scientific **data**

DATA DESCRIPTOR

Check for updates

An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024)

Akash Singh Raghuvanshi ⓘ & Ankit Agarwal ⓘ ✉

A gridded dataset of atmospheric moisture budget components constructed from ERA5 reanalysis over the Indian subcontinent and surrounding ocean regions for the period 1940–2024, referred to as ERA5moistIN. Available at 0.25 degree spatial and hourly temporal resolution, the dataset includes seven components of the column integrated moisture budget: change in storage, horizontal and vertical advection, horizontal and vertical convergence, and associated moisture flux convergence terms. These components are derived from pressure level variables including specific humidity and zonal, meridional, and vertical wind components using central finite difference methods and vertical integration, along with surface pressure. Validation against ERA5 single level outputs, such as total column water vapor and vertically integrated moisture divergence, demonstrates physical consistency and reliability. The diagnostic framework can be adapted to other reanalysis datasets and climate model outputs with minor code changes and is applicable to any region. ERA5moistIN supports high resolution analysis of moisture processes and is suitable for applications including monsoon studies, extreme event attribution, model evaluation, and predictive tool development.

Background & Summary

Atmospheric moisture plays a vital role in the Earth system by influencing energy exchange, sustaining life, and regulating temperature through latent heat release during phase changes of water. Water vapor contributes significantly to the greenhouse effect and modulates surface energy balance, with atmospheric transport of moisture redistributing latent heat across regions and timescales[1,2]. These processes form the foundation of the atmospheric branch of the hydrological cycle, which is closely linked to global and regional climate variability[3–5]. With ongoing anthropogenic climate change, there is growing concern about the intensification of hydroclimatic extremes including droughts, floods, and intense precipitation events which are expected to become more frequent and severe. These extremes are associated with both shifts in the mean state and variability of precipitation, often arising from changes in the atmospheric moisture transport pathway[6–8]. Understanding the mechanisms that drive these events requires a physically consistent quantification of the atmospheric moisture budget and its coupling to circulation dynamics and the energy cycle[6,9,10].

However, deriving moisture budget components from reanalysis data presents challenges. Moisture conservation is maintained in numerical weather prediction models and reanalyses through forward integration of the moisture equation. While these products aim for long-term closure between precipitation, evaporation, and moisture convergence, diagnostic evaluations must rely on archived fields such as three-dimensional winds, specific humidity, and surface pressure. These variables are typically stored at sub-daily or coarser time intervals and interpolated to standard vertical pressure levels, which differ from the native model grid[6]. Although past studies[1,6,11] advocate for budget closure on the native model grid using model-level data, such diagnostics remain impractical for long-term or multi-model comparisons, including those from the Coupled Model Intercomparison Project[12]. Several studies have investigated the atmospheric moisture budget over India[13–15] and other regions globally[9,16–20], often focusing on individual components such as horizontal moisture convergence or advection, especially during extreme events. However, these studies generally use coarse spatial and temporal resolutions and apply simplified assumptions or approximations, which limits their ability to resolve mesoscale and sub-synoptic processes associated with hydroclimatic extreme[6,20].

To address these limitations, we introduce **"ERA5moistIN"**, an hourly dataset of atmospheric moisture budget components over the Indian subcontinent and adjoining ocean regions (see Fig. 1), spanning the period

Department of Hydrology, Indian Institute of Technology, Roorkee, 247667, India. ✉e-mail: ankit.agarwal@hy.iitr.ac.in
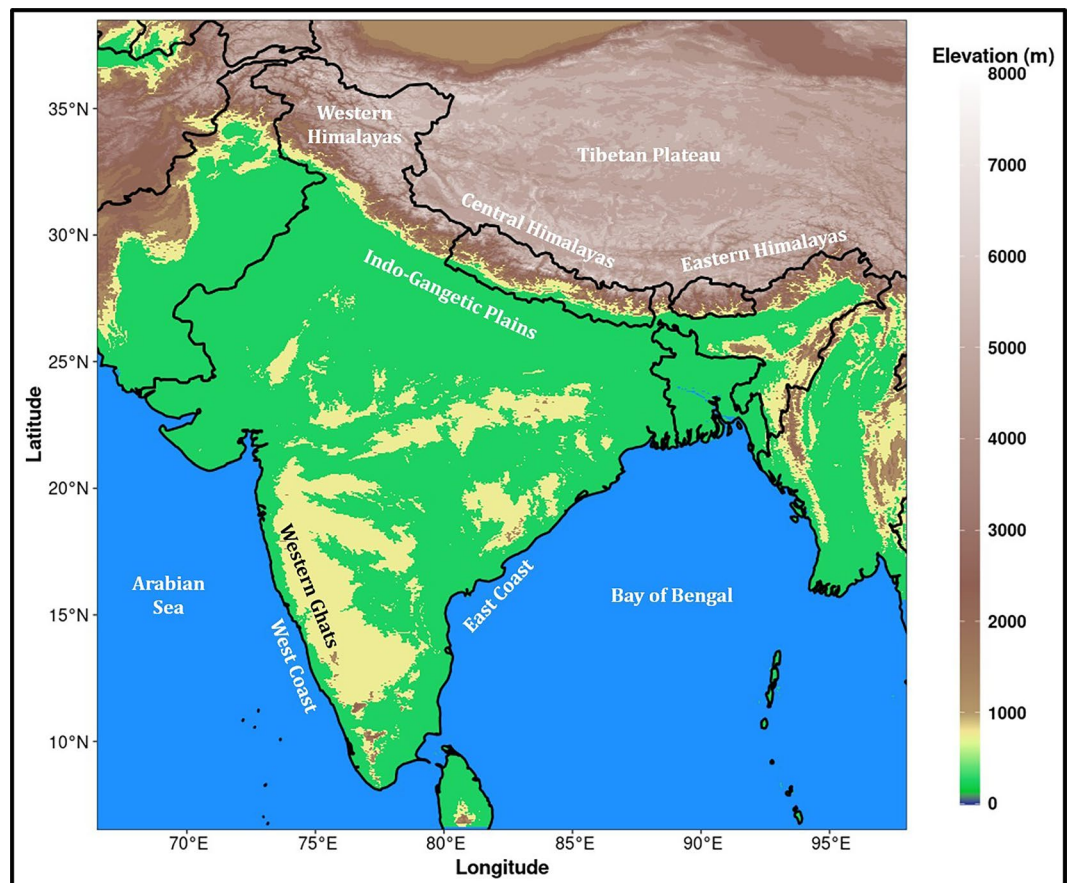
Fig. 1 Topographic map of the Indian subcontinent illustrating the geographical domain used for developing the ERA5moistIN dataset. Major physiographic features include the Western, Central, and Eastern Himalayas, Tibetan Plateau, Indo-Gangetic Plains, Western Ghats, East Coast, and adjoining oceanic regions such as the Arabian Sea and Bay of Bengal. Elevation is shown in meters.

1940 to 2024. The dataset is developed using ERA5 reanalysis data[21] on interpolated pressure levels rather than native model levels, enabling broad accessibility and compatibility with existing observational and modeling frameworks. Using central finite difference approximations, we compute key moisture budget terms including change in storage, horizontal and vertical advection, wind convergence components, and corresponding moisture flux convergences. This dataset offers high-resolution insights into moisture transport dynamics without relying on oversimplified diagnostic approximations. Although **ERA5moistIN** is based on pressure-level data, its spatial ($0.25° \times 0.25°$) and temporal resolution (1-hourly) is sufficient to capture a wide range of atmospheric processes from synoptic to sub-daily scales. The methodological framework provided in this study can be extended to other reanalysis products such as MERRA2[22], JRA55[23], and to global climate model outputs including those from CMIP6. To validate the dataset, we compare selected moisture budget diagnostics with standard ERA5 single-level outputs such as total column water vapor and vertically integrated moisture divergence, demonstrating good agreement. **ERA5moistIN** provides a one-stop data resource to investigate the full atmospheric moisture budget and its role in shaping regional hydroclimate across the Indian subcontinent. It supports the diagnosis of both long-term climatological behavior and short-term hydroclimatic extremes, including droughts, floods, and episodes of intense precipitation. This dataset will enable researchers to uncover the large-scale atmospheric controls and regional feedbacks underlying precipitation variability, land–atmosphere interactions, and moisture transport pathways critical to understanding and predicting future climate risks.

Methods

Acquisition of ERA5 Reanalysis Data. Modern meteorological reanalysis products aim to generate the most accurate possible reconstruction of the historical atmospheric state by integrating vast volumes of observational data with advanced numerical modelling and data assimilation systems[24]. These datasets have become essential tools in climate monitoring and research, offering long-term, physically consistent records of atmospheric conditions. Among these, the ERA5 dataset, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides a comprehensive global reanalysis of atmospheric, surface, and ocean wave variables. Covering the period from 1940 to the present, ERA5 incorporates information from a diverse array of observations using state-of-the-art data assimilation techniques and numerical weather prediction models[21,25]. Relative to its predecessors—ERA-Interim and ERA-40—ERA5 includes enhanced assimilation of

cloud- and rain-affected satellite radiances and features a significantly improved representation of the hydrological cycle[21]. This renders ERA5 particularly suitable for applications involving moisture and precipitation processes.

ERA5 generally yields a reliable depiction of synoptic-scale atmospheric patterns over the Northern Hemisphere, even during the early 1940s, and captures long-term climate variability in line with other established datasets[26]. Currently, ERA5 provides more than 85 years (1940–present) of 1-hourly global data, with three-dimensional atmospheric fields resolved on 137 vertical levels from the surface to ~80 km, and a horizontal grid spacing of approximately 31 km (~0.25°). In addition to atmospheric variables, ERA5 includes fields describing land surface and ocean wave conditions. The dataset is publicly available via the Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu/datasets) and has been widely applied in numerous climate and weather studies, including recent advances in machine learning-based weather forecasting[27,28]. The dataset is openly available under the Copernicus license and can be accessed as hourly single-level data[29] (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview) and hourly pressure-level data[30] (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=overview).

For the present study, we retrieved hourly ERA5 variables relevant to the atmospheric moisture budget. These include specific humidity (q), zonal (u), meridional (v) and vertical (ω) wind components on 20 pressure levels (from 1000 to 300 hPa). Additionally, single-level fields such as surface pressure (p_{sf}), vertically integrated moisture divergence (VIMD) and total column water vapour (TCWV, defined as vertically integrated specific humidity) are obtained to construct and validate the **ERA5moistIN** dataset.

Production process of ERA5moistIN. The conservation of atmospheric moisture is a fundamental physical principle governing Earth's climate system and provides a critical diagnostic for assessing the internal consistency of meteorological reanalysis datasets. The vertically integrated moisture budget, when expressed in pressure coordinates, takes the following general form:

$$\frac{1}{g}\int_{P_{top}}^{P_{sf}}\frac{\partial q}{\partial t}dp + \frac{1}{g}\int_{P_{top}}^{P_{sf}}\nabla\cdot(\vec{V}q)\,dp + \frac{1}{g}\int_{P_{top}}^{P_{sf}}\frac{\partial(\omega q)}{\partial p}dp = E - P \tag{1}$$

Here, g is the gravitational acceleration, $p_{sf}(t)$ denotes the time varying surface pressure, and p_{top} is the top pressure level. The variables q, $\vec{V} = u\hat{i} + v\hat{j}$, and ω represent specific humidity, horizontal wind vector, and vertical velocity (in pressure coordinates), respectively. The terms E and P are surface evaporation and precipitation fluxes ($kg\,m^{-2}\,s^{-1}$). The left-hand side quantifies the sum of the time rate of change in total column water vapor, horizontal moisture divergence, and vertical moisture transport. The right-hand side represents the net surface freshwater flux.

In the **ERA5moistIN** dataset, the vertical integration is conducted from the surface up to 300 hPa, a level near the tropopause. Given that specific humidity diminishes rapidly with altitude, extending the upper integration limit beyond 300 hPa introduces negligible impact on the total column moisture[31]. Thus, Eq. (1) can be rewritten specifically for this application as:

$$\frac{1}{g}\int_{300}^{P_{sf}(t)}\frac{\partial q}{\partial t}dp + \frac{1}{g}\int_{300}^{P_{sf}(t)}\nabla\cdot(\vec{V}q)\,dp + \frac{1}{g}\int_{300}^{P_{sf}(t)}\frac{\partial(\omega q)}{\partial p}dp = E - P \tag{2}$$

To analyze contributions from individual physical processes, we decompose the divergence terms in Eq. (2) using the product rule. This yields a physically interpretable form of the moisture budget, expressed as:

$$\frac{1}{g}\int_{300}^{P_{sf}(t)}\frac{\partial q}{\partial t}dp + \frac{1}{g}\int_{300}^{P_{sf}(t)}(\vec{V}\cdot\nabla q + q\cdot\nabla\vec{V})\,dp + \frac{1}{g}\int_{300}^{P_{sf}(t)}\left(\omega\cdot\frac{\partial q}{\partial p} + q\cdot\frac{\partial\omega}{\partial p}\right)dp = E - P \tag{3}$$

or

$$\underbrace{\frac{1}{g}\int_{300}^{P_{sf}(t)}\frac{\partial q}{\partial t}dp}_{\text{change in storage}} = \underbrace{\underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}(\vec{V}\cdot\nabla q)\,dp\right]}_{\substack{\textit{Horizontal}\\\textit{moisture advection}}} + \underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}(q\cdot\nabla\vec{V})\,dp\right]}_{\substack{\textit{Horizontal}\\\textit{wind convergence}}}}_{\textit{Horizontal Moisture Flux Convergence}\,(\text{HMFC})}$$

$$+ \underbrace{\underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(\omega\cdot\frac{\partial q}{\partial p}\right)dp\right]}_{\substack{\textit{Vertical}\\\textit{moisture advection}}} + \underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(q\cdot\frac{\partial\omega}{\partial p}\right)dp\right]}_{\substack{\textit{Vertical}\\\textit{wind convergence}}}}_{\textit{Vertical Moisture Flux Convergence}\,(\text{VMFC})} + E - P \tag{4}$$

Expressed in component form with $\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j}$ and $\vec{V} = u\hat{i} + v\hat{j}$, the Eq. (4) becomes:

$$\underbrace{\frac{1}{g}\int_{300}^{P_{sf}(t)}\frac{\partial q}{\partial t}dp}_{\substack{\text{change in storage}}} = \underbrace{\left\{\underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y}\right)dp\right]}_{\substack{\text{Horizontal}\\\text{moisture advection}}} + \underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(q\frac{\partial u}{\partial x} + q\frac{\partial v}{\partial y}\right)dp\right]}_{\substack{\text{Horizontal}\\\text{wind convergence}}}\right\}}_{\substack{\text{Horizontal Moisture Flux Convergence(HMFC)}}}$$

$$+ \underbrace{\left\{\underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(\omega \cdot \frac{\partial q}{\partial p}\right)dp\right]}_{\substack{\text{Vertical}\\\text{moisture advection}}} + \underbrace{\left[-\frac{1}{g}\int_{300}^{P_{sf}(t)}\left(q \cdot \frac{\partial \omega}{\partial p}\right)dp\right]}_{\substack{\text{Vertical}\\\text{wind convergence}}}\right\}}_{\substack{\text{Vertical Moisture Flux Convergence(VMFC)}}} + E - P \tag{5}$$

This expanded formulation facilitates the isolation and quantification of different physical processes driving moisture variability. The total moisture budget can therefore be summarized as:

$$\text{Change in Storage} = \underbrace{HMFC + VMFC}_{\substack{\text{Vertically Integrated}\\\text{Moisture Convergence}}} + E - P \tag{6}$$

To derive these components, we use hourly ERA5 pressure-level fields of specific humidity and the three wind components (zonal, meridional, and vertical) across 20 pressure levels from 1000 to 300 hPa, in addition to surface pressure. These fields are used to compute seven physically interpretable components of the moisture budget: horizontal and vertical moisture flux convergence (HMFC and VMFC), each split into contributions from moisture advection and wind convergence, along with the change in storage. Together, these components form the foundation of the **ERA5moistIN** dataset. This formulation enables a physically consistent framework for diagnosing water cycle dynamics, validating reanalysis-based estimates, and assessing moisture transport processes under present and future climate conditions.

Numerical calculation for the moisture budget components.

To accurately compute the atmospheric moisture budget from ERA5 reanalysis, we employ a finite-difference approach over a structured spatiotemporal grid (Fig. 2). While the governing equations of the moisture budget are continuous in form (e.g., Eq. 5), numerical weather prediction and reanalysis systems approximate these using discrete schemes. The Integrated Forecast System (IFS), which underlies ERA5, uses a semi-Lagrangian formulation for advection and a finite-difference scheme for spatial derivatives[32]. These numerical approximations contribute to residual imbalances in the moisture budget and must be considered when interpreting budget diagnostics from reanalysis data[6]. Within the **ERA5moistIN** dataset, we explicitly diagnose seven components of the vertically integrated moisture budget: (1) change in moisture storage, (2) horizontal moisture advection, (3) horizontal wind convergence, (4) vertical moisture advection, (5) vertical wind convergence, (6) horizontal moisture flux convergence (HMFC), and (7) vertical moisture flux convergence (VMFC). HMFC and VMFC are derived as sums of advection and convergence terms.

For specific humidity q, the temporal derivative is computed using a backward difference scheme:

$$\left(\frac{\partial q}{\partial t}\right)_{(i,j,k,t)} = \frac{q_{i,j,k,t} - q_{k,\ i,j,t-1}}{\Delta t} \tag{7}$$
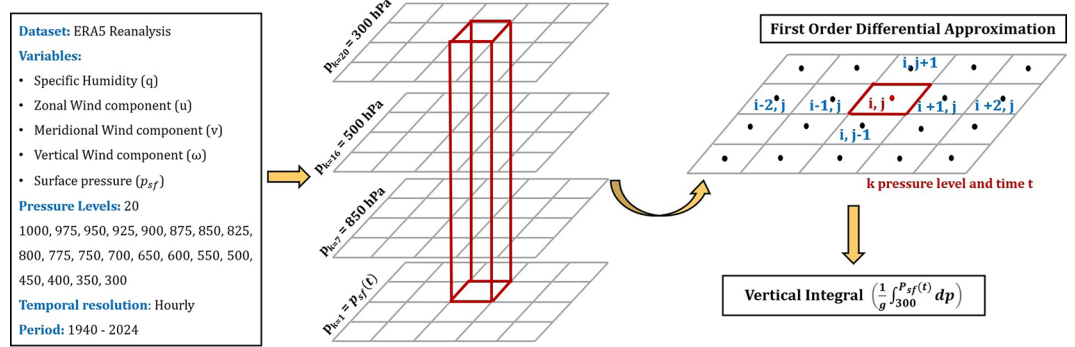
The spatial gradients in the horizontal and vertical directions are calculated using centred finite differences:

$$\left(\frac{\partial q}{\partial x}\right)_{(i,j,k,t)} = \frac{q_{i+1,j,t,k} - q_{i-1,j,t,k}}{x_{i+1,j,t,k} - x_{i-1,j,t,k}} \tag{8}$$

$$\left(\frac{\partial q}{\partial y}\right)_{(i,j,k,t)} = \frac{q_{i,j+1,t,k} - q_{i,j-1,t,k}}{y_{i,j+1,t,k} - y_{i,j-1,t,k}} \tag{9}$$

$$\left(\frac{\partial q}{\partial p}\right)_{(i,j,k,t)} = \frac{q_{i,j,k+1,t} - q_{i,j,\ k-1,t}}{p_{i,j,k+1,t} - p_{i,j,\ k-1,t}} \tag{10}$$

For boundary grid points, forward and backward differences are used:
Forward difference (for initial values):

Fig. 2 Schematic showing gridded data structure and numerical approximations used for moisture budget diagnostics. The ERA5moistIN dataset incorporates atmospheric variables (e.g., wind, humidity, pressure) on a 3D grid spanning multiple pressure levels and hourly time steps. The central 3D view shows the vertical structure used for computing vertical gradients and fluxes. The top-right horizontal grid highlights how finite-difference approximations are applied at each level to estimate horizontal derivatives. The red box marks a representative grid point (i,j,k) at time t, where specific moisture budget terms such as advection, convergence, and flux divergence are evaluated. Central differences are used for interior points, while forward and backward differences are applied at boundaries. Horizontal grid spacing is converted from latitude–longitude coordinates to metric distances, with curvature corrections included for vector derivatives. This framework enables consistent computation of key moisture budget components, including horizontal and vertical advection, wind convergence, and moisture flux convergence.

$$\left(\frac{\partial q}{\partial x}\right)_{(i-2,j,k,t)} = \frac{q_{i-1,j,t,k} - q_{i-2,j,t,k}}{x_{i-1,j,t,k} - x_{i-2,j,t,k}} \tag{11}$$

Backward difference (for last values):

$$\left(\frac{\partial q}{\partial x}\right)_{(i+2,j,k,t)} = \frac{q_{i+2,j,t,k} - q_{i+1,j,t,k}}{x_{i+2,j,t,k} - x_{i+1,j,t,k}} \tag{12}$$

The horizontal wind divergence is computed following established formulations for spherical coordinates[33]:

$$\nabla \vec{V}_{(i,j,k,t)} = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_{(i,j,k,t)} = \underbrace{\frac{u_{i+1,j,t,k} - u_{i-1,j,t,k}}{x_{i+1,j,t,k} - x_{i-1,j,t,k}}}_{\text{Zonal gradient}} + \underbrace{\frac{v_{i,j+1,t,k} - v_{i,j-1,t,k}}{y_{i,j+1,t,k} - y_{i,j-1,t,k}}}_{\text{Meridional gradient}} - \underbrace{\frac{v_{i,j,k,t}}{R}\tan\varnothing_j}_{\substack{\text{Curvature} \\ \text{Correction}}} \tag{13}$$

Similarly, the horizontal gradient of specific humidity is given by:

$$\nabla q_{(i,j,k,t)} = \left(\frac{\partial q}{\partial x} + \frac{\partial q}{\partial y}\right)_{(i,j,k,t)} = \underbrace{\frac{q_{i+1,j,t,k} - q_{i-1,j,t,k}}{x_{i+1,j,t,k} - x_{i-1,j,t,k}}}_{\text{Zonal gradient}} + \underbrace{\frac{q_{i,j+1,t,k} - q_{i,j-1,t,k}}{y_{i,j+1,t,k} - y_{i,j-1,t,k}}}_{\text{Meridional gradient}} \tag{14}$$

To convert grid spacing to metric distances on the globe, the horizontal distances are defined as:

$$x_{i+1,j,t,k} - x_{i-1,j,t,k} = R\cos\varnothing_j \cdot (\lambda_{i+1} - \lambda_{i-1}) \tag{15}$$

$$y_{i,j+1,t,k} - y_{i,j-1,t,k} = R \cdot (\varnothing_{j+1} - \varnothing_{j-1}) \tag{16}$$

where, $\varnothing$ is latitude, λ is longitude, R is Earth's radius.

The curvature correction term in Eq. (13) accounts for the convergence of meridians toward the poles[33], an important consideration for vector fields such as wind velocity ($\vec{V}$). Notably, this correction is not applied in scalar field advection (e.g., specific humidity q) because it does not involve latitudinally varying vector components.

Vertical integration of all moisture budget terms extends up to 300 hPa from the surface pressure. The pressure thickness of the lowest layer is approximated as the difference between surface pressure and the first model pressure level above it. Within this layer, the values of $\vec{V}$, ω and q are taken from the first available level above the surface. Numerical errors may arise due to such approximations, as well as from inconsistencies between the time resolution of diagnostics and the model's native time step. Hence, diagnostics based on higher temporal resolution (e.g., hourly) offer better fidelity compared to daily averages[6].

Data Records

The **ERA5moistIN** dataset is divided into seven components, each available via individual Zenodo repositories: change in storage[34] (vidq_dt.year.nc, https://doi.org/10.5281/zenodo.15751200), horizontal moisture advection[35] (viqadv.year.nc, https://doi.org/10.5281/zenodo.15730248), horizontal wind convergence[36] (viHCM.year.nc, https://doi.org/10.5281/zenodo.15751542), horizontal moisture flux convergence[37] (viHMFC.year.nc, https://doi.org/10.5281/zenodo.15753006), vertical moisture advection[38] (viwdq_dp.year.nc, https://doi.org/10.5281/zenodo.15753073), vertical wind convergence[39] (viqdw_dp.year.nc, https://doi.org/10.5281/zenodo.15753104), and vertical moisture flux convergence[40] (viVMFC.year.nc, https://doi.org/10.5281/zenodo.15753124). Each component is stored as yearly.nc files at hourly resolution, containing three-dimensional data arrays (longitude: 66.5°E–98°E, latitude: 6.5°N–38.5°N, and hourly time steps starting from January 1st at 00:00 UTC). Each file is approximately 549 MB in size, resulting in ~45.4 GB per component and a total dataset size of around 318 GB. Data values are in kg m^{-2} s^{-1} and can be converted to hourly accumulated values (kg m^{-2} hr^{-1} or mm/hr) by multiplying by 3600, which can further be aggregated to daily totals (mm/day) for analyzing relevant hydrometeorological events.

Technical Validation

To assess the accuracy of the reconstructed moisture budget components in **ERA5moistIN**, we compare its diagnostics with standard ERA5 single-level outputs: total column water vapor (TCWV) and vertically integrated moisture divergence (VIMD). The TCWV field enables estimation of the rate of change in column-integrated specific humidity, which corresponds to the change in storage term in the moisture budget. Similarly, the negative of VIMD yields the vertically integrated moisture convergence (VIMC), a proxy for the net inflow of moisture into an atmospheric column (Fig. 2). While these ERA5 diagnostics provide reference values for the quantities we aim to reconstruct from pressure-level variables (q, u, v, ω, p_{sf}), it is important to note that due to the ERA5 data assimilation scheme, these diagnostics do not perfectly close the moisture budget. In particular, the sum of ERA5's diagnosed moisture tendency and VIMD does not exactly balance surface freshwater fluxes (i.e., evaporation minus precipitation, E–P), even after accounting for changes in storage[10].

To evaluate the internal consistency and reliability of **ERA5moistIN**, we compute the root mean square error (RMSE) between its reconstructed moisture budget terms and ERA5's native diagnostics. Specifically, we compare ERA5's VIMC with the horizontal moisture flux convergence (HMFC; Fig. 3A), vertical moisture flux convergence (VMFC; Fig. 3B), and their sum (HMFC + VMFC; Fig. 3C), which together represent the total convergence in **ERA5moistIN**. Additionally, we assess the agreement between the **ERA5moistIN** storage term ($\partial q/\partial t$) and ERA5's time derivative of TCWV ($\frac{d(TCWV)}{dt}$, Fig. 3D). RMSE values (in kg m^{-2} hr^{-1} or mm hr^{-1}) are computed for each season (MAM, JJAS, ON, DJF) based on hourly accumulated data. Errors are lowest when comparing VIMC with the reconstructed total convergence (HMFC + VMFC), confirming that **ERA5moistIN** accurately reproduces ERA5's convergence using independent finite-difference approximations. Likewise, the storage term from **ERA5moistIN** aligns closely with ERA5's $\frac{d(TCWV)}{dt}$, reinforcing the validity of the method. Although **ERA5moistIN** integrates from the surface only up to 300 hPa unlike ERA5, which includes the full atmospheric column (to ~0.01 hPa), the strong agreement implies that moisture above 300 hPa contributes minimally to tropospheric moisture dynamics at the hourly timescale.

To further evaluate consistency with standard reanalysis products, we compute Pearson correlation coefficients between hourly accumulated values of **ERA5moistIN** diagnostics and ERA5's native VIMC (Fig. 4). Since ERA5 provides only total VIMC, we compare it with HMFC (Fig. 4A), VMFC (Fig. 4B), and HMFC + VMFC (Fig. 4C). High correlations across all seasons, particularly in Fig. 4C, where r > 0.8 in most regions demonstrate the ability of **ERA5moistIN** to reliably reconstruct total VIMC using independently computed components. Correlations between **ERA5moistIN**'s storage term and ERA5's $\frac{d(TCWV)}{dt}$ (Fig. 4D) are similarly strong, further supporting the credibility of the diagnostics.

Notably, spatial patterns of correlation vary between HMFC and VMFC. High correlations for HMFC over the Indo-Gangetic Plain and central India (Fig. 4A) suggest dominance of horizontal convergence in lowland regions. In contrast, lower correlations along the Western Ghats and Himalayan foothills likely stem from complex terrain, sharp moisture gradients, and uncertainties in wind fields, which can degrade horizontal convergence estimates. Conversely, VMFC (Fig. 4B) shows better agreement in these orographic zones, indicating the dominant role of vertical transport associated with terrain-induced uplift and vertical velocity structure. Interestingly, this pattern reverses over the Tibetan Plateau, where HMFC exhibits stronger correlations with ERA5 native VIMC than VMFC. This behavior may be attributed to the Plateau's extreme elevation (exceeding 5000 m; Fig. 1), where much of the atmospheric column lies below the surface. ERA5's hybrid sigma-pressure vertical coordinate system struggles to resolve vertical motion accurately over such high topography due to pressure levels intersecting the terrain, introducing uncertainty in vertical velocity (ω) and associated fluxes[41]. Additionally, the model's reduced vertical resolution near the surface in this region (Fig. 1) likely contributes to weaker VMFC correlations[21]. On the other hand, the Plateau's complex terrain significantly modulates horizontal flows and convection patterns[42], leading to stronger agreement in HMFC.

To validate **ERA5moistIN** under extreme precipitation conditions, we assess its performance during four catastrophic flood events: the 26 July 2005 Mumbai floods (Fig. 5A), 17 June 2013 Uttarakhand floods (Fig. 5B), 14 August 2018 Kerala floods (Fig. 5C), and 9 July 2023 Himachal Pradesh floods (Fig. 5D). Each row of Fig. 5 displays daily accumulated precipitation, ERA5 native VIMC, **ERA5moistIN**-derived VIMC, and the spatial bias between the two. **ERA5moistIN** computes VIMC by summing HMFC and VMFC components integrated from the surface up to 300 hPa. Despite differences in vertical integration limits and methodological formulations, both datasets exhibit strong spatial agreement over the flood-affected regions (highlighted by red boxes). While most biases are modest, more pronounced deviations occur over regions with complex topography,
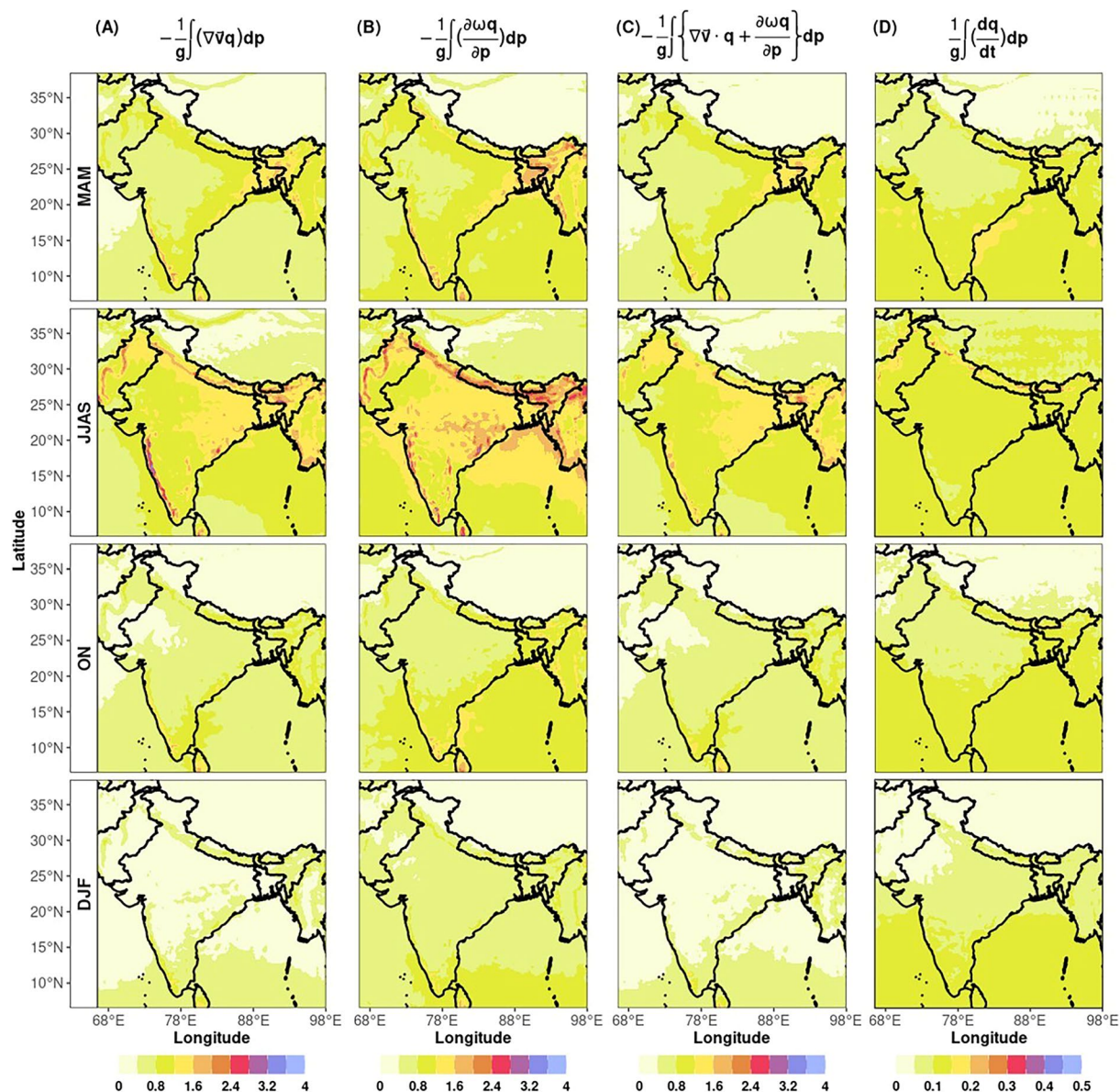
Fig. 3 Evaluation of moisture budget diagnostics in ERA5moistIN against ERA5 reference quantities (RMSE in kg m^{-2} hr^{-1}). Root Mean Square Error (RMSE) maps (in kg m^{-2} hr^{-1}, equivalent to mm hr^{-1}) are shown for four seasons—MAM (March–May), JJAS (June–September), ON (October–November), and DJF (December–February) over the period 1940–2024, comparing ERA5's native vertically integrated moisture convergence (VIMC) with (**A**) horizontal moisture flux convergence (HMFC) from ERA5moistIN; (**B**) vertical moisture flux convergence (VMFC); (**C**) sum of HMFC and VMFC (i.e., reconstructed total moisture convergence from ERA5moistIN); and (**D**) ERA5's time derivative of total column water vapor ($\frac{d(TCWV)}{dt}$) with the change in storage term from ERA5moistIN. All quantities are vertically integrated from the surface to the top of the atmosphere. Notably, ERA5moistIN's vertical integration extends only up to 300 hPa, whereas native ERA5 products (VIMC and TCWV) are integrated up to 0.01 hPa. Despite this difference, RMSE values remain low, particularly in panels (**C,D**) indicating that ERA5moistIN diagnostics effectively capture the dominant components of moisture transport and storage within the troposphere.

particularly along the Himalayas and Western Ghats. These discrepancies primarily stem from the cumulative aggregation of small hourly mismatches into daily values. Additionally, significant errors arise from the use of a simple centered finite-difference scheme to estimate the divergence operator, especially in coastal and mountainous areas where spatial gradients in the moisture convergence field are large[6].

Nonetheless, **ERA5moistIN** captures the key spatial patterns of moisture convergence during these events, underscoring its reliability even under dynamically complex and topographically challenging conditions. The fidelity of these reconstructions highlights the importance of resolving the moisture budget at finer spatio-temporal scales. Prior studies[6,20] have shown that computing VIMC at coarse temporal resolution (e.g., daily or monthly) can introduce significant biases, occasionally even altering the sign of moisture flux (from
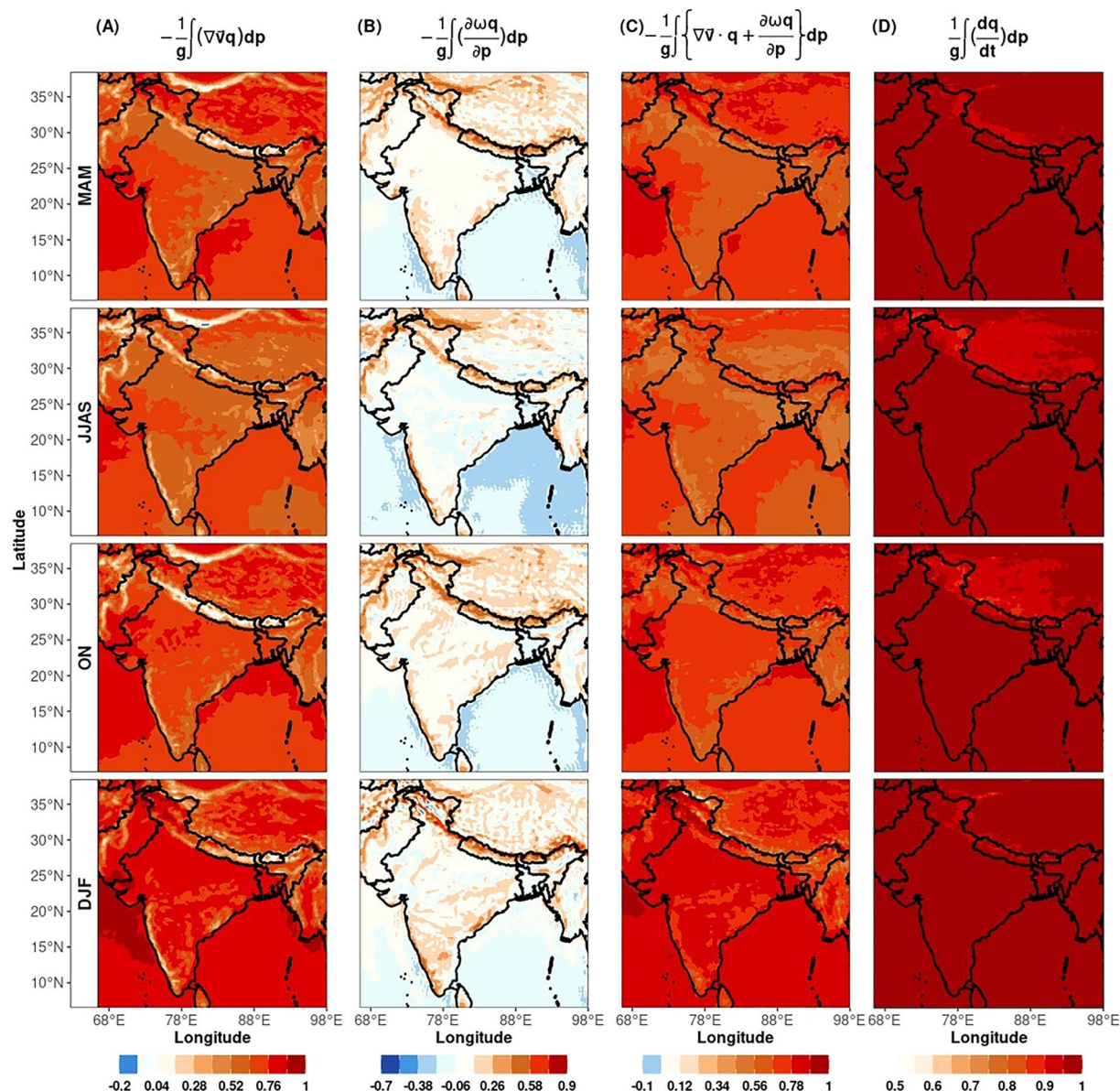
Fig. 4 Spatial correlation between ERA5moistIN moisture budget components and ERA5's native diagnostics of vertically integrated moisture convergence and change in storage. Each panel presents Pearson correlation coefficients between hourly accumulated values of ERA5moistIN-derived terms and the corresponding ERA5 native fields for four climatological seasons: MAM (March–May), JJAS (June–September), ON (October–November), and DJF (December–February) for period 1940-2024. (**A**) Correlation between ERA5moistIN horizontal moisture flux convergence (HMFC) and ERA5's native vertically integrated moisture convergence (VIMC). (**B**) Correlation between ERA5moistIN vertical moisture flux convergence (VMFC) and ERA5's VIMC. (**C**) Correlation between the sum of ERA5moistIN HMFC and VMFC and ERA5's VIMC. (**D**) Correlation between the change in storage term derived from ERA5moistIN and the time derivative of total column water vapor ($\frac{d(TCWV)}{dt}$) from ERA5. All values are calculated after accumulating hourly tendencies and converting units to kg m^{-2} hr^{-1}. ERA5moistIN diagnostics are integrated from the surface up to 300 hPa, whereas ERA5 native fields are up to the top level (~0.01 hPa).

convergence to divergence). These errors stem from the inherent variability in wind direction at sub-daily scales, particularly under mesoscale and sub-synoptic systems. Averaging over longer periods may fail to capture this high-frequency variability, distorting spatial patterns of convergence. Additionally, biases from vertical discretization, physical parameterizations, and data assimilation methods can further compound the error in reanalysis-derived moisture budgets[6,20].

Data availability
The datasets associated with this work are openly accessible via Zenodo repositories[34–40].
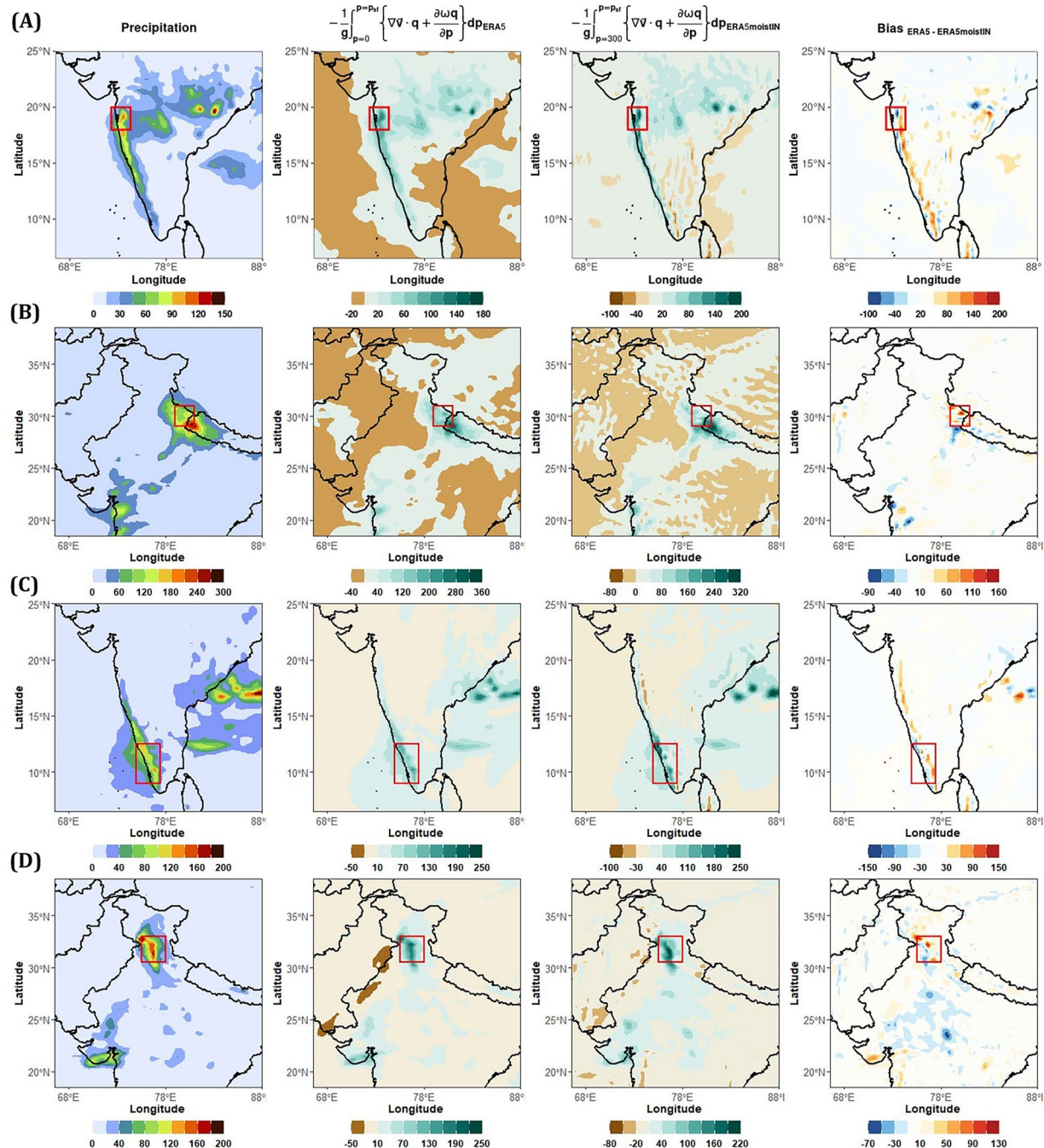
Fig. 5 Comparison of moisture convergence diagnostics during four major flood-inducing extreme precipitation events. Panels show spatial maps of daily accumulated precipitation (1st column), vertically integrated moisture convergence (VIMC) from ERA5 (2nd column), VIMC computed using the ERA5moistIN diagnostic framework (3rd column), and their difference (ERA5 − ERA5moistIN; 4th column) during four catastrophic flood events: (**A**) 26 July 2005 Mumbai floods, (**B**) 17 June 2013 Uttarakhand floods, (**C**) 14 August 2018 Kerala floods, and (**D**) 9 July 2023 Himachal Pradesh floods. VIMC from ERA5 represents the native reanalysis output, while ERA5moistIN values are derived by vertically integrating horizontal and vertical moisture flux convergence (up to 300 hPa) using finite-difference approximations. All quantities are shown in mm day^{-1}. Red boxes denote the core flood-affected regions.

Code availability

The NCL (NCAR Command Language) scripts[43] used to estimate the moisture budget components are available at https://doi.org/10.5281/zenodo.15760435. Two separate scripts are provided: one for computing horizontal components (Moist_H.ncl; change in storage, horizontal advection, horizontal convergence, and horizontal moisture flux convergence) and another for vertical components (Moist_V.ncl; vertical advection, vertical convergence, and vertical moisture flux convergence). For each year, the horizontal script stores all computed

horizontal components in a single NetCDF file, while the vertical script stores all vertical components in a separate NetCDF file, ensuring organized and efficient data handling for reproducible analysis.

References

1. Trenberth, K. E. & Guillemot, C. J. Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses. *Clim Dyn* **14**, 213–231 (1998).
2. Trenberth, K. E. & Guillemot, C. J. Evaluation of the Global Atmospheric Moisture Budget as Seen from Analyses. *J Clim* **8**, 2255–2272 (1995).
3. Gimeno, L. *et al.* Oceanic and terrestrial sources of continental precipitation. *Reviews of Geophysics* **50** (2012).
4. Dominguez, F., Kumar, P. & Vivoni, E. R. Precipitation Recycling Variability and Ecoclimatological Stability—A Study Using NARR Data. Part II: North American Monsoon Region. *J Clim* **21**, 5187–5203 (2008).
5. Patil, C., Prabhakaran, T., Sinha Ray, K. C. & Karipot, A. Revisiting Moisture Transport During the Indian Summer Monsoon Using the Moisture River Concept. *Pure Appl Geophys* **176**, 5107–5123 (2019).
6. Seager, R. & Henderson, N. Diagnostic Computation of Moisture Budgets in the ERA-Interim Reanalysis with Reference to Analysis of CMIP-Archived Atmospheric Model Data*. *J Clim* **26**, 7876–7901 (2013).
7. Seager, R., Naik, N. & Vogel, L. Does Global Warming Cause Intensified Interannual Hydroclimate Variability? *J Clim* **25**, 3355–3372 (2012).
8. Seager, R., Naik, N. & Vecchi, G. A. Thermodynamic and Dynamic Mechanisms for Large-Scale Changes in the Hydrological Cycle in Response to Global Warming*. *J Clim* **23**, 4651–4668 (2010).
9. Neelin, J. D. *et al.* Precipitation Extremes and Water Vapor. *Curr Clim Change Rep* **8**, 17–33 (2022).
10. Mayer, J., Mayer, M. & Haimberger, L. Consistency and Homogeneity of Atmospheric Energy, Moisture, and Mass Budgets in ERA5. *J Clim* **34**, 3955–3974 (2021).
11. Seneviratne, S. I., Viterbo, P., Lüthi, D. & Schär, C. Inferring Changes in Terrestrial Water Storage Using ERA-40 Reanalysis Data: The Mississippi River Basin. *J Clim* **17**, 2039–2057 (2004).
12. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *Bull Am Meteorol Soc* **93**, 485–498 (2012).
13. Pandidurai, D., Attada, R., Chakraborty, R., Kumari, A. & Kumar, P. Widespread monsoon extreme rainfall events over central Indian region: influence of dynamics and thermodynamics. *Clim Dyn* **63**, 151 (2025).
14. Sudharsan, N., Karmakar, S., Fowler, H. J. & Hari, V. Large-scale dynamics have greater role than thermodynamics in driving precipitation extremes over India. *Clim Dyn* **55**, 2603–2614 (2020).
15. Roxy, M. K. *et al.* A threefold rise in widespread extreme rain events over central India. *Nat Commun* **8**, 708 (2017).
16. Tootoonchi, R., Bordoni, S. & D'Agostino, R. Revisiting the moisture budget of the Mediterranean region in the ERA5 reanalysis. *Weather and Climate Dynamics* **6**, 245–263 (2025).
17. Pfahl, S., O'Gorman, P. A. & Fischer, E. M. Understanding the regional pattern of projected future changes in extreme precipitation. *Nat Clim Chang* **7**, 423–427 (2017).
18. Tan, Y., Yang, S., Zwiers, F., Wang, Z. & Sun, Q. Moisture budget analysis of extreme precipitation associated with different types of atmospheric rivers over western North America. *Clim Dyn* **58**, 793–809 (2022).
19. van Zomeren, J. & van Delden, A. Vertically integrated moisture flux convergence as a predictor of thunderstorms. *Atmos Res* **83**, 435–445 (2007).
20. Minallah, S. & Steiner, A. L. Role of the Atmospheric Moisture Budget in Defining the Precipitation Seasonality of the Great Lakes Region. *J Clim* **34**, 643–657 (2021).
21. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).
22. Gelaro, R. *et al.* The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J Clim* **30**, 5419–5454 (2017).
23. Kobayashi, S. *et al.* The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II* **93**, 5–48 (2015).
24. Baatz, R. *et al.* Reanalysis in Earth System Science: Toward Terrestrial Ecosystem Reanalysis. *Reviews of Geophysics* **59** (2021).
25. Bell, B. *et al.* The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society* **147**, 4186–4227 (2021).
26. Soci, C. *et al.* The ERA5 global reanalysis from 1940 to 2022. *Quarterly Journal of the Royal Meteorological Society* **150**, 4014–4048 (2024).
27. Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science (1979)* **382**, 1416–1421 (2023).
28. Bi, K. *et al.* Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
29. Hersbach, H. *et al.* ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* https://doi.org/10.24381/cds.adbb2d47 (2023).
30. Hersbach, H. *et al.* ERA5 hourly data on pressure levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* https://doi.org/10.24381/cds.bd0915c6 (2023).
31. Mo, R. *et al.* Column Relative Humidity and Primary Condensation Rate as Two Useful Supplements to Atmospheric River Analysis. *Water Resour Res* **57** (2021).
32. ECMWF. *IFS Documentation CY41R2 - Part III: Dynamics and Numerical Procedures.* **3**, (ECMWF, 2016).
33. Bluestein, H. B. *Synoptic-Dynamic Meteorology in Midlatitudes: Volume I: Principles of Kinematics and Dynamics.* **1**, (Oxford University Press, 1992).
34. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-1: change in storage] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15751200 (2025).
35. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-2: Horizontal Moisture Advection Term] [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15730248 (2025).
36. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-3: Horizontal Wind Convergence Term] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15751542 (2025).
37. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-4: Horizontal Moisture Flux Convergence Term] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15753006 (2025).
38. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-5: Vertical Moisture Advection Term] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15753073 (2025).

39. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-6: Vertical Wind Convergence Term] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15753104 (2025).

40. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [Component-7: Vertical Moisture Flux Convergence Term] (Version 1) [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.15753124 (2025).

41. Simmons, A. J. & Burridge, D. M. An Energy and Angular-Momentum Conserving Vertical Finite-Difference Scheme and Hybrid Vertical Coordinates. *Monthly Weather Review* **109**, 758–766 (1981).

42. Yang, K. *et al.* Effects of fine terrain complexity on cloud and precipitation changes over the Tibetan Plateau: a modeling study. *NPJ Clim Atmos Sci* **8**, 22 (2025).

43. Raghuvanshi, A. S. & Agarwal, A. ERA5moistIN: An Hourly Dataset of Moisture Budget Components Over the Indian Subcontinent (1940–2024) [NCL Scripts] (Version 1). *Zenodo* https://doi.org/10.5281/zenodo.15760435 (2025).

Acknowledgements

Author contributions

A.S.R. & A.A. conceived the idea and designed the problem. A.S.R. developed codes and analyzed the data with feedbacks from A.A. Both authors discussed and interpreted the results. A.S.R. wrote the first draft of the paper. Both authors reviewed and approved the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-06044-y.

Correspondence and requests for materials should be addressed to A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.