POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Machine Learning Framework for Objective Definition and Projection of Meteorological Seasons: An Application to the Indian Subcontinent

Original
A Machine Learning Framework for Objective Definition and Projection of Meteorological Seasons: An Application to the Indian Subcontinent / Grassi, J., Palazzi, E., Von Hardenberg, J., Davini, P.. - In: JOURNAL OF GEOPHYSICAL RESEARCH. ATMOSPHERES. - ISSN 2169-897X. - 131:20(2026). [10.1029/2025jd045869]

Availability:
This version is available at: 11583/3016680 since: 2026-10-15T10:50:53Z

Publisher:
Wiley

Published
DOI:10.1029/2025jd045869

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 November 2026

Check for updates

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD045869

Special Collection:
Advances in Machine Learning for Earth Science: Observation, Modeling, and Applications

A Machine Learning Framework for Objective Definition and Projection of Meteorological Seasons: An Application to the Indian Subcontinent

Jacopo Grassi[1,2], Elisa Palazzi[2], Jost von Hardenberg[1], and Paolo Davini[3]

[1]Department of Environment, Land, and Infrastructure Engineering, Politecnico di Torino, Turin, Italy, [2]Department of Physics, University of Turin, Turin, Italy, [3]Consiglio Nazionale delle Ricerche, Istituto di Scienze dell'Atmosfera e del Clima (CNR-ISAC), Turin, Italy

Abstract Meteorological seasons are typically defined through fixed calendar intervals or region-specific thresholds, yet such approaches fail to capture the actual variability of the annual cycle. Here we present a machine learning framework that objectively detects meteorological seasons and projects their future evolution. The method combines Radially Constrained Clustering (RCC) to identify seasons from reanalysis data with a supervised Multinomial Logistic Regression Classifier (MLRC) that allows these definitions to be applied also to climate model projections. We use the Indian subcontinent as a detailed proof of concept region since it has established operational references for monsoon onset and withdrawal, and it includes contrasting seasonal regimes, from monsoon dominated lowlands to high mountain regions influenced by winter disturbances. The RCC-derived seasons recover the main operational seasonal phases adopted by the India Meteorological Department, while also providing consistent definitions of transition periods. Evaluation against a CMIP6 multi-model ensemble shows that models capture the seasonal cycle with high skill in Central and Southern India, but exhibit lower performance in complex mountainous regions. Future projections indicate a robust lengthening of the monsoon season, primarily due to delayed withdrawal, and a concurrent shortening of winter, driven by the loss of days in December–January. These findings highlight the asymmetric nature of seasonal shifts under global warming and demonstrate the potential of data-driven approaches to provide objective and reproducible definitions of seasonality.

Plain Language Summary Seasons are often described in calendar terms, such as winter from December to February. In reality, the climate does not follow such fixed rules: the timing and length of seasons vary from year to year, and from region to region. In this study we use a machine learning approach to let the data itself define the seasons, based on temperature, rainfall, and wind. We then apply this method to climate model simulations to see how seasons may change in the future. We focus on the Indian subcontinent, where the summer monsoon brings most of the annual rainfall and winter storms recharge the Himalayan snowpack. Our results show that climate models reproduce the seasonal cycle well in central and southern India, but less reliably in mountainous regions. Looking to the future, we find that the monsoon season is likely to become longer, mainly because it ends later, while the winter season is projected to lose several days, particularly in December and January. This work demonstrates how machine learning can provide an objective and flexible way to track changes in seasonality, supporting better planning for agriculture, water management, and adaptation to a warming climate.

Key Points:

- We develop a machine learning framework to objectively define meteorological seasons
- The method applied to India captures monsoon onset and withdrawal with limited bias
- Future projections show longer monsoon and shorter winter with changes concentrated in withdrawal and early winter

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
J. Grassi,
jacopo.grassi@polito.it

Citation:
Grassi, J., Palazzi, E., von Hardenberg, J., & Davini, P. (2026). A machine learning framework for objective definition and projection of meteorological seasons: An application to the Indian subcontinent. *Journal of Geophysical Research: Atmospheres*, *131*, e2025JD045869. https://doi.org/10.1029/2025JD045869

Received 17 NOV 2025
Accepted 22 SEP 2026

Author Contributions:
Conceptualization: Jacopo Grassi
Data curation: Jacopo Grassi
Methodology: Jacopo Grassi
Software: Jacopo Grassi
Validation: Jacopo Grassi
Writing – original draft: Jacopo Grassi

© 2026 The Author(s). Journal of Geophysical Research: Atmospheres published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The notion of season brings together aspects at once that are both trivial and elusive. Although anyone, from the layperson to the scientist, holds a general understanding of what a season is, attempts to define its boundaries in a rigorous and objective way quickly uncover an inherent complexity. Traditional definition stems from astronomy, which divides the year into four parts based on equinoxes and solstices. Meteorologists, by contrast, usually define seasons as fixed intervals of the year characterized by relatively homogeneous weather conditions. For example, in temperate climates the year is generally divided into four equal 3-month periods (winter, spring, summer, fall) (Trenberth, 1983).

In reality, such strict regularity hardly exists in the climate system. The actual timing and duration of seasonal phases vary considerably, and the climate or ecological periods that define a "season" may be shorter or longer

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

than the conventional 3 months (Littleboy et al., 2024). The annual variation of insolation, which drives the annual climate variations, follows an approximately sinusoidal pattern which by definition does not present discontinuities.

On top of that, the four-season scheme is inherently linked to mid-latitude temperate areas. In low-latitude regions, where annual temperature variations are small, dividing the year into four temperature-based seasons is often meaningless; instead, a two-season classification tied to the precipitation cycle provides a more accurate representation of climatic patterns (Callaghan & Bonell, 2005; Guo et al., 2022). Similarly, the polar annual cycle presents two distinct phases, one dominated by a long, cold winter and one by a short, mild summer (Serreze & Barry, 2014).

This variability is mirrored in how societies conceptualize the annual cycle. Anthropological research shows that "seasons" are flexible, locally constructed frameworks used to organize livelihoods according to environmental rhythms, including the onset of rainfall, river levels, vegetation cycles, the presence of sea-ice, light conditions, and ecological cues (Bremer & Wardekker, 2024; Orlove, 2003).

The very notion of "seasons" appears to be context-dependent: societies define seasons based on the environmental variables and thresholds that matter locally, rather than adhering to universal criteria. There is a need for a framework that accommodates such variability rather than imposing fixed temporal boundaries, especially in these times of rapid climate change. Anthropogenic global warming is already changing the timing and duration of seasonal cycles: in many regions, the spring and summer onset is occurring earlier, winters are becoming shorter and traditional seasonal patterns are shifting (Cayan et al., 2001; Morisette et al., 2009; Sparks & Menzel, 2002; Woolway, 2023). Such changes have wide-ranging impacts on ecosystems (Ren et al., 2024; Uphus et al., 2023), agriculture (Minoli et al., 2022; Tran et al., 2025), water resources (Zhao et al., 2025), and human health (Lüthi et al., 2023; Ruan et al., 2025). In this context, fixed calendar definitions and region-specific thresholds remain useful for many operational applications, but they are less suited to consistently tracking seasonal changes across data sets, periods, and climate model projections. An objective, adaptive, and reproducible approach would be beneficial to consistently track and quantify these shifts.

Traditional approaches for defining meteorological seasons frequently use fixed or quartile-based thresholds applied to one or more atmospheric variables to determine their onset and end dates (Kutta & Hubbart, 2016; J. Wang et al., 2021; Hekmatzadeh et al., 2020; de la Franca et al., 2013). While such methods are suitable for specific regions or purposes, their dependence on local climatology prevents broad applications and generalization.

Data-driven approaches based on machine learning offer a powerful alternative to traditional threshold-based definitions of meteorological seasons. In particular, unsupervised clustering techniques enable the identification of groups of days or months sharing similar atmospheric meteorological states, deriving seasonal regimes directly from observed variability while integrating multiple variables for a multidimensional view of the seasonal cycle (Chattopadhyay et al., 2020; Chidean et al., 2020; Lee et al., 2023). At the same time, clustering provides an objective, reproducible framework that limits a priori assumptions about the number, timing, or drivers of seasons, letting the data determine the most coherent segmentation of the annual cycle by optimizing within-cluster similarity and maximizing between-cluster separation (Xu & Wunsch, 2008).

In this work we introduce a data-driven system to detect meteorological seasons and project their evolution through the 21st century. This study uses Radially Constrained Clustering (RCC; Cannon 2005) to divide the year into distinct, contiguous periods that can be interpreted as seasons, based on atmospheric conditions. While RCC provides an objective way to define current seasons, it cannot be directly applied to new data, such as future climate projections. To address this, we use a Multinomial Logistic Regression Classifier (MLRC; Hastie et al., 2009) to link atmospheric variables to the RCC-defined seasons, allowing the classification to be extended to independent observations and future climate scenarios.

The primary aim of this study is methodological. We propose a workflow for identifying local meteorological seasonal regimes and consistently extending these definitions from reanalysis data to climate model simulations. The Indian subcontinent is used as a proof of concept case study. This region is particularly suitable for this purpose because it combines established reference dates for monsoon onset and withdrawal, strong societal relevance, and substantial internal climatic diversity.

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The seasonal cycle of the Indian subcontinent is primarily shaped by the Indian Summer Monsoon (ISM), which provides the majority of India's annual rainfall. The ISM underpins agricultural cycles and livelihoods (Gadgil, 2003), and its timing significantly affects water resource availability and economic stability (Amrith, 2018). At the same time, the region is not climatically homogeneous and its seasonality is not controlled by the summer monsoon alone. Indeed, winter precipitation over northern India and the surrounding mountain ranges is influenced by Western Disturbances (WDs; Pisharoty and Desai, 1956), which are extratropical cyclonic systems embedded within the subtropical westerly jet that bring cold-season precipitation, contribute to Himalayan snowpack recharge, and influence downstream hydrology (Dimri et al., 2015; Hunt et al., 2025). The coexistence of monsoon-dominated lowlands and mountain regions affected by winter circulation systems makes the Indian subcontinent a suitable domain for assessing whether the proposed framework can identify physically meaningful seasonal regimes under geographically contrasting climatic conditions.

The proposed framework is not intended to replace operational definitions of monsoon onset and withdrawal, which are designed to monitor specific transitions. Rather, its added value lies in providing a multivariate segmentation of the full annual cycle without prescribed variable-specific thresholds, including winter and the pre- and post-monsoon transition seasons, that can be transferred consistently from reanalysis to climate model simulations.

The paper is structured as follows: Section 2.1 introduces the case study region and its climatic peculiarities, and Section 2.3 the machine learning algorithms deployed for the analysis. Finally, Sections 3.1 and 3.2 describe the results obtained respectively for the definition of the seasons in the historical period and their projection in the 21st century.

2. Study Area, Data and Methods

2.1. Study Area

Since the primary objective of this study is methodological, the choice of the test region is critical. We selected the Indian subcontinent (5–40°N, 65–100°E) because it offers extensive information, established operational references for the seasonal cycle, and substantial internal climatic diversity. Here we examine spatial patterns throughout the region with a special focus on four sub-regions, selected to represent contrasting seasonal regimes (Figure 1).

The first region, Southwest India (74–78°E, 8–15°N), mostly includes the Indian state of Kerala and represents the typical early landfall area of the Indian Summer Monsoon. The timing of the onset of the monsoon in this region is operationally monitored and widely used as a reference indicator for the seasonal transition across the subcontinent (Pradhan et al., 2017). The second region, Central India (75–85°E, 15.5–25.5°N), lies within the core monsoon zone and is considered a benchmark for evaluating changes in the onset, duration and intensity of the monsoon season (Desamsetti et al., 2022; Pawar et al., 2012). The third region, Himalayas (78–93°E, 25–32 °N), features a complex topography and is predominantly influenced by the Indian Summer Monsoon, which produces most of the annual precipitation from June to September (Palazzi et al., 2013). Finally, the Hindu Kush-Karakoram region (71–78°E, 32–37°N) spans the north western mountainous belt and is affected by both the summer monsoon and, to a greater extent, winter Western Disturbances, synoptic systems known to trigger precipitation during the cold season (Hunt et al., 2025; Palazzi et al., 2013). As a consequence, this region has a considerably different seasonal cycle from the monsoon-dominated lowlands and is included as an internal benchmark for assessing whether the framework can identify meaningful seasonal structures beyond the canonical Indian monsoon cycle. This within-domain heterogeneity does not constitute a full test of transferability across global climatic regimes, but it provides a first evaluation of the method under contrasting seasonal conditions within South Asia.

2.2. Data

We used both reanalysis and climate model simulations. As reference, we employed ERA5 reanalysis daily data (Hersbach et al., 2020), obtained from hourly outputs at original resolution (0.25° × 0.25°). The variables considered are 2-m temperature (tas), total precipitation (pr), and the zonal (ua850) and meridional (va850) wind components at 850 hPa. These variables are among the most widely used in climatic characterization studies, as they jointly capture the main thermodynamic and dynamic components of the seasonal cycle.

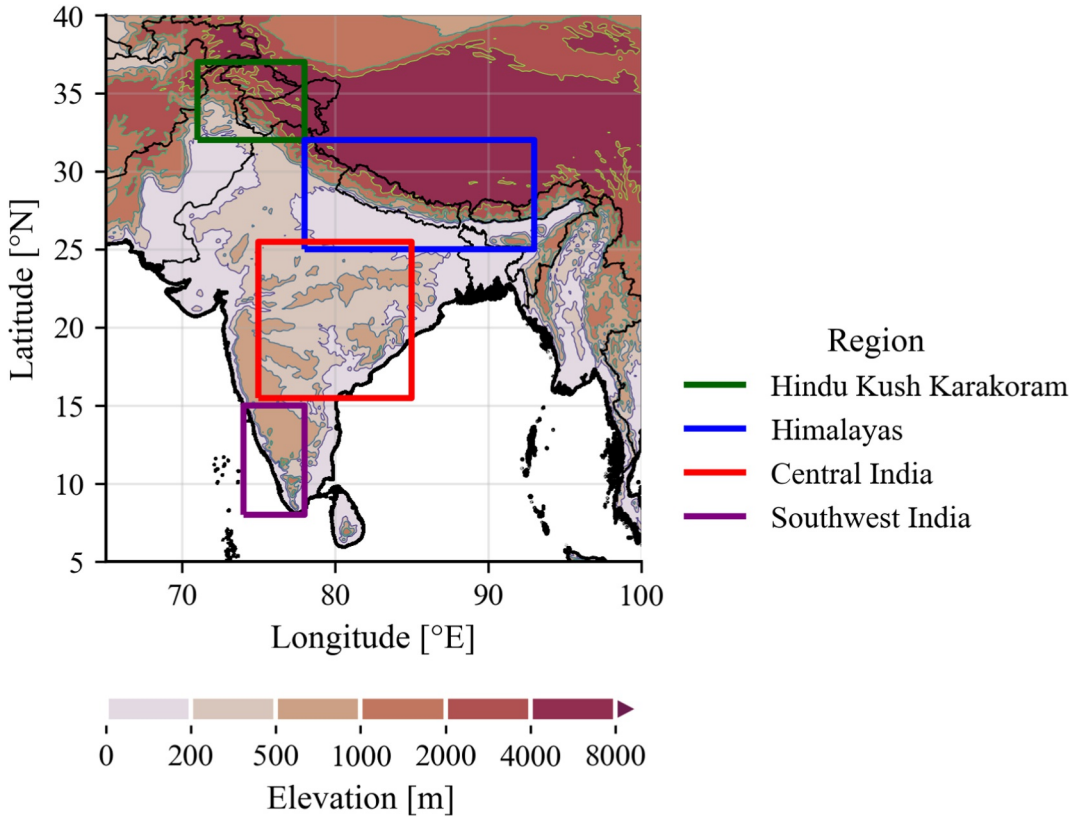


Figure 1. Overview of the selected domain over the Indian subcontinent. Orography is represented by shadings. The four analyzed regions are highlighted by colored boxes.

Reference values for the resulting seasonal transitions were compared with the most recent climatological normal onset and withdrawal dates of the Indian Summer Monsoon (ISM), as provided by the India Meteorological Department (IMD). The climatological normals correspond to the mean onset and withdrawal dates calculated over the period 1961–2019. These reference dates, available on a location-wise basis in Pai et al. (2020), were interpolated onto a 1° × 1° latitude–longitude grid using bilinear interpolation. The IMD dates are derived from a multi-step deterministic algorithm that combines precipitation thresholds with temporal continuity criteria. For a detailed description of this procedure, readers are referred to Pai et al. (2020).

To assess future changes, we analyzed a Coupled Model Intercomparison Project Phase 6 (CMIP6) Multi-Model Ensemble (MME), including both historical simulations (1961–2014) (Eyring et al., 2016) and future projections under the high-emission SSP5-8.5 scenario (2015–2100) (O'Neill et al., 2016), using daily data. The ensemble includes one member for each different model to maximize model spread. The list of models used is reported in Table 1.

ERA5 and CMIP6 data underwent a common pre-processing pipeline. First, all instances of 29th February were removed to ensure a uniform 365-day calendar. The data were then regridded to a common 1° × 1° latitude–longitude grid using bilinear interpolation to ensure spatial consistency across different native resolutions. Finally, a 15-day centered rolling mean was applied to each variable to smooth high-frequency fluctuations and emphasize intra-seasonal variability relevant to the identification of meteorological seasons.

Since one of the goals of the analysis is to evaluate how seasonal patterns are represented in future projections, we additionally applied a common detrending and normalization procedure to both ERA5 and CMIP6 data in order to remove signals that could bias the detection. Specifically, for each variable and each grid point, standardized anomalies (z-scores) were computed within a 10-year moving window for every calendar day by subtracting the running mean and dividing by the corresponding running standard deviation. This normalization was therefore applied consistently across variables, spatial locations, and temporal windows in both data sets. The resulting

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
List of CMIP6 Climate Models Used in This Study

Model	Institution	Resolution [lat x lon]	Member	Reference
ACCESS-CM2	CSIRO-ARCCSS	~1.25° × 1.875°	r1i1p1f1	Bi et al. (2020)
BCC-CSM2-MR	BCC	~1° × 1°	r2i1p1f1	Wu et al. (2019)
CanESM5	CCCma	~2.8° × 2.8°	r1i1p1f1	Swart et al. (2019)
CESM2-WACCM	NCAR	~1.25° × 0.94°	r1i1p1f1	Danabasoglu (2019)
CNRM-CM6-1	CNRM-CERFACS	~0.7° × 0.7°	r1i1p1f2	Voldoire et al. (2019)
EC-Earth3	EC-Earth-Consortium	~0.7° × 0.7°	r1i1p1f1	Döscher et al. (2022)
FGOALS-g3	CAS	~2° × 2°	r1i1p1f1	Li et al. (2020)
GFDL-CM4	NOAA-GFDL	~1° × 1°	r1i1p1f1	Held et al. (2019)
IPSL-CM6A-LR	IPSL	~1.3° × 2.5°	r1i1p1f1	Boucher et al. (2020)
MIROC6	JAMSTEC	~1.4° × 1.4°	r1i1p1f1	Tatebe et al. (2019)
MPI-ESM1-2-HR	MPI-M	~0.94° × 0.94°	r1i1p1f1	Gutjahr et al. (2019)
MRI-ESM2-0	MRI	~1.1° × 1.1°	r1i1p1f1	Yukimoto et al. (2019)
UKESM1-0-LL	Met Office Hadley Centre	~1.25° × 1.875°	r1i1p1f2	Sellar et al. (2019)

standardized fields filter out low-frequency drifts, such as the gradual warming signal, and act as an effective debiasing procedure across models and periods, ensuring that the clustering and classification focus on the internal seasonal dynamics rather than on systematic model biases or long-term trends.

2.3. Methods

To identify meteorological seasons in the current climate, to validate the capabilities of the models to reproduce them and to estimate their future evolution we adopted a two-step approach. First, we defined the baseline seasons from a set of multiple atmospheric variables using ERA5 reanalysis by applying an unsupervised temporal clustering algorithm. Second, we used a supervised classification model to extend this seasonal segmentation to the CMIP6 multi-model ensemble, both in present-day conditions and in future projections.

In the first step, Radially Constrained Clustering (RCC; Cannon, 2005) was applied to identify temporally contiguous seasons. In the second step, a Multinomial Logistic Regression Classifier (MLRC; Hastie et al., 2009) was trained to learn the association between atmospheric states and the RCC-derived seasonal labels. Once trained, the MLRC was applied to classify each day in the CMIP6 MME. This framework allowed us to assess the ability of CMIP6 models to reproduce the seasonal cycle identified in ERA5 and to explore projected changes in seasonal characteristics over the 21st century.

2.3.1. Radially Constrained Clustering

The Radially Constrained Clustering algorithm (RCC; Cannon, 2005) is used to define a number k of seasons by their breakpoints $B = \{b_1, \ldots, b_k\}$, that is the dates that mark the transition from one season to another. In this study we set $k = 4$, corresponding to the four main phases of the seasonal cycle in India: winter, pre-monsoon, monsoon, and post-monsoon. This choice was made to obtain a level of temporal granularity that is consistent with the commonly used seasonal description of the Indian climate and allows a direct comparison with existing literature and operational classifications.

To assess the sensitivity of this methodological choice, we repeated the analysis using $k = 3$ and $k = 5$ and report the results in Supporting Information S1. The comparison shows that the main seasonal organization is broadly preserved across the three configurations, while changes in k mainly affect the degree of aggregation or subdivision of specific regimes in some regions. This indicates that k primarily controls the temporal granularity of the segmentation, supporting $k = 4$ as a physically interpretable compromise for the Indian subcontinent.

Differently with other widely used clustering algorithms that partition the data in the feature space (such as k-means (Lloyd, 1982)), RCC ensures that the clusters are temporally disjoint. Furthermore, RCC treats the last

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

day of the year as continuous to the first, allowing a season to span across the boundary between December 31st and January 1st.

To run the clustering we considered, for each grid point, the four time series of tas, pr, ua850, and va850 from ERA5 for 59 years from 1961 to 2019. The temporal domain was chosen to align with the study by Pai et al. (2020), which defines the monsoon onset and withdrawal normal dates (i.e., the average onset and withdrawal date over the period 1961–2019). These dates were used as a reference for validating the results of the clustering in the present study.

We reshaped the data into a design matrix X with dimensions 365×236. Rows correspond to calendar days, while columns correspond to variable–year combinations. Since four variables (tas, pr, ua850, and va850) were considered over 59 years (1961–2019), the total number of features is $4 \times 59 = 236$. Thus, each column contains the 365-day annual cycle of one variable for one specific year. We then applied a min–max normalization independently to each feature to preserve the seasonal evolution and relative phasing of the variable throughout the year.

The clusters were then defined by the set B of the k breakpoints that minimizes the Within Clusters Sum of Squares (WCSS):

$$WCSS = \sum_{i=1}^{k} \sum_{j \in c_i} \sum_{d \in D} \left(x_{j,d} - \mu_{i,d}\right)^2 \tag{1}$$

where $x_{j,d}$ is the value of the feature $d \in D$ on day j, and $\mu_{i,d}$ is the mean value of feature d within cluster c_i.

The minimum of the WCSS is searched through an iterative algorithm:

1. Generate a set B of k breakpoints equally spaced in the year.
2. Update each breakpoint b_i in B by adding a random value drawn from $U(-\eta, \eta)$. In this work we set $\eta = 3$.
3. Compute WCSS over the design matrix X using the updated breakpoints.
4. If WCSS decreases, retain the new breakpoints, otherwise, revert to the previous ones.
5. Repeat steps 2–4 for a predefined number of iterations.

2.3.2. Multinomial Logistic Regression Classifier

To project the seasons defined on ERA5 with RCC onto each climate model from CMIP6, we employed a Multinomial Logistic Regression Classifier (MLRC; Hastie et al., 2009). For each day, the MLRC estimates the probability of belonging to each season $k \in \{1, \ldots, K\}$ using the SoftMax function:

$$p_{i,k} = \frac{\exp(\mathbf{w}_k \cdot \mathbf{x}_i)}{\sum_{j=1}^{K} \exp(\mathbf{w}_j \cdot \mathbf{x}_i)} \tag{2}$$

where $\mathbf{x}_i$ is the vector of predictors (precipitation, near-surface air temperature, and zonal and meridional wind components), $\mathbf{w}_k$ is the weight vector for class k, and $p_{i,k}$ is the estimated probability that day i belongs to season k.

The model parameters $\mathbf{w}$ are optimized by minimizing the categorical cross-entropy loss over N samples:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log\left(p_{i,k}\right) \tag{3}$$

where $y_{i,k}$ is a binary indicator equal to 1 if sample i belongs to season k, and 0 otherwise. The optimization is performed over multiple epochs using the Adam algorithm (Kingma & Ba, 2017).

We trained the MLRC on ERA5 daily data for the period 1961–2009, using the seasons identified by the RCC as categorical labels. An independent MLRC was trained at each grid point, and the resulting pointwise classifications and coefficients were subsequently aggregated over the four representative regions. As predictors, we used the same four atmospheric variables employed in the clustering step (tas, pr, ua850, va850) thus ensuring full

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

consistency between the unsupervised and supervised phases of the framework. The training data set comprised 17,885 daily samples, which were divided into batches of size 52, and the model was trained over 100 epochs.

Since the MLRC is based on a linear decision function, the learned coefficients provide a direct interpretation of the influence of each predictor on the seasonal classification (Hastie et al., 2009). To compare the relative role of the different predictors, we quantified their importance from the normalized magnitude of the class-specific coefficients:

$$I_{m,k} = \frac{|w_{m,k}|}{\sum_{m=1}^{M} |w_{m,k}|} \times 100(\%), \tag{4}$$

where $w_{m,k}$ denotes the coefficient associated with predictor m in the weight vector $\mathbf{w}_k$. By construction, the resulting importance values sum to 100% for each season and grid point, providing an interpretable measure of the relative contribution of temperature, precipitation, zonal wind, and meridional wind to the classification.

2.3.3. Validation and Skill

To assess the capability of the CMIP6 MME to reproduce the seasonal cycle and the effectiveness of our classification algorithm in identifying the correct season, we classified each day of each model for the period that overlaps ERA5 (1961–2015), and we computed the $F1$ Score (Blair, 1979) for each model and season across the four spatial boxes (Figure 1).

To provide an out-of-sample reference for classifier performance, the trained MLRCs were also applied to the held-out ERA5 period 2010–2019, and the resulting F1 scores were computed against the RCC-derived seasonal labels.

$F1$ score is the harmonic mean of precision and recall, and it is defined as:

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

where the precision measures the fraction of days classified in the season x that actually belong to season x, and the recall the fraction of days classified in the season x over the total number of days in the season x:

$$precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{6}$$

$$recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{7}$$

Values of $F1$ range from 0 to 1, with higher scores indicating better agreement between the predicted and reference seasonal labels. A value close to 1 denotes high classification accuracy, meaning that most days are correctly assigned to their corresponding season. In the present four-class problem, a random assignment would yield an expected $F1$ of approximately 0.25. Therefore, values around 0.5 indicate that the model performs roughly twice as well as random guessing, representing a threshold between poor and moderate skill. In this study, we adopt $F1 = 0.7$ as a practical threshold for good performance. From the definition in Equation 5, this value implies that neither precision nor recall can fall below 0.5 unless the other is very high (e.g., $P = 0.9$ if $R = 0.5$). Hence, setting the threshold at 0.7 ensures that both metrics are simultaneously satisfactory, balancing detection capability and false alarm rate.

3. Results and Discussion

3.1. Clustering Results

We first applied the RCC algorithm to all grid points within the study domain and then aggregated the resulting seasonal breakpoints by averaging them over the four subregions shown in Figure 1. Here, the clustering identifies four distinct seasons that closely reproduce the partitioning adopted by the India Meteorological Department

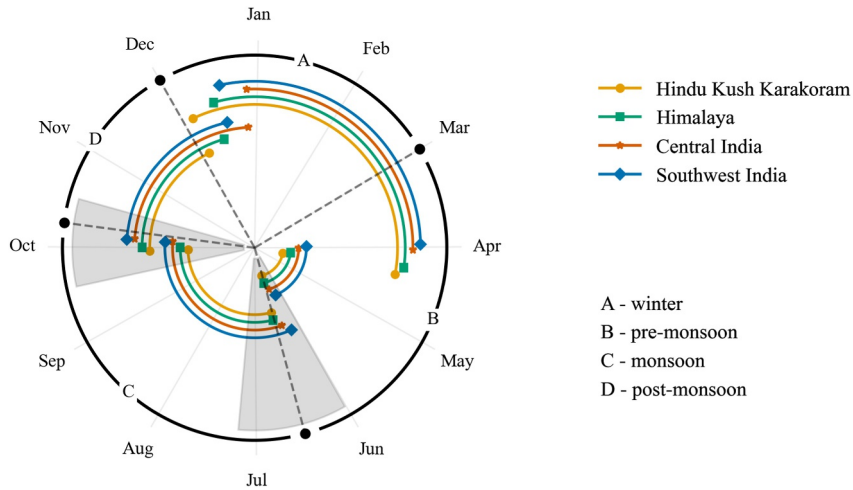


Figure 2. Seasonal coverage for the four focus regions. Each colored arc corresponds to the median duration of the four identified seasons (pre-monsoon, monsoon, post-monsoon, and winter), estimated from RCC clustering. The black outer reference ring highlights the subdivision adopted by the India Meteorological Department (IMD). The shaded gray sectors around the monsoon onset and withdrawal represent the spatial variability (i.e., minimum and maximum dates) of IMD onset and withdrawal across India. The monsoon onset and withdrawal medians are derived from (Pai et al., 2020), using the normal onset and withdrawal dates averaged across the entire Indian landmass. Month labels (e.g., "Feb") indicate the first day of each month, following a clockwise chronological order starting from January.

(IMD): monsoon (June–September), post-monsoon (October–November), winter (December–February), and pre-monsoon (March–May) (Praveen et al., 2020), as shown in Figure 2. The timing of monsoon onset and withdrawal aligns well with the IMD-defined phases of the Indian Summer Monsoon (ISM) (Pai et al., 2020), with the onset in Southwest India (blue line) occurring earlier than in other regions, in agreement with the climatological progression of the monsoon (Figure 2).

For the other three seasons, whose boundaries are not as clearly defined in the literature, the clustering still provides a coherent segmentation (Figure 2). The main discrepancy with the IMD subdivision emerges in the transition from winter to pre-monsoon, where the algorithm identifies a boundary that extends further into pre-monsoon. This behavior is more consistent with the definition of the Western Disturbance season proposed by Palazzi et al. (2013), spanning from December to April. Indeed the Hindu Kush– Karakoram (yellow line) region is the one which present a longer winter season duration (Figure 2).

To further assess the physical consistency of the detected ISM onset and decay, we compared the identified breakpoints with the official monsoon onset and withdrawal normal dates provided by the IMD for each grid point over India. This comparison is intended as an external climatological benchmark rather than as a strict validation target. Indeed, the RCC breakpoints identify transitions between multivariate seasonal regimes in ERA5, whereas the IMD normal dates are based on a different, operationally oriented methodology. Therefore, exact point-wise agreement is not expected. Instead, the purpose of this comparison is to evaluate whether the unsupervised multivariate segmentation, which is neither trained nor constrained by IMD dates, recovers the broad timing and spatial progression of the Indian Summer Monsoon as represented by the most established regional reference. The IMD onset normal dates have been established using precipitation criteria over the period 1961–2019, on a regular grid $1^\circ \times 1^\circ$ (latitude × longitude) (Pai et al., 2020). Although this methodology is precipitation-oriented, it was developed with the explicit goal of reproducing the spatial progression obtained from the IMD operational procedure, which also takes into account circulation at different atmospheric levels (Pradhan et al., 2017). For monsoon withdrawal, the normal dates were established as the average of the operationally declared dates (Pai et al., 2020). The operational withdrawal criteria are based on precipitation, together with low-tropospheric pressure, soil moisture content, and circulation patterns. Since the dates provided by the IMD cover the Indian national territory only, the comparison was limited to this area and is reported in Figure 3.

The RCC-derived monsoon onset is broadly consistent with the IMD climatology, although non-negligible regional differences are present. Positive differences occur in the southeastern part of the domain,

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

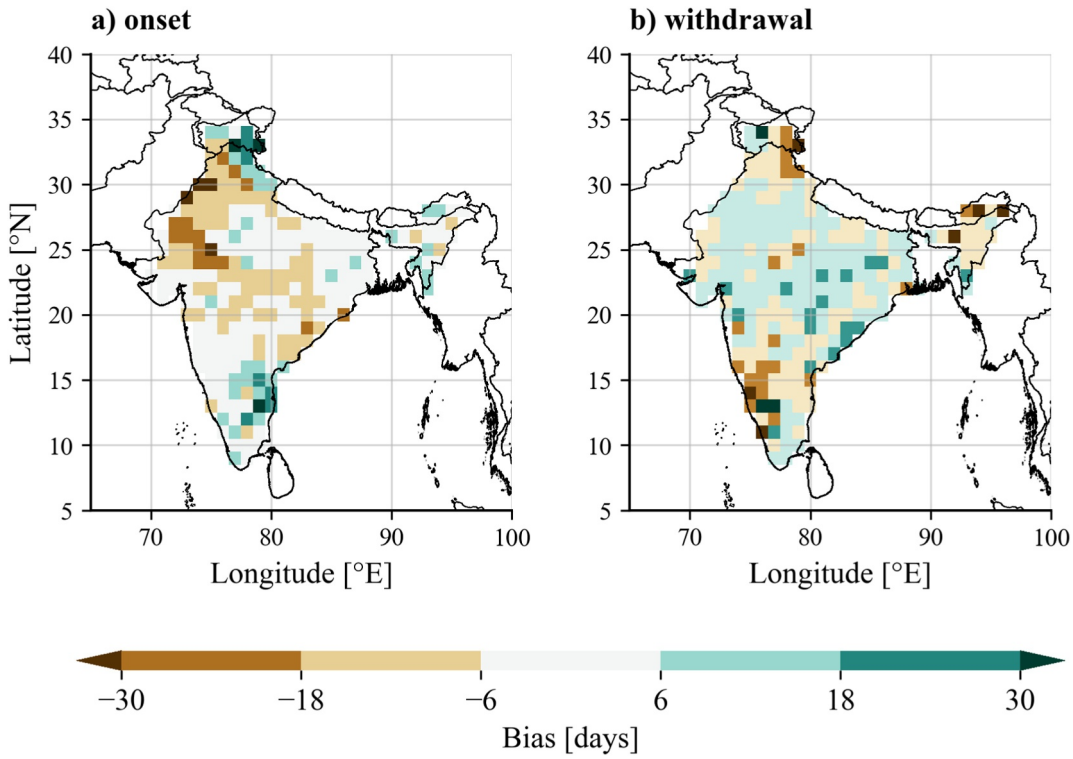


Figure 3. Differences between the ISM phases by the RCC algorithm, compared to the reference dates provided by the IMD. Panel (a) shows the difference in monsoon onset dates between the RCC output and the IMD-defined climatological onset, while Panel (b) shows the difference in withdrawal dates. Positive values indicate a later transition identified by the RCC, and negative values an earlier one. Both fields are clipped over the IMD reference domain.

corresponding to Andhra Pradesh, and in the northern part, corresponding to Ladakh, whereas negative differences are found in the northwestern part, particularly over Punjab and Rajasthan (see Figure 3a). A point-wise comparison of onset dates (Figure 4a) indicates that RCC tends to anticipate the advance of the monsoon relative to the IMD normals. The linear fit between the IMD reference dates (x-axis) and the RCC-derived dates (y-axis) yields a slope $m = 0.39$, indicating that the spatial progression detected by RCC is more compressed in time than in the IMD reference. The spatial correspondence between the two fields is moderate, with a Pearson correlation coefficient $r = 0.35$.

For monsoon withdrawal, the RCC–IMD differences are overall slightly positive, except over the southwestern part of the peninsula, where RCC identifies an earlier transition (Figure 3b). The point-wise comparison (Figure 4b) shows that RCC defines a slightly faster withdrawal progression ($m = 0.77$), largely driven by the negative differences over the southern peninsula. The spatial correspondence is higher than for onset, with $r = 0.53$. This result should nevertheless be interpreted cautiously, since the monsoon withdrawal has been less explored than the onset (Pai et al., 2020), and its assessment remains challenging because of the larger interannual variability (Syroka & Toumi, 2004).

These discrepancies are expected given the different methodologies and data employed in defining and detecting monsoon transitions. They reflect both data set dependence and the fact that the two approaches diagnose related, but not identical, quantities. IMD uses different criteria and data from multiple sources for defining the onset and withdrawal of the monsoon, while the RCC method identifies transitions between multivariate seasonal regimes using temperature, precipitation, and 850 hPa winds from ERA5 reanalysis only. Accordingly, the RCC breakpoints should be interpreted as the onset and decay of a multivariate monsoon-like regime, rather than as precipitation-only onset or withdrawal dates. Furthermore, while some studies have attempted to evaluate the reliability of ERA5 in reproducing regional precipitation patterns (Paul et al., 2024; Singh et al., 2021), the present comparison should not be read as an assessment of the ability of ERA5 to reproduce IMD transition dates, but as a benchmark of consistency for the multivariate seasonal segmentation. Thus, the comparison with IMD is retained

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

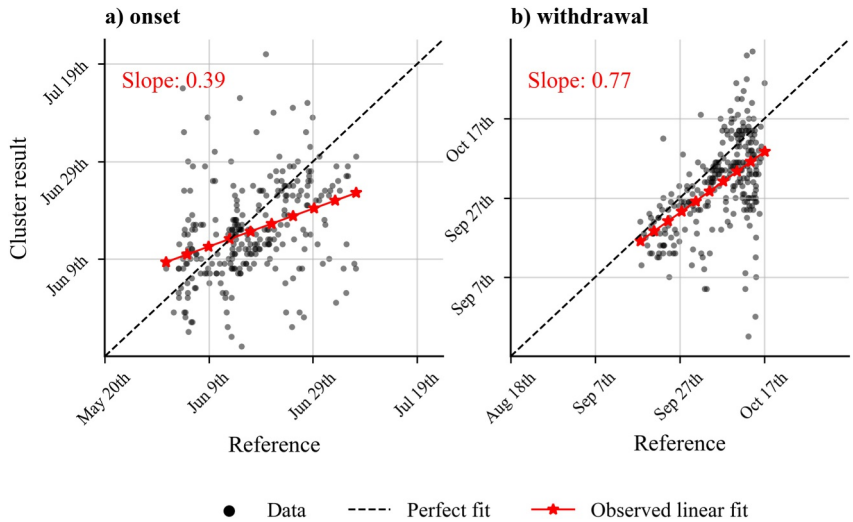


Figure 4. Scatter plots comparing the seasonal transition dates identified by the RCC algorithm with the climatological reference dates from the IMD, across all grid points over the Indian subcontinent. Panel (a) shows the comparison for the monsoon onset, while Panel (b) refers to the monsoon withdrawal. Each point represents a single grid cell, with the reference date on the *x*-axis and the corresponding RCC-based date on the *y*-axis. The black dashed line indicates perfect agreement (1:1 line), while the red dashed line shows the linear regression fit. The slope of the fit is reported in each panel: slope weaker than 1 implies that the onset/withdraw of the monsoon system is progressing in space at a faster pace than in the IMD data set.

as an external benchmark of climatological consistency, while the RCC–IMD differences highlight the sensitivity of monsoon transition dates to both the underlying data set and the adopted definition.

3.2. Classification Results

Before applying the trained MLRC to the CMIP6 ensemble, we analyzed the relative importance of the model predictors in order to provide an interpretation of the classification framework. Using the normalized magnitude of the regression coefficients we quantified the contribution of temperature, precipitation, zonal wind, and meridional wind to the identification of each season. The resulting feature importance values are summarized in Table 2. The relative importance of the MLRC predictors varies substantially across seasons and regions. Temperature is the dominant predictor during the pre-monsoon season in all four regions, accounting for 39%–56% of the total importance. Winter classification shows a stronger regional dependence, with temperature contributing most in HKK and the Himalaya (43%–44%), while precipitation becomes the leading predictor in Central and Southwest India (32%–44%). During the monsoon season, precipitation is the dominant predictor in the Himalaya and Central India (31%–49%), whereas zonal wind contributes most in Southwest India (39%). Post-monsoon classification is characterized by a larger contribution from circulation-related variables, with zonal wind emerging as the dominant predictor in Central and Southwest India (36%–40%). Overall, the results indicate that the relative importance of the predictors is season- and region-dependent, with no single variable dominating the classification across all conditions.

We then apply the MLRC to the CMIP6 data to evaluate the fidelity of the season representation in the climate models. Figure 5 shows, for each season and sub-region, the boxplot distribution of F1 scores obtained using the CMIP6 multi-model ensemble (MME) for the historical period that overlaps the ERA5 data set used in the clustering (1961–2015). For comparison, Figure 5 also reports the F1 scores obtained by applying the trained MLRCs to the held-out ERA5 period 2010–2019. This comparison helps distinguish two contributions to the CMIP6 scores: the intrinsic ability of the linear classifier to separate the RCC-derived seasonal classes, represented by its out-of-sample performance on ERA5, and the additional mismatch introduced when the classifier is applied to climate model data. Thus, CMIP6 scores lower than the ERA5 reference suggest a contribution from model biases, whereas similarly low scores in both data sets indicate that limited class separability also plays a role. In addition, we plotted the monthly distribution of days assigned to the two extreme seasons (winter and monsoon) across the multi-model ensemble, and compared this distribution with that obtained from clustering applied to ERA5 data (Figure 6).

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Relative Importance (%) of the MLRC Predictors

Season	Variable	HKK	Himalaya	Central India	Southwest India
winter	tas	**44**	**43**	19	10
	pr	19	16	**32**	**44**
	ua850	20	29	30	26
	va850	17	13	19	21
pre-monsoon	tas	**39**	**42**	**56**	**45**
	pr	15	16	18	20
	ua850	23	22	14	21
	va850	23	20	13	14
monsoon	tas	**38**	23	15	29
	pr	17	**31**	**49**	18
	ua850	23	30	18	**39**
	va850	22	17	17	13
post-monsoon	tas	**31**	15	30	29
	pr	30	**36**	17	20
	ua850	20	27	**40**	**36**
	va850	19	22	14	14

Note. The dominant predictor for each season and region is shown in bold and underlined.

The models show consistently strong performance across all seasons in Central and Southwest India. The monsoon season, in particular, stands out with high and stable F1 scores, with most models exceeding 0.8 (Central India: 0.86 ± 0.09, Southwest India: 0.91 ± 0.02), indicating a clear and well-captured seasonal signal (Figure 5c). The only exceptions are Central India during the pre-monsoon (0.74 ± 0.10, Figure 5b) and Southwest India during the post-monsoon (0.73 ± 0.07, Figure 5d), where the ensemble average drops slightly below this threshold. The held-out ERA5 score is comparable to the CMIP6 distribution for the Central India monsoon, while it is generally lower than the CMIP6 medians for the other seasons in Central and Southwest India. This may reflect smoother and more regular seasonal cycles in the model. Notably, the MME shows only a

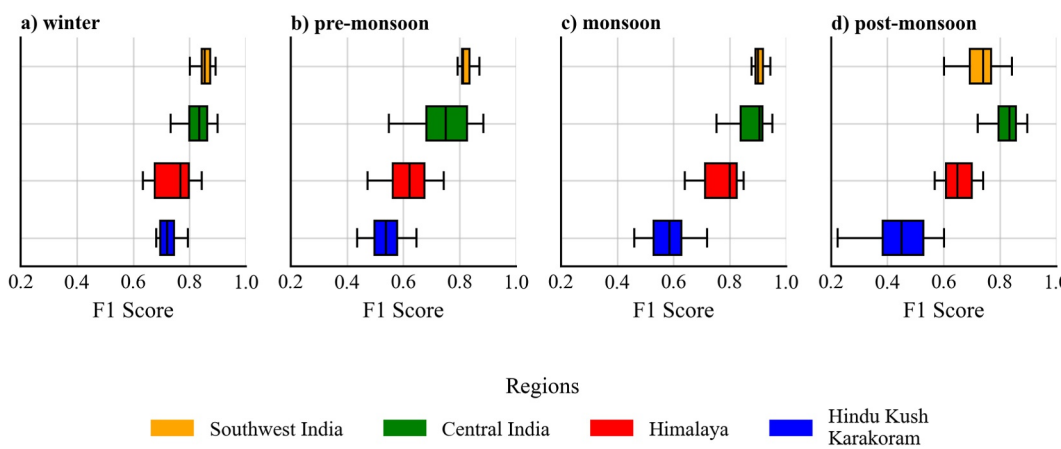


Figure 5. Boxplot distribution of F1 scores computed for each grid point, season, and CMIP6 model over the four representative regions during the historical period (1961–2015). Boxes summarize the distribution across all grid points and 13 CMIP6 models, with the interquartile range, median, and whiskers extending to 1.5 × IQR. Black crosses indicate the corresponding regional mean F1 scores obtained by applying the trained pointwise MLRCs to the held-out ERA5 period 2010–2019. These ERA5 scores provide an out-of-sample reference for interpreting the contribution of class separability and climate-model differences to the CMIP6 scores.

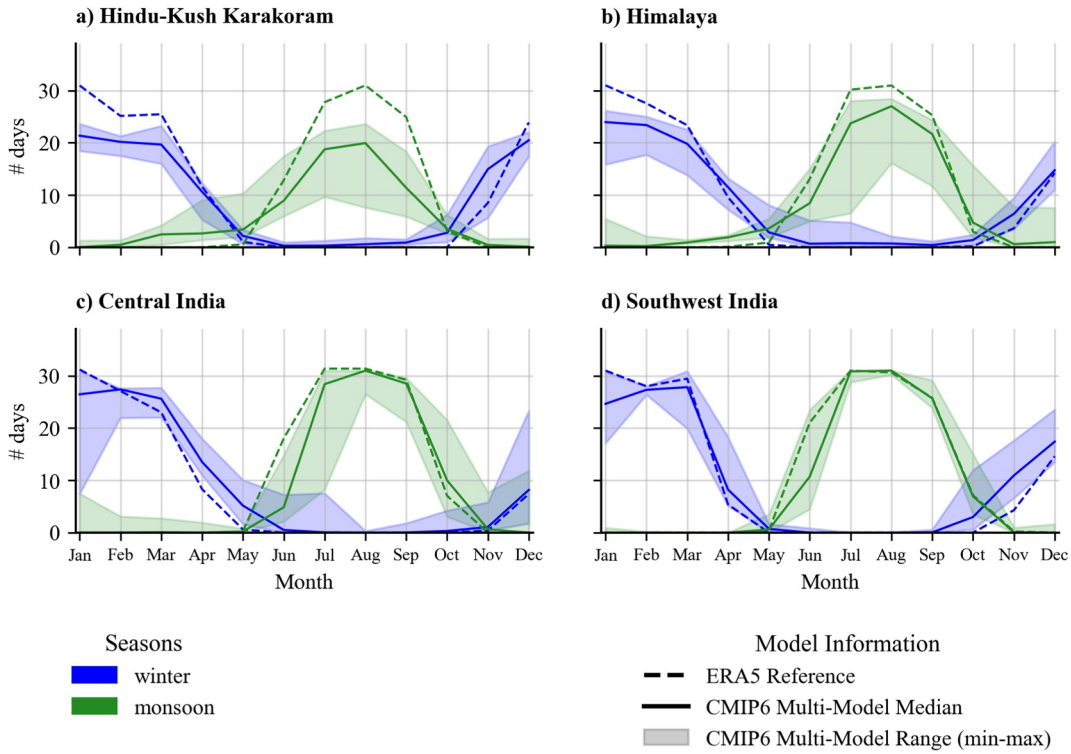


Figure 6. Monthly distribution of days classified as winter (blue) and monsoon (green) in the historical period over (a) Hindu Kush–Karakoram (b) Himalaya (c) Central India (d) Southwest India. Each panel shows the average number of days assigned to a given season, based on the classification performed on the CMIP6 MME (solid lines). The shaded bands represent the inter-model spread (min–max range) of the CMIP6 projections. The result of RCC applied to ERA5 data (dashed lines) is displayed as reference.

slight delay in monsoon onset compared to the reference, while the withdrawal timing is almost identical (Figures 6c and 6d).

In contrast, models over the Hindu Kush–Karakoram are notably less reliable. F1 scores are particularly low during the pre-monsoon (0.53 ± 0.08) and post-monsoon (0.44 ± 0.10), as shown in Figures 5b and 5d. The monsoon season also shows poor results, with an F1 score of 0.56 ± 0.09 (Figure 5c). The held-out ERA5 scores are higher than the CMIP6 distribution in all four seasons, with the largest differences during monsoon and post-monsoon, suggesting an important contribution from model biases. During pre-monsoon, however, the ERA5 score is itself comparatively low, indicating that limited linear separability also contributes to the reduced CMIP6 performance. Although the onset and withdrawal timing are in reasonable agreement with the reference, there is a significant misclassification in July and August, where an average of 10 days are not identified as monsoon days (Figure 6a). This misclassification degrades algorithm performance. Interestingly, winter performance in this region is comparatively more robust (0.72 ± 0.05) (Figure 5a) and more aligned with scores observed in other regions, suggesting that the winter regime is better represented by CMIP6 than the other seasonal regimes over this region, although its score remains below the ERA5 reference.

The Himalaya region shows intermediate performance, with F1 scores ranging between 0.6 and 0.8 depending on the season. The spread among models is particularly wide during winter (0.74 ± 0.07) (Figure 5a) and pre-monsoon (0.61 ± 0.08) (Figure 5b), pointing to larger uncertainty and less consistent seasonal behavior across models. The corresponding ERA5 scores exceed the CMIP6 medians during winter, pre-monsoon, and monsoon, suggesting an additional contribution from model biases. During post-monsoon, ERA5 and CMIP6 show comparable and only moderate performance, indicating that limited separability of this transition season may also contribute.

Overall, these results highlight the models' strong ability to capture the dominant seasonal cycle in lowland regions influenced by the Indian Summer Monsoon, while also revealing substantial challenges in regions with

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Median Number of Models Included in the Significant Multi-Model Ensemble (SMME) for Each Season and Region

Region	Winter	Pre-monsoon	Monsoon	Post-monsoon
Hindu Kush–Karakoram	8 (62%)	2 (15%)	6 (46%)	1 (8%)
Himalaya	9 (69%)	6 (46%)	10 (77%)	7 (54%)
Central India	12 (92%)	10 (77%)	12 (92%)	11 (85%)
Southwest India	13 (100%)	12 (92%)	13 (100%)	12 (92%)

Note. Percentages are relative to the maximum number of models (13).

complex orography. The held-out ERA5 scores also provide context for the fixed F1 > 0.7 threshold used to construct the SMME: where ERA5 performance is itself relatively low, reduced CMIP6 scores should be interpreted cautiously because they may partly reflect limited class separability rather than climate-model differences alone.

It is known that CMIP6 models represent well the monsoon system, despite some limited bias in timing (Khadka et al., 2022). This result has also been proved for precipitation in the mountain area of the Indian subcontinent (Gusain et al., 2020). Our results remain consistent with the known difficulties of GCMs in representing climate characteristics over the Himalayan mountains (Lalande et al., 2021), although the ERA5 comparison indicates that low F1 scores may also partly reflect limited separability of the seasonal classes.

While the historical evaluation offers a robust benchmark for model skills and classification performance, our primary objective was to extend the analysis to future projections, in order to assess the potential evolution of the seasonal cycle in the next decades. To this end, we focused in the main text on changes in the lengths of the winter and monsoon seasons over the 21st century, while the corresponding time series for the pre- and post-monsoon seasons are reported in Supporting Information S1 (Figure S2 in Supporting Information S1).

The transition season anomalies are generally smaller than those found for winter and monsoon. The SMME means indicate modest increases in both transition seasons over the Himalaya and Central India and decreases over Southwest India, while no coherent change emerges over Hindu Kush–Karakoram. Because the four seasonal classes partition the 365-day calendar, their durations sum to 365 days for every model, grid point, and year. The balance is therefore exact before ensemble aggregation; small residuals in the regional SMME means arise because model selection is performed independently for each season and grid point. To account for regional and seasonal variability in model performance, we selected, for each grid point and season, only those models with acceptable classification skill in the historical period, defined by an F1 score greater than 0.7. We refer to this subset as the Significant Multi-Model Ensemble (SMME). The SMME varies across grid points and seasons, with the median number of contributing models summarized in Table 3, showing higher model consistency in Central and Southern India and lower coverage in mountainous regions such as the Himalaya and Hindu Kush—Karakoram.

We focused the analysis on three complementary aspects: the spatial distribution of long-term changes in seasonal duration (Figure 7), the inter-annual evolution of these changes within the four representative spatial boxes (Figure 8), and the intra-annual timing of such variations (Figures 9 and 10). Figures 7 and 8 describe where and when the overall duration of the monsoon and winter seasons is projected to change, but they do not indicate which parts of each season contribute to these variations. To address this aspect, Figures 9 and 10 present a monthly breakdown of the projected changes, showing how the number of days classified within each month evolves across future periods. This monthly perspective highlights the internal structure of the seasonal shifts and helps identify asymmetric extensions or contractions within the annual cycle, clarifying whether the projected changes are concentrated in the early or late phases of the monsoon and winter seasons.

In the monsoon season, most of the subcontinent shows a significant lengthening of the season, especially across Central, Western, and Southern India, with trends exceeding +0.5 days per decade in several regions (Figure 7a). Northern areas and parts of the northeastern region display weaker or slightly negative trends, often without statistical significance.

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

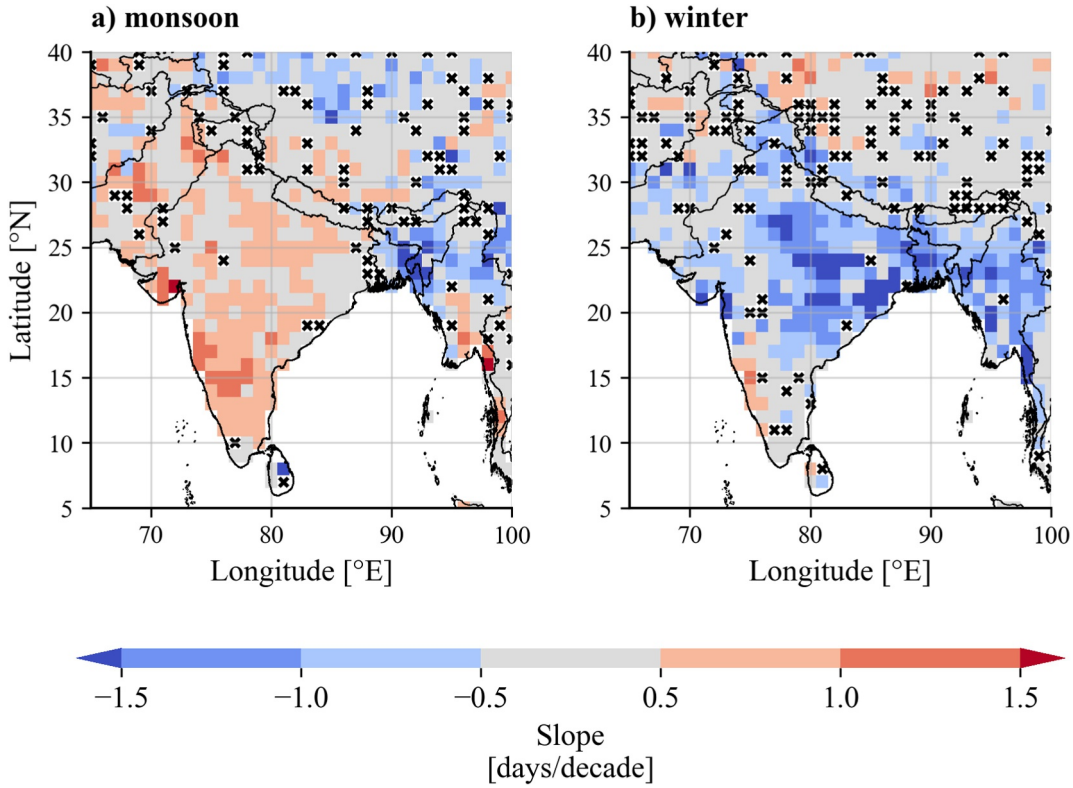


Figure 7. Spatial distribution of linear trends in the number of days classified as part of the monsoon season (a) and winter season (b) over the Indian subcontinent for the period 2020–2100. Trends are expressed in days per decade and computed over the SMME mean. Only trends statistically significant at the 95% confidence level (Mann–Kendall test (Gilbert, 1987), $p < 0.05$) are shown. Grid points with non-significant trends are marked with black crosses.

Central India shows the most prominent monsoon lengthening, with an increase of nearly 5 days projected by the end of the century (Figure 8c). This increase is not uniform throughout the season, but it is largely driven by a delay in the monsoon withdrawal (Figure 9c). The onset timing, in contrast, shows only minor changes, confirming the asymmetry of the seasonal extension. A similar pattern is observed in Southwest India, where an average increase of 5 monsoon days is expected by 2100 (Figure 8d). Also in this case, the extension is mostly attributable to a delayed withdrawal, suggesting a persistent shift toward wetter conditions in the post-monsoon period (Figure 9d). In the Himalayan region, the models project little or no change in the total length of the monsoon season, as shown in Figure 8b. However, the internal structure of the season appears to be evolving, with a redistribution of approximately 2 days from the early to the late phase (Figure 9b). In contrast, Hindu Kush–Karakoram does not exhibit any consistent trend (Figure 8a). The region shows a mix of weak positive and negative signals, with low inter-model agreement, highlighting the difficulty that current GCMs face in representing monsoon processes in an area characterized by complex topography and scarcity of observations (Lalande et al., 2021). In many grid points over the HKK region, the lack of significant trends coincides with low skill scores during the historical period, which further reduces confidence in the projections.

These regional patterns are broadly consistent with the CMIP6-based picture of a wetter Indian monsoon under global warming. It has been assessed that CMIP6 models project an increase in monsoon precipitation (B. Wang et al., 2020) and in the number of wet days (Norgate et al., 2024). Recent CMIP6 analyses further project increases in JJAS precipitation and precipitation extremes over India, although with substantial model biases and spatial heterogeneity (Norgate et al., 2024). A post-AR6 assessment similarly reports a projected increase in India southwest monsoon rainfall, including intensification of heavy precipitation, while emphasizing strong regional variability and persistent inter-model spread (Dhara et al., 2025). Furthermore, the delay of the ISM withdrawal in CMIP6 projections has already been observed in previous studies (Ha et al., 2020) using a threshold method applied to a smoothed precipitation time series (B. Wang & LinHo, 2002). The present study confirms the results

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

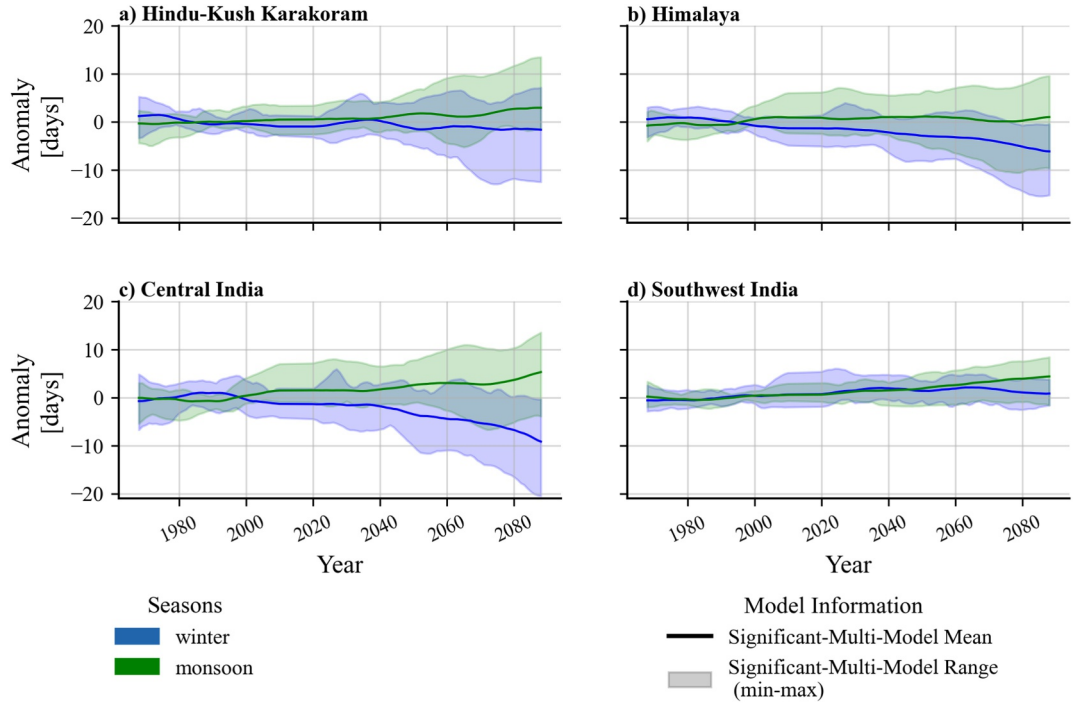


Figure 8. Time series of projected anomalies in the number of days classified as part of the winter (blue) and monsoon (green) seasons over four representative spatial boxes of the Indian subcontinent. Anomalies are computed relative to the historical mean seasonal length (1970–2020) and smoothed using a 15-year centered rolling average. Solid lines show the mean across all models in the SMME; shaded bands show the inter-model spread (min–max range).

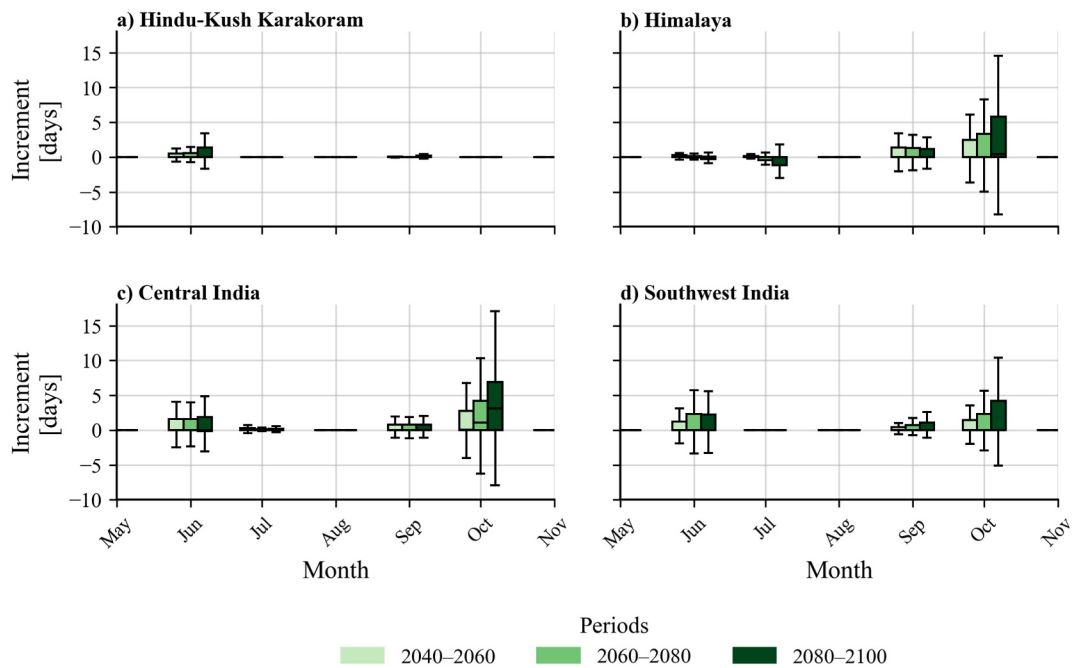


Figure 9. Projected monthly change in the number of days classified as part of the Monsoon season for three future periods (2040–2060, 2060–2080, 2080–2100), relative to the historical climatology (1970–2000). Each panel refers to a specific subregion. Box-plots represent the distribution across grid points and models. Boxes show the interquartile range (IQR), horizontal lines indicate the median, and whiskers extend to 1.5× IQR.

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

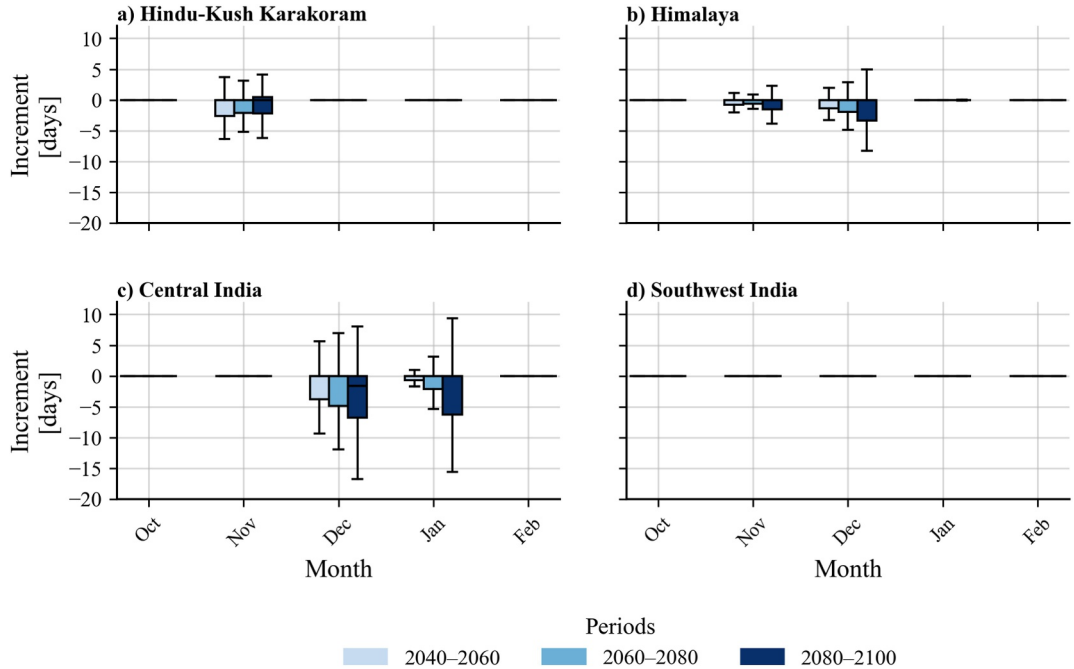


Figure 10. Projected monthly change in the number of days classified as part of the Winter season for three future periods (2040–2060, 2060–2080, 2080–2100), relative to the historical climatology (1970–2000). Each panel refers to a specific subregion. Box-plots represent the distribution across grid points and models. Boxes show the interquartile range (IQR), horizontal lines indicate the median, and whiskers extend to 1.5× IQR.

on this delayed withdrawal signal by a classification-based approach relying on multiple climate variables. At the same time, ISM duration changes should be interpreted as metric-dependent. Using a large-scale definition based on the reversal of the meridional gradient of upper-tropospheric temperature, and applying observational constraints to CMIP6 projections, Cheng et al. (2025) found a shortening of ISM duration under a high-emission scenario, associated with delayed onset and earlier withdrawal. This apparent contrast may reflect the different components and spatial scales of the monsoon diagnosed by the two approaches. The upper-tropospheric temperature gradient provides a thermodynamic proxy for the establishment and decay of the large-scale monsoon circulation, whereas the grid-point classification used here diagnoses the persistence of a regional multivariate hydroclimatic regime. This distinction is particularly relevant for Central India, where precipitation is the leading predictor of the monsoon class, accounting for 49% of its relative feature importance (Table 2). Under warming, these hydrological and dynamical expressions need not evolve synchronously. Model projections indicate that precipitation can increase even when upper-tropospheric circulation weakens, owing to increased moisture availability and changes in lower-tropospheric circulation (Ma & Yu, 2014). Thus, the later persistence of locally defined monsoon-like conditions need not imply a later decay of the large-scale monsoon circulation. This interpretation remains tentative, however, because Cheng et al. (2025) also used observational constraints, a different model ensemble, and a different spatial aggregation. Applying both diagnostics to the same simulations would be required to isolate the contribution of the adopted definition.

For the winter season, a shortening is observed across most of the Indian subcontinent, particularly in Central and Southern India, where trends reach up to −1.5 days per decade, and in the Western part of the Bay of Bengal (Figure 7b). This contraction is statistically significant over a large portion of the domain. Mountain areas exhibit the same trend, but with many grid points showing non-significant trends.

Regarding the spatial averages, Central India exhibits the most pronounced winter shortening, which accelerates after 2040. By the end of the century, the region is projected to lose up to 10 winter days (Figure 8c). This reduction is spread across the entire winter season but is particularly concentrated in December–January, where approximately 4 winter days are expected to be lost (Figure 10c). A similar pattern emerges in the Himalaya region, where the total reduction in winter duration reaches about 5 days by the end of the century (Figure 8b). As in Central India, the loss is most evident in December–January, although with a lower magnitude of around 2 days

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Figure 10b). In contrast, Hindu Kush–Karakoram (HKK) and Southwest India do not exhibit any clear trend in the duration of the winter season (Figures 8a and 8d). In the case of HKK, this absence of trend is associated with a limited number of significant grid points, reflecting a lack of coherence among climate models in reproducing the seasonal cycle in this region. Conversely, in Southwest India, most grid points show statistically significant trends, but with opposing signs, resulting in a spatially heterogeneous pattern and no consistent regional signal.

Projected warming over India could in principle contribute to the apparent loss of winter days, by making winter conditions less distinguishable from the rest of the annual cycle. However, our framework explicitly removes long-term drifts in temperature and other meteorological variables, and therefore does not directly detect the mean warming trend itself. Instead, it focuses on changes in the intra-annual distribution of the multivariate climate state. In this respect, our results are more closely related to studies showing that warming over India is projected to be more pronounced in winter than during the monsoon season, whereas precipitation and wind speed are expected to intensify mainly in the monsoon season (Almazroui et al., 2020; Kumar et al., 2023; Salunke et al., 2023; Zakari et al., 2022).

The feature importance analysis supports this interpretation. In Central India, where the winter shortening is strongest, the winter class is not primarily defined by temperature, but by precipitation and low-level zonal circulation (Table 2). This suggests that the projected reduction in winter days is not simply the direct expression of background warming, but rather reflects a change in the relative position of winter within the annual cycle of precipitation and circulation. In other words, winter-like conditions become less frequent as the seasonal cycle shifts toward states that are dynamically and hydrologically less distinct from the surrounding transition and monsoonal regimes.

The regional differences in feature importance also help explain why the winter-duration signal is not spatially uniform. In the Himalaya and HKK regions, winter classification is dominated by temperature, which accounts for 43% and 44% of the winter feature importance, respectively. However, despite the strong projected warming over the Himalayan–Karakoram sector reported in previous CMIP6 studies (Almazroui et al., 2020; Kumar et al., 2023), this does not translate into a coherent winter shortening over HKK in our results.

4. Conclusions

In this study, we developed a machine learning framework for the objective identification and projection of meteorological seasons, combining Radially Constrained Clustering with a supervised classification scheme. The primary contribution of this work is methodological: the proposed framework defines multivariate seasonal regimes from reanalysis data and consistently extends these definitions to climate model simulations. We used the Indian subcontinent as a climatically heterogeneous and well-documented proof of concept region. Applied to this domain, the method successfully captured the main phases of the seasonal cycle, including the onset and withdrawal of the Indian Summer Monsoon and the winter season associated with Western Disturbances.

The comparison with IMD reference dates indicates that the RCC-derived transitions are broadly consistent with the climatological timing and spatial progression of the Indian Summer Monsoon. This comparison should be interpreted as an external benchmark of climatological consistency rather than as a strict validation target, because RCC identifies transitions between multivariate seasonal regimes whereas the IMD dates rely on a different operational methodology. Accordingly, the framework should be regarded as complementary to, rather than a replacement for, operational onset and withdrawal algorithms. By applying the same learned seasonal definitions to reanalysis and climate-model simulations, it places changes in monsoon timing within the broader reorganization of the annual cycle, including winter and the transition seasons, without relying on prescribed variable-specific thresholds. Larger differences with respect to the IMD seasonal subdivision are found for the winter season, likely reflecting the more heuristic nature of this definition, which lacks a standardized criterion for its onset. In addition, the greater variability in the timing of winter onset identified by our method is consistent with the intrinsically irregular character of Western Disturbances that modulate wintertime circulation over northern India.

Evaluation against CMIP6 simulations revealed that the ability of models to reproduce the observed seasonal cycle strongly depends on the region and season. While performance is robust in Central and Southern India, particularly during the monsoon season, significant limitations remain in mountainous regions such as the Himalaya and Hindu Kush—Karakoram, where model skill is lower and inter-model spread is larger. These

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

results indicate that the proposed framework can provide a complementary diagnostic of climate-model performance, based on the ability of models to reproduce multivariate seasonal regimes rather than individual variables alone.

Future projections indicate a general tendency toward monsoon-season lengthening and winter-season shortening across large parts of the Indian subcontinent, although the magnitude and robustness of these signals vary regionally. The monsoon lengthening is most evident in Central and Southwest India and is primarily associated with delayed withdrawal, while changes are weaker and less robust in the Himalayan and Hindu Kush—Karakoram regions. The projected shortening of winter is strongest in Central India and is concentrated mainly in December–January. Because the framework removes long-term mean trends before classification, this signal should not be interpreted as the direct expression of background warming alone, but rather as a change in the intra-annual organization of the multivariate climate state. The feature importance analysis supports this interpretation, showing that the variables controlling seasonal classification differ across regions and seasons.

More broadly, our findings demonstrate the value of data-driven approaches to seasonality: they reduce the dependence on fixed calendar definitions and variable-specific thresholds, integrate multiple atmospheric variables, and provide a framework that is both objective and reproducible. However, the method also presents some limitations. Because the clustering operates purely on statistical coherence without explicit physical constraints, it may in principle identify seasons that are unusually short or that do not correspond to physically meaningful regimes, especially in regions with weak or noisy seasonal cycles. In this study, the high spatial consistency found across India provides confidence that the detected seasons reflect robust climatic structures rather than spurious classifications. Nevertheless, when applied to areas with less pronounced annual variability, additional checks may be required to ensure physical interpretability. Furthermore, the number of seasonal regimes must be specified a priori.

By projecting such definitions into the future, this framework allows for a consistent assessment of how seasonal patterns may evolve in a changing climate. The methodology is general in its algorithmic formulation and is therefore potentially applicable to other regions, data sets, and climate scenarios. However, the present study demonstrates its performance only over the Indian subcontinent, which serves as a climatically heterogeneous proof of concept domain. Future applications should explicitly test the framework across additional climatic regimes to assess its broader transferability.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

ERA5 reanalysis data on pressure levels were obtained from the Copernicus Climate Data Store (Copernicus Climate Change Service, 2018). CMIP6 simulations were accessed through the Earth System Grid Federation (Earth System Grid Federation, 2025). The following link contains the specific query to retrieve the data from ESGF used in this study: https://metagrid.esgf-west.org/search/cmip6/?project=CMIP6&activeFacets=%7B%22experiment_id%22%3A%5B%22ssp585%22%2C%22historical%22%5D%2C%22source_id%22%3A%5B%22ACCESS-CM2%22%2C%22BCC-CSM2-MR%22%2C%22CanESM5%22%2C%22CESM2-WACCM%22%2C%22CNRM-CM6-1%22%2C%22EC-Earth3%22%2C%22FGOALS-g3%22%2C%22GFDL-CM4%22%2C%22IPSL-CM6A-LR%22%2C%22MIROC6%22%2C%22MPI-ESM1-2-HR%22%2C%22MRI-ESM2-0%22%2C%22UKESM1-0-LL%22%5D%2C%22frequency%22%3A%22day%22%2C%22variable_id%22%3A%5B%22tas%22%2C%22pr%22%2C%22ua%22%2C%22va%22%5D%7D. All scripts and routines used in this study are openly available at Zenodo (Grassi, 2025).

Acknowledgments
Jacopo Grassi acknowledges funding for his PhD scholarship from the NODES project (Nord Ovest Digitale E Sostenibile), Spoke 4—Montagna Digitale e Sostenibile (Digital and Sustainable Mountain). The NODES project received funding from the Italian Ministry of University and Research (MUR) under the National Recovery and Resilience Plan (PNRR), Mission 4, Component 2, Investment 1.5, funded by the European Union—NextGenerationEU (Grant Agreement No. ECS00000036). Open access publishing facilitated by Politecnico di Torino, as part of the Wiley - CRUI-CARE agreement.

References

Almazroui, M., Saeed, S., Saeed, F., Islam, M. N., & Ismail, M. (2020). Projections of precipitation and temperature over the South Asian countries in CMIP6. *Earth Systems and Environment*, *4*(2), 297–320. https://doi.org/10.1007/s41748-020-00157-7

Amrith, S. S. (2018). Risk and the South Asian monsoon. *Climatic Change*, *151*(1), 17–28. https://doi.org/10.1007/s10584-016-1629-x

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., et al. (2020). Configuration and spin-up of ACCESS-CM2, the new generation Australian community climate and Earth system simulator coupled model. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 225–251. https://doi.org/10.1071/ES19040

Blair, D. C. (1979). Information retrieval, 2nd ed. C.J. Van rijsbergen. London: Butterworths, 208 pp. Price: $32.50. *Journal of the American Society for Information Science*, *30*(6), 374–375. https://doi.org/10.1002/asi.4630300621.30

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Bremer, S., & Wardekker, A. (2024). *Changing seasonality: How communities are revising their seasons*. De Gruyter. https://doi.org/10.1515/9783111245591

Callaghan, J., & Bonell, M. (2005). An overview of the meteorology and climatology of the humid tropics. In L. A. Bruijnzeel & M. Bonell (Eds.), *Forests, water and people in the humid tropics: Past, present and future hydrological research for integrated land and water management* (pp. 158–193). Cambridge University Press. https://doi.org/10.1017/CBO9780511535666.016

Cannon, A. J. (2005). Defining climatological seasons using radially constrained clustering. *Geophysical Research Letters*, *32*(14). https://doi.org/10.1029/2005GL023410

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the Western United States. *Bulletin of the American Meteorological Society*, *82*(3), 399–416. https://doi.org/10.1175/1520-0477(2001)082⟨0399:CITOOS⟩2.3.CO;2

Chattopadhyay, A., Hassanzadeh, P., & Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports*, *10*(1), 1317. https://doi.org/10.1038/s41598-020-57897-9

Cheng, Y., Wang, L., Chen, X., Zhou, T., & Turner, A. (2025). A shorter duration of the Indian summer monsoon in constrained projections. *Geophysical Research Letters*, *52*(1), e2024GL112848. https://doi.org/10.1029/2024GL112848

Chidean, M. I., Caamaño, A. J., Casanova-Mateo, C., Ramiro-Bargueño, J., & Salcedo-Sanz, S. (2020). Spatio-temporal climate regionalization using a self-organized clustering approach. *Theoretical and Applied Climatology*, *140*(3–4), 927–949. https://doi.org/10.1007/s00704-019-03082-6

Copernicus Climate Change Service. (2018). ERA5 hourly data on pressure levels from 1940 to present [Dataset]. *Copernicus Climate Data Store*. https://doi.org/10.24381/cds.bd0915c6

Danabasoglu, G. (2019). *NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical*. Earth System Grid Federation. https://doi.org/10.22033/esgf/cmip6.10071

de la Franca, N. L., Sánchez, E., & Domínguez, M. (2013). Changes in the onset and length of seasons from an ensemble of regional climate models over Spain for future climate conditions. *Theoretical and Applied Climatology*, *114*(3), 635–642. https://doi.org/10.1007/s00704-013-0868-2

Desamsetti, S., Dasari, H. P., Langodan, S., Viswanadhapalli, Y., Attada, R., Luong, T. M., et al. (2022). Enhanced simulation of the Indian summer monsoon rainfall using regional climate modeling and continuous data assimilation. *Frontiers in Climate*, *4*, 817076. https://doi.org/10.3389/fclim.2022.817076

Dhara, C., Deshpande, A., Roxy, M. K., Dalpadado, P., & Shrestha, M. S. (2025). A post-AR6 update on observed and projected climate change in India. *PLOS Climate*, *4*(11), e0000724. https://doi.org/10.1371/journal.pclm.0000724

Dimri, A. P., Niyogi, D., Barros, A. P., Ridley, J., Mohanty, U. C., Yasunari, T., & Sikka, D. R. (2015). Western disturbances: A review. *Reviews of Geophysics*, *53*(2), 225–246. https://doi.org/10.1002/2014RG000460

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., et al. (2022). The EC-Earth3 Earth system model for the coupled model intercomparison project 6. *Geoscientific Model Development*, *15*(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

Earth System Grid Federation. (2025). CMIP6 data search portal [Dataset]. *Earth System Grid Federation*. Retrieved from https://esgf-node.llnl.gov/search/cmip6/

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Gadgil, S. (2003). The Indian monsoon and its variability. *Annual Review of Earth and Planetary Sciences*, *31*(1), 429–467. https://doi.org/10.1146/annurev.earth.31.100901.141251

Gilbert, R. O. (1987). *Statistical methods for environmental pollution monitoring*. John Wiley & Sons. (Google-Books-ID: lEo1rvDGUEkC).

Grassi, J. (2025). JGrassi97/XSeas: First stable (v1.0.0). *Zenodo*. [Software]. https://doi.org/10.5281/zenodo.17632210

Guo, J., Hu, S., & Guan, Y. (2022). Regime shifts of the wet and dry seasons in the tropics under global warming. *Environmental Research Letters*, *17*(10), 104028. https://doi.org/10.1088/1748-9326/ac9328

Gusain, A., Ghosh, S., & Karmakar, S. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, *232*, 104680. https://doi.org/10.1016/j.atmosres.2019.104680

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., et al. (2019). Max planck institute Earth system model (MPI-ESM1.2) for the high-resolution model intercomparison Project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. https://doi.org/10.5194/gmd-12-3241-2019

Ha, K.-J., Moon, S., Timmermann, A., & Kim, D. (2020). Future changes of summer monsoon characteristics and evaporative demand over Asia in CMIP6 simulations. *Geophysical Research Letters*, *47*(8), e2020GL087492. https://doi.org/10.1029/2020GL087492

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer-Verlag.

Hekmatzadeh, A. A., Kaboli, S., & Torabi Haghighi, A. (2020). New indices for assessing changes in seasons and in timing characteristics of air temperature. *Theoretical and Applied Climatology*, *140*(3), 1247–1261. https://doi.org/10.1007/s00704-020-03156-w

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of gfdl's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019MS001829

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hunt, K. M. R., Baudouin, J.-P., Turner, A. G., Dimri, A. P., Jeelani, G., Pooja, et al. (2025). Western disturbances and climate variability: A review of recent developments. *Weather and Climate Dynamics*, *6*(1), 43–112. https://doi.org/10.5194/wcd-6-43-2025

Khadka, D., Babel, M. S., Abatan, A. A., & Collins, M. (2022). An evaluation of CMIP5 and CMIP6 climate models in simulating summer rainfall in the Southeast Asian monsoon domain. *International Journal of Climatology*, *42*(2), 1181–1202. https://doi.org/10.1002/joc.7296

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv*. arXiv:1412.6980 [cs]). https://doi.org/10.48550/arXiv.1412.6980

Kumar, T. V. L., Vinodhkumar, B., Rao, K. K., Chowdary, J. S., Osuri, K. K., & Desamsetti, S. (2023). Insights from the bias-corrected simulations of CMIP6 in India's future climate. *Global and Planetary Change*, *226*, 104137. https://doi.org/10.1016/j.gloplacha.2023.104137

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kutta, E., & Hubbart, J. A. (2016). Reconsidering meteorological seasons in a changing climate. *Climatic Change*, *137*(3), 511–524. https://doi.org/10.1007/s10584-016-1704-3

Lalande, M., Ménégoz, M., Krinner, G., Naegeli, K., & Wunderle, S. (2021). Climate change in the high Mountain Asia in CMIP6. *Earth System Dynamics*, *12*(4), 1061–1098. https://doi.org/10.5194/esd-12-1061-2021

Lee, S. H., Tippett, M. K., & Polvani, L. M. (2023). A new year-round weather regime classification for North America. *Journal of Climate*, *36*(20), 7091–7108. https://doi.org/10.1175/JCLI-D-23-0214.1

Li, L., Yu, Y., Tang, Y., Lin, P., Xie, J., Song, M., et al. (2020). The flexible global ocean-atmosphere-land system model grid-point version 3 (FGOALS-g3): Description and evaluation. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002012. https://doi.org/10.1029/2019MS002012

Littleboy, C., Subke, J.-A., Bunnefeld, N., & Jones, I. L. (2024). WorldSeasons: A seasonal classification system interpolating biome classifications within the year for better temporal aggregation in climate science. *Scientific Data*, *11*(1), 927. https://doi.org/10.1038/s41597-024-03732-z

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Lüthi, S., Fairless, C., Fischer, E. M., Scovronick, N., Ben, A., Coelho, M. D. S. Z. S., et al. (2023). Rapid increase in the risk of heat-related mortality. *Nature Communications*, *14*(1), 4894. https://doi.org/10.1038/s41467-023-40599-x

Ma, J., & Yu, J.-Y. (2014). Paradox in South Asian summer monsoon circulation change: Lower tropospheric strengthening and upper tropospheric weakening. *Geophysical Research Letters*, *41*(8), 2934–2940. https://doi.org/10.1002/2014GL059891

Minoli, S., Jägermeyr, J., Asseng, S., Urfels, A., & Müller, C. (2022). Global crop yields can be lifted by timely adaptation of growing periods to climate change. *Nature Communications*, *13*(1), 7079. https://doi.org/10.1038/s41467-022-34411-5

Morisette, J. T., Richardson, A. D., Knapp, A. K., Fisher, J. I., Graham, E. A., Abatzoglou, J., et al. (2009). Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century. *Frontiers in Ecology and the Environment*, *7*(5), 253–260. https://doi.org/10.1890/070217

Norgate, M., Tiwari, P. R., Das, S., & Kumar, D. (2024). On the present and future changes in Indian summer monsoon precipitation characteristics under different SSP scenarios from CMIP6 models. *Climate Dynamics*, *62*(12), 10445–10464. https://doi.org/10.1007/s00382-024-07389-7

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Orlove, B. (2003). How people name seasons. In *Weather, climate, culture* (1st ed., pp. 121–140). Berg. https://doi.org/10.4324/9781003103264-10

Pai, D. S., Arti, B., Sunitha, D., Madhuri, M., R, B. M., P, K. A., et al. (2020). Normal dates of onset/progress and withdrawal of southwest monsoon over India. *Mausam*, *71*(4), 553–570. https://doi.org/10.54302/mausam.v71i4.33

Palazzi, E., von Hardenberg, J., & Provenzale, A. (2013). Precipitation in the Hindu-Kush Karakoram Himalaya: Observations and future scenarios. *Journal of Geophysical Research: Atmospheres*, *118*(1), 85–100. https://doi.org/10.1029/2012JD018697

Paul, S., Sharma, P. J., & Teegavarapu, R. S. V. (2024). Spatial assessment of the reproducibility of Indian summer monsoon rainfall regimes in multiple gridded rainfall products. *Scientific Reports*, *14*(1), 29269. https://doi.org/10.1038/s41598-024-75320-5

Pawar, S. D., Lal, D. M., & Murugavel, P. (2012). Lightning characteristics over central India during Indian summer monsoon. *Atmospheric Research*, *106*, 44–49. https://doi.org/10.1016/j.atmosres.2011.11.007

Pisharoty, P. R., & Desai, B. N. (1956). Western disturbances and Indian weather. *Mausam*. Retrieved 2024-11-18, from https://mausamjournal.imd.gov.in/index.php/MAUSAM/article/view/4555

Pradhan, M., Rao, A. S., Srivastava, A., Dakate, A., Salunke, K., & Shameera, K. S. (2017). Prediction of Indian summer-monsoon onset variability: A season in advance. *Scientific Reports*, *7*(1), 14229. https://doi.org/10.1038/s41598-017-12594-y

Praveen, B., Talukdar, S., Shahfahad, Mahato, S., Mondal, J., Sharma, P., et al. (2020). Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. *Scientific Reports*, *10*(1), 10342. https://doi.org/10.1038/s41598-020-67228-7

Ren, Y., Qiu, J., Zeng, Z., Liu, X., Sitch, S., Pilegaard, K., et al. (2024). Earlier spring greening in Northern Hemisphere terrestrial biomes enhanced net ecosystem productivity in summer. *Communications Earth & Environment*, *5*(1), 122. https://doi.org/10.1038/s43247-024-01270-5

Ruan, W., Liang, Y., Sun, Z., & An, X. (2025). Climate warming and influenza dynamics: The modulating effects of seasonal temperature increases on epidemic patterns. *npj Climate and Atmospheric Science*, *8*(1), 72. https://doi.org/10.1038/s41612-025-00968-3

Salunke, P., Keshri, N. P., Mishra, S. K., & Dash, S. K. (2023). Future projections of seasonal temperature and precipitation for India. *Frontiers in Climate*, *5*, 1069994. https://doi.org/10.3389/fclim.2023.1069994

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the U.K. Earth system model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739

Serreze, M. C., & Barry, R. G. (2014). *The arctic climate system* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139583817

Singh, T., Saha, U., Prasad, V. S., & Gupta, M. D. (2021). Assessment of newly-developed high resolution reanalyses (IMDAA, NGFS and ERA5) against rainfall observations for Indian region. *Atmospheric Research*, *259*, 105679. https://doi.org/10.1016/j.atmosres.2021.105679

Sparks, T. H., & Menzel, A. (2002). Observed changes in seasons: An overview. *International Journal of Climatology*, *22*(14), 1715–1725. https://doi.org/10.1002/joc.821

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth system model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Syroka, J., & Toumi, R. (2004). On the withdrawal of the Indian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, *130*(598), 989–1008. https://doi.org/10.1256/qj.03.36

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Tran, B.-L., Tseng, W.-C., & Chen, C.-C. (2025). Climate change impacts on crop yields across temperature rise thresholds and climate zones. *Scientific Reports*, *15*(1), 23424. https://doi.org/10.1038/s41598-025-07405-8

Trenberth, K. E. (1983). What are the seasons? *Bulletin of the American Meteorological Society*, *64*(11), 1276–1282. https://doi.org/10.1175/1520-0477(1983)064⟨1276:WATS⟩2.0.CO;2

Uphus, L., Uhler, J., Tobisch, C., Rojas-Botero, S., Lüpke, M., Benjamin, C., et al. (2023). Earlier and more uniform spring green-up linked to lower insect richness and biomass in temperate forests. *Communications Biology*, *6*(1), 1052. https://doi.org/10.1038/s42003-023-05422-9

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. https://doi.org/10.1029/2019MS001683

Wang, B., Jin, C., & Liu, J. (2020). Understanding future change of global monsoons projected by CMIP6 models. *Journal of Climate*, *33*(15), 6471–6489. https://doi.org/10.1175/JCLI-D-19-0993.1

Wang, B., & LinHo (2002). Rainy season of the Asian–Pacific summer monsoon. *Journal of Climate*, *15*(4), 386–398. https://doi.org/10.1175/1520-0442(2002)015⟨0386:RSOTAP⟩2.0.CO;2

Wang, J., Guan, Y., Wu, L., Guan, X., Cai, W., Huang, J., et al. (2021). Changing lengths of the four seasons by global warming. *Geophysical Research Letters*, *48*(6), e2020GL091753. https://doi.org/10.1029/2020GL091753

Woolway, R. I. (2023). The pace of shifting seasons in lakes. *Nature Communications*, *14*(1), 2101. https://doi.org/10.1038/s41467-023-37810-4

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The beijing climate center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), 1573–1600. https://doi.org/10.5194/gmd-12-1573-2019

Xu, R., & Wunsch, D. (2008). *Clustering*. John Wiley & Sons. (Google-Books-ID: kYC3YCyl_tkC).

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The meteorological research institute Earth System model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Series II*, *97*(5), 931–965. https://doi.org/10.2151/jmsj.2019-051

Zakari, Y., Michel, A., & Lehning, M. (2022). Future trends in wind resources and their consistency in the Indian sub-continent. *Sustainable Energy Technologies and Assessments*, *53*, 102460. https://doi.org/10.1016/j.seta.2022.102460

Zhao, F., Shi, W., Wu, Y., Lian, Y., & Li, Q. (2025). Counteracting greenhouse gas and aerosol influences intensify global water seasonality over the past century. *Communications Earth & Environment*, *6*(1), 35. https://doi.org/10.1038/s43247-025-02025-6

21698996, 2026, 20, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045869 by CochraneItalia, Wiley Online Library on [15/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License