npj | climate and atmospheric science

Article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-025-01311-6

A deep learning-based land-atmosphere coupled model for heatwave prediction

Check for updates

Dongjin Cho[1], Yoo-Geun Ham[1,2] ✉, Suyeon Jeong[1] & Seon-Yu Kang[1]

Extreme heatwaves are intensifying under climate change, yet their prediction remains limited by inadequate representation of land–atmosphere (L–A) interactions. Most deep learning–based weather models rely solely on atmospheric variables, overlooking the influence of land surface conditions on heat extremes. Here, we present an L–A coupled prediction framework for Northern Hemisphere summer that incorporates multi-layer soil moisture (SM) and temperature into atmospheric forecasting. To better capture delayed land surface feedbacks, the model is trained with a multi-step loss. This approach improved the representation of L–A interactions across 1–7 day lead times. Using multi-step loss, the L–A coupled model achieved a 5.9–11.2% improvement in heatwave forecast accuracy relative to the atmosphere-only model, as measured by root mean squared error, whereas single-step loss achieved only 0.4–2.4% improvement. Skill gain was strongest at short leads (~ 3 day) when both SM and circulation predictability were high, and sustained through 7 days by L–A coupling driven by SM predictability. Case studies of recent heatwaves further demonstrated its ability to capture land surface drying and associated temperature extremes. These findings underscore the importance of incorporating L–A coupling with multi-step optimization for advancing data-driven heatwave prediction.

With ongoing climate change, the frequency, intensity, and duration of heatwaves are projected to increase globally, posing escalating risks to human health, ecosystems, and infrastructure[1,2]. In particular, compound extreme events, such as concurrent heatwaves and droughts, are becoming more frequent and disruptive[3,4]. A notable example is the extreme 2018 heatwave over Western Europe, which caused record-breaking temperatures and substantial socio-economic damage across France, Germany, and other regions[5–7]. These events highlight the urgent need for more reliable heatwave prediction systems to support early warning and impact mitigation efforts.

In recent decades, high-performance computing has significantly advanced numerical weather prediction (NWP) systems[8,9]. These models simulate atmospheric evolution by solving partial differential equations on discretized spatial grids. However, they remain computationally intensive, often requiring hours on supercomputers to generate a single 10-day forecast[10]. Furthermore, unresolved sub-grid processes are represented through parameterizations, which introduce approximation errors that degrade forecast quality[11,12]. These limitations contribute to the underestimation of extreme event intensity and insufficient representation of the nonlinear feedback mechanisms that initiate and sustain heatwaves, posing a significant challenge for reliable forecasting[13,14].

Deep learning-based approaches have shown remarkable progress in weather forecasting in recent years, driven by advances in data availability, computational power, and neural network architectures[15]. Several large-scale models—most notably Pangu-Weather, GraphCast, and FuXi—have demonstrated that data-driven systems can serve as viable alternatives to traditional NWP models[16–22]. These models leverage historical reanalysis data and spatiotemporal learning to forecast future atmospheric states directly, bypassing the need for explicit physical simulations. Consequently, they can generate global forecasts within seconds while achieving superior forecast skill than those of state-of-the-art dynamical systems[23,24]. Despite their success, most existing efforts have concentrated primarily on atmospheric variables, with limited integration of land–atmosphere (L–A) feedback, an essential component for reliable heatwave prediction.

L–A coupling is critical in developing and persisting extreme heat events. A central mechanism is the soil-moisture–controlled land surface energy flux partitioning[25,26]. When soil moisture is depleted, latent heat flux through evapotranspiration is suppressed, and a greater portion of surface energy is converted into sensible heat. This intensifies near-surface air temperatures, which increases evaporative demand and accelerates further soil drying, forming a positive feedback loop that amplifies and sustains heatwaves[27,28]. Numerous observational and modeling studies have

---

[1]Environmental Planning Institute, Seoul National University, Seoul, Republic of Korea. [2]Department of Environmental Managements, Graduate School of Environmental Studies, Seoul National University, Seoul, Republic of Korea. ✉e-mail: yoogeun@snu.ac.kr

identified this feedback as a key driver of extreme heat events across various climatic regimes[6,29,30]. These findings suggest that L–A interactions do not emerge instantaneously; rather, they develop progressively over time through accumulated feedback involving surface energy fluxes, radiative processes, and boundary-layer dynamics[25,26,29]. Because atmospheric states evolve rapidly while land surface variables change more gradually and retain memory over extended periods[31–33], representing L–A coupling in prediction models is important to enhance the physical realism and predictive capability of heatwave forecasting systems.

This study aimed to develop a deep learning-based L–A coupled model for heatwave prediction, with a particular focus on capturing the evolving dynamics of L–A interactions across multiple lead times. The proposed framework incorporated atmospheric and land surface variables as input and employed an autoregressive forecasting structure optimized through multi-step learning, which updates the model using errors from multiple lead times. To assess the impact of multi-step optimization, we compared the proposed coupled model with an identical coupled architecture trained using single-step learning, which optimizes only the next-day prediction. Additionally, we evaluated the impact of L–A coupling itself by comparing the coupled model with an atmosphere-only baseline, highlighting the novelty of integrating autoregressive multi-step learning with L–A coupling for improved heatwave prediction.

Results

Assessment of simulated L–A coupling strength

Figure 1 illustrates the spatial distribution of L–A coupling strength, calculated as the Pearson correlation between daily anomalies of top-layer soil moisture (SM, Lev 1) and two key atmospheric variables—daily maximum temperature (Tmax) and latent heat flux (LH)—during June to August from 2018 to 2023. The observational L–A coupling strength revealed physically consistent patterns, with strong negative SM–Tmax correlations across key continental regions (e.g., central United States, central Europe, and southeastern Asia), where reduced soil moisture limits latent heat flux and enhances sensible heating, thereby intensifying temperatures[34–36]. The SM–LH coupling field similarly exhibited strong positive correlations across vegetated midlatitude areas, consistent with evapotranspiration-limited regimes where drier soils reduce latent heat flux[26]. These physically grounded patterns provide a robust reference for evaluating model performance. Notably, coupling strength was weaker across the Southern Hemisphere (e.g., South America, southern Africa, and Australia) due to reduced L–A interactions during the austral winter.

We compared spatial coupling strength derived from 7-day forecasts of the coupled model under multi-step and single-step optimization (Fig. 2). The multi-step model more faithfully reproduced the observed coupling patterns across known transition zones, capturing both the magnitude and spatial coherence of SM–Tmax and SM–LH relationships. In contrast, the single-step model consistently underestimated coupling strength, particularly in the SM–LH field, reflecting a limited capacity to represent lagged soil influence on surface fluxes. The results are consistent with other forecast lead times, while they are less pronounced in short lead times (Figs. S3 and S4). These results demonstrate that multi-step training better preserves physically meaningful soil–atmosphere feedback relevant to heatwave development, particularly by maintaining the delayed influence of soil conditions over extended forecasts.

Evaluation of forecast skill in the L–A coupled model

Figure 2 summarizes forecast skill under heatwave conditions, defined as grid points exceeding the 90th percentile of Tmax from June to August from 2018 to 2023 in the Northern Hemisphere. Across all models, forecast accuracy gradually declined with increasing lead time, as expected. While performance was comparable at shorter lead times, models trained with multi-step loss exhibited notably greater skill as lead time increased, achieving higher anomaly correlation coefficient (ACC) and lower root mean squared error (RMSE) than models trained with single-step loss. This finding is consistent with previous studies demonstrating that multi-step optimization improves the learning of temporal evolution patterns across consecutive forecast steps, thereby enhancing the accuracy of the forecasts at the extended lead times[37,38].

More importantly, the coupled model consistently outperformed the atmosphere-only baseline regardless of training strategy, underscoring the importance of incorporating land surface processes for reliable heatwave prediction. It should be highlighted that, over the 1–7 day forecast window, the coupled model trained with single-step loss achieved moderate skill improvements compared to the atmosphere-only model, ranging from 0.4% to 2.4% (Fig. 2c). In contrast, the coupled model achieved substantially higher improvement in skill scores with multi-step training, ranging from 5.9% to 11.2%, with the largest gains (exceeding 10%) observed at lead times of 2 to 4 days. These results highlight that multi-step optimization captures

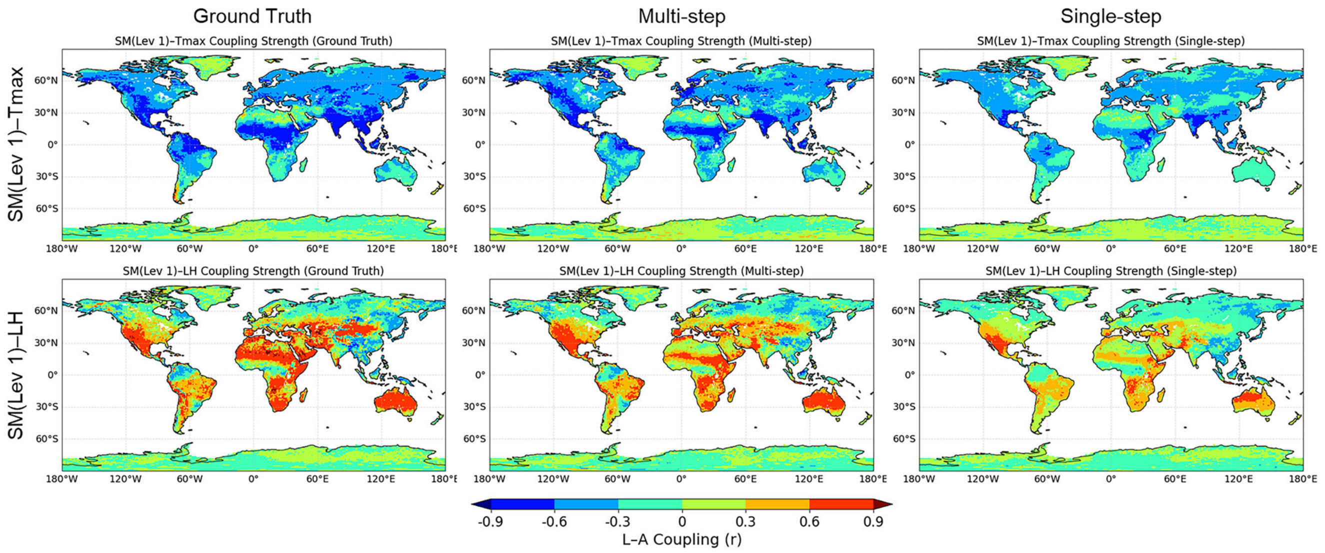

Fig. 1 | Spatial distribution of land–atmosphere (L–A) coupling strength during June–August (JJA) of 2018–2023. Coupling strength is defined as the Pearson correlation between daily anomalies of soil moisture at the first layer (SM, Lev 1) and (top) maximum temperature (Tmax), and (bottom) latent heat flux (LH). Results are shown for (left) observational data (Ground Truth), and for the coupled model trained with (middle) multi-step and (right) single-step loss. Coupled model results correspond to 7-day lead forecasts.

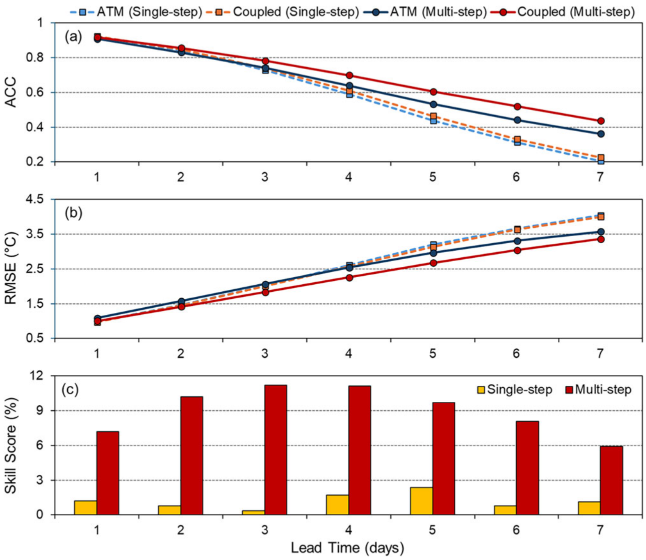

Fig. 2 | Forecast performance under heatwave conditions, defined as grid points exceeding the 90th percentile of daily maximum temperature (Tmax) during June–August (JJA) from 2018 to 2023 in the Northern Hemisphere. Results are presented for the atmosphere-only models (ATM) and the land–atmosphere coupled models (coupled), each trained using either single-step or multi-step loss functions. Panels depict **a** anomaly correlation coefficient (ACC) and **b** root mean square error (RMSE) across forecast lead times, and **c** the skill score of the coupled model relative to the ATM baseline, computed separately for each training strategy. Dashed lines represent models trained with the single-step loss function, and solid lines represent those trained with the multi-step loss function.

delayed L–A feedback more effectively, enabling the model to better leverage land state (e.g., SM) to improve heatwave prediction.

Next, to understand the predictable source of the coupled model, we compared the lead-time–dependent spatial distribution of the Tmax skill improvement under heatwave conditions to the forecast skill of the SM (Lev 1) (Fig. 3). At 1-day lead time, the spatial pattern of skill score of Tmax was largely noisy, while the coupled model's SM ACC exhibited uniformly high and stable values, with no discernible correlation between the two. In contrast, coherent regions of both Tmax skill improvement and SM ACC emerged as lead increased, and a clear positive correlation was established: at 3-day lead, higher SM ACC consistently corresponded to higher skill scores, with a Pearson correlation coefficient (r) of 0.81 and the steepest positive slope (Fig. 3c). Beyond this peak at 3-day lead, the slope of the relationship diminished with increasing lead time, but significant positive correlations still remain at 5-day ($r = 0.79$) and 7-day ($r = 0.73$). As large-scale atmospheric variables strongly influence heatwave occurrence, we also evaluated the association between Z500 forecast skill and the coupled model's Tmax skill score under heatwave conditions (Fig. S5). A positive correlation was found only at 3-day lead, when large-scale circulation anomalies remain relatively well predicted. Beyond this lead time, however, the association largely vanished, and no meaningful relation with the coupled model's Tmax skill score was detected. These findings suggest that the coupled model derives the strongest forecast skill gains at short leads (around 3 days), where both SM and large-scale circulation skill are high and act in a complementary way. At longer leads, however, as atmospheric forecast skill declines, the advantage is sustained primarily by L–A coupling which mainly involves surface variables (i.e., Tmax and SM).

The overall performance of the L–A coupled model was comprehensively evaluated in terms of its predictive skill across a wide range of meteorological and land surface variables and forecast lead times, relative to the atmosphere-only model (Figs. 4 and S6, S7). The coupled model showed skillful forecasts (ACC ≥ 0.3) up to approximately 7–10 days for most atmospheric variables, including temperature, pressure, and surface fluxes, except for the precipitation (Figs. 4a and S6). Notably, land surface variables (i.e., SM (Lev 1)) maintained high ACC values exceeding 0.8 up to a week forecast, and remained skillful even at 21 days, reflecting their strong autocorrelation and long memory characteristics[26,39]. The implies that the inherent long persistency of land surface states offers complementary predictability, contributing to improved medium-range forecasts in coupled models.

To quantify the added value of coupling, Fig. 4b presents ACC differences between the coupled and atmosphere-only models across lead times. The coupled model generally outperformed the atmosphere-only baseline across most variables and within the skillful forecast range (up to 7–10 days), with particularly large improvements for near-surface variables such as Tmax, daily minimum temperature (Tmin), LH, and sensible heat flux (SH). Spatial analysis of coupling impacts (Fig. 4c) further revealed widespread improvements in ACC for key heatwave-related variables (Tmax, LH, and geopotential at 500 hPa (Z500)) at a 7-day lead time, indicating that the coupled model effectively captures L–A interactions across diverse geographic and climatic regimes. However, as shown in Fig. S7, forecast skill improvement for humidity-related variables remained comparatively limited; this may be attributed to their stronger dependence on free-atmospheric dynamics and vertical mixing processes, which are less directly influenced by land surface forcing[40].

To demonstrate the improvement in the temporal evolution of individual heatwave event, a heatwave event over Western Europe in 2018, which was one of the most devastating extreme heat events[5,7], was compared (Fig. 5). This event is known to be characterized by persistently high temperatures co-occurring with anomalously dry soil conditions and the resultant increased sensible heat, which amplified the intensity and duration of the heatwave through strong L–A feedbacks[6,41]. These conditions were confirmed by observed temperature and soil moisture anomalies that deviated substantially from climatology.

Both the atmosphere-only and coupled models adequately captured mid-tropospheric dynamics, as indicated by their accurate prediction of Z500 (Fig. 5a). However, notable differences emerged for the surface variables. The coupled model reasonably captured the exceptionally dry soil moisture conditions (SM, Lev 1) observed during the heatwave (Fig. 5b), which in turn enabled a more realistic simulation of surface energy fluxes; the coupled model captured the observed increase in SH and decrease LH more realistically than the atmosphere-only model (Fig. 5c, d), though it showed a slight overestimation early in the heatwave period. In contrast, the atmosphere-only model underestimated both fluxes due to its inadequate representation of surface drying and associated energy partitioning. This improved representation of land surface processes in the coupled model led to more realistic evolution of near-surface temperatures: while the atmosphere-only model exhibited a cold bias by underestimating Tmax and Tmin during the peak of the heatwave, the coupled model captured their evolution and temporal variation more accurately (Fig. 5e, f).

The 2022 East Asian heatwave was additionally examined to evaluate the performance of the coupled and atmosphere-only models under extreme heatwave conditions (Fig. 6). Unlike the Western European case, this event was predominantly atmosphere-driven, with an amplified Z500 ridge initiating the extreme conditions (Fig. 6a) and subsequently inducing enhanced surface fluxes and soil moisture depletion (Fig. 6b–d). Here, soil drying acted as a consequence rather than a cause of the heatwave. Despite these atmosphere-dominated conditions, the coupled model outperformed the atmosphere-only model in reproducing the temporal evolution of near-surface temperatures, capturing both Tmax and Tmin more accurately (Fig. 6e, f). At the same time, the coupled model reasonably reproduced the progression of soil moisture depletion (Fig. 6b); however, the magnitude was underestimated, consistent with the muted simulation of enhanced SH, LH, and peak temperatures (Tmax and Tmin) (Fig. 6c–f). These results highlight the benefits of L–A coupling for enhancing heatwave prediction skill across different types of extremes. While it improves the representation of soil drying and L–A feedbacks in land-driven events, it also reduces temperature biases and enables the simulation of concurrent drought evolution under

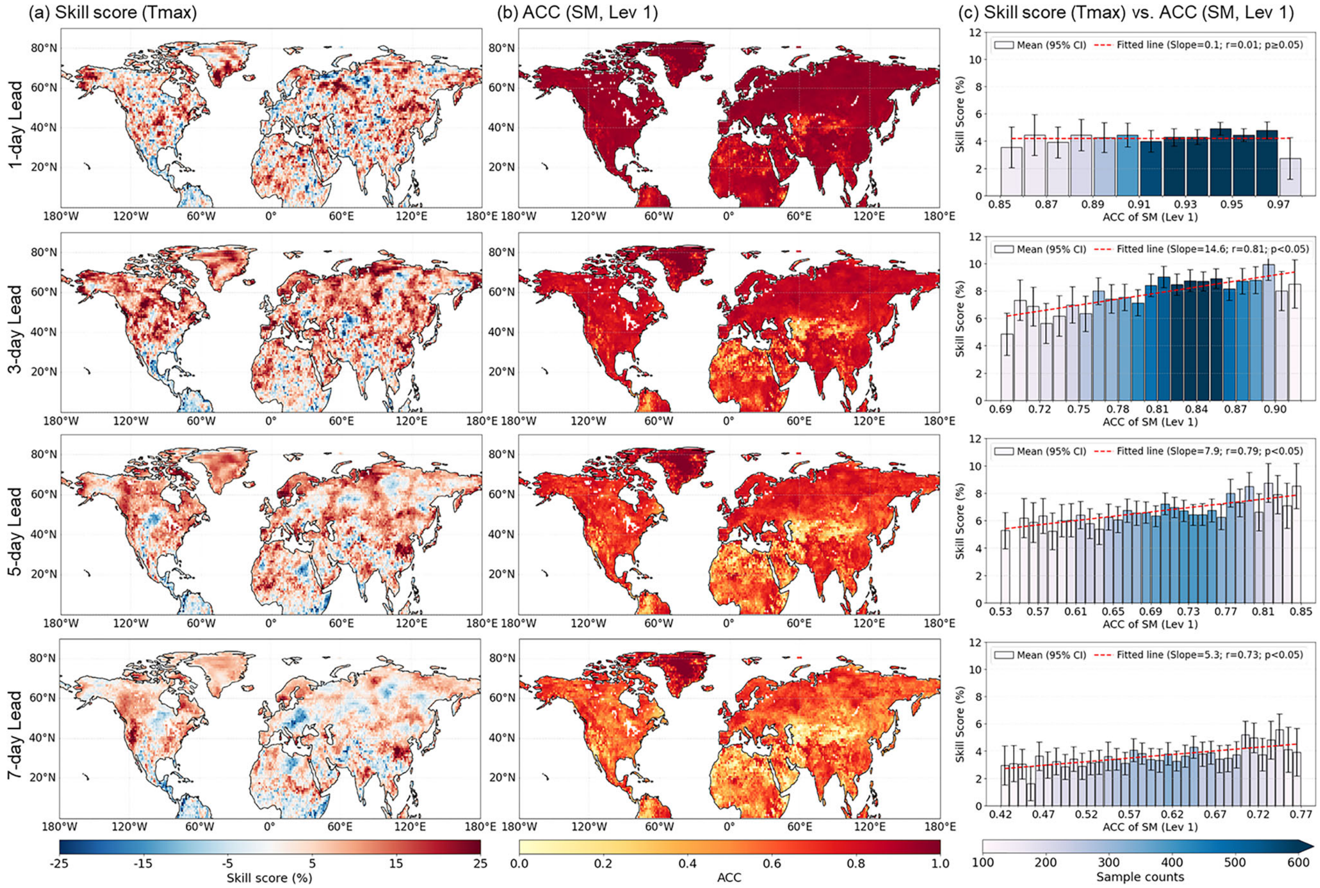

Fig. 3 | a Spatial distribution of daily maximum temperature (Tmax) skill score of the coupled model relative to the atmosphere-only baseline under heatwave conditions (defined as grid points exceeding the 90th percentile of Tmax during June–August (JJA) from 2018 to 2023). **b** Spatial distribution of JJA soil moisture (SM, Level 1) forecast skill (ACC) from the coupled model. **c** Skill score as a function of SM ACC (bin-averaged with 95% confidence intervals) and fitted regression slopes. In **c**, the analysis was restricted to regions with statistically meaningful SM–Tmax coupling strength ($r \le -0.3$) based on ground-truth observations, and bins with fewer than 100 samples were excluded.

atmosphere-dominated extremes, supporting the anticipation of compound heat–drought risks.

Discussion

In this study, we developed a deep learning-based L–A coupled modeling framework designed to advance the prediction of extreme heat events, focusing on Northern Hemisphere summer conditions. While most existing deep learning weather prediction models have primarily focused on atmospheric dynamics without incorporating land surface states[23,24], our coupled model jointly captures land surface variability and atmospheric processes by integrating SM and ST across three layers. To better reflect the L–A interactions on sub-seasonal time scale, the coupled model was trained with a multi-step loss, enhancing its ability to capture essential temporal dependencies in the forecasting process. When comparing coupling strength across training strategies, the coupled model trained with multi-step loss more effectively preserved physically meaningful L–A interactions (Fig. 1), demonstrating its clear advantage in maintaining their magnitude and spatial structure over longer lead times.

To assess the contribution of L–A coupling, we compared the coupled model with atmosphere-only model under different training strategies. The results indicated that combining L–A coupling with multi-step loss yields substantial improvements in overall forecast skill and the prediction of extreme heat events. Specifically, the coupled model trained with multi-step loss outperformed the atmosphere-only model trained under the same multi-step setting in predicting Tmax under heatwave conditions, achieving RMSE-based skill score gains of 5.9–11.2% over 1–7 day lead times, whereas single-step training produced smaller gains of 0.4–2.4% (Fig. 2). The coupled model's forecast skill gains showed distinct lead-time dependence: at ~3 days, high SM and circulation predictability acted complementarily, producing the strongest improvements; at longer leads, as circulation skill declined, the gains were sustained primarily by surface L–A coupling between Tmax and SM (Figs. 3 and S5). Furthermore, case studies of the 2018 Western European and 2022 East Asian heatwaves further demonstrated the coupled model's superior performance in simulating heat extremes. Specifically, it captured land-driven soil drying and feedbacks, and enabling drought simulation under atmosphere-dominated extremes. Therefore, the integration of L–A coupling with multi-step loss emerges as a key strategy for enhancing the reliability of heatwave forecasts.

It is worth noting that the proposed framework is not intended to compete with recent global AI forecasting systems. Rather, it builds upon architectural practices that have proven effective in these models and adapts them for a process-level investigation of heatwave evolution and land–atmosphere coupling. By maintaining closely aligned structural designs between the coupled and atmosphere-only configurations, the framework enables a controlled assessment of coupling effects while remaining complementary to broader forecasting efforts. The proposed framework demonstrates promising results, yet several limitations remain. First, the coupled model exhibited limited skill in predicting humidity and precipitation, which may be attributed to their stronger dependence on free-atmospheric processes and convective variability[40], less directly influenced by land surface conditions. Future work could address this by integrating additional vertical atmospheric information or refining coupling

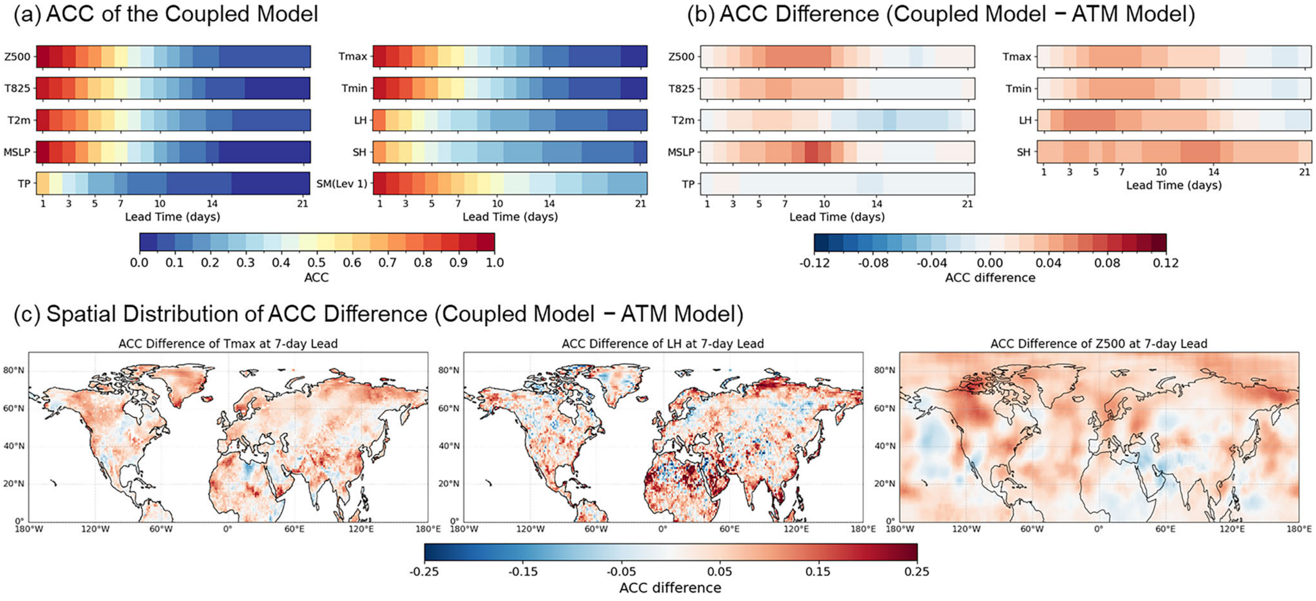

Fig. 4 | Forecast performance of the land–atmosphere coupled model (Coupled) and its comparison with the atmosphere-only model (ATM) over the Northern Hemisphere (0°–90°N). a shows the anomaly correlation coefficient (ACC) for key meteorological variables predicted by the Coupled model, and **b** presents the corresponding ACC differences between the Coupled and ATM models. Soil moisture at the first layer (SM, Lev 1) is excluded from (**b**) as it is not predicted by the ATM model. **c** Spatial distribution of ACC differences at 7-day lead time for maximum temperature (Tmax), latent heat flux (LH), and geopotential at 500 hPa (Z500).

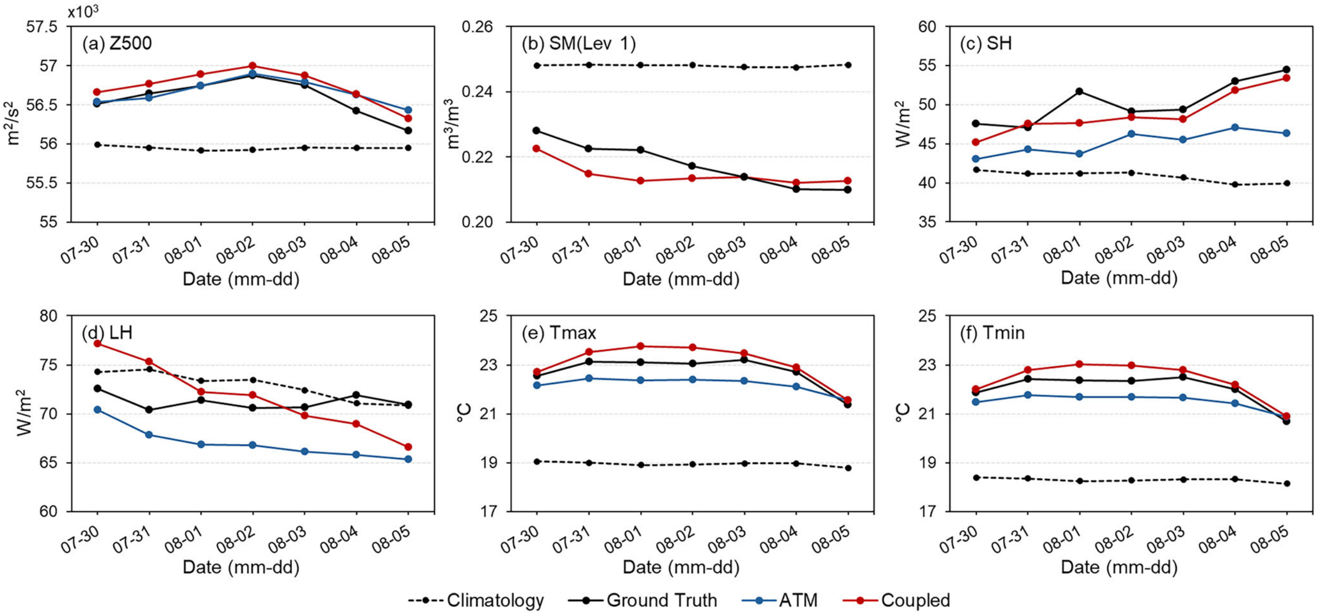

Fig. 5 | Case study of the 2018 extreme heatwave over Western Europe (36°–72°N, 350°–20°E), initialized on 29 July. The plots show 7-day lead time predictions from the atmosphere-only model (ATM, blue) and the coupled land–atmosphere model (Coupled, red). Black solid lines represent the observational ground truth, and black dashed lines indicate climatology. Panels show **a** geopotential at 500 hPa (Z500), **b** soil moisture at the first layer (SM, Lev 1), **c** sensible heat flux (SH), **d** latent heat flux (LH), **e** maximum temperature (Tmax), and **f** minimum temperature (Tmin).

mechanisms near the planetary boundary layer. Second, although multi-step loss enhances temporal consistency, it tends to induce a shrinkage effect that suppresses forecast variance and sharp transition[42], a particular limitation for precipitation due to its high spatiotemporal variability. Incorporating probabilistic forecasting techniques or leveraging spectral operator–based models such as the Fourier Neural Operator, which can better preserve amplitude and phase characteristics, may help address this limitation[22,43]. Lastly, as this study focused on boreal summer conditions in the Northern Hemisphere, further evaluation is needed to assess the framework's applicability to heatwave prediction in the Southern Hemisphere. Addressing these aspects could further improve the generalizability and robustness of L–A coupled prediction systems.

Methods

Data

We used the ERA5 reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) to construct globally consistent atmospheric and land surface datasets[44]. A total of 42 variables

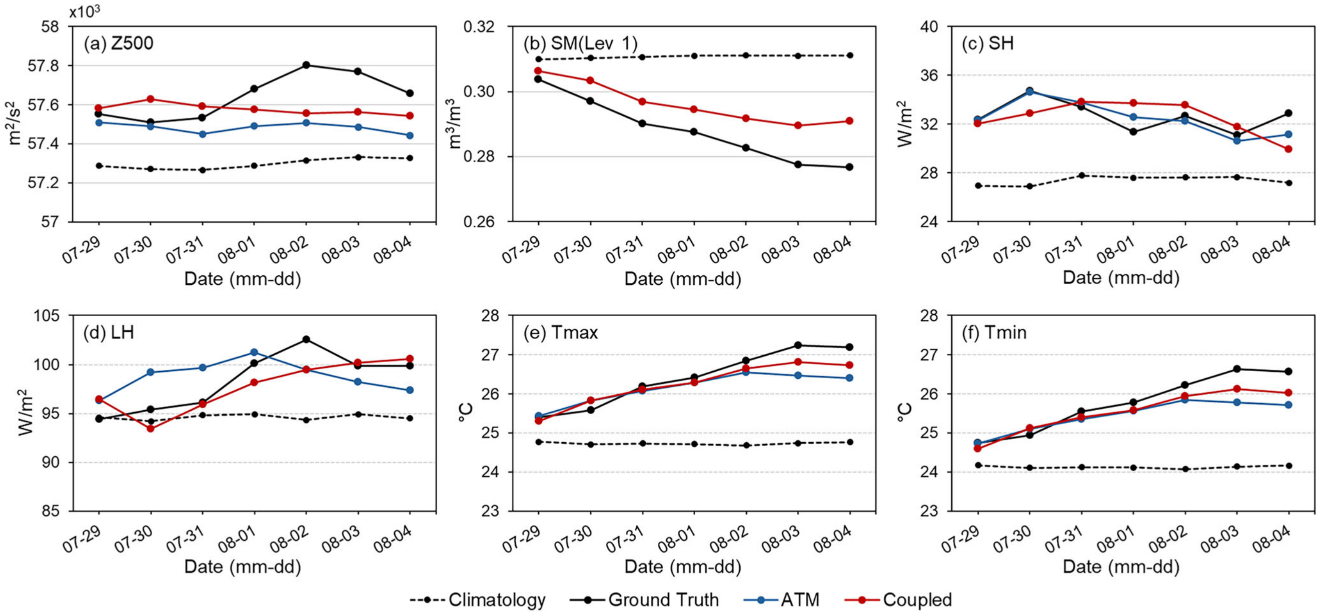

Fig. 6 | Case study of the 2022 extreme heatwave over East Asia (25°–45°N, 105°–140°E), initialized on 28 July. The plots show 7-day lead time predictions from the atmosphere-only model (ATM, blue) and the land–atmosphere coupled model (Coupled, red). Black solid lines represent the observational ground truth, and black dashed lines indicate climatology. Panels show **a** geopotential at 500 hPa (Z500), **b** soil moisture at the first layer (SM, Lev 1), **c** sensible heat flux (SH), **d** latent heat flux (LH), **e** maximum temperature (Tmax), and **f** minimum temperature (Tmin).

Table 1 | ERA5 variables used in this study, categorized by variable type, along with their abbreviations and preprocessing methods

Variable type	Variable name	Abbreviation	Preprocessing
Atmospheric Pressure (1000, 825, 700, 500, 200 hPa)	Geopotential	Z	Daily anomaly
	Temperature	T	
	Specific humidity	Q	Total field
	U-component of wind	U	
	V-component of wind	V	
Atmospheric surface	2 m temperature	T2m	Daily anomaly
	Mean sea level pressure	MSLP	
	Mean surface net long-wave radiation flux	NLWR	
	Mean surface net short-wave radiation flux	NSWR	
	Maximum 2 m temperature	Tmax	
	Minimum 2 m temperature	Tmin	
	Mean surface sensible heat flux	LH	
	Mean surface latent heat flux	SH	
	10 m U-component of wind	U10m	Total field
	10m V-component of wind	V10m	
	Total precipitation	TP	
Land (Lev 1: 0–7 cm, Lev 2: 7–28 cm, Lev3: 28–100 cm)	Soil temperature	ST	Daily anomaly
	Volumetric soil water	SM	Total field

known to influence heatwave evolution were selected, including upper- and surface-level atmospheric variables and multi-layer land state variables (Table 1)[25,29]. The original dataset is provided at an hourly temporal resolution and a spatial resolution of 0.25° × 0.25°. In this study, all variables except for Tmax and Tmin were aggregated to daily means. Tmax and Tmin were computed as the daily maximum and minimum values from hourly data, respectively. All variables were subsequently regridded to a 1° × 1° spatial resolution using bicubic interpolation. The analysis focused on the boreal summer season (May to September) from 1979 to 2023, which corresponds to the peak heatwave period in the Northern Hemisphere.

To optimize the model for extreme heatwave prediction, variable-specific preprocessing was applied to preserve physically relevant signals. For variables representing large-scale thermodynamic conditions (e.g., temperature, pressure, and radiative fluxes), daily anomalies were calculated by subtracting the seasonal cycle (i.e., calendar-day mean) at each grid cell. This approach emphasized deviations from typical seasonal values and improved sensitivity to extreme anomalies[45,46]. In contrast, total fields were retained for humidity and wind components to preserve mass continuity and moisture transport essential for representing precipitation processes[47,48]. This preprocessing strategy was designed to balance the detection of extreme thermodynamic signals with the conservation properties necessary for hydrological consistency.

L–A coupled model

The proposed model employed a convolutional neural network-based encoder–decoder architecture that reduces and restores spatial dimensions for computational efficiency[49]. The architecture is illustrated in Fig. 7. It was trained to produce forecasts for the next day (i.e., a 1-day lead time), treating all variables as prognostic except for precipitation, which is handled as a diagnostic output due to its highly intermittent and noisy nature[50,51]. Separate encoders were designed to process atmospheric and land variables independently, with each further split into total and anomaly components to reflect their differing physical properties (i.e., four independent encoders). Each encoder applied convolutional operations tailored to the input type: standard convolutions were used for atmospheric variables, whereas partial convolutions were employed for land surface inputs to address missing values over ocean regions.

Geocyclic padding was applied along longitudinal and latitudinal boundaries to preserve spatial continuity[52]. Step-aware layers were applied between the encoder and decoder to modulate the encoded features to allow the optimal fusion of heterogeneous features across the different forecast lead times. The decoder restored spatial resolution through upsampling, merged skip-connected features from the encoder, and refined them via further convolutional layers. Lastly, the decoder outputs were combined with the corresponding inputs via residual connections to form the final predictions, except for precipitation, which is predicted directly without residual addition.

The proposed model was trained within a multi-step forecasting framework to capture the evolving dynamics of L–A interactions on sub-seasonal time scale. Multi-step loss formulations have been adopted in recent state-of-the-art forecasting systems (e.g., FuXi and GraphCast), and have shown notable improvements in predictive performance[17,20]. The model was optimized in this setting to predict atmospheric and land surface states across several consecutive lead times, rather than at a single target horizon. This approach accounts for the temporal evolution of forecast errors and enables improved long-range accuracy by optimizing loss over a Markov chain of intermediate predictions. It is particularly well-suited for autoregressive prediction schemes, where the model recursively uses its forecasts as inputs for subsequent steps. In this study, the model was trained to forecast up to 7 days of lead time, based on previous findings that heatwave predictability remains meaningful within this horizon[53,54]. To account for increasing uncertainty with lead time, a time-decaying weighting approach was applied to the total loss:

$$\mathcal{L}_{total} = \sum_{t=1}^{7} w_t \cdot \mathcal{L}_t,\ w_t = \frac{e^{-0.1t}}{\sum_{k=1}^{7} e^{-0.1k}} \tag{1}$$

where $\mathcal{L}_t$ denotes the prediction loss at lead time t, and the weight w_t which exponentially decreases over forecast lead time. This weighting scheme

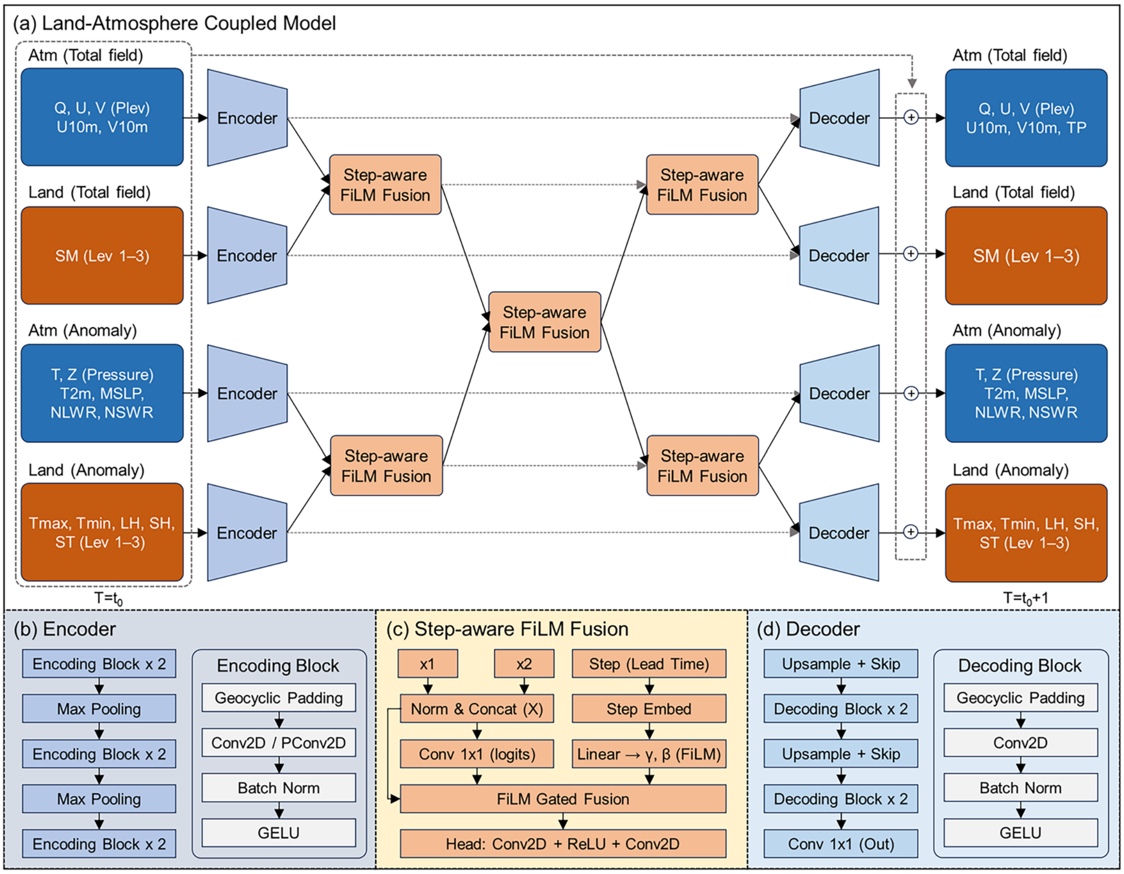

Fig. 7 | Architecture of the proposed convolutional neural network (CNN)-based land–atmosphere coupled model. a The model integrates total and anomaly fields from atmospheric and land variables, fused through Step-aware Feature-wise Linear Modulation (FiLM) modules. **b** CNN-based encoder with geocyclic padding and convolutional blocks. **c** Step-aware FiLM Fusion block incorporating lead time information. **d** Decoder that reconstructs output features through upsampling and skip connections from the encoder. Final predictions are obtained by adding the decoder outputs to the input variables via residual connections, except for precipitation, which is predicted directly without residual addition.

emphasizes near-term accuracy while incorporating longer-range forecasts into the training objective.

To account for the varying importance of atmospheric and land surface variables across different forecast lead times, our model incorporated a step-aware layer based on Feature-wise Linear Modulation (FiLM)[55]. First, the lead time was encoded using a sinusoidal embedding, which was then projected into scale and shift parameters (γ and β). These parameters modulated a set of learned gating logits, enabling dynamic fusion of heterogeneous encoded features conditioned on forecast lead time. The fused features were subsequently refined through convolutional layers with geocyclic padding to model the nonlinear interactions between fused features.

The model was trained using the GELU activation function and AdamW optimizer with a learning rate of 10^{-4} to support stable convergence and smooth nonlinearity[56,57]. A ReduceLROnPlateau scheduler adaptively halved the learning rate when the validation loss plateaued (minimum 10^{-6}), enabling gradual fine-tuning during training. The loss function combined mean squared error (MSE) and Fourier spectrum loss with a 0.7 to 0.3 weighting. While MSE penalizes pointwise errors in the spatial domain, the spectrum loss complements it by aligning the amplitude components in the frequency domain, helping to reduce spectral distortion and suppress spurious high-frequency signals[58,59]. To ensure stable forecasting before introducing sequential dependencies, the model was first pretrained for 100 epochs (~70k steps) using single-step loss, allowing it to learn a reliable 1-day input–target mapping. Multi-step loss training was then performed for 50 epochs (~34k steps), optimizing predictions across a 7-day autoregressive forecast window. A batch size of 8 was used throughout training to balance convergence stability and computational efficiency. A small number of hyperparameters (e.g., learning rate and batch size) were manually adjusted based on validation performance, but these adjustments did not materially influence the overall results. The final model corresponding to the epoch with the minimum validation loss was retained for subsequent evaluation. The corresponding training and validation loss curves are provided in Fig. S1 for reference.

Evaluation

The model was trained using data from 1979 to 2015, with validation on 2016–2017 and evaluation on an independent test set spanning 2018–2023. To assess the effects of model design choices, we compared our proposed coupled model against three baselines: a coupled variant trained with single-step loss, and two atmosphere-only models trained with either single-step or multi-step loss. All single-step models were trained for 150 epochs to match the training budget of the multi-step setting, enabling a fair comparison across training strategies. The atmosphere-only model excludes land variables entirely and adjusts fusion settings to maintain architectural complexity comparable to the coupled model (Fig. S2). This comparative setup was designed to disentangle the contributions of multi-step optimization and land surface coupling, providing insight into how each factor influences the model's ability to capture L–A interactions and enhance predictive accuracy.

Model evaluation was conducted primarily during June–August, when extreme heat events are most frequent in the Northern Hemisphere. To assess the overall predictability, including that of land variables, model evaluation was conducted for lead times up to 21 days, whereas the analysis of heatwave prediction skill focused on forecasts within 7 days. Before calculating forecast skill, bias correction was applied by removing the lead-time-specific mean bias of each model from the test forecasts, ensuring that the anomaly means are zero across the test period. Forecast skill was assessed using the ACC and RMSE. The ACC was computed as:

$$ACC(t) = \frac{\sum_{i,j,d}\left(A^{pred}_{i,j}(t,d)\cdot A^{obs}_{i,j}(d)\right)}{\sqrt{\sum_{i,j,d}\left(A^{pred}_{i,j}(t,d)\right)^{2}}\cdot\sqrt{\sum_{i,j,d}\left(A^{obs}_{i,j}(d)\right)^{2}}} \tag{2}$$

where $A^{pred}_{i,j}(t,d)$ denotes the bias-corrected forecast anomaly at grid point (i,j), lead time t, and forecast day d, and $A^{obs}_{i,j}(d)$ is the observed daily anomaly on day d. The RMSE for each lead time t was computed as:

$$RMSE(t) = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i,j,d}}\left(\mathrm{A}^{\mathrm{pred}}_{\mathrm{i,j}}(\mathrm{t,d}) - \mathrm{A}^{\mathrm{obs}}_{\mathrm{i,j}}(\mathrm{d})\right)} \tag{3}$$

where N denotes the total number of spatiotemporal grid points and forecast days included in the summation. In addition, to quantify the relative improvement in prediction accuracy of the L–A coupled model compared to the atmosphere-only baseline, we adopted a skill score based on RMSE, defined as:

$$Skil\ lscore(t)(\%) = \left(1 - \frac{RMSE_{Coupled}(t)}{RMSE_{ATM}(t)}\right)\times 100 \tag{4}$$

where $RMSE_{Coupled}(t)$ and $RMSE_{ATM}(t)$ represent the RMSEs of L–A coupled model and the atmosphere-only baseline, respectively, at lead time t. Higher values of ACC and skill score indicate better predictive performance, whereas lower RMSE values denote smaller forecast errors and thus higher accuracy.

Data availability

The datasets used in this study were obtained from the ERA5 reanalysis archive and are publicly available at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5.

Code availability

The code used for analysis in this study is available from the corresponding author upon reasonable request.

Received: 29 August 2025; Accepted: 22 December 2025;
Published online: 02 February 2026

References

1. Brown, S. J. Future changes in heatwave severity, duration and frequency due to climate change for the most populous cities. *Weather Clim. Extrem.* **30**, 100278 (2020).
2. Campbell, S., Remenyi, T. A., White, C. J. & Johnston, F. H. Heatwave and health impact research: a global review. *Health Place* **53**, 210–218 (2018).
3. Mukherjee, S. & Mishra, A. K. Increase in compound drought and heatwaves in a warming world. *Geophys. Res. Lett.* **48**, e2020GL090617 (2021).
4. Tripathy, K. P., Mukherjee, S., Mishra, A. K., Mann, M. E. & Williams, A. P. Climate change will accelerate the high-end risk of compound drought and heatwave events. *Proc. Natl. Acad. Sci. USA.* **120**, e2219825120 (2023).
5. Bastos, A. et al. Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity. *Sci. Adv.* **6**, eaba2724 (2020).
6. Dirmeyer, P. A., Balsamo, G., Blyth, E. M., Morrison, R. & Cooper, H. M. Land-atmosphere interactions exacerbated the drought and heatwave over northern Europe during summer 2018. *AGU Adv*. **2**, (2021).
7. García-León, D. et al. Current and projected regional economic impacts of heatwaves in Europe. *Nat. Commun.* **12**, 5807 (2021).
8. Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
9. Alley, R. B., Emanuel, K. A. & Zhang, F. Advances in weather prediction. *Science* **363**, 342–344 (2019).
10. Ritchie, H. et al. Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Weather Rev.* **123**, 489–514 (1995).

11. Allen, M., Frame, D., Kettleborough, J. & Stainforth, D. Model error in weather and climate forecasting. in *Predictability of Weather and Climate* (eds. Palmer, T. & Hagedorn, R.) 391–427 (Cambridge University Press, 2006).
12. Palmer, T. N. et al. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* **33**, 163–193 (2005).
13. Ferranti, L., Corti, S. & Janousek, M. Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Q. J. R. Meteorol. Soc.* **141**, 916–924 (2015).
14. Matsueda, M. Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.* **38**, L06801 (2011).
15. Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021).
16. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
17. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *Npj Clim. Atmos. Sci.* **6**, 1–11 (2023).
18. Chen, K. et al. FengWu: pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv* https://doi.org/10.48550/arXiv.2304.02948 (2023).
19. Hu, Y. et al. Spherical multigrid neural operator for improving autoregressive global weather forecasting. *Sci. Rep.* **15**, 11522 (2025).
20. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
21. Pathak, J. et al. FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv* https://doi.org/10.48550/arXiv.2202.11214 (2022).
22. Price, I. et al. Probabilistic weather forecasting with machine learning. *Nature* **637**, 84–90 (2025).
23. Ben Bouallègue, Z. et al. The rise of data-driven weather forecasting: a first statistical assessment of machine learning–based weather forecasts in an operational-like context. *Bull. Am. Meteorol. Soc.* **105**, E864–E883 (2024).
24. Ren, X. et al. Deep learning-based weather prediction: a survey. *Big Data Res.* **23**, 100178 (2021).
25. Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D. & Schär, C. Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *J. Clim.* **20**, 5081–5099 (2007).
26. Seneviratne, S. I. et al. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**, 125–161 (2010).
27. Hao, Z. et al. Compound droughts and hot extremes: characteristics, drivers, changes, and impacts. *Earth Sci. Rev.* **235**, 104241 (2022).
28. Miralles, D. G., Gentine, P., Seneviratne, S. I. & Teuling, A. J. Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Ann. N. Y. Acad. Sci.* **1436**, 19–35 (2019).
29. Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C. & Vilà-Guerau de Arellano, J. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.* **7**, 345–349 (2014).
30. Tak, S., Seo, E., Dirmeyer, P. A. & Lee, M.-I. The role of soil moisture-temperature coupling for the 2018 Northern European heatwave in a subseasonal forecast. *Weather Clim. Extrem.* **44**, 100670 (2024).
31. Durre, I., Wallace, J. M. & Lettenmaier, D. P. Dependence of extreme daily maximum temperatures on antecedent soil moisture in the contiguous United States during summer. *J. Clim.* **13**, 2641–2651 (2000).
32. Koster, R. D. et al. GLACE: the global land–atmosphere coupling experiment. Part I: overview. *J. Hydrometeorol.* **7**, 590–610 (2006).
33. Wu, W. & Dickinson, R. E. Time scales of layered soil moisture memory in the context ofLand–atmosphere interaction. *J. Clim.* **17**, 2752–2764 (2004).
34. Mei, R. & Wang, G. Summer land–atmosphere coupling strength in the United States: comparison among observations, reanalysis data, and numerical models. *J. Hydrometeorol.* **13**, 1010–1022 (2012).
35. Trok, J. T., Davenport, F. V., Barnes, E. A. & Diffenbaugh, N. S. Using machine learning with partial dependence analysis to investigate coupling between soil moisture and near-surface temperature. *J. Geophys. Res.* **128**, e2022JD038365 (2023).
36. Seneviratne, S. I., Lüthi, D., Litschi, M. & Schär, C. Land-atmosphere coupling and climate change in Europe. *Nature* **443**, 205–209 (2006).
37. Benechehab, A., Thomas, A., Paolo, G., Filippone, M. & Kégl, B. A multi-step loss function for robust learning of the dynamics in model-based reinforcement learning. *arXiv* https://doi.org/10.48550/arXiv.2402.03146 (2024).
38. Somalwar, A., Lee, B. D., Pappas, G. J. & Matni, N. Learning with imperfect models: when multi-step prediction mitigates compounding error. *arXiv* https://doi.org/10.48550/arXiv.2504.01766 (2025).
39. McColl, K. A., He, Q., Lu, H. & Entekhabi, D. Short-term and long-term surface soil moisture memory time scales are spatially anticorrelated at global scales. *J. Hydrometeorol.* **20**, 1165–1182 (2019).
40. van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Moene, A. F. & Holtslag, A. A. M. Interactions between dry-air entrainment, surface evaporation and convective boundary-layer development. *Q. J. R. Meteorol. Soc.* **135**, 1277–1291 (2009).
41. Rousi, E. et al. The extremely hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective. *Nat. Hazards Earth Syst. Sci.* **23**, 1699–1718 (2023).
42. Svetunkov, I., Kourentzes, N. & Killick, R. Multi-step estimators and shrinkage effect in time series models. *Comput. Stat.* **39**, 1203–1239 (2024).
43. Bonev, B. et al. Spherical Fourier neural operators: learning stable dynamics on the sphere. *arXiv* https://doi.org/10.48550/arXiv.2306.03838 (2023).
44. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
45. Horton, D. E. et al. Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature* **522**, 465–469 (2015).
46. Seneviratne, S. I. et al. Changes in climate extremes and their impacts on the natural physical environment. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds. Field, C. B., Barros, V., Stocker, T. F. & Dahe, Q.) 109–230 (Cambridge University Press, 2012).
47. Koster, R. D. et al. Regions of strong coupling between soil moisture and precipitation. *Science* **305**, 1138–1140 (2004).
48. Wacker, U., Frisius, T. & Herbert, F. Evaporation and precipitation surface effects in local mass continuity laws of moist air. *J. Atmos. Sci.* **63**, 2642–2652 (2006).
49. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, 2015).
50. Lang, S. et al. AIFS -- ECMWF's data-driven forecasting system. *arXiv* https://doi.org/10.48550/arXiv.2406.01465 (2024).
51. Watt-Meyer, O. et al. ACE2: accurately learning subseasonal to decadal atmospheric variability and forced responses. *arXiv* https://doi.org/10.48550/arXiv.2411.11268 (2024).
52. Cheon, M. et al. KARINA: an efficient deep learning model for global weather forecast. *arXiv* https://doi.org/10.48550/arXiv.2403.10555 (2024).
53. Lavaysse, C., Naumann, G., Alfieri, L., Salamon, P. & Vogt, J. Predictability of the European heat and cold waves. *Clim. Dyn.* **52**, 2481–2495 (2019).
54. Oertel, A. et al. Everything hits at once: how remote rainfall matters for the prediction of the 2021 north American heat wave. *Geophys. Res. Lett.* **50**, e2022GL100958 (2023).

55. Perez, E., Strub, F., De Vries, H., Dumoulin, V. & Courville, A. FiLM: visual reasoning with a general conditioning layer. *AAAI Artif. Intell.* **32**, 3942–3951 (2018).
56. Hendrycks, D. & Gimpel, K. Gaussian error linear units (GELUs). *arXiv* https://doi.org/10.48550/arXiv.1606.08415 (2016).
57. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv* https://doi.org/10.48550/arXiv.1711.05101 (2017).
58. Guan, H., Arcomano, T., Chattopadhyay, A. & Maulik, R. LUCIE: a lightweight uncoupled climate emulator with long-term stability and physical consistency. *J. Adv. Model. Earth Syst.* **17**, e2025MS005152 (2025).
59. Wang, J., Wang, X., Guan, J., Zhang, L. & Zhou, J. TAFFNet: Time-aware adaptive feature fusion network for very short-term precipitation forecasts. *Geophys. Res. Lett.* **50**, e2023GL104370 (2023).

Acknowledgements

This research was supported by the ASTRA Project through the National Research Foundation (NRF) funded by the Ministry of Science and ICT (No. RS-2024-00440482).

Author contributions

D. Cho contributed to data curation, methodology development, and formal analysis, and wrote the manuscript. Y.-G. Ham designed the study, supervised the project, and acquired funding. S. Jeong and S.-Y. Kang contributed to data collection and formal analysis. All authors discussed the results and reviewed and approved the final manuscript.

Competing interests

Y.-G.H. is an Associate Editor of this journal. All other authors declare no competing financial or non-financial interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-025-01311-6.

Correspondence and requests for materials should be addressed to Yoo-Geun Ham.