MAUSAM, **76**, 4 (October 2025), 1297-1314

MAUSAM

DOI: https://doi.org/10.54302/mausam.v76i4.7052
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No.551.509.313.1:551.501

Comparative verification of GDAS and NCUM analyses using In-situ radiosonde observations

SUJATA PATTANAYAK[#], ASHISH ROUTRAY*, SURYAKANTI DUTTA[$] and V. S. PRASAD[&]

National Centre for Medium Range Weather Forecasting, Ministry of Earth Sciences (MoES),

A-50, Sector-62, NOIDA-201309, UP, India. ([#]sujata.pattanayak@gov.in,

[$]suryakanti.dutta@gov.in, [&]ps.vijapurapu@nic.in)

(*Received 17 June 2025, Accepted 2 August 2025*)

***Corresponding author's email: ashish.routray@gov.in**

सार – इस अध्ययन ने राष्ट्रीय मध्यम अवधि मौसम पूर्वानुमान केंद्र (NCMRWF) के वैश्विक डेटा एसिमिलेशन सिस्टम (GDAS) और एकीकृत मॉडल (NCUM) द्वारा उत्पन्न उच्च-रिज़ॉल्यूशन वाले वैश्विक विश्लेषणों के प्रदर्शन का मात्रात्मक मूल्यांकन किया। अवलोकन नेटवर्कों के बढ़ते घनत्व ने संख्यात्मक मॉडल के प्रदर्शन में सुधार लाने में योगदान दिया है, लेकिन संख्यात्मक मौसम पूर्वानुमान (NWP) प्रणालियों की क्षमता को मान्य और बढ़ाने के लिए विश्लेषण और पूर्वानुमान उत्पादों के कठोर सत्यापन की आवश्यकता है। 2024 के ग्रीष्मकालीन मानसून के दौरान, बेसिक सिस्टम्स कमीशन (CBS) के दिशानिर्देशों के आधार पर, दुनिया भर के 12 अलग-अलग क्षेत्रों में GDAS और NCUM विश्लेषणों के बीच एक तुलनात्मक अध्ययन शुरू किया गया था। रेडियोसॉन्ड (RS/RW) प्रेक्षणों के विरुद्ध दो विश्लेषणों से दैनिक, साप्ताहिक और मासिक समय-सीमाओं पर विभिन्न सत्यापन कौशल अंकों की गणना की गई। मॉडल प्रदर्शन में क्षेत्रीय विविधताएँ देखी गईं, जिनमें दक्षिणी ध्रुव (SP), उष्णकटिबंधीय (TR), क्षेत्रीय विशिष्ट मौसम विज्ञान केंद्र (RSMC), और भारत (IN) क्षेत्रों के लिए भू-संभावित ऊँचाई और तापमान में उच्च त्रुटियाँ देखी गईं। SP क्षेत्र में पवन घटक अधिक महत्वपूर्ण त्रुटियाँ प्रदर्शित करते हैं। हालाँकि, GDAS पियर्सन सहसंबंध और विसंगति सहसंबंध गुणांकों में कम त्रुटियाँ प्रदर्शित करता है। ये निष्कर्ष दुनिया भर के परिचालन केंद्रों पर मॉडल प्रदर्शन की अंतर-तुलना को सुगम बनाते हैं, जिससे NWP प्रणालियों में भविष्य में होने वाले सुधारों की जानकारी मिलती है।

ABSTRACT. This study quantitatively evaluated the performance of high-resolution global analyses generated by the National Centre for Medium Range Weather Forecasting (NCMRWF) Global Data Assimilation System (GDAS) and the Unified Model (NCUM). The growing density of observational networks has contributed to improvements in numerical model performance, but there is a need for rigorous verification of analysis and forecast products to validate and enhance the competency of Numerical Weather Prediction (NWP) systems. A comparative study between GDAS and NCUM analyses was commenced during the 2024 summer monsoon over 12 distinct regions globally based on Commission for Basic Systems (CBS) guidelines. Various verification skill scores were calculated daily, weekly and monthly timescales from two analyses against the radiosonde (RS/RW) observations. Regional variations in model performance are observed, with higher errors in geopotential height and temperature noted for the South Pole (SP), Tropics (TR), Regional Specialized Meteorological Centre (RSMC), and India (IN) regions. Wind components exhibit more significant errors over the SP region. However, the GDAS exhibits lower errors in Pearson correlation and anomaly correlation coefficients. These findings facilitate the inter-comparison of model performance at operational centers worldwide, informing future improvements in NWP systems.

Key words – GDAS/NCUM analyses, Radiosonde, Summer monsoon, Pearson correlation, CBS.

1. Introduction

The National Centre for Medium Range Weather Forecasting (NCMRWF), one of the leading operational centres in India and serves as a pivotal entity in providing analyses from various numerical models by assimilating the quality-controlled observations. Therefore, it is thereby playing a critical role in the country's weather forecasting and climate monitoring endeavours. To fulfil this responsibility, NCMRWF has operationally integrated two distinct modelling systems *viz*. Global Data Assimilation and Forecasting System (GDAS) and the NCMRWF Unified Model (NCUM). These two systems are designed to provide high-resolution global analyses along with 10-day weather forecasts, leveraging advancements in observational networks and modeling techniques. The

GDAS system, which has been operational, since 1994, utilizes the National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) model to generate high-resolution global analysis (Prasad *et al*., 2011). Over the years, this system has undergone significant upgrades and transitions, incorporating advancements in observational networks, including both conventional and non-conventional observations. These upgrades have enabled the GDAS system to better capture complex weather phenomena and improve its predictive capabilities (Prasad *et al*., 2013; 2021). In contrast, the NCUM, which became operational in 2012, is based on the UK Met Office (UKMO) Unified Model to produce analyses (Rajagopal, *et al*., 2012) and real-time 10-day weather forecasts since 2015 (George *et al*., 2016). This model has also undergone substantial developments, enabling it to capture complex weather phenomena better and improve its predictive capabilities over the Indian monsoon region.

Despite the advancements in observational networks and modeling systems, numerical models still face limitations, underscoring the need for rigorous verification of analysis products and ultimately enhancing their predictive capabilities. The skill of Numerical Weather Prediction (NWP) models can vary across different parts of the globe due to several factors. One primary reason is the density and quality of observational data, which significantly impacts model performance. Regions with dense observation networks tend to have better model accuracy, whereas areas with limited data, such as oceans or remote land areas, often experience reduced model performance. Additionally, the complexity of terrain and weather patterns also affects model skill.

The World Meteorological Organisation (WMO) established the Lead Centre for Deterministic NWP Verification (LC-DNV) at the European Centre for Medium-Range Weather Forecasts (ECMWF) to facilitate standardized verification of deterministic NWP forecasts (http://apps.ecmwf.int/ wmolcdnv/). The LC-DNV enables consistent evaluation and comparison of forecast performance across WMO Global Data Processing and Forecasting System (GDPFS) centres, providing operational forecasters with valuable insights and centres with opportunities for improvement. Standardized verification procedures and scores are exchanged between participating centres *via* the LC-DNV, ensuring routine and consistent comparisons of forecast results.

The 16th WMO Congress (2011) designated nine regions for deterministic NWP verification, including the Northern Hemisphere, Southern Hemisphere, Tropics, Asia, North America, Europe/North Africa, Australia/New Zealand, North Pole & South Pole (WMO Tech report, 2017). In the present study, the verification domain was expanded to include three additional regions: Global, Regional Specialized Meteorological Centre (RSMC), and India region. This expanded regional framework enables a more comprehensive evaluation of NWP forecast performance, allowing for a detailed assessment of model accuracy across various spatial scales and geographical domains.

In collaboration with participating centers, including the China Meteorological Administration (CMA), the UKMO, the Tanzania Meteorological Agency (TMA), and the Japan Meteorological Agency (JMA), the LC-DNV generates comprehensive verification reports for surface parameters. Additionally, multiple centers contribute to upper-air verification efforts, including the Australian Bureau of Meteorology (BoM), the Canadian Meteorological Centre (CMC), the NCEP, the German Weather Service (DWD), Météo-France, the Korean Meteorological Administration (KMA), and JMA. The resulting verification scores and metrics are archived and made accessible through the ECMWF ftp server (ftp://wmolcdnv@ftp.ecmwf.int/).

A comparison of five numerical weather prediction analyses climatology showed that ECMWF, UKMO, and NCEP-NCAR reanalysis performed well for sea level pressure and 500-hPa height. ECMWF, Environment and Climate Change Canada (ECCC), KMA, NCEP, and UKMO showed better prediction skills for monthly precipitation forecasts, with ECMWF performing the best among 11 models analysed (Endris, *et al*., 2021; Jolliffe and Stephenson, 2003; 2012; Connolley and Harangozo, 2001). Keep in mind that model performance can vary depending on the specific weather pattern, region, and forecasting timescale. These studies highlight the importance of ongoing evaluation and improvement of weather prediction models.

Radiosonde observations remain a vital component of upper-air observations, providing critical data that significantly impacts NWP systems. It provides a detailed vertical profile of atmospheric conditions, including temperature, humidity, wind speed, and wind direction. These observations significantly impact NWP systems by offering high-altitude and high-resolution data, particularly over regions where other observation platforms may be limited. By validating NWP model performance, radiosonde observations also help identify areas for improvement and optimize model configurations, ultimately contributing to more accurate and reliable weather forecasting.

This study conducts a comprehensive verification of GDAS and NCUM analyses using radiosonde data, generating daily, weekly, and monthly statistics for 12 distinct regions covering the entire globe. These regions include 9 WMO-approved regions for deterministic NWP verification, as well as three additional regions: global, RSMC, New Delhi, and India. The verification process involves a detailed analysis of various metrics, including mean error, root mean square error, and anomaly correlation, to assess the performance of both models.

The Indian summer monsoon season, which spans from June to September (JJAS), is a critical period for understanding its variability in the tropical region. This season plays a pivotal role in modulating global climate variability. The present study focuses on the performance of NCMRWF global analyses over different regions of the globe during these months, with a particular emphasis on understanding the strengths and limitations of both models through verification statistics.

The study is structured as follows: Section 2 outlines the data and methodology employed with a brief description of the models used, highlighting their key features and differences, including the sources of radiosonde data and the software used along with the verification metrics used. Section 3 highlights key findings showcasing the models performances in daily, monthly and seasonal timescales. Finally, Section 4 summarizes the conclusions drawn from this study, providing insights into the strengths and limitations of both models and informing future improvements in NWP systems.

2. Data and methodology

This section provides an overview of the assimilation frameworks employed by the GDAS and NCUM and illustrates in-depth the description of the data utilized, the methodology employed, and the areas of evaluation and validation for the study.

2.1. *Global data assimilation system*

The GDAS at NCMRWF has undergone significant advancements over three decades, with notable upgrades in resolution, assimilation techniques, and computational infrastructure. The system progressed from a T80L18 configuration with Spectral Statistical Interpolation (SSI) in 1994 to higher resolutions (T1534 L64) and more sophisticated assimilation methods, including Grid-point Statistical Interpolation (GSI), Hybrid GSI, and 4D-EnVAR. These updates, aligned with the NCEP GFS operational suite, have enhanced the system's accuracy and reliability (Rajagopal *et al*., 2007; Prasad *et al*., 2011). A detailed description of the analysis system can be found in Kleist *et al*., 2009, while the results of pre-implementation tests conducted at NCMRWF are presented in Rajagopal *et al*., 2007. These studies provide a comprehensive understanding of the GDAS system's development, implementation, and performance over the years. The continuous upgrades and improvements to the GDAS system have enabled NCMRWF to provide high-quality analyses and support a wide range of applications. Over the years, this system has undergone significant upgrades and transitions, incorporating advancements in observational networks, including both conventional and non-conventional observations. These upgrades have enabled the GDAS system to better capture complex weather phenomena and improve its predictive capabilities (Prasad *et al*., 2013; 2014; 2017; 2021). The system's ability to assimilate data from various sources and generate accurate analyses has made it an essential tool for weather forecasting and research in India.

2.2. *NCMRWF unified model*

The NCUM is a state-of-the-art, seamless prediction system designed to provide 10-day forecasts using initial conditions from 00 and 12 UTC on a routine basis, since 2012. The NCUM is an adaptation of the Unified Model (UM), which is available under the UM Partnership on the Met Office Shared Repository Service. The NCUM assimilation-forecast system has undergone periodic upgrades to incorporate new scientific and technological advancements, with the goal of improving numerical forecasts. A study by Rajagopal *et al*., 2012 and George, *et al*., 2016 highlights the major components of the NCUM system, adapted from the UK Met office Unified Model (UM11.2). Kumar *et al*., 2018; 2020 discuss the NCUM system's hybrid data assimilation, combining 4D Variational Analysis (4DVAR) with an Ensemble Transform Kalman Filter (ETKF). The details of 4D-Var system used in the NCUM can be found from the study by Rawlins *et al*., 2007.

The NCUM global system is a comprehensive, end-to-end system that encompasses several critical components, including an observation processing system, Hybrid 4D-Var data assimilation system, Surface data assimilation/preparation system, high spatial resolution, *etc*., all functioning at a high spatial resolution of 12 km. The in-house development of the observation pre-processing system and data assimilation with new observations is a major component of the NCUM system, allowing for customized processing of observational data used in the analysis system. The NCUM system has been designed to provide accurate and reliable forecasts, leveraging the latest scientific and technological advancements in NWP. The system's end-to-end architecture and high spatial resolution enable it to capture complex weather phenomena and provide detailed forecasts, making it a valuable tool for weather

TABLE 1

Domains used for verification in this study

Sl. No.	DOMAINS	DEFINITION
1	Global (GL)	90° S - 90° N; inclusive all longitudes
2	Northern Hemisphere (NH)	20° N - 90° N; inclusive all longitudes
3	Southern Hemisphere (SH)	20° S - 90° S; inclusive all longitudes
4	Tropics (TR)	20° S - 20° N; inclusive all longitudes
5	Asia (AS)	25° N - 65° N; 60° E - 145° E
6	North America (NA)	25° N - 60° N; 145° W - 50° W
7	Europe _N. Africa (EA)	25° N - 70° N; 10° W - 28° E
8	Australia _New Zealand (AN)	10° S - 55° S; 90° E - 180° E
9	North Pole (NP)	60° N - 90° N; inclusive all longitudes
10	South Pole (SP)	60° S - 90° S; inclusive all longitudes
11	RSMC, New Delhi (RS)	10° S - 50° N; 40° E - 110° E
12	India (IN)	5° S - 40 ° N; 50 ° E - 100 ° E

TABLE 2

The maximum number of RS/RW observations as per WMO over different verification area

Domains	Max. No. Radiosonde Observations
GL	718
NH	508
SH	76
TR	135
AS	162
NA	96
EA	89
AN	37
NP	70
SP	12
RS	124
IN	52

forecasting and research applications (Routray *et al*., 2012; 2017; 2019 and Singh *et al*., 2018).

2.3. *Study area*

The models' performances are assessed globally and across various regions, including the Northern Hemisphere, Southern Hemisphere, Tropics, Asia, North America, Europe/North Africa, Australia/New Zealand, North Pole, South Pole, RSMC and India. A detailed description of these verification domains is provided in Table 1. This comprehensive evaluation framework enables a thorough assessment of the models' performance across various spatial scales and regions.

2.4. *Data and software used*

The analyses obtained from GDAS and NCUM for the JJAS period are used for verification. Radiosondes are the primary source of upper-air observations, providing critical data on wind speed, wind direction, pressure, temperature, and humidity. The integration of GPS technology has significantly enhanced the accuracy and reliability of these measurements. This study utilizes radiosonde observation stations adhering to WMO Commission for Basic Systems (CBS) standard verification protocols, ensuring consistency and comparability of data across different regions and countries, as outlined in Table 2. The validation of both analyses is conducted by comparing them with Radiosonde/Rawinsonde (RS/RW) observations received at NCMRWF, which are available every 12 hours interval over the specific regions (Table 1), utilizing the Model Evaluation Tool (MET) for evaluation. The regional distribution of RS/RW observations reveals a maximum frequency of 124 observations over RS region and 52 observations over the IN region. These observational data are extracted from the global WMO distribution dataset, highlighting the regional variability in atmospheric sounding measurements. Currently, IMD maintains an operational network comprising 39 RS/RW observatories throughout India, facilitating the acquisition of comprehensive upper-air observational data.

The MET is a cutting-edge verification framework developed by the Developmental Testbed Center (DTC) at the National Oceanic and Atmospheric Administration (NOAA) to evaluate NWP products (Clark *et al*., 2011; 2012; Bullock *et al*., 2016), primarily designed for the Weather Research and Forecasting (WRF) community. However, its versatility allows for adaptation to other models, provided specific format requirements are met. MET provides a comprehensive framework for assessing model accuracy, skill, and reliability. It calculates various verification metrics, such as bias, mean error, and root mean square error, and enables spatial and temporal analysis of model performance. MET's key features include verification metrics, spatial and temporal analysis, multi-model comparison, and object-based verification. These features enable users to evaluate model performance from different perspectives, identify strengths and weaknesses, and compare models. By leveraging MET's capabilities, the improvements in forecasting accuracy and reliability, ultimately supporting better decision-making in various applications. To facilitate validation, both GDAS and NCUM analyses are converted to the requisite format. The MET tool, accessible at (https://dtcenter.org/community-code/model-evaluation-tools-met), generates statistical evaluations using the point-stat package, enabling comprehensive assessment of model performance. The MET verification package requires gridded model output on a standard de-staggered grid with pressure levels in the vertical as input. To achieve this, the unified post-processor in the WRF source code is utilized. Additionally, the "copygb" utility is employed to re-grid analysis and observation datasets in GRIB version 1 format to a common verification grid. Specifically, both GDAS and NCUM analyses are de-staggered using the "copygb" utility prior to validation. The Point-stat Evaluation Package within MET is then used to generate verification statistics, taking the re-gridded and de-staggered datasets as input and producing statistical outputs for model evaluation.

2.5. *Verification statistics*

A suite of verification statistics, as recommended by Wilks (2011) and utilized by WMO lead centers for deterministic verification of NWP products (JWGFVR, 2009), is approved by the WMO congress (2011) and employed to evaluate model performance (https://confluence.ecmwf.int/display/WLD/Standard+verification+procedures). The statistics include mean error (ME), root mean square error (RMSE), Pearson correlation coefficient (*PR_CORR)* and anomaly correlation coefficient (*ANOM_CORR).* These metrics are assessed over 12 distinct regions to comprehensively evaluate model performance.

The Mean Error (ME) is a measure of overall bias for continuous variables, which can be presented as the average difference between predicted and observed values, indicating the accuracy of a model or forecasting system. A positive ME suggests over-prediction, while a negative ME indicates under-prediction. The ME provides an overall indicator of accuracy, with smaller values signifying better performance. A ME close to zero indicates an unbiased and accurate model, while a large ME highlights the need for adjustment or improvement. It is defined as,

$$ME = \frac{1}{n} \sum_{i=1}^{n} (f_i - o_i) \tag{1}$$

The daily, weekly, and monthly bias is estimated for the JJAS period from both GDAS and NCUM analysis.

The variance in the model analysis/forecast is defined as,

$$SD_f = \sqrt{\frac{1}{T+1} \sum_{i=1}^{T} \left(f - \bar{f}\right)^2}, \quad \bar{f} = \frac{1}{n} \sum_{i=1}^{n} f_i \tag{2}$$

The variance in the observation is defined as,

$$SD_o = \sqrt{\frac{1}{T+1} \sum_{i=1}^{T} (o - \bar{o})^2}, \quad \bar{o} = \frac{1}{n} \sum_{i=1}^{n} o_i \tag{3}$$

The Root Mean Square Error (RMSE) is a statistical metric that measures the magnitude of errors in a model's predictions by calculating the square root of the average squared differences between predicted and observed values. RMSE provides a sense of the average size of the errors, with larger values indicating greater discrepancies, and is sensitive to outliers, making it useful for identifying models with occasional large mistakes. The RMSE can be defined as,

$$RMSE = \sqrt{\frac{1}{n} \sum (f_i - o_i)^2} \tag{4}$$

The *PR_CORR* is a statistical measure that calculates the strength and direction of the linear relationship between two continuous variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear correlation. By quantifying the linear relationship, *PR_CORR* helps identify whether variables are positively, negatively, or not correlated, making it a widely used tool in various fields. Values close to 1 indicate strong positive linear relationships, while values close to -1 indicate strong negative linear relationships, and values near 0 suggest weak or no linear relationships between the model analysis/forecast and observations. It is defined as,

$$PR\_CORR = \frac{\sum_{i=1}^{T} (f - \bar{f})(o - \bar{o})}{\sqrt{\sum (f_i - \bar{f})^2} \sqrt{\sum (o_i - \bar{o})^2}} \tag{5}$$

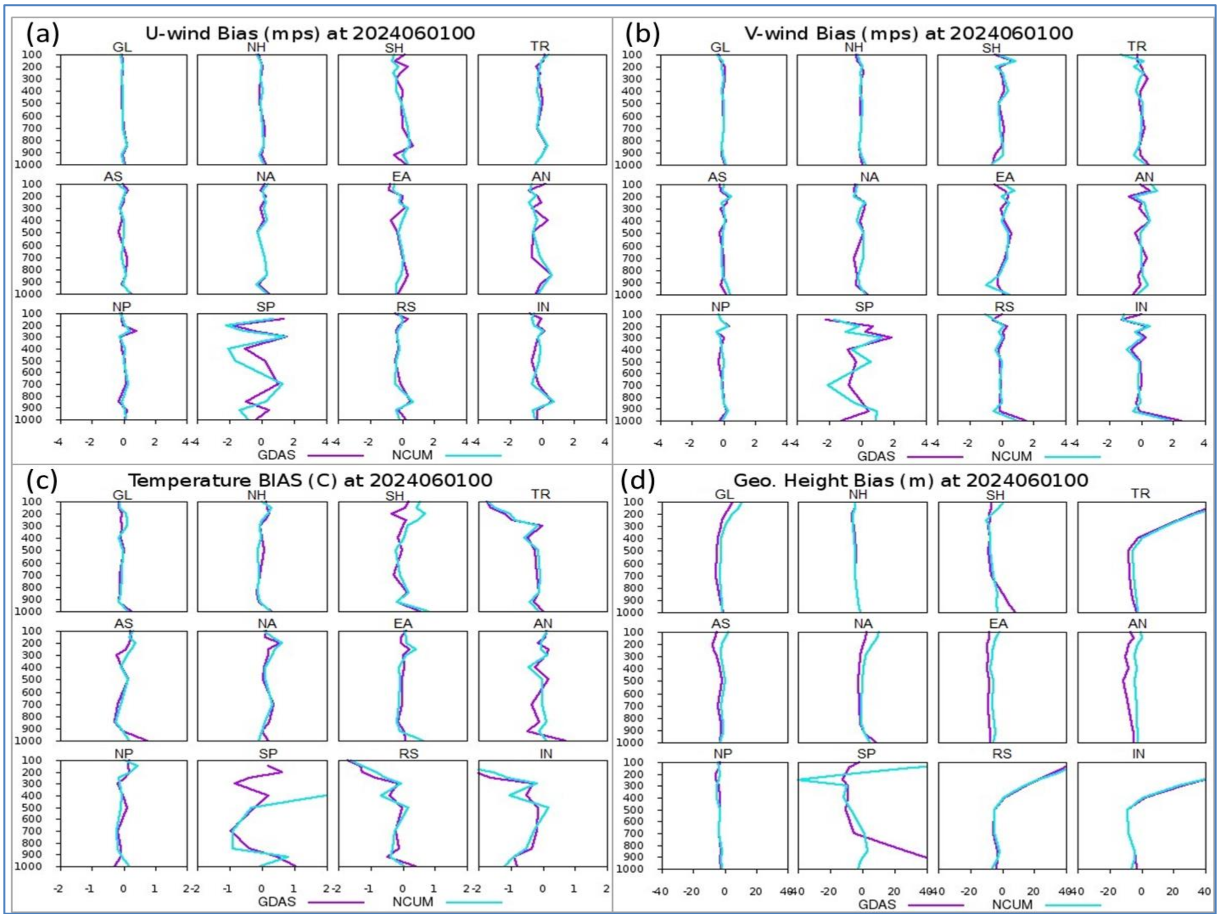
Also, the ($ANOM\_CORR$) is estimated, which is the equivalent of the Pearson correlation coefficient, but both the observation and forecast are first adjusted to the climatological value. It measures the strength of linear association between forecast anomalies and observed anomalies. It is defined as,

$$ANOM\_CORR = \frac{\sum (f_i - c)(o_i - c)}{\sqrt{\sum (f_i - c)^2} \sqrt{(o_i - c)^2}} \tag{6}$$

The skill of the analysis evaluated through Gilbert Skill Score (GSS). It is defined as follows.

$$GSS = \frac{a - \frac{(a+b)(a+c)}{a+b+c+d}}{a+b+c - \frac{(a+b)(a+c)}{a+b+c+d}} \tag{7}$$

where: a = number of correct forecasts (hits), b = number of false alarms, c = number of misses and d = number of correct negatives.



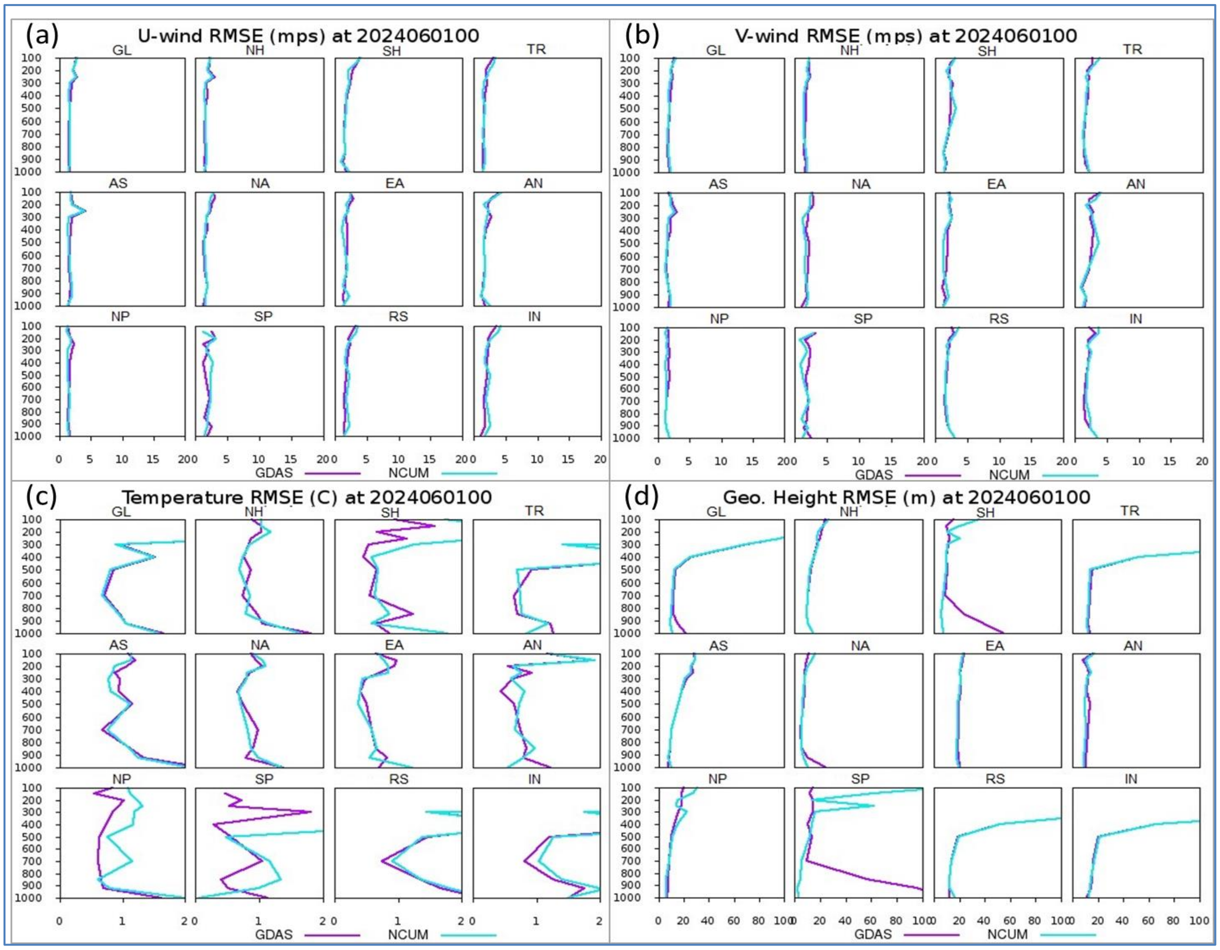
Figs. 1(a-d). Vertical distribution of daily bias from GDAS and NCUM analyses for (a) zonal wind (u; m/s); (b) meridional wind (v; m/s); (c) temperature (t; °C) and (d) geopotential height (m) valid at 00Z 20240601 over 12 specific domains of verification [x-axis: Bias; y-axis: Pressure level]

Error statistics are computed at standard pressure levels, specifically 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, and 100 hPa. Additionally, select metrics, including $ANOM\_CORR$ and GSS, are evaluated at specific pressure levels, namely 850, and 500 hPa.

3. Results and analysis

The performances of both analyses are assessed on daily, weekly, and monthly intervals, employing a range of statistical metrics to evaluate the accuracy of basic meteorological parameters, including temperature, geopotential height, and zonal and meridional wind components. The metrics used include mean error, standard deviation, correlation coefficient, anomaly correlation coefficient, and skill score, which are computed at standard pressure levels to comprehensively assess model performance across various atmospheric conditions and temporal scales.

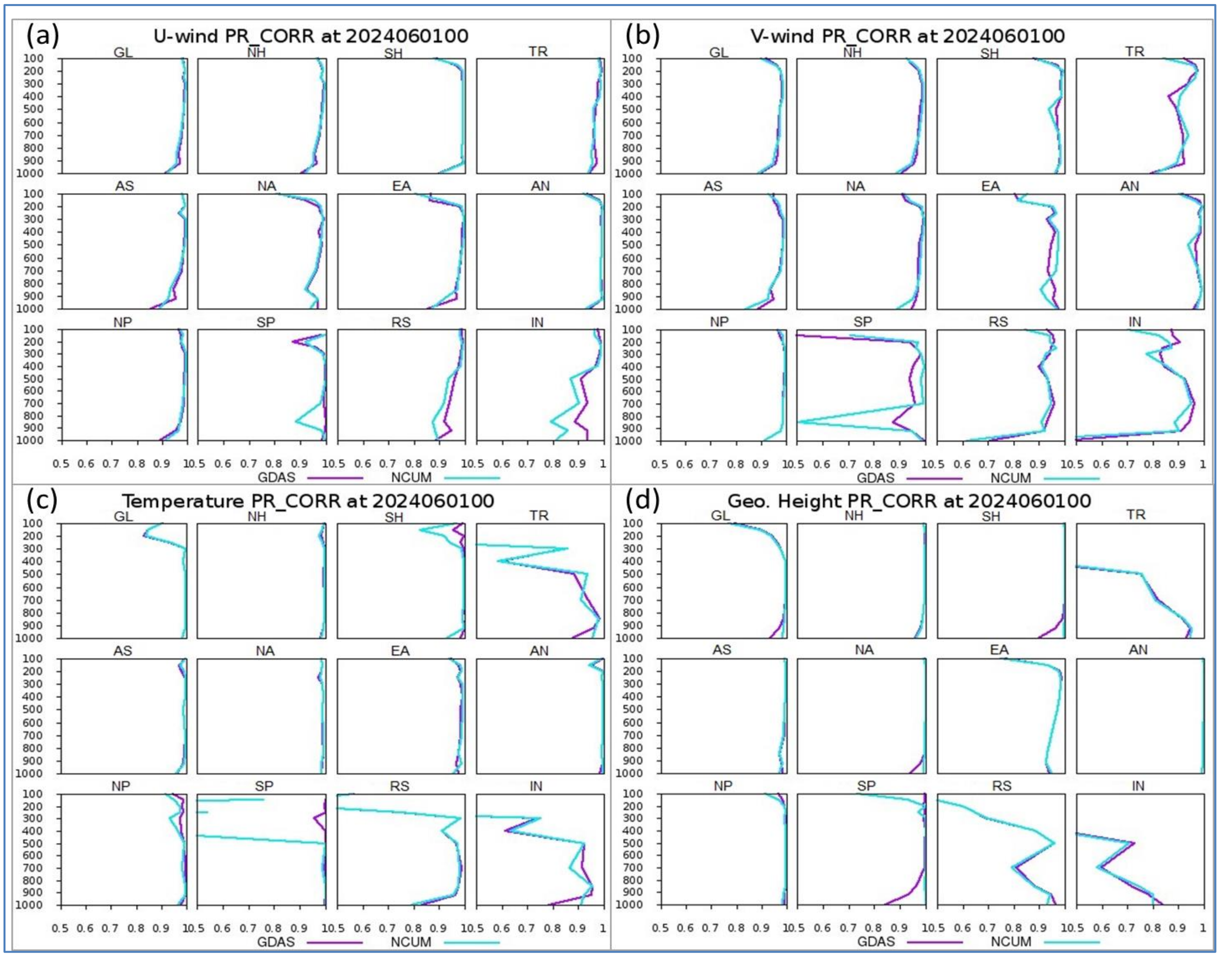
A comprehensive assessment of model performance is conducted across 12 distinct regions, adhering to CBS guidelines. The analysis depicted in Figs. 1(a-d) reveals the bias in wind, temperature, and geopotential height predictions from both GDAS and NCUM models. Notably, both models exhibit comparable performance across most regions, with relatively lower biases observed over the Global (GL), Northern Hemisphere (NH), Asia (AS), North America (NA), Europe (EA), and North Pole (NP) regions. This suggests that the models demonstrate robust predictive capabilities in these regions. Conversely, the Southern Pole (SP), Antarctica (AN), and India (IN) regions display slightly higher biases, which may be attributed to the limited sample sizes available for these regions. Furthermore, temperature biases are marginally higher over SP, AN, and IN regions, indicating potential challenges in accurately capturing thermal dynamics in these areas. The SP region also exhibits a pronounced geopotential height bias, highlighting the need for improved model representation of atmospheric circulation patterns in this



Figs. 2(a-d). Vertical distribution of RMSE from GDAS and NCUM analyses for (a) zonal wind (u; m/s); (b) meridional wind (v; m/s); (c) temperature (T; °C) and (d) geopotential height (m) valid at 00Z 20240601 over 12 specific domains of verification [x-axis: RMSE; y-axis: Pressure level]

region. SP region is characterized by a severely limited radiosonde observation network, with a paucity of data points (maximum of 12 observations as per WMO guidelines; (https://community.wmo.int/en/forecast - verifications). This scarcity of observational data leads to inadequate sampling, resulting in reduced model accuracy and reliability over the SP region. The insufficient data availability hampers the model's ability to capture complex atmospheric processes, ultimately contributing to degraded forecast performance.

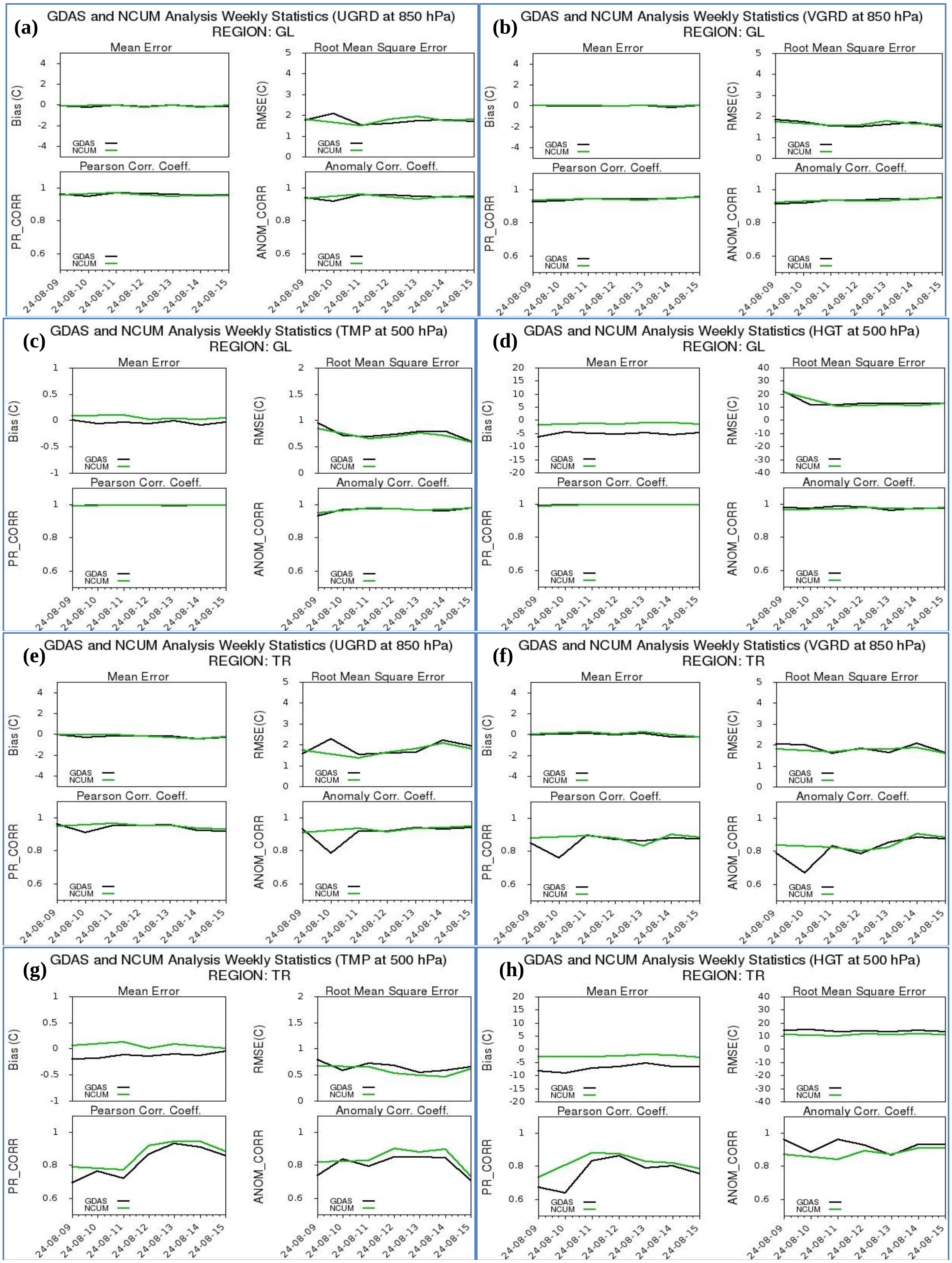
Figs. 2(a-d) presents the RMSE for key meteorological parameters, providing insight into model performance. The RMSE measures the average magnitude of the difference between predicted and observed values. RMSE is widely used to evaluate the accuracy of models, with lower values indicating better performance. The RMSE in wind components reveals a notable trend, with higher errors observed in the lower troposphere (up to 850 hPa) compared to the middle and upper levels. However, the GL, NH, and NA regions exhibit higher RMSE at 200 hPa, suggesting increased uncertainty in wind predictions at this altitude. For meridional wind components, the RMSE is significantly higher over the TR, RS, and IN regions in the lower troposphere, indicating challenges in capturing wind dynamics in these areas. The temperature RMSE displays a characteristic pattern, with higher values observed in the lower and upper troposphere, and lower values in the middle troposphere. This suggests that the model struggles to accurately predict temperature profiles, particularly near the surface and in the upper atmosphere. Furthermore, the geopotential height RMSE is higher in the lower troposphere over the SH and SP regions, indicating difficulties in representing atmospheric circulation patterns in these areas. These findings highlight the need for improved model representation of atmospheric dynamics and thermodynamics, particularly in regions with complex terrain or unique meteorological conditions.



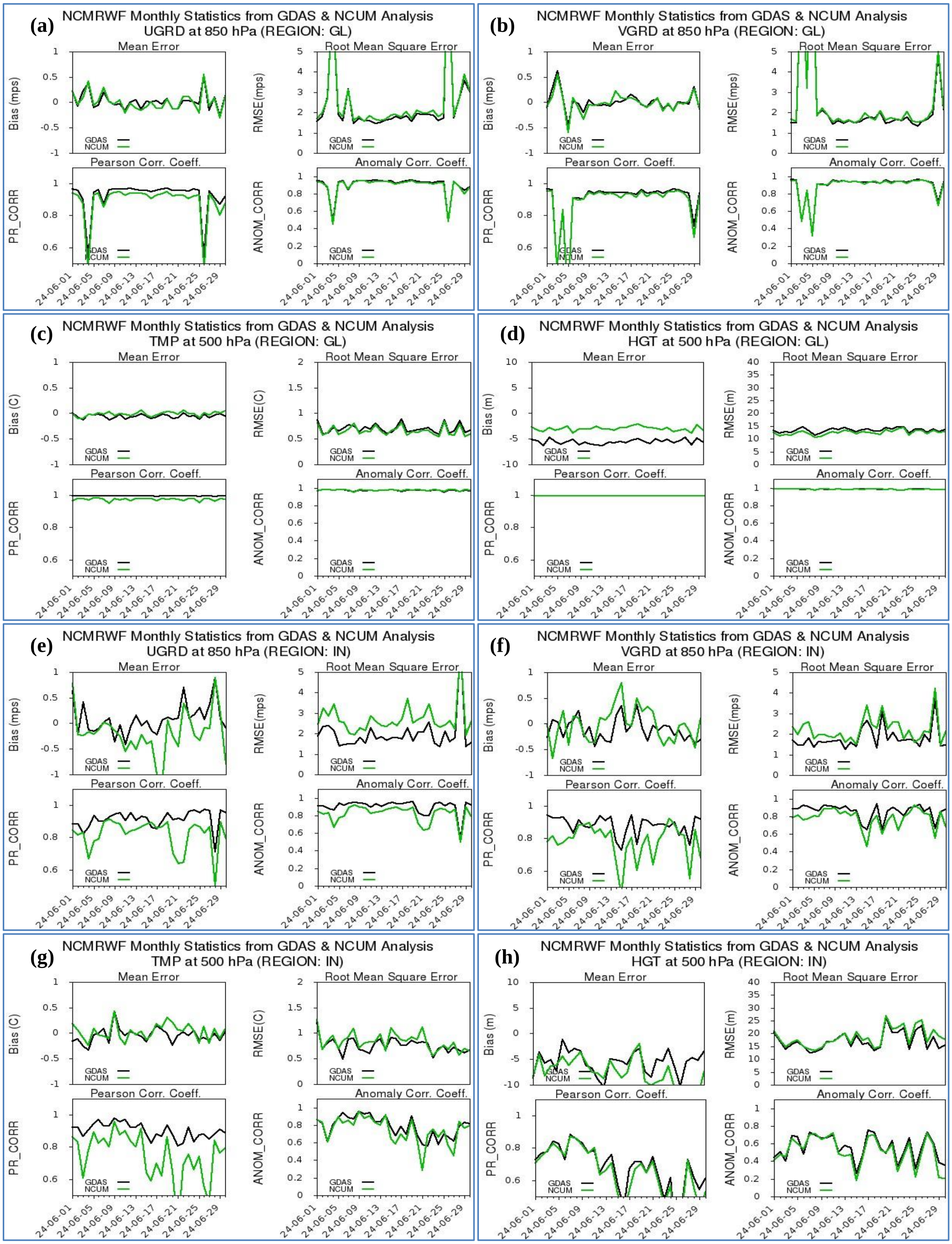
Figs. 3(a-d). Vertical distribution of PR_CORR from GDAS and NCUM analyses for (a) zonal wind (u; m/s); (b) meridional wind (v; m/s); (c) temperature (T; °C) and (d) geopotential height (m) valid at 00Z 20240601 over 12 specific domains of verification [x-axis: PR_CORR; y-axis: Pressure level]

Figs. 3(a-d) illustrates the vertical distribution of the Pearson correlation coefficient for key meteorological parameters. In the context of meteorological analysis, it assesses the strength and direction of relationships between predicted and observed variables, such as temperature, wind components, and geopotential height. The analysis reveals a high degree of consistency between the two models, with both exhibiting similar correlation patterns as a function of altitude. This suggests that both models demonstrate comparable skill in capturing the relationships between predicted and observed meteorological variables, with correlation coefficients varying similarly with height.

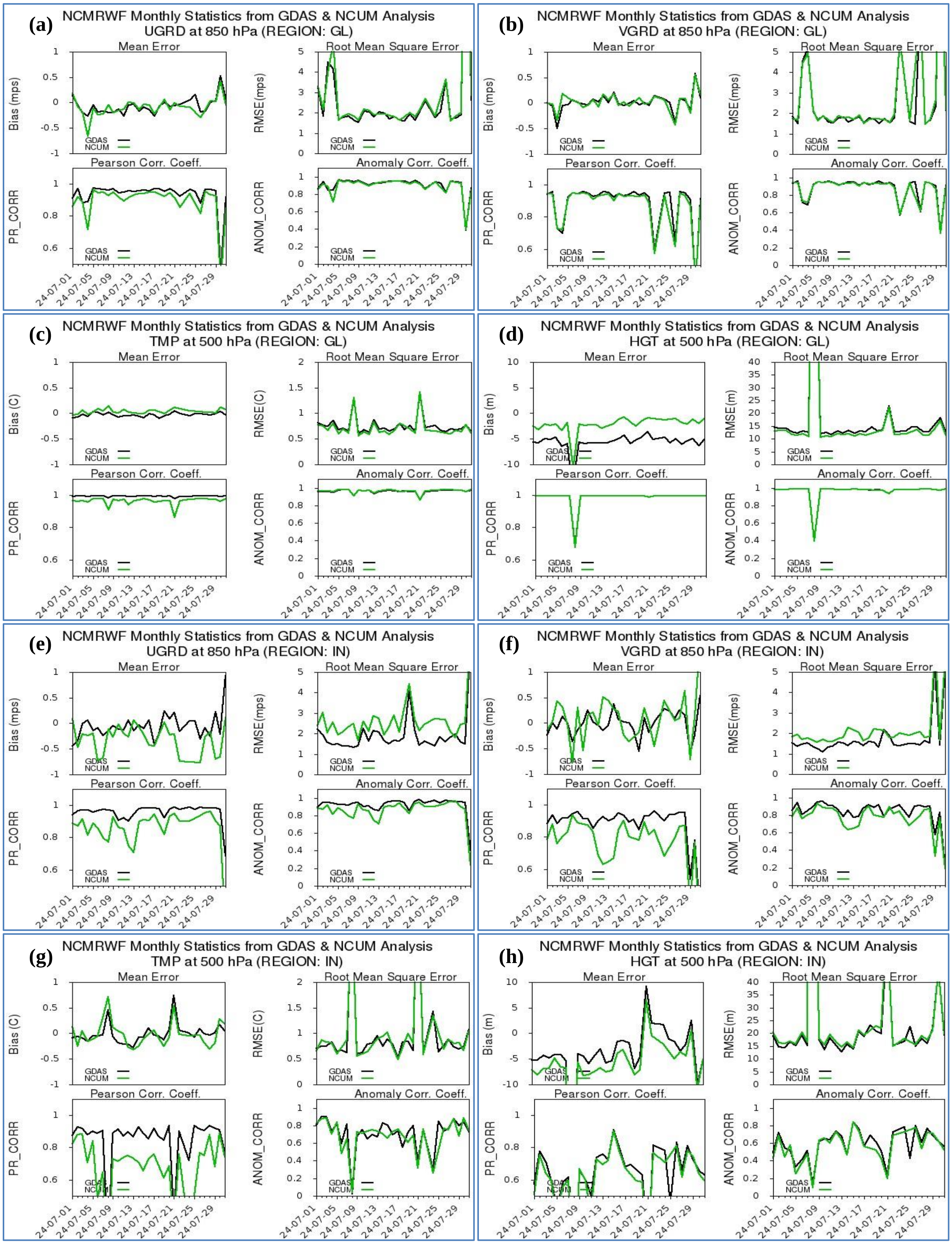
In addition to daily assessments, the performance of the analysis is evaluated on weekly and monthly timescales at 850, 500 and 250 hPa pressure levels. The weekly evaluation involves generating statistics for the last 7 days, updated daily, and providing a rolling assessment of model performance over a short-term period. This approach enables the estimation of analysis accuracy over a few days, offering insights into the model's ability to capture recent atmospheric trends. Furthermore, the weekly statistics facilitate the identification of recent patterns and anomalies, informing potential modifications to the analysis system to improve its performance. By examining the analysis performance on multiple timescales, researchers can gain a more comprehensive understanding of the model's strengths and limitations, ultimately contributing to the development of more accurate and reliable forecasting systems. Figs. 4 (a-d) presents a comprehensive assessment of weekly statistics, including mean error, RMSE, Pearson correlation, and anomaly correlation, for U- and V-components of wind at 850 hPa, temperature, and geopotential height at 500 hPa, respectively over the Global region. The analysis reveals that GDAS and NCUM exhibit similar performance characteristics, with minor deviations observed in specific metrics, indicating comparable skill in capturing global atmospheric patterns.



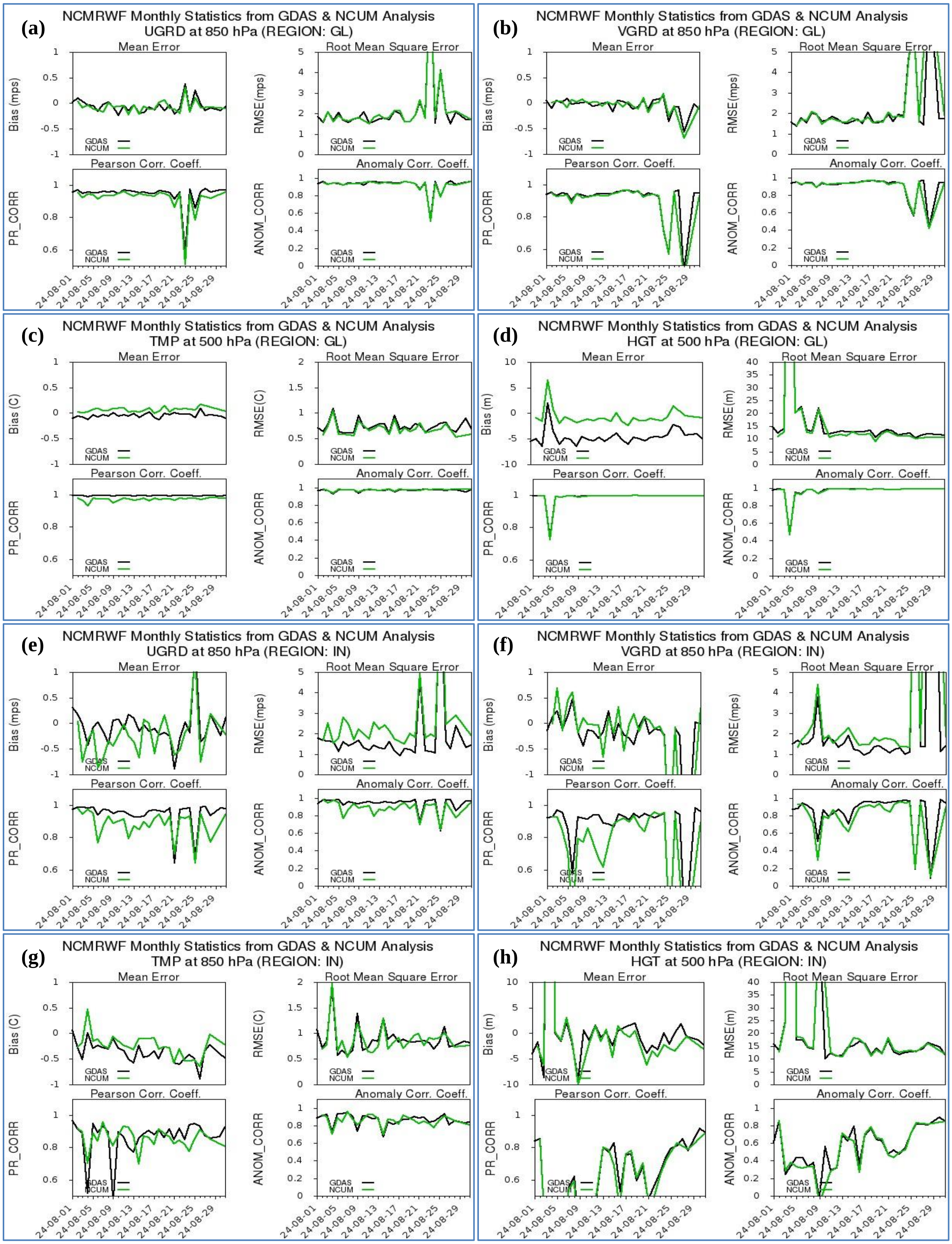
Figs. 4(a-h). Weekly statistics for bias, RMSE, Pearson correlation and anomaly correlation for (a) zonal wind at 850 hPa; (b) meridional wind at 850 hPa; (c) temperature at 500 hPa and (d) Geopotential at 500 hPa over global region during 09-15 August 2024. (e) – (h) is same as (a) – (d), but valid over Tropics region [x-axis: yy-mm-dd]



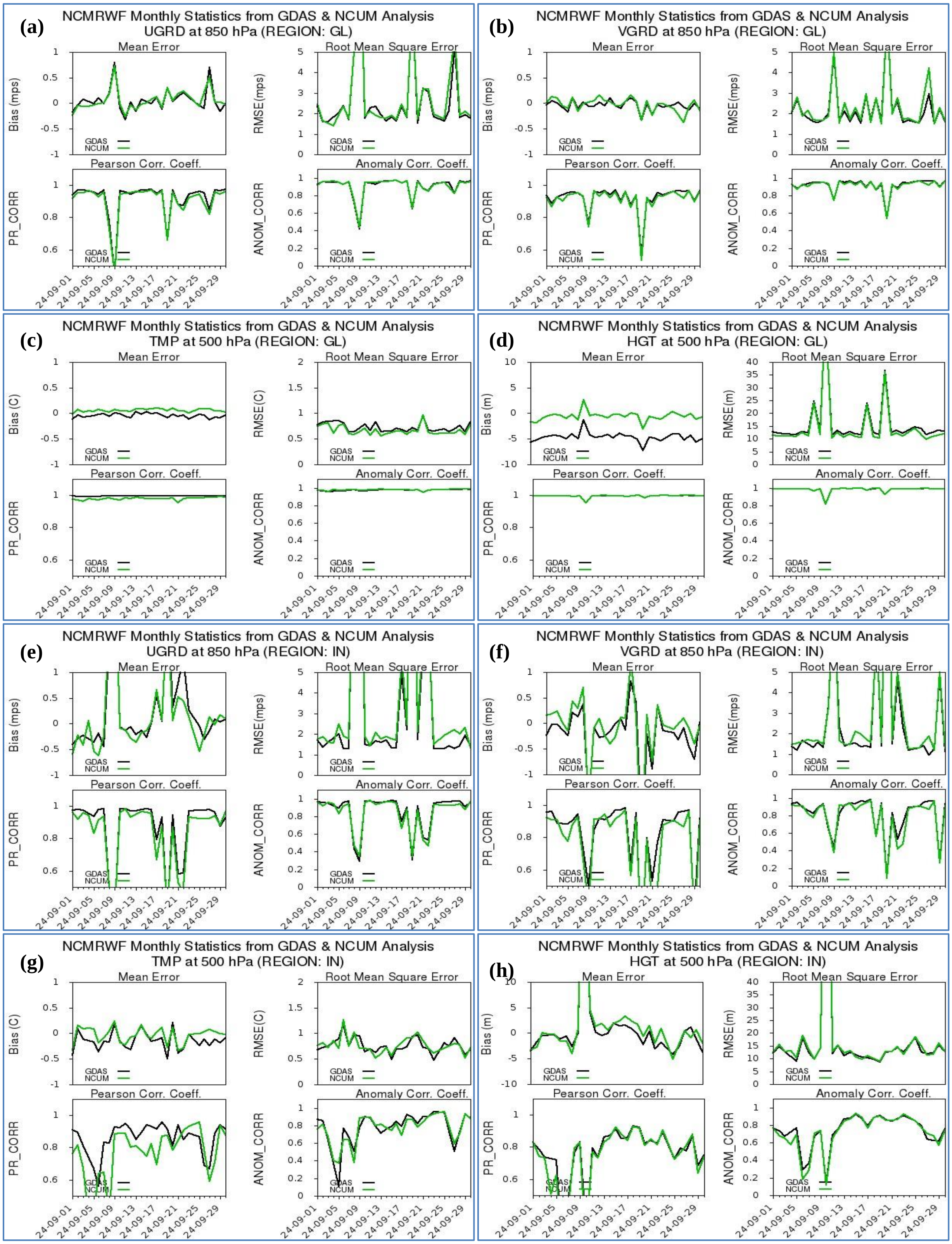
Figs. 5(a-h). Monthly statistics for mean bias, RMSE, Pearson correlation and anomaly correlation for (a) zonal wind at 850 hPa; (b) meridional wind at 850 hPa; (c) temperature at 500 hPa and (d) Geopotential at 500 hPa over global region for the month of June 2024. (e) – (h) is same as (a) – (d), but valid over India region [x-axis: yy-mm-dd]



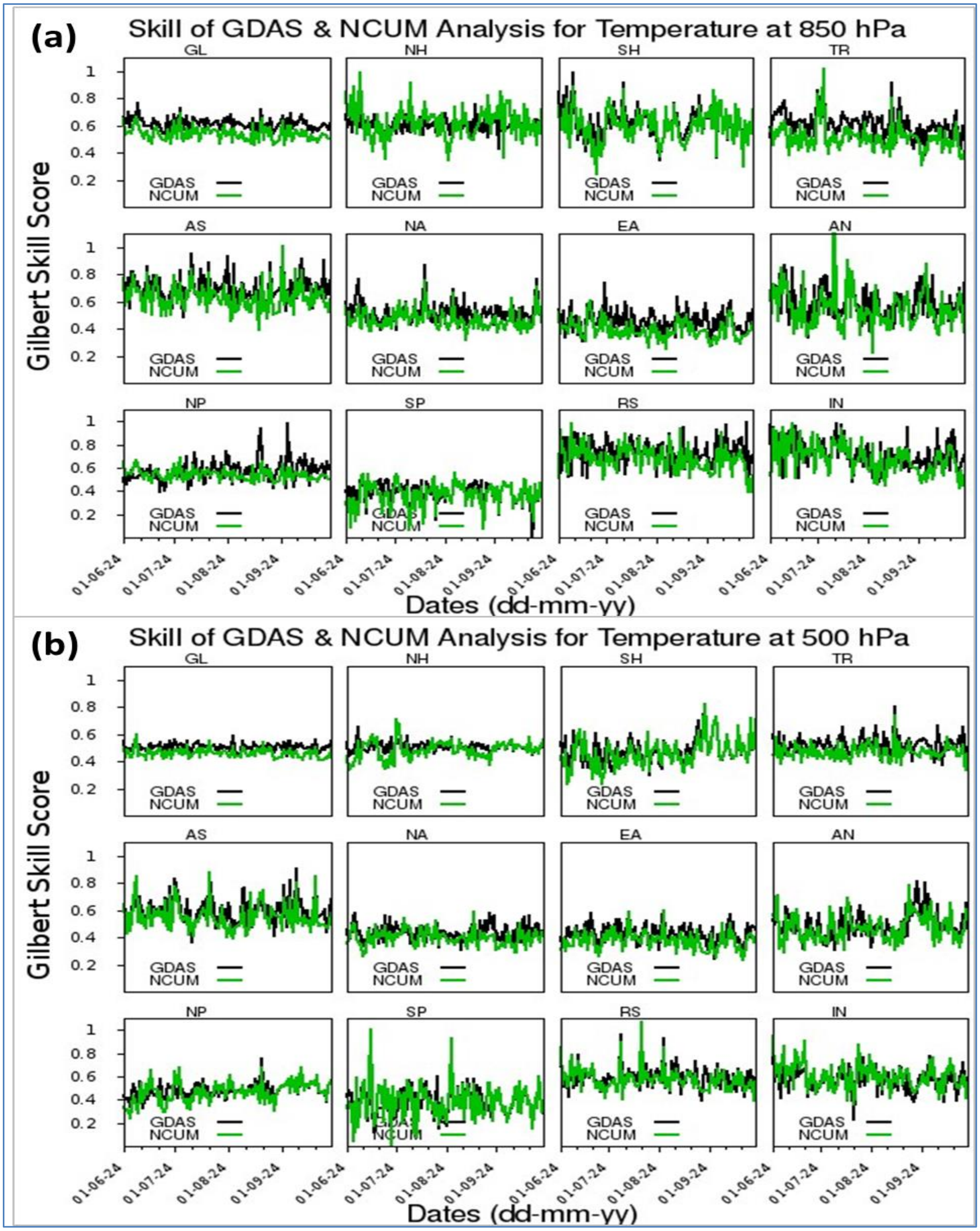
Figs. 6(a-h). Monthly statistics for mean bias, RMSE, Pearson correlation and anomaly correlation for (a) zonal wind at 850 hPa; (b) meridional wind at 850 hPa; (c) temperature at 500 hPa and (d) Geopotential at 500 hPa over global region for the month of July 2024. (e) – (h) is same as (a) – (d), but valid over India region [x-axis: yy-mm-dd]



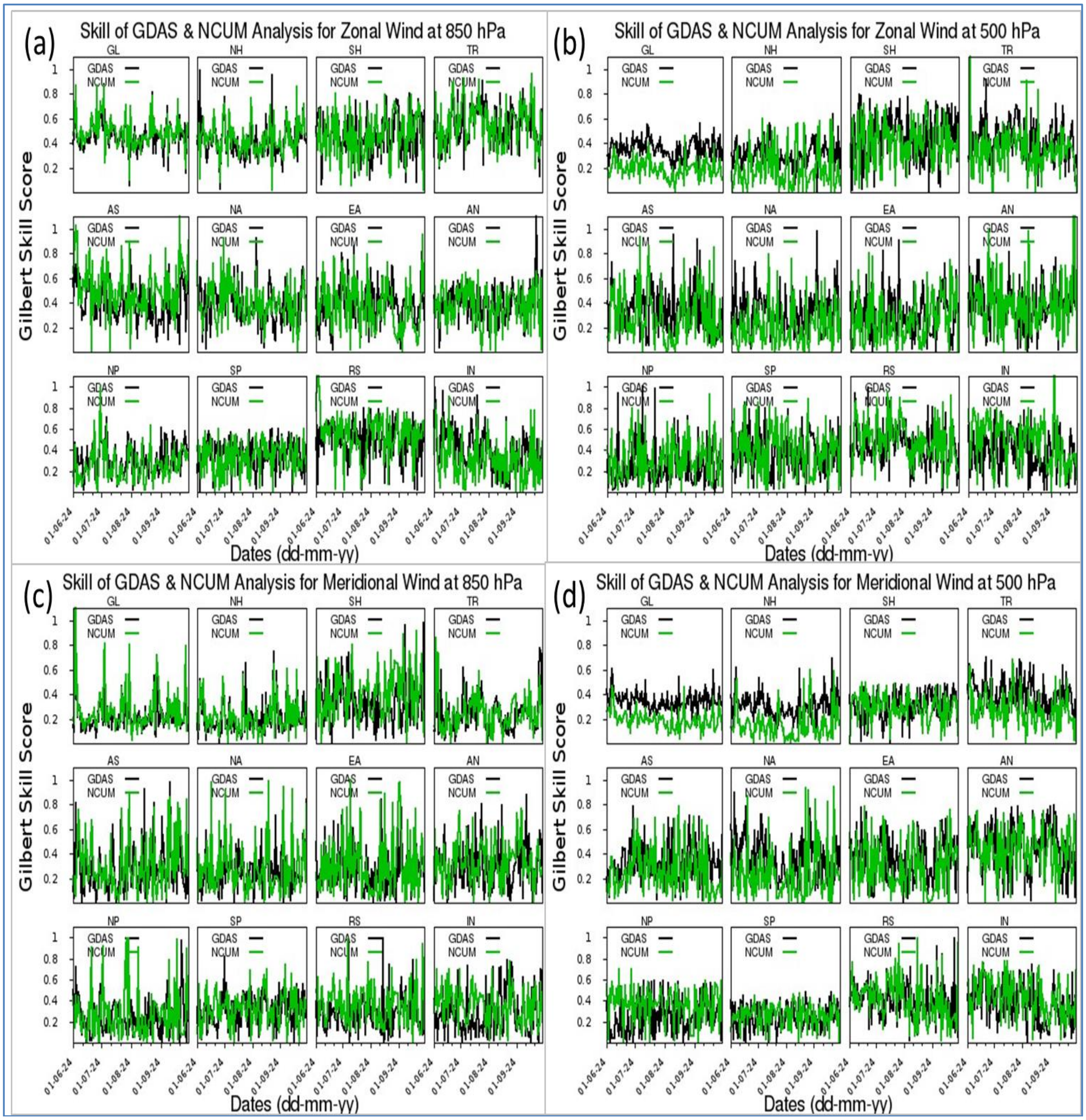
Figs. 7(a-h). Monthly statistics for mean bias, RMSE, Pearson correlation and anomaly correlation for (a) zonal wind at 850 hPa; (b) meridional wind at 850 hPa; (c) temperature at 500 hPa and (d) Geopotential at 500 hPa over global region for the month of August 2024. (e) – (h) is same as (a) – (d), but valid over India region [x-axis: yy-mm-dd]



Figs. 8(a-h). Monthly statistics for mean bias, RMSE, Pearson correlation and anomaly correlation for (a) zonal wind at 850 hPa; (b) meridional wind at 850 hPa; (c) temperature at 500 hPa and (d) Geopotential at 500 hPa over global region for the month of September 2024. (e) – (h) is same as (a) – (d), but valid over India region [x-axis: yy-mm-dd]



Figs. 9 (a-b). Skill of GDAS and NCUM analyses for the prediction of temperature at (a) 850 hPa and (b) 500 hPa over all the specified regions for JJAS 2024



Figs. 10(a-d). Skill of GDAS and NCUM analyses for the prediction of zonal wind at (a) 850 hpa and (b) 500 hPa over all the specified regions for JJAS 2024. (c-d) is same as (a-b), but for meridional wind

Figs. 4(e-h) represents the weekly statistics over the tropics. The error value dramatically increased over the tropics compared to the global assessment. However, GDAS analysis performs better than NCUM analysis in the tropic regions. This comparative assessment is extended to other regions, with weekly statistics generated for 850 and 500 hPa levels, facilitating a comprehensive evaluation of model performance across various spatial and vertical domains.

The $ANOM\_CORR$ measures the correlation between predicted and observed anomalies from their respective means. It assesses a model's ability to predict deviations from normal conditions, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). $ANOM\_CORR$ is widely used in meteorological forecasting to evaluate model performance in predicting anomalies, such as temperature or precipitation patterns. Thus, for weekly and monthly statistics, the $ANOM\_CORR$ is also evaluated along with the PR_CORR. A comprehensive monthly assessment is conducted at the end of each month, with Fig. 5 illustrating the statistical metrics for June 2024. The analysis includes mean error, RMSE, Pearson correlation, and anomaly correlation for U- and V-components of wind at 850 hPa, as well as temperature and geopotential height at 500 hPa, over the Global region. A notable increase in error is observed for wind components over the global region during early and late June, indicating potential challenges in capturing specific atmospheric dynamics during these periods. In contrast, temperature and geopotential height predictions demonstrate higher accuracy, characterized by lower RMSE values and stronger correlations, suggesting that the models effectively capture the large-scale thermal and geopotential patterns. The Indian region Figs. 5(e-h) shows comparatively higher bias throughout the month. Fig. 6 illustrates the statistical metrics for July 2024, revealing that both analyses exhibit comparable performance characteristics. Notably, higher errors are observed for wind and geopotential height predictions, whereas temperature forecasts demonstrate relatively lower errors, indicating varying levels of model skill across different atmospheric variables. In August (Fig. 7), a notable distinction in model performance is observed between global and regional scales. Globally, the analyses exhibit relatively lower errors across parameters. However, the tropics and Indian region display higher errors, particularly pronounced for wind and geopotential height predictions, suggesting region-specific challenges in model accuracy. The statistical evaluation for September (Fig. 8) reveals that both global and regional analyses exhibit significant errors in wind field predictions. In contrast, temperature forecasts demonstrate higher accuracy, indicating that the models effectively capture thermal patterns, while struggling with wind field dynamics during this period.

Both analyses present more or less a similar way for all the meteorological parameters except geopotential height, which exhibits a very sharp increase in bias, RMSE, and with less correlation coefficient a few days in daily, weekly, and monthly statistics. It may be noted that temperature is better predicted parameter than other fields. The proficiency of GDAS and NCUM analysis systems is quantitatively evaluated for wind and temperature at various verification levels utilizing the Gilbert Skill Score (GSS) methodology for the JJAS period. This approach provides a robust metric for assessing the categorical forecast skills of these models. The skill scores for temperature at 850 hPa and 500 hPa are depicted in Figs. 9 (a&b), respectively. A comparative analysis reveals that both GDAS and NCUM exhibit similar skill score patterns. At the 850 hPa level, the GDAS analysis system exhibits a marginally higher skill score compared to NCUM for temperature predictions over the GL region. Notably, the SP region demonstrates relatively lower skill scores compared to other regions, which can be attributed to the limited number of observational data points available over this region. The skill scores for the zonal wind component at 850 hPa and 500 hPa are presented in Figs. 10(a&b), respectively. A comparative analysis reveals that both GDAS and NCUM exhibit similar performance patterns. Notably, at the 500 hPa level, GDAS demonstrates a marginally higher skill score compared to NCUM over the GL and NH regions, whereas NCUM outperforms GDAS over the IN region. The skill scores for the meridional wind component at 850 hPa and 500 hPa are illustrated in Figs. 10(c&d), respectively. Both GDAS and NCUM display similar skill score patterns. However, at the 500 hPa level, GDAS exhibits enhanced skill over the GL region, while performance over other regions remains comparable. This evaluation provides insight into the models' performance and skill in capturing large-scale atmospheric patterns during the specified period.

4. Conclusions

A thorough assessment of the GDAS and NCUM analysis performance is conducted over 12 distinct regions, including additional domains such as global, RSMC, and India, beyond those specified by ECMWF and WMO Lead Centre for Deterministic NWP Verification. The CBS is computed on daily, weekly, and monthly timescales using the MET, which enables validation of model analysis against gridded or point observations. The MET's flexibility allows for verification of key meteorological parameters, including temperature, wind, and geopotential height, providing a comprehensive insight into model performance.

The results indicate that both analyses exhibit similar performance characteristics on a daily basis, suggesting that both models capture short-term atmospheric dynamics with comparable accuracy. Similarly, on weekly and monthly timescales, the global region shows analogous statistics for GDAS and NCUM analyses, implying that both models demonstrate equivalent skill in predicting large-scale patterns.

However, a notable distinction emerges when focusing on the Indian region, where the weekly and monthly verification reveals a superior performance of the GDAS analysis compared to NCUM. This region-specific difference in performance highlights the strengths of the GDAS model in capturing local atmospheric dynamics and nuances, potentially due to differences in model formulation, data assimilation, or initialization. The superior performance of GDAS over the Indian region has implications for regional weather forecasting and decision-making.

Data Availability

The RS/RW observations used in the study are taken from NCMRWF internal repositories. The GDAS and NCUM model analysis are used from the NCMRWF internal repositories.

Funding

This work receives no specific grant from any funding agency in the public, commercial, or not-for-profit organizations.

Acknowledgments

Appreciation is extended to the India Meteorological Department (IMD) for generously providing RS/RW observations for the study period. Furthermore, the authors acknowledge the DTC, NOAA for providing the MET system, a crucial resource that enabled the verification analysis presented in this study.

Authors' Contributions

Sujata Pattanayak: Conceptualization, Conducting experiments, Investigation, Visualization, Writing-Original draft preparation.
Ashish Routray: Supervised, Writing-reviewing & editing.
Suryakanti Dutta: Reviewing and Editing.
V. S. Prasad: Provide infrastructure facilities.

Disclaimer: The contents and views presented in this research article/paper are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Bullock, R. G., Brown, B. G., and Fowler, T. L., 2016, "Method for Object-Based Diagnostic Evaluation", NCAR Technical Note NCAR/TN-532+STR, 1-84.

Clark, W. L., Wick, G. A., Tollerud, E. I., Jensen, T. L., Gotway, J. H., Sukovich, E., and Yuan, H., 2012, "Assessing and Comparing Real-Time Northeast Pacific Atmospheric River IWV and IVT Intensities Provided by Model (GFS) and Satellite (SSM/I, SSMIS) Using DTC's MET/MODE Object Attributes", 21st Conference on Probability and Statistics in the Atmospheric Sciences, 22-27 January, New Orleans LA, American Meteorological Society (Boston).

Clark, W. L., Yuan, H., Jensen, T. L., Wick, G., Tollerud, E. I., Bullock, R. G., and Sukovich, E., 2011, "Evaluation of GFS water vapor forecast errors during the 2009–2010 West Coast cool season using the MET/MODE object analyses package", 25th Conference on Hydrology. January 22–27, Seattle, WA. American Meteorological Society (Boston).

Connolley, W. M. and Harangozo, S. A., 2001, "A Comparison of Five Numerical Weather Prediction Analysis Climatologies in Southern High Latitudes", Journal of Climate, 30-44. https://doi.org/10.1175/1520-0442(2001)014<0030:ACO FNW>2.0.CO;2

Endris, H. S., Hirons, L., Tessema, Z., Gudoshava, M., Woolnough, S., and Artan, G. A., 2021, "Evaluation of the Skill of Monthly Precipitation Forecasts from Global Prediction Systems over the Greater Horn of Africa", *Weather and Forecasting, 36*, 1275-1298.

George, J. P., Indira Rani, S., Jayakumar, A., Mohandas, S., Mallick, S., Rakhi, R., Sreevathsa, M. N. R. and Rajagopal, E. N., 2016, "NCUM Data Assimilation System", NCMRWF Technical Report, NMRF/TR/01/2016, 1-21.

Jolliffe, I. T., and Stephenson, D. B., 2003, "Forecast Verification: A Practitioner's Guide in Atmospheric Science", John Wiley and Sons, 1-240.

Jolliffe, I. T., and Stephenson, D. B., 2012, "Forecast Verification: A Practitioner's Guide in Atmospheric Science", John Wiley and Sons (2nd edition), pp. 1-292.

JWGFVR, 2009, "Recommendations for the Verification and Inter-comparison of QPFs and PQPFs from Operational NWP Models", WMO TD No. 1485, WWRP 2009-1.

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu W. –S., and Lord, S., 2009, "Introduction of the GSI into the NCEP Global Data Assimilation System", Weather Forecasting **24**, 1691–1705. http://dx.doi.org/10.1175/2009WAF2222201.1.

Kumar, S., Bushair, M. T., Jangid, B. P., Lodh, A., Sharma, P., George, G. and Indira Rani, S. Jayakumar, A., Mohandas, S., Kumar, S., Sharma, K., Karunasagar S., and Rajagopala, E. N., 2020, "NCUM Global NWP System: Version 6 (NCUM-G:V6)", NMRF/TR/06/2020.

Kumar, S., Jayakumar, A., Bushair, M. T., Jangid, B. P., George, G., Lodh, A., Indira Rani, S., Mohandas, S., George, J. P., and Rajagopal, E. N., 2018, "Implementation of New High Resolution NCUM Analysis-Forecast System in Mihir HPCS", NMRF/TR/01/2018, 1-17.

Prasad V. S., Mohandas, S., Das Gupta, M., Rajagopal, E.N. and Dutta, S. K., 2011, "Implementation of Upgraded Global Forecasting Systems (T382L64 and T574L64) at NCMRWF", NCMRWF Technical Report, NCMR/TR/5/2011. https://www.ncmrwf.gov.in/gfs_report_final.pdf.

Prasad, V. S., Gupta, A., Rajagopal, E. N. and Basu, S., 2013, "Impact of OScat surface wind data on T574L64 assimilation and forecasting system -A study involving tropical cyclone Thane", Current Science, 104, 5, 627-631.

Prasad, V. S., Dutta, S. K., Pattanayak, S., Johny, C.J., George, J. P., Kumar, S and Indira Rani S., 2021, "Assimilation of satellite and other data for forecasting tropical cyclones over NIO", Mausam, 72, 1, 107-118.

Prasad, V. S., Johny, C. J., Mali, P., Singh, S. K., and Rajagopal, E. N., 2017, "Global retrospective analysis using NGFS for the period 2000-2011", Current Science, 112, 2, 370-377.

Prasad, V. S., Mohandas, S., Dutta, S. K., Das Gupta, M., Iyengar, G. R., Rajagopal, E. N., and Basu, S., 2014, "Improvements in medium range weather forecasting system of India", Journal of Earth System Sciences, **123**, 2, 247–258.

Rajagopal, E. N., Iyengar, G.R., George, J. P., Das Gupta, M., Mohandas, Saji., Siddharth, R., Gupta, A., Chourasia, M., Prasad, V. S., Aditi, Sharma, K., and Ashish, A., 2012, "Implementation of Unified Model based Analysis-Forecast System at NCMRWF", *NMRF/TR/2/2012*.

Rajagopal, E.N., Das Gupta, M., Mohandas, S., Prasad, V.S., George, J. P., Iyengar G. R., and Preveen Kumar, D., 2007, "Implementation of T254L64 Global Forecast System at NCMRWF", NCMRWF Technical Report,NMRF/TR/1/2007, 1-42.

Rawlins, F., Ballard, S. P., Bovis, K. J., AM, Clayton, Li, D., Inverarity, G. W., 2007, "The Met Office global four-dimensional variational data assimilation scheme", Q. J. R. Meteorological Society, **133**, 347–362.

Routray, A., 2012, "Implementation of the UM model based analysis–forecast system at NCMRWF", NMRF/TR/2012, 1-45.

Routray, A., Singh, V., Gupta, A., Dutta, D., and George, J. P., 2019, "Impact of Vortex Initialization in Prediction of Tropical Cyclones over Bay of Bengal with NCUM Model", Marine Geodesy, **42**, 2, 201–226. (https://doi.org/10.1080/01490419.2018.1552219).

Routray, A., Singh, V., Singh, H., Dutta, D., George, J. P., and Rakhi, R., 2017, "Evaluation of different versions of NCUM global model forsimulation of track and intensity of tropical cyclones over Bay of Bengal", Dynamics of Atmospheres and Oceans, **78**, 71–88.

Singh, V., Routray, A., Dutta, D., George, J. P., 2018, "Impact of Vortex Initialization with Satellite Data Assimilation for Predicting Track and Intensity of Tropical Cyclone Chapala", *Remote Sensing in Earth Systems Sciences, 1*, 39–52. https://doi.org/10.1007/s41976-018-0003-3.

Wilks, D., 2011, "Statistical methods in Atmospheric Sciences", 3rd Edition, Academic Press, ISBN 9780123850225.

WMO, 2017, "WMO Manual on the Global data-processing and Forecasting system: Annex IV to the WMO technical regulations", WMO-no.485.