scientific reports

Check for updates

OPEN

Deep learning-based prediction of cold surge frequency over South Korea

Eung-Sup Kim[1], Joonlee Lee[2,4]✉, Jina Hur[1], Sera Jo[1], Yong-Seok Kim[1], Kyo-Moon Shim[1] & Joong-Bae Ahn[3]

Operational seasonal prediction models have limited skill in predicting number of wintertime cold surge days over South Korea. Here, we present a hybrid prediction framework that combines a coupled general circulation model with a Long Short-Term Memory neural network. Using data from 1980/81 to 2021/22, the framework incorporates model predictions together with 24 climate indices whose robustness was confirmed through leave-one-year-out cross-validation (LOYOCV). Model performance was evaluated using both LOYOCV and an independent train–test split, ensuring robust assessment. Compared with dynamical-only predictions, the hybrid system achieved substantially higher correlation and lower root mean square error, demonstrating improved prediction skill. Shapley Additive exPlanations analysis identified key contributors such as the Scandinavia pattern, Western Pacific pattern, and Southern Oscillation Index, and regression confirmed that enhanced skill after 2000 was linked to stronger contributions from atmospheric indices. These results indicate a temporal shift in dominant teleconnection drivers from oceanic to atmospheric control, reflecting evolving sources of predictability. The framework highlights the complementary value of combining dynamical and statistical approaches. As cold surges over South Korea are embedded within larger-scale East Asian circulation anomalies, underscoring its broader regional significance and offering a pathway to improved early prediction of extremes and better climate risk management.

Keywords Cold surge days, Seasonal prediction, Long Short-Term memory (LSTM), Coupled general circulation model (CGCM), SHapley additive explanations (SHAP), Teleconnection patterns

The global mean temperature has exhibited a persistent upward trend under ongoing global warming, and this warming has accelerated in recent decades[1]. Despite this trend, extreme cold events such as cold surges continue to occur frequently, and in some regions their severity has even intensified[1]. In South Korea, winter mean temperatures have increased significantly, but the frequency, duration, and annual number of cold surge days (CSDs) have not shown clear long-term changes[2–4]. Moreover, projections based on future climate scenarios suggest that cold surges will persist into the future[5], indicating that these events are shaped not only by long-term warming trends but also by complex atmospheric dynamics (e.g.,[6]).

The occurrence of cold surges is influenced by various large-scale circulation patterns, including the Arctic Oscillation (AO), the Siberian High, the Aleutian Low, and blocking events in the mid- to high-latitudes such as Ural and Okhotsk blocking. A negative phase of AO intensifies the East Asian trough, leading to the southward advection of cold Arctic air and the occurrence of cold surges[7–10]. Cohen et al.[11] and Kim et al.[12] pointed out that the reduction of Arctic sea ice, particularly in the Barents–Kara Seas, can enhance the negative phase of AO and thereby increase the probability of cold surge occurrence. Park et al.[13,14] classified Ural and Okhotsk blocking as factors contributing to cold surge occurrence over South Korea, while Wang et al.[15] demonstrated that Ural blocking can intensify the Siberian High, which in turn enhances cold surges in East Asia. In addition, the simultaneous occurrence of Ural and Okhotsk blocking, commonly referred to as double blocking, has been reported as a major factor leading to longer-lasting and more intense cold surges over South Korea[2,16].

Previous efforts to predict cold surges in East Asia have relied on either dynamical or statistical approaches. Dynamical model–based studies have contributed to elucidating the physical mechanisms of cold surges by

[1]Climate Change Division, National Institute of Agricultural Sciences, Wanju, South Korea. [2]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Urban, Ulsan, South Korea. [3]Department of Atmospheric Sciences, Pusan National University, Busan, South Korea. [4]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, UNIST-gil 50, UNIST-gil 50, Ulsan 44919, Republic of Korea. ✉email: leejl@unist.ac.kr

simulating atmospheric circulation and energy budgets[17,18]. In addition, Kim et al.[2] analyzed multiple cold surge cases to evaluate the reproducibility of their spatial structures and statistical characteristics using seasonal prediction from a coupled general circulation model (CGCM). Heo et al.[5] further examined projected future changes in cold surge occurrence and the role of thermal structures using models from the Coupled Model Intercomparison Project Phase 5 (CMIP5). Nevertheless, dynamical models are often limited in accurately predicting temporal variability, intensity, and regional distribution of cold surges, due to sensitivity to initial conditions, coarse resolution, and challenges in representing high-latitude circulation[19,20]. Statistical model–based studies, on the other hand, have predicted cold surge occurrence by employing empirical relationships with observed climate indices, and in some cases demonstrating reasonable performance[21,22]. Yet their prediction skill tends to deteriorate under nonstationary climate conditions or extreme events beyond the training sample[21,23,24].

Recently, advanced statistical approaches, particularly machine learning and deep learning techniques capable of capturing nonlinear and complex atmosphere–ocean interactions, have been increasingly applied in weather and climate prediction research (e.g.,[25–27]). In particular, Long Short-Term Memory (LSTM), a recurrent neural network (RNN) architecture designed for time-series data, is widely used to predict various climate variables by mitigating long-term dependency issues and enhancing training stability[28,29]. LSTM has shown improved performance in predicting the North Atlantic Oscillation (NAO)[30], the Madden–Julian Oscillation (MJO)[31], and summer rainfall[27], and has also been applied to temperature distribution forecasts during heatwave and cold surge events[32]. However, LSTM-based studies specifically targeting seasonal prediction of CSDs remain scarce, and few attempts have combined dynamical prediction with deep learning while also interpreting the underlying predictors. Furthermore, the extent to which evolving teleconnection patterns influence cold surge predictability remains unclear, despite growing evidence that Arctic amplification and jet stream changes are altering the dynamics of East Asian winters.

Accordingly, this study has two main objectives. First, we aim to enhance the seasonal prediction skill of winter CSDs in South Korea by combining a dynamical model (CGCM) with a deep learning model (LSTM), thereby demonstrating the complementarity between physical simulation and data-driven learning. Second, we seek to examine changes in prediction performance and explore their possible connections with shifts in the dominant teleconnection drivers, focusing on the relative roles of oceanic and atmospheric patterns. By integrating prediction performance with interpretability, this study provides both a practical hybrid prediction framework and new evidence on the evolving sources of cold surge predictability. To our knowledge, few studies have applied LSTM to seasonal prediction of Cold Surge Days (CSD) in South Korea, and this approach remains relatively unexplored more broadly in East Asia. Because cold surges over South Korea are closely embedded within larger-scale East Asian circulation anomalies, the findings of this study may also provide implications beyond the Korean Peninsula. We therefore expect that the results will provide a novel perspective on the evolving drivers of cold surge predictability and contribute to advancing seasonal prediction research with relevance to East Asian winter climate.

Data and method

Data

The dynamical model used in this study is the Pusan National University/Rural Development Administration (PNU/RDA) CGCM, which was developed at PNU and is currently operated by the National Institute of Agricultural Sciences under the RDA. The model also contributes to the Asia-Pacific Economic Cooperation Climate Center (APCC) Multi-Model Ensemble system, providing monthly seasonal predictions. The PNU/RDA CGCM consists of atmospheric, oceanic, sea-ice, and land component models (Table 1), with detailed descriptions provided in[2,38]. Its performance has been evaluated and validated in several previous studies (e.g.,[2,39–42]). In this study, we used winter (December–February) predictions initialized in November, comprising 10 ensemble members for 42 winters from 1980/81 to 2021/22.

To derive CSD, we used the CGCM-predicted daily mean 2-m temperature, with mean bias correction applied using ERA5, the fifth-generation reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF;[43]), at 0.1° horizontal resolution. Observed CSD was calculated from daily temperature data at 56

Component model	Atmosphere		Community Climate Model (CCM3, [34])
	Land		Land Surface Model (LSM, [35])
	Ocean		Modular Ocean Model (MOM3, [36])
	Sea-Ice		Elastic-Viscous-Plastic Model (EVP, [37])
Resolution	Horizontal	Atmosphere	Spectral truncation T42
		Land	
		Ocean	longitude: 2.8125°, low latitude: ~0.7°, mid latitude: ~1.4°, high latitude: ~2.8°
		Sea-Ice	
	Vertical	Atmosphere	18 hybrid sigma-pressure level (top: 2.917 hPa)
		Land	6 levels
		Ocean	40 levels (top: 10 m, bottom: 5258 m)
		Sea-Ice	3 levels

Table 1. Description of the PNU/RDA CGCM and its components[2,33].

Automated Synoptic Observing System (ASOS) stations operated by the Korea Meteorological Administration (KMA).

Definition of cold surge

In this study, the daily temperatures from 56 ASOS stations were averaged to obtain the daily mean temperature over South Korea. Following previous studies, the onset of a cold surge was defined as a decrease exceeding 1.5 standard deviations (σ) of the daily mean temperature within two days, whereas its termination was determined as the first day when the temperature returned above −0.5σ after the onset[2,5,8,9,13,14,16]. For the model, cold surges were identified using the same method, with South Korea represented by the region 125–130°E and 35–40°N[2].

Predictors for cold surge

To enhance both the diversity and quality of input data for the deep learning model, we incorporated climate indices that showed strong correlations with observed winter CSD in South Korea, in addition to the model-predicted CSD. A lag-correlation analysis was conducted for 30 candidate climate indices, which cover lead times from 9 months to 2 months prior to the target winter season (i.e., March to October) to ensure a realistic seasonal prediction, and 12 indices (24 index–month combinations) with statistically significant correlations at the 95% and 99% confidence levels were selected (Fig. 1a). The analysis revealed that long-memory variables related to sea surface temperature (SST), such as the Trans-Niño Index (TNI), El Niño Modoki Index (EMI), and Niño 1.2 index, exhibited strong correlations with winter CSD, indicating their persistent influence on cold surge occurrence several months in advance. In addition, the Pacific–Japan (PJ) index in September–October (PJ_9 and PJ_10), the East Atlantic–Western Russia (EAWR) index in August (EAWR_8), and the Circumglobal Teleconnection (CGT) index in August (CGT_8) also showed strong correlations.

The predictive stability of the selected indices was assessed using the Leave-One-Year-Out Cross-Validation (LOYOCV) scheme. In this procedure, each winter in the 42-year record (1980–2021) was excluded once from training and used solely for validation, while the remaining 41 years served as the training set. This ensured that every winter was predicted independently of its own data, thereby reducing the chance of overfitting and providing a reliable estimate of predictor–CSD relationships[26,31]. Their correlations with observed CSD were then summarized using boxplots (Fig. 1b). Most indices consistently exceeded the 95% confidence threshold even in cross-validation, confirming their robustness as predictors. To further address potential concerns about predictor selection bias, we emphasize that the LOYOCV procedure confirmed the stability of key predictors. Among them, key predictors such as PJ_10, TNI_5, and CGT_8 maintain median correlations above 0.45, with some even surpassing the 99% confidence level. This persistence indicates that their predictive contributions are not simply a result of the full-record correlation, but instead reflect stable signals that hold across independent validation years. Based on these results, 12 climate indices, represented by 24 index–month combinations (Fig. 1a), were employed as predictors in the deep learning model. Figure S1 presents the pairwise Pearson correlations among the selected indices. Although the pairwise correlations indicate some multicollinearity, they do not capture lead-dependent phase and propagation, conditional (partial) dependencies among indices, or cross-lagged and nonlinear interactions that can contribute to prediction. Such temporal and nonlinear structure was

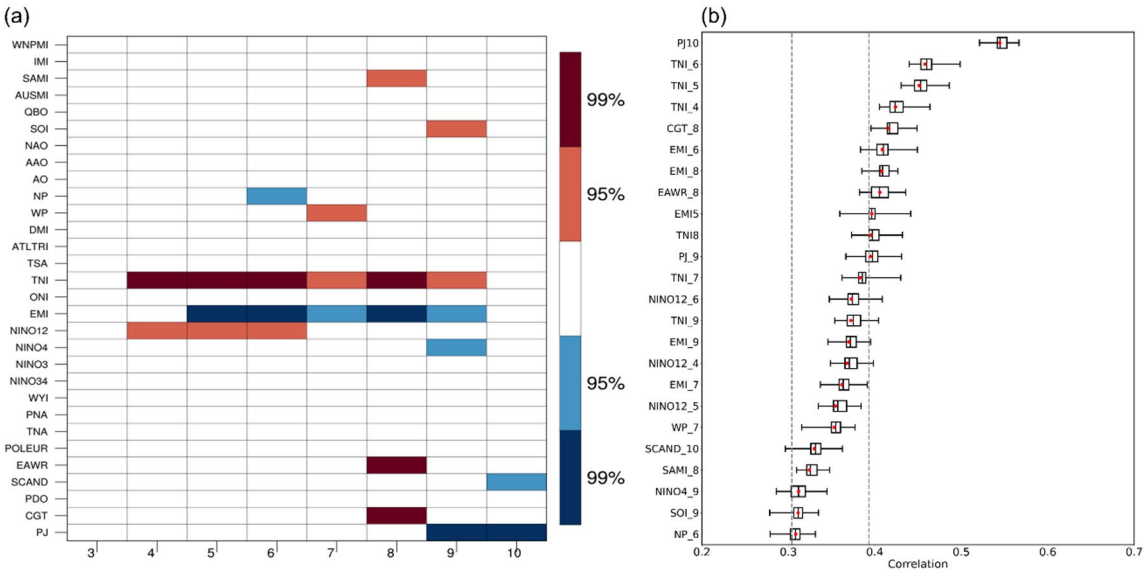

Fig. 1. (**a**) Statistical significance of correlation between observed winter (DJF) number of cold surge days (CSD) in South Korea and 30 climate indices. Red (blue) colors indicate positive (negative) correlation at 95% and 99% confidence levels. (**b**) Distributions of cross-validated correlation coefficients between observed CSD and selected climate indices. Red dots indicate the mean of each distribution. Gray dashed lines denote the 95% and 99% confidence levels.

exploited by the LSTM, and redundancy was mitigated by dropout regularization (p = 0.1), thereby capturing lead–lag nonlinear dependencies and potentially improving predictive skill. Because the sample comprises 42 winters, the fully nested cross-validation required for embedded feature discovery was judged impractical, and physically interpretable predictor set verified via LOYOCV was adopted, comprising 24 index–month combinations. The calculation methods for the climate indices are provided in Table S1.

Deep learning framework

In this study, we applied LSTM, a widely used deep learning technique for time-series prediction. LSTM is a type of RNN that mitigates the long-term dependency problem of conventional RNNs, enabling stable learning and prediction for extended time series[28,29]. LSTM is trained in a supervised learning framework using input–target pairs, and predictions are generated from the trained model.

As the input data, we used winter CSD predictions from 10 ensemble members of the PNU/RDA CGCM initialized in November, together with the 24 climate indices related to cold surges selected earlier. To evaluate prediction skill objectively, we employed two complementary validation strategies. First, LOYOCV was applied so that each winter was withheld once for validation while the remaining years were used for training, ensuring independence between training and test data[31]. Second, an independent validation period was performed, using 1980–2001 for training and 2002–2021 for validation, to assess whether the improvements were robust when the validation years were temporally separated from the training years[26]. Together, these approaches provided a comprehensive assessment of predictive performance.

Each model generated predictions through 100 repeated training runs with different random seeds. Hyperparameters were configured based on extensive testing and are detailed in Supplementary Table S2. The final architecture consisted of four stacked LSTM layers (100, 50, 30, and 20 hidden units) with dropout (0.1) on the first layer, followed by a dense output layer. Training used the Adam optimizer (learning rate 0.001), mean squared error loss, and batch size of 41. Early stopping (patience = 20 epochs) and dropout regularization were employed to prevent overfitting. All predictor variables were standardized using z-score normalization, and data shuffling was disabled to preserve temporal order. Detailed implementation settings, including software versions and computing environment, are provided in the Supplementary information (Table S2) to ensure reproducibility. Final validation was conducted based on the ensemble mean and distribution of the predictions, and the overall experimental procedure is presented in Fig. 2. Evaluation metrics included the Pearson correlation coefficient (CORR) and the root-mean-square error (RMSE). Statistical significance of individual Pearson correlations was assessed using a two-tailed Student's t-test at the 95% and 99% levels, and differences between dependent correlations (same samples) were evaluated with Steiger's test[44].

To quantitatively assess the relative contributions of the input variables, we applied SHAP, a widely used eXplainable Artificial Intelligence (XAI) method[45]. SHAP assigns an importance value to each predictor by computing the marginal contribution of that predictor to the model output across all possible feature coalitions, thereby providing an interpretable estimate of feature importance (e.g.,[26,46–48]). In this study, SHAP values were calculated for each predictor variable in the trained LSTM models. It is important to note that SHAP values provide model-based attributions of prediction skill, reflecting statistical associations rather than direct physical causality. To ensure stability, we repeated model training 100 times with different random seeds and averaged the absolute SHAP values across these runs, thereby reducing sensitivity to stochastic variations and yielding robust estimates of predictor importance. The resulting SHAP values were then aggregated into summary plots (Figs. 3 and 4) to visualize the distribution and ranking of predictor contributions.

To further evaluate the impact of input configuration on predictive performance, we designed three complementary experiments. EXP1 (the primary experiment) used both the CGCM predictions and the 24 climate indices as inputs, EXP2 used only the climate indices, and EXP3 used only the CGCM predictions. An overview of the experimental design is provided in Figure S2. This design allowed us to test whether improvements were consistently obtained across different predictor sets and to evaluate the complementarity of dynamical and statistical information.

Results

Cold surge prediction skill

We first evaluate the prediction skill of wintertime CSDs using the PNU/RDA CGCM and the LSTM developed in this study. Figure 3a presents a comparison between the observed CSD, the CGCM prediction, and the LSTM prediction. Figure 3b shows the same comparison normalized by climatology, while Fig. 3c summarizes the correlation coefficients (CORR) and the normalized root mean square error (nRMSE) for the two models over the entire period (ALL) and two periods, P1 from 1980 to 2000 and P2 from 2001 to 2021. The PNU/RDA CGCM exhibited limited skill, with a CORR of 0.33, statistically significant at the 95% confidence level, and an RMSE of 8.33 with an nRMSE of 1.16, consistent with previous findings[2]. By contrast, the LSTM improved skill substantially, achieving a CORR of 0.62 (99% significance), with RMSE reduced to 6.95 and nRMSE to 0.87, indicating enhanced correlation and reduced error. Using Steiger's test for dependent correlations, the improvement from correlation of 0.33 (CGCM) to 0.62 (LSTM) over 1980–2021 was statistically significant at the 99% confidence level[44].

For the two periods, CORR improved in P2 relative to P1 for both models, increasing from 0.21 to 0.41 for the CGCM and from 0.27 to 0.88 for the LSTM. The LSTM also showed a clear reduction in nRMSE, from 1.18 in P1 to 0.51 in P2. These results indicate that prediction skill was higher in the recent period (P2), with greater improvement in the LSTM compared to the CGCM. The superiority of the LSTM was especially evident during the mid-2000s, when CSD was below climatology, and in capturing the increased variability after the mid-2010s. Moreover, the decrease in CSD during 2019–2020 followed by an increase in 2021 was not reproduced by the CGCM, whereas the LSTM successfully captured this feature (Fig. 3b). Despite improvements in initialization

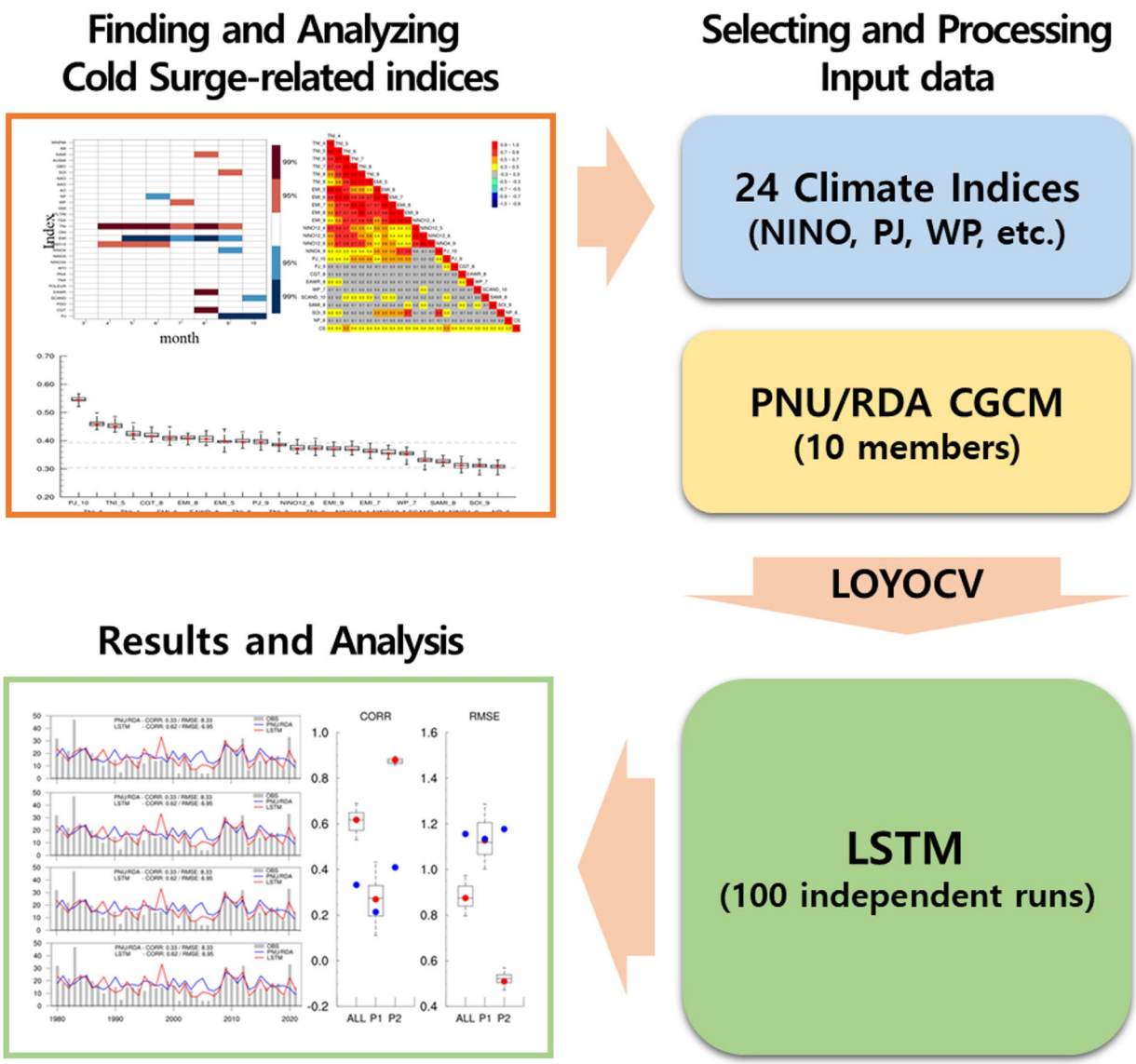

Fig. 2. Schematic diagram of the experimental workflow: (1) analyze 30 climate indices to identify those related to cold-surge days (CSD); (2) prepare inputs from the 24 selected index–month pairs and the 10-member PNU/RDA CGCM; (3) train the LSTM under leave-one-year-out cross-validation (LOYOCV), performing 100 runs and using the mean or distribution; (4) evaluate and diagnose model performance (e.g., CORR, RMSE).

and physics parameterizations, the CGCM still underestimates the temporal variability of CSD and often predicts the onset of cold surges inaccurately or misses events due to inherent systematic biases[2,19,20]. In contrast, the LSTM can learn cross-variable and cross-temporal dependencies that statistically compensate for these errors, suggesting complementary roles for dynamical and data-driven approaches.

To further assess robustness, we conducted an independent validation period in addition to LOYOCV. Training was performed with winters from 1980 to 2001 (22 years), and validation with winters from 2002 to 2021 (20 years). Figure 4a and b present the time series and normalized time series of the observed CSD and the LSTM-predicted CSD, while Fig. 4c illustrates the distribution of the performance metrics. In this experiment, the CORR of the LSTM predictions improved substantially from 0.40 to 0.88, while the RMSE decreased from 8.04 to 4.11 and the nRMSE decreased from 1.19 to 0.49, confirming consistent improvement across validation schemes. For the 2002–2021 validation period, the difference between correlation of 0.40 (CGCM) and 0.88 (LSTM) was also significant at the 99% confidence level[44]. Overall, the results from both LOYOCV and the independent validation period demonstrate that the LSTM-based prediction framework provides consistent and robust skill improvements over the CGCM, thereby overcoming limitations of conventional dynamical models in long-term CSD prediction.

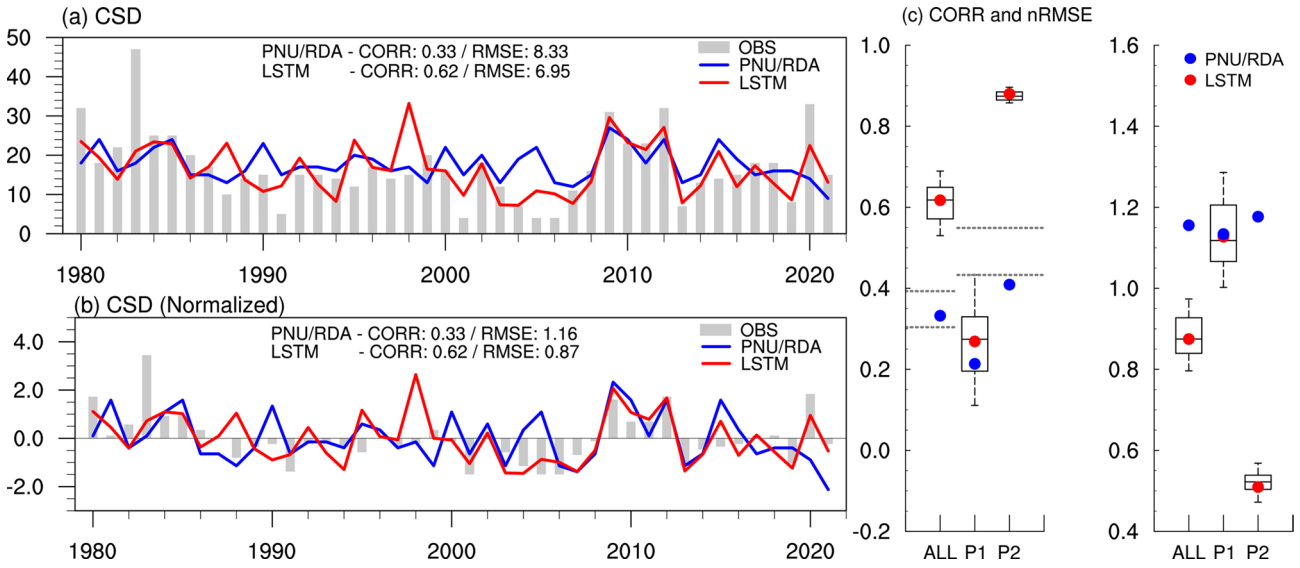

Fig. 3. (**a**) Time series of observed (gray bars), CGCM-predicted (blue line), and LSTM-predicted (red line) number of cold surge days (CSD). (**b**) Same as (a), but normalized by climatology. (**c**) Correlation coefficients (CORR) and normalized RMSE (nRMSE) during ALL (1980–2021), P1 (1980–2000), and P2 (2001–2021). Red and blue dots denote LSTM and CGCM mean values, respectively. Time series show the mean of 100 repeated runs. Boxplots summarize their distribution, with boxes representing the interquartile range and whiskers denoting the 5th–95th percentile. Horizontal dashed lines denote 95% and 99% significance levels.

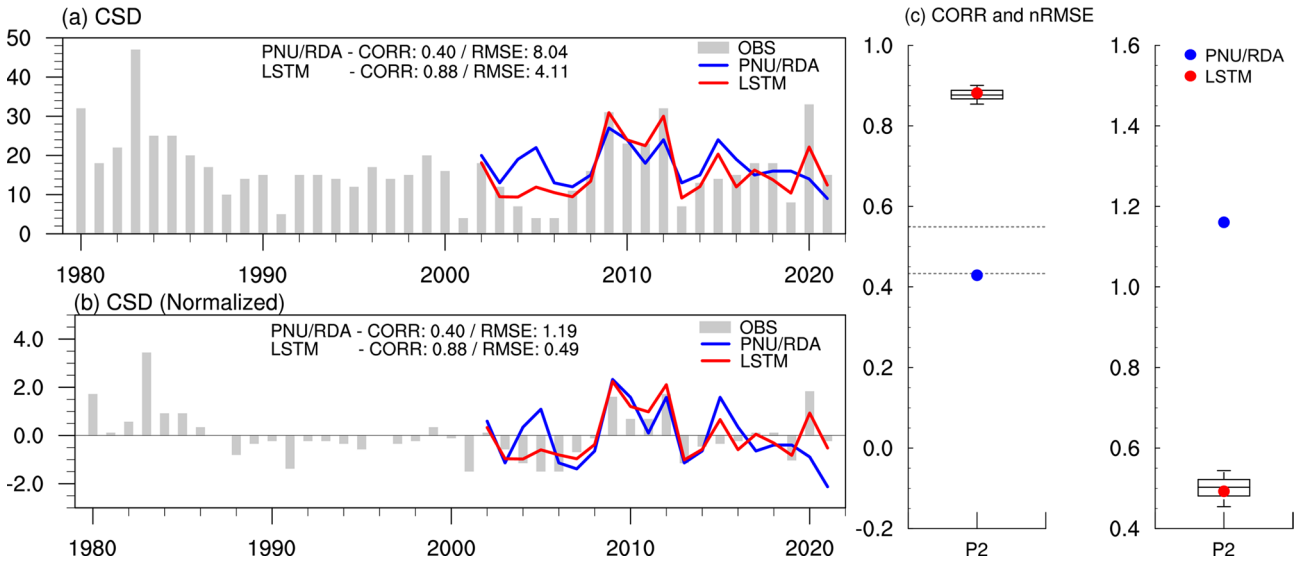

Fig. 4. (**a**) Time series of observed (gray bars), CGCM-predicted (blue line), and LSTM-predicted (red line) number of cold surge days (CSD) for the recent 20-year period (2002–2021). (**b**) Same as (a), but normalized by climatology. (**c**) Correlation coefficients (CORR) and normalized RMSE (nRMSE) during the 20-year period. Red and blue dots denote LSTM and CGCM mean values, respectively. Time series show the mean of 100 repeated runs. Boxplots summarize their distribution, with boxes representing the interquartile range and whiskers denoting the 5th–95th percentile. Horizontal dashed lines denote 95% and 99% significance levels.

Factors influencing changes in cold surge prediction skill

To identify the drivers of the improvement in LSTM-based CSD prediction skill, we applied SHAP analysis (Fig. 5). We averaged absolute SHAP values across repeated runs to obtain robust estimates of predictor importance and to avoid underestimation from cancellation of positive and negative contributions. The analysis revealed that SCAND_10, WP_7, and SOI_9 consistently exerted strong influences across the entire period (Fig. 5a–c). These teleconnection patterns are well established as key drivers of East Asian winter climate variability (e.g.,[41,49–52]). In particular, the SCAND pattern, a dominant mode over the North Atlantic–Eurasia sector, induces southward

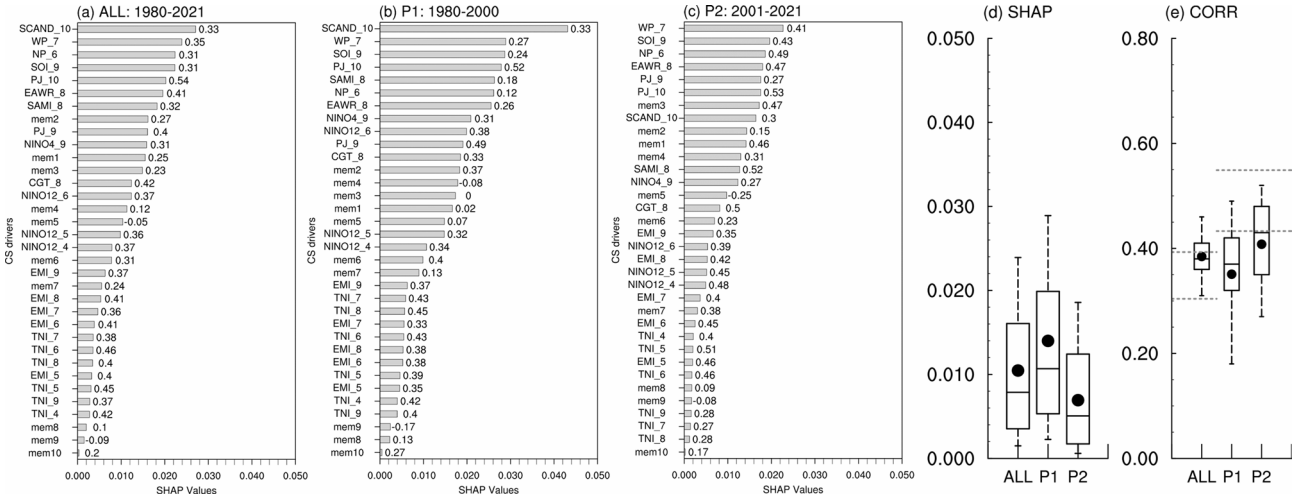

Fig. 5. Mean absolute SHapley Additive exPlanations (SHAP) values of input variables for number of cold surge days (CSD) prediction during (**a**) ALL (1980–2021), (**b**) P1 (1980–2000), and (**c**) P2 (2001–2021). (**d**) Distribution of SHAP values. (**e**) Correlation coefficients (CORR) between each input variable and observed CSD. Black dots denote mean values. Horizontal dashed lines denote 95% and 99% significance levels.

cold advection during its positive phase, triggering East Asian cold surges[49,50]. The WP pattern, characterized by a north–south dipole over the North Pacific, modulates winter temperature variability and the intensity of cold surges[41,49,51], while the SOI reflects ENSO-related sea level pressure differences across the tropical Pacific and influences East Asian winter temperature and precipitation[52–55].

SCAND_10 and WP_7 were the dominant predictors during P1, with SCAND_10 in particular showing stronger contributions than in the entire period (Fig. 5b). In contrast, WP_7 and SOI_9 exerted relatively greater influence during P2 (Fig. 5c–d). A comparison of the distributions of absolute SHAP values (Fig. 5d) shows that P1 had the highest median and mean, while P2 exhibited the lowest among all periods. These results indicate a shift in the dominant predictors of CSD between P1 and P2, reflecting changes in large-scale climate patterns. Interestingly, this contrasts with the correlations between CSD and the indices (Fig. 5e), where WP_7, SOI_9, and NP_6 were strongly correlated with CSD in P2 but showed limited SHAP contributions. This difference underscores that SHAP reflects the internal attribution of skill by the model, which may diverge from simple correlation measures, and suggests that in P2 the LSTM relied more on a subset of strongly correlated indices as a group, thereby reducing the apparent contribution of individual predictors.

To validate this interpretation, we conducted a Multiple Linear Regression (MLR) analysis using only the input indices to reconstruct the CSD. To address the multicollinearity evident in Figure S1, we used TNI_6 as the representative index among the highly pairwise-correlated TNI, EMI, and NINO indices. Accordingly, ten indices (TNI_6, PJ_9, PJ_10, CGT_8, EAWR_8, WP_7, SCAND_10, SAMI_8, SOI_9, and NP_6) were selected as input predictors for the MLR experiment. The results showed that the CSD time series could be accurately reconstructed using only the indices, with the coefficient of determination (R^2) reaching 0.87 in P2 (Fig. 6a). This supports the conclusion that the improvement in prediction skill in P2 is closely linked to the enhanced reproducibility of the indices. In addition, the indices were classified into atmospheric and oceanic groups, and their correlation coefficients with CSD were compared, as shown in Fig. 6b and c. Oceanic indices exhibited a relatively narrow correlation range in P1, but the spread widened in P2, indicating reduced consistency and greater uncertainty among individual variables (Fig. 6b). In contrast, the atmospheric indices generally displayed higher correlations and improved consistency in their distribution during P2 (Fig. 6c). These results suggest that the enhancement of CSD prediction skill in P2 was largely attributable to the increased contribution of atmospheric indices relative to oceanic indices. As confirmed earlier in Fig. 1b, this relative shift in importance from oceanic to atmospheric drivers is not simply a result of predictor selection, but represents a robust feature of the predictor–CSD relationship.

This interpretation is consistent with previous studies reporting that Arctic amplification has strengthened the Siberian High and altered jet stream patterns, thereby modifying the mechanisms of cold surges over East Asia (e.g.,[6]). Kumar[6] further demonstrated that, during the post-2000 period, the increased pressure difference between the Siberian High and the Aleutian Low has intensified the southward advection of cold air, while the westward shift or weakening of the jet stream has led to the prolonged persistence of cold surges. These circulation changes also directly affected the Korean Peninsula, with the prolonged cold surge in January–February 2008 and the extreme cold surge in January 2016 being closely associated with these atmospheric pattern changes. Overall, these results suggest that the increased role of atmospheric circulation patterns, shaped by Arctic amplification, has likely become an important factor influencing CSD predictability in recent decades. Consistent with this view, the spatial correlation between the CSDs in South Korea and 2-m temperature (Figure S3) exhibits significant correlations extending across a broad domain from Central Asia to the western North Pacific, covering much of East Asia including northeastern China and the Korean Peninsula. Because cold surges are typically associated with stationary Rossby wave trains, blocking events, and jet stream fluctuations, the

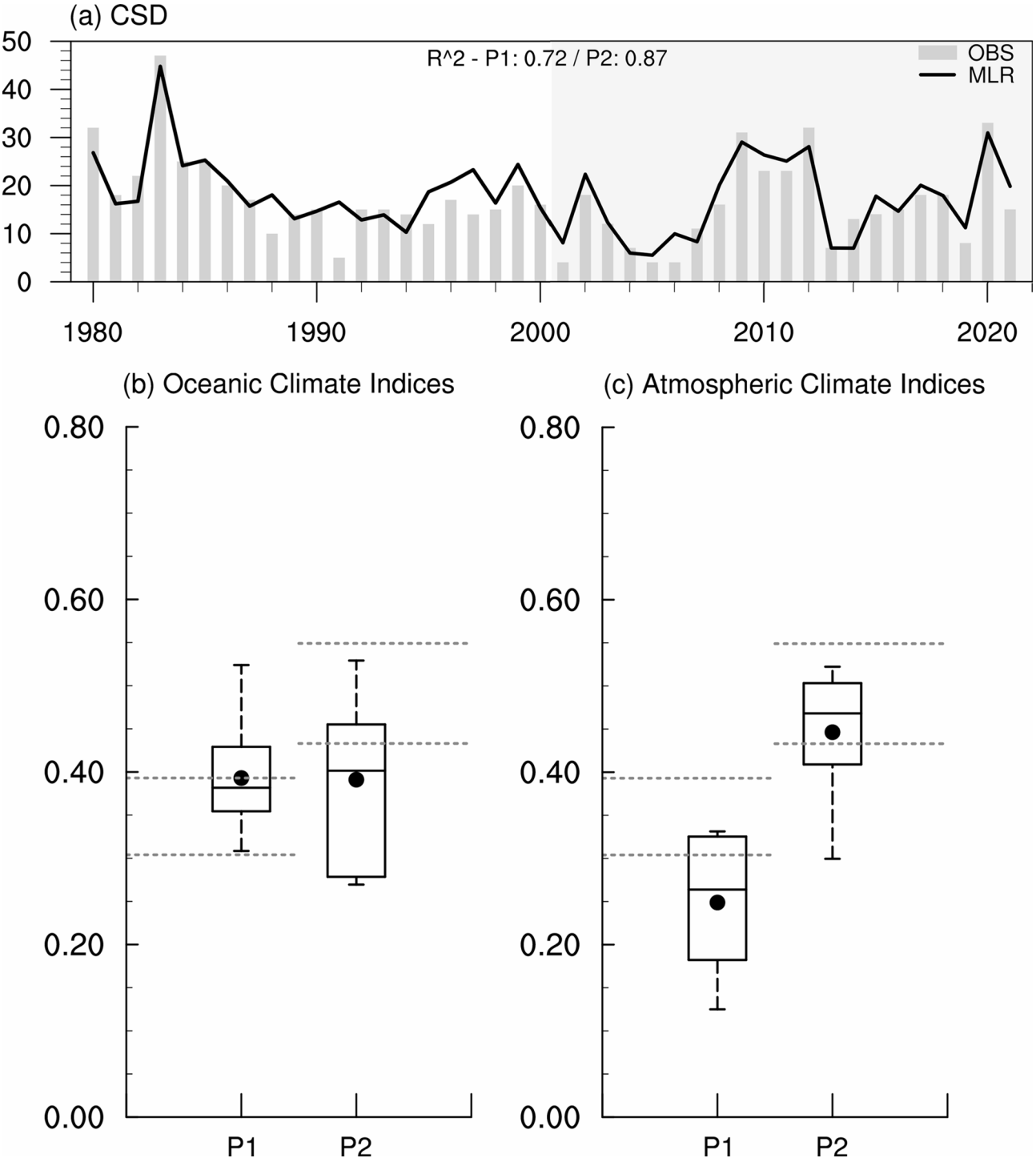

Fig. 6. (**a**) Time series of observed (gray bars) and MLR-reconstructed (black line) number of cold surge days (CSD) during ALL (1980–2021) periods. (b, c) Correlation coefficients between CSD and (**b**) oceanic and (**c**) atmospheric climate indices during P1 (1980–2000), and P2 (2001–2021). Black dots denote mean values. Horizontal dashed lines denote 95% and 99% significance levels.

mechanisms identified here can be reasonably interpreted within a broader East Asian context. This indicates that cold surges over South Korea are manifestations of larger-scale East Asian circulation anomalies rather than local variability, suggesting that the proposed hybrid prediction framework has potential applicability beyond South Korea.

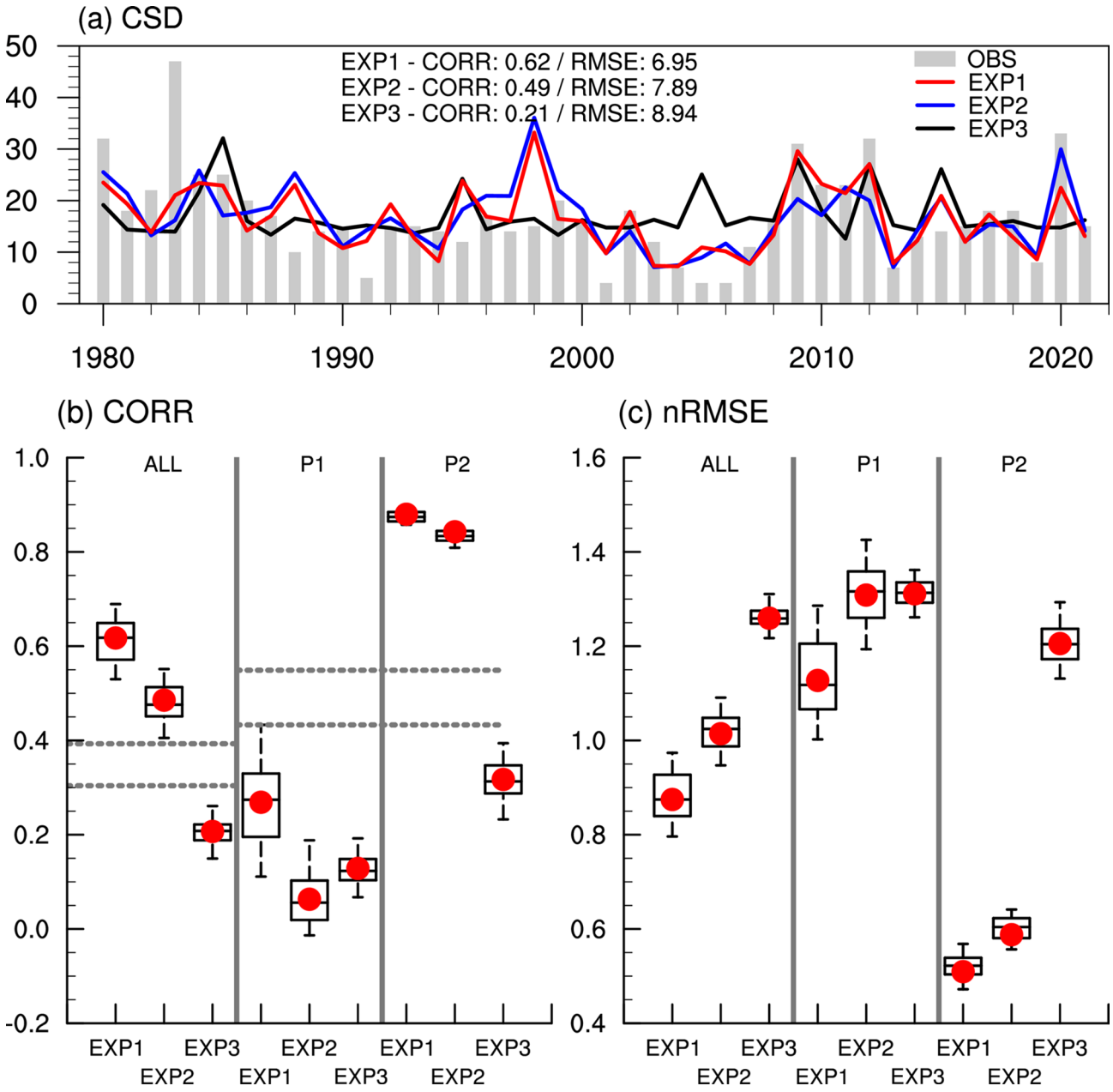

Fig. 7. (**a**) Time series of observed (gray bars) and predicted number of cold surge days (CSD) from three experiments: EXP1 (red; uses both climate indices and CGCM predictions), EXP2 (blue; uses indices only), and EXP3 (black; uses CGCM only). (**b**, **c**) Correlation coefficients (CORR) and normalized RMSE (nRMSE) during ALL (1980–2021), P1 (1980–2000), and P2 (2001–2021). Red dots denote mean values. Time series show the mean of 100 repeated runs. Boxplots summarize their distribution, with boxes representing the interquartile range and whiskers denoting the 5th–95th percentile. Horizontal dashed lines denote 95% and 99% significance levels.

Sensitivity experiment for cold surge prediction to input data

The performance of a deep learning–based prediction model is inherently sensitive to the type and configuration of input data. In particular, CGCM predictions and climate indices encapsulate distinct sources of climate information, and their combined use can provide complementary signals that reduce prediction uncertainties compared to using either source alone. To assess the sensitivity of the proposed CSD prediction model to input configuration, three experiments were conducted (Fig. 7). EXP1 combined CGCM-predicted CSD with the 24 climate indices, EXP2 used only the climate indices, and EXP3 relied solely on the CGCM predictions (Figure S2). In the time series comparison, EXP1 most accurately and consistently reproduced the observed CSD variability (Fig. 7a). It also showed the best performance across all metrics, achieving the highest CORR and the lowest RMSE over the entire period and in both P1 and P2 (Fig. 7b and c). Compared with EXP3, EXP2 yielded better results, confirming that climate indices provided more predictive information than the CGCM-only predictions.

Improvements were particularly pronounced during P2, when both EXP1 and EXP2 performed better relative to P1, consistent with the tendencies identified in Figs. 3, 4, 5 and 6. This suggests that the growing importance of atmospheric teleconnection indices in recent decades contributed substantially to enhanced prediction skill.

Over the entire period, neither EXP2 nor EXP3 was able to reproduce key features, such as the pronounced peaks in the early 1980s, the reduction of CSD in the mid-2000s, and the recent variability including the decline in 2019–2020. These features were captured only when dynamical model predictions were included in EXP1, highlighting the complementary value of combining dynamical and statistical information.

Finally, to benchmark the proposed model, we trained two MLR configurations under LOYOCV: one using a reduced predictor set of ten indices to address multicollinearity (with TNI/EMI/NINO clustered as TNI_6, Figure S4a) and another using all 24 selected indices (Figure S4b). The reduced MLR achieved CORR of 0.46, 0.80, and 0.64 for P1, P2, and ALL, respectively, whereas the full MLR yielded 0.30, 0.70, and 0.54. In both cases, the MLR skill was lower than that of the LSTM across all periods, indicating that purely linear mappings do not fully exploit the combined indices-CGCM information, while the LSTM benefits from learning cross-variable and cross-temporal dependencies.

Discussion and conclusion

This study developed an LSTM–based prediction framework that integrates CGCM predictions with diverse climate indices and achieved substantial improvements in the prediction skill of winter CSDs over South Korea. The input data to the LSTM consisted of 24 climate indices with statistically significant correlations with CSD, together with CSD predictions from 10 ensemble members of the PNU/RDA CGCM. The results demonstrated clear increases in correlation and reductions in prediction errors relative to the dynamical model, highlighting the enhanced prediction performance of the LSTM framework across both qualitative patterns and quantitative accuracy (Figs. 3 and 4). In particular, during the recent 21 years (P2: 2001–2021), CORR rose from 0.27 to 0.88, while nRMSE declined from 1.18 to 0.51, indicating a much greater improvement than in the earlier 21 years (P1: 1980–2000).

The SHAP analysis further revealed that climate indices from the mid–high latitudes of the Northern Hemisphere and the tropical Pacific, such as SCAND_10, WP_7, and SOI_9, were consistently influential across the entire period, although the dominant contributing indices varied between P1 and P2 (Fig. 5). A MLR experiment confirmed that the explanatory power of climate indices was much stronger in P2 than in P1, indicating that climate predictors played an increasingly important role in recent decades. Furthermore, separating the indices into atmospheric and oceanic groups showed that atmospheric indices exhibited stronger correlations with CSD than oceanic indices during P2 (Fig. 6). The apparent shift from oceanic to atmospheric indices after the early 2000s is consistent with previous findings linking Arctic amplification and jet stream changes to East Asian winter extremes[6,11]. Our results therefore provide statistical evidence that the relative importance of atmospheric predictors has increased in recent decades, in agreement with existing literature. Taken together, these findings suggest that the increasing role of atmospheric circulation has become a key factor shaping recent cold surges and their predictability.

A sensitivity experiment was also conducted to examine the role of input configuration (Fig. 7). The results revealed that the prediction skill was significantly higher when using the combined case of climate indices and dynamical model predictions as input data (EXP1), compared with cases that used only climate indices (EXP2) or the CGCM (EXP3). This highlights that integrating diverse climate indices with dynamical model predictions maximizes CSD prediction skill and underscores the value of strategically designing input configurations to further improve LSTM-based seasonal prediction.

Although the CGCM reasonably reproduces the spatial structures and statistical characteristics of cold surges, it tends to underestimate their temporal variability[2]. This underperformance likely arises from inherent systematic biases, as reported in previous studies[19,20]. Such limitations often manifest as phase biases, whereby the model tends to predict the onset of cold surges inaccurately or fails to capture them altogether. In contrast, the LSTM does not explicitly correct these biases. Instead, it learns cross-variable and cross-temporal dependencies among predictors, thereby statistically mitigating timing errors and other biases inherent to the CGCM. This contrast underscores the complementary roles of dynamical and data-driven approaches in the seasonal prediction of CSD.

It should be noted that the interpretability results presented here, particularly those based on SHAP and correlation analysis, reflect statistical associations rather than direct physical causality. The LSTM framework attributes prediction skill to input features in a statistical sense, and thus predictor contributions should not be interpreted as dynamical explanations. Despite these limitations, the framework presented here shows promise for extending to other extremes such as heatwaves and heavy rainfall. Future work should incorporate additional predictors (e.g., land surface and cryospheric variables) and test generalizability across regions, thereby advancing seasonal prediction capability.

Overall, these findings indicate that the LSTM-based framework not only improves prediction skill but also clarifies how the sources of cold surge predictability have evolved in recent decades. The robustness of the results, confirmed through multiple validation schemes and sensitivity experiments, highlights the reliability of the conclusions. In addition, spatial correlation patterns (Figure S3) suggest that cold surges over South Korea are embedded within broader East Asian circulation anomalies, indicating that the proposed framework may also be relevant beyond the Korean Peninsula. Such improvements have practical implications, as better seasonal cold surge forecasts can support agricultural risk management, energy demand planning, and disaster preparedness in South Korea and across East Asia. This framework may further be extended to other extreme climate events, with potential relevance for operational applications in agricultural planning, disaster preparedness, and policy decision-making.

Data availability

The data that support the findings of this study are available from the authors upon reasonable request.

Received: 12 September 2025; Accepted: 11 November 2025
Published online: 03 December 2025

References

1. IPCC & Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. in *Climate Change 2021: the Physical Science Basis*. 2391 (eds Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R. & Zhou, B.) (Cambridge University Press, 2021). https://doi.org/10.1017/9781009157896
2. Kim, E. S., Kryjov, V. N. & Ahn, J. B. Seasonal prediction and simulation of the cold surges over the Korean Peninsula using a CGCM. *Theor. Appl. Climatol.* **155**, 1793–1806. https://doi.org/10.1007/s00704-023-04731-7 (2024).
3. Kim, K. Y., Lee, S. & Kim, M. K. Long-term variability of cold surges in Korea. Asia-Pac. *J. Atmos. Sci.* **50**, 519–529. https://doi.org/10.1007/s13143-014-0041-6 (2014).
4. Ryoo, S. B., Kwon, W. T. & Jhun, J. G. Characteristics of wintertime daily and extreme minimum temperature over South Korea. *Int. J. Climatol.* **24**, 145–160. https://doi.org/10.1002/joc.990 (2004).
5. Heo, J. W. et al. Changes in cold surge occurrence over East Asia in the future: role of thermal structure. *Atmosphere* **9**, 222. https://doi.org/10.3390/atmos9060222 (2018).
6. Kumar, A. A new approach to cold surge classification in East Asia. *Sci. Rep.* **11**, 23659. https://doi.org/10.1038/s41598-021-02873-0 (2021).
7. He, S., Gao, Y., Li, F., Wang, H. & He, Y. Impact of Arctic Oscillation on the East Asian climate: A review. *Earth-Sci. Rev.* **164**, 48–62. https://doi.org/10.1016/j.earscirev.2016.10.014 (2017).
8. Jeong, J. H. & Ho, C. H. Changes in occurrence of cold surges over East Asia in association with Arctic Oscillation. *Geophys. Res. Lett.* **32**, L14704. https://doi.org/10.1029/2005GL023024 (2005).
9. Park, T. W., Ho, C. H. & Yang, S. Relationship between the Arctic Oscillation and cold surges over East Asia. *J. Clim.* **24**, 68–83. https://doi.org/10.1175/2010JCLI3529.1 (2011).
10. Woo, S. H., Kim, B. M., Jeong, J. H., Kim, S. J. & Lim, G. H. Decadal changes in surface air temperature variability and cold surge characteristics over Northeast Asia and their relation with the Arctic Oscillation during 1979–2011. *J. Geophys. Res. Atmos.* **117**, D18117. https://doi.org/10.1029/2011JD016929 (2012).
11. Cohen, J. et al. Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* **7**, 627–637. https://doi.org/10.1038/ngeo2234 (2014).
12. Kim, B. M. et al. Weakening of the stratospheric Polar vortex by Arctic sea-ice loss. *Nat. Commun.* **5**, 4646. https://doi.org/10.1038/ncomms5646 (2014).
13. Park, T. W., Ho, C. H. & Deng, Y. A synoptic and dynamical characterization of wave-train and blocking cold surge over East Asia. *Clim. Dyn.* **43**, 753–770. https://doi.org/10.1007/s00382-013-1817-6 (2014).
14. Park, T. W., Ho, C. H., Jeong, J. H., Heo, J. W. & Deng, Y. A new dynamical index for classification of cold surge types over East Asia. *Clim. Dyn.* **45**, 2469–2484. https://doi.org/10.1007/s00382-015-2483-7 (2015).
15. Wang, L. et al. Effect of the climate shift around the mid-1970s on the relationship between wintertime ural blocking circulation and East Asian climate. *Int. J. Climatol.* **30**, 153–158. https://doi.org/10.1002/joc.1876 (2010).
16. Kim, E. S. & Ahn, J. B. Study on the classification and characteristics of cold surge in South Korea. *Int. J. Climatol.* **43**, 720–735. https://doi.org/10.1002/joc.7820 (2023).
17. Dai, G. & Mu, M. Arctic influence on the East Asian cold surge forecast: A case study of January 2016. *J. Geophys. Res. Atmos.* 125, e2020JD033298 (2020). https://doi.org/10.1029/2020JD033298
18. Lin, R., Dong, X., Zhang, H., Wu, C. & Jin, J. Simulation of the boreal winter East Asian cold surge by IAP AGCM4.1. *Atmosphere* **13**, 1176. https://doi.org/10.3390/atmos13081176 (2022).
19. Scaife, A. A. & Smith, D. A signal-to-noise paradox in climate science. *Npj Clim. Atmos. Sci.* **1**, 28. https://doi.org/10.1038/s41612-018-0038-4 (2018).
20. Shepherd, T. G. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708. https://doi.org/10.1038/ngeo2253 (2014).
21. Luo, X. & Wang, B. How predictable are winter extremely cold days over temperate East asia? *Clim. Dyn.* **48**, 2557–2568. https://doi.org/10.1007/s00382-016-3222-4 (2017).
22. Luo, X. & Wang, B. Predictability and prediction of the total number of winter extremely cold days over China. *Clim. Dyn.* **50**, 1769–1784. https://doi.org/10.1007/s00382-017-3720-z (2018).
23. Siegert, S., Stephenson, D. B., Sansom, P. G. & Scaife, A. A. A bayesian framework for verification and recalibration of ensemble forecasts: how uncertain is NAO predictability? *J. Clim.* **29**, 995–1012. https://doi.org/10.1175/JCLI-D-15-0196.1 (2016).
24. Vigaud, N., Robertson, A. W. & Tippett, M. K. Multimodel ensembling of subseasonal precipitation forecasts over North America. *Mon Weather Rev.* **145**, 391–403. https://doi.org/10.1175/MWR-D-16-0037.1 (2017).
25. Lee, J. H. & Ahn, J. B. A new statistical correction strategy to improve long-term dynamical prediction. *Int. J. Climatol.* **39**, 2173–2185. https://doi.org/10.1002/joc.5943 (2019).
26. Lee, Y. et al. I. Unveiling teleconnection drivers for heat-wave prediction in South Korea using explainable artificial intelligence. *Npj Clim. Atmos. Sci.* **7**, 176. https://doi.org/10.1038/s41612-024-00722-1 (2024).
27. Seo, G. Y. & Ahn, J. B. Comparison of bias-correction methods for summertime daily rainfall in South Korea using quantile mapping and a machine-learning model. *Atmosphere* **14**, 1057. https://doi.org/10.3390/atmos14071057 (2023).
28. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).
29. Sherstinsky, A. Fundamentals of recurrent neural networks and long short-term memory networks. *Phys. D.* **404**, 132306. https://doi.org/10.1016/j.physd.2019.132306 (2020).
30. Yuan, S., Luo, X., Mu, B. & Li, J. Prediction of the North Atlantic Oscillation index with convolutional LSTM based on ensemble empirical mode decomposition. *Atmosphere* **10**, 252. https://doi.org/10.3390/atmos10050252 (2019).
31. Kim, H., Ham, Y. G., Joo, Y. S. & Son, S. W. Deep learning for bias correction of MJO prediction. *Nat. Commun.* **12**, 3087. https://doi.org/10.1038/s41467-021-23406-3 (2021).
32. Chung, J., Lee, Y., Jang, W., Lee, S. & Kim, S. Correlation analysis between air temperature and MODIS land surface temperature and prediction of air temperature using tensorflow long short-term memory for the period of cold and heat waves. *Remote Sens.* **12**, 3231. https://doi.org/10.3390/rs12193231 (2020).
33. Ahn, J. B. & Lee, J. Comparative study on the seasonal predictability dependency of boreal winter 2-m temperature and sea surface temperature on CGCM initial conditions. *Atmosphere* **25**, 353–366. https://doi.org/10.14191/ATMOS.2015.25.2.353 (2015).
34. Kiehl, J. T. et al. *Description of the NCAR Community Climate Model (CCM3).* NCAR Tech. Note NCAR/TN-420 + STR, National Center for Atmospheric Research, (Boulder, CO, USA., 1996).

35. Bonan, G. B. The land surface climatology of the NCAR land surface model coupled to the NCAR community climate model. *J. Clim.* **11**, 1307–1326. https://doi.org/10.1175/1520-0442(1998)011%3C1307:TLSCOT%3E2.0.CO;2 (1998).
36. Pacanowski, R. C. & Griffies, S. M. *MOM 3.0 Manual* p 08542 (NOAA/Geophysical Fluid Dynamics Laboratory, 1998).
37. Ahn, J. B. & Lee, J. A. Numerical study on the role of sea-ice using ocean general circulation model. *Sea: J. Korean Soc. Oceanogr.* **6**, 225–233 (2001). (In Korean with English abstract) https://koreascience.kr/article/JAKO200120417585143.page
38. Jo, S. & Ahn, J. B. Improvement of CGCM prediction for wet-season precipitation over the maritime continent using a bias-correction method. *Int. J. Climatol.* **35**, 3721–3732. https://doi.org/10.1002/joc.4232 (2015).
39. Bayasgalan, G. & Ahn, J. B. Seasonal prediction of high-resolution 2-m temperature over Mongolia during boreal winter using both coupled general circulation model and artificial neural network. *Int. J. Climatol.* **38**, 5418–5429. https://doi.org/10.1002/joc.5848 (2018).
40. Kim, G. et al. Assessment of MME methods for seasonal prediction using WMO LC-LRFMME hindcast dataset. *Int. J. Climatol.* **41**, E2462–E2481. https://doi.org/10.1002/joc.6858 (2021).
41. Kim, E. S., Kryjov, V. N. & Ahn, J. B. Predictability of the wintertime Western Pacific pattern in the APEC climate center multi-model ensemble. *Atmosphere* **13**, 1772. https://doi.org/10.3390/atmos13111772 (2022).
42. Lee, J., Lee, M. I. & Ahn, J. B. Importance of ocean initial conditions of late autumn on winter seasonal prediction skill in an atmosphere–land–ocean–sea-ice coupled forecast system. *Clim. Dyn.* **58**, 3427–3440. https://doi.org/10.1007/s00382-021-06106-y (2022).
43. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R Meteorol. Soc.* **146**, 1999–2049. https://doi.org/10.1002/qj.3803 (2020).
44. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251. https://doi.org/10.1037/0033-2909.87.2.245 (1980).
45. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30** https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (2017).
46. Gao, J., Meng, Q., Zhang, L. & Hu, D. How does the ambient environment respond to industrial heat Island effects? An innovative and comprehensive methodological paradigm for quantifying the varied cooling effects of different landscapes. *GISci Remote Sens.* **59**, 1643–1659. https://doi.org/10.1080/15481603.2022.2127463 (2022).
47. Kang, Y., Kim, M., Kang, E., Cho, D. & Im, J. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia. ISPRS J. Photogramm. *Remote Sens.* **183**, 253–268. https://doi.org/10.1016/j.isprsjprs.2021.11.016 (2022).
48. Lv, Z., Nunez, K., Brewer, E. & Runfola, D. Mapping the tidal marshes of coastal virginia: A hierarchical transfer learning approach. *GISci Remote Sens.* **61**, 2287291. https://doi.org/10.1080/15481603.2023.2287291 (2024).
49. Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon Weather Rev.* **115**, 1083–1126. https://doi.org/10.1175/1520-0493(1987)115%3C1083:CSAPOL%3E2.0.CO;2 (1987).
50. Bueh, C. & Nakamura, H. Scandinavian pattern and its Climatic impact. *Q. J. R Meteorol. Soc.* **133**, 2117–2131. https://doi.org/10.1002/qj.173 (2007).
51. Wallace, J. M. & Gutzler, D. S. Teleconnections in the geopotential height field during the Northern hemisphere winter. *Mon Weather Rev.* **109**, 784–812. https://doi.org/10.1175/1520-0493(1981)109%3C0784:TITGHF%3E2.0.CO;2 (1981).
52. Wang, B., Wu, R. & Fu, X. Pacific–East Asian teleconnection: how does ENSO affect East Asian climate? *J. Clim.* **13**, 1517–1536. https://doi.org/10.1175/1520-0442(2000)013%3C1517:PEATHD%3E2.0.CO;2 (2000).
53. Trenberth, K. E. Signal versus noise in the Southern Oscillation. *Mon Weather Rev.* **112**, 326–332. https://doi.org/10.1175/1520-0493(1984)112%3C0326:SVNITS%3E2.0.CO;2 (1984).
54. Trenberth, K. E. The definition of El Niño. *Bull. Am. Meteorol. Soc.* **78**, 2771–2777. https://doi.org/10.1175/1520-0477(1997)078%3C2771:TDOENO%3E2.0.CO;2 (1997).
55. Troup, A. J. The Southern Oscillation. *Q. J. R Meteorol. Soc.* **91**, 490–506. https://doi.org/10.1002/qj.49709139009 (1965).

Acknowledgements

This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01746002)" Rural Development Administration, Republic of Korea.

This study was supported by 2025 the RDA Fellowship Program of National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea.

Author contributions

Conceptualization, J. Lee and E. S. Kim; Data Curation, E. S. Kim; Formal analysis, E. S. Kim, J. Lee, J. Hur, S. Jo, and Y. S. Kim; Investigation, E. S. Kim, J. Hur, S. Jo, and Y. S. Kim; Methodology, J. Lee, and E. S. Kim; Resources, J. Hur, and K. M. Shim; Software, E. S. Kim; Supervision, J. Lee; Validation, J. Lee, K. M. Shim, and J. B. Ahn; Visualization, E. S. Kim; Writing—original draft preparation, E. S. Kim; Writing—review and editing, J. Lee, and E. S. Kim. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Supplementary Information The online version contains supplementary material available.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-28608-z.

Correspondence and requests for materials should be addressed to J.L.