



NMRF/TR/06/2023



TECHNICAL REPORT

**IMDAA reanalysis like products from
NCUM operational global NWP system**

**Gibies George, Syam Sankar, S. Indira Rani
and John P. George**

2023

***National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector-62, NOIDA, India 201 309***

IMDAA reanalysis like products from NCUM operational global NWP system

Gibies George, Syam Sankar, S. Indira Rani and John P. George

NMRF/TR/06/2023

**National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences, Government of India
A-50, Sector 62, NOIDA-201309, INDIA
www.ncmrwf.gov.in**

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/06/2023
3	Month of publication	November 2023
4	Title of the document	IMDAA reanalysis like products from NCUM operational global NWP system
5	Type of Document	Technical Report
6	No of pages, Figures and Tables	27 pages, 8 figures and 11 tables
7	Number of References	12
8	Author (S)	Gibies George, Syam Sankar, S. Indira Rani and John P. George
9	Originating Unit	NCMRWF
10	Abstract	<p>Indian Monsoon Data Assimilation and Analysis (IMDAA) is a long-term (1979-2018) satellite era regional reanalysis over Indian monsoon domain developed at NCMRWF in collaboration with Met Office, UK under the Monsoon Mission Project of the Ministry of Earth Sciences (MoES). With the request from many users, NCMRWF extended the IMDAA further two more years up to December 2020. There is a recurring request from many of our valuable users for the near-real-time generation of IMDAA. Compared to the current NCMRWF Unified Model (NCUM) global operational data assimilation system (NCUM-G-DA), which can assimilate many of the new generation satellite data (radiance and wind profiles), IMDAA data assimilation system is relatively old, which cannot use many of the new satellite data. Thus, continuing IMDAA beyond 2020 with original setup may not yield better products than the operational system. As an alternative, to satisfy the user request, NCMRWF decided to extract the IMDAA equivalent products from the 12 km NCUM global operational Numerical Weather Prediction (NWP) system analysis from January 2021. A software package has been developed to extract and process the IMDAA equivalent variables from the operational global NCUM NWP system (hereafter called “IMDAA-like” products). This technical report describes the details of the data fields included in IMDAA-like products as well as the salient features of the processing package developed to prepare the IMDAA-like products.</p>
11	Security classification	
12	Distribution	Unrestricted Distribution
13	Key Words	IMDAA, NCUM, NCUM-G-DA, Data Assimilation,

Table of Contents

	Abstract	1
1	Introduction	2
2	Data description	3
3	Processing package and its components	4
4	Discussion on the scope of the products	17
5	Conclusion	25
	References	26

IMDAA reanalysis like products from NCUM operational global NWP system

Gibies George, Syam Sankar, S. Indira Rani and John P. George

सारांश

भारतीय मानसून डेटा आकलन और विश्लेषण ('इम्डा') पृथ्वी विज्ञान मंत्रालय की मानसून मिशन परियोजना के तहत यू.के. के मौसम कार्यालय के सहयोग से राष्ट्रीय मध्यम अवधि मौसम पूर्वानुमान केंद्र में विकसित भारतीय मानसून डोमेन पर एक दीर्घकालिक (1979-2018) उपग्रह युग का क्षेत्रीय विश्लेषण है। कई उपयोगकर्ताओं के अनुरोध के साथ, राष्ट्रीय मध्यम अवधि मौसम पूर्वानुमान केंद्र ने इम्डा को दिसंबर 2020 तक और दो साल बढ़ा दिया। 'इम्डा' की निकट-वास्तविक पीढ़ी के लिए हमारे कई मूल्यवान उपयोगकर्ताओं से एक आवर्ती अनुरोध है। वर्तमान रा.म.अ.मौ.पू.के. यूनिफाइड मॉडल (एन.सी.यू.एम.) ग्लोबल ऑपरेशनल डेटा एसिमिलेशन सिस्टम (एन.सी.यू.एम.-जी-डी.ए., जो कई नई पीढ़ी के उपग्रह डेटा (विकिरण और पवन प्रोफाइल) को आत्मसात कर सकता है) की तुलना में, 'इम्डा' डेटा एसिमिलेशन सिस्टम (जो कई नए उपग्रह डेटा का उपयोग नहीं कर सकता है) अपेक्षाकृत पुराना है, इस प्रकार, मूल सेटअप के साथ 2020 के बाद 'इम्डा' को जारी रखने से परिचालन प्रणाली की तुलना में बेहतर उत्पाद नहीं मिल सकते हैं। एक विकल्प के रूप में, उपयोगकर्ता के अनुरोध को पूरा करने के लिए, रा.म.अ.मौ.पू.के. ने जनवरी 2021 से 12 कि.मी. एन.सी.यू.एम. वैश्विक परिचालन एन.डब्ल्यू.पी. प्रणाली विश्लेषण से 'इम्डा' समकक्ष उत्पादों को निकालने का निर्णय लिया। परिचालन वैश्विक एन. सी. यू. एम. एन. डब्ल्यू. पी. प्रणाली (जिसे इसके बाद 'इम्डा-जैसे' उत्पाद कहा जाता है) से 'इम्डा' समकक्ष चरों को निकालने और संसाधित करने के लिए एक सॉफ्टवेयर पैकेज विकसित किया गया है। यह तकनीकी रिपोर्ट 'इम्डा-जैसे' उत्पादों में शामिल डेटा क्षेत्रों के विवरण के साथ-साथ 'इम्डा-जैसे' उत्पादों को तैयार करने के लिए विकसित प्रसंस्करण पैकेज की मुख्य विशेषताओं का वर्णन करती है।

Abstract

Indian Monsoon Data Assimilation and Analysis (IMDAA) is a long-term (1979-2018) satellite era regional reanalysis over Indian monsoon domain developed at NCMRWF in collaboration with Met Office, UK under the Monsoon Mission Project of the Ministry of Earth Sciences (MoES). With the request from many users, NCMRWF extended the IMDAA further two more years up to December 2020. There is a recurring request from many of our valuable users for the near-real-time generation of IMDAA. Compared to the current NCMRWF Unified Model (NCUM) global operational data assimilation system (NCUM-G-DA), which can assimilate many of the new generation satellite data (radiances and wind profiles), IMDAA data assimilation system is relatively old, which cannot use many of the new satellite data. Thus, continuing IMDAA beyond 2020 with original setup may not yield better products than the operational system. As an alternative, to satisfy the user request, NCMRWF decided to extract the IMDAA equivalent products from the 12 km NCUM global operational Numerical Weather Prediction (NWP) system analysis from January 2021. A software package has been developed to extract and process the IMDAA equivalent variables from the operational global NCUM NWP system (hereafter called "IMDAA-like" products). This technical report describes the details of the data fields included in IMDAA-like products as well as the salient features of the processing package developed to prepare the IMDAA-like products.

1. Introduction

Indian Monsoon Data Assimilation and Analysis (IMDAA) reanalysis system produced a reanalysis dataset for the period 1979-2018 for the whole Indian monsoon region (Rani et al., 2020, 2021). The IMDAA reanalysis domain extends from 30°E to 120°E and 15°S to 45°N. Further due to the increasing demand from the users an extension of the IMDAA for the period 2019-2020 was also completed. The major difference between the IMDAA and IMDAA-extension is the use of Lateral Boundary Conditions (LBC) from the NCMRWF Unified Model (NCUM) 12 km Global Data Assimilation System (NCUM-G-DA) in the place of ERA-interim LBC used till December 2018. IMDAA regional reanalysis products from 1979 to 2020 can be freely downloaded using the Uniform Resource Locator (URL) address <https://rds.ncmrwf.gov.in/> from the Reanalysis Data Server (RDS) of NCMRWF. A brief description of IMDAA single and multi-level variables with their temporal frequency is also available in the web portal. The IMDAA products are very useful for various applications like (1) the training of hydrology models specific for the Indian Monsoon Domain, (2) survey of potential zones for wind power harvesting, (3) calibration and validation of automatic weather station instruments, (4) study the history of tropical synoptic weather systems like monsoon depressions, tropical cyclones, etc.

NCMRWF has been receiving requests from our users to continue the IMDAA products beyond 2020. The coverage of the IMDAA-domain including Tibetan plateau and equatorial Indian ocean is very important to capture all features of the Indian Monsoon domain. Many of the satellite mission datasets used in the IMDAA reanalysis stopped because of the end of those satellite missions. In the meantime, the global NWP centres started assimilating data from new generation satellites (MetOp-C, NOAA-20, etc.) in 2020 (Ingleby et al., 2021, Kumar et al., 2021). Many operational centres, including NCMRWF, started assimilating wind data from a novel lidar instrument onboard the Aeolus satellite, which has a good impact on the analysis and forecast (Rennie et al, 2021; George et al., 2021). The IMDAA reanalysis system of 12 km horizontal resolution, comparatively older than the current NCUM global operational system of 12 km resolution, cannot assimilate the information from these new generation satellites unless we make many changes in the reanalysis system. However, changing the DA system in the course of reanalysis will not serve the purpose, as the reanalysis system is designed to use the same DA and model throughout the reanalysis period. The model climatology and the observation impact of the current operational system may be different from that of the system used for the IMDAA products available for the period until 2020. Setting up a new reanalysis configuration may require a long time and dedicated manpower. The effort also demands a lot of computational resources, which is not affordable on the existing capacity. However, as a short-term alternative, to satisfy the user requests, NCMRWF decided to produce the IMDAA equivalent products from the 12 km NCUM global operational configuration from January 2021 (Table-1) and made them available through the IMDAA reanalysis web portal to the users. In this context, NCMRWF developed a software package to extract and process the IMDAA equivalent data fields from the operational NCUM-G-DA. The processing software package for generating IMDAA-like products is operational at Mihir HPCs and has one day lag synchronisation with a dedicated data server for the dissemination of the data. The IMDAA-like product contains more than forty fields which are classified into various groups (Table-2). The IMDAA-like product

generation software includes (1) a data extraction script (2) a visualisation toolkit and (3) an automated slideshow file generator.

A description of the single and multi-level data that has to be extracted from the NCUM-G-DA system over the IMDAA-domain is provided in section 2 of this report. The package developed for the preparation of IMDAA-like products and its various components are described in section 3. An overview of the various products extracted over the IMDAA-domain and a discussion on the wide scope of the products is given in section 4, followed by a summary of the current technical package in section 5.

2. Data description

The operational version of NCUM-G-DA system is based on Unified Model version 11.2 (UM-11.2) which is an upgraded version of that used in IMDAA regional reanalysis system (UM-10.2). This is a stand-alone global atmospheric general circulation model based on EndGame dynamical core with 12 km horizontal resolution (N1024) and 70 vertical levels (L70) where the model top is at 80 km. The Observation Processing System version 2020 (OPS-2020) and VARIational data assimilation system version 2020 (VAR-2020) with capabilities to assimilate the latest available observations (Kumar et al. 2020, 2021; George et al. 2021), are used in the NCUM-G-DA system in contrast with their older versions used in the IMDAA reanalysis system. NCMRWF routinely runs the NCUM-G-DA, four times a day with six hourly intermittent cycles during 00, 06, 12, and 18 Z. Table 1 compares the NCUM-G-DA operational system components with their corresponding versions used in the IMDAA reanalysis system.

Table 1: Details of the NCUM-G-DA system components used for IMDAA products.

Dataset Name	System description	Start date	End date
IMDAA reanalysis	<ul style="list-style-type: none"> ● Unified Model version 10.2 (Davies et al., 2005; Brown et al., 2012; Wood et al., 2014; Rani et al. 2021) ● Regional (30°E to 120°E, and 15°S to 45°N) ● 63 vertical levels ● Model top at 40 km ● ops-2016.03.0 ● var-32.0.0-t2704 ● surf-32.0.0 ● jules-4.8 ● Lateral Boundary Conditions from ERA-interim 	01/01/1979	31/12/2018
IMDAA-Extension	<ul style="list-style-type: none"> ● Unified Model version 10.2 (Rani et al., 2022) ● Regional (30°E to 120°E, and 15°S to 45°N) same as above. ● Lateral Boundary Conditions from 	01/01/2019	31/12/2020

	operational NCUM-G-DA		
Operational NCUM-G-DA (IMDAA- Like products are prepared from this system)	Please refer Kumar et al. 2020 <ul style="list-style-type: none"> ● Unified Model version 11.2 (Global) ● Global domain ● 70 vertical levels ● Model top at 80 km ● ops-2019.02.0 ● var-2019.02.1-r5533 ● surf-2019.02-r1653 ● jules-5.1-r14778 	01/01/2021	17/05/2021
	Please refer Kumar et al. 2021 <ul style="list-style-type: none"> ● Unified Model version 11.2 (Global) ● 70 vertical levels ● Model top at 80 km ● ops-2020.01.2-r14057 ● var-2020.01.1-r7442 ● surf-2019.02-r1653 ● jules-5.1-r14778 	18/05/2021	till date

3. Processing package and its components

Short range operation forecasts (9 hours) are generated based on each data assimilation cycle (00, 06, 12, and 18 Z), which consist of instantaneous model fields at hourly intervals (Kumar et al. 2020). The initial six hour (00 - 05, 06 - 11, 12 - 17, and 18 - 23 Z) fields from the short-range forecasts generated in each DA cycle of the NCUM-G-DA are used for extracting the IMDAA-like data products which are consistent with the actual IMDAA reanalysis variables. The hourly accumulated variables are valid for the middle of the hourly intervals (say 00:30, 01:30, 02:30, 03:30, ... etc) unlike the instantaneous variables.

This IMDAA-like product processing package has three main components: data extraction, visualisation tool kit, and the slideshow generator. These components are briefly described in the following sub-sections. More than forty variables are extracted from the global NCUM analysis/short-forecast and they are categorised into 12 groups, A-L. The various groups are Isobaric level data, Planetary Boundary Layer (PBL) related data, Flux at the top of the atmosphere (TOA), Flux at the surface, Instantaneous hourly data at Mean Sea Level (MSL), Screen level data (1.5 m height), Pole level data (10 m height), Tower level data (50 m height), Instantaneous cloud data, Accumulated hydrology data, Instantaneous soil layer data, and Atmospheric model-level data. General Regularly distributed Information in Binary form: Edition 2 (GRIB2) format is used for the final data, which is transferred to the RDS of NCMRWF. The basic information about each data field including time interval, type of data acquisition (instantaneous or accumulated or averaged), classification based on targeted user group, name of UM source file, UM STASH code, and GRIB code information are listed in Table 2.

Table 2: Groupwise listing of data fields extracted from NCUM-G-DA over the IMDAA domain.

A <u>Isobaric level data</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	Grib code		
							Discip.	Categ.	No.
1	HGT-prl	3 hours	inst	isobaric	pd000	s16i202	0	3	5
		Geopotential height				geopotential meter (gpm)			
2	RH-prl	3 hours	inst	isobaric	pd000	s16i256	0	1	1
		Relative humidity at isobaric levels				%			
3	TMP-prl	3 hours	inst	isobaric	pd000	s16i203	0	0	0
		Temperature at isobaric levels				Kelvin			
4	DZDT-prl	3 hours	inst	isobaric	pd000	s15i242	0	2	9
		Vertical velocity at isobaric level				Pascals per second			
5	UGRD-prl	3 hours	inst	isobaric	pd000	s15i201	0	2	2
		Zonal wind at isobaric levels				m s ⁻¹			
6	VGRD-prl	3 hours	inst	isobaric	pd000	s15i202	0	2	3
		Meridional wind at isobaric levels				m s ⁻¹			
B <u>Planetary boundary layer (PBL) related data</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
7	TMP-sfc	3 hours	inst	skin	pb000	s00i024	0	0	17
		Skin temperature				Kelvin			
8	HPBL-sfc	3 hours	inst	pbl	pb000	s00i025	0	3	18
		Planetary boundary layer height				meters			
9	ROHLENABL-sfc	3 hours	inst	pbl	pb000	s03i026	0	15	193
		Surface roughness length				meters			
10	GUST-10m	3 hours	inst	gust	pb000	s03i463	0	2	22
		Wind speed (Gust) at 10 m height				m s ⁻¹			

11	LAND-sfc	3 hours	inst	lsm	pb000	s00i030	0	0	0
		Land Sea Mask				unitless binary mask			
C <u>Flux at the top of the atmosphere (TOA)</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
11	USWRF-toa	3 hours	ave	toa	pf000	s01i205	0	4	8
		Outgoing (Upward) shortwave radiation flux at TOA				W m ⁻²			
12	DSWRF-toa	3 hours	ave	toa	pf000	s01i207	0	4	7
		Incoming (Downward) shortwave radiation flux at TOA				W m ⁻²			
13	ULWRF-toa	3 hours	ave	toa	pf000	s02i205	0	5	4
		Net Outgoing (Upward) longwave radiation (OLR) flux at TOA				W m ⁻²			
D <u>Flux at the surface</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
14	DSWRF-sfc	3 hours	ave	surface	pf000	s01i235	0	4	7
		Downward shortwave radiation flux at surface				W m ⁻²			
15	DLWRF-sfc	3 hours	ave	surface	pf000	s02i207	0	5	3
		Downward longwave radiation flux at surface				W m ⁻²			
16	NSWRF-sfc	3 hours	ave	surface	pf000	s01i202	0	4	9
		Net shortwave radiation flux at surface				W m ⁻²			
17	NLWRF-sfc	3 hours	ave	surface	pf000	s02i201	0	5	5
		Net longwave radiation flux at surface				W m ⁻²			
18	SHTFL-sfc	3 hours	ave	surface	pf000	s03i217	0	0	11
		Sensible heat flux				W m ⁻²			
19	LHTFL-sfc	3 hours	ave	surface	pf000	s03i234	0	0	10
		Latent heat flux				W m ⁻²			

20	EVARSS-sfc	3 hours	ave	surface	pf000	s03i296	0	1	193
		Evaporation from soil				kg m ⁻² s ⁻¹			
21	EVARCA-sfc	3 hours	ave	surface	pf000	s03i297	0	1	194
		Evaporation from Canopy				kg m ⁻² s ⁻¹			
E <u>Instantaneous hourly data at mean sea level (MSL)</u>									
SI No.	Data Field	Time	type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
22	PRMSL-msl	hourly	inst	sea level	pp006	s16i222	0	3	1
		Pressure at mean sea level				Pascal			
F <u>Screen level data (1.5 m height)</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
23	PRES-sfc	hourly	inst	screen	pp006	s00i409	0	3	0
		Surface pressure				Pascal			
24	TMP-2m	hourly	inst	screen	pp006	s03i236	0	0	0
		Temperature at screen level (1.5 m)				Kelvin			
25	RH-2m	hourly	inst	screen	pp006	s03i245	0	1	1
		Relative Humidity at screen level (1.5 m)				%			
26	VIS-2m	hourly	inst	screen	pp006	s03i247	0	19	0
		Visibility				meters			
G <u>Pole level data (10 m height)</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
27	UGRD-10m	hourly	inst	pole	pe000	s03i225	0	2	2
		Zonal wind at pole height (10 m)				m s ⁻¹			
28	VGRD-10m	hourly	inst	pole	pe000	s03i226	0	2	3
		Meridional wind at pole height (10 m)				m s ⁻¹			

H	<u>Tower level data (50 m height)</u>								
Sl No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
29	UGRD-50m	hourly	inst	tower	pe000	s15i212	0	2	2
		Zonal wind at tower height (50m)					m s ⁻¹		
30	VGRD-50m	hourly	inst	tower	pe000	s15i213	0	2	3
		Meridional wind at tower height (50m)					m s ⁻¹		
I	<u>Instantaneous cloud data</u>								
Sl No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
31	VLCDC-atmc	hourly	inst	cloud	pe000	s09i202	0	6	201
		Very low cloud cover					%		
32	LCDC-atmc	hourly	inst	cloud	pe000	s09i203	0	6	3
		Low cloud cover					%		
33	MCDC-atmc	hourly	inst	cloud	pe000	s09i204	0	6	4
		Medium cloud cover					%		
34	HCDC-atmc	hourly	inst	cloud	pe000	s09i205	0	6	5
		High cloud cover							
35	TCDCRO-atmc	hourly	inst	cloud	pe000	s09i216	0	6	202
		Total cloud cover (assuming random overlap)					%		

J <u>Accumulated hydrology data (hourly)</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
36	LSWP-sfc	hourly	accu	hydrology	pe000	s04i201	0	1	47
		Large scale rain amount					kg m ⁻²		
37	SNOL-sfc	hourly	accu	hydrology	pe000	s04i202	0	1	15
		Large scale snow amount					kg m ⁻²		
38	CWP-sfc	hourly	accu	hydrology	pe000	s05i201	0	1	48
		Convective rain amount					kg m ⁻²		
39	SNOC-sfc	hourly	accu	hydrology	pe000	s05i202	0	1	14
		Convective snow amount					kg m ⁻²		
40	APCP-sfc	hourly	accu	hydrology	pe000	s05i226	0	1	8
		Total precipitation					kg m ⁻²		
K <u>Accumulated hydrology data (3 hourly)</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
41	WEASD-sfc	3 hours	accu	ice	pb000	s00i023	0	1	13
		Water equivalent of accumulated snow depth (derived from snowfall amount)					kg m ⁻²		
42	ICEC-sfc	3 hours	accu	ice	pb000	s00i031	0	2	0
		Sea ice cover					unitless area fraction		
43	ICETK-sfc	3 hours	accu	ice	pb000	s00i032	0	2	1
		Sea ice thickness					meters		
L <u>Instantaneous soil layer data</u>									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
44	TSOIL-soil	hourly	inst	Soil layer	pp006	s08i225	2	0	25
		Soil temperature at soil layers equivalent to 10cm, 35cm, 1m and 2m					Kelvin		

45	VSOILM-soil	hourly	inst	Soil layer	pp006	s08i223	2	0	3
		Soil moisture (volumetric) at soil layers equivalent to 10cm, 35cm, 1m and 2m				m ³ m ⁻³			
M Atmospheric model-level data									
SI No.	Data Field	Time	Type	Class / group	Source file	STASH code	GRIB code		
							Discip.	Categ.	No.
46	PRES-mdl	hourly	inst	model level	pp006	s00i407	—	—	—
		Pressure at 63 rho levels				Pascal			

3.1 Method of extraction

As mentioned in section 2, the required data fields to be extracted over the IMDAA-domain from the NCUM-G-DA are classified into data groups, listed in Table 2. Each data field in the same group has a common source file, temporal frequency, and dimension. The extraction (Table-3), file splitting and compression (Table-4), format conversion (Table-7), and variable renaming (Table-8) procedures have been completed in ten different steps using standard tools such as MULE, UMRider, wgrib2 and NCO respectively.

3.1.1 Spatio-temporal domain extraction of Unified Model Field Files

The four steps described in table 3, used MULE (this is a Python package providing an interface to various files used and produced by Unified Model) for the data extraction of the required spatio-temporal domain. MULE operates on Unified Model Field Files (UMFF) and produces domain-specific UMFF datasets of the same spatio-temporal resolution.

Table 3: Algorithms for Spatio-temporal domain extraction.

MULE:	Algorithms for Spatio-temporal domain extraction of UMFF using mule
Step 1:	Extraction of required fields from the data source file:
	<code>mule-select \${filein} \${fileout} --include luser4=\${STASH_LIST}</code>
Step 2:	Cropping of the Indian Monsoon Domain from the global dataset:
	<code>mule-cutout coords \${filein} \${fileout} \${SWLON} \${SWLAT} \${NELON} \${NELAT}</code>
Step 3:	Time slicing the dataset for the initial six hour window:

	mule-select \${filein} \${fileout} --include lbhr=\${SELECT_HOUR}
Step 4:	Level selection of the data fields as per the requirement:
	mule-select \${filein} \${fileout} --include lblev=\${SELECT_LEVEL}

3.1.2 UMRider-2022

An improved version of the operational UMRider (Arulalan et al. 2020) application toolkit is included in the package for generating compressed GRIB files for storage and dissemination through the web data server. UMRider is an IRIS based python package used to read UMFF and produce compressed GRIB files. Unlike the original UMRider package, the improved version (UMRider-2022) does not require any NetCDF intermediate file before generating the final GRIB files and thereby streamlines the data flow uniformity across data variables. The removal of NetCDF package dependency reduces the risk of system environment conflicts and thereby improves the robustness of the package.

Table 4: Algorithms for generating compressed GRIB dataset for web based dissemination

UMRider:	Algorithms for generating compressed GRIB for web based dissemination using UMRider-2022
Step 5:	Generate compressed GRIB files using UMRider-2022 toolkit
	<code>\${HOMEDIR}/jobs/submit_umrider.sh</code>
Step 6:	Split up GRIB data into single variable single level two-dimensional files.
	<code>\${HOMEDIR}/jobs/submit_webdata.sh</code>
Step 7:	Regrid to uniform lat-lon grid.
	<code>wgrib2 \${INFILE} -new_grid latlon 30:751:0.120 -15:501:0.120 \${OUTFILE}</code>
Step 8:	Compress the GRIB file to reduce storage space requirement.
	<code>wgrib2 \${INFILE} -ncpu 2 -set_bin_prec 12 -set_grib_type complex2 -grib_out \${OUTFILE}</code>

The UMRider-2022 has four steps of data processing (steps 5 to 8 in Table 4), the first stage (step 5) generates GRIB files with multiple variables at multiple levels and the second stage (step 6) splits up these data to be distributed among small size GRIB files with a single variable and single level, to make it easy to handle with the web-based data server. Step 7 regrid the processed data to the same resolution of original IMDAA reanalysis products and in step 8, the GRIB files are compressed for the better storage efficiency.

3.1.3 Local GRIB Table

Table 2 (last columns) provide the required information for the identification of variables by the GRIB data processing packages (like wgrib2). However for some variables, the details are not available in the default GRIB table. In such cases we can use a local GRIB table for the identification of the data fields. Table-6 provides the details of entries in local GRIB table corresponding to ROHLENABL-sfc, EVARSS-sfc, EVARCA-sfc, VLCDC-atmc and TCDCRO-atmc. Similar to the previous version of UMRider (Arulalan et al. 2020), the current version also uses local GRIB table to handle variables which are not included in the GRIB table by wgrib developers. An example of a record on the local GRIB table is provided below for the demonstration purpose.

0:1:0:10:29:1:1:193:EVARSS:Evaporation Rate From Soil Surface:kg m-2 s-1

Individual information on a record in the GRIB table is separated with colons. There are eleven segments in the GRIB table entry (above example) which are explained in table 5 and required information for individual data fields, such as discipline, category and number (provided in Table 2).

Table 5: Details of local GRIB table segments and syntax.

Sl. No.	Information type	Entries in the example	Identifier	Type
1	Parameter Discipline	0	disc	Integer
2	Master Table Version Set	1	mtab_set	Integer
3	Master Table Version Start	0	mtab_low	Integer
4	Master Table Version End	10	mtab_high	Integer
5	Centre Code	29	cntr	Integer
6	Local Tables Version	1	ltab	Integer
7	Parameter Category	1	pcat	Integer
8	Parameter Number	193	pnum	Integer
9	Variable Short Name (Standard name)	EVARSS	name	String
10	Variable Description	Evaporation Rate From Soil Surface	desc	String
11	Variable Unit	kg m-2 s-1	unit	String

The environment variable “GRIB2TABLE” can be used to provide the local GRIB table to the wgrib package to enable it to identify the variable information and thereby assign proper variable names to the data field.

Table 6: Variables definition entries in local GRIB table

Sl. No.	Grib numerical code	Variable name	Description	Unit
1	0:1:0:10:29:1:15:193	ROHLENABL	Roughness length after boundary layer calculation	m
2	0:1:0:10:29:1:1:193	EVARSS	Evaporation Rate From Soil Surface	kg m-2 s-1
3	0:1:0:10:29:1:1:194	EVARCA	Evaporation Rate From Canopy	kg m-2 s-1
4	0:1:0:10:29:1:6:201	VLCDC	Very low cloud cover	%
5	0:1:0:10:29:1:6:202	TCDCRO	Total cloud cover (assuming random overlap)	%

3.1.4 NetCDF dataset for visualisation

The NetCDF file format is widely accepted worldwide, because of being a self-explanatory data file format with all the metadata and the header information stored within the file along with the data. Most of the visualisation packages support NetCDF file format. Tables 7 and 8 describe the preparation of NetCDF structured binary files for quick visualisation applications.

Table 7: wgrib2 command syntax

wgrib2:	wgrib2 command syntax
Step 9:	Create NetCDF file from GRIB2 file:
	wgrib2 \${INFILE} -netcdf \${OUTFILE}

The data file in GRIB2 format is converted to NetCDF format using the wgrib2 package with the help of a local GRIB table. For the quick processing, UMFF can be directly converted to NetCDF using other relevant applications. However, in the case of the unavailability of a local GRIB table or missing of relevant entries in the local GRIB table, the output netcdf file may not contain meaningful variable name information.

NetCDF Operators (NCO) command ‘ncrename’ is a useful tool to set meaningful variable names for the data field. Table 8 describes the NCO command syntax used for variable

renaming. The details of the renamed variables within the NetCDF files are listed in table 9 against the field number which is used as an identifier in the absence of the variable name information in the input file.

Table 8: NCO command syntax.

NCO :	NCO command syntax
Step 10:	Rename UM field variable to get meaningful nomenclature:
	<code>ncrename -v \${oldname},\${newname} \${outfile_nc}</code>

Table 9: Details of the renamed variables used in NCO

Sl. No.	Field No.	Description	Data Group	Variable Name
1	25	Visibility	screen	VIS
2	324	Roughness length after boundary layer calculation	pbl	ROHLENA BL
3	1694	Gust wind at 10m height	pbl	GUST
4	200	Shortwave radiation flux (downward) at the top of the atmosphere.	flux (toa)	DSWRFtoa
5	203	Shortwave radiation flux (downward) at the surface.	flux (sfc)	DSWRFsfc
6	1526	Evapouration over soil surface	flux	EVARSS
7	1527	Evapouration over canopy	flux	EVARCA
8	1090	Cloud amount at very low altitude	cloud	VLDC
9	33	Cloud amount at low altitude	cloud	LCDC
10	32	Cloud amount at medium altitude	cloud	MCDC
11	31	Cloud amount at high altitude	cloud	HCDC
12	30	Total cloud cover (assuming random overlap)	cloud	TCDCRO

3.2 Data visualisation toolkit

Once the dataset is ready with proper meaningful names for data variables it can be used for visualisation and dissemination as per the requirement. This section provides some details of a custom-made data visualisation toolkit. The package also consists of a customised toolkit “daview” for the visualisation of the domain-extracted dataset described above. The customised toolkit is developed using the python libraries (Pyngl and Pynio) of the National Center for Atmospheric Research (NCAR). NCAR python libraries enable most of the features of the legacy NCAR Command Language (NCL) on the python platform. The daview toolkit consists of a customised library module and an associated dictionary file which enable the reusability of the code with minimum data specific and user-level modifications in the dictionary file. The user can easily modify the dictionary file and use the daview toolkit without any prior knowledge of NCL syntax or its python equivalent. The algorithm of the daview toolkit is described in Table 10.

Table 10: Algorithm of data visualisation toolkit “daview” inside the package.

Sl. No.	Component	Description
1.	Bash edited calling script on python	<ul style="list-style-type: none"> ❖ Read the data directory path from operating system environment variable ❖ Read the plot directory path from operating system environment variable ❖ Get the list of the data fields to be plotted. ❖ Call the library function for generating the plot with arguments such as. <ul style="list-style-type: none"> ➤ Data directory path ➤ Plot directory path ➤ Date information ➤ Analysis hour ➤ Short forecast lead time (to be considered for plotting)
2.	Generate Workspace for plotting	<ul style="list-style-type: none"> ❖ Prepare the file name for which the plot is to be saved ❖ Select an appropriate file type (eg:- png/jpeg/...) ❖ Prepare the string of the date time information to be written on the plot (top-right corner outside the plot) to identify the time of validity of the information. ❖ Define a workspace using the NCAR Graphics Library function (Ngl.open_wks). ❖ Store the address of the workspace on the dictionary variable “plotinfo” for use at a later stage.
3.	Read the data	<ul style="list-style-type: none"> ❖ Prepare a string variable that holds the complete path of the data file. ❖ Prepare a string variable to handle the data slicing

		<ul style="list-style-type: none"> ❖ for the two-dimensional plotting. ❖ Store the information on a dictionary variable “datainfo” for use at a later stage. ❖ Open the data file using the NCAR input-output library function (Nio.open_file) ❖ Read the data as per the requirement along with necessary attributes and the dimensions ❖ Store the data to an appropriate key (data / zonal_data / merid_data) in the dictionary variable.
4.	Generate the plot	<ul style="list-style-type: none"> ❖ Identify an appropriate plotting function from the NCAR Graphics Library (a few commonly used examples are listed below) <ul style="list-style-type: none"> ➤ Ngl.vector_scalar_map ➤ Ngl.vector_map ➤ Ngl.contour_map ❖ Prepare an appropriate set of resource control variables as per the specification of NCAR Graphics Library. ❖ Generate the plot using the library functions listed above or any other as per the requirement.
5.	Closing the NCAR Graphics Library.	<ul style="list-style-type: none"> ❖ Frame the plot (either single or panel) to the appropriate file name and path specification referring to the dictionary variable “plotinfo” ❖ Release the workspace by closing NCAR Graphics Library (Ngl.end)

The algorithm in Table 10 provides a brief overview of the major components of the “daview” toolkit without finer details of the entire source code. The control flow of the toolkit is through the dictionary variables “datainfo” and “plotinfo” which are initialised through a user-facing dictionary file “datadic.py” inside the “pydic” folder of the package. Any common user can customise the plot appearance and details by editing the dictionary file without touching the source code of the python library file (“pylib/daview.py”).

3.3 Presentation of weather map

The package also contains a “TeX” based automated presentation generator within it. The TeX file (“tex/presentation.tex”) is invoked and modified using a bash script (“jobs/presentation.sh”) to get the correct date and time information. Tex-macros defined in the header file (“tex/header.tex”) control the appearance and details of the individual slides. Slides are sorted based on the validity hour of the day.

3.4 Package overall structure and data flow sequence

The package generates three types of products (1) GRIB2 data files for web based dissemination (2) a set of standard plots for quick visualisation (3) a slideshow file in PDF format, with optimum designing and coding of the package, nurtured by a bunch of standard resources. The primary input for the package is the Unified Model short-range forecast output

and the final product is the presentation file in PDF format. GRIB dataset and a bunch of useful plots are preserved for the detailed research scope beyond the nowcasting time window. The package contains five rose-applications for its integration with the operational rose suite. The list of rose-applications is provided in Table 11. The operational package implementation procedure is sophisticated using an installation script inside the jobs folder at the package home.

Table 11. Details of specific applications in the package and the list of external dependencies.

Sl. No.	Application Name	Purpose	Dependency
1	extract_imdaa_nrt	1. MULE based spatio-temporal domain extraction	MULE,
		2. Prepare UMFF output for UMRider-2022	
2	umrider_imdaa_nrt	generate compressed GRIB files using UMRider-2022 (internal within the package)	IRIS, GRIB
3	webdata_imdaa_nrt	1. split up data into single variable single level GRIB files for web-based data servers.	wgrib2
		2. Prepare NetCDF output for visualisation.	wgrib2, TCL, NCO
4	plot_imdaa_nrt	Visualisation toolkit on python using NCAR Graphics Library (NGL)	PyNIO, PyNGL, Numpy Pandas
5	slides_imdaa_nrt	TeX based tools for generating slideshow PDF file	TeX

4. Discussion on the scope of the products

This section provides a gist of the plots which can be generated from the package, for example for a typical day (here we present the plots valid for 00UTC (05.30 IST) and 05 UTC (11.30 IST) of 6 December 2021). Figure 1 shows the cloud amount (a,b) and wind vector (c,d) at ~ 200 hPa extracted from the NCUM-G-DA over the IMDAA domain. The left/right panel is valid for 00/05 UTC of 6 December 2021. The high-level cloud amount over the IMDAA-domain valid for 00 UTC and 05 UTC on 6 December 2021 indicates an eastward developing equatorial cluster over the maritime continent along with the remnant of Cyclonic Storm ‘JAWAD’ over the eastern part of India and Bangladesh.

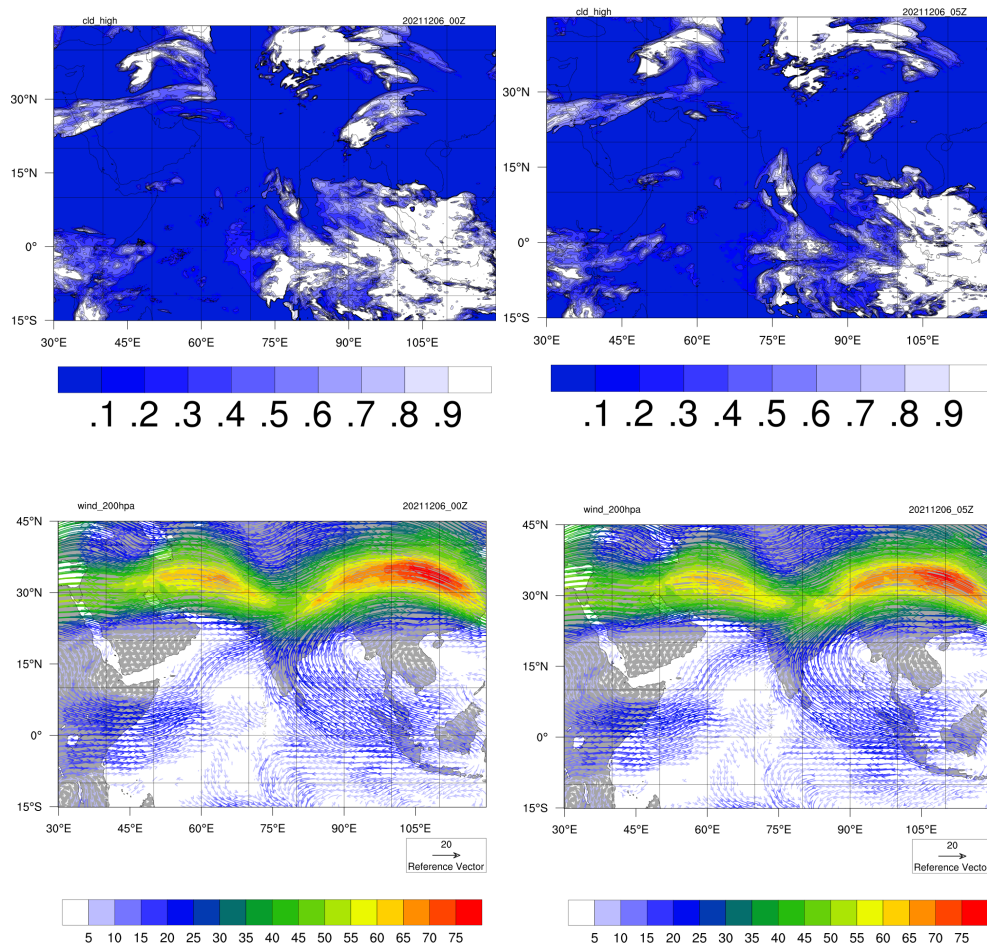


Figure 1: The cloud amount (a,b) and wind vector (c,d) at high altitude valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

Figure 2 is similar to Figure 1 but for the middle level (~ 500 hPa) valid for 00 and 05 UTC of 6 December 2021. A trough located in the westerlies at about 45°E can be identified using the mid-tropospheric cloud (Figures 2 a,b) and wind (Figures 2 c,d) at 500 hPa level along with another active trough over the 80°E. The remnant of Cyclonic Storm ‘JAWAD’ in the cloud pattern alone without any signature in the wind can also be seen.

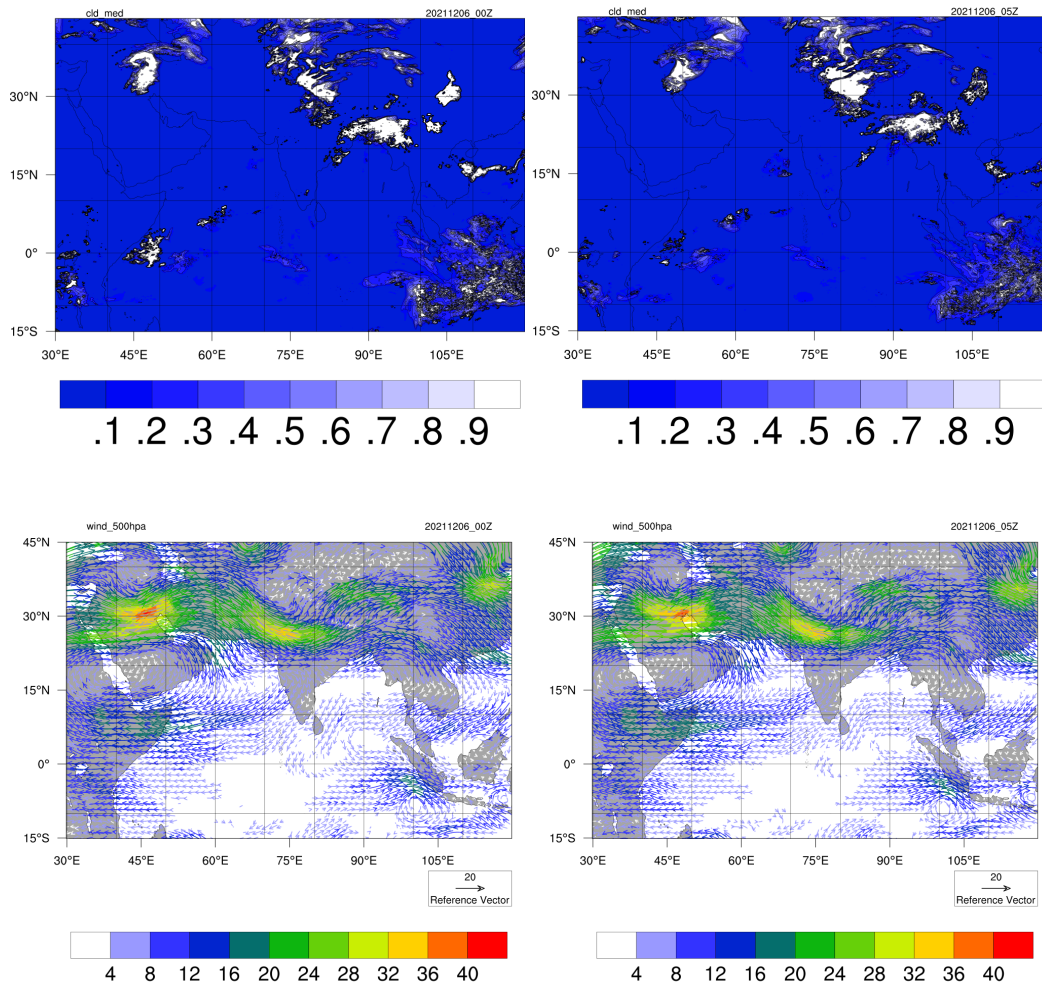


Figure 2: The cloud amount (a,b) and wind vector (c,d) at mid-troposphere valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

Figures 3 and 4 show the cloud amount and wind vectors valid for 00 and 05 UTCs over low (~850 hPa) and lower (~50m) altitudes. Dissipating remnants of Cyclonic Storm ‘JAWAD’ can be seen in the lower troposphere (850 hPa ~ 1.5 km) cloud cover (Figures 3 a, b) and in the wind (Figures 3 c,d) pattern.

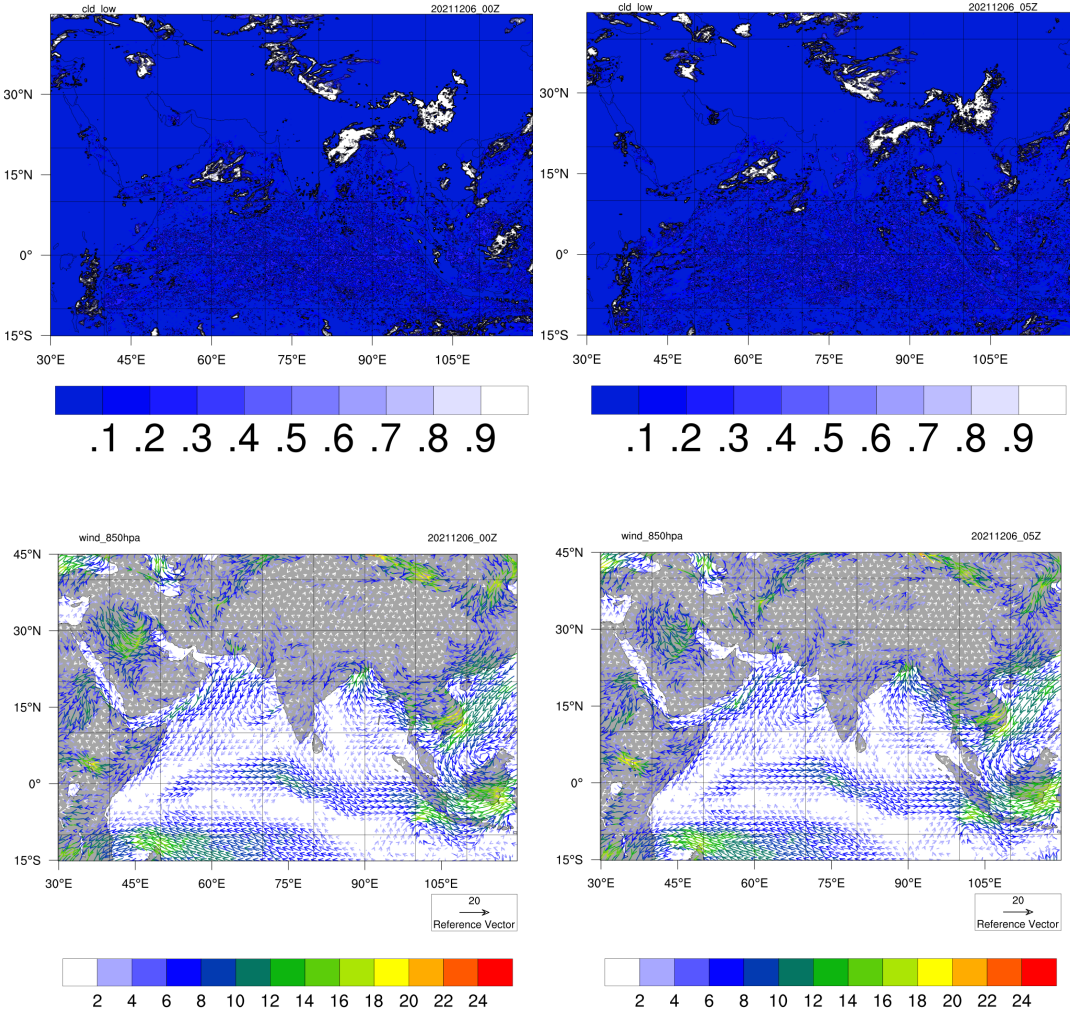


Figure 3: The cloud amount (a,b) and wind vector (c,d) at lower troposphere valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

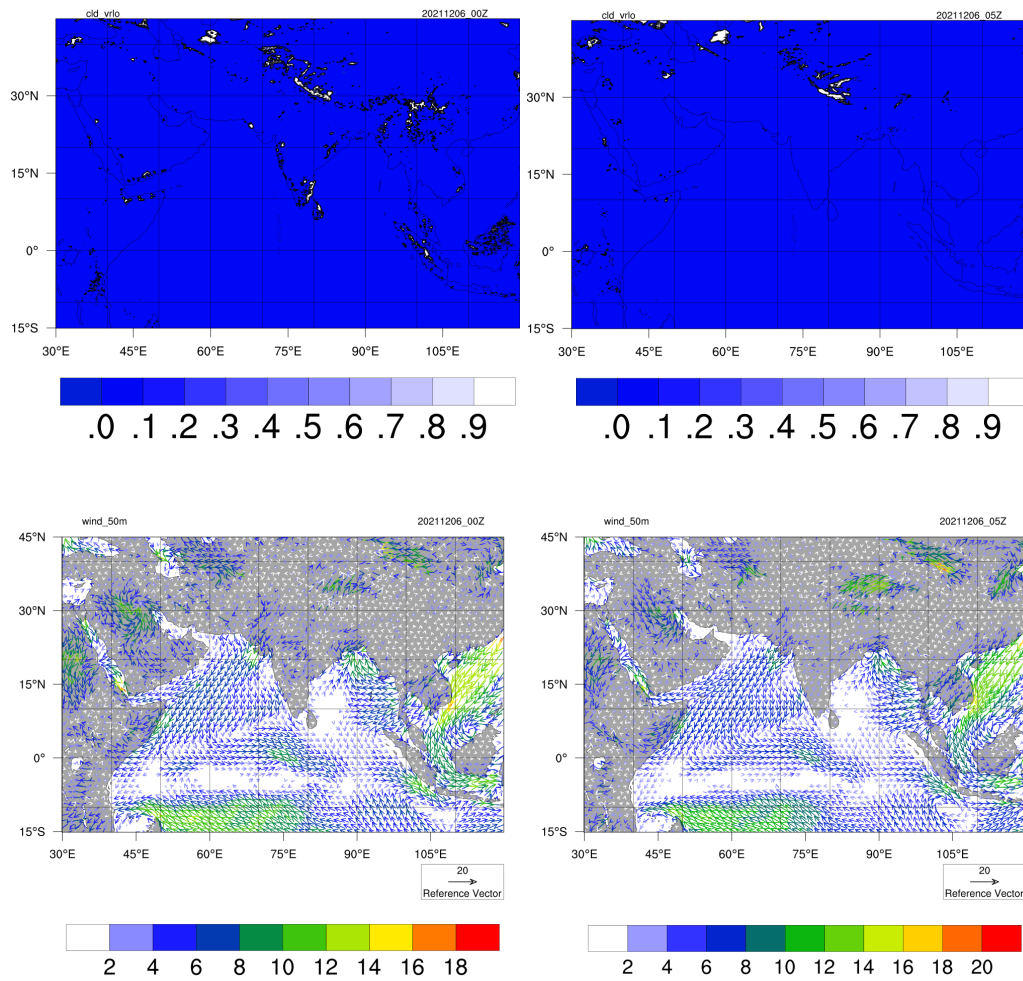


Figure 4: The cloud/fog amount (a,b) and wind vector (c,d) at very low altitude (~ 50 m) valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCU-M-G-DA.

Figure 5 is the relative humidity and visibility at the surface (2m height screen level) valid for 00 and 05 UTCs of 6 December 2021. The relative humidity over the Indian land region and southeast Asia is more than 80% before the sunrise (Figure 5a) which is reduced to nearly 50 % (Figure 5b) after the sunrise. The relative humidity over the Arabian desert regions is nearly 50% even during 0 UTC (Figure 5a).

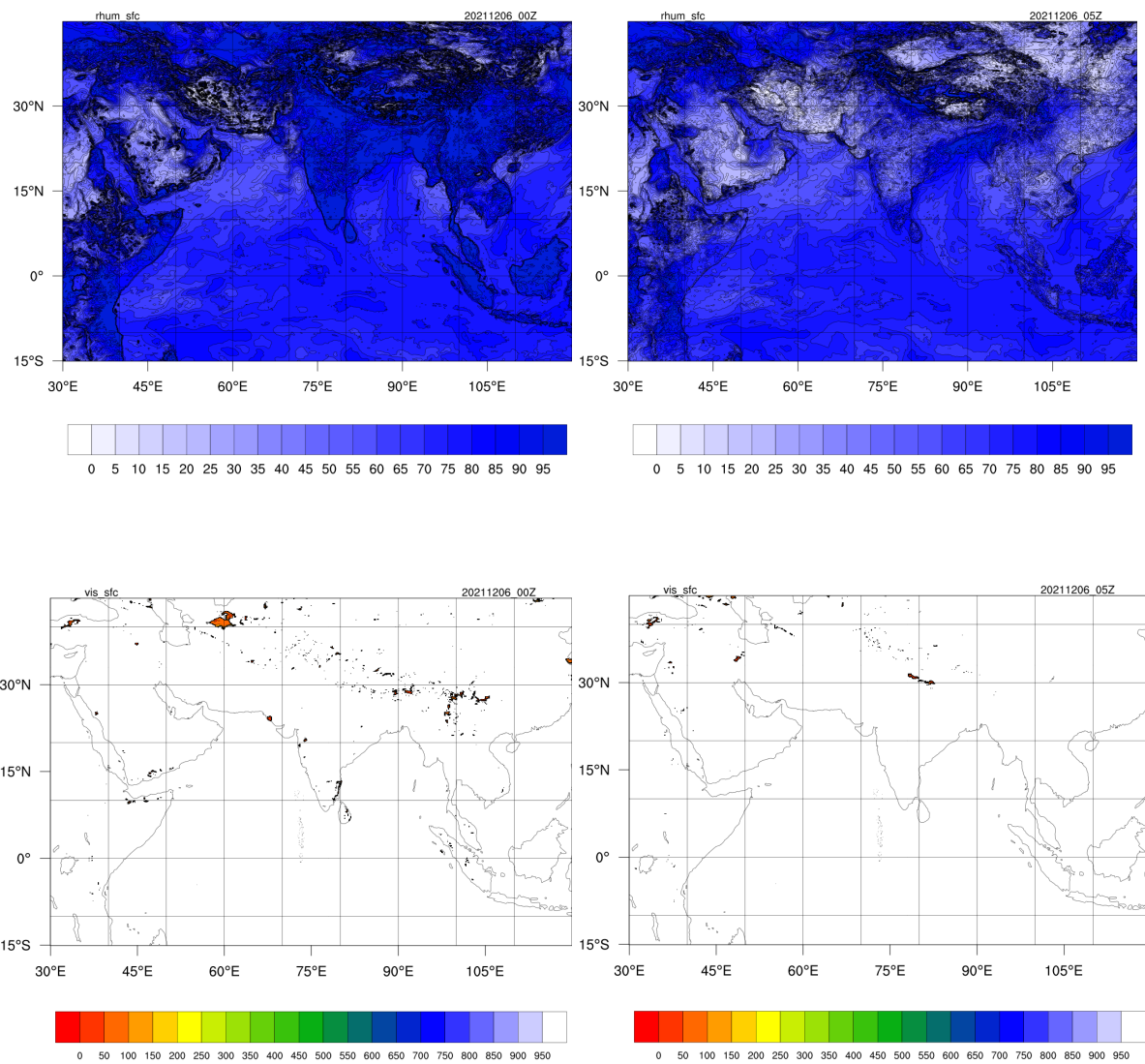


Figure 5: The relative humidity (a,b) and visibility (c,d) at surface, valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

Figure 6 shows the screen level (2m) temperature (a,b) and skin layer temperature (c, d) over land and ocean, valid for 00 UTC (a, c) and 05 UTC (b, d) on 6 December 2021. A clear land-ocean contrast of the surface temperature (Figure 6 a) can be seen at 00 UTC while the land over India and Southeast Asia started warming up immediately after sunrise (Figure 6 b).

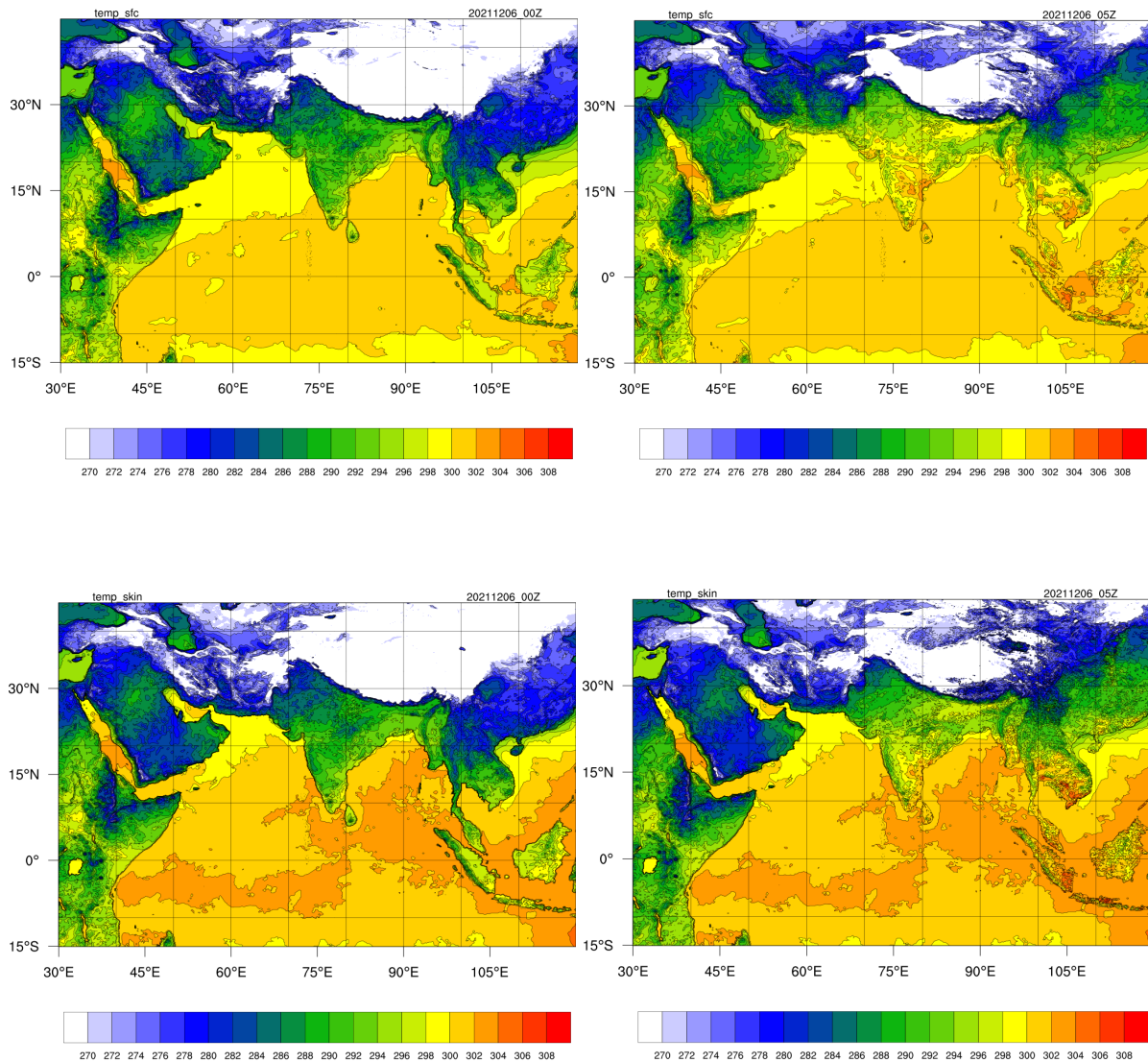


Figure 6: The atmospheric temperature at surface (a,b) and skin layer (c,d) over land and ocean, valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

Figure 7 shows the surface pressure (a, b) and the mean sea level pressure (c,d), valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021. The atmospheric pressure at the surface during 00 UTC of 6th December 2021 has the signature of the remnants of Cyclonic Storm ‘JAWAD’ as a single closed contour of about 1010 hPa in surface pressure (Figure 7a) and the same projected to mean sea level (Figure 7c), but almost vanishing five hours later (Figures 7 b,d).

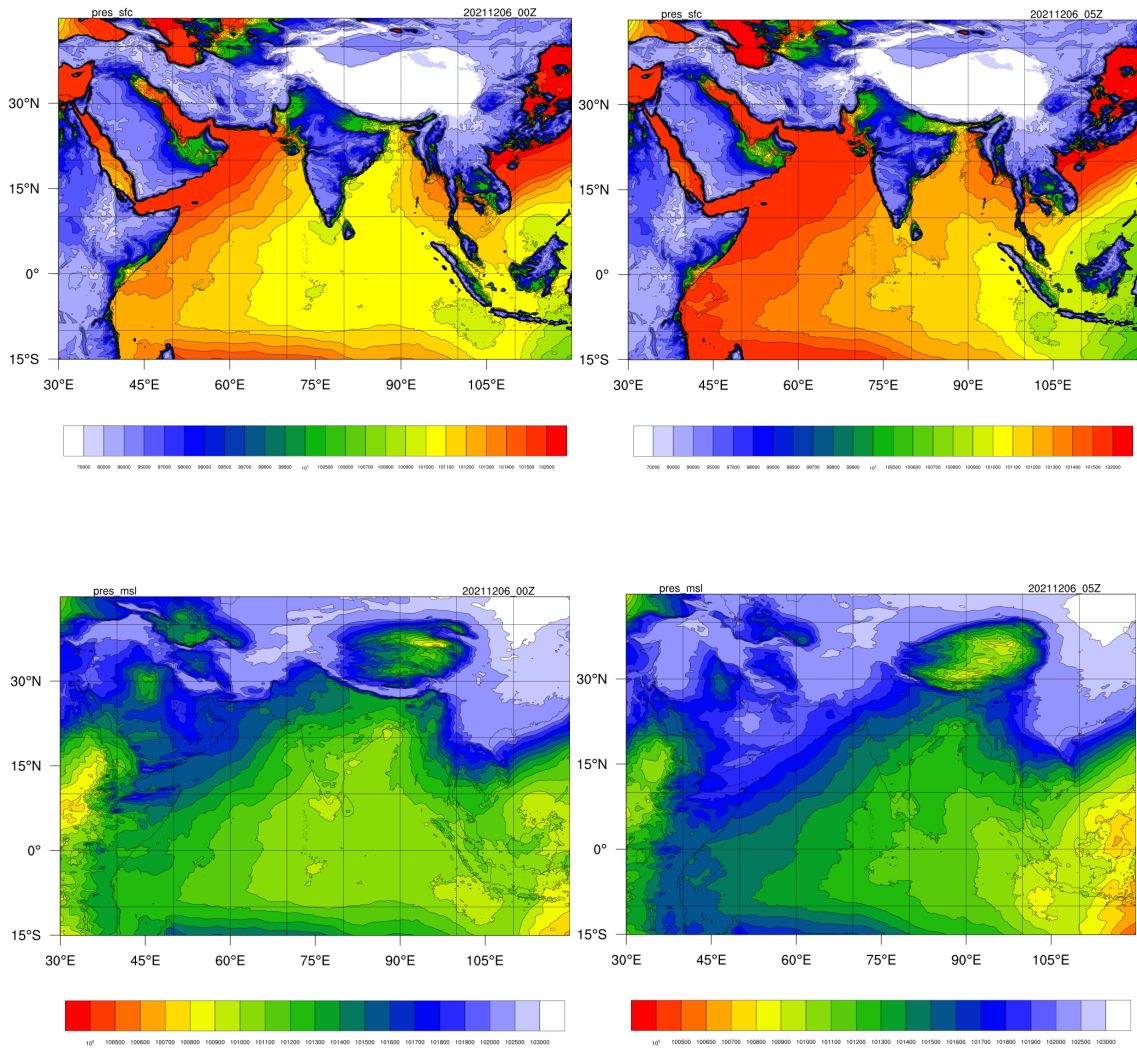


Figure 7: The atmospheric pressure at the surface (a,b) and that projected to mean sea level (c,d), valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

Figure 8 is the wind at 10 m (a,b) and soil moisture (c,d), valid for 00 UTC (a) and 05 UTC (b) on 6 December 2021. A weak signature of the remnants of Cyclonic Storm ‘JAWAD’ is evident in the wind vector at 10 m for 00 UTC (Figure 8a) as well as 05 UTC (Figure 8b). The surface wind over the western half of the Bay of Bengal is very calm (less than 2 m/s) during the early morning hours. The map of soil moisture reveals that the soil of the maritime continent region of southeast Asia, Sri Lanka, and the southernmost states of India (Kerala and Tamil Nadu) are highly saturated with moisture (Figures 8 c,d).

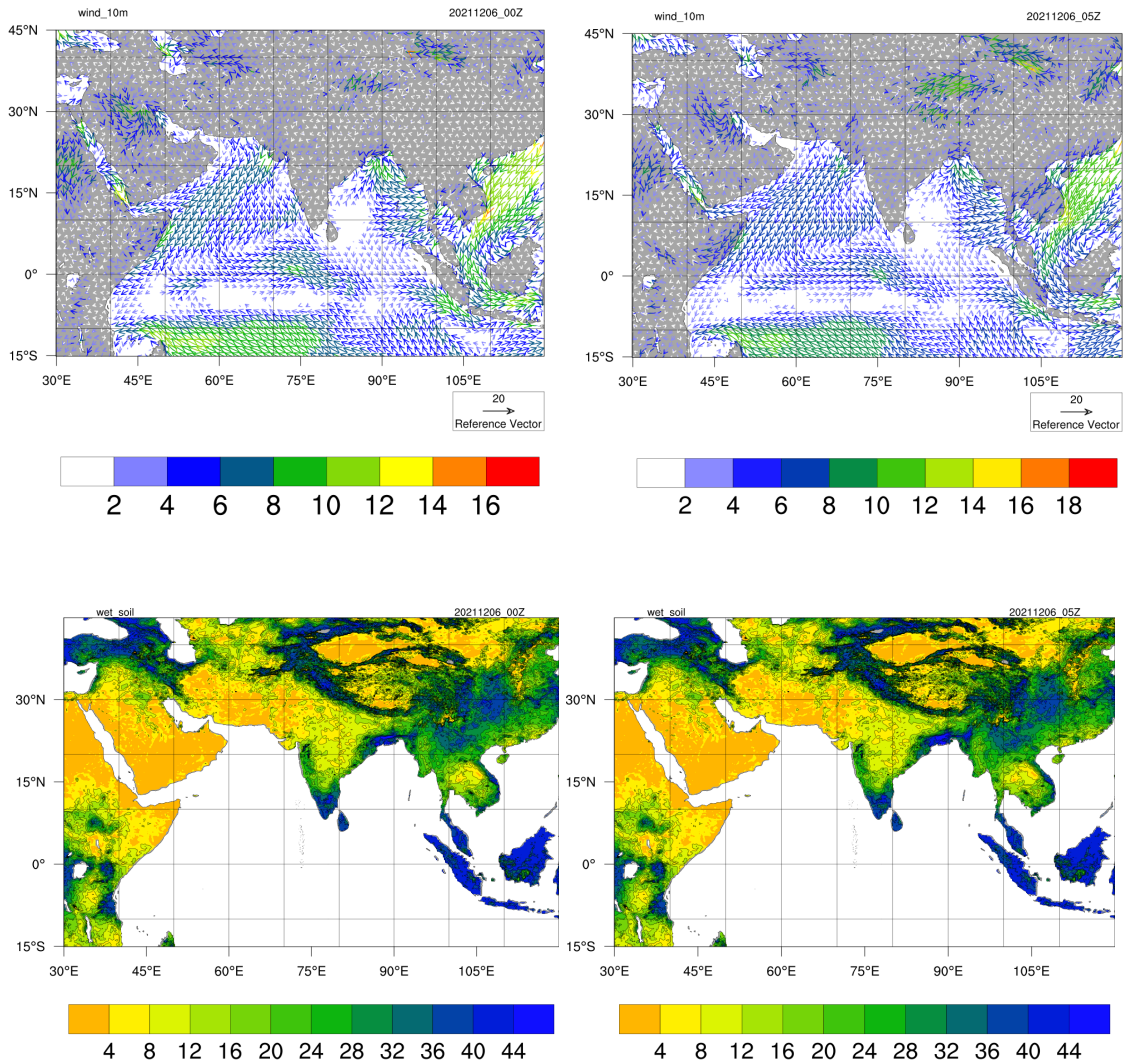


Figure 8: 10m wind (a,b) and soil moisture (c,d), valid for 00 UTC (a,c) and 05 UTC (b,d) on 6 December 2021 over the IMDAA-domain extracted from NCUM-G-DA.

5. Conclusion

An automated package has been developed to prepare IMDAA-like near-real-time products from the operational global NCUM analysis-forecast system. These IMDAA-like products are available at the NCMRWF reanalysis portal since 2021 for the users. The IMDAA-like product generation package has three main components: extract, visualise, and slideshow generator. The salient features of each component of this package are briefly described in this report. The major difference in IMDAA-like product when compared to the original IMDAA products available until 2020 is that, IMDAA-like product is derived from the operational global data assimilation system, while the original IMDAA products are derived from a regional reanalysis setup.

Acknowledgments

Authors acknowledge the Met Office, UK team for developing and releasing the MULE utility package through GitHub for public use. Authors also acknowledge the SciTools

team for developing and distributing IRIS packages. Authors acknowledge Arulalan T for the previous version of the UMRider post processing application toolkit and helpful interaction during the process of upgradation of the toolkit. Authors acknowledge NCAR for developing and releasing PyNGL and PyNIO on public domain. Authors acknowledge the comments and suggestions by Dr. Saji Mohandas, Dr. Imranali M. Momin, Dr. A. Jayakumar, Dr. Mohan S. Thota, Dr. Anurose T. J., Mr. Ankur Gupta, Dr Sumit Kumar, Dr. Vivekananda Hazra, and Dr. Lakshmikant Dhage during various stages of this work. Authors are thankful to Ms. Ruchika Malik for her support for the hindi translation of the abstract. Authors thankfully acknowledge NCMRWF internal report review committee and Head NCMRWF for timely support and guidance.

References

Arulalan, T.; (2020) UMRider –A Parallel Post Processing Utility for Unified Model. NCMRWF Technical Report, NMRF/TR/04/2020.

https://www.ncmrwf.gov.in/Reports-eng/UMRider-Report_2020.pdf

Brown, A.; Milton, S.; Cullen, M.; Golding, B.; Mitchell, J.; Shelly, A.; (2012) Unified modeling and prediction of weather and climate: a 25 year journey, Bulletin of the American Meteorological Society, 93, 1865– 1877, <https://doi.org/10.1175/BAMS-D-12-00018.1>

Davies, T.; Cullen, M.J.P.; Malcolm, A.J.; Mawson, M.H.; Staniforth, A.; White, A.A.; Wood, N.; (2005) A new dynamical core for the Met Office’s global and regional modelling of the atmosphere, Quarterly Journal of the Royal Meteorological Society, 131, 1759–1782, [doi:10.1256/qj.04.101](https://doi.org/10.1256/qj.04.101).

George, G.; Halloran, G.; Kumar, S.; Rani, S.I.; Bushair, M.T.; Jangid, B.P.; George, J.P.; Maycock, A.; (2021) Impact of Aeolus horizontal line of sight wind observations in a global NWP system. Atmospheric Research Volume 261, 15 October 2021, 105742, [doi:10.1016/j.atmosres.2021.105742](https://doi.org/10.1016/j.atmosres.2021.105742)

Ingleby, B.; Candy, B.; Eyre, J.; Haiden, T.; Hill, C.; Isaksen, L.; et al. (2021) The impact of COVID-19 on weather forecasts: A balanced view. *Geophysical Research Letters*, 48, e2020GL090699. <https://doi.org/10.1029/2020GL090699>

Kumar, S.; George, G.; Jangid, B.P.; Bushair, M.T.; Rani, S.I.; George, J.P.; (2021) NCUM Global DA System: Highlights of the 2021 upgrade, NCMRWF Technical Report, NMRF/TR/05/2021

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

Rani, S.I.; Arulalan T., George, J.P.; Rajagopal, E.N.; Renshaw, R.; Maycock, A.; Barker, D.; Rajeevan, M.; (2021) IMDAA: High Resolution Satellite-era Reanalysis for the Indian Monsoon Region, *Journal of Climate*, <https://doi.org/10.1175/JCLI-D-20-0412.1>

Rani, S.I.; George, J.P.; Arulalan, T.; Kumar, S.; Gupta, M.D.; Rajagopal, E.N.; Renshaw, R.; (2020) Evaluation of High Resolution IMDAA Regional Reanalysis Precipitation over India during Summer Monsoon Season, *CLIVAR Exchanges - Special Issue: India's Monsoon Mission*, No.79, Pages 31-33, November 2020 DOI: <https://doi.org/10.36071/clivar.79.2020>.

Rani, S.I.; Jangid, B.P.; Francis, T.; Sharma, P.; George, G.; Kumar, S.; Thota, M.S.; George, J.P.; Nath, S.; Gupta, M.D.; Mitra A.K.; (2022) Assimilation of aircraft observations over the Indian monsoon region: Investigation of the effects of COVID-19 on a reanalysis. *Quarterly Journal of the Royal Meteorological Society*; Volume 149, Issue 752, April 2023, Part A Pages 894-910, <https://doi.org/10.1002/qj.4439>

Rennie, M.P.; Isaksen, L.; Weiler, F.; Kloe, J.de; Kanitz, T.; Reitebuch, O.; (2021) The impact of Aeolus wind retrievals on ECMWF global weather forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147(740),3555–3586. <https://doi.org/10.1002/qj.4142>

Wood, N.; Staniforth, A.; White, A.; Allen, T.; Diamantakis, M.; Gross, M.; Melvin, T.; Smith, C.; Vosper, S.; Zerroukat, M.; Thuburn, J.; (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep- atmosphere global non-hydrostatic equations, *Quarterly Journal of the Royal Meteorological Society*, 140, 1505–1520

Contribution by authors:

- 1) Dr. Gibies George :
 - a) Development of the post processing package.
 - b) Writing of initial draft, editing, updating and final formatting of the manuscript.
- 2) Dr. Syam Sankar
 - a) Updation of the package source code for addition of extra variables.
 - b) Editing and updating the draft of manuscript.
- 3) Dr. S. Indira Rani
 - a) Editing and updating the draft of manuscript.
- 4) Dr. John. P. George
 - a) Supervision and guidance
 - b) Editing and updating the draft of manuscript.