NMRF/TR/08/2024

TECHNICAL REPORT

Real-time Monitoring of GTS Data Reception at NCMRWF

Upal Saha, S. Indira Rani, D. Srinivas and V. S. Prasad

October 2024

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA-201309, INDIA

Real-time Monitoring of GTS Data Reception at NCMRWF

Upal Saha, S. Indira Rani, D. Srinivas and V. S. Prasad

October 2024

National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences
A-50, Sector 62, NOIDA-201309, INDIA

Ministry of Earth Sciences
National Centre for Medium Range Weather Forecasting
Document Control Data Sheet

1	Name of the Institute	National Centre for Medium Range Weather Forecasting
2	Document Number	NMRF/TR/08/2024
3	Date of publication	October 2024
4	Title of the document	Real-time Monitoring of GTS Data Reception at NCMRWF
5	Type of Document	Technical Report
6	No. of pages & Figures	12 & 4
7	Number of References	2
8	Author (s)	Upal Saha, S. Indira Rani, D. Srinivas and V. S. Prasad
9	Originating Unit	NCMRWF
10	Abstract	This report explores the real-time monitoring of ASC and BIN files received at NCMRWF through the Global Telecommunication System (GTS), highlighting the benefits of automating the monitoring process. The system tracks file sizes at regular intervals and calculates departures from 30-day averages to detect anomalies in the current data file sizes. By leveraging automated plotting and visualizations, the system provides a clear representation of file size trends for both ASC and BIN data types. Color-coded departure alerts indicate deviations from the normal, allowing for rapid identification of potential issues in data reception via the GTS. The strengths of this real-time monitoring approach include enhanced data integrity and early detection of irregularities, which are critical for maintaining the smooth operation of NCMRWF global and regional data assimilation systems to prepare initial conditions for generating forecasts. This proactive system reduces the risk of data corruption or loss and ensures the reliability of data consistency. However, the system does require substantial computational resources and must carefully manage thresholds to avoid false alarms. Despite these challenges, the real-time monitoring system significantly improves the timeliness, accuracy, and integrity of meteorological data processing, making it a valuable tool in real-time data management.
11	Security classification	Non-Secure
12	Distribution	Unrestricted Distribution
13	Key Words	GTS; RTH; IMD; WMO; Monitoring.

Contents

Abstract

This report explores the real-time monitoring of ASC and BIN files received at NCMRWF through the Global Telecommunication System (GTS), highlighting the benefits of automating the monitoring process. The system tracks file sizes at regular intervals and calculates departures from 30-day averages to detect anomalies in the current data file sizes. By leveraging automated plotting and visualizations, the system provides a clear representation of file size trends for both ASC and BIN data types. Color-coded departure alerts indicate deviations from the normal, allowing for rapid identification of potential issues in data reception via the GTS. The strengths of this real-time monitoring approach include enhanced data integrity and early detection of irregularities, which are critical for maintaining the smooth operation of NCMRWF global and regional data assimilation systems to prepare initial conditions for generating forecasts. This proactive system reduces the risk of data corruption or loss and ensures the reliability of data consistency. However, the system does require substantial computational resources and must carefully manage thresholds to avoid false alarms. Despite these challenges, the real-time monitoring system significantly improves the timeliness, accuracy, and integrity of meteorological data processing, making it a valuable tool in real-time data management.

Keywords: GTS; RTH; IMD; WMO; Monitoring.

सारांश

यह रिपोर्ट ग्लोबल टेलीकम्यूनिकेशन सिस्टम (जी.टी.एस) के माध्यम से रा. म. अ. मौ. पू. के. में प्राप्त ASC और BIN फाइलों की वास्तविक समय की निगरानी की पड़ताल करती है, निगरानी प्रक्रिया को स्वचालित करने के लाभों पर प्रकाश डालती है। सिस्टम नियमित अंतराल पर फ़ाइल आकार को ट्रैक करता है और वर्तमान डेटा फ़ाइल आकार में विसंगतियों का पता लगाने के लिए 30-दिन के औसत से विचलन की गणना करता है। स्वचालित प्लॉटिंग और विज़ुअलाइज़ेशन का लाभ उठाकर, सिस्टम ए.एस.सी और बिन दोनों डेटा प्रकारों के लिए फ़ाइल आकार के रुझान का स्पष्ट प्रतिनिधित्व प्रदान करता है। रंग-कोडित प्रस्थान अलर्ट सामान्य से विचलन का संकेत देते हैं, जिससे जी.टी.एस के माध्यम से डेटा रिसेप्शन में संभावित मुद्दों की तेजी से पहचान की जा सकती है। इस वास्तविक समय निगरानी दृष्टिकोण की ताकत में बढ़ी हुई डेटा अखंडता और अनियमितताओं का शीघ्र पता लगाना शामिल है, जो पूर्वानुमान उत्पन्न करने के लिए प्रारंभिक स्थितियां तैयार करने के लिए रा. म. अ. मौ. पू. के. वैश्विक और क्षेत्रीय डेटा एसिमिलेशन सिस्टम के सुचारू संचालन को बनाए रखने के लिए महत्वपूर्ण हैं। यह सक्रिय प्रणाली डेटा भ्रष्टाचार या हानि के जोखिम को कम करती है और डेटा स्थिरता की विश्वसनीयता सुनिश्चित करती है। हालाँकि, सिस्टम को पर्याप्त कम्प्यूटेशनल संसाधनों की आवश्यकता होती है और मिथ्य अलार्म से बचने के लिए थ्रेसहोल्ड को सावधानीपूर्वक प्रबंधित करना चाहिए। इन चुनौतियों के बावजूद, वास्तविक समय की निगरानी प्रणाली मौसम संबंधी डेटा प्रसंस्करण की समयबद्धता, सटीकता और अखंडता में काफी सुधार करती है, जिससे यह वास्तविक समय डेटा प्रबंधन में एक मूल्यवान उपकरण बन जाता है।

कुंजी शब्द: GTS; RTH; IMD; WMO; Monitoring

1. Introduction

The National Centre for Medium Range Weather Forecasting (NCMRWF) receives atmospheric and ocean observational data through the Global Telecommunications System (GTS) network of the World Meteorological Organisation (WMO). NCMRWF has established data links with global satellite operators such as NOAA, EUMETSAT, CMA, KMA, and ISRO to receive satellite datasets. NCMRWF generates initial conditions for atmospheric models based on NCEP's Global Forecasting System (GFS), known as NCMRWF-GFS (NGFS) or UK Met Office's Unified Model (UM) system, known as NCMRWF-UM (NCUM), which are used by various other centres of the Ministry of Earth Sciences and other public and private stakeholders in India and nearby countries. The conventional observations are received through the Regional Telecommunication Hub (RTH) at the India Meteorological Department (IMD) in New Delhi.

Initially, NCMRWF utilized ECMWF decoders to decode the GTS data and stored the decoded data in a specific RDBMS format. These decoders were replaced with those from NCEP (USA) in 2007, establishing a new data processing system wherein processed data is stored in a specialized directory structure termed "BUFR-TANK" (Rajagopal et al., 2007). In both the systems, the received GTS data undergoes initial binning into half-hourly subdirectories, retaining original filenames transmitted by IMD RTH-New Delhi, before being passed through a set of data decoders tailored to handle specific datasets. Each GTS file (either ASCII or Binary) contain reports from different observing systems like Surface, upper air, aircraft, etc. The decoding of these files incurs a delay due to indiscriminate reading of all files by each decoder, causing unnecessary processing overhead. NCMRWF implemented early cycle runs with a two-hour cut-off time to ensure timely availability of Numerical Weather Prediction (NWP) products to operational forecasters and stakeholders. For faster decoding of the GTS observations, a new real-time data separation is introduced in 2018 (Prasad, 2020). The GTS messages are separated according to the

GTS headers, and the observation reports are stored in two letter directories as per the observing system. These messages are passed to the respective decoders avoiding the indiscriminate reading of all messages by every decoder, reducing the decoding time.

2. Brief description on Global Telecommunication System (GTS)

The GTS serves as the coordinated framework of telecommunication facilities and arrangements for enabling the rapid exchange of meteorological data and processed information worldwide (Figure 1). Within the World Meteorological Organization (WMO) Information System (WIS), the GTS plays a pivotal role in ensuring the timely, reliable, and cost-effective transmission of data and processed products to meet the requirements of the World Weather Watch (WWW) system, providing access to all WMO Members (National Meteorological Services) and International organisations such as ECMWF and EUMETSAT.

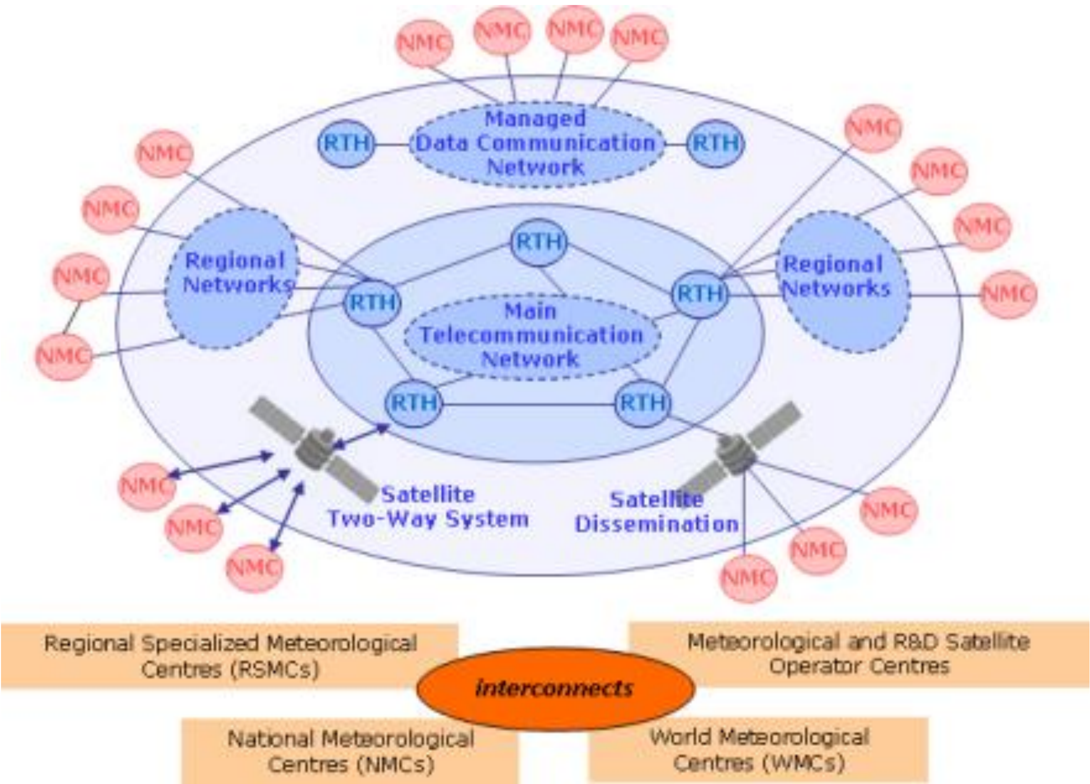

Figure 1. *Structure of the GTS*. [Image Courtesy: https://community.wmo.int/en/activity-areas/global-telecommunication-system-gts]

Organized based on the principles of the WIS, the GTS operates on a hierarchical three-level structure. This structure comprises the Main Telecommunication Network (MTN), which interconnects three World Meteorological Centers (WMCs) [Melbourne, Moscow and Washington] as well as eighteen designated Regional Telecommunication Hubs (RTHs) [Alger, Beijing, Brasilia, Buenos Aires, Cairo, Dakar, Exeter, Jeddah, Melbourne, Moscow, Nairobi, New Delhi, Offenbach, Prague, Sofia, Tokyo, Toulouse, Washington]; and regional and national telecommunication networks.

NCMRWF accesses this GTS data from RTH-New Delhi via the National Knowledge Network (NKN) connectivity, with normal internet FTP access available as a backup option. The GTS data received by NCMRWF consists of two types of continuous messages: alphanumeric numerical channels and binary channels. Both message types commence with an "abbreviated header" and adhere to a fixed structure. In addition to transmitting GTS messages, RTH-New Delhi also forwards specific data to NCMRWF using the same communication channel. This data encompasses RARS data sourced from Australia and Japan, observations from the Indian Doppler Weather Radar (DWR), and radiances from the Himawari satellite operated by JMA, Japan. The details of the GTS structure and header can be obtained from Prasad (2020).

3. Archival of GTS data files

The GTS data from RTH-New Delhi are synchronized with the Mihir HPCs, enabling the automated transfer of GTS files into separate directories—ASC and BIN. These files are stored under "/home/gfsprod/data/gts" on Mihir HPCs, organized by the date of origin in YYYYMMDD format. The GTS data files are pushed continuously and they are stored in 48 directories each day comprising of data received in every 30 minutes for both ASC and BIN files (e.g., ASC00 to ASC47 and BIN00 to BIN47). The data archival initiates at 00:15 Hours every day and is being stored half-hourly in the respective directories, i.e. for ASC and BIN. Table 1 shows the half-hourly archival directory structure of the GTS data received at NCMRWF. Within each directory, the ASCII files (starting with "DEMS" and ending in “.a”) and Binary files (starting with "DEMS" and ending in “.b”) are stored in real-time. The file sizes vary each half-hour depending on the number of files received at NCMRWF via GTS. If data is not transmitted during a specific hour, it is pushed in the subsequent hour accordingly. An example of the storage structure in NCMRWF Mihir HPCs is as follows:

- **ASC Files**: /home/gfsprod/data/gts/YYYYMMDD/ASC/ASC{00..47}/DEMS*.a
- **BIN Files**: /home/gfsprod/data/gts/YYYYMMDD/BIN/BIN{00..47}/DEMS*.b

In addition to GTS messages, RTH-New Delhi’s data stream includes Asia-Pacific RARS data, radiance data from HIMAWARI-8, and Doppler Weather Radar (DWR) data from IMD’s network. These are organized in DBNET, MTSAT, and station-specific subdirectories, respectively.

Table 1. *Half-hourly Archival Directory Structure of GTS Data received at NCMRWF on a daily basis*:

Start Time (Hours) in UTC	Time Window (Hours)	Data Types	
		ASC (Directories)	BIN (Directories)
00:15	00:00-00:30	ASC00	BIN00
00:45	00:30-01:00	ASC01	BIN01
01:15	01:00-01:30	ASC02	BIN02
.. .. ..	.. .. ..	.. .. ..	.. .. ..
22:45	22:30-23:00	ASC45	BIN45
23:15	23:00-23:30	ASC46	BIN46
23:45	23:30-00:00	ASC47	BIN47

4. Real-time monitoring of data files obtained through GTS

Real-time monitoring of ASC (ASCII) and BIN (binary) files receiving through GTS is crucial for assimilation purpose. This part of the report focuses on the real-time monitoring of ASC and BIN files using shell scripting, detailing the processes, benefits, and limitations. The steps involved include automated size monitoring, calculation of departures, and interpreting the outcomes for decision-making.

A number of steps involved in real-time monitoring of these files, which is described as below:

A. Master Script for ASC and BIN File Monitoring:

Script Name:

a) master_ASC_size_half_hourly.sh,

b) master_BIN_size_half_hourly.sh

These master shell scripts automate the size monitoring of ASC and BIN files every half hour. These scripts are executed with a predefined schedule based on the time parameter UTC_HHMM, which identifies specific time windows such as 00:55, 01:25, etc. The primary objective of these scripts is to execute size departure calculations by passing the date and ASC or BIN type to the file-size calculation script at designated times. The script takes care of the following:

a) **Time-based Execution:** Since the ASC and BIN files are pushed through GTS every half hourly, the master script executes every 30 minutes to ensure frequent monitoring.

b) **File Directory Setup:** The master script defines directories for storing and retrieving GTS ASC and BIN files and monitors their sizes.

c) **Daily Adjustment:** The master script adjusts for day transitions, making it dynamic and able to handle new daily files automatically.

B. Calculation of ASC and BIN File Size Departure:

Script Name:

a) calculate_asc_size_departures.sh,

b) calculate_bin_size_departures.sh

The scripts compute file size departures by comparing the current file size against the last 30-day mean. This provides insights into whether file sizes are normal or exhibit anomalies. The script takes care of the following:

a) **Data Retrieval**: The script retrieves the ASC and BIN data from respective directories corresponding to the current and past 30 days.

b) **Size Calculation**: It then calculates the file sizes in megabytes (MB) and computes the percentage of departure using the following formula:

ASC/BIN Departure = ((Today's ASC/BIN size - 30-day Mean ASC/BIN size) / 30-day Mean ASC/BIN size) * 100

c) **Average Over 30 Days**: Also, it ensures any outliers are smoothened out, providing a more reliable baseline for departure calculations.

The following are the advantages of the real-time monitoring of the ASC or BIN file sizes:

- **Improved Data Integrity:** Constant monitoring helps ensure that data inconsistencies, missing files, or corrupt files are flagged in real time, allowing for rapid troubleshooting.
- **Automated timely alerts**: Real-time alerts and visualizations ensure that anomalies are quickly identified and addressed, minimizing the risk of data loss or system malfunctions.

The main disadvantage of the real-time monitoring requires constant system engagement, disk space, and processing power as real-time monitoring is resource-intensive.

C. Real-time plotting of ASC and BIN File Sizes:

Script Name:

a) plot_gts_asc_size_monit.py,

b) plot_gts_bin_size_monit.py

These scripts are used to visualize the size of the ASC and BIN files and their departures in real-time. The bars indicate the departure of today's file size from the 30-day mean, color-coded by severity:

- **Blue**: Departure ≥ 30% (high/excess)
- **Red**: Departure ≤ -30% (low/deficient)
- **Green**: Departure between -30% and 30% (normal)

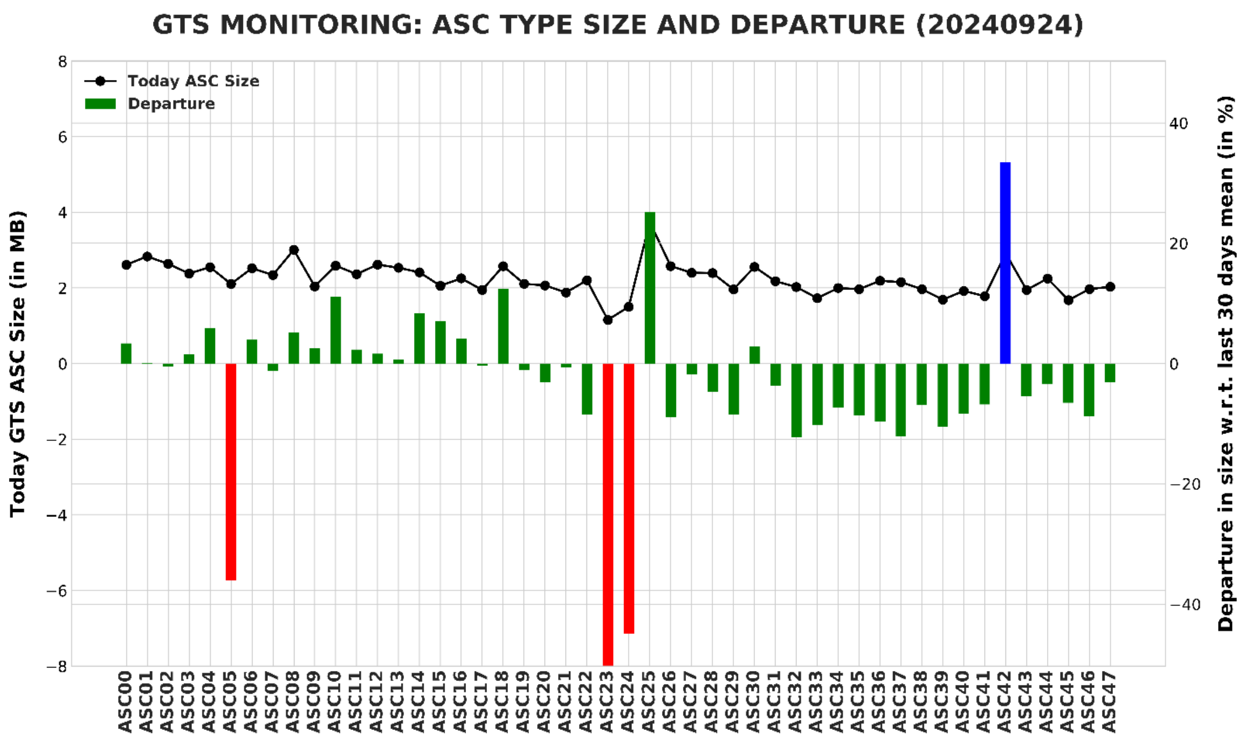

Figure 2. *Real-time monitoring of ASC file size and its departure w.r.t 30 days mean for a particular date (20240924).*

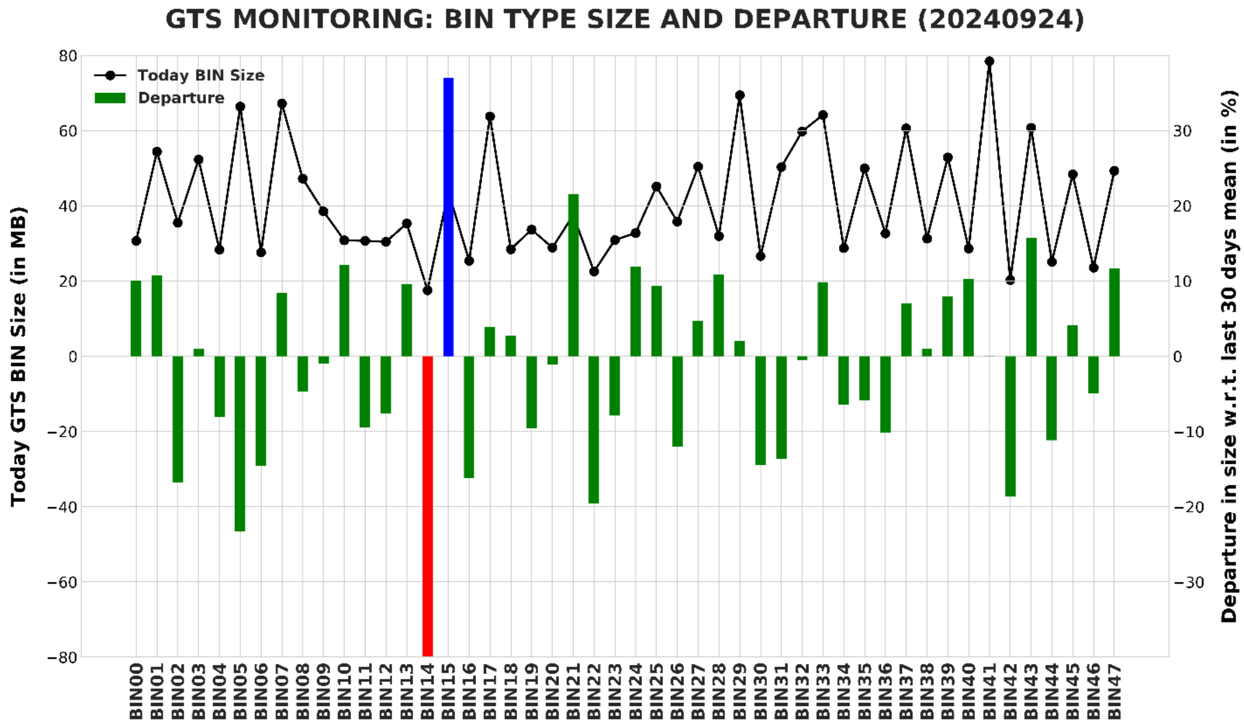

Figure 3. *Real-time monitoring of BIN file size and its departure w.r.t 30 days mean for a particular date (20240924).*

Figures 2 and 3 shows the real-time monitoring of ASC and BIN file sizes (black line) and their departure with respect to last 30-days mean (colour bars) for 24 September 2024. The blue, red and green bars represent the excess, deficient and normal data file size for the current day with respect to the previous 30-day mean sizes. This allows a quick, visual interpretation of anomalies in the file size. These figures are invaluable in ensuring that no significant deviations in file sizes go unnoticed. The output figure file names are dynamically created from the input file, ensuring proper identification and storage.

D. ASC and BIN file size time series of previous 30-days:

Script Name:

a) calc_plot_asc_bin_size_timeseries_daily.sh,

b) plot_gts_asc_bin_monit_size_daily.py

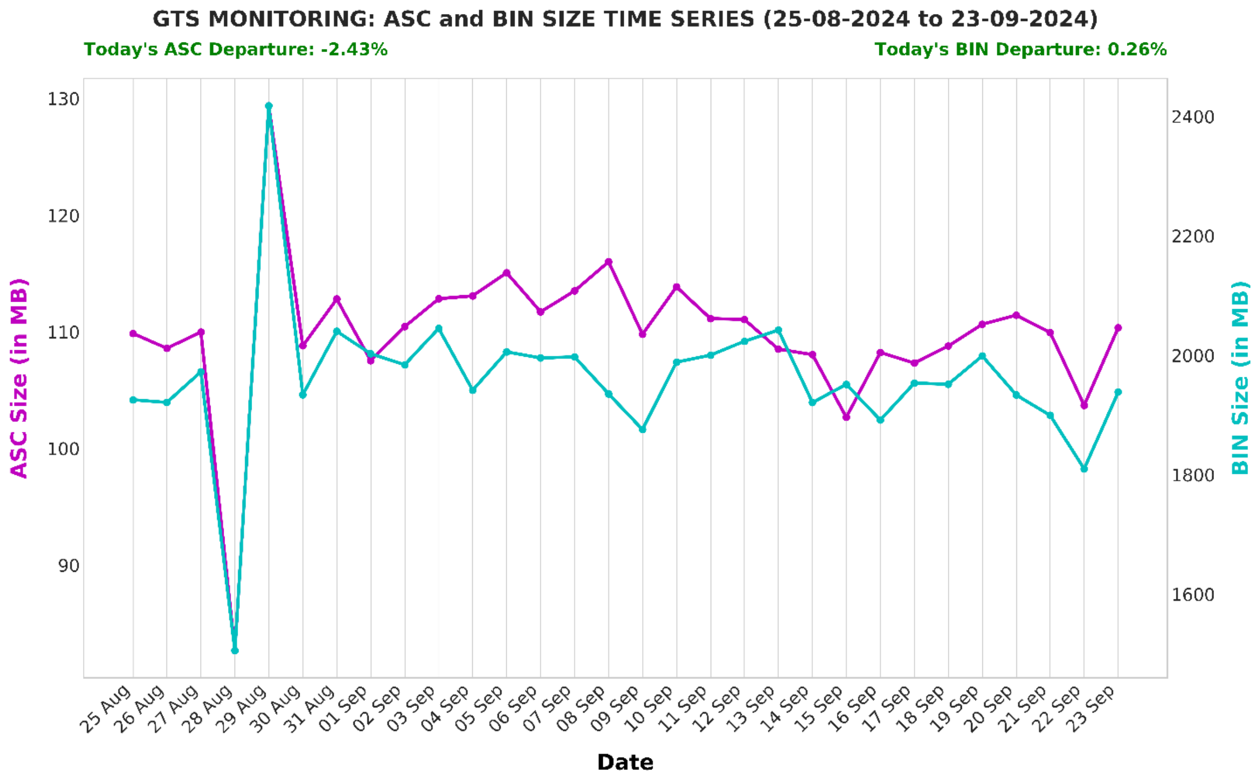

Figure 4. *Real-time dynamic time-series for the ASC and BIN file sizes for the previous 30 days.*

A daily shell script plots the time series of both ASC and BIN file sizes for the past 30 days. Another python script compares the daily file sizes and overlays today's size departure information at the top of the generated figure for both file types (Figure 4). This allows for long-term trend analysis, making identify recurring anomalies or seasonal changes in data sizes easier. This script generates reports consolidating ASC and BIN file size behaviours, offering a broader view of the GTS data flow. The output file names are dynamically created from the input file, ensuring proper identification and storage. The time-series plot provides a concise, visual representation of both ASC and BIN file types' sizes over time, making it easy to track changes and spot trends.

5. Conclusions

The real-time monitoring of ASC and BIN files received at NCMRWF through GTS is a robust process, using dynamic plotting and color-coded departure alerts, that ensures data integrity and timely response to anomalies. The system allows for effective decision-making and ensures the continuity of GTS data flows through automated monitoring, size departure calculations, dynamic time-series generation, and real-time visualizations. The ability to compare file size trends and observe potential anomalies ensures the robustness of GTS data transfer and processing systems. Despite its resource demands, the benefits of ensuring data accuracy and system health far outweigh the cons, making real-time monitoring an essential tool in data management.

Acknowledgments

We would like to express our sincere appreciation to the MIHIR HPC support team for their unwavering support and dedication in facilitating data archiving at NCMRWF, Noida.

Author Contributions

Upal Saha: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing.
S. Indira Rani: Conceptualization, Methodology, Supervision, Writing – Review & Editing.
D. Srinivas: Resources, Data Curation.
V. S. Prasad: Resources, Data Curation, Project administration.

References

Prasad, V.S., 2020. GTS Data Pre-Processing and Archival System at NCMRWF. *NCMRWF Technical Report* (No. NMRF/TR/01/2020), pp. 1-48.

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. In NCMRWF Technical Report (No. NMRF/TR/01/2007), pp. 1-42.