

Data Assimilation System Applied to Short-Range Forecast System [†]

Pedro M. González *, Maibys Sierra and Adrián L. Ferrer 

Center for Atmospheric Physics, Institute of Meteorology, La Habana 10400, Cuba; maibysl@gmail.com (M.S.); aluislfh@gmail.com (A.L.F.)

* Correspondence: pedro.met90@gmail.com

[†] Presented at the 5th International Electronic Conference on Atmospheric Sciences, 16–31 July 2022;

Available online: <https://ecas2022.sciforum.net/>.

Abstract: An evaluation of 3DVAR, 3DEnVAR and 4DEnVAR methods is carried out, assimilating in a combined way prepbufr and radiances data, applied to the Short-range Forecast System to determine which scheme is more suitable for short-term forecasting purposes. The results suggest that hybrid schemes tend to generate more accurate forecasts than 3DVAR; however, 4DEnVAR is the most robust scheme and therefore the one that provides more realistic solutions. The forecast of the accumulated rainfall in 24 h constitutes the greatest difficulty since all the assimilation schemes generate underestimates related to the 24 h rainfall forecast.

Keywords: SisPI; WRFDA; short-range; assimilation



Citation: González, P.M.; Sierra, M.; Ferrer, A.L. Data Assimilation System Applied to Short-Range Forecast System. *Environ. Sci. Proc.* **2022**, *19*, 20. <https://doi.org/10.3390/ecas2022-12806>

Academic Editor: Anthony Lupo

Published: 14 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the aim of obtaining more realistic solutions in relation to short- and very-short-term numerical forecasting, the SisPI project (Short-range Forecast System, acronyms in Spanish), whose fundamental core is the WRF-ARW (Weather Research and Forecasting-Advanced Research WRF) mesoscale model, is working operationally at the Meteorology Institute of Cuba. One of the lines that is currently in full development within the framework of SisPI is the implementation of data assimilation techniques, seeking a universal design that meets the growing demands for forecasts and numerical products, all of this framed within the project “Development of the data assimilation module for the Short-range Forecast System”, belonging to the “Meteorology and sustainable development” program.

Efforts to operationally include a data assimilation scheme that would meet the demands in terms of forecast quality and current technological capabilities began from the very beginning of the project [1]. However, they were limited since these studies involved only the 3DVAR scheme and, although experiments were carried out with different covariance matrices [2,3], the use of the generic matrix prevailed. Recently, aspects related to the extension of the temporary window for the construction of the matrices [4], the impact of the application of multiple OL (outer loops), as well as the application of hybrid methods have been included in the experiments in order to achieve the most robust scheme.

In this research, three assimilation schemes are evaluated, 3DVAR, 3DEnVAR and 4DEnVAR, all available in the WRFDA module. The impact of combined prepbufr and radiances data assimilation, these lately obtained from the microwave channel from polar orbit satellites, is analyzed. For the investigation, three study cases are used: Hurricane ETA (Tropical Storm passing over Cuba) on 7–8 November 2020; a mesoscale convective system (SCM) that affected the central region of the country on 23 May 2020 and a thunderstorm over the western north coast that occurred on 29 September 2020.

2. Materials and Methods

2.1. Short-Range Forecast System (SisPI)

To develop the experiments, the WRF model was used with the ARW dynamic core in its version 3.8.1 [1,5]. The main objective of this system is very short and short-term forecasting. SisPI's design has two-way nested domains of 27 and 9 km (kilometers), respectively, and a one-way nested domain with 3 km of grid resolution (Figure 1). The model was initialized from GFS (Global Forecast System) forecast data with 0.5° horizontal resolution, the same that is used at the operative design.

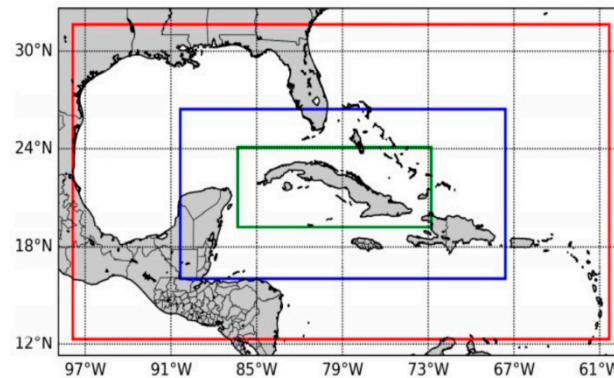


Figure 1. SisPI's domains. Red box represents the parent domain; blue box represents the 9 km nested domain and green box the high-resolution domain (3 km).

The configuration proposed by SisPI includes 28 vertical levels, the Mellor–Yamada–Nakanishi and Niino2.5 PBL scheme and the RRTM longwave parameterization for all domains. In the case of low-resolution domains, it contains the microphysics of WSM5, the Grell–Freitas cumulus parameterization, and the shortwave Dudhia scheme. For the high-resolution domain, the Morrison double-moment microphysics is used, the cumulus parameterization is deactivated and the Goddard shortwave radiation scheme is used. These differences have been supported by sensitivity studies developed in the project [1,5].

2.2. Assimilation Methods

2.2.1. Three-Dimensions Variational (3DVAR)

The 3DVAR method can be summarized as the iterative solution to find the state x that minimizes the cost function (Equation (1)). This solution represents the maximum probability (least variance) estimate of the atmosphere true state given two a priori data sources: the background field and observations [6].

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - H(x))^T R^{-1}(y - H(x)) \quad (1)$$

2.2.2. Hybrid Assimilation Methods (3DEnVAR-4DEnVAR)

Recent studies suggest that hybrid formulations have some advantages over 3DVAR, highlighting the fact that the covariances that weigh the model errors extracted from an ensemble, being flow-dependent, can better represent the error of the day, as opposed to the isotropic and static characteristics that contain the 3DVAR algorithm [7]. This means that, in the purely variational case, when the error contained in the covariance matrix differs from the flow of the day, the results may not be satisfactory [7].

In the 3DEnVAR formulation (Equation (2)), the ensemble is valid only for the analysis time:

$$J(x; \alpha) = B_s \frac{1}{2}(x_1 - x_b)^T B^{-1}(x_1 - x_b) + B_e \frac{1}{2} \sum_{i=1}^n (a_i^T C^{-1} a_i) + \frac{1}{2}[y - H(x_1 + x_e)]^T R^{-1}[y - H(x_1 + x_e)] \quad (2)$$

In this case, B is the weight attributed to the ensemble, with C being the correlation matrix for the effective location of the ensemble's perturbations.

The algorithm used for 4DEnVAR is very similar; the difference lies in the fact that it requires ensemble perturbations and observations for multiple time steps (Equation (3)) [8].

$$J(x; \alpha) = B_s \frac{1}{2} (x_1 - x_b)^T B^{-1} (x_1 - x_b) + B_e \frac{1}{2} \sum_{i=1}^n (a_i^T C^{-1} a_i) + \frac{1}{2} [y - H(x_1 + x_{e;k})]^T R^{-1} [y - H(x_1 + x_{e;k})] \quad (3)$$

where k represents the multiple perturbations used in the algorithm.

2.3. Experiments Design

To carry out the data assimilation experiments, the WRFDA version 3.9.1 package was used (https://www2.mmm.ucar.edu/wrf/src/WRFDA_V3.9.1.tar.gz), because the 4DEnVAR method is unavailable for previous versions. The assimilation was executed only on the domain with the highest horizontal resolution (3 km). For all experiments, a statistical background error domain-dependent was used, using the forecasts generated up to 15 days prior to initialization at interest day. The construction of background error was carried out using the `gen_be_wrapper.ksh` program, available in the WRFDA.

For 3DVAR's experiments, multiple OLs (outer loops) were applied. This was performed following the results shown in similar studies by [7,9]; for these experiments where multiple OLs are used, it was decided to empirically modify the multiplicative weight of the standard deviation and the scale length too, reducing them in each OL (Table 1). This strategy aims to give greater influence to the first guess, improved in the successive OL.

Table 1. Multiplicative weight and control variables modification.

| Control Variables (CV7 Background Error Covariance) | Modification on Variance Scaling (1ro-2do-3ro OL) | Modification on Length Scaling (1ro-2do-3ro OL) |
|---|--|--|
| U component | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| V component | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Temperature | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Pseudo relative humidity | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Surface pressure | 1.5-1.0-0.5 | 1.0-0.5-0.2 |

Regarding the application of the hybrid schemes, the ensemble designed to obtain the flow-dependent perturbations was build using previous SisPI runs, under the criterion that they included members at least to 12 h prior to the initialization moment, with the purpose of mitigating the possible effects of the model spin-up on the calculation of flow-dependent perturbations. This allowed one to make a small ensemble with 5 members.

For this research, a weight of 75% was arbitrarily assigned to the ensemble contribution and 25% to the static covariance matrix, giving thus a greater relevance to flow-dependent errors on the assimilation process in the case of hybrid schemes. This weight ratio has also been established in other studies with satisfactory results [7,9].

2.4. Data Employed for Assimilation and Verification

For this investigation, we used combined prepbufr and radiances data. The information contained in the PREPBUFR files includes observations in FM-12 codes, METAR, ship data, buoys and soundings. Regarding the assimilation of radiances, data in bufr format of microwave channel were used: AMSU-A (NOAA-15/16/18/19), MHS (NOAA-18/19), SSMIS (DMSP-16) and ATMS (Suomi-NPP), in all cases extracted from the site <https://rda.ucar.edu/datasets> (accessed on 20 August 2021).

For the evaluation of the experiments, the satellite precipitation estimation data of the GPM product [10] were used, as well as the data from the surface meteorological stations of Meteorology Institute.

3. Results

3.1. Differences Related with the Volume of Data Assimilated by the Methods

In relation to data in prepbufr format, 3DVAR and 3DENVAR are usually close, while 4DENVAR assimilates a significantly higher volume. The fundamental differences seem to be in the radiances, where it was the hybrid schemes that maintained high volumes of information of this type assimilated in contrast to 3DVAR. These differences seem to be associated with the algorithms inside the assimilation methods, because the available observations and the assimilation domain was the same in all cases.

3.2. Cost Function Reduction and Application of Multiple OLs

The use of multiple OLs leads to an increase in the computational cost requiring a greater number of iterations. It is found that this technique does not always guarantee variational control, since in some cases a $\Delta J_o > 0$ is obtained, which indicates that the number of rejected observations in the last OL was less than at the beginning, which is an indicative of the emphasis diminishing of the guess in later OLs. Hybrid methods lead to the minimization of the cost function in a similar number of iterations, initializing it at higher values as a result of a greater number of assimilated observations. The reduction rate is similar to 3DVAR for these experiments, although it can be seen that 4DENVAR achieves a more effective minimization than 3DENVAR (Figure 2). These results are considered to coincide with [8], who state that hybrid methods can work effectively with the application of only one OL.

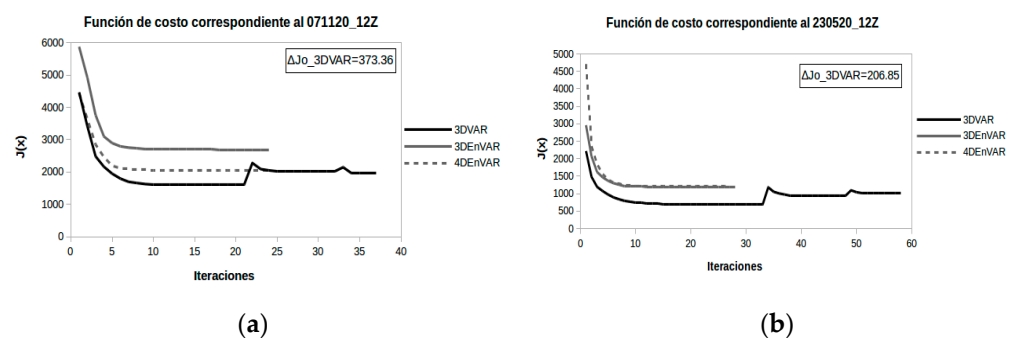


Figure 2. Cost function reduction obtained by the different methods, (a) experiment initialized on 7 November 2020 at 12 UTC corresponding to tropical storm ETA; (b) experiment initialized on 23 May 2020 at 12 UTC corresponding to mesoscale convective system.

3.3. Incremental Analysis

The incremental analysis, a result of the difference between the analysis field and background, seen through cross-sections, allows understanding the differences shown by the different assimilation schemes and the impact on the numerical forecast of different meteorological fields (Figure 3). For all cases, 3DVAR method contribution to the background field is very small and local, coinciding with other results [8,11].

In the case of ETA the incremental analysis shows a change in latitudinal position of the storm core at day 7 experiment (omitted figure), expressed in a presence of a thermal dipole. There are also signs of growth of convective cells with respect to the background field, characterized by cooling at low levels and heating at high levels. The aforementioned differences become more noticeable in the runs of day 8 initialized at 0000 UTC.

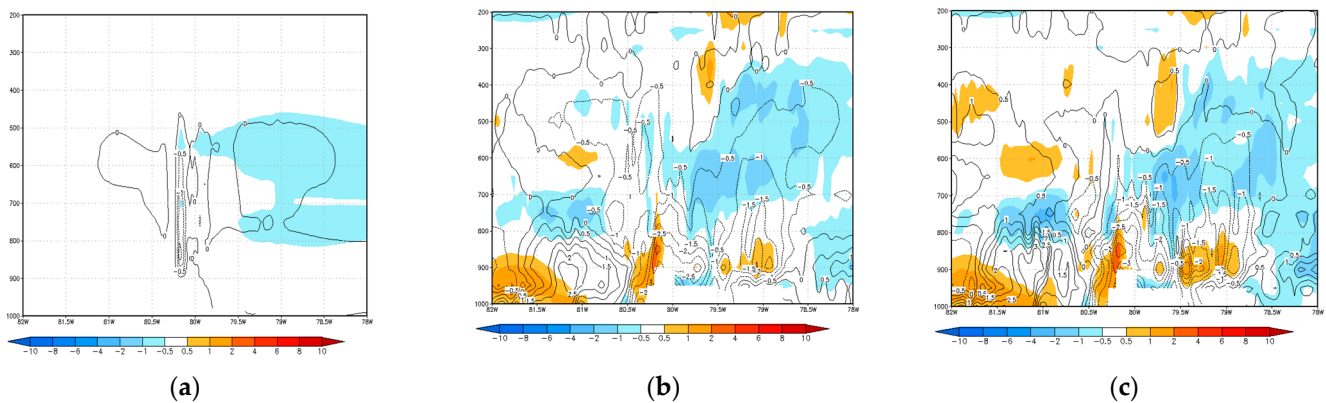


Figure 3. Cross-section showing the incremental analysis corresponding to 00:00 UTC initialization for mesoscale convective system case. Shaded represent the temperature increments and contour lines the specific humidity increments. (a) 3DVAR; (b) 3DEnVAR; (c) 4DEnVAR.

The incremental analysis made in SCM indicate that the contribution was characterized by an increase in humidity at low levels, which was more marked in the case of 3DEnVAR, which exceeds thresholds between 1.5 and 5 g/kg to 4DEnVAR. The latter proposes an initial condition with a lower and middle troposphere cooler than SisPI, while 3DEnVAR limits this cooling to the lower layers (where increases in specific humidity occur) and exhibits warming zones with respect to SisPI at high levels, suggesting, mainly towards the Guamuhaya massif region, the early growth of convective cells. The contributions observed at the initialization of 0000 UTC support the theory of an accelerated stratification by the model, based on the additional cooling provided by the assimilation schemes (mainly the hybrids) in the layer between the 800 and 400 hPa surfaces.

In relation to the TE in the experiments initialized at 1200 UTC, it is observed that the hybrid methods tend to heat the layer between 600 and 300 hPa approximately. An increase in humidity is also observed at medium and high levels as a result of the contribution of flow disturbances contained in these methods. There is also a certain tendency to a slight cooling in the lower layers with respect to the background field, a solution that led to the early development of convective cells in these experiments. The main difference between both methods is that 4DEnVAR shows a slightly less humid atmosphere in some portions between the 850 and 700 hPa layers.

3.4. Rainfall Forecast

In general the model tends to underestimate the total amount of precipitation in 24 h. The behavior of this variable in ETA show as likely cause of errors in spatial coverage and rainfall intensity in the hybrid schemes, the predicted interaction of ETA with an upper low, as these solutions locate the center of circulation of the upper low closer to the core of the upper low. ETA. The intrusion of dry air at mid-levels, favored by the presence of the aforementioned system, may have been the determining factor in the forecast of the distribution of the storm's rainfall areas (Figure 4).

In relation to the SCM, the 3DVAR scheme is the one that produces a somewhat superior solution to the rest of the experiments. On the other hand, the hybrid schemes lead to very similar forecasts with errors in the spatial coverage of the precipitation areas higher than those exhibited by 3DVAR or SisPI itself. In the opposite way, in the experiments initialized at 00:00 UTC, because although the underestimation trend continues, the solution proposed by 3DVAR is very similar to that of SisPI and it is the hybrid schemes that produce significant improvements, especially towards the region western country.

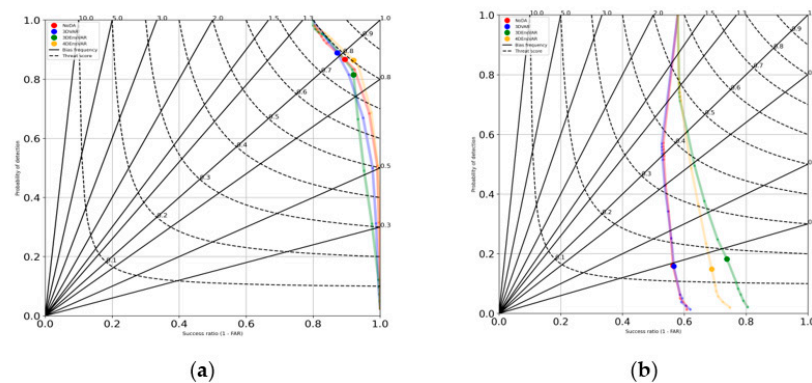


Figure 4. Performance evaluation of 24 h accumulated precipitation forecast derived from SisPI and the different assimilation schemes. (a) Eta's case, (b) thunderstorm's case, both initialized at 00:00 UTC.

Finally at the TE case, SisPI and 3DVAR solutions were closer, although 3DVAR improved the spatial coverage of precipitation, especially towards the eastern half of the country where convection was stimulated by the presence of a tropical wave. Towards the western half (where the study case was developed) the differences between 3DVAR and SisPI were not very significant. On the other hand, the hybrid methods increased the prediction error towards the eastern part, significantly reducing the precipitation areas. However, a slight improvement is observed towards the west, although both schemes predicted the region of convergence of the mesoscale flow just overland on the north coast of Pinar del Río and Artemisa, a solution derived from the forecast of a weaker southeast region flow, in relation to reality. On the contrary, the solutions of the hybrid schemes corresponding to the initialization of 0000 UTC turned out to be much superior to the forecasts emanating from SisPI and 3DVAR, improving the spatial coverage and the intensity of the precipitation in numerous points of the domain.

4. Conclusions

The results described in the previous section indicate that the model shows a clear tendency to underestimate the accumulated rainfall in 24 h, which is only partially corrected by the assimilation schemes as they have been designed.

In relation to the purely variational proposal, it is obtained that it poorly modifies the background field as a result of the isotropic and static characteristics of the algorithm, which causes its results to quickly converge to the solution without assimilation. In this sense, the application of multiple LOs modifying the multiplicative coefficients of the scale length and the variance in order to give greater representativeness to the first corrected approximation resulting from the initial LOs is not satisfactory and increases the computational cost of the method.

On the other hand, the contribution of the flow-dependent errors in combination with the static errors contained in the covariance matrix, in the hybrid schemes, manages to palpably modify the field of the first approximation, leading to the assimilation effect being prolonged in an approximate threshold of 6 to 12 h. The performance of the 3DENVAR method is unstable as it can lead to very realistic forecasts or others comparable with 3DVAR in the same situation. The 4DENVAR method allows assimilating a significantly larger volume of conventional data, however it does not exhibit the same superiority in relation to radiances. The 4DENVAR scheme turns out to be the most robust of the three because, although it does not always turn out to be the one that exhibits the most realistic solutions, it is the only one whose forecast is always superior to the run without assimilation.

Author Contributions: Conceptualization, P.M.G. and M.S.; methodology, M.S.; software, A.L.F.; validation, P.M.G., M.S. and A.L.F.; formal analysis, P.M.G.; investigation, P.M.G.; resources, M.S.; data curation, A.L.F.; writing—original draft preparation, P.M.G.; writing—review and editing, M.S.; visualization, A.L.F.; supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sierra, M.; Ferrer, M.; Hernández, R.; González, Y.; Borrajero, I.; y Rodríguez, R.F. Informe del Resultado No.1 del Proyecto: Sistema de Predicción a muy corto plazo basado en el Acoplamiento de Modelos de Alta Resolución y Asimilación de Datos. In *2014: Sistema Automático Depredicción a Mesoescala de Cuatro Ciclos Diarios*; Instituto de Meteorología de Cuba: La Habana, Cuba, 2014.
2. Armas, E. *Asimilación de Observaciones de Satélites Meteorológicos en el Modelo WRF*; Trabajo Diploma; Instituto Superior de Tecnologías y Ciencias Aplicadas: La Habana, Cuba, 2015.
3. Ferrer, A.F.; Borrajero, I. Estudio Preliminar Para La Implementación De La Asimilación De Datos De Radares Y Satélites Meteorológicos En Pronósticos Numéricos A Muy Alta Resolución Sobre La Región Occidental De Cuba. 2019. Available online: <http://repositorio.geotech.cu/jspui/bitstream/1234/4186/1/Estudio%20preliminar%20para%20la%20implementaci%C3%B3n%20de%20la%20asimilaci%C3%B3n%20de%20datos%20de%20radares.pdf> (accessed on 20 August 2021).
4. Fernández, D. *Corrección de Errores Sistemáticos en la Inicialización del Modelo WRF Utilizando Asimilación de datos de Radar*; Trabajo Diploma; Instituto Superior de Tecnologías y Ciencias Aplicadas: La Habana, Cuba, 2019.
5. Sierra, M.; Borrajero, I.; Ferrer, A.L.; Morfa, Y.; Morejón, Y.; Hinojosa, M. Estudios de sensibilidad del SisPI a cambios de la PBL, la cantidad de niveles verticales y, las parametrizaciones de microfísica y cúmulos, a muy alta resolución. In *Informe de Resultado*; Instituto de Meteorología: La Habana, Cuba, 2017.
6. Barker, D.M.; Huang, W.; Guo, Y.R.; Xiao, Q.N. A Three-Dimensional (3DVAR) Data Assimilation System For Use With MM5: Implementation and Initial Results. *Mon. Weather Rev.* **2004**, *132*, 897–914. [\[CrossRef\]](#)
7. Malakar, P.; Kesarkar, A.P.; Bate, J.; Deshamukhya, A. Apraisal of data assimilation techniques for dynamical downscaling of the structure and intensity of tropical cyclones. *Earth Space Sci.* **2020**, *7*, e2019EA000945. [\[CrossRef\]](#)
8. Liu, C.; Xiao, Q.; Wang, B. An Ensemble based four dimensional-variational data assimilation scheme. Part I Technical formulation and preliminary test. *Mon. Weather Rev.* **2008**, *136*, 3363–3373. [\[CrossRef\]](#)
9. Schwartz, C.S.; Liu, Z.; Huang, X.; Kuo, Y.; Fong, C. Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: Sensitivity to outer loops and vortex relocation. *Mon. Weather Rev.* **2013**, *141*, 4350–4372. [\[CrossRef\]](#)
10. Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The Global Precipitation Measurement Mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [\[CrossRef\]](#)
11. Tong, M.; Sippel, J.A.; Tallapragada, V.; Liu, E.; Kieu, C.; Kwon, I.; Wang, W.; Liu, Q.; Ling, Y.; Zhang, B. Impact of Assimilating Aircraft Reconnaissance Observations on Tropical Cyclone Initialization and Prediction Using Operational HWRF and GSI Ensemble-Variational Hybrid Data Assimilation. *Mon. Weather Rev.* **2018**, *146*, 4155–4177. Available online: <https://journals.ametsoc.org/view/journals/mwre/146/12/mwr-d-17-0380.1.xml> (accessed on 3 October 2021). [\[CrossRef\]](#)