Environmental Science: Atmospheres

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d6ea00029k

Machine learning-based 1 year-lead summer precipitation prediction over the DPRK using regionally optimized spatial-contrast indices

Yong-Sik Ham, *[ab] Hyon-Su Ri,[a] Kyong-Bok Sonu,[ab] Sang-Il Jong[a] and Un-Sim Paek[a]

The northern part of the Korean Peninsula (NPKP) is a region highly dependent on summer monsoon precipitation for water resources and disaster risk management, and reliable one-year-ahead prediction of summer precipitation is pivotal for formulating climate adaptation strategies. Using the NCEP/NCAR reanalysis dataset spanning 75 years (1948–2022) and precipitation data from 37 meteorological stations in the Democratic People's Republic of Korea, this study developed regionally optimized spatial-contrast indices to explain the interannual variability of summer precipitation. The results show that the circulation patterns with high concurrent correlation with summer precipitation do not persist into the following year. However, the deepened low-pressure system over inland Eurasia, particularly around Lake Baikal, in the preceding summer ($t-1$) exhibits a significant lagged correlation with summer precipitation (t). Based on this physical insight, the study designed spatial-contrast indices that quantify the regional circulation anomaly differences between Siberia and eastern China. A back-propagation neural network (BPNN) trained on these indices achieved an anomaly correlation coefficient (ACC) of 0.59, a root mean square error (RMSE) of 134.61 mm, and a mean bias error (MBE) of −4.22 mm during the 5-fold cross-validation period (1949–2022), significantly outperforming traditional linear regression models. These findings demonstrate that regionally optimized atmospheric circulation indices provide robust predictive skill for long-lead precipitation forecasting in data-scarce regions.

Received 24th February 2026
Accepted 9th April 2026

DOI: 10.1039/d6ea00029k

rsc.li/esatmospheres

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Environmental significance

This study advances atmospheric science by developing a machine learning framework for 1 year-lead summer rainfall prediction over the northern Korean Peninsula, a data-sparse region highly vulnerable to climate variability. By introducing regionally optimized spatial-contrast indices that capture large-scale circulation gradients between Siberia and eastern China, the work enhances understanding of mid-latitude atmospheric drivers of regional precipitation. The approach addresses critical gaps in long-lead forecasting for Northeast Asia, supporting climate adaptation in agriculture, water resources, and disaster risk management. Methodologically, integrating spatial pattern analysis with neural networks offers a transferable strategy for improving seasonal prediction in observation-limited regions, contributing to global environmental risk assessment and resilience planning under a changing climate.

1 Introduction

Reliable long lead forecasting of summer precipitation is a cornerstone of climate resilience in monsoon dependent societies. Over the northern part of the Korean Peninsula (NPKP), where water security hinges on a narrow seasonal window, summer precipitation delivers more than half of the annual total.[1,2] Yet this concentration also renders the region acutely vulnerable: intense precipitation episodes, often amplified by typhoon interactions or stalled frontal systems, can trigger flash floods and landslides, disrupting food systems, infrastructure, and human safety. Conversely, a weak Jangma (Meiyu/Baiu) season may precipitate agricultural drought within weeks. In this context, annual lead prediction, issued before the onset of the preceding winter, holds exceptional operational value, enabling early allocation of reservoir capacity, crop planning, and disaster preparedness.

The Jangma front, recognized as the regional manifestation of the East Asian summer monsoon (EASM) over the Korean Peninsula, lies climatologically between the Meiyu (China) and Baiu (Japan) subsystems.[3–5] Unlike idealized monsoons driven by land–sea thermal contrast, Jangma is maintained through a delicate convergence of four distinct air mass sources: (i) warm, moist inflow from the South China Sea; (ii) dry, cool

[a]Department of Meteorology, Faculty of Global Environmental Science, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea. E-mail: hys12211126@163.com; wanglong_animation@163.com; sukb0819@163.com; horangyi11@163.com; pus19920427@163.com

[b]School of Water Resources and Environmental Science, China University of Geosciences (Beijing), China

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

outflow from the Siberian high; (iii) subsiding air under the western Pacific subtropical high (WPSH); and (iv) cold, moist advection from the Okhotsk Sea.[6–9]

Crucially, the intensity and persistence of these moisture-laden flows are influenced not only by atmospheric circulation but also by slowly varying boundary forcings such as oceanic or land-surface conditions in some monsoon systems.[10–13] While oceanic precursors including mixed-layer temperature and mean sea level anomalies exhibit high skill for Indian monsoon precipitation forecasting, their utility diminishes over the northern East Asian monsoon domain due to the increasing dominance of mid-latitude dynamics.[14–16]

ENSO, though a leading driver of tropical–subtropical monsoon variability on interannual timescales, exerts a highly modulated influence over the Korean Peninsula. Its teleconnection is nonlinear, contingent not only on ENSO's spatial flavor but also on the background state of the extratropical circulation.[17] As ref. 18 emphasized, skillful prediction of the East Asian summer monsoon requires explicit representation of mid-latitude precursors such as Eurasian snow cover or the Silk Road pattern that can override or amplify remote ENSO signals depending on their phase and evolution. Consequently, even remotely forced SST anomalies that enhance predictability in the Indian or southern East Asian sectors.[13] are strongly attenuated by the time they reach the NPKP, where local atmospheric configurations, particularly the Baikal-centered low and the extended Northern East Asian low (NEAL), emerge as the dominant indicators of interannual predictability linked to boundary memory.

This diagnostic gap motivates a shift from global index local response paradigms toward regionally tailored predictor design. A growing body of work suggests that predictive skill at multi-month to annual leads may reside not in the magnitude of individual anomalies, but in their spatial configuration; for example, zonal asymmetries in the WPSH, meridional shear across the mid-latitude jet, or dipole structures in Eurasian surface pressure.[19–21] Notably, ref. 22 identified NEAL as a cyclonic feature whose persistence modulates moisture transport into the NPKP. However, its predictive utility remains underexplored, particularly in lagged contexts. Moreover, cyclone tracks originating from the Huai River basin, a key moisture conduit, show marked decadal shifts that are poorly represented by coarse-scale indices. These observations point to a critical need: to develop circulation metrics that encode dynamic gradients, not just area means.

From a methodological standpoint, the choice of modeling framework must align with both the nonlinear nature of monsoon precipitation and the operational demand for skill, robustness, and interpretability. While linear regression remains a standard benchmark, as demonstrated by ref. 23, who achieved significant skill in forecasting East Asian summer monsoon precipitation using oceanic heat sources in a multiple linear regression framework, its performance degrades markedly in the presence of threshold effects, regime shifts, and tail events, which are intrinsic to Korean Peninsula hydroclimate.[24,25] In contrast, artificial neural networks, particularly the back-propagation neural network (BPNN), require no prespecified functional form and can approximate arbitrarily complex, nonlinear input–output mappings.[26,27] Recent studies confirm their superiority in capturing high-magnitude precipitation events, where linear models systematically underperform.[21,28] Though neuro-fuzzy systems show promise,[28] BPNN retains advantages in training stability, scalability, and sensitivity to extreme quantiles,[13] qualities that are essential for operational forecasting of flood- and drought-triggering precipitation over the NPKP. Notably, even in mid-latitude settings where dynamical memory is weak, purely statistical approaches remain competitive:[29] showed that a simple linear model using North Atlantic precursors achieved skill comparable to state-of-the-art dynamical ensembles in predicting winter wind speeds over China. This supports our strategy of prioritizing regionally optimized statistical predictors over global dynamical outputs, particularly when long-lead skill is sought in data-scarce regions.[19,30,31]

Accordingly, this study aims to develop a reliable annual lead forecasting framework for summer precipitation over the northern part of the Korean Peninsula (SR-NPKP). To this end, we focus on three core objectives. First, we will investigate the interannual variabilities of SR-NPKP, WPSH index (WPSHI), and NEAL index (NEALI). Second, we will analyze the simultaneous and one-year-lagged correlations between large-scale circulation patterns and SR-NPKP and explore the underlying physical mechanisms. Third, we will assess the performance of one-year-ahead prediction experiments for SR-NPKP. This assessment will involve a comparative evaluation of multiple model classes, including BPNN and standard linear regression models. The goal is to rigorously benchmark their forecast skill for both mean conditions and extreme precipitation quantiles.

The remainder of this paper is structured as follows. Section 2 details the data sources and methodology employed. In Section 3, we present the results. We begin by examining the observed variabilities of the key indices. Then, we analyze the diagnostic relationships and plausible mechanisms linking circulation patterns to SR-NPKP. Finally, we evaluate and discuss the comparative skill of the 1 year-lead prediction experiments using BPNN and linear models. The paper concludes with a summary and discussion of the main findings in Section 4.

2 Data and methodology

2.1 Data and study area

This study employs two complementary datasets to investigate the long-lead predictability of summer (June–August) precipitation over the NPKP. The spatial domain of the study region and the station distribution are shown in Fig. 1.

For the analysis of large-scale atmospheric circulation, monthly mean fields from the NCEP/NCAR Reanalysis I dataset.[32] are used. The analysis covers the 75 year period from 1948 to 2022 and is restricted to the boreal summer season. The following variables were extracted: sea level pressure (SLP), geopotential height at 500 hPa, and zonal wind (u), meridional wind (v), and air temperature (T) at four standard pressure levels (1000, 850, 700, and 500 hPa). All reanalysis data are provided

 © 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

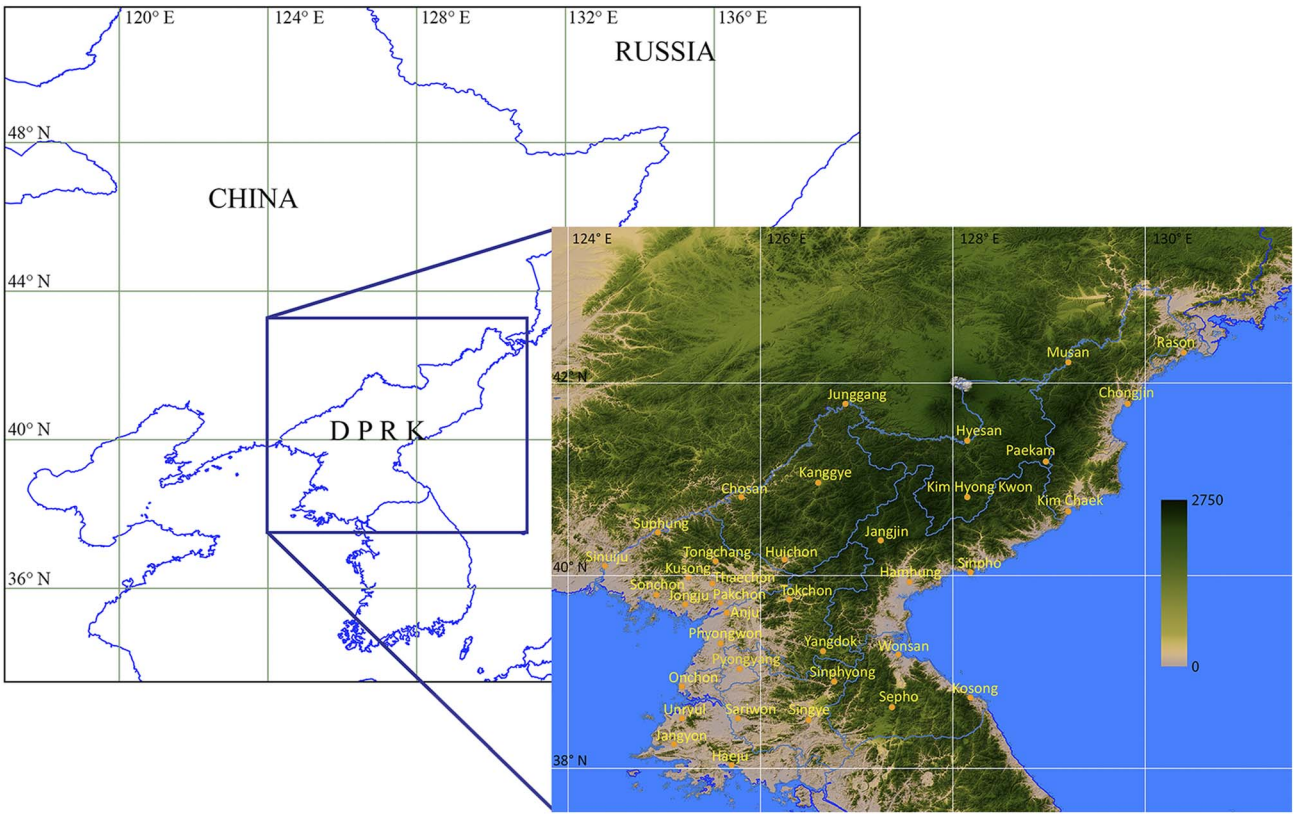

Fig. 1 The spatial domain of the study region and the station distribution.

on a 2.5° × 2.5° horizontal grid. The vertical resolution was intentionally limited to these levels to ensure robustness in the lead–lag correlation analysis and to mitigate the risk of overfitting during predictor selection.

For forecast verification, observed summer precipitation data were obtained from the State Hydro-Meteorological Administration (SHMA) of the Democratic People's Republic of Korea. An initial pool of 183 stations was screened for temporal completeness and internal consistency. A final network of 37 stations was retained, each possessing uninterrupted June–August precipitation records spanning the full 75 year period (1948–2022). Prior to data release, the SHMA applied standardized quality assurance procedures, including inter-station consistency checks, homogenization for instrument and observing practice changes, and correction for wind-induced undercatch. The final dataset was further subjected to independent verification using breakpoint detection and inter-station coherence analysis, with no major unresolved inconsistencies identified.

2.2 Methodology

This study aims to establish a 1 year lead prediction system for SR-NPKP. To this end, the atmospheric circulation field was quantified using two key indices:

WPSHI: average of 500 hPa geopotential height anomalies over 20–25°N, 125–140°E.[8,33]

NEALI: −1 × (average of 850 hPa geopotential height anomalies over 45–60°N, 110–130°E).[22] The negative sign ensures that higher NEALI values correspond to stronger cyclonic activity (lower geopotential height).

A linear least squares fit was applied to the 1948–2022 time series of SR-NPKP, WPSHI, and NEALI to remove the long-term linear trend. Subsequent analyses (*e.g.*, correlation, predictor construction) were performed only on the detrended interannual components to isolate year-to-year variability from secular trends. A 21 year moving mean[34] was computed for SR-NPKP to isolate low-frequency fluctuations. Abrupt shifts were detected using the Mann–Kendall nonparametric test,[35,36] a rank-based method robust to non-normality; a statistically significant change point is identified when the forward (UF) and backward (UB) cumulative rank statistics intersect above the 95% confidence envelope.

To identify predictors for 1 year lead forecasting, pointwise Pearson correlations were calculated between SR-NPKP in year t and summer (June–August) anomalies of sea level pressure (SLP), geopotential height, air temperature (T), zonal (u), and meridional (v) wind at 1000, 850, 700, and 500 hPa from the NCEP/NCAR Reanalysis I.[32] over 10–70°N in the preceding year ($t - 1$). Only detrended interannual components were used. The 1 year lead correlation thus corresponds to the relationship between atmospheric variables for 1948–2021 and SR-NPKP for 1949–2022.

Based on these maps, an area-shift experiment[37] was conducted, shifting the analysis domain in 2.5° increments to compute regional averages. Spatial-contrast indices—primarily differences (and select linear combinations) of anomalies between dynamically linked subregions (NEAL core *vs.* WPSH

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

edge)—were constructed to enhance physical interpretability and reduce redundancy. Though historically termed "spatial-contrast indices", the selected indices predominantly represent zonal/meridional gradients or dipole structures, consistent with evidence that spatial configuration often outperforms single-area means in seasonal prediction.[19–21]

Only spatial-contrast indices whose Pearson correlation coefficients with SR-NPKP satisfy $|r| \geq 0.4$ ($p < 0.001$, $n = 74$) were retained as candidate predictors. Here, the abbreviations denote: SLPA (sea level pressure anomaly), GPHA (geopotential height anomaly), TA (temperature anomaly), UA (zonal wind anomaly), and VA (meridional wind anomaly); numerical suffixes indicate pressure level (GPHA850) and analysis domain.

A backpropagation neural network (BPNN) was employed as the prediction model. As a feedforward neural network trained *via* gradient descent to minimize prediction errors,[26,27] BPNN requires no predefined functional relationship and thus performs well in capturing the strong nonlinearity and threshold sensitivity characteristic of monsoon precipitation systems.[38] The network architecture was structured as follows: the input layer consisted of the selected spatial-contrast indices; three hidden layers were adopted, each containing 10 neurons with a sigmoid activation function; and the output layer contained a single neuron representing the predicted SR-NPKP.

In this study, the full dataset covering 1949–2022 was used for model development. To avoid overfitting and ensure the stability and reliability of model performance, 5-fold cross-validation was applied to the entire dataset. In this procedure, the full dataset was randomly divided into five subsets of equal size. In each fold, four subsets were used for training and the remaining one for validation; this process was repeated five times to determine the optimal model parameters.

Predictive performance was evaluated based on the results of 5-fold cross-validation. Three metrics were employed to comprehensively quantify forecast skill: the anomaly correlation coefficient (ACC) to assess anomaly prediction skill, root mean square error (RMSE) to quantify the overall magnitude of prediction errors, and mean bias error (MBE) to evaluate systematic over- or under-prediction.

3 Results and discussion

3.1 Variabilities of SR-NPKP, and WPSH & NEAL intensities

We investigated the annual variability and abrupt change characteristics of SR-NPKP, WPSHI and the NEALI over the period 1948–2022.

Anomalies of SR-NPKP (deviations from the long-term mean of 633.5 mm) show strong interannual variability, with a standard deviation of 164.9 mm—accounting for 26% of the mean. The 75 year SR-NPKP time series has a statistically significant linear decreasing trend ($p = 0.03621$), with a decrease rate of 18.3 mm per decade (Fig. 2a). As a result of spectral analysis,[39] the three most prominent periodicities were found to be 8.3 years, 2.8 years, and 15.0 years.

The 21 year moving average (1958–2012) shows a distinct regime shift around the mid-1970s (Fig. 2a). Over the 27 years from 1948 to 1974, positive anomalies accounted for 16 years (59%) and negative anomalies for 11 years (41%), with above-average precipitation persisting overall, and there were no extreme drought years with anomalies less than −200 mm. In contrast, over the 48 years from 1975 to 2022, negative anomalies dominated at 32 years (67%), and extreme precipitation events exceeding +200 mm (5 years) and less than −200 mm (6 years) occurred with similar frequencies. This suggests that amid the overall decreasing trend, climate change has increased the variability of SR-NPKP, leading to more frequent extreme events such as droughts and floods.

The Mann–Kendall test was applied for statistical significance verification. The intersection of the forward statistic (UF) and backward statistic (UB) was confirmed within the 95% confidence interval in the mid-to-late 1970s, indicating that the abrupt shift in SR-NPKP is statistically significant. This result is consistent with the two regimes (1948–1974 and 1975–2022) derived from the moving average analysis, with the transition around 1974–1975.

Fig. 3a shows the long-term variation trend of the WPSHI. The long-term average value of the WPSHI is 5872.5 gpm, with a standard deviation of 11.1 gpm. It showed a significant increasing trend of +3.5 gpm per decade ($p < 0.001$) over the entire period, and the intensity regime changed starting in the late 1970s. Over the 30 years from 1949 to 1978, negative anomalies were overwhelmingly prevalent at 28 years (93%), but over the 44 years from 1979 to 2022, positive anomalies shifted to dominate at 31 years (70%), and strong negative anomalies of −10 gpm or less rarely occurred. Mann–Kendall test results showed that the abrupt change in WPSH intensity was statistically significant at the 95% or higher confidence level after 1986.

The NEALI exhibits a long-term mean of 1424.6 gpm and a standard deviation of 19.6 gpm (Fig. 3b). The time series exhibited a highly significant decreasing trend of −6.98 gpm per decade ($p < 10^{-15}$). Additionally, an abrupt regime shift occurred around 1974, transitioning from a positive phase (strong low pressure) to a negative phase (weak low pressure), which was also significant at the 95% or higher confidence level.

Simultaneous correlation analysis results showed that the correlation coefficient between WPSHI and SR-NPKP was −0.18 ($p > 0.05$), which is not statistically significant. This indicates that the interannual variability of western Pacific subtropical high intensity does not exhibit a direct and stable linear relationship with SR-NPKP. In contrast, the correlation coefficient between NEALI and SR-NPKP was +0.47 ($p < 0.001$), demonstrating a robust concurrent relationship: stronger Northeast Asian low pressure activity (*i.e.*, decreased geopotential height anomalies → increased NEALI) is associated with greater summer precipitation in the same year. This result aligns with the physical mechanism proposed by ref. 22, which holds that the interannual variability of East Asian summer precipitation is primarily regulated by the activity of mid-latitude low pressure systems in Northeast Asia.

However, for 1 year lead forecasting, lag-1 correlation (1 year delayed correlation) is more critical than simultaneous correlation. As confirmed in this study, neither WPSHI nor NEALI showed significant lag-1 correlations, making them unsuitable

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

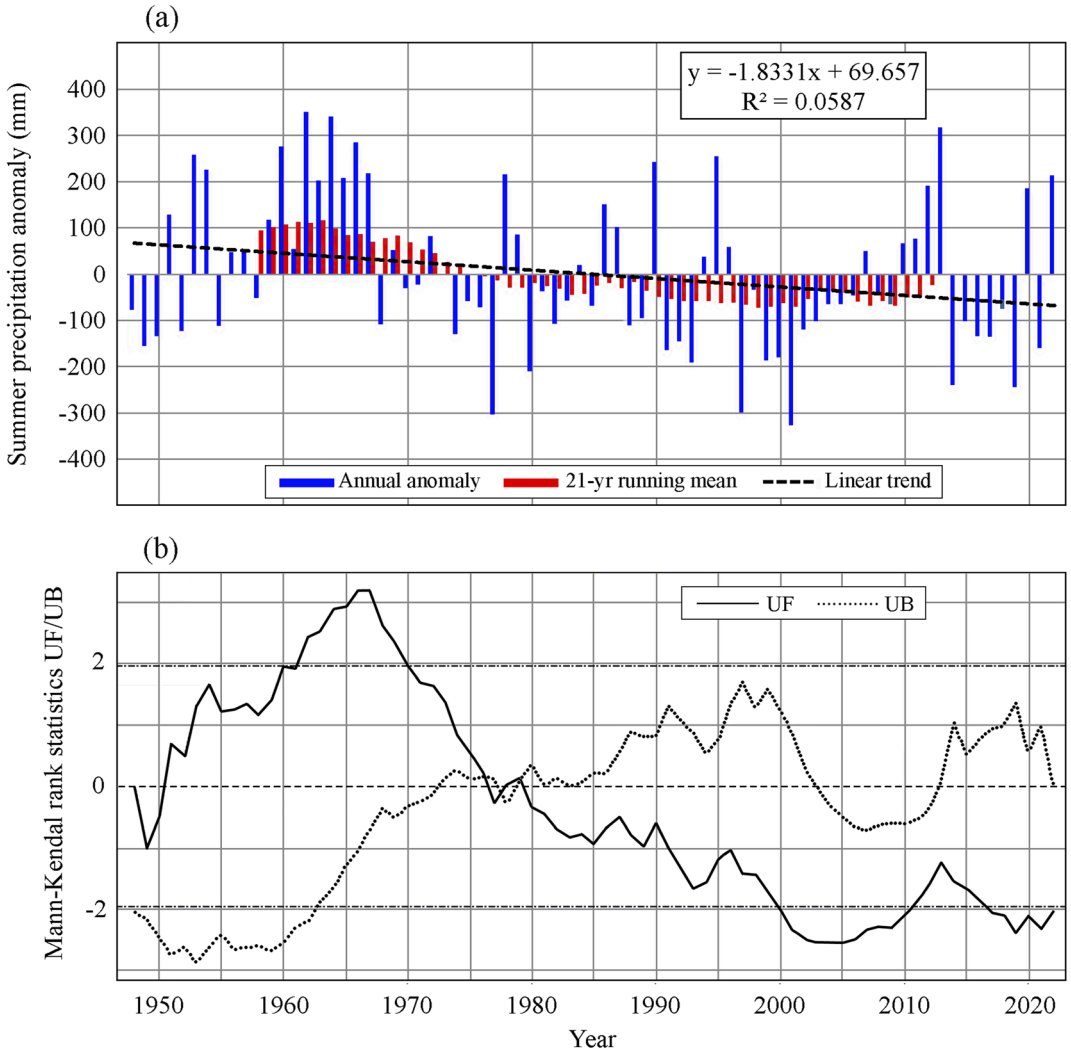

Fig. 2 The variability (a) and abrupt change (b) in SR-NPKP anomaly time series for the period of 1948–2022.

as predictors. This may be due to the weak annual persistence of high/low pressure intensity itself or the involvement of complex remote forcing. Therefore, in this study, the spatial-contrast index of regional anomalies was introduced as a new predictive factor to capture more comprehensive and persistent atmospheric circulation signals.

3.2 Simultaneous and 1 year-lagged correlations between summer atmospheric circulation and SR-NPKP: mechanisms and predictor design

To objectively identify optimal predictor domains, we performed an area shift experiment:[24,37,40] for each atmospheric variable and pressure level, area-averaged anomalies were computed over the Eastern Hemisphere (10–70°N), systematically shifting a 2.5° × 2.5° grid box in latitude and longitude. For each variable, the region yielding the maximum absolute 1 year-lagged correlation with SR-NPKP was selected as the optimal predictor domain, and these regions correspond precisely to the green rectangles in Fig. 4–7.

3.2.1 Spatial structure of correlations. Fig. 4–7 show the spatial patterns of simultaneous and 1 year-lagged correlations between SR-NPKP and anomalies of various atmospheric variables over the Eastern Hemisphere (10–70°N, 0–180°E). In the following figures, climate variable abbreviations are defined as follows: SLPA-sea level pressure anomaly; GPHA – geopotential height anomaly; TA – air temperature anomaly; UA – zonal wind anomaly; and VA – meridional wind anomaly. Numerical suffixes indicate the isobaric level; for example, GPHA850 denotes the geopotential height anomaly at 850 hPa.

As shown in Fig. 4a, a broad region of strong negative simultaneous correlation ($r \leq -0.50$) exists between SR-NPKP and SLPA over continental East Asia (70–130°E), centered near the Huai River Valley (35°N, 112.5°E). This reflects the well-documented link, highlighted in previous studies, between enhanced low-pressure/cyclonic activity over northern China and increased SR-NPKP.[22,41] A relatively weak positive correlation ($r \approx +0.30$) is observed southeast of Japan (30°N, 135°E), which is thought to be associated with the westward extension of the WPSH. As the vertical level increases (from SLP to 500

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

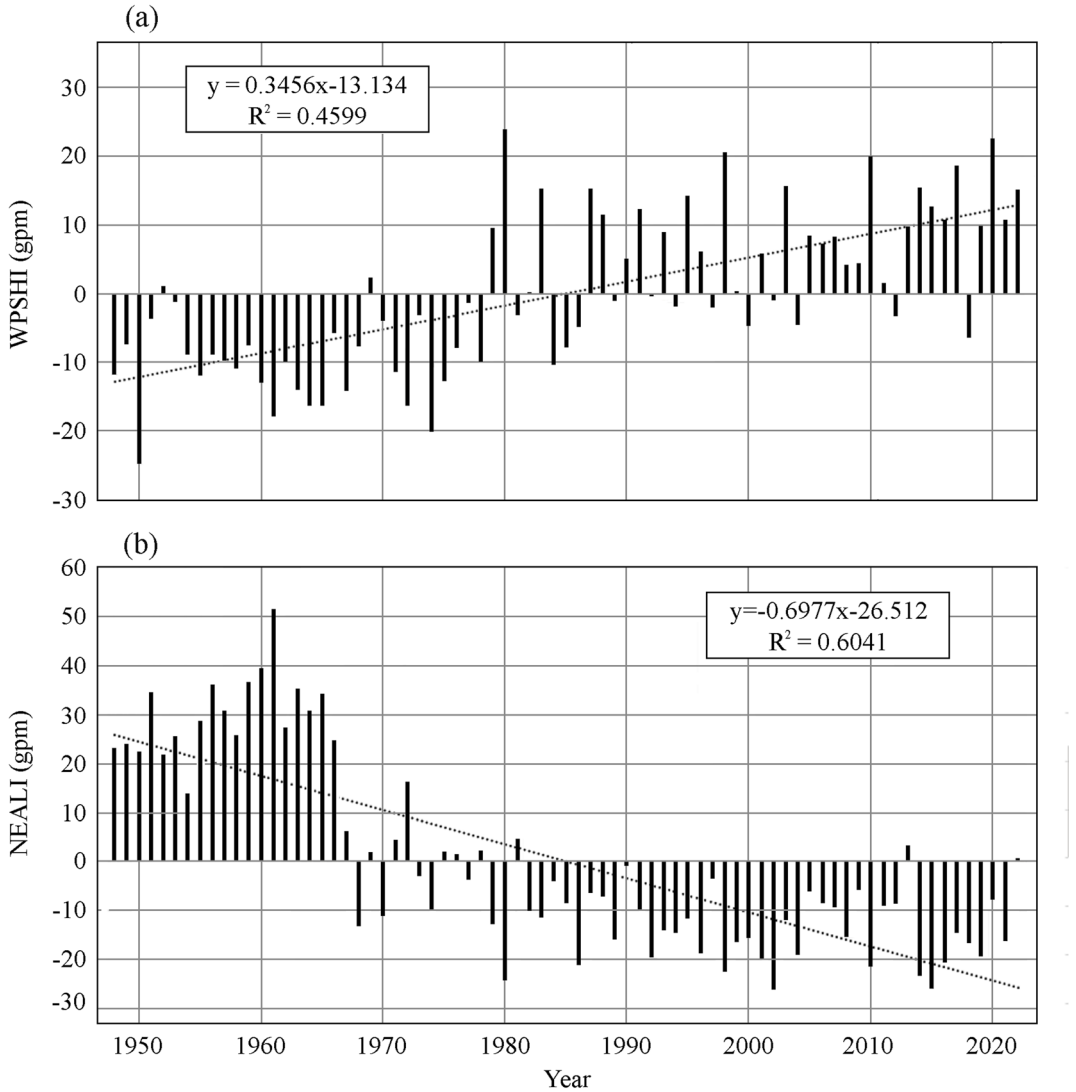

Fig. 3 The interannual variabilities of (a) WPSHI and (b) NEALI time series for 1948–2022. Dotted lines indicate the linear trends of WPSHI and NEALI time series for the study period.

hPa), both correlation centers shift toward the core region of the NEAL (45–60°N, 110–130°E), though the absolute magnitudes of the correlations tend to decrease (Fig. 4a, c, e and g). This indicates a vertically attenuating signal with baroclinic characteristics. In contrast, the 1 year-lagged correlation patterns (Fig. 4b, d, f and h) show that the primary negative correlation center shifts northwestward and locates over the Lake Baikal region (approximately 50°N, 95°E). Additionally, no significant positive lagged correlation is found in the northwestern Pacific. This confirms that the intensity of the WPSH in summer t does not persist into summer $t + 1$, thus limiting its utility as a 1 year-lead predictor.

Similar spatial displacement phenomena are observed for other atmospheric variables (Fig. 5–7). For temperature, the strongest positive simultaneous correlation appears at 850 hPa near Chengdu (30°N, 105°E) with $r = +0.39$, and the lagged correlation pattern shifts northwestward to peak at 850 hPa with $r = +0.42$. For zonal wind, enhanced westerlies over eastern China (115–125°E) are associated with wetter summers, showing a correlation coefficient of $r = +0.46$ at 850 hPa; the lagged correlation pattern persists despite a slight weakening, with $r = +0.44$. For meridional wind, both strengthened southerlies over the East Asia (132.5°E, 37.5°N; $r = +0.45$) and strengthened northerlies over Krasnoyarsk (55°N, 97.5°E; $r \approx -0.55$) create favorable conditions for increased SR-NPKP (Fig. 7a and g).

3.2.2 Physical basis for 1 year memory. Fig. 8 schematically synthesizes the dominant atmospheric anomaly patterns in the previous summer ($t - 1$) that exhibit statistically significant 1 year lag correlations with SR-NPKP in year t. Specifically:

– Negative SLP and geopotential height anomalies over inland Asia (40–50°N, 65–105°E) reflect a deepened low-pressure system, enhancing cyclonic circulation and low-level moisture convergence toward the Korean Peninsula. Crucially,

 © 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

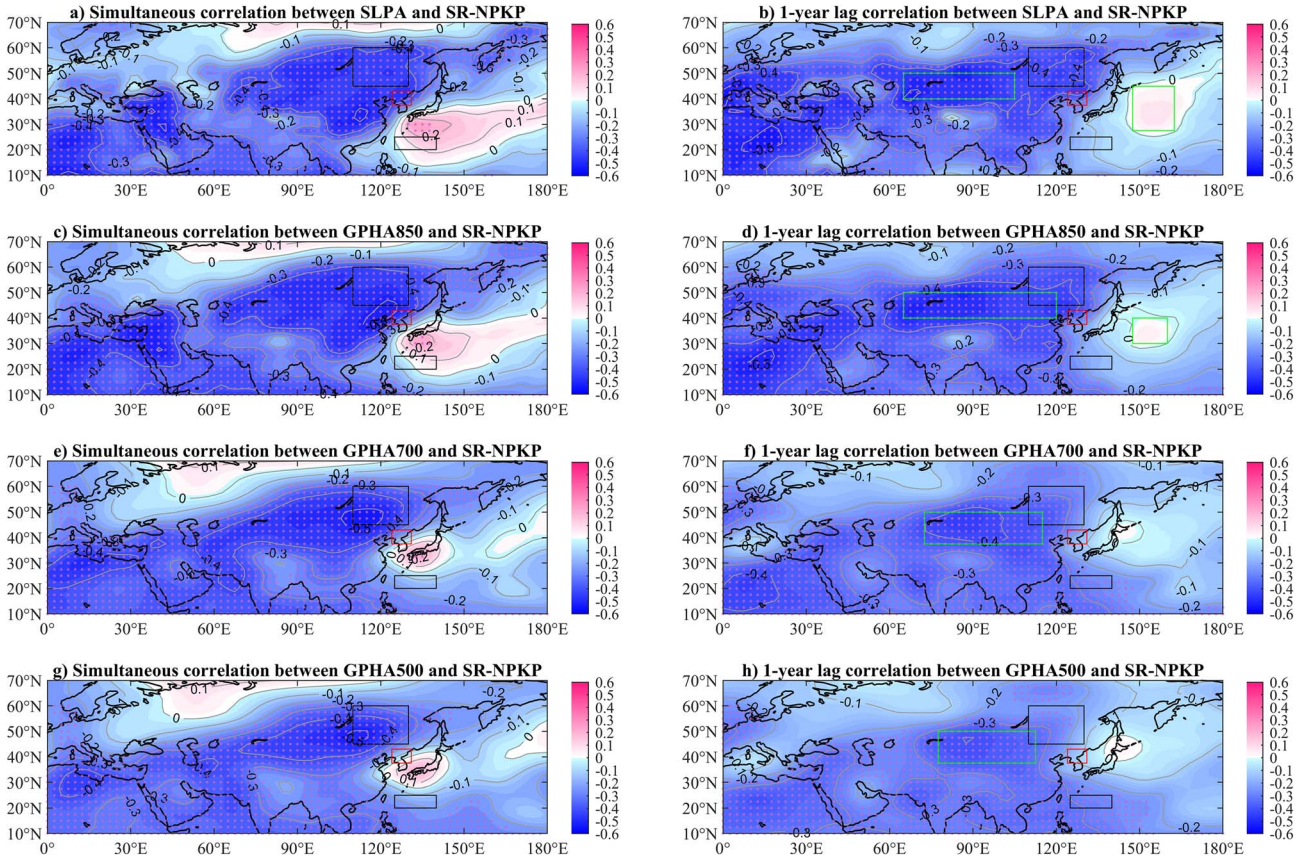

Fig. 4 Simultaneous and 1 year lag correlation maps of SR-NPKP with SLP and GPHA in summer at 850, 700, and 500 hPa for the region of 10–70°N in the Eastern Hemisphere. Red dots mark grid points where correlations are statistically significant ($p < 0.05$). The black rectangles denote the core areas of NEAL (large) and WPSH (small), respectively. The red rectangle denotes the NPKP. The green rectangle denotes the high 1 year lag correlation area of SR-NPKP with SLP in summer.

this circulation pattern is often anchored by antecedent Eurasian snow cover anomalies that persist through the subsequent winter and spring, providing a memory bridge to summer (t).[42]

– Positive temperature anomalies over the eastern Tibetan Plateau (37.5–42.5°N, 77.5–90°E) weaken the meridional thermal contrast, reducing the barrier effect of the plateau and facilitating the northward extension and persistence of the Jangma rainband.[22] This thermal anomaly is sustained by soil moisture memory and subsurface heat capacity over the plateau, which decay slowly compared to atmospheric variables.

– A zonal wind dipole—enhanced westerlies over East China (35–40°N, 115–125°E) and weakened westerlies over East Siberia (50–55°N, 85–100°E)—strengthens the East Asian subtropical jet waveguide, promoting upper-level divergence over the study region.[22] The stability of this waveguide is maintained by large-scale boundary forcings (*e.g.*, Eurasian snow cover and soil moisture anomalies) that lock the jet stream position across seasons.

– Strengthened southerly flow along a northeast–southwest corridor (Russian Far East to southern China) and strengthened northerly flow over East Siberia (52.5–55°N, 92.5–102.5°E) enhance low-level moisture transport from the western Pacific while suppressing dry continental air intrusion, favoring prolonged Jangma-front activity and heavy precipitation.[2] This moisture transport efficiency is further amplified by local soil moisture-precipitation feedback, where wet soils from year $t-1$ enhance evapotranspiration in year t.[43]

To test whether these conceptual patterns reflect robust, recurrent large-scale states rather than statistical artifacts, we conducted a composite analysis of summer ($t-1$) atmospheric fields for high-minus-low SR-NPKP years (upper *vs.* lower terciles of 1949–2022, $n = 25$ per group). As shown in Fig. 9, the composite anomalies reveal a coherent circulation structure that closely matches the schematic in Fig. 8:

- A pronounced negative SLP anomaly (up to −2.0 hPa) centered over Lake Baikal confirms the intensification of a deep inland low-pressure system (Fig. 9a);

- Positive T850 anomalies (+0.8 K) over the eastern Tibetan Plateau support thermal forcing of the monsoon (Fig. 9b);

- Enhanced westerlies (+0.7 m s^{-1}) over eastern China and strengthened southerlies (+0.6 m s^{-1}) along the Russian Far East coast reflect intensified low-level moisture transport (Fig. 9c and d).

All these features are statistically significant at the 90% confidence level (stippling, $p < 0.10$, Welch's t-test), providing strong observational validation of the proposed mechanism.

While the intrinsic memory of the free atmosphere is limited to intraseasonal timescales,[44] the atmospheric anomalies in

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

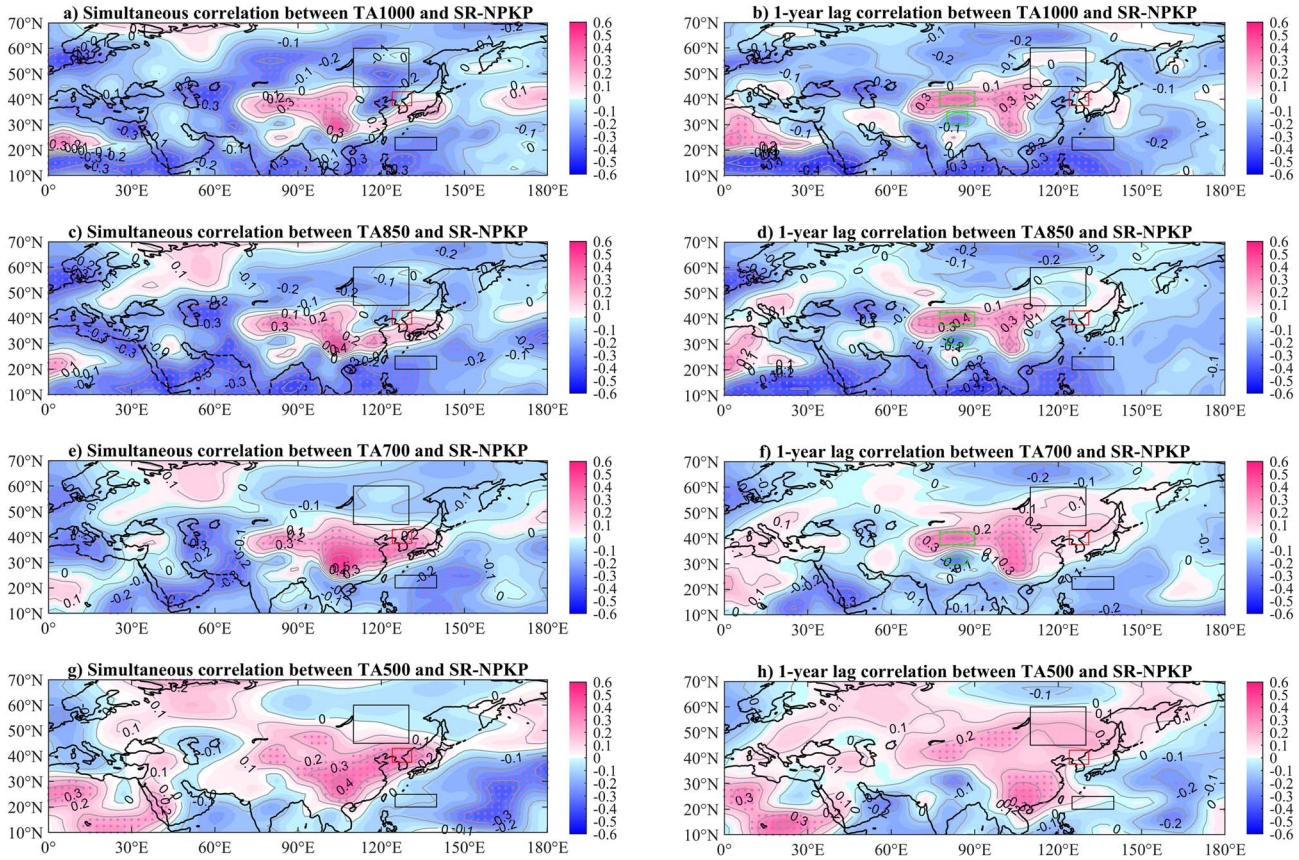

Fig. 5 The same as in Fig. 4, but with air temperature anomalies (TA) in summer at 1000, 850, 700, and 500 hPa.

summer ($t-1$) identified in this study act as proxies for slowly varying boundary conditions. Specifically, the deepened low-pressure system over Lake Baikal likely reflects antecedent anomalies in soil moisture, Eurasian snow cover, or sea surface temperature, which possess memory extending from months to years.[42] These boundary forcings persist through the winter and spring, modulating the overlying atmospheric circulation in summer (t) *via* land–atmosphere coupling and Rossby waveguide adjustments.[22] Thus, the atmospheric indices selected herein capture the integrated dynamical response to these persistent boundary forcings, rather than representing intrinsic atmospheric memory.

In particular, enhanced summer precipitation in year $t-1$ increases soil moisture over the Korean Peninsula and surrounding Eurasian landmass. Owing to reduced evapotranspiration during the cold season (October–April), this soil moisture anomaly decays slowly and persists into early summer of year t, elevating low-level specific humidity and convective available potential energy (CAPE). As demonstrated by,[24] such land–atmosphere coupling provides a physically sound basis for interannual predictability of the East Asian summer monsoon. Supporting this, ref. 22 showed that NEAL-related anomalies, when anchored by orography (Baikal–Okhotsk region), exhibit enhanced persistence *via* Rossby waveguide adjustment—further stabilizing the large-scale environment favorable for moisture transport and ascent.

3.2.3 Construction of spatial-contrast indices for 1 year prediction. The 1 year lag correlations between area-averaged climate variable anomalies (SLPA, GPHA, TA, UA and VA) in the preceding summer ($t-1$) and SR-NPKP in year t are presented in Table 1.

These anomalies were obtained over the 10–70°N band in the Eastern Hemisphere by shifting calculation regions at 2.5° grid intervals. This technique is called an area shift experiment, as described in ref. 21 and 37, and it serves to objectively identify robust correlation centers.

As shown in Table 1, individual anomalies exhibit 1 year lag correlations with SR-NPKP that range from −0.48 to +0.48. However, some regions show negligible predictive skill. A typical example is GPHA850-2 over the domain 30–40°N, 147.5–160°E, which has a correlation coefficient of only +0.03. This poor performance likely reflects local variability that is unconnected to the Korean Peninsula. In contrast, spatial-contrast indices consistently enhance the predictive signal. These indices are defined as differences or weighted sums between regionally averaged anomalies, and the methodology has been validated for typhoon prediction in ref. 21. In that study, differences such as SLPA-4 minus SLPA-1, which represents the pressure contrast between low pressure over North America and high pressure over the Indian Ocean, or U1000-3 minus U1000-2, which captures westerly shear, were shown to better capture large-scale oscillation patterns. These patterns include

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

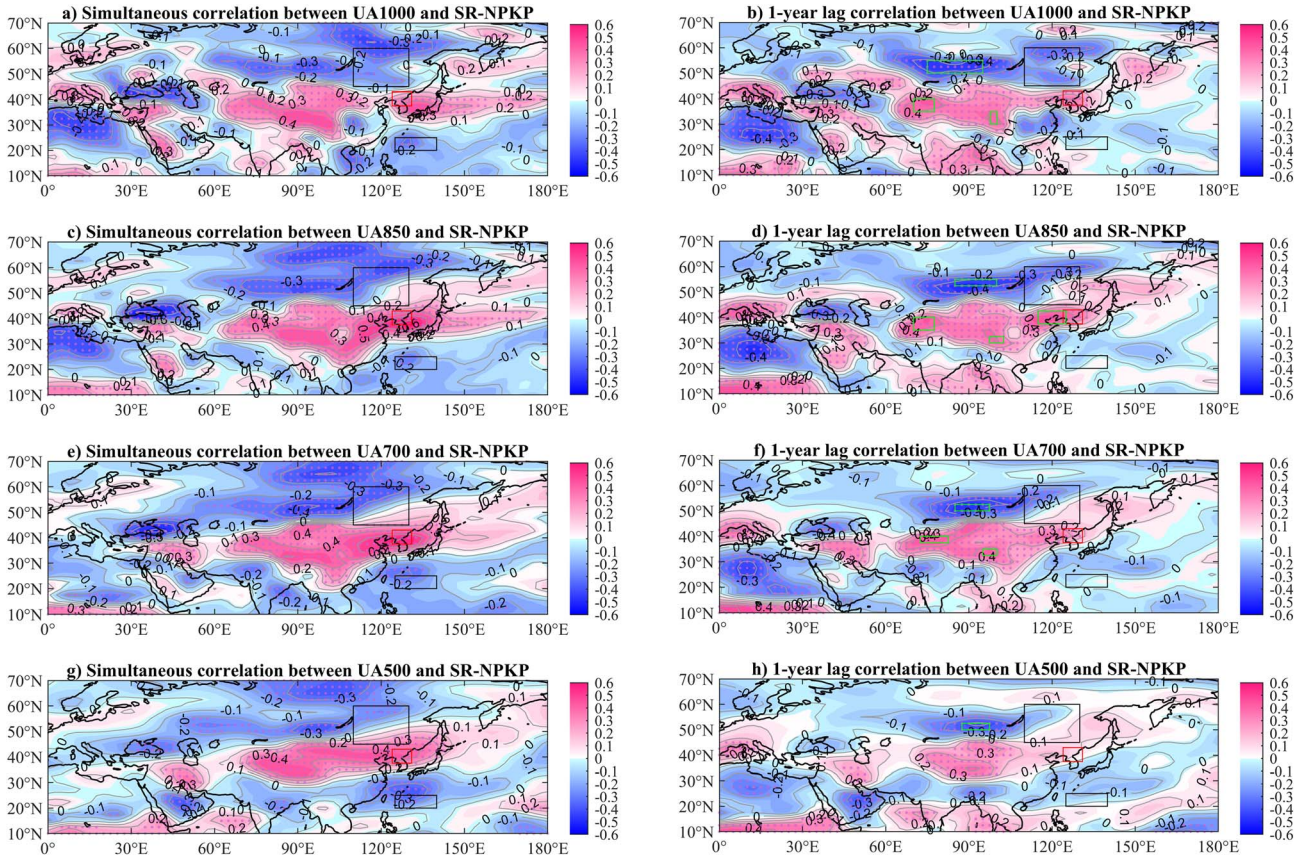

Fig. 6 The same as in Fig. 4, but with zonal wind anomalies in summer at 1000, 850, 700, and 500 hPa.

phenomena like land–ocean pressure seesaws and meridional wind shear, and the indices perform better than single-region averages by suppressing local noise and improving the signal-to-noise ratio.

Following a physics-based approach, this study generated candidate indices *via* linear combinations of regionally averaged anomalies—using differences for dipole patterns and weighted sums for tripole patterns. All mathematically possible pairwise and tripole combinations among the 13 candidate regions identified through the area-shift experiment were systematically evaluated against the statistical significance threshold ($|r| \geq 0.40, p < 0.001, n = 74$). Exactly 13 combinations satisfied this criterion and were retained without further subjective filtering. Notably, all retained indices exhibited physically coherent spatial patterns (*e.g.*, pressure gradients between inland Eurasia and the Northwest Pacific), indicating that physical interpretability emerged naturally as a property of statistically significant predictors rather than serving as an additional selection criterion (Table 2).

This data-driven refinement is exemplified by the first index (SLP_D12), which quantifies the sea level pressure contrast not between the *a priori* defined WPSH and NEAL domains, but between the two regions exhibiting the strongest lagged signals: inland Eurasia (40–50°N, 65–105°E) and the Northwest Pacific (27.5–45°N, 147.5–162.5°E). While the conventional WPSH–NEAL intensity difference aligns reasonably well with simultaneous circulation–precipitation relationships (Fig. 4a), its 1 year-lagged correlation is markedly limited, as the dominant signal shifts northwestward toward the Lake Baikal region in lagged correlation fields (Fig. 4b, d, f and h). In contrast, SLP_D12 captures this spatially extended pressure gradient involving inland Eurasia and achieves an enhanced absolute correlation coefficient of $|r| = 0.49$. This illustrates how regionally optimized spatial contrasts, grounded in observed lagged teleconnections, can improve upon traditional indices that rely on fixed climatological domains.

The eighth index (U850_D1234) quantifies the spatial asymmetry between strengthened low-level westerlies over three adjacent regions in eastern China (35–42.5°N, 70–125°E) and weakened background westerlies over central Siberia (52.5–55°N, 85–100°E). This structure physically aligns with the mechanism of NEAL intensification accelerating jet waveguide acceleration as proposed by ref. 22. It also exhibits the highest one-year-lagged correlation with SR-NPKP ($r = +0.52$, $p < 0.001$). The eleventh index (V1000_D12) represents the combination of enhanced southerly winds along the eastern coast of the Korean Peninsula and enhanced northerly winds over central Siberia. This pattern is consistent with the low-level moisture transport mechanism sustaining the Jangma front, as described by ref. 2. It is thus interpreted as a low-level moisture transport index.

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

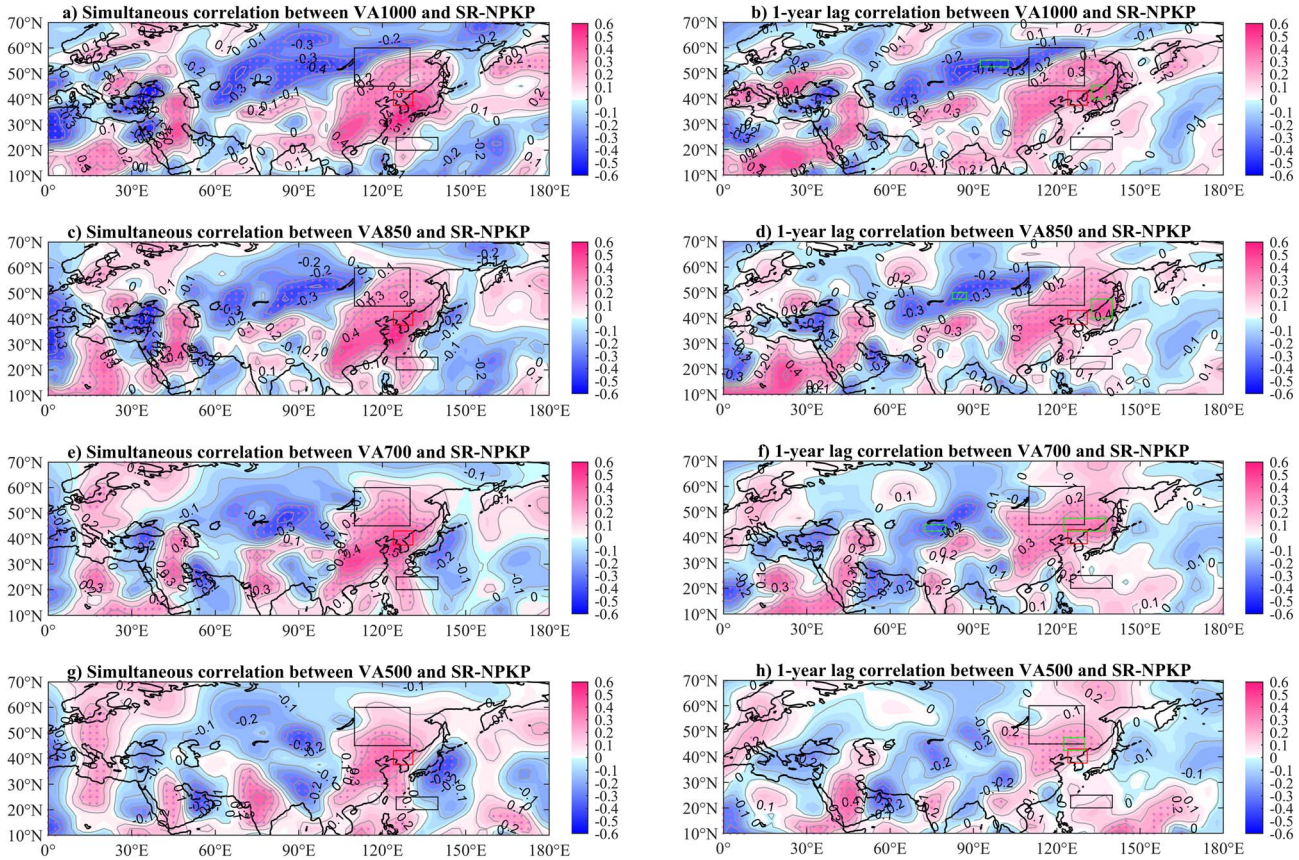

Fig. 7 The same as in Fig. 4, but with meridional wind anomalies in summer at 1000, 850, 700, and 500 hPa.

3.3 Prediction experiments of SR-NPKP for annual lead time and analyses of results

The central question of this study is which subset of predictors yields the highest prediction skill for SR-NPKP at an annual lead time using BPNN. Predictor selection for the neural network model was performed based on predictive performance derived from 5-fold cross-validation. Various predictor combinations were constructed using the 13 potential spatial-contrast indices listed in Table 2, and performance metrics including anomaly correlation coefficient (ACC), root-mean-square error (RMSE),

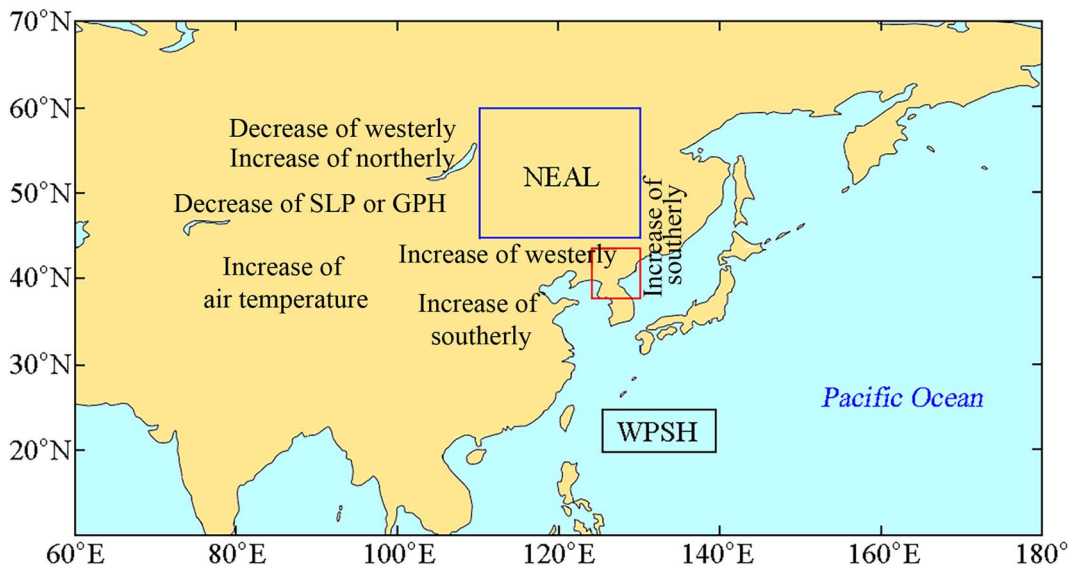

Fig. 8 The schematic showing the main circulation patterns in previous summer that are favorable for the increase in SR-NPKP. The black rectangle indicates the core region of WPSH and the blue rectangles denote the core region of NEAL. The red rectangle denotes the NPKP.

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

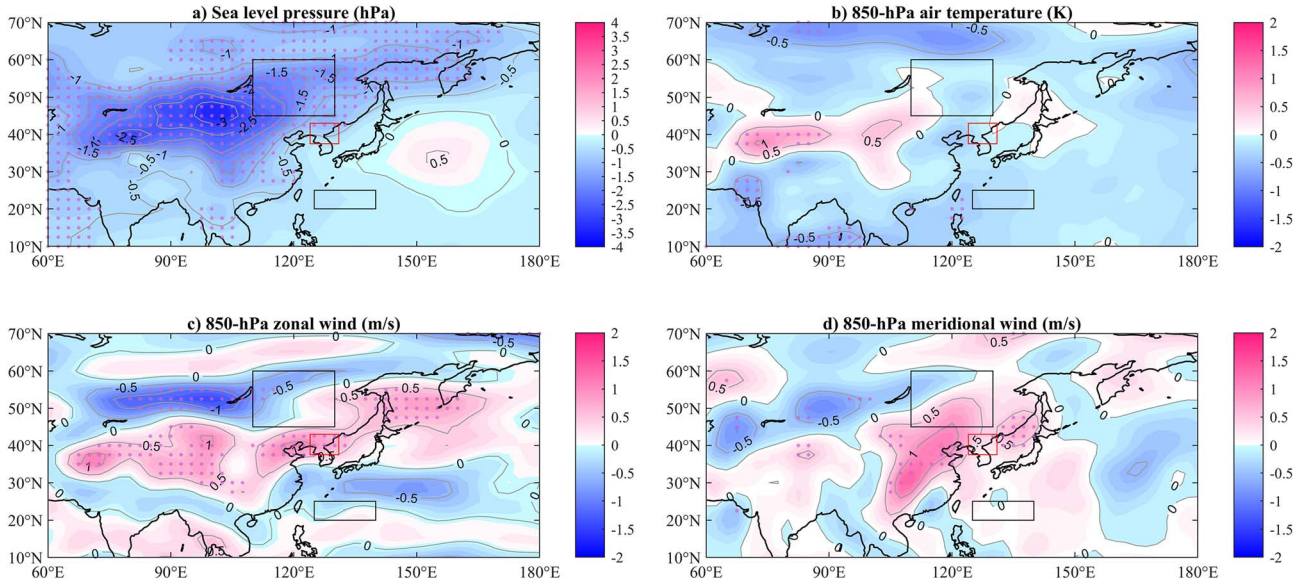

Fig. 9 Composite anomalies of summer atmospheric variables in the preceding year ($t-1$) for high minus low SR-NPKP years (upper *vs.* lower terciles of 1949–2022, $n = 25$ each). Red dots mark grid points where correlations are statistically significant ($p < 0.05$). The black rectangles denote the core areas of NEAL (large) and WPSH (small), respectively. The red rectangle denotes the NPKP.

Table 1 Potential area-averaged climate variable anomalies in previous summer affecting the interannual variability of SR-NPKP. The numbers behind a hyphen represent the focusing areas

Climate variable anomaly symbol	Area	Correlation coefficient
SLPA-1	40–50°N, 65–105°E	−0.47
SLPA-2	27.5–45°N, 147.5–162.5°E	+0.07
GPHA850-1	40–50°N, 65–120°E	−0.45
GPHA850-2	30–40°N, 147.5–160°E	+0.03
GPHA700	37.5–50°N, 72.5–115°E	−0.44
GPHA500	37.5–50°N, 7.5–112.5°E	−0.38
TA1000-1	37.5–42.5°N, 77.5–90°E	+0.37
TA1000-2	30–37.5°N, 80–87.5°E	−0.23
TA850-1	37.5–42.5°N, 77.5–90°E	+0.39
TA850-2	30–32.5°N, 80–87.5°E	−0.25
TA700-1	37.5–42.5°N, 77.5–90°E	+0.37
TA700-2	30–32.5°N, 80–87.5°E	−0.26
UA1000-1	35–40°N, 70–77.5°E	+0.45
UA1000-2	30–35°N, 97.5–100°E	+0.42
UA1000-3	50–55°N, 75–95°E	−0.48
UA850-1	35–40°N, 70–77.5°E	+0.46
UA850-2	30–35°N, 97.5–100°E	+0.42
UA850-3	37.5–42.5°N, 115–125°E	+0.46
UA850-4	52.5–55°N, 85–100°E	−0.47
UA700-1	37.5–40°N, 72.5–82.5°E	+0.48
UA700-2	32.5–35°N, 95–100°E	+0.44
UA700-3	50–52.5°N, 85–92.5°E	−0.47
UA500-1	32.5–37.5°N, 85–95°E	+0.39
UA500-2	50–52.5°N, 87.5–97.5°E	−0.41
VA1000-1	40–45°N, 132.5–137.5°E	+0.45
VA1000-2	52.5–55°N, 92.5–102.5°E	−0.46
VA850-1	40–47.5°N, 132.5–140°E	+0.45
VA850-2	47.5–50°N, 82.5–87.5°E	−0.41
VA700-1	42.5–47.5°N, 122.5–137.5°E	+0.42
VA700-2	42.5–45°N, 72.5–80°E	−0.36
VA500	42.5–47.5°N, 122.5–130°E	+0.33

and mean bias error (MBE) were evaluated. Accordingly, the top 5 best-performing predictor sets were selected for detailed analysis (Table 3), in which the performance metrics of the linear regression model using the same predictors are also presented for comparison.

For each predictor set, a BPNN with the logistic sigmoid activation function was employed. Model training and validation were implemented *via* 5-fold cross-validation over the full period 1949–2022, where each fold uses 4/5 of the time series for training and 1/5 for validation. Considering the sensitivity of neural network training to initial random weights, 100 repeated runs with different random seeds were conducted for each predictor set. The model achieving the lowest cross-validation RMSE was selected as the representative model for each set.

Performance metrics (ACC, RMSE, MBE) were calculated by concatenating the predicted and observed values from all five validation folds of the 5-fold cross-validation (1949–2022), thereby representing the model's skill over the entire period. The top 5 predictor combinations are ranked by BPNN ACC. ΔACC and ΔRMSE denote the improvement of BPNN over MLR (positive ΔACC and negative ΔRMSE indicate BPNN superiority). The optimal predictor set (case 3) is highlighted in bold.

As shown in Table 3, in terms of the ACC and RMSE metrics, the BPNN consistently outperforms the MLR across all five predictor combinations. Under the optimal predictor set (case 3: V850_D12, SLP_D12, T850_D12, U850_D1234), the BPNN achieves the highest ACC of 0.59 and the lowest RMSE of 134.61 mm, whereas the MLR yields an ACC of 0.53 and an RMSE of 140.41 mm with the same predictors. The performance gap (ΔACC = +0.06, ΔRMSE = −5.80 mm) confirms that even when well-selected predictors are used, nonlinear modeling provides additional predictive performance improvement. At this time, the optimal predictor set obtained above includes the

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Potential spatial-contrast indices of area-averaged climate variable anomalies connected with the interannual variability of SR-NPKP

Number	Definition	Symbol	Correlation coefficient
1	SLPA-1 − SLPA-2	SLP_D12	−0.49
2	GPHA850-1 − GPHA850-2	GPHA850_D12	−0.47
3	GPHA700	GPHA700	−0.44
4	TA1000-1 − TA1000-2	T1000_D12	+0.43
5	TA850-1 − TA850-2	T850_D12	+0.45
6	TA700-1 − TA700-2	T700_D12	+0.43
7	UA1000-1 + UA1000-2 − UA1000-3	U1000_D123	+0.50
8	UA850-1 + UA850-2 + UA850-3 − UA850-4	U850_D1234	+0.52
9	UA700-1 + UA700-2 − UA700-3	U700_D123	+0.52
10	UA500-1 − UA500-2	U500_D12	+0.43
11	VA1000-1 − VA1000-2	V1000_D12	+0.49
12	VA850-1 − VA850-2	V850_D12	+0.48
13	VA700-1 − VA700-2	V700_D12	+0.43

meridional wind index (V850_D12), sea-level pressure index (SLP_D12), low- to mid-tropospheric temperature index (T850_D12), and zonal wind index (U850_D1234).

Notably, the advantage of BPNN becomes more pronounced for suboptimal predictor combinations. For instance, in case 1, BPNN achieves $\Delta ACC = +0.22$ and $\Delta RMSE = -18.73$ mm relative to MLR, suggesting that BPNN better captures nonlinear interactions among atmospheric variables that linear models cannot represent. Regarding systematic bias, both models exhibit small MBE values ($|MBE| < 8$ mm) across all cases, indicating minimal systematic over- or under-prediction. These results confirm that the superiority of BPNN stems from its ability to model complex, nonlinear relationships among circulation predictors, rather than merely from using multiple inputs.

Fig. 10 illustrates the prediction performance of BPNN using the optimal predictor set (case 3: V850_D12, SLP_D12, T850_D12, U850_D1234) over the full period (1949–2022). Panel (a) shows the time series of observed precipitation anomalies alongside BPNN (red squares) and MLR (green triangles) predictions, while panels (b) and (c) present scatter plots with corresponding ACC values.

The time series comparison (Fig. 10a) reveals that both BPNN and MLR models reasonably capture the general phase of interannual variability, demonstrating the practical utility of these models for seasonal to interannual rainfall prediction. However, both models exhibit systematic limitations in predicting extreme events. While BPNN shows slightly greater variability than MLR, both models substantially underestimate precipitation anomalies during extreme wet years (*e.g.*, early 1960s: 1960, 1962, 1964, 1966; mid-1990s: 1990, 1995; 2013) and substantially overestimate precipitation anomalies during extreme dry years (*e.g.*, 1974, 1977, 1980, 1997, 2001, 2014, 2019). The differences for these extreme events often exceed 150 mm, indicating that neither model adequately captures the full amplitude of precipitation variability. Notably, however, the prediction errors for extreme events are notably reduced in BPNN compared to MLR.

To complement the continuous metrics and provide more operationally relevant guidance, we also evaluated categorical prediction skill using the Heidke Skill Score (HSS).[45] Following the WMO standardized verification system for long-range forecasts,[46] observed and predicted precipitation anomalies were classified into three equiprobable tercile categories: below normal (<33rd percentile), near normal (33rd–67th percentile), and above normal (>67th percentile).

The results show that BPNN achieves an overall HSS of 0.236, which represents a ΔHSS of +0.10 compared to MLR (HSS =

Table 3 Comparison of prediction performance between BPNN and Multiple Linear Regression (MLR) models using identical predictor sets

		BPNN			MLR			Δ(BPNN–MLR)
Case	Set of predictors	ACC	RMSE	MBE	ACC	RMSE	MBE	ΔACC/ΔRMSE
1	V850_D12, SLP_D12, T700_D12, U1000_D123	0.54	139.08	−7.64	0.32	157.81	3.29	+0.22/−18.73
2	V850_D12, T850_D12, U850_D1234, U500_D12	0.53	140.15	2.13	0.41	152.03	1.46	+0.12/−11.88
3	**V850_D12, SLP_D12, T850_D12, U850_D1234**	0.59	134.61	−4.22	0.53	140.41	1.83	+0.06/−5.80
4	V850_D12, T850_D12, GPHA700, U850_D1234	0.54	138.96	−6.63	0.38	154.64	5.27	+0.16/−15.68
5	V850_D12, T850_D12, GPHA700, U850_D1234, U500_D12	0.53	140.19	3.04	0.40	153.74	4.33	+0.13/−13.55

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

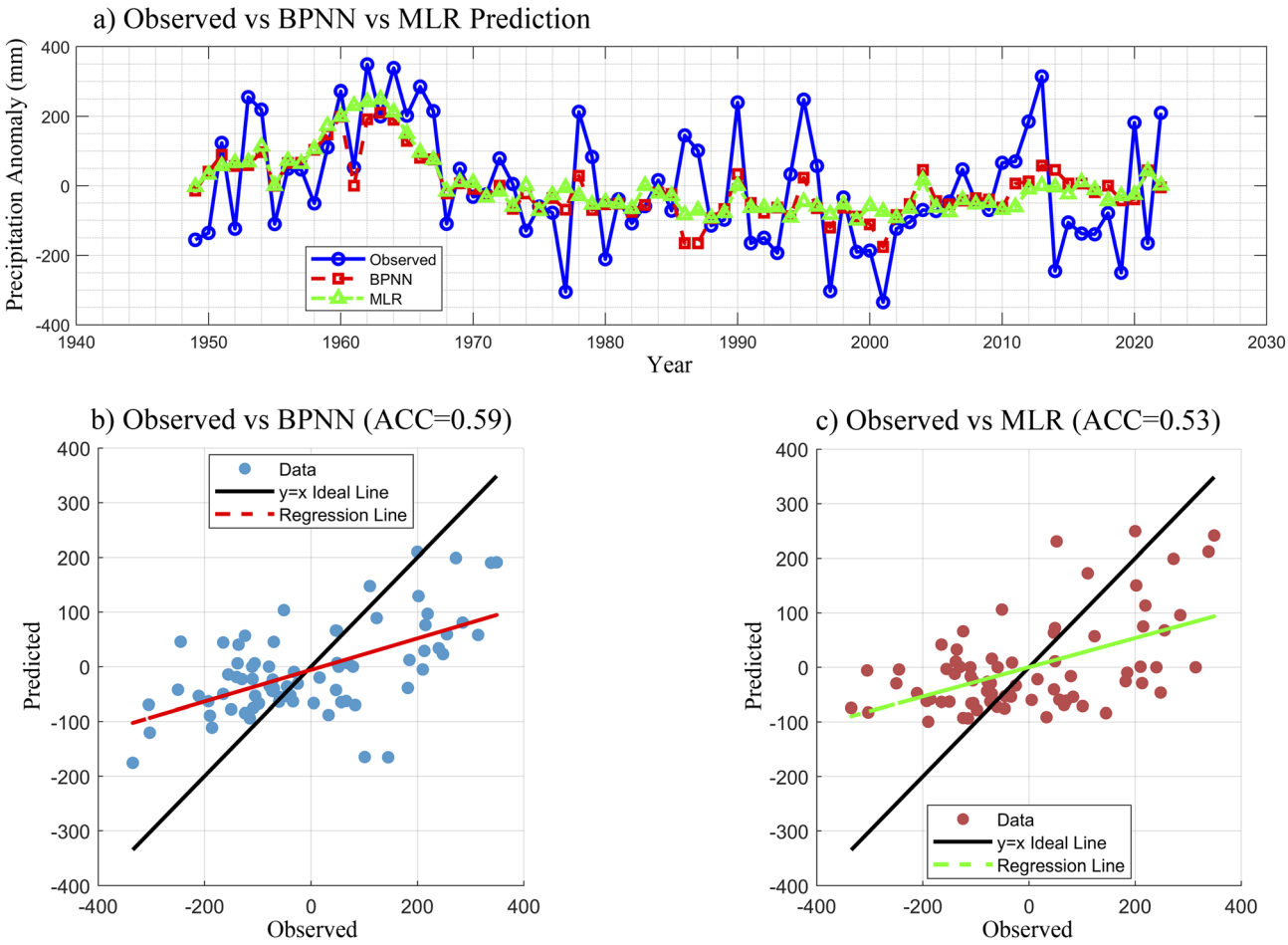

Fig. 10 Prediction performance of the BPNN model using the optimal predictor set (case 3: V850_D12, SLP_D12, T850_D12, U850_D1234) over the full period 1949–2022.

0.136), corresponding to a 73.5% relative improvement. This indicates that the BPNN model provides statistically superior predictive skill for seasonal rainfall categorical forecasting compared to MLR. This categorical assessment reveals an important nuance: while both models struggle to capture the exact amplitude of extreme precipitation events in continuous space (as evidenced by the regression-to-the-mean effect in Fig. 10b and c), they nonetheless provide skillful guidance for predicting the general direction of seasonal anomalies (above or below normal). For operational users—such as agricultural planners, water resource managers, and disaster preparedness agencies—this categorical information is often more actionable than precise quantitative forecasts. The BPNN's modest but meaningful improvement over MLR (ΔHSS = +0.10, or 74% relative increase) further underscores the value of nonlinear modeling for capturing complex circulation–precipitation relationships that linear approaches may miss.

Nevertheless, users should exercise caution when interpreting predictions for years with potentially extreme rainfall conditions, as both models exhibit systematic underestimation of extreme wet events and overestimation of extreme dry events. Future enhancements, such as incorporating direct moisture predictors (*e.g.*, specific humidity at 850 hPa, precipitable water) or adopting probabilistic forecasting frameworks, may further improve skill for extreme-event prediction.

The scatter plots (Fig. 10b and c) confirm this limitation. Although BPNN achieves a higher ACC (0.59) compared to MLR (0.53), the marginal improvement of +0.06 indicates that while BPNN captures slightly more variance than MLR, the improvement is limited. Both models exhibit regression lines with slopes substantially less than unity, reflecting a strong tendency to shrink predictions toward the climatological mean. This "regression to the mean" effect is particularly severe for extreme events (|anomaly| > 150 mm), where data points deviate substantially from the 1 : 1 line in both models.

Despite these limitations, the ACC values of 0.53–0.59 are noteworthy for seasonal to interannual prediction, as precipitation is inherently more difficult to predict than temperature due to its high spatiotemporal variability. For operational seasonal forecasting purposes, these models provide operationally useful information for predicting the general direction of seasonal rainfall anomalies (above or below normal), which is valuable for agricultural planning, water resource management, and disaster preparedness. Nevertheless, users should exercise

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

caution when interpreting predictions for years with potentially extreme rainfall conditions.

Although the optimal predictor set (case 3) does not include explicit moisture-related predictors, the wind, temperature and pressure-based indices (V850_D12, SLP_D12, T850_D12, U850_D1234) are dynamically linked to atmospheric circulation patterns. However, the systematic misrepresentation of extremes suggests that these circulation indices alone may not fully capture the complex boundary-layer atmosphere interactions (*e.g.*, soil moisture feedback, typhoon interactions, convective processes) that amplify extreme precipitation events. The absence of direct moisture predictors (*e.g.*, specific humidity at 850 hPa, precipitable water) may further limit the models' ability to capture extreme wet events, which often depend critically on moisture availability and convective instability.

While this study demonstrates that atmospheric circulation precursors—particularly regionally optimized spatial-contrast indices—provide statistically significant skill for 1 year-lead prediction of SR-NPKP, we acknowledge that boundary forcings may contribute additional predictability at this lead time. In particular, sea surface temperature anomalies (*e.g.*, ENSO), Eurasian snow cover, and soil moisture memory have been documented as sources of interannual-to-decadal climate predictability in the East Asian monsoon system.[24,42,43]

However, two practical considerations motivated our focus on atmospheric indices in the present work. First, the observational record for reliable land surface variables (*e.g.*, *in situ* soil moisture, snow depth) over inland Eurasia remains sparse prior to the satellite era (pre-1980s), limiting robust multi-decadal analysis with our 74 year dataset (1949–2022). Second, preliminary analysis revealed that direct lag-1 correlations between SR-NPKP and canonical ENSO indices (*e.g.*, Niño3.4) are weak, consistent with previous findings that ENSO's teleconnection to East Asian climate is highly nonlinear, shows complex variations, and is strongly modulated by extratropical circulation patterns.[47] In contrast, the spatial-contrast indices developed herein capture circulation configurations that integrate both tropical and extratropical influences, yielding substantially higher lag-1 correlations ($|r| \geq 0.40$).

We therefore interpret atmospheric circulation not as a substitute for boundary forcings, but as an integrated dynamical response that partially reflects the memory of prior boundary conditions. Future work will explicitly incorporate reanalysis-based boundary variables (*e.g.*, ERA5-Land soil moisture, EASE-Grid snow water equivalent) and phase-resolved ENSO indices into a hybrid predictor system. Such integration may further enhance prediction skill, particularly for extreme events where boundary-atmosphere coupling plays a dominant role.

In addition to predictor limitations, spatial scale limitations exist. While this study demonstrates skillful prediction for the regional mean precipitation over the NPKP, there are limitations regarding sub-regional scales. Precipitation variability at smaller spatial scales is often governed by local topography and convective processes that are not fully captured by large-scale circulation indices. Additionally, the limited sample size (74 years) poses challenges for developing robust independent models for multiple sub-regions without overfitting. Consequently, future work will aim to address these issues by incorporating higher-resolution observational data and local terrain factors to enhance predictive skill at sub-regional scales.

4 Conclusion

This study establishes a regionally tailored framework for 1 year-lead prediction of summer precipitation over the northern part of the Korean Peninsula (NPKP). The principal findings and their implications are summarized below.

First, traditional circulation indices such as WPSHI and NEALI exhibit negligible lag-1 correlations ($|r| < 0.2$) with SR-NPKP, confirming their limited utility for annual-lead forecasting. However, simultaneous and 1 year-lagged correlation analyses reveal a pronounced northwestward shift of the dominant signal from the canonical WPSH–NEAL dipole in summer (t) to a Baikal-centered low-pressure anomaly in summer ($t - 1$). This shift indicates that interannual predictability resides in spatially displaced pressure–wind configurations over inland Eurasia rather than in fixed climatological domains.

Second, guided by this physical insight, an objective area-shift experiment identified robust lagged precursors across the Eastern Hemisphere (10–70°N). Thirteen spatial-contrast indices constructed as linear combinations of regionally averaged circulation anomalies were screened for statistically significant lag-1 correlations ($|r| \geq 0.40$, $p < 0.001$). The sea level pressure contrast index SLP_D12, quantifying the pressure difference between the Baikal low (40–50°N, 65–105°E) and the Northwest Pacific high (27.5–45°N, 147.5–162.5°E), achieves $|r| = 0.49$ and substantially outperforms raw area-mean predictors.

Third, a BPNN using four optimal predictors (V850_D12, SLP_D12, T850_D12, U850_D1234) yields an anomaly correlation coefficient (ACC) of 0.59, a root mean square error (RMSE) of 134.61 mm, and a mean bias error (MBE) of −4.22 mm over the 5-fold cross-validation period (1949–2022), significantly outperforming multiple linear regression benchmarks (ACC = 0.53, RMSE = 140.41 mm) using identical predictors. The performance gap (ΔACC = +0.06, ΔRMSE = −5.80 mm) confirms that nonlinear modeling provides additional predictive skill even when well-selected predictors are used. Wind-based indices dominate the optimal set, consistent with their dynamical linkage to underlying pressure gradients and moisture transport pathways.

Fourth, while this study demonstrates skillful prediction for regional mean precipitation, we acknowledge limitations regarding sub-regional scales and boundary variable integration. Precipitation variability at smaller spatial scales is often governed by local topography and convective processes that are not fully captured by large-scale circulation indices. Additionally, the limited sample size (74 years) poses challenges for developing robust independent models for multiple sub-regions without overfitting. Future work will aim to address these issues by incorporating higher-resolution observational data, reanalysis-based boundary variables (*e.g.*, ERA5-Land soil

© 2026 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

moisture, EASE-Grid snow water equivalent), and phase-resolved ENSO indices into a hybrid predictor system.

Finally, this work demonstrates that regionally optimized spatial contrasts of atmospheric circulation provide superior predictive skill for monsoon precipitation compared to global-scale indices, offering a transferable strategy for long-lead forecasting in data-scarce regions. The methodology developed herein can be adapted to other monsoon-dependent regions where observational networks are sparse but reanalysis data are available, contributing to improved climate resilience and disaster preparedness in vulnerable communities.

Author contributions

Yong-Sik Ham: visualization, investigation, data curation. Hyon-Su Ri: software, writing – original draft. Kyong-Bok Sonu: software, visualization. Sang-Il Jong: conceptualization, methodology. Un-Sim Paek: supervision.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability

Data are available from the corresponding author upon request.

Acknowledgements

We are grateful to the State Hydro-Meteorological Administration (SHMA) of the DPRK for providing the observational monthly precipitation data of the DPRK.

References

1 R. Zhang, Changes in East Asian summer monsoon and summer precipitation over eastern China during recent decades, *Sci. Bull.*, 2015, **60**(13), 1222–1224.
2 K. C. Om, G. Ren, S. Li and K. C. O, Climatological characteristics and long-term variation of rainy season and torrential rain over DPR Korea, *Weather Clim. Extremes*, 2018, **28**, 667–678.
3 Y. Fang, Y. Zhang, A. Huang and B. Li, Seasonal and intraseasonal variations of East Asian summer monsoon precipitation simulated by a regional air–sea coupled model, *Adv. Atmos. Sci.*, 2013, **30**, 315–329.
4 Y. H. Ding and J. C. L. Chan, The East Asian summer monsoon: an overview, *Meteorol. Atmos. Phys.*, 2005, **89**, 117–142.
5 W. N. Guan, X. J. Ren, W. Shang, *et al.*: Subseasonal zonal oscillation of the western Pacific subtropical high during early summer, *J. Meteorol. Res.*, 2018, **32**(5), 768–780.
6 R. H. Kripalani and A. Kulkarni, Monsoon precipitation variations and teleconnections over South and East Asia, *Int. J. Climatol.*, 2001, **21**, 603–616.
7 B. Wang and H. Lin, Rainy season of the Asian-Pacific summer monsoon, *J. Clim.*, 2002, **15**, 386–396.
8 Y. Huang and X. Li, The interdecadal variation of the western Pacific subtropical high as measured by 500 hPa eddy geopotential height, *Atmos. Oceanic Sci. Lett.*, 2015, **8**(6), 371–375.
9 Y. B. Huang, X. Wang, X. Li and H. Wang, Changes in the influence of the western Pacific subtropical high on Asian summer monsoon precipitation in the late 1990s, *Clim. Dyn.*, 2018, **51**, 443–455.
10 M. M. Ali, P. V. Nagamani, N. Sharma, R. T. Venu Gopal, M. Rajeevan, G. J. Goni and M. A. Bourassa, Relationship between ocean mean temperatures and Indian summer monsoon precipitation, *Atmos. Sci. Lett.*, 2015, **16**(3), 408–413, DOI: 10.1002/asl2.576.
11 T. Venugopal, M. M. Ali, M. A. Bourassa, Y. Zheng, G. J. Goni, G. R. Foltz and M. Rajeevan, Statistical evidence for the role of southwestern Indian Ocean heat content in the Indian summer monsoon precipitation, *Sci. Rep.*, 2018, **8**, 12092, DOI: 10.1038/s41598-018-30552-0.
12 T. Venugopal, H. Rahaman, R. Anna, S. Erik, M. Ravichandran and S. S. V. S. Ramakrishna, Quantifying the role of antecedent Southwestern Indian Ocean capacitance on the summer monsoon precipitation variability over homogeneous regions of India, *Sci. Rep.*, 2023, **13**, 5553, DOI: 10.1038/s41598-023-32840-w.
13 N. Kurian, T. Venugopal, J. Singh and M. M. Ali, A soft-computing ensemble approach (SEA) to forecast Indian summer monsoon precipitation, *Meteorol. Appl.*, 2017, **24**(2), 308–314, DOI: 10.1002/met.1650.
14 Z. Z. Hu, Interdecadal variability of summer climate over East Asia and its association with 500hPa height and global sea surface temperature, *J. Geophys. Res.*, 1997, **102**, 19403–19412.
15 Y. Chen, H. Zhang, R. Zhou and H. Wu, Relationship between the ground surface temperature in Asia and the intensity and location of subtropical high in the western Pacific, *Chin. J. Atmos. Sci.*, 2001, **25**, 515–522.
16 Y. Yuan, H. Gao, W. Li, Y. Liu, L. Chen, B. Zhou and Y. Ding, The 2016 summer floods in China and associated physical mechanisms: A comparison with 1998, *J. Meteorol. Res.*, 2017, **31**, 261–277.
17 H. L. Ren, J. Zuo and Y. Deng, Statistical predictability of Niño indices for two types of ENSO, *Clim. Dyn.*, 2019, **52**, 5361–5382.
18 K. Fan, Y. Liu and H. Chen, Improving the Prediction of the East Asian Summer Monsoon: New Approaches, *Weather Forecast.*, 2012, **27**, 1017–1030, DOI: 10.1175/WAF-D-11-00092.1.
19 H. Gong, L. Wang, W. Chen, R. Wu, K. Wei and X. Cui, The climatology and interannual variability of the East Asian winter monsoon in CMIP5 models, *J. Clim.*, 2014, **27**, 1659–1678, DOI: 10.1175/JCLI-D-13-00039.1.
20 W. Hu, P. Liu, Q. Zhang and B. He, Dominant patterns of winter-time intraseasonal surface air temperature over the CONUS in response to MJO convections, *Clim. Dyn.*, 2019, **53**, 3917–3936.

Open Access Article. Published on 10 April 2026. Downloaded on 5/21/2026 7:45:45 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

21 S. I. Jong, Y. S. Ham, K. C. Om, U. S. Paek and S. S. O, Seasonal prediction of typhoons approaching the Korean Peninsula using several statistical methods, *Nat. Hazards*, 2022, **114**, 1857–1877.
22 Z. Lin and B. Wang, Northern East Asian low and its impact on the interannual variation of East Asian summer rainfall, *Clim. Dyn.*, 2016, **46**, 83–97.
23 E. Lee, N. Thomas, T. N. Chase and B. Rajagopalan, Seasonal forecasting of East Asian summer monsoon based on oceanic heat sources, *Int. J. Climatol.*, 2007, **28**, 667–678.
24 Z. Wu, B. Wang, J. Li and F. F. Jin, An empirical seasonal prediction model of the east Asian summer monsoon using ENSO and NAO, *J. Geophys. Res.*, 2009, **114**, D18120, DOI: 10.1029/2009JD011733.
25 L. Wang, M. Ting, D. Chapman, D. E. Lee, N. Henderson and X. Yuan, Prediction of northern summer low–frequency circulation using a high–order vector auto–regressive model, *Clim. Dyn.*, 2016, **46**, 693–709.
26 K. Hornik, M. Stichcombe and H. White, Multi-layer feedforward networks are universal approximators, *Neural Networks*, 1989, **2**, 359–366.
27 G. Zhang, B. E. Patuwo and M. Y. Hu, Forecasting with artificial neural networks: The state of the art, *Int. J. Forecast.*, 1998, **14**, 35–62.
28 E. Yakut and S. Süzülmüş, Modelling monthly mean air temperature using artificial neural network, adaptive neuro-fuzzy inference system and support vector regression methods: A case of study for Turkey, *Comput. Neural Syst.*, 2020, **31**, 1–36.
29 J. F. Lockwood, H. E. Thornton, N. Dunstone, A. A. Scaife, P. E. Bett, C. Li and H. L. Ren, Skilful seasonal prediction of winter wind speeds in China, *Clim. Dyn.*, 2019, **53**, 3937–3955.
30 P. Woli, J. Jones, K. Ingram and J. Paz, Forecasting drought using the agricultural reference index for drought (ARID): A case study, *Weather Forecast.*, 2013, **28**, 427–443.
31 L. Wang, X. Yuan and C. Li, Subseasonal forecast of Arctic sea ice concentration via statistical approaches, *Clim. Dyn.*, 2019, **52**, 4953–4971.
32 E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, *et al.*: The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 1996, **77**, 437–471.
33 X. He and D. Gong, Interdecadal change in western Pacific subtropical high and climatic effects, *J. Geogr. Sci.*, 2002, **12**(2), 202–209.
34 WMO, *Climatic Change, WMO Technical Note No. 79, WMO No. 195-TP-100*, World Meteorological Organization, 1966, Geneva.
35 H. B. Mann, Nonparametric tests against trend, *Econometrica*, 1945, **13**, 245–259.
36 M. G. Kendall *Rank Correlation Methods*, Charles Griffin, London, 1975, p. 202.
37 S. I. Jong, K. C. Om and Y. I. Pak, Influences of atmospheric circulation patterns on interannual variability of winter precipitation over the northern part of the Korean Peninsula, *Clim. Res.*, 2021, **85**, 35–50.
38 M. M. Mokhtarzad, F. Eskandari, N. J. Vanjani and A. Arabasadi, Drought forecasting by ANN, ANFIS, and SVM and comparison of the models, *Environ. Earth Sci.*, 2017, **76**, 729, DOI: 10.1007/s12665-017-7064-0.
39 D. S. Wilks *Statistical Methods in the Atmospheric Sciences: an Introduction*, Academic Press, San Diego, CA, 1995.
40 D.-Y. Gong, J. Yang, D. Guo, M. Hu, S.-J. Kim, Y. Gao, D. Guo, T. Zhou and M. Hu, Spring Arctic Oscillation–East Asian Summer Monsoon Connection through Circulation Changes over the Western North Pacific, *Clim. Dyn.*, 2011, **41**, 3085–3096.
41 B. Z. Shen, Z. D. Lin, R. Y. Lu and Y. Lian, Circulation anomalies associated with interannual variation of early- and late-summer precipitation in Northeast China, *Sci. China:Earth Sci.*, 2011, **54**, 1095–1104.
42 B. Wu and J. Wang, Winter Arctic Oscillation, Siberian High and East Asian winter monsoon, *Geophys. Res. Lett.*, 2002, **29**(19), 1897, DOI: 10.1029/2002GL015373.
43 R. D. Koster, *et al.*: Regions of strong coupling between soil moisture and precipitation, *Science*, 2004, **305**(5687), 1138–1140.
44 E. N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.*, 1963, **20**, 130–141, DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
45 D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press, Amsterdam, 3rd edn, 2011, p. 100.
46 WMO, *Standardized Verification System (SVS) for Long-Range Forecasts (LRF). WMO-No. 1091*, World Meteorological Organization, Geneva, 2012, p. 47.
47 T. Yu, W. Chen, P. Huang, G. Huang and X. Yang, Weakened influence of ENSO on the East Asian summer monsoon since the early 2000s, *npj Clim. Atmos. Sci.*, 2025, **8**, 114, DOI: 10.1038/s41612-025-00983-4.

© 2026 The Author(s). Published by the Royal Society of Chemistry