JOURNAL OF
HYDRO-METEOROLOGY

Research Article

Data assimilation for tropical cyclone-induced rainfall forecasting for central Viet Nam

Nguyen Duc Nam[1], Cong Thanh[2], Vu Van Thang[1], Truong Ba Kien[1]*

[1] Viet Nam Institute of Meteorology, Hydrology and Climate Change; ducnam.mi@gmail.com; vvthang26@gmail.com; kientb@imh.ac.vn
[2] Hanoi University of Sciences-Vietnam National University; congthanh1477@gmail.com

*Corresponding author: kientb@imh.ac.vn; Tel.: +84–989903773

Received: 15 November 2024; Accepted: 24 December 2024; Published: 25 March 2025

Abstract: Data assimilation plays a particularly important role in numerical weather prediction models (NWPs). It ingeste observational data into the initial fields of NWPs, improving their initial conditions to better represent the actual atmospheric conditions. This enhancement consequently leads to improved forecast accuracy of NWP. Globally, three-dimensional variational assimilation (3DVAR) and four-dimensional variational assimilation (4DVAR) methods have been widely used, with 4DVAR considered the most advanced technique. This study will be focused on data assimilation methods using the WRFDA system with WRF-ARW Core. This arrtice was investigated cases without data assimilation of 3DVAR and 4DVAR with different assimilation windows. Evaluations the initial fields demonstrate that both 3DVAR and 4DVAR methods effectively improve the model's initial fields by assimilating observational data. This is evidenced by the fact that the RMSE of the 3DVAR and 4DVAR analysis fields is consistently smaller than the RMSE of the original model initial fields when compared to observations and for all fields, including wind, temperature, moisture, and pressure. This study further reveals that the 4DVAR method offers superior improvements in the initial fields compared to the 3DVAR method. Under the same conditions, the RMSE of 4DVAR analysis fields is generally smaller than that of 3DVAR analysis fields. The evaluation based on the POD and FAR indices indicates that the forecast skill of the 4DVAR case is better than that of 3DVAR, especially at larger rainfall thresholds above 50 mm, which is clearly reflected in the 24-hour accumulated rainfall forecast.

Keywords: Data assimilation; 3DVAR; 4DVAR; Tropical cyclone-induced rainfall; Central Viet Nam.

1. Introduction

Data assimilation systems play a critical role in improving the accuracy of numerical weather prediction models. These systems provide optimal estimates of initial conditions an essential prerequisite for precise weather forecasts by processing massive amounts of atmospheric observation data. Notably, modern data assimilation methods, such as three-dimensional variational assimilation (3DVAR) and four-dimensional variational assimilation (4DVAR), have been developed based on pioneering studies [1, 2]. Lorenc transformed the statistical estimation problem into a variational framework, while Talagrand and Courtier employed an adjoint model to produce initial fields consistent with observation data over a finite time window. These two approaches were subsequently integrated to form the 4DVAR data assimilation method, which has been widely implemented by major meteorological centers such as ECMWF (European Centre for Medium-Range Weather Forecasts) and Météo-France.

In Canada, the 3DVAR system was first operationalized in 1997 [3–4] and continuously improved to enhance background error covariance and analytical precision [4]. In 2005, the 4DVAR system was introduced, incorporating advancements related to tangent linear and adjoint models [5]. In the United States, the National Centers for Environmental Prediction (NCEP) developed and operationalized the Gridpoint Statistical Interpolation (GSI) system, initially as a 3DVAR system in 2007 [6–7] and later upgraded to 4DVAR in 2016. Similarly, Japan has been a pioneer in applying 4DVAR, implementing the Meso-4DVar system for regional weather forecasting in 2002 and subsequently updating it for the JMA-NHM model in 2004. Moreover, the 4DVAR method has demonstrated significant advantages over 3DVAR in forecasting rainfall and extreme weather events [8–9]. By leveraging dynamic and physical constraints to refine initial conditions, 4DVAR yields more accurate forecasts, particularly in model systems like the Weather Research and Forecasting (WRF) model.

Three-dimensional variational assimilation (3DVar) [10] and four-dimensional variational assimilation (4DVar) are methods used in the WRF model to improve meteorological analysis and forecasts. Studies [11, 12] show 4DVar significantly enhances rainfall predictions compared to 3DVar by using dynamic and physical constraints to refine initial conditions. Unlike 3DVar, which processes observations at a single analysis time with lower computational demands, 4DVar [13] incorporates observations across a time window, utilizing tangent and adjoint models [14] to propagate analysis increments. 4DVar also ensures better balance in initial conditions, making it a promising tool for future advancements [15]. Both methods minimize a cost function to reduce errors between observations and forecasts. Variational techniques enable assimilation of conventional and radar data [10, 16]. Comparisons show 4DVar outperforms 3DVar in cyclone and short-term rainfall forecasting [17, 18].

In Vietnam, the National Center for Hydro-Meteorological Forecasting has applied the 3DVAR method within the High-Resolution Model (HRM) since 2007, utilizing MTSAT satellite data [19]. Recently, the combination of 3DVAR and Local Ensemble Transform Kalman Filter (LETKF) in the WRF model has been adopted to enhance forecasts of heavy rainfall and tropical cyclone formation. The study [20] has shown that integrating 3DVAR and LETKF improves initial conditions while accurately predicting the location and timing of tropical depression formation over the Bien Dong Sea. The study [21] used satellite wind data combined with the Ensemble Kalman Filter method to improve the initial conditions of the WRF model. The study [22] used the WRF model with radar data assimilation via 3DVAR to simulate heavy summer rainfall in Ho Chi Minh City. They tested cold and warm start modes with three assimilation configurations, finding that warm start with reflectivity and radial wind improved rainfall forecasts. The study [23] applied data assimilation for heavy rainfall forecasts in the Central Highlands, showing warm start assimilation with AMVs yielded the best results for 24-hour forecasts. The study [24] used 3DVAR to improve temperature, humidity, and rainfall forecasts by assimilating satellite wind and observational data. The study [25] applied 4DVAR for rainfall forecasting in Southern Vietnam, improving forecasts for both light and heavy rainfall. However, 4DVAR required significantly more computational resources than 3DVAR.

Finally, modern data assimilation systems have continually evolved, offering significant potential to advance global and regional weather forecasting accuracy. Future developments integrating variational methods with emerging techniques such as ensemble Kalman filtering hold promise for further breakthroughs in meteorology.

2. Materials and methods

2.1. Data collection

The GFS data is a high-resolution global weather dataset provided by the United States National Centers for Environmental Prediction (NCEP). This dataset includes meteorological variables such as temperature, pressure, humidity, wind, and precipitation at the surface and

various atmospheric levels. It is updated daily at 6-hour intervals and offers global coverage. In this study, GFS data with a horizontal resolution of 0.25° latitude-longitude is utilized to provide initial and boundary conditions for the WRF model.

The data used for assimilation includes various types of observational datasets, specifically SOUND, SONDE_SFC, SYNOP, GEOAMV, METAR, AIREP, SHIPS, BUOY, and PILOT, obtained from the Global Telecommunications Systems (GTS) from NCEP.

The GPM (Global Precipitation Measurement mission) rainfall data is used to evaluate the rainfall forecast results of the model. IMERG (Integrated Multi-satellite Retrievals for GPM) is a NASA product that estimates global surface rainfall at a high resolution of 0.1° every half hour, starting from 2000. It is part of the NASA-JAXA Global Precipitation Measurement project, utilizing the GPM Core Observatory satellite as a standard to integrate rainfall observations from an international satellite network. IMERG can be applied globally or in areas with little or no reliable surface observation data. It offers three versions with different latency to meet varying application needs: Early Run (4-hour latency) for rapid response applications, Late Run (14-hour latency) for same-day/next-day applications, and Final Run (3.5-month latency) for post-real-time research. In this study, the author uses the Final Run dataset.

2.2. Methods

Data assimilation (DA) refers to the process of combining observational data with numerical models to estimate the most accurate representation of a system's state. In this artice, three-dimensional variational assimilation (3DVAR) and four-dimensional variational assimilation (4DVAR) for WRF [10] will be appliced for tropical cyclone-induced rainfall forecasting for central Viet Nam.

3DVAR is a variational data assimilation approach that seeks to minimize a cost function based on a three-dimensional snapshot of the atmosphere or system state. The background error covariance of 3DVAR is stationary and does not vary with time; Observations are assimilated at a single time step, and no temporal evolution is explicitly considered; The objective is to minimize a cost function, typically defined as:

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

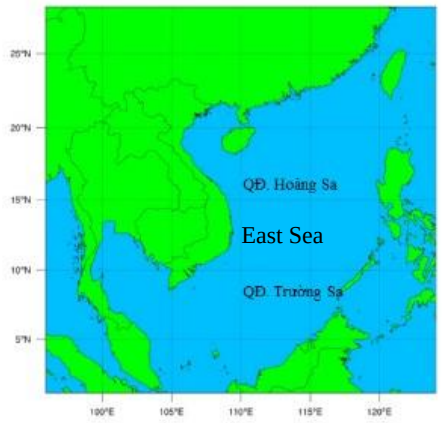
where x is the model state, x_b is the background state, B is the background error covariance matrix, H(x) is the observation operator, y represents observations, and R is the observation error covariance matrix.

4DVAR extends the principles of 3DVAR by incorporating the temporal evolution of the system and observations. It uses adjoint integration to compute the sensitivity of the cost function over time, optimizing initial conditions across a time window. It minimizes a cost function over a time window by using the model's dynamics as a constraint. Both 3DVAR and 4DVAR are integral to modern NWP systems. While 3DVAR is commonly used for operational forecasting due to its computational efficiency, 4DVAR offers higher accuracy for retrospective analysis and research by accounting for the dynamic evolution of the atmosphere.

2.3. Experimental design

The WRF model version 4.3.3 was used with a domain resolution of 1:18 km with 169×169 grid points (Figure 1). The initial and boundary conditions were taken directly from the GFS forecast data at 0.25 degrees latitude and longitude. The WRFDA assimilation program was used to assimilate 3DVAR and 4DVAR in the case of heavy rain due to storms. The physical parameters are presented below.

In this study, Typhoon Noru in 2022 was choiced for investigation. Typhoon Noru made landfall in the Central region in the early morning of September 28. The storm caused heavy rain in the Central Central region on September 27-28 with range of 100-200 mm, particularly, Thua Thien Hue to Quang Ngai, 200-400 mm, some parts over 450mm such as Bach Ma (Thua Thien Hue) and Hoa Binh 597.6 mm, Dau Moi Ho Viet An (Quang Nam) 699.6 mm, Nui Thanh (Quang Nam) 487.2 mm, Ly Son (Quang Ngai) 518.8 mm, Tra Phu (Quang Ngai) 419.6 mm.



Microphysics	WRF single-moment 3-class (Hong et al., 2004)
Shortwave parameters	Dudhia shortwave scheme (Dudhia , 1989)
Long wave parameters	RRTM longwave scheme (Mlawer et al., 1997)
Surface layer parameters	Revised MM5 Scheme (Jimenez et al., 2012)
Surface physical parameters	5–layer Thermal Diffusion Scheme (Dudhia , 1996)
Convection parameters	Grell 3D Ensemble Scheme (Grell, 1993; Grell and Devenyi , 2002)
Boundary layer parameters	Yonsei University Scheme (Hong et al., 2006)

Figure 1. Domain (left) and Physical parameter choices (right).

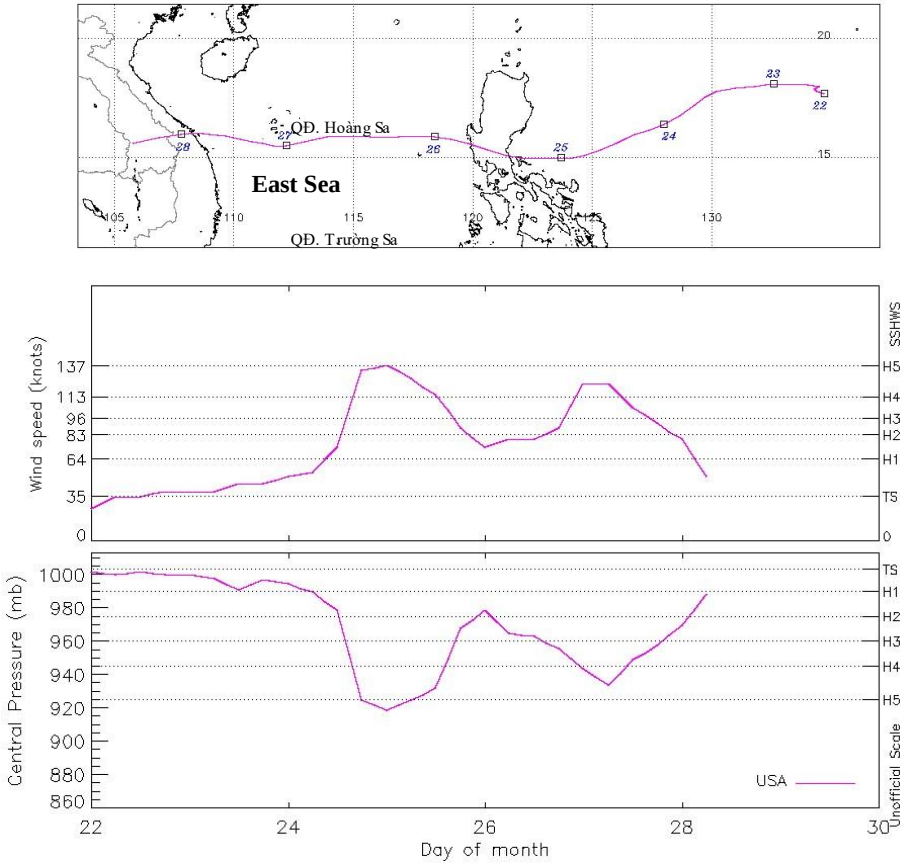


Figure 2. Track (top) and maximum wind speed and minimum pressure of Typhoon Noru in 2022 (Source: JTWC).

The study conducted experiments with the 3DVAR and 4DVAR assimilation schemes with NODA (no data assimilation), 3DVAR (3DVAR with 1 hour assimilation window), 4DVAR_1h (4DVAR with 1 hour assimilation window); 4DVAR_3h (4DVAR with 3-hour assimilation window); 4DVAR_6h (4DVAR with 1 hour assimilation window). The initial time of the model is 00z on September 27, 2022.

3. Results and discussion

The number of observation data with different types of observation data used in 3DVAR and 4DVAR is shown in Table 1. With the 4DVAR assimilation window of 3 hours and 6 hours, it shows that the number of data used for 4DVAR assimilation is more than that of 3DVAR. The number of SYNOP data in 4DVAR increased to 322 data points at 3 hours and 648 data points at 6 hours. The number of METAR data also increased significantly. The number of AMV satellite wind data increased to 3775 and 6294 data points at 3 hours and 6 hours respectively, because this data is updated every hour. In particular, in 4DVAR assimilation, there is an additional type of QSCAT data, which is sea surface wind data, which is not available in 3DVAR. Besides, there are also data types that are almost unchanged in quantity in 4DVAR assimilation compared to 3DVAR, which are data from SOUND, PILOT, BOGUS, SONDE, SHIPS sounding balloons. This shows the ability to assimilate more data in 4DVAR compared to 3DVAR.

Table 1. Number of observation data used in 3DVAR and 4DVAR assimilation.

Data type	3DVAR	4DVAR_1h	4DVAR_3h	4DVAR_6h
Synopsis	316	316	322	648
METAR	105	137	347	664
SHIPS	20	20	21	27
SONDE_SFC	20	20	20	20
GEOAMV	1746	1746	3775	6294
SOUND	2368	2368	2368	2368
PILOT	77	77	77	77
BOGUS	258	258	258	258
QSCAT	0	0	1780	2007

Table 2 shows the average RMSE errors of the wind, temperature, humidity, and pressure fields of the observed data and the data from the initial field (first guess field of the model) of the model (o–g); the average RMSE errors between the observed fields and the 3DVAR and 4DVAR analysis fields. Table 2 shows that almost all RMSEs in (o–a) are always smaller than RMSE og for all wind, temperature, humidity, and pressure fields. This proves that 3DVAR and 4DVAR data assimilation both help improve the initial field of the model.

Table 2. RMSE errors of the wind, temperature, humidity, and pressure between observed data and first guess for 3DVAR and 4DVAR at stations.

	3DVAR		4DVAR 1h		4DVAR_3h		4DVAR_6h	
	o–g	o–a	o–g	o–a	o–g	o–a	o–g	o–a
SYNOP								
U (m/s)	2.0534	1.3619	1.5695	0.2454	1.5453	1.2684	1.7064	1.4143
V (m/s)	1.7026	1.2949	1.5481	0.296	1.5338	1.3607	1.6665	1.4076
T (K)	1.1606	0.9642	1.8106	0.5712	1.8062	1.6332	1.8484	1.7041
Q (kg/kg)	0.0019	0.0015	0.0017	0.0014	0.0017	0.0014	0.0017	0.0015
P (hPa)	1.0976	1.0171	1.1334	0.0354	1.173	1.0618	1.5589	1.1293
METAR								
U (m/s)	1.667	1.4456	1.5965	1.378	1.5427	1.2209	1.6796	1.2037
V (m/s)	1.306	1.1382	1.2688	1.0883	1.4864	1.2131	1.5817	1.2208
T (K)	1.1923	0.9916	1.5553	1.345	1.4737	1.3348	1.711	1.5495
Q (kg/kg)	0.0021	0.0018	0.0019	0.0016	0.0019	0.0016	0.0019	0.0017
P (hPa)	0.6986	0.6167	0.8582	0.6939	1.0961	0.7979	1.1106	0.7676
SHIPS								
U (m/s)	2.4637	2.7829	2.6044	2.9012	2.6044	2.9942	2.6421	3.0712
V (m/s)	3.5266	3.5759	3.628	3.7033	3.628	3.8104	3.4098	3.7151
T (K)	1.4026	1.3305	1.258	1.2079	1.258	1.2237	1.3846	1.3361
Q (kg/kg)	0.0021	0.0019	0.0018	0.0018	0.0018	0.0018	0.0019	0.0018
P (hPa)	3.7344	3.5327	3.6893	3.4035	3.605	3.3545	3.2894	3.0142
SONDE_SFC								

	3DVAR		4DVAR 1h		4DVAR_3h		4DVAR_6h	
	o–g	**o–a**	**o–g**	**o–a**	**o–g**	**o–a**	**o–g**	**o–a**
U (m/s)	1.8139	1.2318	1.7749	1.378	1.7752	1.5209	1.7751	1.4833
V (m/s)	1.3559	1.1624	1.5134	1.1784	1.5116	1.3723	1.5118	1.3963
T (K)	1.0185	0.9504	1.3606	1.304	1.3606	1.2974	1.3606	1.3367
Q (kg/kg)	0.0014	0.0012	0.0012	0.0011	0.0012	0.0012	0.0012	0.0011
P (hPa)	1.2079	1.0791	1.14	1.1328	1.1416	1.167	1.1418	1.15
				GEOAMV				
U (m/s)	3.4532	3.1418	3.4532	3.1405	3.3912	2.9737	3.9322	3.4939
V (m/s)	3.1724	2.911	3.1724	2.9225	3.138	2.8217	3.3681	3.0068
				SOUND				
U (m/s)	3.2573	2.4783	3.2656	2.4728	3.2653	2.4917	3.2656	2.5219
V (m/s)	3.1076	2.3887	3.1321	2.4267	3.1313	2.4542	3.1311	2.4841
T (K)	1.3052	0.8775	1.301	0.8852	1.3013	0.8968	1.3014	0.8943
Q (kg/kg)	0.0009	0.0007	0.0009	0.0008	0.0009	0.0008	0.0009	0.0008
				PILOT				
U (m/s)	2.6442	2.3732	2.6442	2.386	2.6442	2.4743	2.6442	2.5872
V (m/s)	2.3547	2.093	2.3547	2.0938	2.3547	2.1346	2.3547	2.158
				BOGUS				
U (m/s)	7.1416	5.6395	7.1416	5.659	7.1416	5.756	7.1416	5.8042
V (m/s)	4.4724	3.6813	4.4724	3.6563	4.4724	3.7306	4.4724	3.769
				QSCAT				
U (m/s)					1.6867	1.1001	1.7253	1.1178
V (m/s)					2.0798	1.1713	2.0911	1.2949

The RMSE (o–a) in the 4DVAR_1h case is noticeably smaller than in the 3DVAR case for SYNOP data. For other data types, the RMSE values are similar. In some instances, the RMSE (o–a) for 4DVAR is larger than for 3DVAR, possibly due to differences in the amount of data assimilated by the two methods, resulting in variations in average errors. A notable exception is observed with AMV satellite wind data: despite 4DVAR_3h utilizing nearly 2,000 more data points than 3DVAR (Figure 2). An average RMSE (o–a) in 4DVAR is smaller than in 3DVAR (Table 2). This suggests the significant role of AMV wind data in enhancing the 4DVAR data assimilation process in this case.

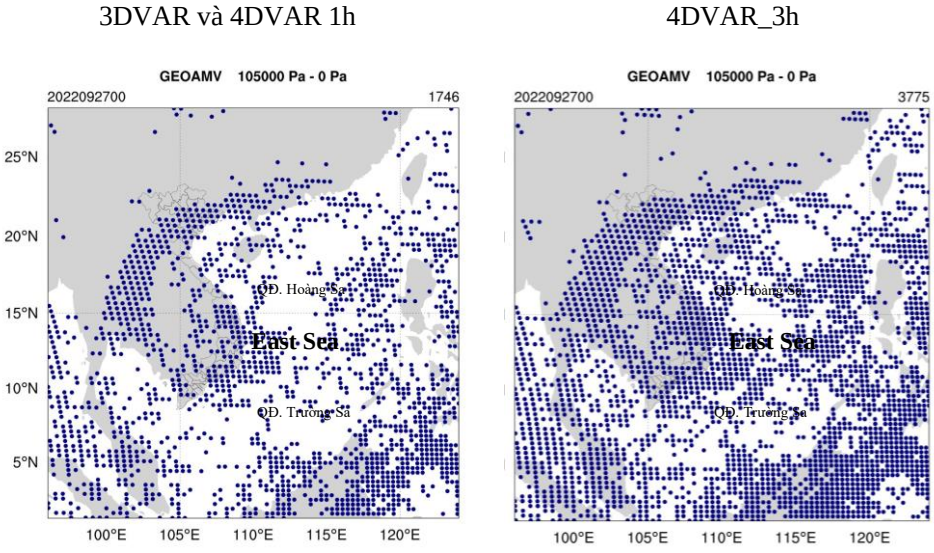


Figure 2. Distribution of AMV wind data points in the case of 3DVAR and 4DVAR_3h.

The above analysis compares the errors of the 3DVAR and 4DVAR background and analysis fields with each type of observation data. The analysis has shown the improvement of the assimilated analysis field compared to the initial background field of the model with the errors in the analysis fields being lower than the errors of the initial background field. Next, the author analyzes the changes of the entire domain of the analysis field compared to

the initial background field of the model. This change is shown in the figures below. The figures below show the increase of the temperature variables T (K), wind U, V (m/s), Q (g/kg) of the assimilated analysis field compared to the initial field of the model, also known as the increment. Specifically, it is the calculation of the analysis field minus the initial background field of the model. The analyses were performed at level 5 of the model, which is the fifth level in a total of 34 vertical levels dividing the atmosphere from the surface to the top of the troposphere. At this level, the author finds that the values of the increments show relatively clearly the difference between the analysis field and the background field.

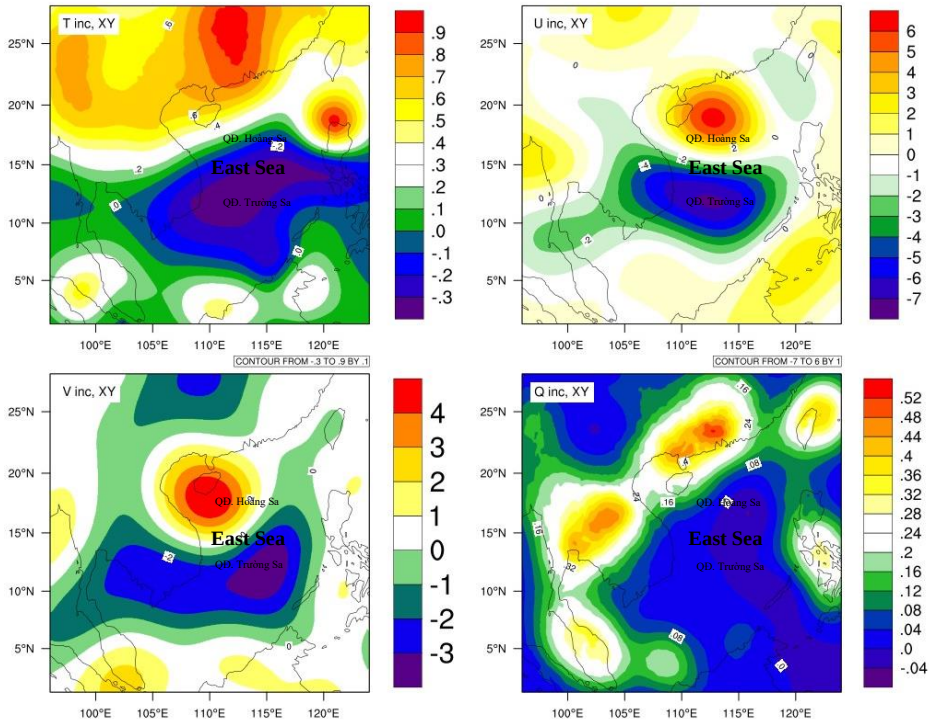


Figure 3. 3DVAR analysis results minus the background field at level 5 of the model with variables T (K), U (m/s), V (m/s), Q (g/kg).

Figure 3 shows the increase in the 3DVAR analysis field compared to the initial field at level 5 of the model. The temperature field here does not change much, the increase in temperature ranges from -0.3 to 0.9 degrees K. The temperature decreases in the central area of the East Sea and increases in the north of the calculation domain, over the mainland of China. Similar to the humidity field Q, the change compared to the initial field is not much, ranging from -0.04 g/kg to 0.52 g/kg. The humidity field mainly increases compared to the initial field of the model, but the increase is not much. With the wind field, the meridional wind speed U and latitudinal wind speed V have changes mainly around the center of the storm. The wind field U has an increase of about -7 to 6 m/s, the wind field V has an increase of about -3 to 4 m/s compared to the base field.

Figures 3, 4 shows that the increase of the analysis field using the 4DVAR method with an assimilation window of 1 hour. With the dynamic window similar to the 3DVAR case. The amount of observation data used in this case is almost the same as the 3DVAR case. It can be seen that the change of the 4DVAR analysis field with an assimilation window of 1 hour has a distribution quite similar to the 3DVAR case in Figure 3. There is only a slight change in the temperature and humidity distribution in Figure 4.

Figure 5 for 4DVAR with 3h window time, the observation data used in this assimilation case is larger than the two cases above. Thus, it can be seen that the change in the temperature field T and the humidity field Q is larger in this case. The temperature field has a change in the distribution of the additional values. The humidity field in this case has an increased value from about -1.4 to 1.6 g/kg. The increase in the humidity field accounts for the majority, the

decrease in the humidity field appears in some points in the Southeast of the storm center, west of the Philippines. The wind fields U, V in this case have a similar distribution of changes to the 3DVAR case but the value is about -4 to 5 m/s for the meridian direction U and -4 to 2.5 m/s for the latitude wind direction V.

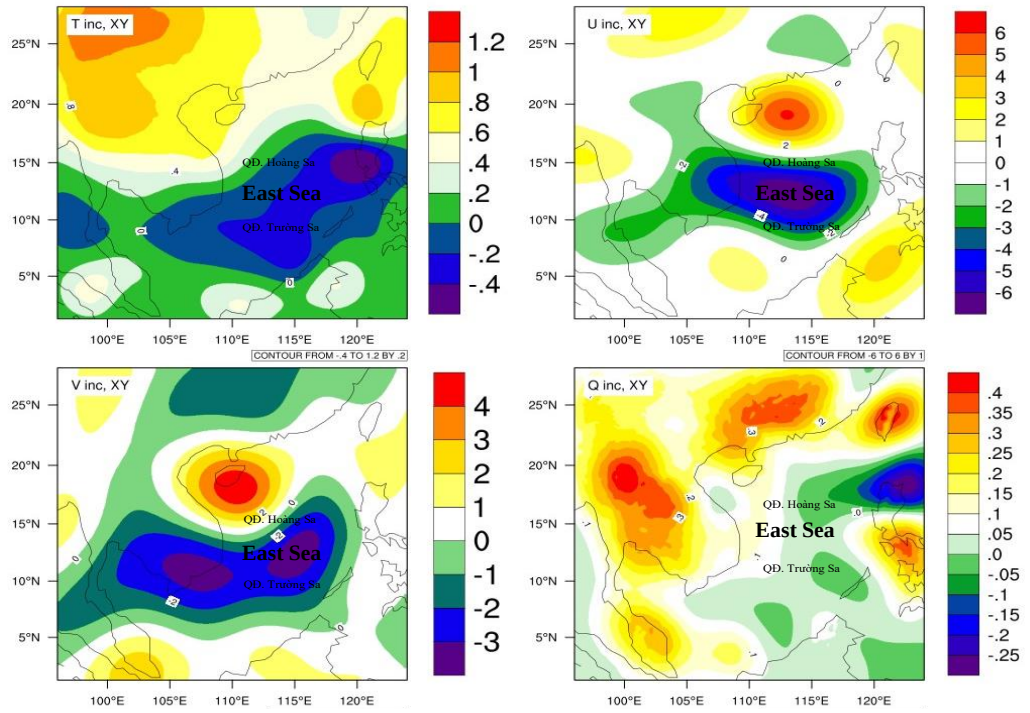


Figure 4. 4DVAR_1h analysis results minus the background field at level 5 of the model with variables T (K), U (m/s), V (m/s), Q (g/kg).

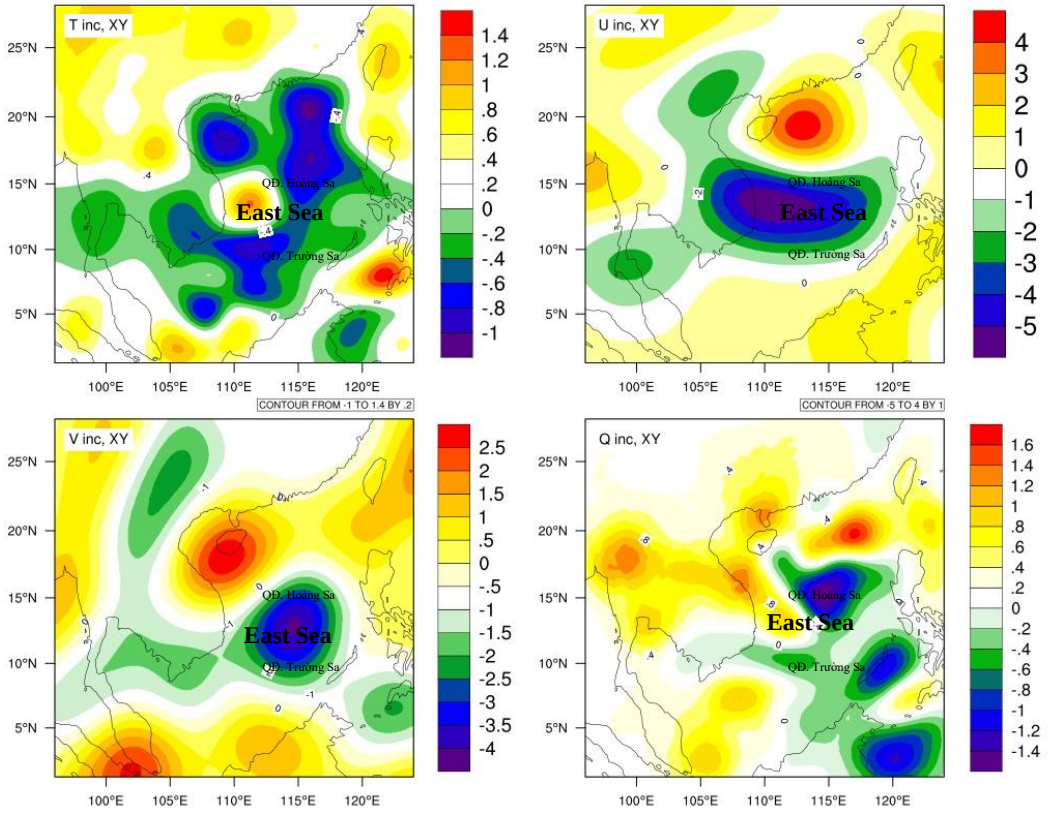


Figure 5. 4DVAR_3h analysis results minus the background field at level 5 of the model with variables T (K), U (m/s), V (m/s), Q (g/kg).

Figure 6 with a 4DVAR assimilation window of 6 hours with distribution of changes in variables is not much different from that in the case of a 3-hour assimilation window. The temperature and wind fields in the 3-hour and 6-hour cases have relatively similar distributions. With a more obvious change in the humidity field, the humidity in the North Central region has a clear increase with a value of up to 2 g/kg. This increase in value is larger than in the case of a 3-hour assimilation window with a value of about 1.2 g/kg. In the North East Sea region, the increase in humidity is also lower than in the case of a 3-hour assimilation window.

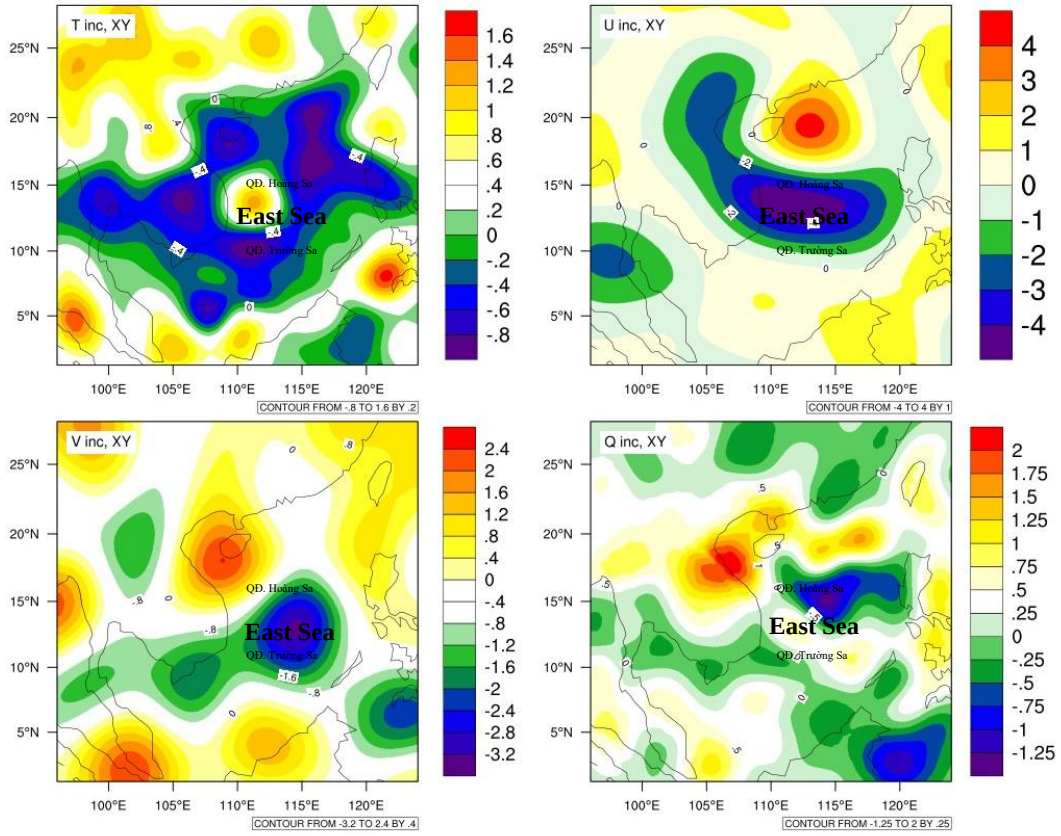


Figure 6. 4DVAR_6h analysis field minus the background field at level 5 of the model with variables T (K), U (m/s), V (m/s), Q (g/kg).

3.2. Rainfall forecast results of the 3DVAR and 4DVAR methods

This section presents the evaluation of the rainfall forecast results from the model under three different cases: no data assimilation, 3DVAR data assimilation, and 4DVAR data assimilation. Figure 7 shows the model's rainfall forecast for the first hour in cases with no data assimilation (NODA), 3DVAR data assimilation, and 4DVAR data assimilation, compared to the corresponding hourly rainfall values from GPM data. The rainfall distribution in the first hour for NODA and 3DVAR cases is relatively similar. According to the GPM data, the rainfall associated with Typhoon Noru has affected the coastal region of Central Vietnam. Both NODA and 3DVAR fail to capture this. In the 4DVAR case, there is an improvement in the rainfall pattern in the western part of the storm center over the coastal region of Central Vietnam. It is evident that this rainfall area has thickened and caused rainfall in this region.

At the time after the model's 15-hour integration (Figure 8), the hourly rainfall at this point shows that the storm is approaching landfall in Central Vietnam. The storm's rainfall area has covered the Central region. In the NODA, 3DVAR, and 4DVAR cases, the 4DVAR forecast still performs better, with a thicker rainfall area over Central Vietnam compared to NODA and 3DVAR, which aligns with the GPM rainfall distribution.

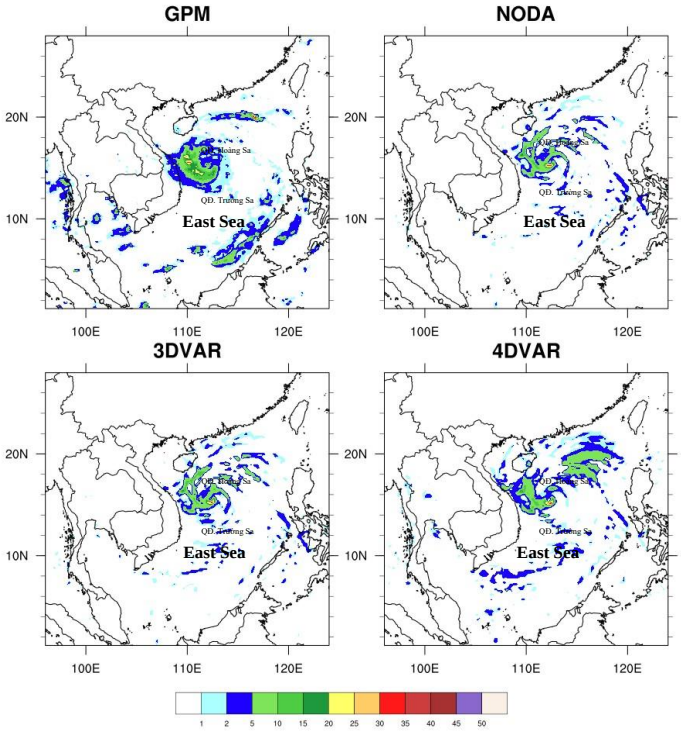


Figure 7. Hourly rainfall forecast results of the model in the first hour for different cases compared with GPM rainfall data (mm).

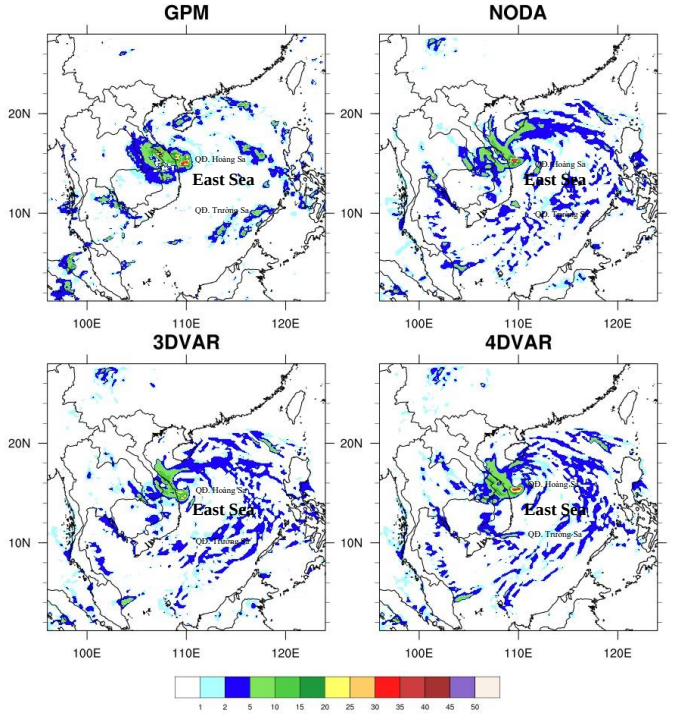


Figure 8. Hourly rainfall forecast results from 15z to 16z on September 27th of the model, compared with GPM rainfall data (mm).

The 24-hour accumulated rainfall for NODA, 3DVAR, and 4DVAR cases is compared with the GPM rainfall data on September 27, 2022 (Figure 9). Overall, the 24-hour accumulated rainfall for NODA, 3DVAR, and 4DVAR appears relatively similar. However, upon closer inspection, the 4DVAR assimilation case shows a larger rainfall distribution over the sea compared to NODA and 3DVAR. On land, the rainfall over Central Vietnam and Southern Laos is also higher in the 4DVAR case compared to NODA and 3DVAR. This is more consistent with the GPM rainfall data.

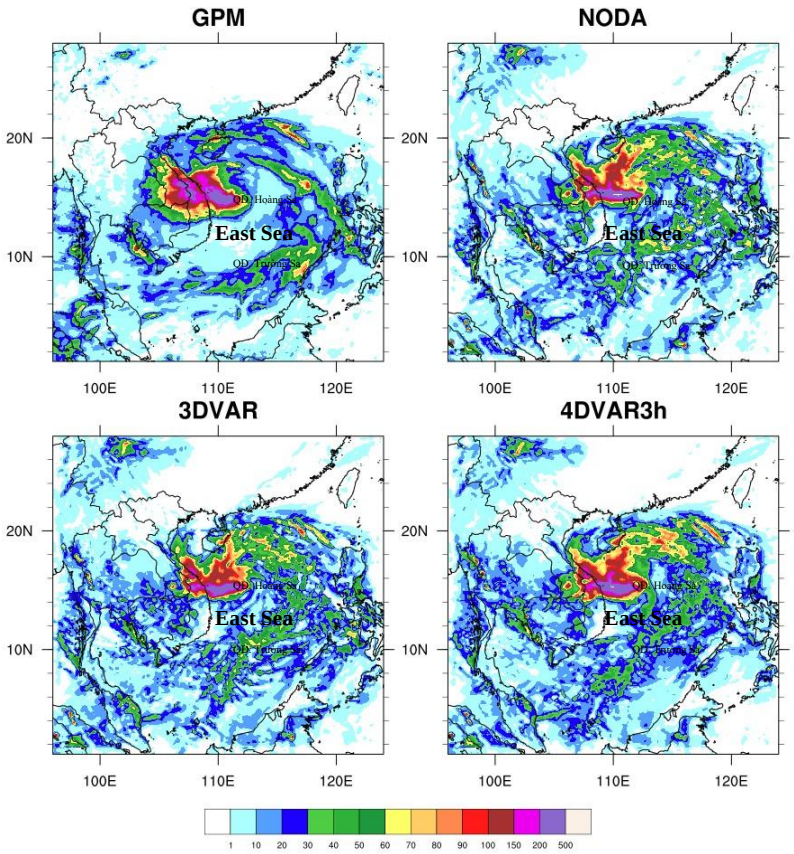


Figure 9. 24-hour accumulated rainfall forecast from 00z on September 27th to 00z on September 28th, 2022 for NODA, 3DVAR, and 4DVAR3h cases compared with GPM rainfall data.

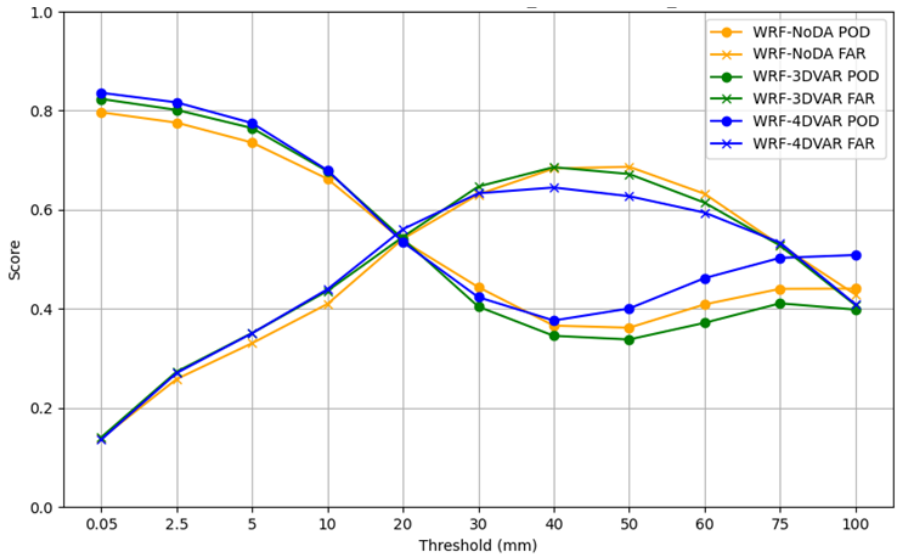


Figure 10. Plot showing POD and FAR indices evaluating the 24-hour accumulated rainfall forecast on September 27, 2022 from the model compared to GPM grid rainfall data.

To assess forecast skill in detail and compare between different cases, two indices, POD and FAR, were used for evaluation. Figure 10 shows the POD and FAR indices evaluating the 24-hour accumulated rainfall forecast on September 27, 2022, for the model compared with the GPM grid rainfall data across the entire region. The yellow line represents the NODA case, the green line represents the 3DVAR case, and the blue line represents the 4DVAR case. The lines marked with "o" represent the POD index, and those marked with "x" represent the FAR index. The POD scores for the 4DVAR assimilation case are better

than the other two cases at most rainfall thresholds. Especially for higher rainfall thresholds above 50 mm, the 4DVAR case shows clear improvement compared to 3DVAR and NODA. When comparing 3DVAR and NODA, 3DVAR performs better at smaller rainfall thresholds, but for larger rainfall thresholds, 3DVAR has worse performance than NODA. Regarding the FAR index, 4DVAR also performs better at higher rainfall thresholds. The false alarm rate (FAR) for 3DVAR is lower than NODA at higher thresholds, but not significantly.

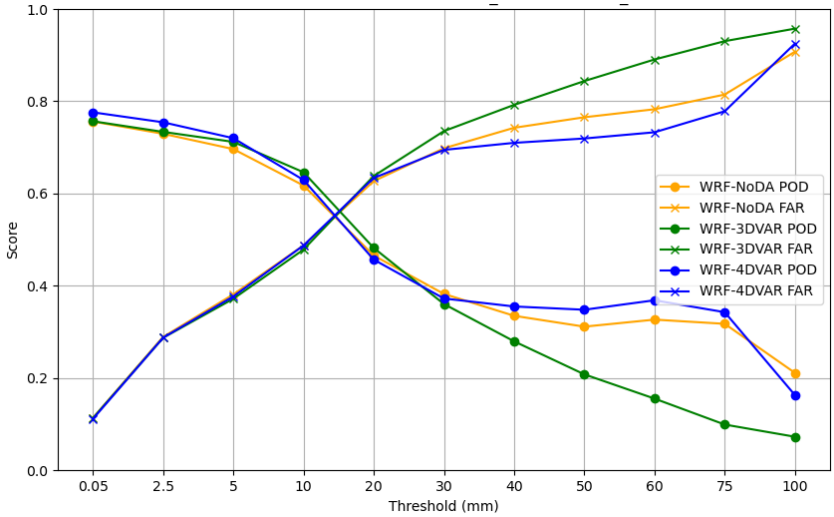


Figure 11. Plot showing POD and FAR indices evaluating the 24–hour accumulated rainfall forecast on September 28, 2022 from the model compared to GPM grid rainfall data.

The evaluation of the forecast for September 28th is shown in Figure 11. Similar to Figure 10, the results are evaluated using the POD and FAR indices for the 3DVAR, 4DVAR, and NODA cases. The evaluation results show that the 4DVAR assimilation case still performs better in both POD and FAR indices compared to the other cases. However, these indices are lower compared to the forecast results for September 27th. The results for 4DVAR and NODA are closer to each other, while 3DVAR shows the poorest forecast skill.

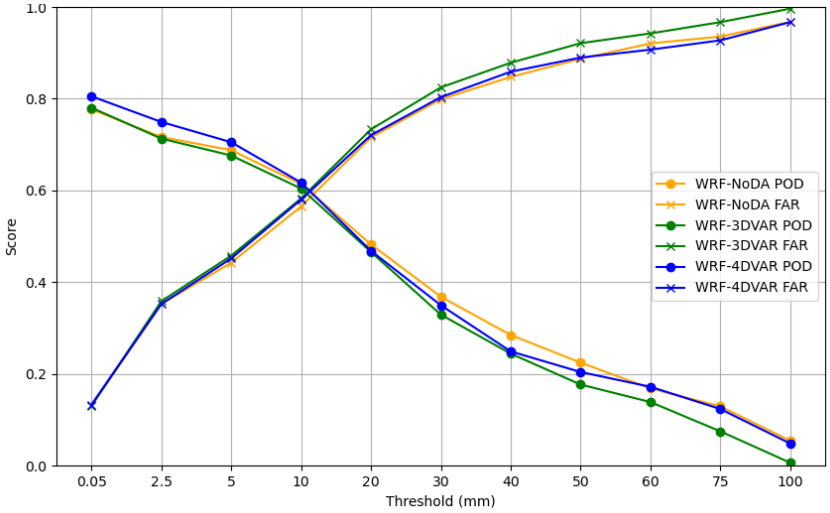


Figure 12. Plot showing POD and FAR indices evaluating the 24-hour accumulated rainfall forecast on September 29, 2022 from the model compared to GPM grid rainfall data.

Figure 12 is similar to Figures 10 and 11. The accumulated rainfall for September 29th is evaluated using the POD and FAR indices. In this case, the 4DVAR assimilation shows better performance only for small rainfall thresholds below 10 mm. For larger rainfall thresholds, the 4DVAR performance is similar to NODA. The forecast skill of 3DVAR remains worse than both 4DVAR and NODA.

4. Conclusion

The results of the study evaluating the 3DVAR and 4DVAR data assimilation methods for improving the initial fields of the numerical model, as presented above, highlight the critical role of data assimilation in enhancing numerical modeling accuracy. Based on these findings, the author draws several conclusions from the test cases analyzed:

Both 3DVAR and 4DVAR data assimilation methods significantly enhance the initial fields of the numerical model by incorporating observational data. This is evident as the Root Mean Square Error (RMSE) values for the analysis fields obtained from both methods are consistently lower than those of the model's initial fields across all variables, including wind, temperature, humidity, and pressure.

Among the two methods, 4DVAR demonstrates a superior ability to improve the model's initial fields compared to 3DVAR. Under identical conditions such as the same set of observational data and assimilation window the RMSE of the 4DVAR analysis fields is generally smaller than that of the 3DVAR fields. This advantage is attributed to 4DVAR's ability to assimilate a larger volume of observational data with finer temporal resolution. Notably, data sources with high time resolution, such as AMV (Atmospheric Motion Vector) satellite wind observations, are better utilized in 4DVAR.

The evaluation results of the rainfall forecast for the cases using the 3DVAR and 4DVAR methods show an improvement in the forecast when using the 4DVAR method compared to the 3DVAR method. The evaluation of the POD and FAR indices indicates that the forecast skill of the 4DVAR case is better than that of 3DVAR, especially at larger rainfall thresholds above 50 mm, which is clearly reflected in the 24-hour accumulated rainfall forecast.

Author contribution statement: Conceived and designed the experiments; Analyzed and interpret the data; contributed reagents, materials, analysis tools or data; data; manuscript editing: T.B.K., C.T., V.V.T., N.D.N.; Performed the experiments; contributed reagents, materials, analyzed and interpret the data: N.D.N.; write the draft manuscript: T.B.K.

Acknowledgements: This study was supported by the project "Research on the application of the Cloud-resolving model integrated with the regional numerical model to a 6-hour accumulated quantitative precipitation forecast with 24-48 hours lead time for Mid- Central Viet Nam", which is funded by the Ministry of Natural Resources and Environment (MoNRE) under grant no. TNMT.2023.06.07. The authors would also like to thank the members of the Vietnam Institute of Meteorology, Hydrology and Climate Change for discussions that improved the quality of the publication.

Competing interest statement: The authors declare no conflict of interest.

References

1. Lorenc, A.C. Analysis methods for numerical weather prediction. *Q. J. R. Meteorolog. Soc.* **1986**, *112(474)*, 1177–1194.
2. Talagrand, O.; Courtier, P. Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory. *Q. J. R. Meteorolog. Soc*. **1987**, *113(478)*, 1311–1328. https://doi.org/10.1002/qj.49711347812.
3. Gauthier, P.; Thépaut, J. N. Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Weather Rev*. **2001**, *129(8)*, 2089–2102.
4. Laroche, S.; Gauthier, P.; St–James, J.; Morneau, J.; Laroche, S. Implementation of a 3D variational data assimilation system at the Canadian meteorological centre. Part II: The regional analysis. *Atmos. Ocean* **1999**, *37(3)*, 281–307. https://doi.org/10.1080/07055900.1999.9649630.
5. Laroche, S.; Tanguay, M.; Zadra, A.; Morneau, J. Use of adjoint sensitivity analysis to diagnose the CMC global analysis performance: A case study. *Atmosphere Ocean*

2002, *40(4)*, 423–443. https://doi.org/10.3137/ao.400404.
6. Kleist, D.T.; Parrish, D.F.; Derber, J.C.; Treadon, R.; Errico, R.M.; Yang, R. Improving incremental balance in the GSI 3DVAR analysis system. *Mon. Weather Rev*. **2009**, *137(3)*, 1046–1060.
7. Derber, J.C.; Parrish, D.F.; Lord, S.J. The new global operational analysis system at the national meteorological center. *Weather Forecasting* **1991**, *6(4)*, 538–547.
8. Ishikawa, Y.; Koizumi, K. Meso–scale analysis. Outline of the operational numerical weather prediction at the Japan Meteorological Agency. Appendix to WMO Technical Progress Report on the Global Data–Processing and Forecasting System (GDPFS) and Numerical Weather Prediction (NWP). Japan Meteorological Agency, Tokyo, Japan, 2002, pp. 26–31.
9. Sato, Y. Introduction of variational bias correction technique into the JMA global data assimilation system. CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modeling, Rep, 2007, 37, pp. 0126–0127.
10. Barker, D.M.; Huang, W.; Guo, Y.R.; Bourgeois, A.J.; Xiao, Q.N. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Weather Rev*. **2004**, *132(4)*, 3241–3258. https://doi.org/10.1175/1520–0493(2004)132<0897:ATVDAS>2.0.CO2.
11. Huang, X.Y.; Xiao, Q.; Barker, D.M.; Zhang, X.; Michalakes, J.; Huang, W.; Henderson, T.; Bray, J.; Chen, Y.; Ma, Z.; Dudhia, J.; Guo, Y.; Zhang, X.; Won, D.J.; Lin, H.C.; Kuo, Y.H. Four-dimensional variational data assimilation for WRF: Formulation and preliminary results. *Monthly Weather Rev.* **2009**, *137(1)*, 299–314. https://doi.org/10.1175/2008MWR2577.1.
12. Zhang, F.; Zhang, M.; Poterjoy, J. E3DVar: Coupling an ensemble kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Monthly Weather Rev.* **2013**, *141(3)*, 900–917. https://doi.org/10.1175/MWR-D-12-00075.1.
13. Errico, R.M. What is an adjoint model? *Bull. Am. Meteorol. Soc.* **1997**, *78(11)*, 2577–2591. https://doi.org/10.1175/1520-0477(1997)078<2577:wiaam>2.0.co;2.
14. Sun, J.; Xue, M.; Wilson, J.W.; Zawadzki, I.; Ballard, S.P.; Onvlee-Hooimeyer, J.; Joe, P.; Barker, D.M.; Li, P.W.; Golding, B.; Xu, M.; Pinto, J. Use of nwp for nowcasting convective precipitation: Recent progress and challenges. *Bull. Am. Meteorol. Soc.* **2014**, *95(3)*, 409–426. https://doi.org/10.1175/BAMS-D-11-00263.1.
15. Sun, J.; Crook, N.A. Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.* **1997**, *54(12)*, 1642–1661. https://doi.org/10.1175/1520-0469(1997)054<1642:DAMRFD>2.0.CO;2.
16. Chu, K.; Xiao, Q.; Liu, C. Experiments of the WRF three-/four-dimensional variational (3/4DVAR) data assimilation in the forecasting of Antarctic cyclones. *Meteorol. Atmos. Physics*, 2013, *120(3–4)*, 145–156. https://doi.org/10.1007/s00703-013-0243-y.
17. Sun, J.; Wang, H. Radar data assimilation with WRF 4D-var. Part II: Comparison with 3D-var for a squall line over the U.S. great plains. *Monthly Weather Rev.* **2013**, *141*(7), 2245–2264. https://doi.org/10.1175/MWR-D-12-00169.1.
18. Duc, L.; Thuy, D.L.; Trung, L.H. Construction of moisture fields for the HRM model using geostationary satellite data based on the three-dimensional variational (3DVAR) method. *VN J. Hydrometeorol.* **2007**, *558*, 22–32.
19. Tien, T.T.; Hoa, D.N.Q. Experiments on using WRF model data assimilation of coupled 3DVAR – LETKF in predicting the geneses of tropical cyclones in the Vietnamese East Sea. *VNU J. Sci.: Earth Environ. Sci.* **2018**, *34(1S)*, 77–89. https://doi.org/10.25073/2588-1094/VNUEES.4338.

20. Du, T.D.; Duc, T.N.; Kieu, C. Initializing the WRF model with tropical cyclone real-time reports using the ensemble Kalman filter algorithm. *Pure Appl. Geophys.* **2017**, *174(7)*, 2803–2825. https://doi.org/10.1007/S00024-017-1568-0/METRICS.
21. Thuc, T.T.; Thanh, C. Radar data assimination in WRF model to forecast heavy rainfall at Ho Chi Minh City. *VNU J. Sci.: Earth Environ. Sci.* **2018**, *34(1S)*, 59–70. https://doi.org/10.25073/2588-1094/vnuees.4336.
22. Phuong, L.L.; Nam, P.Q.; Duc, T.Q.; Tan, P.V. An experiment for assimilating different type of data observations in forecasting heavy rainfall over central highlands region due to the impact of hurricane Damrey. *VNU J. Sci.: Earth Environ. Sci*. **2019**, *35(4)*, 121–129. https://doi.org/10.25073/2588-1094/VNUEES.4478.
23. Nam, N.D.; Manh, N.T.; Anh, N.X.; Khuung, P.L.; Linh, N.T.; Hiep, N.V. Application of Data Assimilation for High–Resolution Simulation of Meteorological Variables over Than Uyen area (Lai Chau). *Vietnam J. Hydrometeorol.* **2021**, *724(4)*, 59–71. https://doi.org/10.36335/vnjhm.2021(724).59-71.
24. Thang, V.V.; Thuc, T.D.; Trung, N.Q. Data assimilation with WRF 4d-Var for rainfall forecasting over the south of Vietnam. *VN J. Hydrometeorol.* **2019**, *EME2*, 174–185. https://doi.org/10.36335/vnjhm.2019(eme2).174-185.