Dual-Scale Temporal Fusion Reveals Structured Predictability in Subseasonal-to-Seasonal Temperature Prediction

Elnaz Bashir[1], **Jiali Wang**[2], **and Lin Yan**[1,*]

[1]Department of Computer Science, Iowa State University, Ames, IA 50011, USA
[2]Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, USA
[*]linyan@iastate.edu

ABSTRACT

Subseasonal-to-seasonal (S2S) temperature forecasts, spanning several weeks to a few months, are critically needed in agriculture practice, energy planning, and extreme-weather induced risk management, yet their reliability varies substantially across seasons and regions. Forecast skill is often attributed primarily to lead time, but this perspective does not fully explain the spatiotemporal patterns of predictability. Here we show that S2S predictability is organized across interacting temporal components, spatial heterogeneity, and large-scale pattern coherence, and that this structure can be explicitly characterized and exploited. We develop a dual-scale learning framework that separates calendar-aligned historical climate context from lead-time matched recent weather evolution, combining them through spatially adaptive fusion to enable stable temperature forecasts across the 30 to 90-day window. The learned fusion weights reveal that the balance between these two temporal scales shifts systematically with season and geography: during winter, interannual context dominates over high latitudes and complex terrain where forecast is the most difficult, while summer predictions reflect a more balanced temporal contribution across the domain. This spatially explicit reorganization of predictability, rather than simple lead-time decay, emerges as the primary determinant of forecast skill within the subseasonal window. Topology-aware structural constraints further improve spatial coherence of predicted temperature fields, stabilizing large-scale pattern organization particularly over complex terrain. These results reframe S2S predictability as a structured, multi-scale phenomenon, providing a more interpretable foundation for improving forecast systems and informing their use in practice.

Introduction

Subseasonal-to-seasonal (S2S) prediction is widely recognized as a frontier problem in both scientific and operational practices because it must bridge the gap between medium-range weather forecasting and seasonal outlooks[1,2]. At these lead times, forecast skill inherited from initial conditions decays rapidly, while boundary-forced signals and slowly varying modes have not yet emerged as robust and locally actionable sources of predictability[3–7].

The importance of S2S prediction extends beyond theoretical interest. Many high-consequence societal decisions are made on a 4–12 week planning horizon, where day-to-day weather forecasts are no longer sufficient and seasonal averages are too coarse to support action. In agriculture and food systems, decisions such as planting and harvest timing, fertilization, and irrigation scheduling often require actionable outlooks beyond two weeks[1,8]. In water management, reservoir operations and allocation planning similarly rely on S2S guidance to anticipate drought and unusual demand[1,9,10]. Energy demand planning, heat-risk preparedness, and compound-risk management also benefit from skillful temperature outlooks on the S2S horizon[11,12].

Despite decades of progress in numerical weather prediction, deterministic skill at S2S lead times remains limited, and uncertainty quantification typically relies on large ensemble systems[6,13,14]. To accelerate progress on this timescale, the S2S Prediction Project, led by the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP), assembled a multi-center database of near-real-time ensemble forecasts and reforecasts extending to about 60 days, enabling more consistent evaluation across models[4,5]. Even with this infrastructure, operational skill in near-surface temperature usually declines substantially beyond two weeks and exhibits strong seasonality and spatial dependence, underscoring the challenge of representing multiscale variability and teleconnections as lead time increases. These limitations motivate two broad research directions: improving numerical models and developing data-driven or hybrid systems that can better exploit historical information and systematic structure[13–15].

Artificial Intelligence (AI) has recently transformed medium-range weather forecasting. Models such as Pangu-Weather and GraphCast have demonstrated striking gains in multi-variable global prediction out to about 10 days, while benchmark efforts such as WeatherBench, WeatherBench 2, and Extreme Weather Bench have enabled more transparent and reproducible

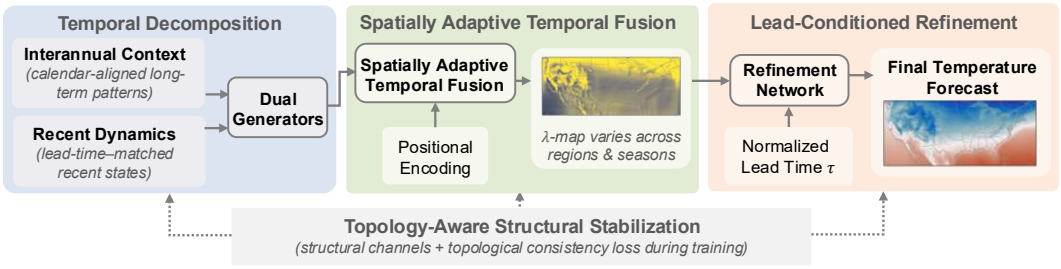

Figure 1. Conceptual overview of a dual-scale S2S temperature forecasting architecture. Historical climate context and recent weather evolution are processed by parallel generators and combined through spatially adaptive temporal fusion, producing regionally varying fusion weights. A lead-conditioned refinement network improves coherence of the fused forecast. Topology-aware structural descriptors are incorporated within the network, with additional consistency constraints applied during optimization.

comparison of data-driven methods[16–20]. These advances have stimulated growing interest in extending AI-based prediction beyond the two-week horizon. FuXi-S2S reported global daily-mean forecasts out to roughly 42 days for a broad suite of variables, representing one of the clearest demonstrations of AI competitiveness at the S2S scale[21]. More recent diffusion-based approaches have also shown progress toward stable 70-90-day forecasts in large-scale generative settings, highlighting the rapid evolution of AI for S2S prediction, although these systems remain new and require careful interpretation with respect to task design and evaluation protocols[22,23].

One reason is that skill degradation is often framed primarily as a function of forecast horizon, that is, as a roughly monotone decay with lead time. This encourages one-dimensional evaluation and can obscure other regime-dependent factors that matter to end users[1,7]. However, S2S predictability is modulated not only by season but also by background state and large-scale modes that reorganize error growth and teleconnection pathways. For example, the Madden–Julian Oscillation is widely recognized as a major source of subseasonal predictability, with teleconnections that influence extratropical circulation and surface temperature patterns[24–26]. These considerations suggest that S2S temperature skill should be examined not only as a function of lead time, but also through the lens of seasonally varying spatial structure.

In this work, we revisit S2S temperature prediction over the continental United States at 30–90-day lead times from a structural perspective. We adopt a dual-scale fusion model that separates calendar-aligned historical climate context from lead-time–matched recent weather evolution and integrates them through spatially adaptive fusion. This design enables explicit analysis of how temporal information is weighted across seasons and geographic regions. To promote stable spatial organization, we incorporate topology-aware structural descriptors together with consistency constraints that encourage preservation of large-scale spatial coherence. The overall architecture is illustrated in Fig. 1.

Here, we use the term structured predictability to refer to the explicit representation of temporal decomposition, spatial heterogeneity, and geometric coherence in S2S temperature forecasting. Temporally, temperature evolution is represented through separable components operating on distinct scales, including interannual context versus recent day-to-day evolution. Spatially, predictability weighting varies across regions. Geometrically, large-scale pattern coherence is encouraged through topology-aware constraints. This formulation moves beyond uniform horizon-based modeling by explicitly accounting for multi-scale temporal behavior, region-dependent skill variation, and spatial dependency structure.

Our analysis yields three main findings. First, within the 30–90-day window, forecast errors exhibit a pronounced seasonal organization: winter errors are systematically larger and more spatially heterogeneous than summer errors, whereas within-season differences across lead times are comparatively modest. Second, the learned fusion weights adapt to local forecast difficulty in a season-dependent manner, indicating that the balance between recent weather evolution and long-term climate context shifts across forecasting regimes. Third, topology-aware structural information improves spatial coherence and accelerates convergence, providing a practical mechanism for stabilizing learning without altering the central adaptive fusion behavior.

Results

Building on the dual-scale architecture outlined in Fig. 1, we first examine its temporal structure (Fig. 2), then its seasonal and spatial manifestations of predictability (Fig. 3 and Fig. 4), and finally the role of topology-aware cues in stabilizing spatial coherence (Fig. 5).

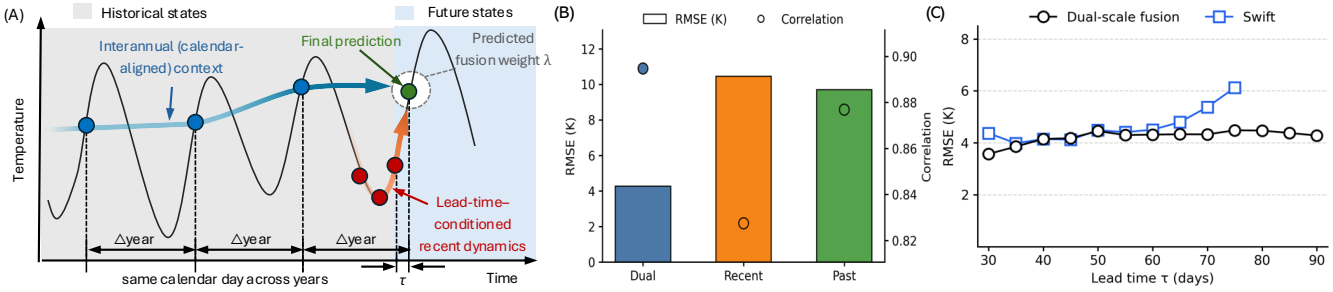

Figure 2. Dual-scale formulation and forecast characteristics. (A) Conceptual schematic illustration of the dual-scale fusion framework, combining calendar-aligned interannual context and lead-time–matched recent dynamics. A spatially adaptive fusion weight λ is predicted by the model to modulate their relative contributions and produce the final forecast at lead time τ. (B) Forecast error and spatial correlation metrics for the dual-scale formulation and corresponding single-scale variants. RMSEs and correlations are averaged over lead times from 30 to 90 days sampled at 5-day intervals. (C) Forecast skill of our dual-scale fusion model across lead times τ over the held-out test period, compared with an external S2S stochastic AI foundation forecasting baseline (Swift[23]). Swift results are available up to 75 days and are computed from ensemble predictions (32 initial times, 15 members each).

Dual-scale formulation and temporal stability

Fig. 2 highlights the temporal decomposition at the core of the dual-scale fusion framework. At each grid point, temperature evolution is represented as the combination of a recurring annual background pattern aligned across years and lead-time-matched information from recent weather states. Rather than blending these influences implicitly, the framework represents them as complementary temperature estimates derived from historical climate context and recent weather evolution(Fig. 2A). These two components are combined through a spatially adaptive fusion weight, λ, predicted from the input fields. Unlike fixed or spatially uniform combinations, this design allows the relative contribution of long-term climate context and recent weather evolution to vary across regions, making λ a spatially explicit indicator of how temporal information is weighted.

To evaluate the benefit of temporal decomposition, we compare the dual-scale fusion framework with two single-scale variants that rely exclusively on either historical climate context or recent weather evolution. As shown in Fig. 2B, combining both components consistently reduces forecast error and improves spatial correlation relative to either single-scale variant, indicating that the two timescales provide complementary predictive information.

To further contextualize model performance, we compare our results with those from an external, stochastic AI foundation model, Swift[23] in Fig. 2C. Swift provides forecasts up to 75 days at a spatial resolution of about 150 km. To ensure a fair comparison, our predictions were downsampled to match this resolution prior to evaluation. Within the overlapping lead-time range (approximately 30–70 days), our dual-scale fusion model shows comparable performance while maintaining relatively stable error behavior across the full 30–90-day window. Representative comparisons are provided in Fig. S3.

Seasonal prediction quality and spatial–distributional structure

Fig. 3 contrasts representative winter (DJF) and summer (JJA) forecasts averaged across lead times from 30 to 90 days. Transitional seasons are shown in Fig. S4. Seasonal statistics are summarized in Table S1, while Table S2 reports seasonal contrast and lead-time sensitivity relative to a simple day-of-year climatology baseline. Pronounced seasonal differences emerge in both spatial structure and error magnitude.

In DJF, the ground-truth temperature field exhibits sharper meridional gradients and more intricate spatial texture. The predicted field preserves the large-scale organization but displays amplified errors across high latitudes and complex terrain. In contrast, the JJA ground-truth temperature fields are smoother and more spatially coherent, and the predictions align closely with regional temperature patterns, especially terrain-driven localized gradients. Mean RMSE decreases from approximately 5.8 K in DJF to 2.7 K in JJA, highlighting a strong seasonal contrast that exceeds variation across lead times within the same 30–90-day window.

Spatial error maps indicate that this contrast is systematic rather than localized. Winter errors extend across broad continental regions, whereas summer errors are smaller in magnitude and confined mainly to areas with strong spatial gradients (e.g., mountains). The difference, therefore, reflects a shift in the overall error structure rather than a few isolated hotspots.

Distributional analysis reinforces this picture. Kernel density estimates (KDEs) show substantial overlap between predicted and ground truth temperatures in both seasons, with an overlap ratio of 0.95 and 0.93 in summer and winter, respectively; however, winter distributions exhibit broader tails and greater spread. The seasonal contrast thus arises predominantly from differences in variability rather than from systematic mean bias.

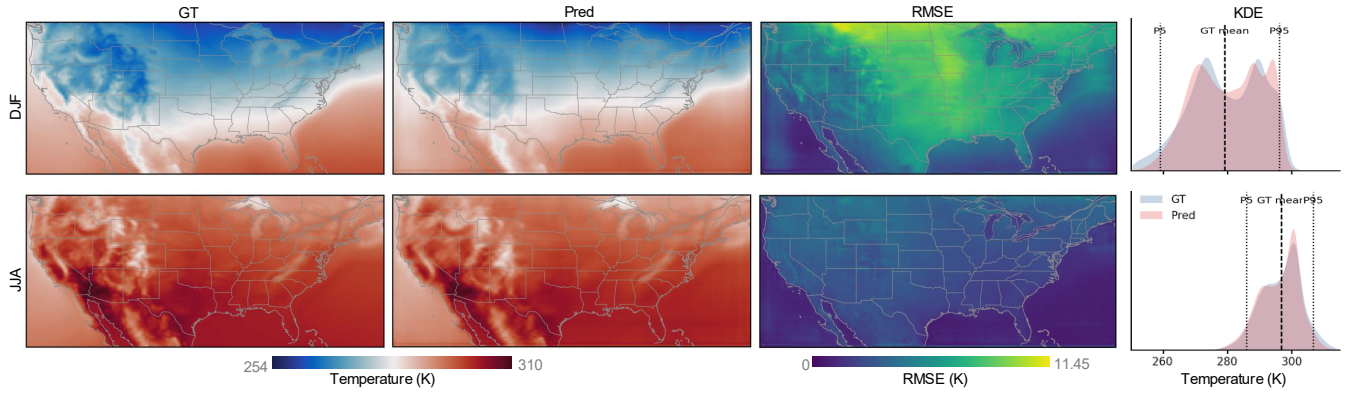

Figure 3. Seasonal spatial structure and distributional consistency of S2S temperature prediction. Seasonal mean near-surface temperature over CONUS comparing ground truth (GT), model prediction (Pred), and spatial root-mean-square error (RMSE) averaged across lead times from 30–90 days (5-day intervals). Winter (DJF) and summer (JJA) are shown in separate rows. GT and Pred share identical color scales within each season to facilitate direct structural comparison, while RMSE is shown using a fixed scale across seasons. Kernel density estimates (right) summarize the temperature distributions and illustrate the distributional agreement between GT and Pred. The dashed line indicates the GT mean, and the dotted lines mark the 5th and 95th percentiles of GT. The overlap ratios between the GT and Pred tails are 0.48 for values below P5 and 0.69 for values above P95 in winter, and 0.99 and 0.70, respectively, in summer. Despite increased intrinsic variability in winter, the model preserves both large-scale spatial organization and bulk distributional characteristics across seasons. Seasonal summary statistics are reported in Table S1.

This seasonal contrast remains evident across lead times, with the winter–summer separation exceeding within-season variation over the 30–90-day window (Fig. S5A). This indicates that, within the subseasonal range, seasonal structure exerts a stronger influence on forecast quality than horizon-dependent degradation[27]. This seasonal organization motivates the next analysis of how adaptive temporal fusion responds across seasons and geographic regions.

Seasonal contrasts and space-aware adaptation of the predicted fusion weight

Having established that seasonal differences dominate lead-time effects in forecast error, we next examine how the fusion weight λ varies across space. Fig. 4A–C contrasts winter (DJF) and summer (JJA) λ fields and their difference ($\Delta\lambda = \lambda_{\text{winter}} - \lambda_{\text{summer}}$), revealing pronounced and spatially coherent seasonal contrasts. Larger λ values correspond to greater reliance on historical climate context, whereas smaller values place more weight on recent weather evolution.

During winter, λ is systematically elevated across large portions of the continental United States, particularly over the northern U.S. and mountainous regions, suggesting greater reliance on interannual context under more variable midlatitude environment. In summer, λ is lower and more spatially homogeneous, remaining near 0.5 across much of the domain, indicating a more balanced contribution from historical context and recent weather. The difference map (Fig. 4C) shows that regions with amplified winter λ largely coincide with regions of increased forecast error (Fig. 3), indicating that the fusion strength intensifies where predictability is lower rather than shifting uniformly with season.

Beyond these seasonal contrasts, elevated λ values are also consistently observed over complex terrain, including during summer when forecast errors remain relatively low. This indicates that higher λ does not uniformly correspond to larger errors, but instead reflects region-dependent temporal weighting. For example, during summer, regions with elevated λ, particularly over mountainous areas, still exhibit relatively low RMSE (Fig. 3). These patterns suggest that the fusion mechanism adapts differently across spatial and seasonal conditions, reflecting the interplay between temporal variability and geographic structure rather than a simple monotonic relationship between λ and forecast error.

Error-stratified analysis (Fig. 4D; Fig. S6; Table S3) further clarifies this behavior. In winter, mean λ increases monotonically across RMSE bins, with higher-error regions assigning greater weight to interannual context. This monotonic increase persists across substantial sample sizes, indicating that the pattern is not driven by isolated extremes. In summer, by contrast, λ remains comparatively stable across error bins, with only weak sensitivity to local RMSE. Transitional seasons display intermediate behavior. Importantly, λ also remains largely stable across forecast horizons within the 30–90-day window (Fig. S5B), with nearly identical central tendencies at 30, 60, and 90 days. Thus, although forecast error increases modestly with lead time (Fig. 2C), the fusion weight does not vary systematically with forecast horizon.

Taken together, these results show that adaptive temporal fusion responds primarily to seasonal regime and local forecast difficulty, rather than to lead time alone. In particular, elevated winter λ over northern latitudes and complex terrain suggests

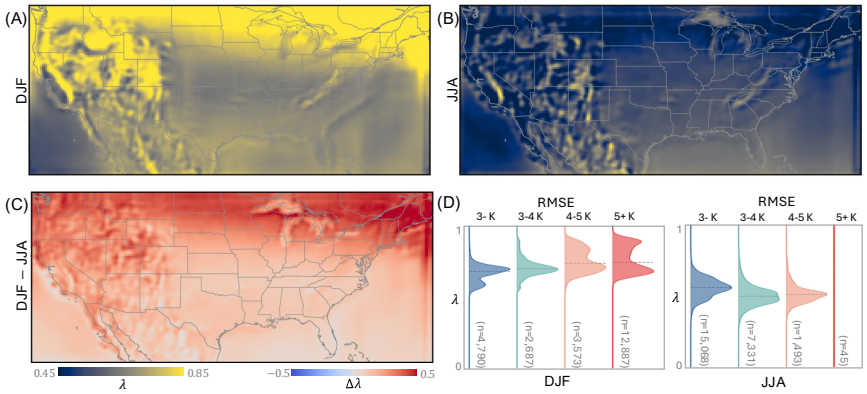

Figure 4. Seasonal reorganization of the learned fusion weight λ and its dependence on predictability regimes. (A) Spatial distribution of λ during winter (DJF). (B) Spatial distribution of λ during summer (JJA). (C) Seasonal difference, $\Delta\lambda = \lambda_{\text{winter}} - \lambda_{\text{summer}}$. Winter exhibits enhanced λ over inland and higher-latitude regions, indicating increased reliance on interannual variability under reduced subseasonal predictability. (D) Distributions of λ stratified by seasonal mean RMSE bins (3- K, 3–4 K, 4–5 K, 5+ K) for DJF (left) and JJA (right). In winter, λ shifts systematically toward higher values with increasing RMSE, whereas summer shows substantially weaker regime-dependent modulation. These results indicate that λ encodes seasonally structured predictability organization rather than reflecting direct pixelwise error sensitivity.

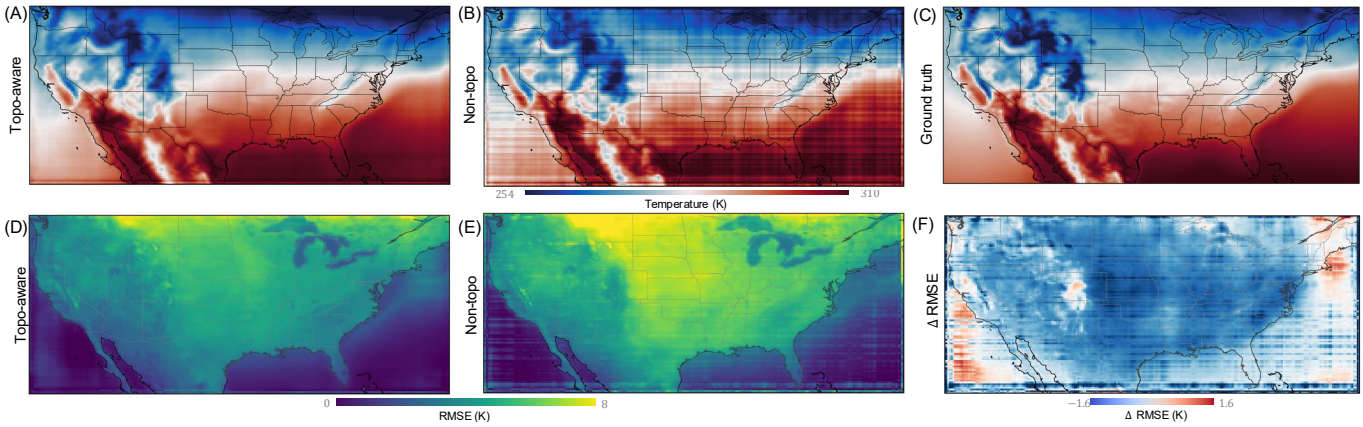

Figure 5. Topology-aware modeling improves spatial coherence and predictive accuracy. The first row (A–C) shows the annual mean temperature fields over the evaluation period, comparing the topology-aware prediction (A), the non-topological baseline (B), and the ground truth (C). The second row (D–F) presents the corresponding spatial RMSE for the topology-aware model (D), the baseline model (E), and their difference (ΔRMSE, F). Both models capture the large-scale temperature distribution, but the non-topological baseline exhibits grid-aligned artifacts and reduced spatial coherence. In contrast, the topology-aware model produces more realistic and physically consistent spatial patterns. The ΔRMSE map indicates broadly distributed error reduction across most regions.

that the model leans more heavily on historical climate context in settings where recent weather evolution is less reliable. This behavior is physically plausible and consistent with well-known challenges in numerical weather prediction: midlatitude winter is characterized by stronger baroclinic instability and richer mesoscale variability, while the local manifestations of weather, such as snowfall, cold extremes, and wind gusts, are often much harder to predict than the larger-scale synoptic pattern. The learned λ field, therefore, provides a spatially explicit indicator of when and where interannual context offers a stabilizing benefit relative to recent dynamics. We next examine how topology-aware structure further shapes spatial coherence within this adaptive system.

Topology-aware modeling improves spatial coherence and optimization stability

Fig. 5 examines the effect of incorporating topology-aware structural descriptors on spatial prediction behavior. Spatial RMSE maps averaged across lead times from 30 to 90 days reveal clear differences between the topology-aware model (Fig. 5D) and the matched baseline without topological inputs (Fig. 5E). Although the overall reduction in RMSE is moderate, the spatial

organization of the error field changes noticeably.

The baseline exhibits visible grid-aligned banding patterns across portions of the domain, whereas the topology-aware formulation produces more realistic spatial patterns and generally smaller errors. In particular, regional gradient structures become more clearly resolved, and the predicted fields align more closely with large-scale terrain-related temperature variations. Fig. 5F further shows that error reductions are broadly distributed rather than confined to isolated regions, indicating that the topology-aware representation improves spatial consistency of predicted temperature fields.

Importantly, incorporating topological information also improves training behavior and predictive consistency (Fig. S7). These results align with the notion of structured predictability introduced earlier: by helping preserve large-scale spatial coherence while interacting with temporally decomposed inputs and season- and region-dependent fusion weighting, the topology-aware formulation promotes more stable learning of spatial dependencies and more consistent forecast behavior.

Discussion

Our study complements physically based understanding of S2S variability and highlights the potential role of data-driven models as diagnostic tools for identifying where and when different sources of predictability become most influential[16]. In particular, the results reveal a structured organization of predictability that can be examined through the proposed dual-scale fusion framework.

A structured view of S2S predictability

This structured behavior is reflected in three main aspects. First, forecast errors exhibit a clear seasonal contrast, with the largest errors occurring in winter. Winter is characterized by broader error distributions and stronger spatial heterogeneity, particularly over northern and mountainous regions, whereas summer forecasts remain comparatively stable and spatially consistent. This pattern aligns with the stronger variability and reduced predictability of midlatitude winter conditions, and is further corroborated by consistent improvements over a day-of-year climatology baseline (the historical mean temperature for each calendar day, a standard reference in S2S forecasting[1,4]): RMSEs are reduced by 2.26 K in DJF and 3.76 K in JJA (Table S2), suggesting that the observed seasonal contrast reflects physically meaningful structure rather than model-specific artifacts.

Second, the learned fusion weights show systematic spatial and seasonal variation. Higher λ values are associated with regions of larger forecast error in winter (e.g., the northern U.S. and complex terrain), while lower values are observed under more stable summer conditions. Elevated λ values also appear consistently over mountainous regions, indicating spatially heterogeneous temporal weighting. Notably, higher λ does not uniformly correspond to larger errors, suggesting that the fusion mechanism adapts to both seasonal variability and geographic structure. This relationship should be interpreted as an empirical association rather than a causal mechanism.

Third, incorporating topology-aware structural information improves the coherence of predicted temperature fields. Compared with models without structural constraints, the resulting predictions exhibit more consistent large-scale spatial organization while preserving localized variability, suggesting that topology-aware cues act as an effective inductive bias for stabilizing spatial dependencies.

Implications, Limitations, and Future Directions

These findings provide a complementary perspective on how S2S predictability is structured, with implications for both interpretation and model design. From a modeling perspective, future AI-based approaches may benefit from incorporating multi-scale structure, in which slowly varying background conditions help stabilize longer-range predictions. From an interpretability perspective, the learned fusion weights provide a more interpretable view of data-driven model behavior by revealing how multi-scale information is utilized across seasons and regions.

Several limitations should be noted. First, while the learned fusion behavior exhibits physically plausible patterns, its relationship to established climate modes and large-scale circulation features has not been explicitly quantified. Future work could investigate connections between the learned representations and known sources of subseasonal predictability, such as teleconnection patterns.

Second, direct quantitative comparison with other data-driven S2S systems is constrained by the limited public availability of model outputs over the full 30–90-day window. Most existing approaches, including FuXi-S2S[21], report results up to approximately 42 days, leaving the extended subseasonal range without established AI reference points. Swift represents the closest available comparison; to ensure a fair evaluation, our predictions were coarsened to match Swift's[23] spatial resolution (approximately 150 km) prior to comparison. Establishing standardized benchmarks for the full subseasonal window remains an open challenge for the community, and future evaluation against operational numerical ensemble systems such as ECMWF S2S reforecasts would provide a more complete picture of relative skill.

Third, this study relies primarily on near-surface air temperature as the input variable. Extending the current framework to incorporate additional atmospheric variables could further improve temperature prediction by providing complementary

information beyond temperature alone. It may also enable extensions to more challenging settings, such as prolonged extreme events (e.g., heatwaves), where higher temporal resolution and richer inputs are needed to capture evolving dynamics.

Fourth, this framework is developed for a single-variable setting and a regional domain. Extending this approach toward a more comprehensive Earth system setting, in which multiple atmospheric variables are jointly modeled and predicted, would enable a more complete characterization of S2S predictability. In particular, the framework may be useful for variables governed by multi-scale temporal processes, such as components of the water cycle (e.g., precipitation, soil moisture, and evaporation), where different memory timescales and responses to atmospheric forcing interact. In such settings, separating slowly varying background influences from recent forcing may provide further insight into how predictability is organized.

More broadly, this perspective suggests a potential pathway for developing more integrated S2S forecasting systems, in which learning is organized around structured variability rather than forecast lead time alone, potentially improving both interpretability and forecast stability.

Methods

This section summarizes the datasets, forecasting formulation, and modeling framework used to analyze S2S temperature predictability. Detailed preprocessing procedures, architectural diagrams, and training specifications are provided in Supporting Information.

Data

We use the European Centre for Medium-Range Weather Forecasts phase 5 reanalysis dataset (ERA5)[28], which provides globally consistent atmospheric fields generated through advanced data assimilation techniques[29] and has been widely adopted in ML model training for weather predictability[16–18]. Our analysis focuses on daily mean 2-m air temperature over the continental United States (CONUS). Hourly ERA5 fields are aggregated to daily means and cropped to a 101×237 spatial region covering the CONUS domain across all experiments . The dataset spans 2010–2024, with the first 10 years (2010–2019) for training and the last 5 years (2020–2024) for evaluation.

Problem formulation and forecasting task

We consider S2S prediction of daily mean near-surface air temperature at lead times ranging from 30 to 90 days, a window that bridges the scales between medium-range weather forecasting and seasonal forecasting [1,7].

Forecasts across the 30–90-day range are trained within a unified model rather than as separate horizon-specific networks. Lead time τ is incorporated as a conditioning variable within the refinement module, enabling horizon-dependent adjustment through shared parameters.

The dual-scale fusion mechanism, including the predicted λ-map, does not receive τ as an explicit input. However, lead time is implicitly encoded through the horizon-conditioned recent context, whose temporal sampling depends on τ. This design allows the model to learn both common structures (e.g., temporal and spatial organization) and time horizon-dependent variations, while preserving the interpretability of the fusion weight. Training samples are organized across lead-time groups spanning the 30–90-day window, without imposing globally fixed temporal combinations[18,21].

At S2S scales, predictability reflects contributions from both slowly varying background climate conditions and more rapidly evolving recent atmospheric variability[3,5,26]. To reflect these distinct temporal influences, model inputs are organized into two complementary contexts:

- *Historical climate context:* calendar-aligned temperature fields from the same day in previous years, capturing climatological baselines and low-frequency variability.
- *Recent weather evolution:* horizon-conditioned observations immediately preceding forecast initialization, capturing short-term weather evolution relevant to the target horizon.

This separation is calendar-aligned and horizon-conditioned rather than a conventional seasonal decomposition. The two contexts encode distinct predictive signals and are not assumed to contribute equally across space, season, or forecast horizon.

Dual-scale learning framework

Building on the dual-scale architecture summarized in Fig. 1, we implement the temporal decomposition through two parallel predictive branches that process historical and recent weather contexts separately (detailed in Fig. S2). Each branch produces a candidate temperature forecast reflecting the structure emphasized by its respective temporal input. Both branches are implemented as U-Net–style encoder–decoder networks with skip connections, trained with auxiliary adversarial objectives.

Rather than prescribing a fixed or globally uniform temporal weighting, the two candidate forecasts are combined through a spatially resolved fusion λ-map predicted conditionally on the input fields. Unlike scalar or globally shared gating mechanisms,

λ is learned as a dense map at the same spatial resolution as the output temperature field. This design enables location-specific modulation of interannual and recent weather contributions. By explicitly predicting a spatially resolved fusion field, the formulation treats temporal fusion as an interpretable spatial representation rather than an internal network operation. The λ-map functions as a learned gating variable rather than a physical quantity, but its explicit representation allows post hoc analysis of how temporal information is balanced across space and season. The fused prediction is subsequently processed by a residual refinement network[30], which improves spatial coherence while preserving large-scale structure.

Lead-conditioned refinement

Forecast lead time τ is encoded as a normalized scalar and broadcast spatially as a conditioning map. This lead-conditioned representation is injected into intermediate layers of a U-Net–based refinement module[31], allowing the model to adjust its residual correction according to forecast horizon.

This refinement stage performs a single learned update rather than iterative diffusion. Its role is to reduce noise and enhance spatial consistency without altering the primary dual-scale fusion mechanism.

Topology-aware structural constraint

Predictions from data-driven models may exhibit spatial artifacts or inconsistent patterns, particularly in regions with complex variability. To mitigate these issues, we incorporate topology-aware structural descriptors derived from the temperature field as auxiliary input channels (see Fig. S1). These descriptors encode critical-point type, scalar magnitude, and saddle-level contour structure, providing explicit information about spatial organization.

Topological summaries provide compact representations of spatial connectivity and feature persistence that are robust to small perturbations[32,33]. In this setting, topology-aware channels act as an inductive bias that guides how spatial dependencies are organized during learning.

The topology-aware component does not alter the core forecasting architecture or impose explicit physical constraints. It influences spatial representation learning by providing structural cues. While alternative regularization strategies may achieve related stabilization effects, explicitly encoding structural information offers a principled mechanism for improving spatial coherence within the modeling framework considered here.

Training objective

The model is trained end-to-end using a composite objective that integrates pixel-level reconstruction, structural similarity[34], adversarial consistency[35], and structural regularization terms[36,37]. The topological regularization component is activated after an initial warm-up period to avoid interfering with early-stage optimization.

Funding

This work was supported in part by startup funds provided by Iowa State University. Grant number: Not applicable. Computational resources were also provided by Iowa State University.

Acknowledgment

We thank Dr. Jason Stock and Dr. Rao Kotamarthi of Argonne National Laboratory for sharing the Swift forecasting results, which enabled direct comparison with our study.

Author contributions statement

E.B. designed and implemented the experiments, conducted data analysis, and prepared the figures. J.W. provided domain-specific interpretation of temperature prediction results. L.Y. conceived the study, developed the topology-aware dual-scale modeling framework, and supervised the research. All authors contributed to the discussion and writing of the manuscript.

References

1. White, C. J. *et al.* Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Am. Meteorol. Soc.* **103**, E1448–E1472 (2022).
2. Becker, E. J., Kirtman, B. P., L'Heureux, M., Muñoz, Á. G. & Pegion, K. A decade of the north american multimodel ensemble (nmme): Research, application, and future directions. *Bull. Am. Meteorol. Soc.* **103**, E973 – E995, DOI: 10.1175/BAMS-D-20-0327.1 (2022).
3. Hoskins, B. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q. J. Royal Meteorol. Soc.* **139**, 573–584 (2013).

4. Vitart, F. & Robertson, A. W. The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj climate atmospheric science* **1**, 3 (2018).
5. Vitart, F. Madden—julian oscillation prediction and teleconnections in the s2s database. *Q. J. Royal Meteorol. Soc.* **143**, 2210–2220 (2017).
6. Robertson, A. W., Vitart, F. & Camargo, S. J. Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones. *J. Geophys. Res. Atmospheres* **125**, e2018JD029375 (2020).
7. Robertson, A. & Vitart, F. *Sub-seasonal to seasonal prediction: the gap between weather and climate forecasting* (Elsevier, 2018).
8. Klemm, T. & McPherson, R. A. Assessing decision timing and seasonal climate forecast needs of winter wheat producers in the south-central united states. *J. Appl. Meteorol. Climatol.* **57**, 2129–2140 (2018).
9. Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K. & Mackey, L. Improving subseasonal forecasting in the western us with machine learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2325–2335 (2019).
10. Sengupta, A. *et al.* Advances in subseasonal to seasonal prediction relevant to water management in the western united states. *Bull. Am. Meteorol. Soc.* **103**, E2168–E2175 (2022).
11. Zscheischler, J. *et al.* Future climate risk from compound events. *Nat. climate change* **8**, 469–477 (2018).
12. Pendergrass, A. G. *et al.* Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Clim. Chang.* **10**, 191–199 (2020).
13. Leutbecher, M. & Palmer, T. N. Ensemble forecasting. *J. computational physics* **227**, 3515–3539 (2008).
14. Palmer, T. N. *et al.* Stochastic parametrization and model uncertainty. *ECMWF Reading, UK* (2009).
15. Mouatadid, S. *et al.* Adaptive bias correction for improved subseasonal forecasting. *Nat. Commun.* **14**, 3482 (2023).
16. Bi, K. *et al.* Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).
17. Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
18. Rasp, S. *et al.* Weatherbench: a benchmark data set for data-driven weather forecasting. *J. Adv. Model. Earth Syst.* **12**, e2020MS002203 (2020).
19. Rasp, S. *et al.* Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *J. Adv. Model. Earth Syst.* **16**, e2023MS004019 (2024).
20. Extreme weather bench. https://github.com/brightbandtech/extremeweatherbench (2024).
21. Chen, L. *et al.* A machine learning model that outperforms conventional global subseasonal forecast models. *Nat. Commun.* **15**, 6425 (2024).
22. Hatanpää, V. *et al.* Aeris: Argonne earth systems model for reliable and skillful predictions. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 72–85 (2025).
23. Stock, J., Arcomano, T. & Kotamarthi, R. Swift: An autoregressive consistency model for efficient weather forecasting. *arXiv preprint arXiv:2509.25631* (2025).
24. Wheeler, M. C. & Hendon, H. H. An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. *Mon. weather review* **132**, 1917–1932 (2004).
25. Zhang, C. Madden-julian oscillation. *Rev. Geophys.* **43** (2005).
26. Kim, H., Vitart, F. & Waliser, D. E. Prediction of the madden–julian oscillation: A review. *J. Clim.* **31**, 9425–9443 (2018).
27. Domeisen, D. I. *et al.* The role of the stratosphere in subseasonal to seasonal prediction: 1. predictability of the stratosphere. *J. Geophys. Res. Atmospheres* **125**, e2019JD030920 (2020).
28. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. journal royal meteorological society* **146**, 1999–2049 (2020).
29. Dee, D. P. *et al.* The era-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. royal meteorological society* **137**, 553–597 (2011).
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
31. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).

32. Yan, L. Topology-based visualization techniques for scientific data exploration. *IEEE Comput. Graph. Appl.* **45**, 89–98 (2025).

33. Yan, L. *et al.* Scalar field comparison with topological descriptors: Properties and applications for scientific visualization. In *Computer Graphics Forum*, vol. 40, 599–633 (Wiley Online Library, 2021).

34. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**, 600–612 (2004).

35. Goodfellow, I. J. *et al.* Generative adversarial nets. *Adv. neural information processing systems* **27** (2014).

36. Edelsbrunner, H., Harer, J. *et al.* Persistent homology-a survey. *Contemp. mathematics* **453**, 257–282 (2008).

37. Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, 263–271 (2005).