from disaggregating IR and MW instruments are noticeable in the mid-lower troposphere in the TR and SH. This suggests a slight advantage in spreading out the IR and MW sounders so that the global spatial coverage is more uniform (fewer spatial gaps or higher horizontal resolution as defined by the SPRWG).

Small differences between aggregated and disaggregated MW/IR instruments are also found in terms of RMS wind error (Figs. 11a–c). Again, the overall added benefits of increasing the number of satellite radiances, regardless of the platform orbit configuration, is found in the SH and TR.

Finally, vertical cross sections of the RMS temperature and wind errors are plotted in Figs. 12a–c. Positive impact over the relatively poor baseline (ST) exists when the number of satellite radiances is increased, but the difference in impact of combining IR and MW instruments in the same platform or disaggregating the IR and MW into six separate platforms is small, with slightly better results obtained when the spatial coverage is increased by disaggregating the sensors across different evenly spaced orbits.

7. Discussion

The NSOSA study generated questions that could be partially addressed through OSSEs. Trade-off analyses have been conducted in this study to quantify the impact of potential scenarios for IR, MW, and RO sounding capabilities. More specifically, the quantity of IR, MW, and RO observations and several different orbits of the IR and MW sounders were investigated. The results from these OSSEs show that a large positive impact is found when the number of RO soundings per day is increased to 50 000 profiles per day globally, well beyond current values, while the other observational systems remain the same, a result consistent with previous studies. Increasing the revisit rate from 12 to 1 h for satellite radiance soundings produces the largest benefits, consistent with previous studies showing the large impact of low data latency (high refresh rate) but at the high cost of a large increase in the number of polar orbiters from 1 to 12. Together these two results indicate that increasing the number of RO observations can be used as a mitigation strategy for a limited MW/IR sounding revisit rate. Negligible benefits result from increasing the along-track horizontal resolution of MW satellite radiances.

The benefits from combining pairs of IR and MW sounders on the same platform versus evenly disaggregating the orbits, and so having IR and MW sensors in different platforms, showed little difference in impact in terms of verification skill. However, a small benefit was noted in the disaggregated orbits experiment. Potential advantages from a technological or cost perspective of having pairs of IR/MW instruments in the same orbit configuration were not taken into account in the design of these OSSEs.

Finally, we note that the results of this study contain the same uncertainties and limitations of all OSSEs as summarized in the introduction. These have been mitigated by using well-tested models and observation simulation methods. Despite these limitations, the main results have been evaluated against other independent studies and expert judgment and shown to be credible.

Results presented in this study focused on global numerical weather prediction applications, and the number of experiments was limited. In addition, it is important to emphasize that OSSEs provide quantitative information on relative NWP skill between configurations rather than on the absolute value of a given individual configuration. Finally, OSSEs need to be repeated and updated often, so enhancements in the NWP systems can be incorporated into the quantitative assessments in a timely manner.

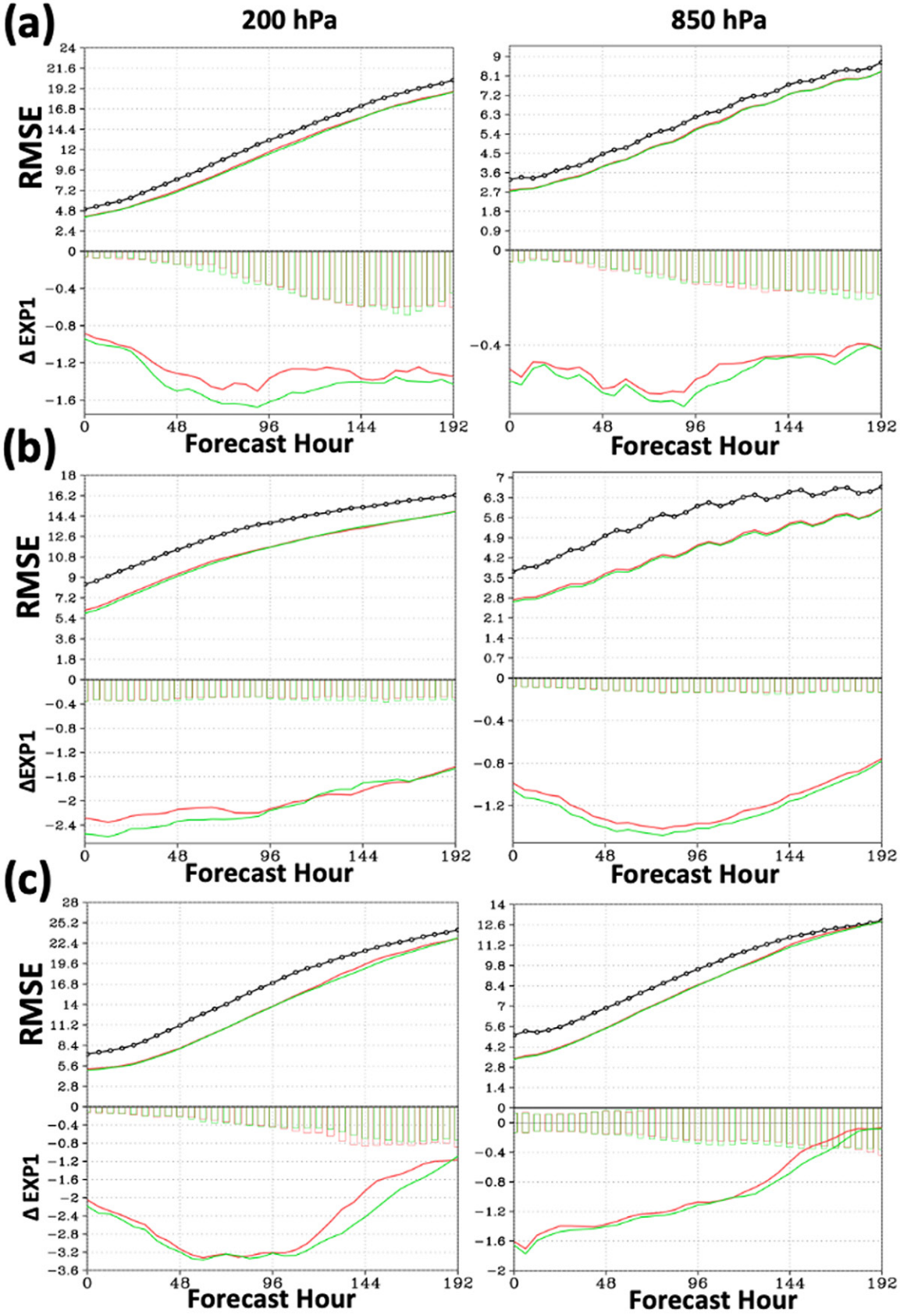

FIG. 11. Upper- and lower-level RMS wind vector error (m s^{-1}) as a function of forecast lead time for EXP1 (ST; black), EXP7 (MWIR 3 orbits, red), and EXP8 (MWIR 6 orbits, green) for (a) NH, (b) TR, and (c) SH. The lower parts of each panel show differences with respect to ST. Bars show limits of statistical significance at the 95% confidence level; values outside bars are statistically significant.

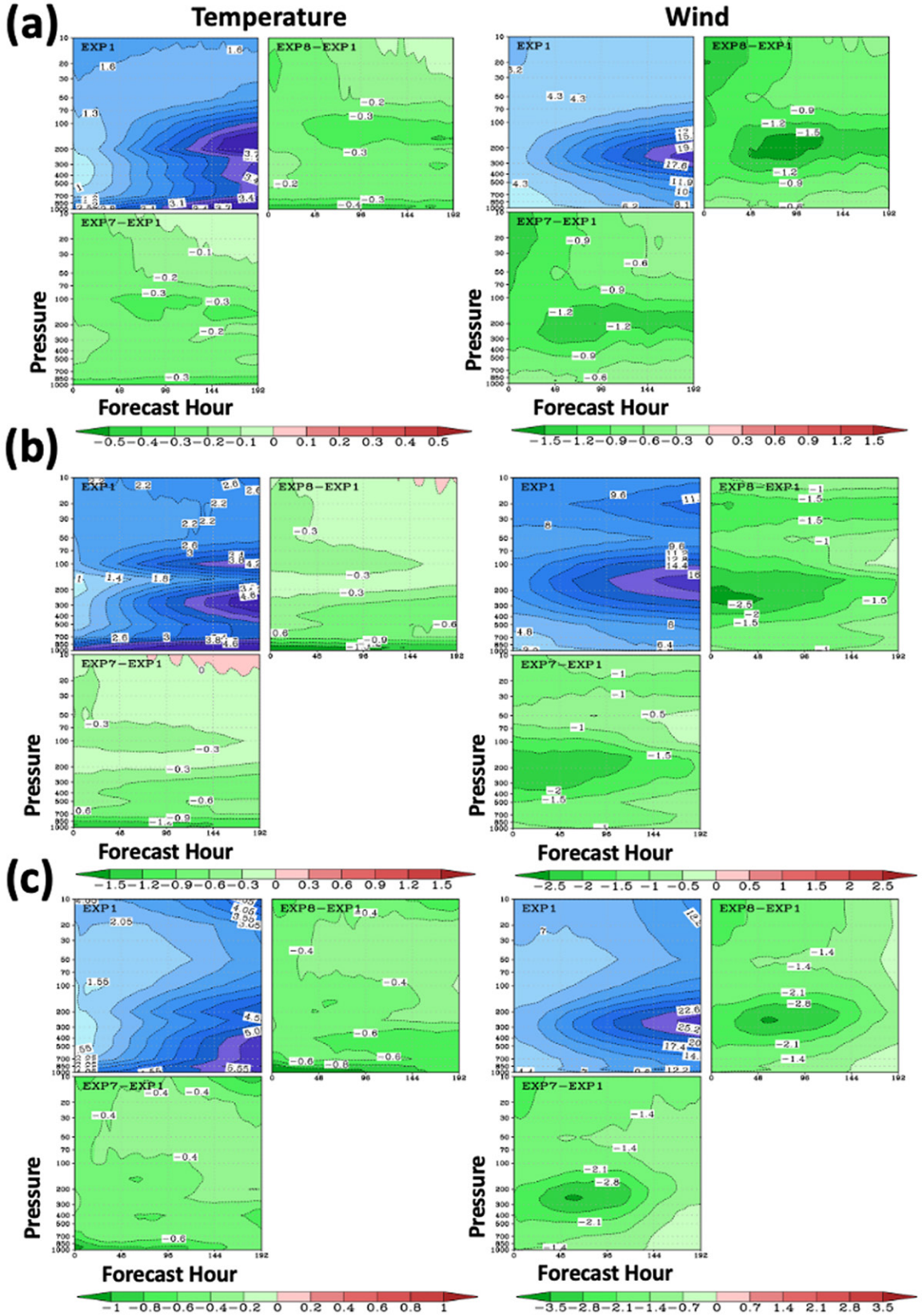

Fig. 12. Vertical cross section of the mean RMS temperature error (K) and wind error (m s^{-1}) as a function of the forecast lead time for (a) NH, (b) TR, and (c) SH. In each figure, the top left panel shows the RMS error of EXP1 (ST), while the other panels show the difference in RMS error between EXP7 (MWIR 3 orbits, lower left) and EXP8 (MWIR 6 orbits, upper right) and ST. Green shading represents an improvement over EXP1 (ST). Note the different ranges of the color bar for each latitudinal region.

Acknowledgments. We thank ECMWF for producing and CIRA/CSU for distributing the ECO1280 nature run. Ben Johnson from the Joint Center for Satellite Data Assimilation (JCSDA) helped us modify the CRTM version available in COSS to simulate satellite radiances under cloudy conditions. Two anonymous reviewers provided excellent comments. This work was supported by the NOAA Cooperative Agreement NA21OAR4310383 (Anthes) and NA21OAR4310473 (Casey, Mueller, and Vidal).

Data availability statement. The nature run (ECO1280) is available from the Cooperative Institute for Research in the Atmosphere at Colorado State University (CIRA/CSU). The ERA5 dataset (Hersbach et al. 2020) is available through the Copernicus Climate Data Center: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview. The OSSE dataset for all the experiments on which this paper is based is too large to be retained or publicly archived with available resources. Documentation and methods used to support this study are available in the references.

References

Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.*, **89**, 313–334, https://doi.org/10.1175/BAMS-89-3-313.

——, and Coauthors, 2019: Developing priority observational requirements from space using multi-attribute utility theory. *Bull. Amer. Meteor. Soc.*, **100**, 1753–1774, https://doi.org/10.1175/BAMS-D-18-0180.1.

Atlas, R., E. Kalnay, and M. Halem, 1985: Impact of satellite temperature sounding and wind data on numerical weather prediction. *Opt. Eng.*, **24**, 242341, https://doi.org/10.1117/12.7973481.

——, and Coauthors, 2015: Observing System Simulation Experiments (OSSEs) to evaluate the potential impact of an Optical Autocovariance Wind Lidar (OAWL) on numerical weather prediction. *J. Atmos. Oceanic Technol.*, **32**, 1593–1613, https://doi.org/10.1175/JTECH-D-15-0038.1.

Boukabara, S., K. Garrett, and V. Krishna-Kumar, 2016: Potential gaps in the satellite observing system coverage: Assessment of impact on NOAA's numerical prediction overall skills. *Mon. Wea. Rev.*, **144**, 2547–2563, https://doi.org/10.1175/MWR-D-16-0013.1.

Bouttier, F., and G. Kelley, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **127**, 1469–1488, https://doi.org/10.1002/qj.49712757419.

Cardinali, C., and S. Healy, 2014: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics. *Quart. J. Roy. Meteor. Soc.*, **140**, 2315–2320, https://doi.org/10.1002/qj.2300.

Casey, S. P. F., and L. Cucurull, 2022: The impact of data latency on operational global weather forecasting. *Wea. Forecasting*, **37**, 1211–1220, https://doi.org/10.1175/WAF-D-21-0149.1.

Cucurull, L., R. Atlas, R. Li, M. Mueller, and R. N. Hoffman, 2018: An observing system simulation experiment with a constellation of radio occultation satellites. *Mon. Wea. Rev.*, **146**, 4247–4259, https://doi.org/10.1175/MWR-D-18-0089.1.

Errico, R. M., R. Yang, N. Prive, K.-S. Tai, R. Todling, M. E. Sienkiewicz, and J. Guo, 2013: Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1162–1178, https://doi.org/10.1002/qj.2027.

Han, Y., P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, 2006: JCSDA Community Radiative Transfer Model (CRTM)—Version 1. NOAA Tech. Rep. NESDIS 122, 33 pp.

Harnisch, F., S. B. Healy, P. Bauer, and S. J. English, 2013: Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations. *Mon. Wea. Rev.*, **141**, 4395–4413, https://doi.org/10.1175/MWR-D-13-00098.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoffman, R. N., and R. Atlas, 2016: Future observing system simulation experiments. *Bull. Amer. Meteor. Soc.*, **97**, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00200.1.

——, S. Malardel, and T. Peevey, 2019: New 14-month forecast available for research. *ECMWF Newsletter*, No. 158, ECMWF, Reading, United Kingdom, 12–13, https://www.ecmwf.int/en/newsletter/158/news/new-14-month-forecast-available-research.

IROWG, 2022: Summary of the Ninth International Radio Occultation Workshop. IROWG, 25 pp., https://irowg.org/wpcms/wp-content/uploads/2023/04/IROWG9_Minutes_Summary.pdf.

Langland, R. H.., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56**, 189–201, https://doi.org/10.1111/j.1600-0870.2004.00056.x.

Lonitz, K., N. Bowler, E. Holm, and S. Healy, 2021: Assimilating Spire and COSMIC-2 data. *ECMWF Newsletter*, No. 169, ECMWF, Reading, United Kingdom, 25–32, https://www.ecmwf.int/en/newsletter/169/meteorology/assimilating-spire-and-cosmic-2-data-ifs.

Maier, M. W., and Coauthors, 2021: Architecting the future of weather satellites. *Bull. Amer. Meteor. Soc.*, **102**, E589–E610, https://doi.org/10.1175/BAMS-D-19-0258.1.

McNally, A. P., 2019: On the sensitivity of a 4D-Var analysis system to satellite observations located at different times within the assimilation window. *Quart. J. Roy. Meteor. Soc.*, **145**, 2806–2816, https://doi.org/10.1002/qj.3596.

Mueller, M. J., A. C. Kren, L. Cucurull, R. Hoffman, R. Atlas, and T. Peevey, 2020: Impact of refractivity profiles from a proposed GNSS-RO constellation on tropical cyclone forecasts in a global modeling system. *Mon. Wea. Rev.*, **148**, 3037–3057, https://doi.org/10.1175/MWR-D-19-0360.1.

NOAA, 2018a: The NOAA Satellite Observing System architecture study. Building a plan for NOAA's 21st Century Satellite Observing System, 36 pp., https://www.regulations.gov/document/NOAA-NESDIS-2018-0053-0002.

——, 2018b: Our satellites: Flyout charts. Accessed 1 October 2018, https://www.nesdis.noaa.gov/content/our-satellites.

Noh, Y.-C., A. H. N. Lim, H.-L. Huang, and M. D. Goldberg, 2020: Global forecast impact of low data latency infrared and microwave sounders observations from polar orbiting satellites. *Remote Sens.*, **12**, 2193, https://doi.org/10.3390/rs12142193.

Privé, N. C., R. M. Errico, and A. E. Akkraoui, 2022: Investigation of the potential saturation of information for the Global Navigation Satellite System radio occultation observations with an observing system simulation experiment. *Mon. Wea. Rev.*, **150**, 1293–1316, https://doi.org/10.1175/MWR-D-21-0230.1.

Tan, D. G. H., E. Andersson, M. Fisher, and L. Isaksen, 2007: Observing-system impact assessment using a data assimilation ensemble technique: Application to the ADM–Aeolus wind profiling mission. *Quart. J. Roy. Meteor. Soc.*, **133**, 381–390, https://doi.org/10.1002/qj.43.

Volz, S., M. W. Maier, and D. Di Pietro, 2016: The NOAA Satellite Observing System architecture study. *2016 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Beijing, China, Institute of Electrical and Electronics Engineers, 5518–5521, https://doi.org/10.1109/IGARSS.2016.7730439.

Wang, P., J. Li, and T. J. Schmit, 2020: The impact of low latency satellite sounder observations on local severe storm forecasts in regional NWP. *Sensors*, **20**, 650, https://doi.org/10.3390/s20030650.

Zeng, X., and Coauthors, 2020: Use of observing system simulation experiments in the United States. *Bull. Amer. Meteor. Soc.*, **101**, E1427–E1438, https://doi.org/10.1175/BAMS-D-19-0155.1.