2021 DTC Visitor Program Final Report:

Assessing the Rapid Refresh Forecast System data assimilation capability to represent an Amazon coastal squall line

DTC Host: DTC Data Assimilation team

Ivette Hernández Baños
Graduate Program in Meteorology, National Institute for Space Research, São José dos Campos, São Paulo, Brazil (ivette@ucar.edu)
Ph.D. Advisor: Luiz Fernando Sapucci

1 Background and introduction

The Amazon offers favorable conditions for the development of deep convection that is organized in mesoscale convective systems (MCS). In northern South America, MCS occurrence is related to Amazon coastal squall lines. Amazon coastal squall lines are observed in satellite imagery as discontinuous clusters of convective cells along the northern coast of South America, developed on the sea-breeze circulation, organized on the mesoscale, and considered one of the main rain-producing systems in the region (Garstang et al., 1994; Cohen; SILVA DIAS; Nobre, 1995). Because of the role the Amazon plays on the regional weather and climate, numerous studies have been conducted using observations derived from field campaigns held in the Amazon, meteorological satellites, *in situ* sources, reanalyses, and numerical models providing the scientific community with a better understanding of processes at various scales. In particular, the data collected during field campaigns, such as the Amazon Boundary Layer Experiment (ABLE) field campaign (Garstang et al., 1990), the Large-Scale Biosphere-Atmosphere Experiment (LBA) (Keller et al., 2004) and more recently the GoAmazon2014/5 experiment (Martin et al., 2016), have been invaluable in providing large amounts of data that otherwise would not be possible to obtain. The results from these field campaigns have greatly improved the knowledge and understanding of the atmospheric chemistry of the Amazon as well as the dynamics, environmental conditions, structure, rainfall characteristics, and life cycle of convective systems that occur in the Amazon basin. However, despite all these efforts, the deep convection and diurnal cycles of precipitation and convection in the Amazon region are not satisfactorily simulated by the models. As global numerical models increase the horizontal resolution and more numerical weather prediction (NWP) centers are concentrating their efforts on unified models, it is important to examine what is the current capability of regional and convection-allowing models to represent a wide variety of phenomena. Currently, the National Oceanic and Atmospheric Administration (NOAA) is transitioning toward a unified model in which the same dynamical core is used for all applications, from global to convective scale through the Unified Forecast System effort (UFS[1]). The Rapid Refresh Forecast System (RRFS), the UFS application for regional and convective scales, is intended to cover a similar domain as the operational Rapid Refresh (RAP), imposing the need to investigate the capability of the prototype

[1] ⟨https://ufscommunity.org/⟩

RRFS to represent convection over part of northern South America and the Atlantic Ocean. In particular, the initiation and development of Amazon coastal squall lines can give insight to the capabilities of RRFS over areas with low data density.

In this study, the data assimilation framework for the prototype RRFS is assessed and impacts on forecasts of an Amazon coastal squall line case study are investigated. Overall, this study examines to what extent the assimilation of few and sparse data can have a positive impact in the RRFS analyses over this region. Due to data availability issues along with a lack of severe weather reports in this region, a methodology is employed for the selection of a case study. This methodology follows Oliveira and Oyama (2015) in an attempt to create an objective algorithm to identify squall lines in the outputs of the tracking system ForTraCC (Vila et al., 2008). Once the case is selected, sensitivity tests using RRFS are performed. Different configurations in the Gridpoint Statistical Interpolation (GSI; Wu, Purser and Parrish (2002)) are tested, such as various ensemble background error covariance weights in hybrid analyses, supersaturation removal, the planetary boundary layer (PBL) pseudo-observations function, as well as varied observation types. Two physics suites are tested: one based on the Global Forecast System (GFS) version 15 physics (GFSv15) and a suite based on the High Resolution Rapid Refresh (HRRR) physics (RRFS_PHY_v1a). Forecasts are assessed using the Model Evaluation Tools (MET), which is the unified verification package that will be used by UFS applications (Brown et al., 2021). Although the area studied has low density data coverage, results show that large scale patterns are well captured in all experiments and the forecasts are improved when using data assimilation.

2 Methodology

2.1 Case study selection

Unlike in the mid-latitudes, seasons in the tropics are defined following rainfall frequency rather than temperature. The methodology proposed in Marengo et al. (2001), based on the computation of pentads (5-days average) of accumulated precipitation, is applied in this research. It focuses on determining the onset of the rainy (dry) season as the pentad with daily average precipitation greater (less) than 4.5 mm day^{-1} and this value remains above (below) 4.5 mm day^{-1} in the 6 to 8 subsequent pentads. Prior to pentads indicating the initial (ending) date, the values of pentads with daily average precipitation must be more (less) than 3.5 mm day^{-1} in the subsequent 6 to 8 pentads. This computation is performed for the area in the mouth of the Amazon River using the MERGE product from CPTEC (Rozante, 2017). Following this methodology, the dry season of 2020 near the mouth of the Amazon is first found in the pentad centered on 21 June and ends in the pentad centered on 29 October. This means that the onset of the 2020 dry season occurred inside the earliest limits suggested by Marengo et al. (2001), while the end occurred slightly earlier than the range considered in the climatological values. ForTraCC is then executed for the period between 21 June and 29 October 2020.

ForTraCC is an algorithm for tracking and predicting the morphological and radiative characteristics of convective systems using infrared channel images from GOES satellites (Vila et al.,

2008). Images from the GOES-16 infrared channel 13 with spatial resolution of 2 $\times$ 2 km and temporal resolution of 10 minutes are used for the tracking of the convective systems. Prior to the execution of ForTraCC, a quality control procedure is applied to the GOES imagery in which images with temperatures below 180 K are removed. Convective systems with initiation inside the limits of the Amazon basin are selected for further analyses (Fig. 1). The monthly spatial distribution for July, August, September, and October is presented in Fig. 1. The few days from June that are inside the dry season range are not considered in this monthly analysis. During the four months analyzed, the preferred region of initiation varies slightly from the north and northwest in July, to be more concentrated on the northwest in September, to spread toward the west and slightly to the central and southern region, to then completely spread toward the central and eastern Amazon with some spread over the south. Over the northern Amazon, there are convective systems initiating throughout the four months, however, July is the month with the most convective genesis. These results are in agreement with previous studies such as Cohen, SILVA DIAS and Nobre (1995).

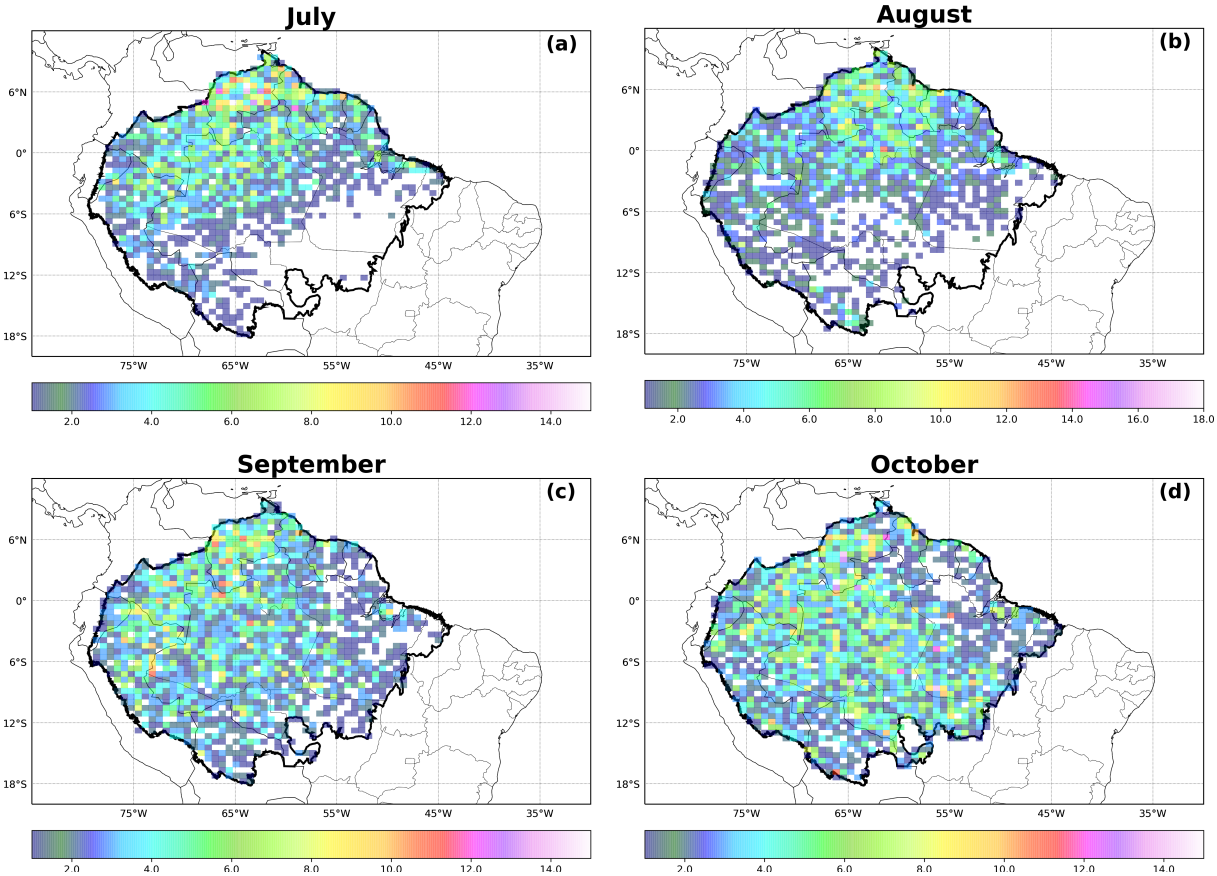

Figure 1: Spatial distribution of the preferred regions of genesis of convective systems for (a) July, (b) August, (c) September, and (d) October.

According to Cohen, SILVA DIAS and Nobre (1995), the formation region for Amazon coastal squall lines is between the latitudes 10°N and 5°S, south of the Intertropical Convergence Zone (ITCZ) during the dry season. The results shown in Fig. 1 confirm that finding and therefore determines the source region considered in this study. After it is verified that convective systems had genesis inside the source region, they are filtered according to their morphological

characteristics as proposed by Oliveira and Oyama (2015). The area, tilting, and eccentricity are examined in order to find tracked convective systems that are more linear and tilted. Figure 2 shows the positioning and extent of three tracked convective systems during its evolution making evident that this methodology is able to identify the main convective systems associated with cases of Amazon squall lines. The linear organization along the coastline as the system evolves is well identified by the adapted Oliveira and Oyama (2015) algorithm. However, there are two points that need to be considered. The first is that Amazon coastal squall lines can reach the synoptic scale with the form of a discontinuous or arc of discrete clusters of cells (Garstang et al., 1994), and second is that new cells can be developed as part of the squall line circulation (interaction between the updrafts and downdrafts), and be identified by ForTraCC as a separate system rather than as part of the whole system. Two examples are shown in Figs. 2b–c of convective systems tracked in ForTraCC for which hour and longitude genesis are very close to the main system in Fig. 2a. With a visual analysis of the satellite imagery, it is possible to identify that these other systems are part of the same Amazon coastal squall line that reached synoptic dimensions, with new convection being developed at different times. Therefore, with this current methodology, a subjective analysis is still needed to complement the information obtained with ForTraCC.

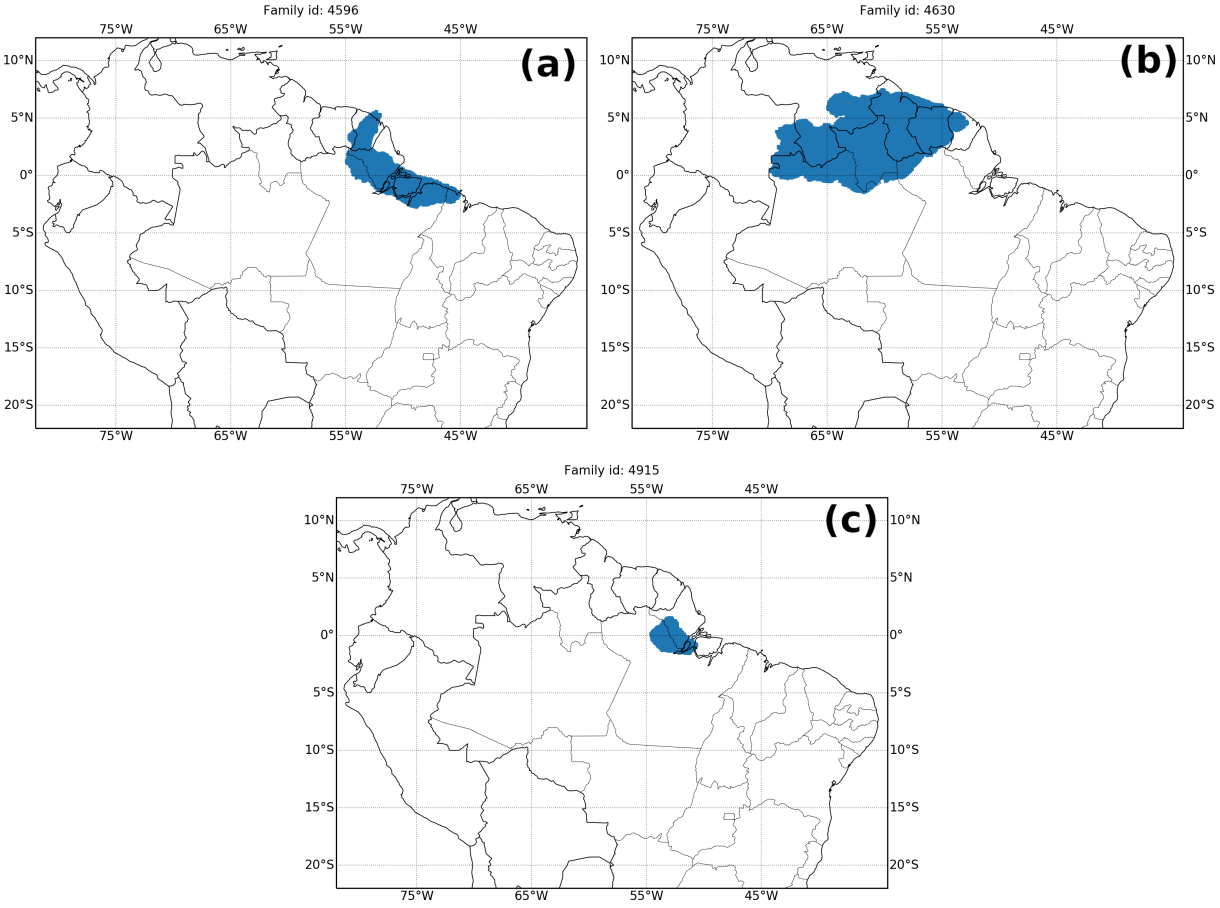

Figure 2: Three tracked convective systems identified by ForTracc using GOES-16 infrared observations which are associated with an Amazon coastal squall line. The blue clusters represent the positioning and extent of the tracked system during its evolution. Panels present the evolution of (a) the convective system identified with number 4596, (b) the convective system identified with number 4630, and (c) the convective system identified with number 4915.

2.2 Case overview

The Amazon coastal squall line case previously presented is selected for the numerical simulations of this study. Figure 3 shows the initiation and evolution of the convection associated with this system. Some of the initial cells are observed between French Guiana, the state of Macapa, and northern Para in Brazil. Between 19:00 UTC and 21:00 UTC, more cells develop and a line of discontinuous convective storms is observed along the coast in the satellite images. The strongest convection occurs at 23:00 UTC, and it is also observed that the stratiform part has widened. This cloud band, with a northwest-southeast orientation, slowly propagates inland from northeast to southwest. Some inner clusters have very cold cloud tops (white shade indicates temperatures lower than -75°K) indicating possible overshootings and that deep convection is occurring. At 01:00 UTC the line has propagated farther inland and some clusters have decreased their intensity while some are still very deep. 4 hours later, at 05:00 UTC, the squall line is farther into the continent and later seems to merge with other convective systems in the area. The convection associated with the squall line continues to propagate toward the southwest Amazon, but it loses its linear characteristics. Over the Atlantic ocean, there is strong

convection associated with the ITCZ.

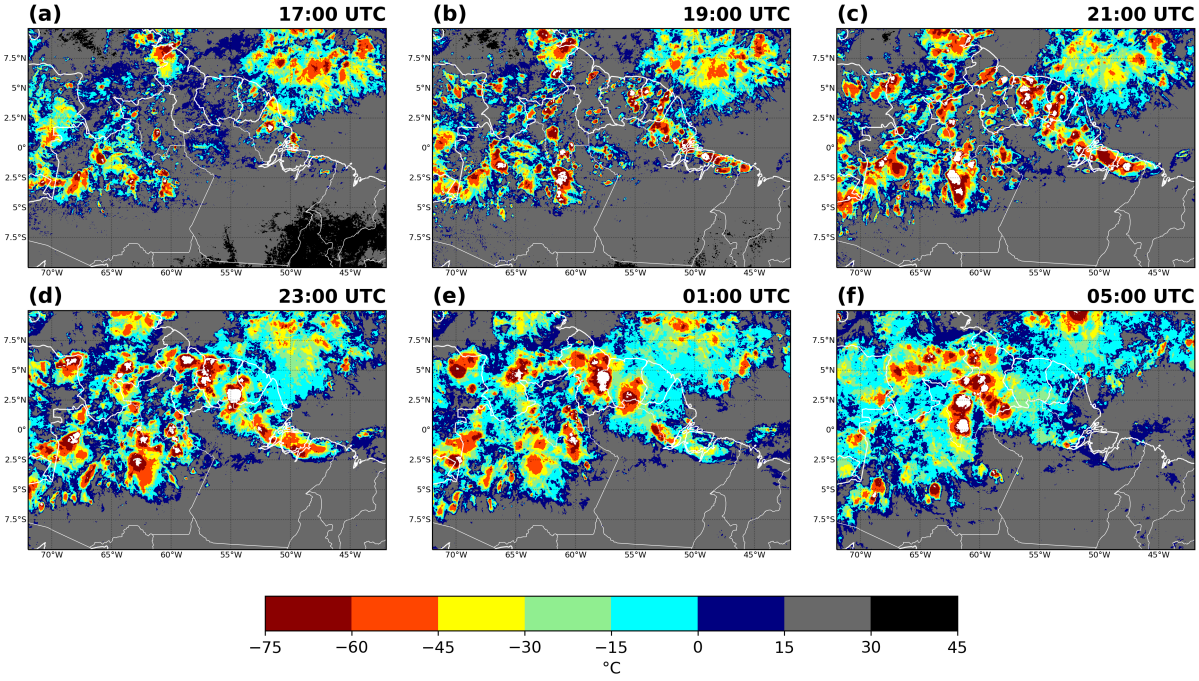

Figure 3: Brightness temperature from the GOES-16 infrared 13 channel from 17:00 UTC on 5 July 2020 through 01:00 UTC on 6 July 2020 every 2 hours (a, b, c, d, and e) and at 05:00 UTC on 6 July 2020 (f).

2.3 Setup of the experiments

For the simulation of this case, a domain is configured using 1200 $\times$ 700 grid cells centered on the coordinate point at 1°N and 57°W with 3 km horizontal grid spacing and 64 vertical layers. The domain covers the area of formation and propagation of the squall line through the western Amazon. All simulations start at 00:00 UTC on 4 July 2020 and run 3-hourly cycles until 21:00 UTC on 6 July 2020. Analyses and forecasts from the GFS at 0.25° resolution are used to generate the initial and lateral boundary conditions (ICs and LBCs) for the limited area model capability based on the Finite Volume Cubed-Sphere (FV3) dynamical core (FV3LAM). Global Data Assimilation System (GDAS) observations are assimilated in each experiment along with Global Navigation Satellite System (GNSS) radio occultation (GNSSRO) bending angles and satellite radiances. The time window used is 3 hours, allowing for observations within 1 hour and 30 minutes before to 1 hour and 30 minutes after the analysis time to be assimilated. Experiments are conducted testing the GSI 3DVar and 3DEnVar systems. For the hybrid 3DEnVar analysis, the GDAS 80 member ensemble forecasts (9 h forecasts) are used to provide the ensemble background error covariance (e.g., Wu et al. (2017)).

2.4 Cycling configuration

The current cycling configuration of the prototype RRFS is similar to the one used in RAP, i.e. cold starts are performed every 12 hours and warm starts are performed at all other cycles using the 1 h forecast from the previous cycle as background for the analysis. However, instead

of hourly cycles, 3 hourly cycles are performed considering the available resources (observations, analyses, and forecasts) to provide model initial and boundary conditions. The 3 hourly cycling strategy is configured using GFS ICs and LBCs and GDAS observations. Cold starts are similarly performed every 12 hours at 00:00 and 12:00 UTC and warm starts are performed at all other cycles using the FV3LAM 3 h forecast from the previous cycle as background for the analysis. In each cycle, a 24 h free forecast is launched following the analysis, with hourly outputs. Figure 4 illustrates the RRFS cycling configuration from cycles initialized between 00:00 UTC through 18:00 UTC.

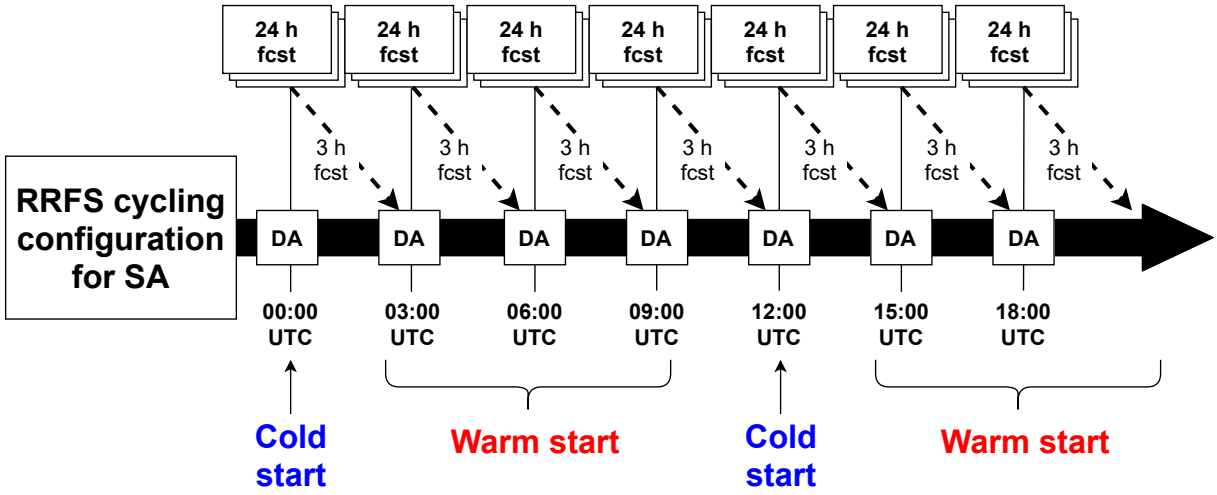

Figure 4: Diagram of the 3-hourly cycling configuration for RRFS for the case study over northern South America (SA).

2.5 Sensitivity experiments

A series of experiments are designed to examine the impact of different configurations on the analyses and forecasts. An experiment with no data assimilation is provided, acting as the baseline for all other experiments. This baseline experiment is called NoDAsa and uses the same cycling configuration as experiments with data assimilation. In order to select the appropriate physics suite for this study, two NoDAsa experiments are conducted testing the RRFS_PHY_v1a and GFSv15 physics suites available in the UFS Short Range Weather Application (SRW; UFS Development Team (2021)). Figure 5 shows precipitation forecasts from these experiments when run without data assimilation. 1 h accumulated precipitation estimations from CMORPH are used for comparison. The ccpp_HRRR physics suite shows smaller coverage than ccpp_GFSv15, but it captures the main precipitating patterns shown in the observations and with more accurate intensity and scattered patterns compared to the precipitation estimates. The ccpp_GFSv15 physics suite overestimates the accumulated precipitation in terms of coverage and the intensity, especially at the 6 h forecasts. Therefore, the suite based on HRRR physics is selected for all experiments.

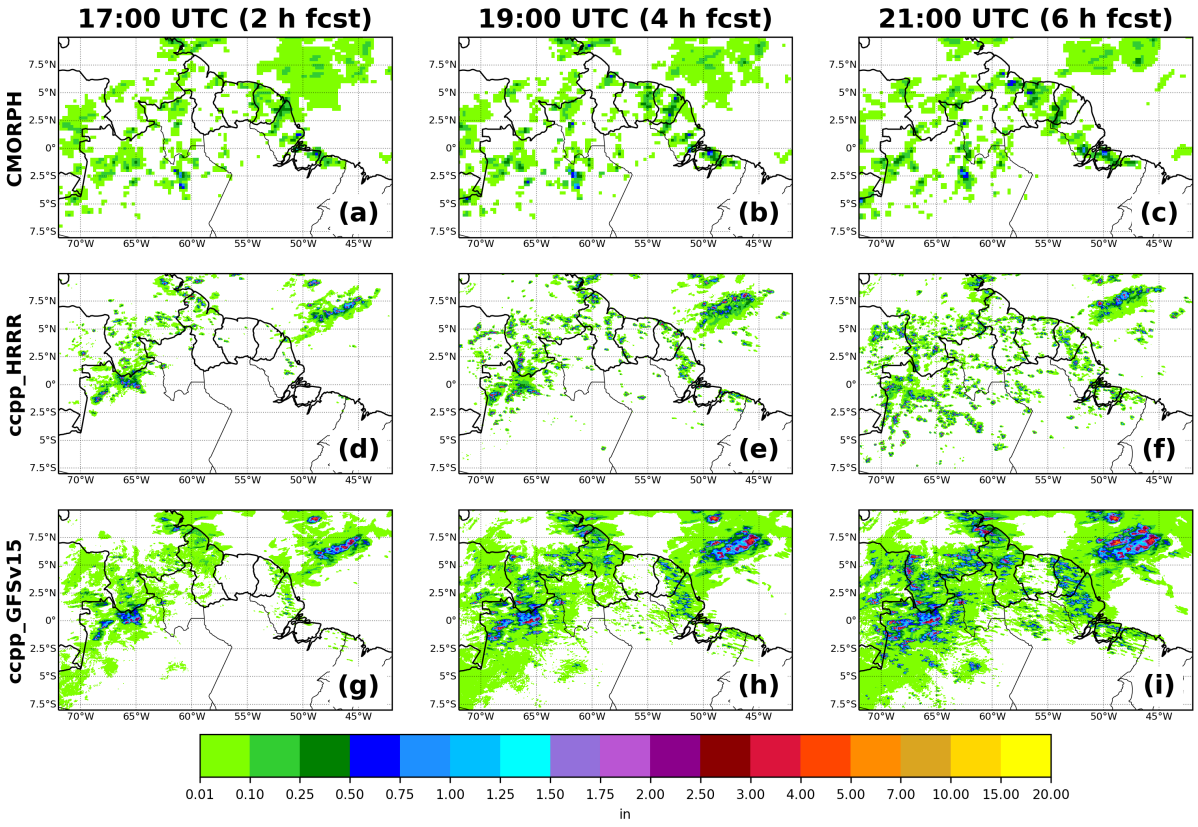

Figure 5: 2, 4, and 6 h forecasts of 1 h accumulated precipitation from experiments ccpp_HRRR (d, e, and f) and ccpp_GFSv15 (g, h, and i), initialized at 19:00 UTC on 4 July 2020 and the precipitation estimates from CMORPH at valid hours 17:00, 19:00, and 21:00 UTC (a, b, and c), respectively.

Table 1 lists all experiments in which GSI options are tested. Pure ensemble background error covariance, a combination of 85 % ensemble and 15 % static background error covariance, and 3DVar are examined in experiments 100EnBECsa, 85EnBECsa, and 3DVarsa, respectively. The experiment HL330sa is conducted by changing the horizontal localization radius to 330 km and keeping a vertical localization radius of 3 layers. The experiment VL9sa is then conducted by increasing the vertical localization radius to 9 layers and maintaining the original horizontal localization radius of 110 km. Finally, a third experiment is conducted in which both parameters are modified adopting a horizontal localization radius of 330 km and a vertical localization radius of 9 layers. This experiment is called VL9HL330sa. The supersaturation removal function in GSI is tested for this case study with the experiment CLIPSATsa. The PBL pseudo-observations function in GSI is tested in PSEUDOsa, which showed similar results to 85EnBECsa in all forecasts and statistics analyzed, and therefore the results are only shown in the analysis of the quantitative precipitation in Sect. **??**.

Table 1: List of experiments conducted testing different options in GSI in this study.

Experiments	**BEC weights**	**Supersat. removal**	**PBL pseudo-obs.**	**Localization scales**
NoDAsa	**No data assimilation**			
3DVarsa	**0 % ensemble** 100 % static	false	false	hloc=110 km vloc=3 layers
100EnBECsa	**100 % ensemble** 0 % static	false	false	hloc=110 km vloc=3 layers
85EnBECsa	**85 % ensemble** 15 % static	false	false	hloc=110 km vloc=3 layers
VL9sa	**85 % ensemble** 15 % static	false	false	hloc=110 km **vloc=9 layers**
HL330sa	**85 % ensemble** 15 % static	false	false	**hloc=330 km** vloc=3 layers
VL9HL330sa	**85 % ensemble** 15 % static	false	false	**hloc=330 km vloc=9 layers**
CLIPSATsa	**85 % ensemble** 15 % static	**true**	false	hloc=110 km vloc=3 layers
PSEUDOsa	**85 % ensemble** 15 % static	false	**true**	hloc=110 km vloc=3 layers

3 Summary of results

3.1 Examination of analyses

Data availability and coverage are the backbone of rapid updated analyses. Figure 6 presents the spatial distribution of assimilated temperature (Fig. 6a), wind (Fig. 6b and c), and radiance observations (Fig. 6d) at the 15:00 UTC cycle on 5 July 2020 for experiment 3DVarsa. The analysis residuals are shown at each point. Temperature observations in Fig. 6a are from conventional sources including radiosondes, surface marine observations such as buoys, synoptic observations over land, and METAR reports. At most of the same locations, humidity and surface pressure observations are also available. The conventional data is sparse and scattered throughout the domain but have the lowest analysis residuals. In Fig. 6b, the winds are from the same conventional sources and from scatterometers over the ocean (ASCATW). Satellite-derived winds are shown in Fig. 6c. Radiance observations from the MHS sensor from the MetOp B satellite covering the western side of the domain are presented in Fig. 6d. The root mean square (RMS) of the analysis residuals (OmA) in the bottom of each figure indicates that radiances present the larger values. At upper levels the coverage is improved by nonconven-

tional sources. All these data combined have great potential to positively impact the analyses and forecasts.

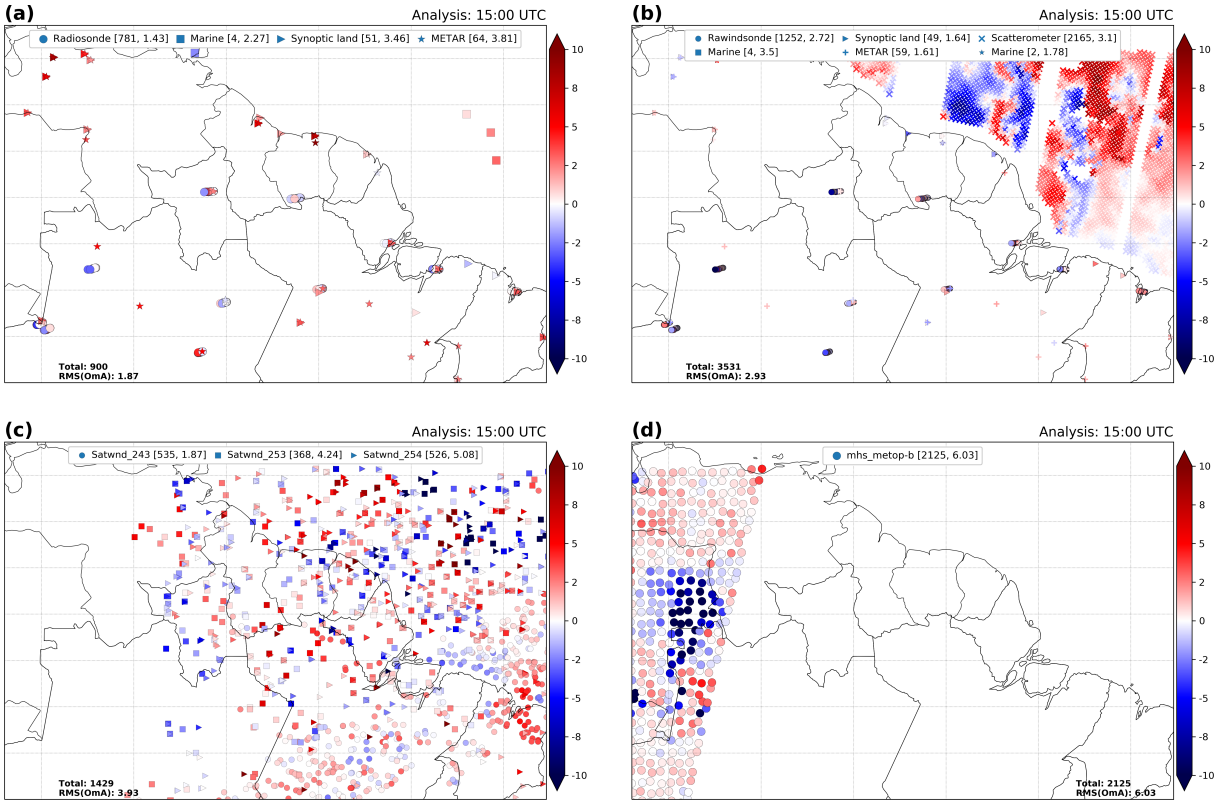

Figure 6: Spatial distribution of temperature (a), winds (b and c), and radiance (d) observations and analysis residuals (OmA) for the analysis at 15:00 UTC on 5 July 2020 from the experiment 3DVarsa. The color scale on the right indicates the magnitude of analysis residuals. The legend of observation type markers is shown at the top along with brackets listing associated counts and RMS error for the OmA. In the bottom of each panel is presented the total and averaged RMS of the OmA of all observations.

The RMS error and bias of the OmA and observation innovations (OmB) of the temperature from all observation types for analyses in all cycles performed for experiments 85EnBECsa and 3DVarsa are shown in Fig. 7. The analyses created at each cycle are closer to the observations with lower RMS and bias of the OmA values. Especially, the hybrid 3DEnVar with 85 % of the ensemble error covariance shows less biased analyses when compared to 3DVarsa. However, there is an evident diurnal cycle with an increase of the errors during the afternoon hours and a decrease during the night and early morning. The RMS of the OmB shows an increase in the first cycle after the model is cold started which may be related to the need of spin-up in the cycling configuration. This occurs at cycles initialized at 03:00 UTC and 15:00 UTC followed by a noticeable increase in the RMS of OmB values at 18:00 UTC. These results corroborate the difficulty of predicting convection during the afternoon hours, which is more evident in these experiments because of the lower coverage of surface data, as shown in the previous figure.

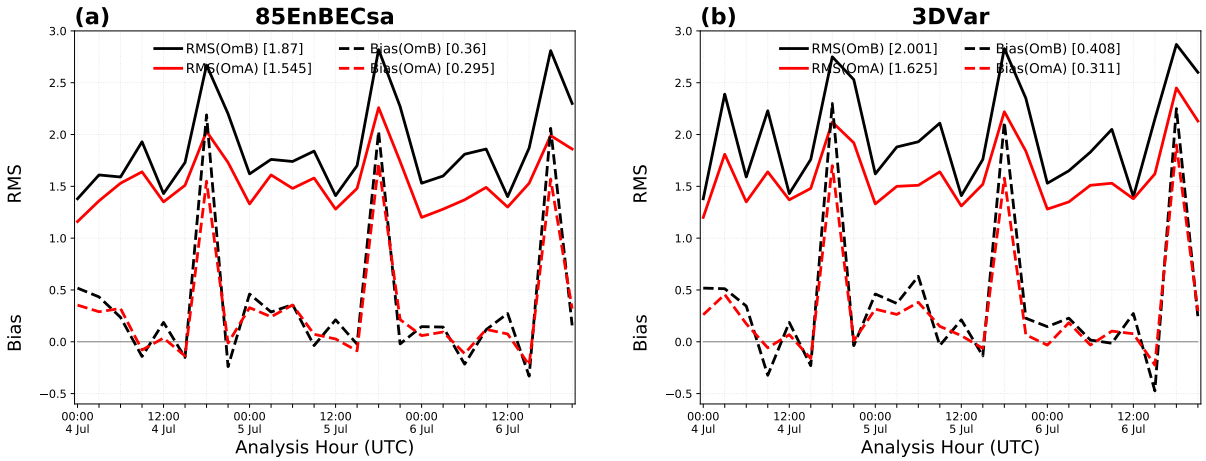

Figure 7: RMS and bias of the temperature background (OmB) and analysis (OmA) against all observation types for analyses in all cycles performed for experiments (a) 85EnBECsa and (b) 3DVarsa.

3.2 The impact of hybrid ensemble weights

The benefits of using a hybrid 3DEnVar analysis is investigated for this case study. Figure 8 presents the temperature and specific humidity analysis increments for the 15:00 UTC cycle on 5 July 2020 for experiments 100EnBECsa, 85EnBECsa, and 3DVarsa. The increments are small and more concentrated over certain points and some spread is observed in the surrounding areas. This is expected due to the lower data coverage. In 3DVarsa, the increments are smoother than in 85EnBECsa and 100EnBECsa. Meanwhile, the increments in 100EnBECsa are slightly noisier than those in 85EnBECsa. This indicates the effects of using the contribution from an ensemble background error covariance in producing analyses with more flow-dependent characteristics. Although not shown here, the results of the OmA and OmB statistics for these experiments show lower RMS of the OmA in 85EnBECsa and 100EnBECsa when compared to 3DVar, with 85EnBECsa slightly better than 100EnBECsa.

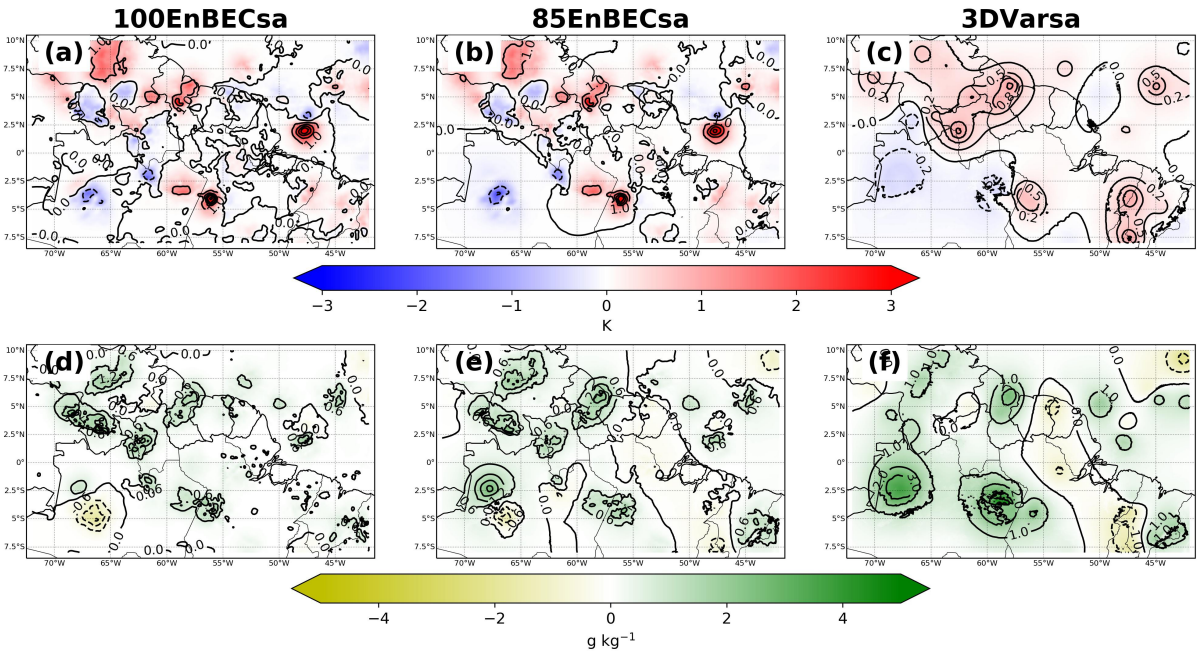

Figure 8: Analysis increment for temperature (K) (a, b, and c) and specific humidity (g kg^{-1}) (d, e, and f) at the first level above the surface for 15:00 UTC on 5 July 2020, for experiments 100EnBECsa (a and d) 85EnBECsa (b and e), and 3DVarsa (c and f).

The 2, 4, and 6 h forecasts of 1 h accumulated precipitation from the 15:00 UTC cycle on 5 July 2020 are examined in Fig. 9 for experiments 100EnBECsa, 85EnBECsa, 3DVarsa, and NoDAsa along with the precipitation estimates from the Climate Prediction Center (CPC) morphing technique (CMORPH) satellite precipitation estimates at valid hours 17:00, 19:00, and 21:00 UTC, respectively. The experiments using RRFS correctly capture the the precipitation along the coast from northern Para, Brazil to eastern Venezuela and the convection occurring over northern Amazonas and Roraima, Brazil, and southeastern Venezuela. This indicates the ability of the system in representing large scale conditions, which are better represented in experiments using data assimilation. Data assimilation has a greater impact in the first 2 and 4 h forecasts where those experiments show positive impacts over the experiment NoDAsa. Among the experiments with data assimilation, 3DVarsa shows an overestimate of the intensity and coverage of the precipitation at all forecasts lengths. Meanwhile, the experiment 85EnBECsa shows a better agreement with the precipitation estimates than 100EnBECsa or 3DVarsa. At 4 h and 6 h forecasts, the experiment 100EnBECsa shows improvements in the precipitation coverage along the coast as in other parts of the domain, but 85EnBECsa shows slightly better results. All experiments overproduce precipitation over the ocean, especially in 3DVarsa.

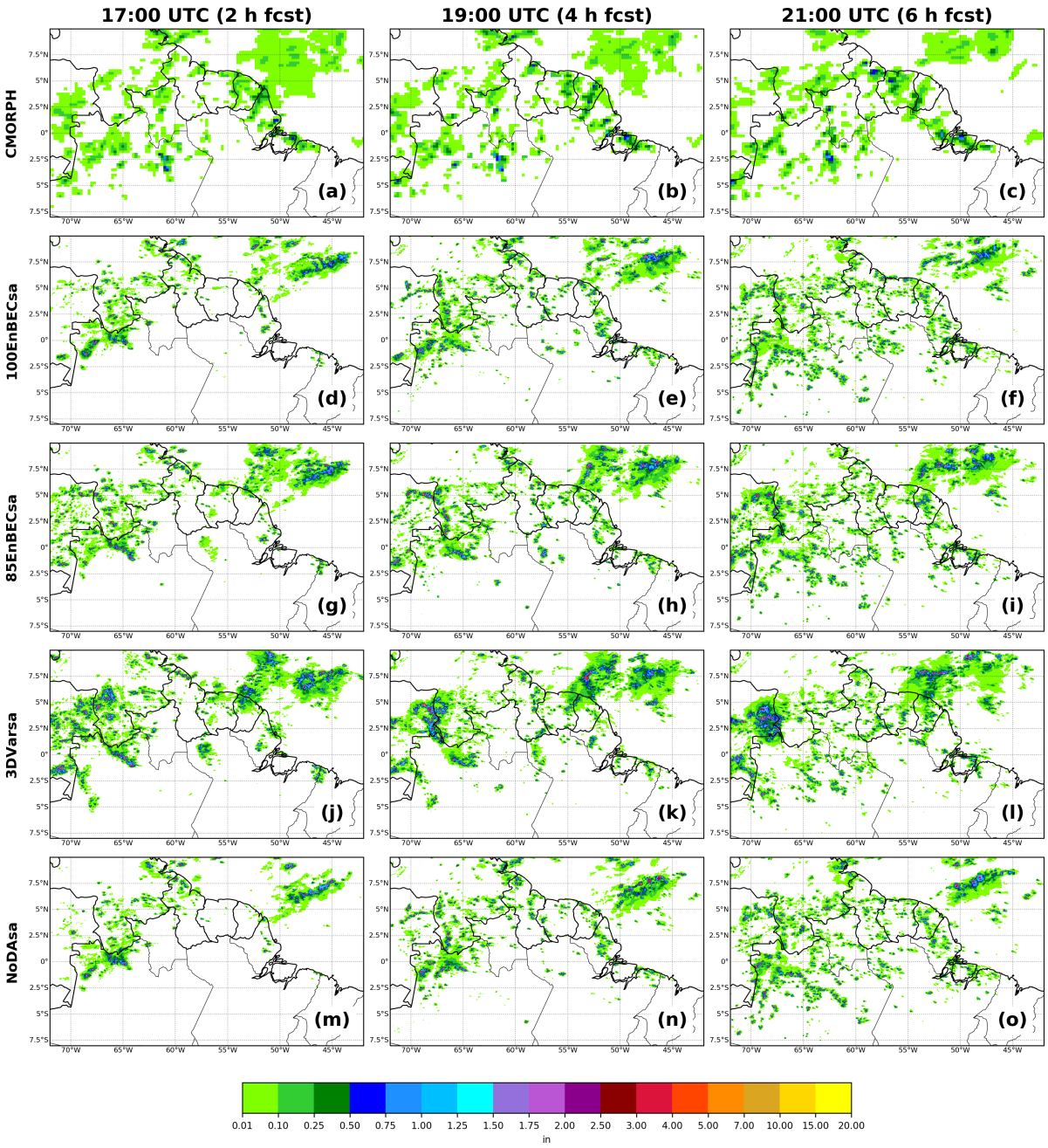

Figure 9: As in Fig. 5, but for experiments 100EnBECsa (d, e, and f), 85EnBECsa (g, h, and i), 3DVarsa (j, k, and l), and NoDAsa (m, n, and o).

3.3 The impact of covariance localization

Similar to Fig. 8, Fig. 10 presents the analysis increments of temperature and specific humidity for the 15:00 UTC cycle on 5 July 2020 but for experiments VL9sa, HL330sa, and VL9HL330sa. For this case study, the increase of the vertical localization radius from 3 to 9 layers shows almost neutral impact in the analysis when compared to 85EnBECsa. However, increasing the horizontal localization from 110 to 330 km shows a clear impact in the temperature and specific humidity analysis increments (Fig. 10b and e). Larger analysis increments, in coverage and magnitude, are observed in many parts of the domain. Increments are also more

detailed (less smooth) than increments in VL9sa. In the experiment VL9HL330sa (Fig. 10c and f), in which both length scales are modified, the analysis increments are similar to HL330sa but slightly less intense due to the change in the vertical localization. These results corroborate the importance of covariance localization.

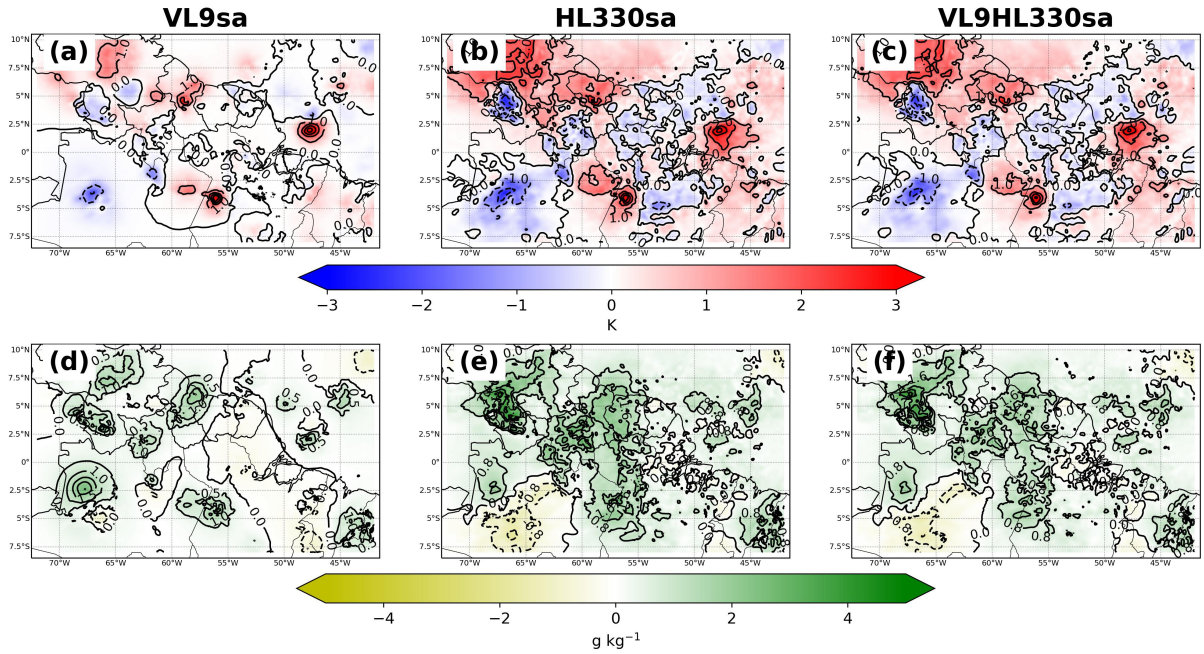

Figure 10: As in Fig. 8, but for experiments VL9sa (a and d), HL330sa (b and e), and VL9HL330sa (c and f) at 15:00 UTC on 5 July 2020.

.

In order to investigate whether or not the adjustments in the temperature and specific humidity analysis increments have also improved the forecasts, the precipitation forecasts are analyzed. Figure 11 presents the 2, 4, and 6 h forecasts of 1 h accumulated precipitation as well as the estimates from CMORPH. Overall, the results are similar, but improvements in precipitation forecast along the coast are observed in the experiments with the increase of the horizontal localization at all forecast lengths. A slightly better coverage of the precipitation is shown in HL330sa and VL9HL330sa at 2 h forecast when compared to 85EnBECsa. At 4 and 6 h forecasts, the experiment VL9HL330sa show results that better match the precipitation estimates.

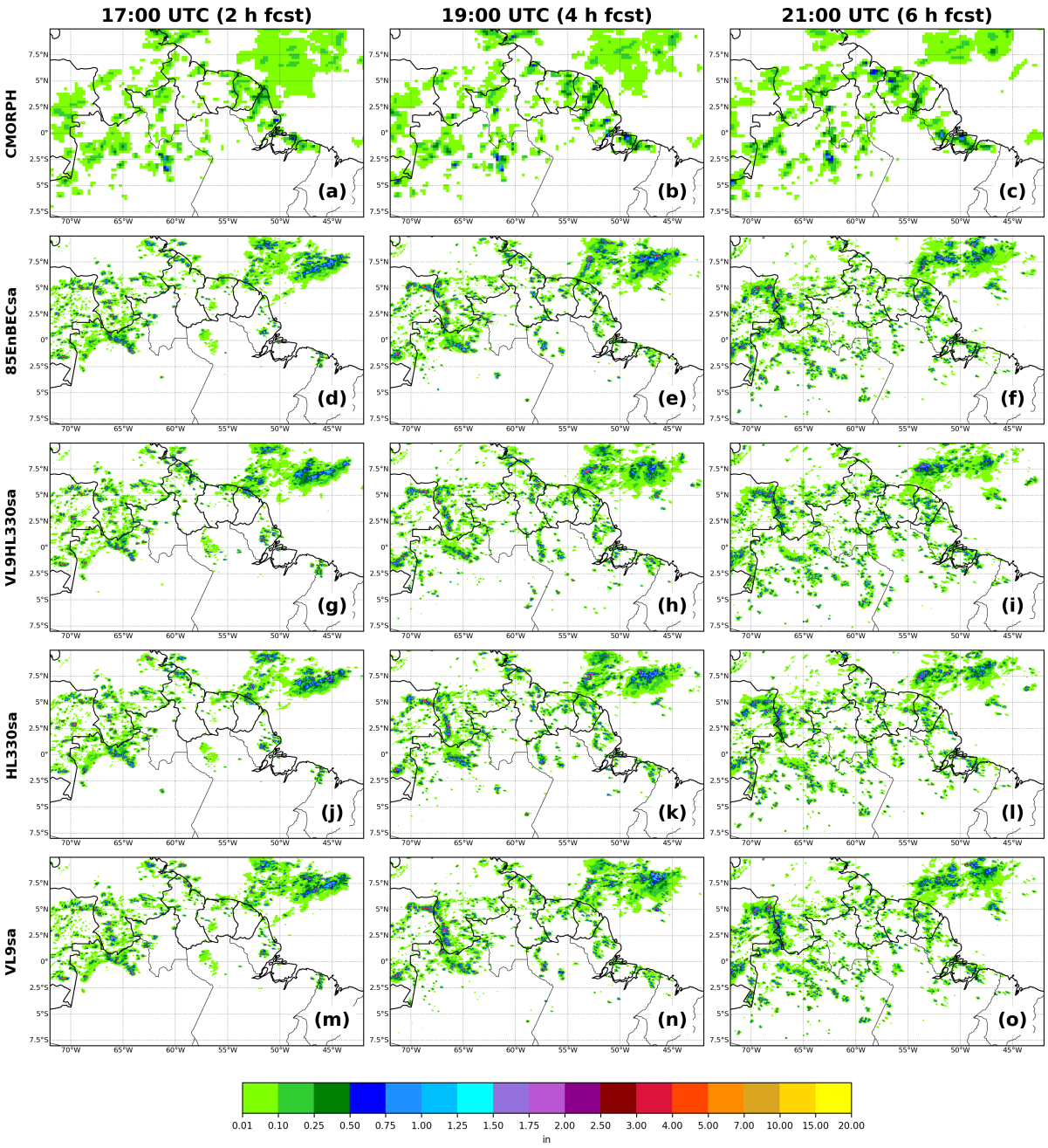

Figure 11: As in Fig. 9, but for experiments 85EnBECsa (d, e, and f), VL9HL330sa (g, h, and i), HL330sa (j, k, and l), and VL9sa (m, n, and o).

The RMSE and bias for the 3 h forecast of 2 m temperature (Fig. 12a and c) and 2 m dew point temperature (Fig. 12b and d) verified against the the ECMWF Reanalysis v5 (ERA5) data, for experiments 85EnBECsa, VL9HL330sa, HL330sa, VL9sa, and NoDAsa are presented in Fig. 12. The RMSE and bias are computed using a bootstrap resampling technique of 1000 replications with replacement at each forecast lead hour in every cycle. The sample size for the statistics computation is 8829 for each cycle and the sample uncertainty for both RMSE and bias is very small and therefore not shown in the figure. The RMSE and bias show the diurnal cycle, with lower RMSE and bias values during the night and larger values during the afternoon, which is similar to results in Fig. 7. However, the differences between experiments with data

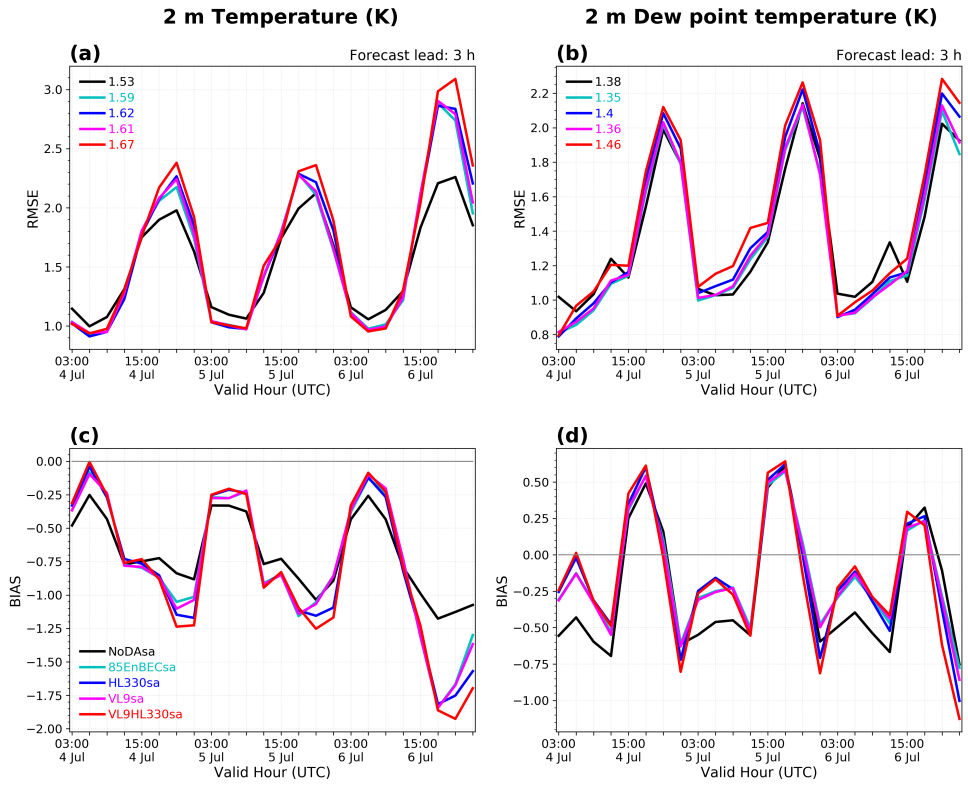

Figure 12: RMSE and bias for the 3 h forecast of 2 m temperature (a and c) and 2 m dew point temperature (b and d) against the ERA5 reanalysis for experiments 85EnBECsa, VL9HL330sa, HL330sa, VL9sa, and NoDAsa. The legend for each experiment is shown at the bottom of panel (c). The RMSE averaged over all cycles is shown in panels (a) and (b) for each experiment.

assimilation and without it are more marked. The impact of the variation in the horizontal and vertical localization radius is neutral to negative in most of the cycles and for both variables when compared to 85EnBECsa. The experiment VL9HL330sa shows the worst performance at cycles initialized between 15:00 UTC and 21:00 UTC, when NoDAsa performs best.

3.4 The impact of supersaturation removal and PBL pseudo observations

The function to remove supersaturation in the background, available during the analysis process in GSI, is activated in the experiment CLIPSATsa. Results from the difference in the specific humidity (g kg^{-1}) analyses between experiments with and without activating this option show more positive and negative differences with a larger magnitude in the southwestern part of the domain over the state of Amazonas and northwestern Para in Brazil, and also over Guyana which corresponds with the area where precipitation is occurring according to the CMORPH estimates (figure not shown). The 2, 4, and 6 h forecasts of 1 h accumulated precipitation from the 15:00 UTC cycle on 5 July 2020 for experiments CLIPSATsa and 85EnBECsa are shown in Fig. 13 along with the precipitation estimates from CMORPH at valid hours 17:00, 19:00, and 21:00 UTC, respectively. CLIPSATsa shows improvements when compared with 85EnBECsa, especially over the ocean at all forecast lengths. However, over land, the improvements are mainly at 4 and 6 h forecasts with slightly better coverage of the precipitation along the coast.

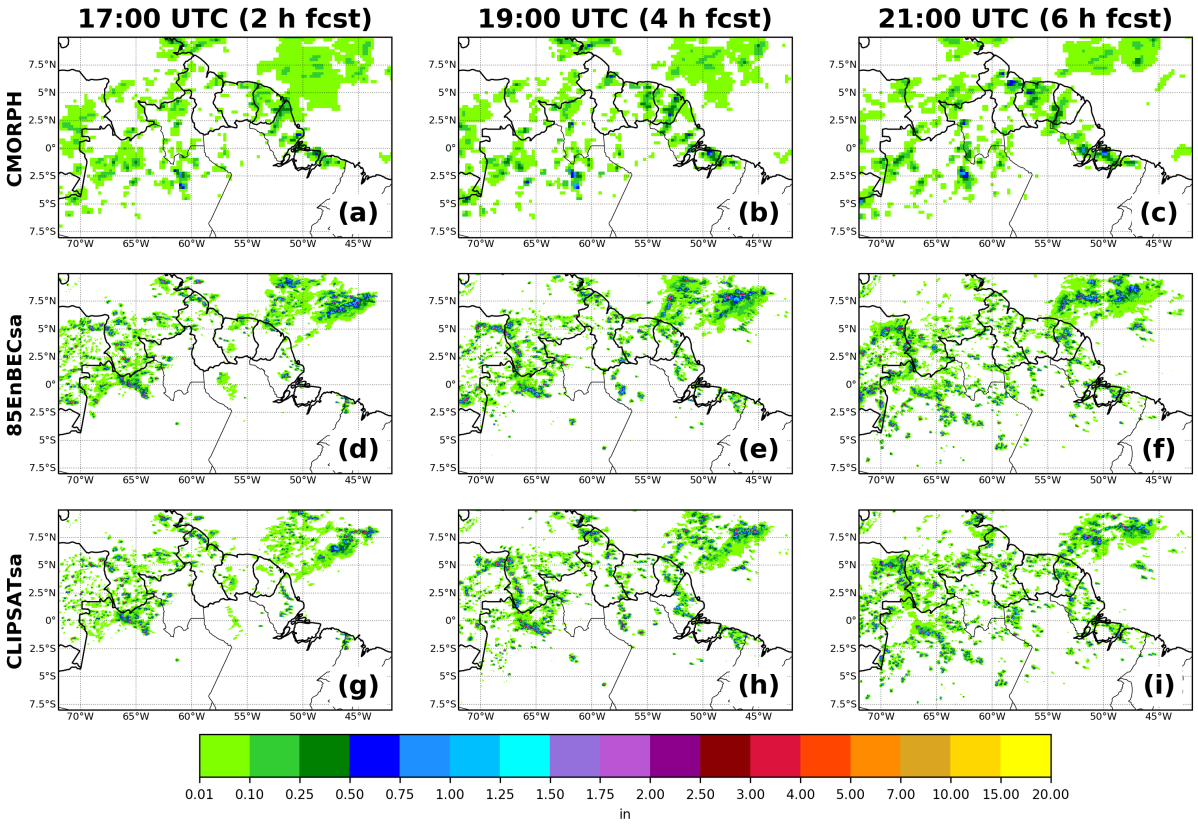

Figure 13: As in Fig. 9, but for experiments 85EnBECsa (d, e, and f) and CLIPSATsa (g, h, and i).

4 Conclusions and future work

In this DTC visitor project, the data assimilation framework for the prototype RRFS is investigated through the simulation of weather systems often seen over tropical latitudes, such as a typical Amazon coastal squall line case during the 2020 Amazon dry season. Sensitivities to various configurations and algorithms available in GSI are analyzed in order to find the best configuration to produce more realistic convection forecasts and provide guidance on convective scale data assimilation over regions of the globe with low data density, as in the Amazon region. The initial challenge for this study is the case selection, and a methodology based on multiple past efforts is developed and applied. The Amazon coastal squall line case that initiated during the afternoon of 5 July 2020 is selected and studied through the execution of various numerical experiments. Overall, results suggest that RRFS can provide reasonably good guidance for the tropical region. The main findings of the study are listed below:

a) RRFS is able to capture the main large scale patterns with a correct positioning of the precipitating systems as analyzed using the CMORPH precipitation estimates;

b) A CCPP physics suite based on HRRR physical parameterizations shows a better representation of the precipitation, while a GFS-based physics suite shows larger coverage and intensity than the precipitation estimates;

c) Despite the low coverage of available data for experiments with data assimilation, the data

assimilation system performs adequately over this region, with RRFS analyses closer to the observations in all cycles;

d) Precipitation coverage along the coast and other parts of the domain are improved when using data assimilation. Experiments 85EnBECsa and 100EnBECsa show a closer agreement with the precipitation estimates of the domain at 4 and 6 h forecast, with 85EnBECsa showing better results at all forecast lengths;

e) There is a notable diurnal cycle in the RMSE and bias values, with the experiment NoDAsa outperforming the experiments with data assimilation during the afternoon hours. Errors associated with the convection occurrence and sparser coverage of surface observations in the domain may have contributed to these results;

f) The experiment 3DVarsa shows larger RMSE and bias when comparing against observations of 2 m temperature and 2 m dew point temperature in all cycles, while the experiment 85EnBECsa performs better than the others during the night and early morning hours;

g) When increasing the localization radius in ensemble-based error covariance, analysis increments show more flow-dependent characteristics, particularly when increasing the horizontal localization radius from 110 km to 330 km which also allows for improvements in the RMS of the OmA values and slightly better representation of the satellite-derived precipitation estimates. Nevertheless, when comparing against surface observations, the impact is neutral to negative;

h) When activating supersaturation removal in GSI for this case study, the results show positive impacts against 85EnBECsa with a slightly better representation of the precipitation along the coast. However, the improvements are small and mainly concentrated at larger forecast lengths

Despite all the options tested in this study, it is clear that there is still much to investigate regarding the capabilities of RRFS over the tropics. Future studies may be focused on investigating the use of GOES-16 GLM data as a proxy for reflectivity in the data assimilation system. This approach would be particularly important for convective scale data assimilation over regions that rely on satellite data, such as the Amazon.

References

Brown, B. et al. The Model Evaluation Tools (MET): More than a Decade of Community-Supported Forecast Verification. *Bulletin of the American Meteorological Society*, American Meteorological Society, Boston MA, USA, v. 102, n. 4, p. E782 – E807, 2021. Available at: ⟨https://journals.ametsoc.org/view/journals/bams/102/4/BAMS-D-19-0093.1.xml⟩.

Cohen, J. C. P.; SILVA DIAS, M. A. F.; Nobre, C. A. Environmental Conditions Associated with Amazonian Squall Lines: A Case Study. *Monthly Weather Review*, v. 123, n. 11, p. 3163–3174, 1995.

Garstang, M. et al. Amazon coastal squall lines. part i: Structure and kinematics. *Monthly Weather Review*, v. 122, n. 4, p. 608–622, 1994.

Garstang, M. et al. The amazon boundary-layer experiment (able 2b): A meteorological perspective. *Bulletin of the American Meteorological Society*, American Meteorological Society, Boston MA, USA, v. 71, n. 1, p. 19 – 32,

1990. Available at: ⟨https://journals.ametsoc.org/view/journals/bams/71/1/1520-0477_1990_071_0019_tablea_2_0_co_2.xml⟩.

Keller, M. et al. The large-scale biosphere-atmosphere experiment in amazonia: Analyzing regional land use change effects. *Ecosystems and Land Use Change Geophysical Monograph Series 153*, 2004. Available at: ⟨https://www.fs.usda.gov/treesearch/pubs/30201⟩.

Marengo, J. A. et al. Onset and End of the Rainy Season in the Brazilian Amazon Basin. *Journal of Climate*, v. 14, n. 5, 2001.

Martin, S. T. et al. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmospheric Chemistry and Physics*, v. 16, n. 8, p. 4785–4797, 2016. ISSN 16807324.

Oliveira, F. P. de; Oyama, M. D. Antecedent atmospheric conditions related to squall-line initiation over the northern coast of brazil in july. *Weather and Forecasting*, American Meteorological Society, Boston MA, USA, v. 30, n. 5, p. 1254 – 1264, 2015. Available at: ⟨https://journals.ametsoc.org/view/journals/wefo/30/5/waf-d-14-00120_1.xml⟩.

Rozante, J. R. *Produto de precipitação MERGE*. [S.l.], 2017. Available at: ⟨http://ftp.cptec.inpe.br/modelos/io/produtos/MERGE/read_me_MERGE.pdf⟩.

UFS Development Team. *Unified Forecast System (UFS) Short-Range Weather (SRW) Application*. Zenodo, 2021. Available at: ⟨https://doi.org/10.5281/zenodo.4534994⟩.

Vila, D. A. et al. Forecast and tracking the evolution of cloud clusters (fortracc) using satellite infrared imagery: Methodology and validation. *Wea. Forecasting*, v. 23, p. 233–245, 2008. Available at: ⟨https://doi.org/10.1175/2007WAF2006121.1⟩.

Wu, W.; Purser, R. J.; Parrish, D. Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Monthly Weather Review*, v. 130, p. 2905–2916, 2002. Available at: ⟨https://doi.org/10.1175/1520-0493(2002)130⟨2905:TDVAWS⟩2.0.CO;2⟩.

Wu, W.-S. et al. Regional ensemble–variational data assimilation using global ensemble forecasts. *Weather and Forecasting*, v. 32, n. 1, p. 83–96, 2017. ISSN 0882-8156. Available at: ⟨https://doi.org/10.1175/WAF-D-16-0045.1⟩.