https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

Increasing Resolution and Accuracy in Sub-Seasonal Forecasting through 3D U-Net: the Western US

Jihun Ryu[1], Hisu Kim[2], Shih-Yu (Simon) Wang[3], and Jin-Ho Yoon[1]

[1]School of Environment and Energy Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea
[2]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea
[3]Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA

Correspondence: Jin-Ho Yoon (yjinho@gist.ac.kr)

Abstract. Sub-seasonal weather forecasting is a major challenge, particularly when high spatial resolution is needed to capture complex patterns and extreme events. Traditional Numerical Weather Prediction (NWP) models struggle with accurate forecasting at finer scales, especially for precipitation. In this study, we investigate the use of 3D U-Net architecture for post-processing sub-seasonal forecasts to enhance both predictability and spatial resolution, focusing on the western U.S. Using the ECMWF ensemble forecasting system and high-resolution PRISM data, we tested different combinations of ensemble members and meteorological variables. Our results demonstrate that the 3D U-Net model significantly improves temperature predictability, and consistently outperforms NWP models across multiple metrics. However, challenges remain in accurately forecasting extreme precipitation events, as the model tends to underestimate precipitation in coastal and mountainous regions. While ensemble members contribute to forecast accuracy, their impact is modest compared to the improvements achieved through downscaling. This study lays the groundwork for further development of neural network-based post-processing methods, showing their potential to enhance weather forecasts at sub-seasonal timescales.

1 Introduction

Sub-seasonal forecasting based on numerical weather prediction (NWP) models has made significant advances over the past few decades, with the ability to predict extreme events such as heat waves up to four weeks in advance (Ardilouze et al., 2017; Vitart and Robertson, 2018). However, limitations still exist, which has led to increasing interest in deep learning models as alternative approaches for weather forecasts. Some models directly generate the forecasts from the input data. Weyn et al. aimed to provide ensembles similar to those in NWP systems. Two deep learning models, Graphcast and Pangu, have outperformed NWP in weather and medium-range forecasts, from 1 day to 10 days (Bi et al., 2023; Lam et al., 2023). More recently, deep learning models such as Fuxi-S2S have been reported to surpass NWP in sub-seasonal forecasting (Chen et al., 2024). Among them, Graphcast does not provide precipitation forecasting, while these models only generate deterministic forecasts and struggle with predicting extreme weather events (Olivetti and Messori, 2024).

On the other hand, post-processing NWP outputs have also been explored as a means of improving forecast accuracy (Woolnough et al., 2024). In recent years, neural network-based post-processing methods have gained traction. U-Net architecture

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

has been widely utilized for weather forecast post-processing due to its ability to capture fine details through contracting and expanding layers (Horat and Lerch, 2024; Faijaroenmongkol et al., 2023; Deng et al., 2023; Xin et al., 2024). U-Net has also shown potential in probabilistic forecasting for sub-seasonal predictions (Horat and Lerch, 2024). Furthermore, U-Net was employed to correct biases in seasonal precipitation forecasts in Thailand (Faijaroenmongkol et al., 2023).

Moreover, generating high-resolution NWP outputs demands significant computational resources, so deep learning has been applied to downscale sub-seasonal forecasts and simultaneously improve predictability efficiently. For example, studies in wildfire weather forecasting in the western United States have successfully downscaled predictions to the county level (Son et al., 2022). Another example is the improved predictability and downscaling of temperature and precipitation in China, achieved by using a weighted combination of multiple models based on U-Net (Xin et al., 2024).

A key consideration in these studies is the selection of input data. Some studies use only target variables (Xin et al., 2024), while others use a broader subset of variables (Horat and Lerch, 2024; Weyn et al., 2021). The extent to which inputs significantly affect sub-seasonal forecasting remains undetermined and case-sensitive. Even though studies on weather forecasts have found that sub-variables play a limited role in temperature forecasting, they have demonstrated improvements in wind gust predictions (Rasp and Lerch, 2018; Schulz and Lerch, 2022). Additionally, attempts to utilize each ensemble member of the NWP for U-Net training resulted in only marginal improvements in weather forecasting accuracy. (Höhlein et al., 2024).

This study enhances predictability in the Western United States through the 3D U-Net-based post-processing plus downscaling forecasts to higher spatial resolutions. In doing so, we identify the role played by ensemble members and sub-variables in enhancing predictability and investigate whether downscaling with neural networks leads to meaningful improvements at smaller scales such as the county level. Section 2 describes the data, 3D U-Net architecture, pre-processing, and evaluation metrics, while Section 3 discusses the results and analysis. Lastly, conclusions are presented in Section 4.

2 Data and Method

2.1 Data

This study employs two primary datasets: the European Centre for Medium-Range Weather Forecasts (ECMWF) real-time perturbed forecasts and the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset. First, as the ECMWF forecast model from sub-seasonal to seasonal (S2S) prediction project continues to evolve, providing an increasing number of ensemble members, forecast periods, and forecast cycles, we select forecasts from CY40R1 with a $1.5^\circ \times 1.5^\circ$ resolution, 50 ensemble perturbation forecasts, twice-weekly forecast cycles, and 32-day lead times (Roberts et al., 2018). We utilize forecasts from CY40R1 to CY48R1, covering the period from January 2015 to December 2023. These forecasts span weather to sub-seasonal time scales, offering a comprehensive range of meteorological variables essential for our neural network post-processing model. Next, we utilize the daily PRISM dataset, developed by Oregon State University, which provides high-resolution climate data for the United States (Daly et al., 2008) for the sake of model validation and high-resolution reference data. PRISM offers grid estimates of variables at a fine spatial resolution of $0.042^\circ \times 0.042^\circ$ (approximately 4 km). Only data from January 2015 to January 2024 is used, corresponding to the period of ECMWF forecasts utilized in this study.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

We chose the Western United States because it is a diverse region, ranging from coastal areas to high mountain ranges, and the importance of water management emerges in the face of hydrological changes driven by the climate crisis (Siirila-Woodburn et al., 2021). To evaluate the model's performance at finer spatial scales, we select five diverse regions in the Western United States, each representing different climatological socio-economic characteristics. These regions include three highly populated urban areas and two important agricultural zones. In detail, we choose (1) San Francisco, California, a major high-populated metropolitan area with a unique coastal climate; (2) Orange County, California, known for its citrus farming and Mediterranean climate; (3) the area around the Great Salt Lake in Utah, which combines high population density with a distinctive lake-effect climate; (4) Seattle, Washington, representing the Pacific Northwest's urban environment and maritime climate; and (5) a vast wheat farming region in eastern Washington, exemplifying the inland agricultural areas of the West.

2.2 3D U-Net Architecture

The post-processing approach utilizes U-Net architecture, originally proposed by Ronneberger et al. for biomedical image segmentation. The U-Net is particularly well-suited for our task of enhancing sub-seasonal forecasts due to its ability to capture multi-scale features and preserve spatial information through skip connections (Horat and Lerch, 2024). We leverage this architecture to process the data above, as shown in Figure 1 properly.

The 3D U-Net structure consists of a contracting path (encoder) and an expanding path (decoder), connected by a bottleneck layer. Our implementation features three contracting-expanding cycles, optimized for the spatial scales relevant to sub-seasonal forecasting. The contracting path progressively reduces spatial dimensions while increasing feature channels, allowing the model to capture broader contextual information. Conversely, the expanding path restores spatial resolution, enabling precise localization of weather patterns. In short, this structure concatenates feature maps from the contracting path to the expanding path so that the model retains fine-grained spatial information that might otherwise be lost during downsampling.

We train the model using ECMWF forecast fields as input and high-resolution PRISM observations as the target output. To investigate the impact of ensemble forecasting on post-processing performance, we conduct experiments with different combinations of ensemble members; Using only the first ensemble member (E01), utilizing all 50 ensemble members (E50), and merely exploiting the mean of all 50 ensemble members (E50M). Further, we explore the impact of input variable selection on model performance by testing configurations with varying numbers of meteorological sub-variables (V1, V2, V4, V8). This exploration aims to determine whether incorporating sub-variables beyond the target variable could enhance the model's predictive capabilities.

In our specific implementation, we integrate the ensemble members and variables into a channel, utilizing a 3D U-Net structure with forecast lead time, latitude, and longitude as the dimensions. The forecast period ranges from 1 day to 32 days ahead, with a longitude range of $235.5°$ to $253.25°$ and a latitude range of $31.25°$ to $49°$, consisting of 72 grid points in each. Based on the forecast start date, The training period spans from January 2015 to December 2020, the validation period from March 2021 to February 2022, and the test period from January 2023 to December 2023. The loss function incorporates mean squared error and pattern correlation, and we use the Adam with a learning rate of 0.001 for model optimization.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

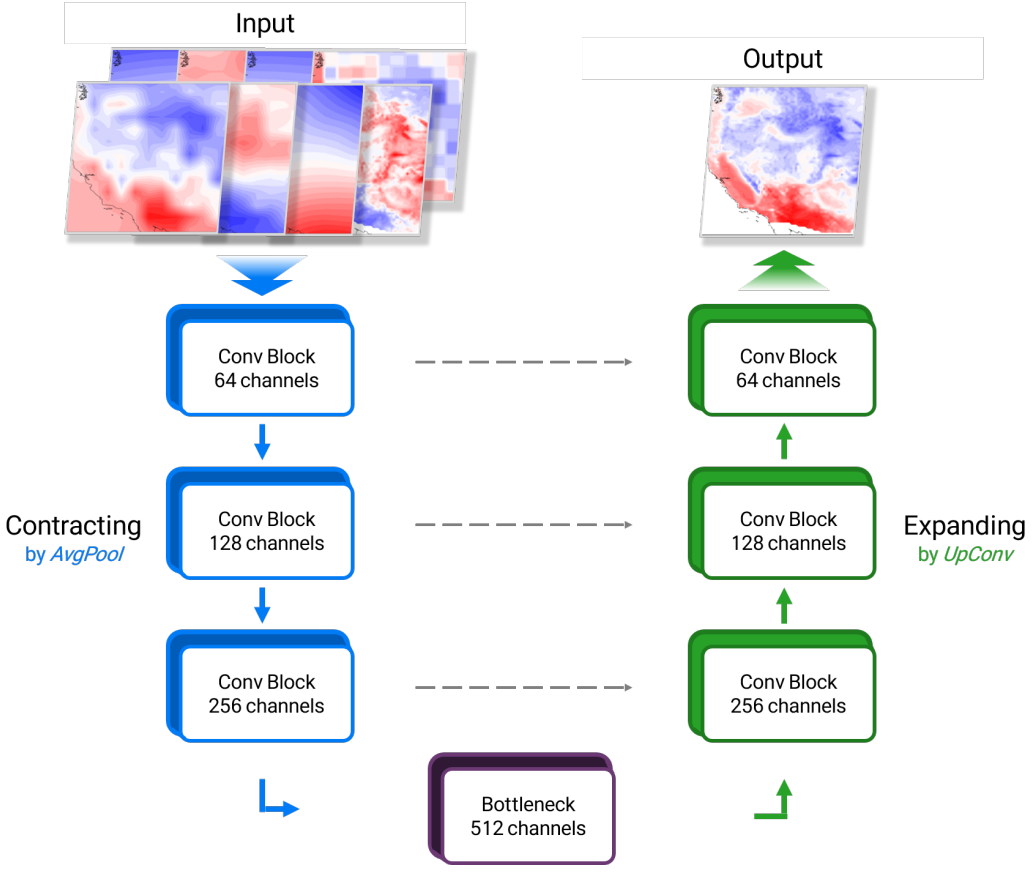

Figure 1. Schematic of the 3D U-Net architecture adapted for weather forecast post-processing. The model consists of a contracting path (left, blue), an expanding path (right, green), and a bottleneck layer (center, purple), with skip connections (dashed gray arrows) preserving spatial information.

2.3 Pre-processing

To assess the sensitivity of the sub-variables used in the learning process, we select 16 variables: 2m temperature, precipitation, total column water (TCW), mean sea level pressure (MSLP), 10m u-wind (u10), 10m v-wind (v10), elevation, and geopotential height (z), along with u-wind (u) and v-wind (v) at the 850hPa, 500hPa, and 200hPa levels. The dataset is split into two pre-processing groups, one being precipitation and TCW, and the other being topography and the remaining atmospheric variables. For precipitation and TCW, any negative values are set to zero, as they are non-physical for these types of data. We then apply conservative interpolation to ensure the accurate preservation of values during spatial adjustments. For the remaining variables, linear interpolation was applied. All variables are interpolated onto a $0.25° \times 0.25°$ latitude-longitude grid. Given the established relationship between the mean state and predictability (Ryu et al., 2024), we calculate the mean state of each sub-variable across both weather and sub-seasonal timescales. The pattern correlation coefficient between the mean state of

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

each sub-variable and the target variable is then determined. The absolute values of these correlations are averaged, and the top eight variables with the highest correlations are selected for further analysis (Figure S1, and Table S1).

The interpolated dataset is further processed for input into deep learning models. For precipitation and TCW, to handle zero values, we transform the data by adding 1 and applying a log10 transformation, following established methods (Aich et al., 2024). The transformed data is then standardized by calculating the mean and standard deviation, making it suitable for use in the 3D U-Net architecture. For the other variables, we follow standardization by computing the mean and standard deviation, similar to the pre-processing approach used in Graphcast (Lam et al., 2023). This normalization step ensures that all variables are prepared for efficient training in the 3D U-Net model.

2.4 Evaluation Metrics

Intending to assess the performance of our 3D U-Net-based post-processing model comprehensively, we employ the following three key evaluation metrics; pattern correlation, root mean square error (RMSE), and E_{pre} (Ryu et al., 2024). The first two metrics quantify the model's ability to capture the spatial patterns of temperature and precipitation and the average magnitude of forecast errors. Both metrics are commonly selected to evaluate sub-seasonal predictions. Lastly, we incorporate the E_{pre} metric, which builds upon the concept of Taylor diagrams and has been utilized in several studies for evaluating forecast performance (Ryu et al., 2024; Wang et al., 2021; Yang et al., 2013). This metric offers a comprehensive assessment by integrating both the variance ratio and the correlation between predictions and observations. We measure E_{pre} per lead time by averaging values of all initial dates.

$$E_{pre} = \frac{1}{N}\sum_{i=1}^{N} \log\left[\frac{\left(\frac{\sigma_{obs,i}}{\sigma_{pre,i}} + \frac{\sigma_{pre,i}}{\sigma_{obs,i}}\right)^2 (1+r_0)^4}{(1+r_i)^4}\right] \tag{1}$$

Here, $\sigma_{obs,i}$ and $\sigma_{pre,i}$ denote the standard deviations of observed and predicted values respectively. r_0 represents an ideal correlation (set to 1), and r_i is the actual correlation at time step i. N stands for the number of initial dates. The E_{pre} metric is designed to yield a value of 0 for perfect predictions, with increasing values indicating greater discrepancies between forecasts and observations. By incorporating both spread and accuracy considerations, this metric proves particularly valuable for evaluating the nuanced performance of ensemble predictions in sub-seasonal forecasting contexts.

3 Results and Discussion

3.1 Role of Ensemble and Variables

The performance of the 3D U-Net model in comparison to traditional NWP forecasts was evaluated across different ensemble configurations and input variable sets. A comprehensive overview of all cases in Figure S2 in the supplementary material demonstrates a clear superiority of 3D U-Net models over NWP in both forecasts across all three evaluation metrics. The focus of this analysis is the relationship between ensemble size and forecast skill. Next, we clustered experiments with the same sub-variables and ensemble configurations to validate the role of auxiliary variables and ensembles.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

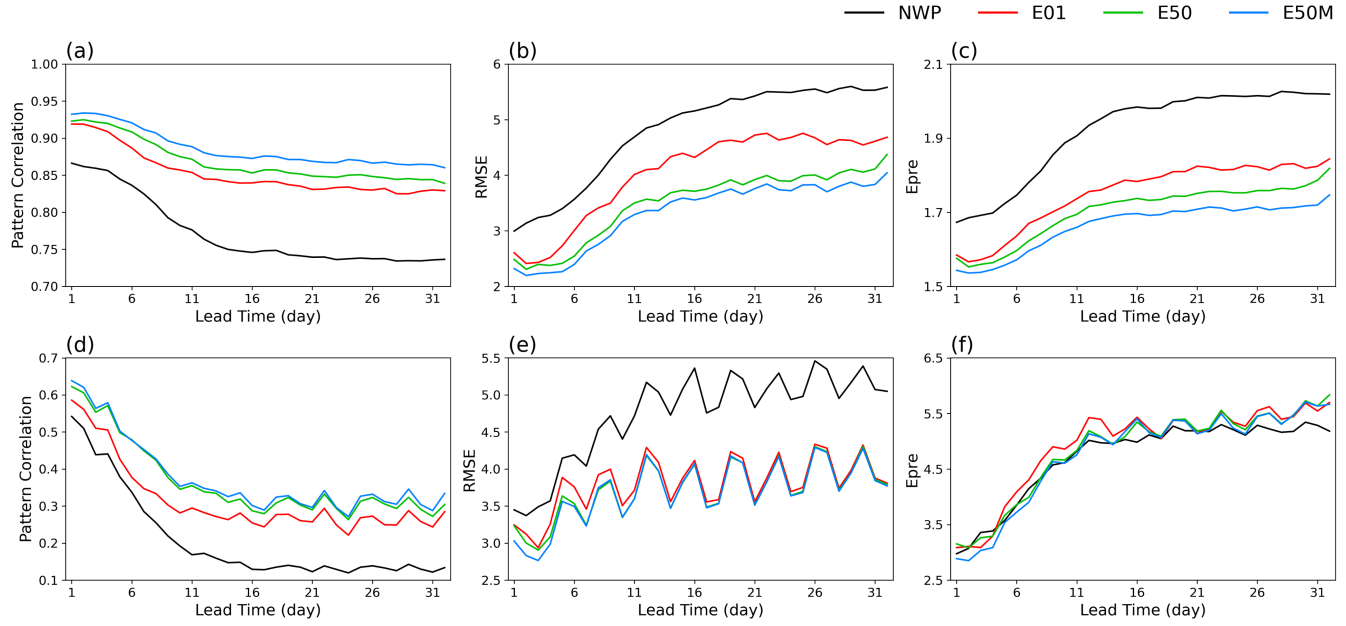

Figure 2. Ensemble sensitivity benchmark scores for the Western U.S., comparing NWP and 3D U-Net models (E01, E50, E50M) for temperature (top row) and precipitation (bottom row) forecasts over 32 days. Columns show (a, d) pattern correlation, (b, e) RMSE, and (c, f) E_{pre}, respectively.

Figure 2 illustrates benchmark scores with respect to the ensemble configurations. In temperature predictions, E50M (see Section 2.2) shows the best performance and E01 is the most deficient in all matrices. Precipitation predictions also exhibit analogous operational characteristics; nonetheless, E50 and E50M exhibit significantly aligned trajectories and the overall disparity among all configurations has diminished in both pattern correlation and RMSE metrics. In contrast, E_{pre} for precipitation does not show significant differences between NWP, likely due to limitations in precipitation variance. Results in Figure 2 interestingly imply that E01 proved insufficient for effective learning by only the first ensemble member, resulting in performance lagging behind the other ensemble configurations. However, the performance difference between using E50 and E50M was negligible. To support these findings, we performed additional experiments on ensembles of 10, 20, and 25 members, which showed similar performance to E50 (Figure not shown). This suggests that while post-processing significantly improves forecast skill, the benefits of increasing ensemble members beyond the mean are limited for both temperature and precipitation predictions in the current setting. This is consistent with previous research that ensemble spread plays a limited role in improving weather forecast accuracy, and these findings suggest that this limitation extends to sub-seasonal forecasts as well (Höhlein et al., 2024). In other words, using the ensemble mean could be sufficient for achieving optimal performance with the 3D U-Net model.

The impact of input variables on model performance is further explored in Figure 3. The 3D U-Net models consistently outperform NWP across all lead times for both temperature and precipitation forecasts. This superiority reinforces the ro-

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

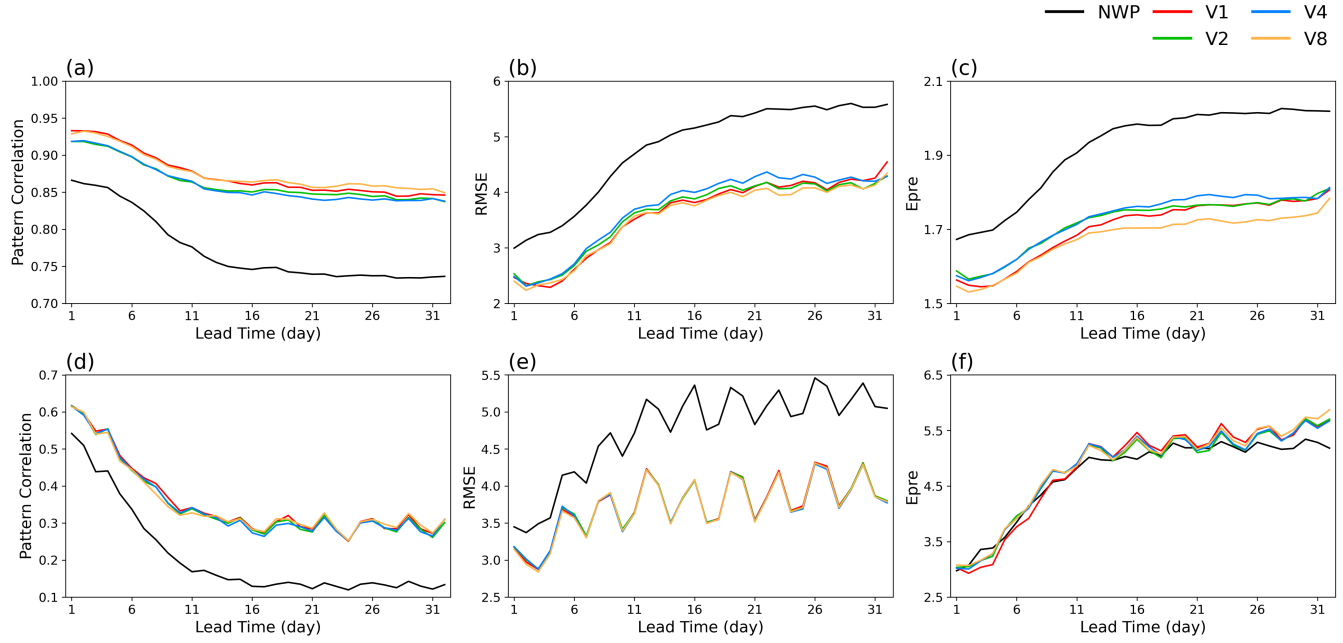

Figure 3. Sub-variable sensitivity benchmark scores for the Western U.S., comparing NWP and 3D U-Net models with varying input variables (V1, V2, V4, V8) for temperature and precipitation forecasts over 32 days. Layout is the same as in Figure 2.

bustness of the neural network approach. Specifically, V8 slightly surpasses the others in all temperature metrics. This result aligns with previous studies that suggest elevation can enhance temperature post-processing accuracy (Rasp and Lerch, 2018). However, for precipitation forecasts, V1, V2, V4, and V8 reveal insignificant variations in terms of their predictability scores. The E_{pre} values for precipitation exhibited comparable patterns to those observed in Figure 2(f), attributable to analogous underlying mechanisms. An intriguing observation is that the performance differences among the 3D U-Net models with varying numbers of input variables are minimal for both target variables. This contrasts with prior research, which has suggested that sub-variables contribute to forecast improvement (Schulz and Lerch, 2022). However, our finding is consistent with studies that indicate sub-variables may contribute only marginally or in a limited role, particularly when used mean state (Rasp and Lerch, 2018; Höhlein et al., 2024). This indicates that increasing the number of sub-variables in the 3D U-Net model does not significantly enhance its ability to extract relevant information or improve forecast skills in this context. Such a result challenges the conventional wisdom that more input data invariably leads to better predictions, and suggests that the 3D U-Net architecture in the current setting may be efficiently capturing the most relevant features for the prediction even with a limited set of input variables. Thus we use E50M V1 and V8 for the following analysis.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

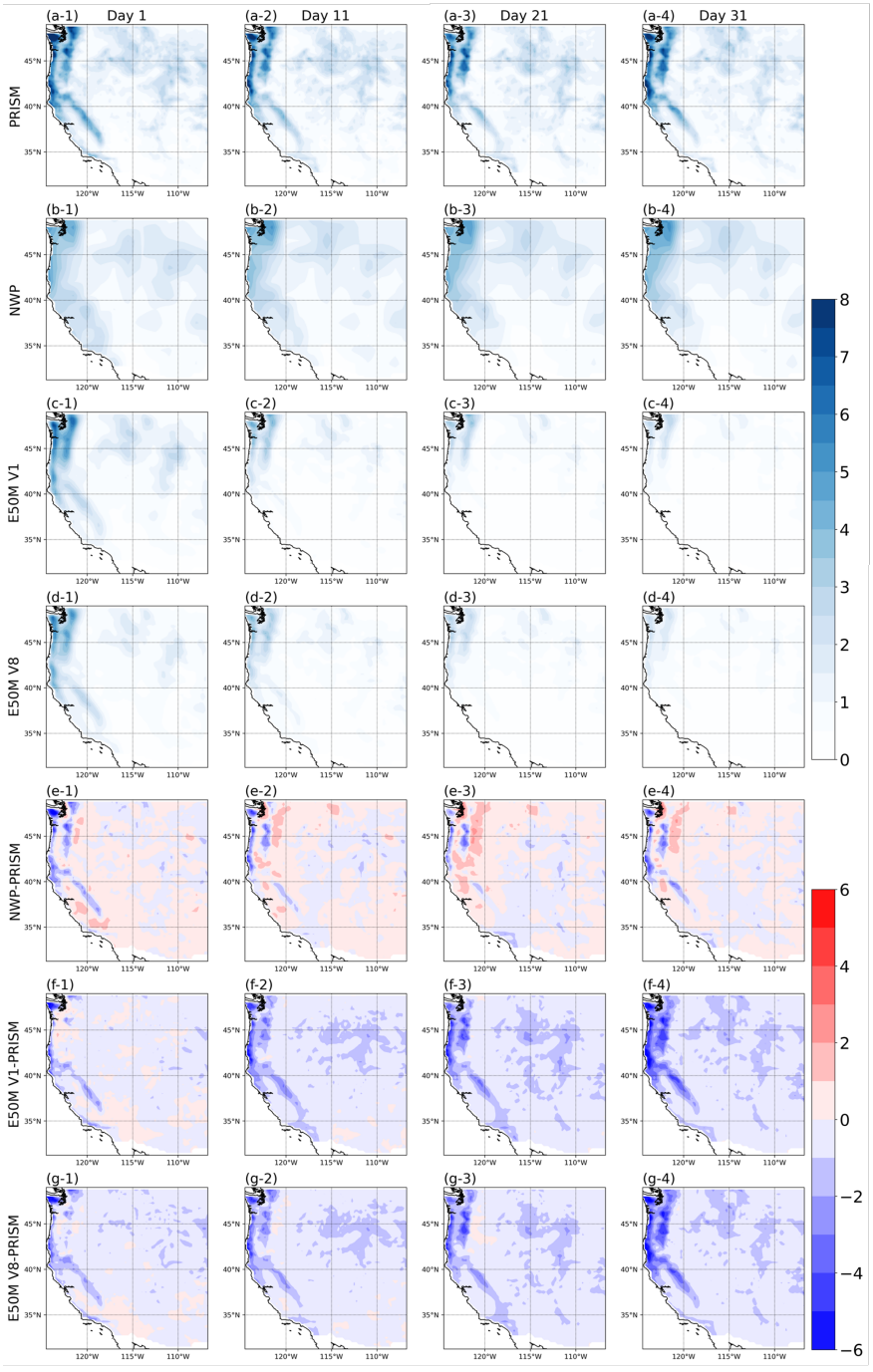

Figure 4. Comparison of precipitation forecasts (104 forecasts in 2023 are averaged) and differences across lead times for the Western U.S. Rows represent (a) PRISM observations, (b) NWP forecasts, (c) 3D U-Net E50M V1 predictions, (d) 3D U-Net E50M V8 predictions, (e) differences between NWP and PRISM, (f) differences between E50M V1 and PRISM, and (g) differences between E50M V8 and PRISM. Precipitation amounts (rows a-d) are shown using the scale in the upper color bar (0-8 mm/day), while differences (rows e-g) are depicted using the scale in the lower color bar (-6 to +6 mm/day). Columns show forecasts for Days 1, 11, 21, and 31.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

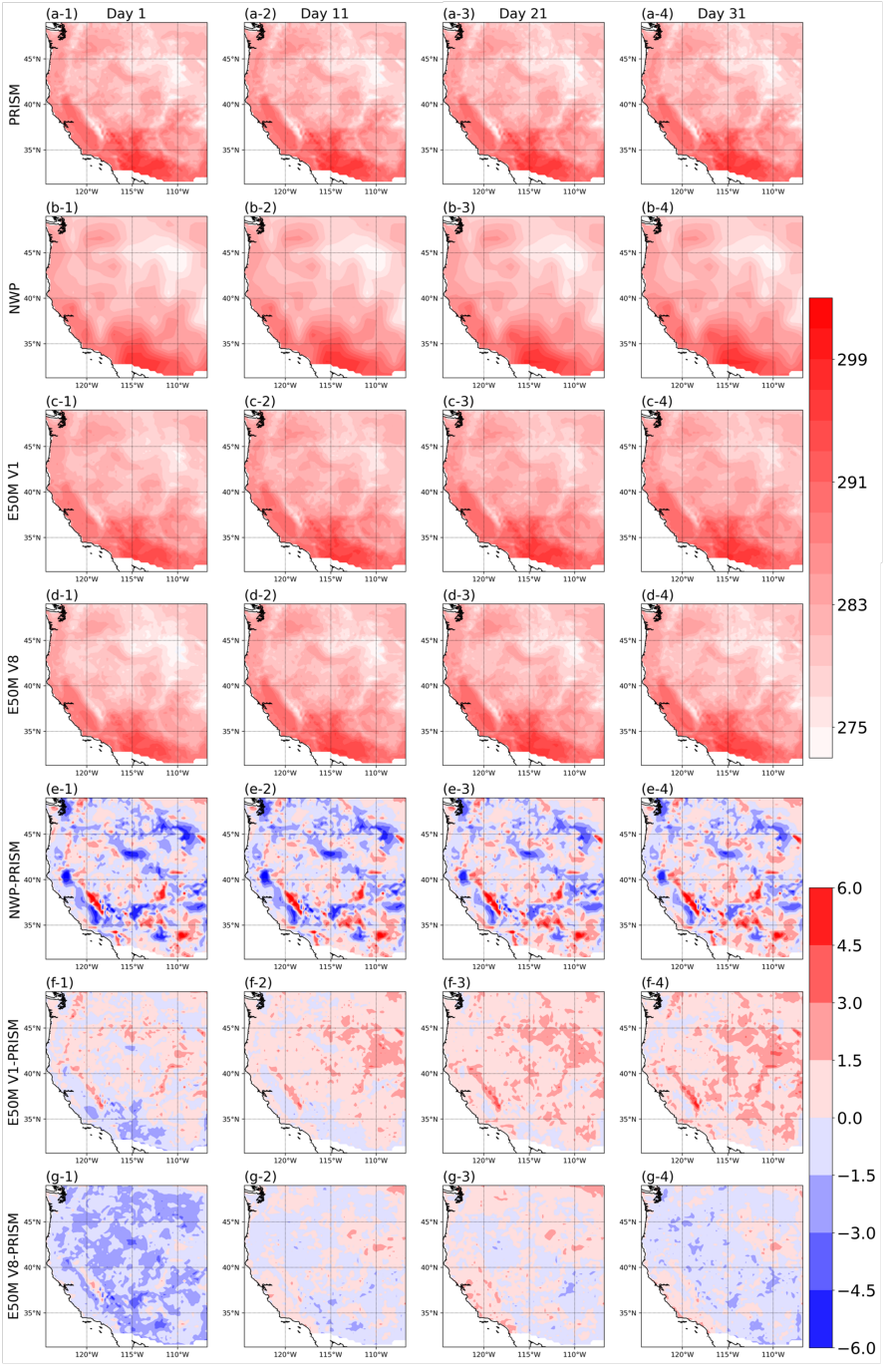

Figure 5. Temperature forecasts and differences similar to Figure 4. The layout is the same, with temperature values and differences shown in Kelvin (K).

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

3.2 Predictability and Downscaling

Next, we compare the spatial pattern of the forecast between NWPs and E50M 3D U-Net with both V1, which uses only the target variables, and V8, which includes all variables. The 3D U-Net model demonstrates significant improvements in both predictability and downscaling capabilities for temperature forecasts. While precipitation forecasts also show improvement, the gains are less pronounced than for temperature. For precipitation (Figure 4), the 3D U-Net models achieve higher spatial resolution compared to NWP, revealing fine-scale patterns. However, a consistent underestimation of precipitation is observed across all lead times, particularly in coastal and mountainous regions, regardless of the number of input variables used. This reduction in precipitation is also observed in other regions during downscaling and U-Net-based post-processing (Xin et al., 2024). Temperature forecasts (Figure 5) showcase more substantial improvements. The 3D U-Net models significantly enhance spatial resolution and reduce overall forecast errors compared to NWP. The 3D U-Net approach, especially E50M V8, captures fine-scale temperature patterns effectively, showing reduced biases across various terrain types.

The performance of the 3D U-Net model in extreme cases provides further insights into its capabilities and limitations. Figure 6 presents an extreme precipitation event in California from March 7 to March 13, 2023. The 3D U-Net models (E50M V1 and V8) demonstrate improved spatial detail compared to NWP. On March 10, the 3D U-Net model captures the rainfall that NWP doesn't (Figure 6(b-4, c-4, d-4)) and specifies the location, both coastal area and inland, more accurately on 2023-03-11. Even so, the models still struggle with accurately capturing the intensity of heavy precipitation events. Increasing the training data can be one alternative to improve precipitation extremes (Hu et al., 2023). Alternatively, this limitation may stem from the post-processing technique itself and warrants further investigation.

Even in extreme temperature cases, Figures S3 and S4 confirm results that 3D U-Nets are superior to NWP. The overall performance of E50M V1 in the high-temperature case and E50M V8 in the low-temperature case, as well as the overall differences and recovery from cold waves, appear to outperform NWP. However, limitations are evident, highlighting the persistent challenges in predicting extreme events despite the improved spatial resolution.

The contrasting performance between precipitation and temperature forecasts underscores the varying complexities in predicting these two variables. Although some challenges are left in precipitation forecasting, the 3D U-Net model's ability to capture fine-scale patterns and improve spatial resolution for both variables represents a significant advancement. These results suggest that with further refinement, particularly in handling extreme events and complex terrain interactions, neural network-based post-processing methods like 3D U-Net have the potential to substantially improve both temperature and precipitation forecasts at sub-seasonal timescales.

3.3 Predictability in County-scale

To assess the model's performance at finer spatial scales, crucial for local decision-making and resource management, we evaluate forecasts for five selected county-level regions in the Western U.S. Figure 7 presents comprehensive performance metrics for temperature and precipitation forecasts across these 5 regions, comparing NWP and various 3D U-Net model configurations over a 32-day lead time. For temperature forecasts at the county scale, 3D U-Net models generally demonstrate improved

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

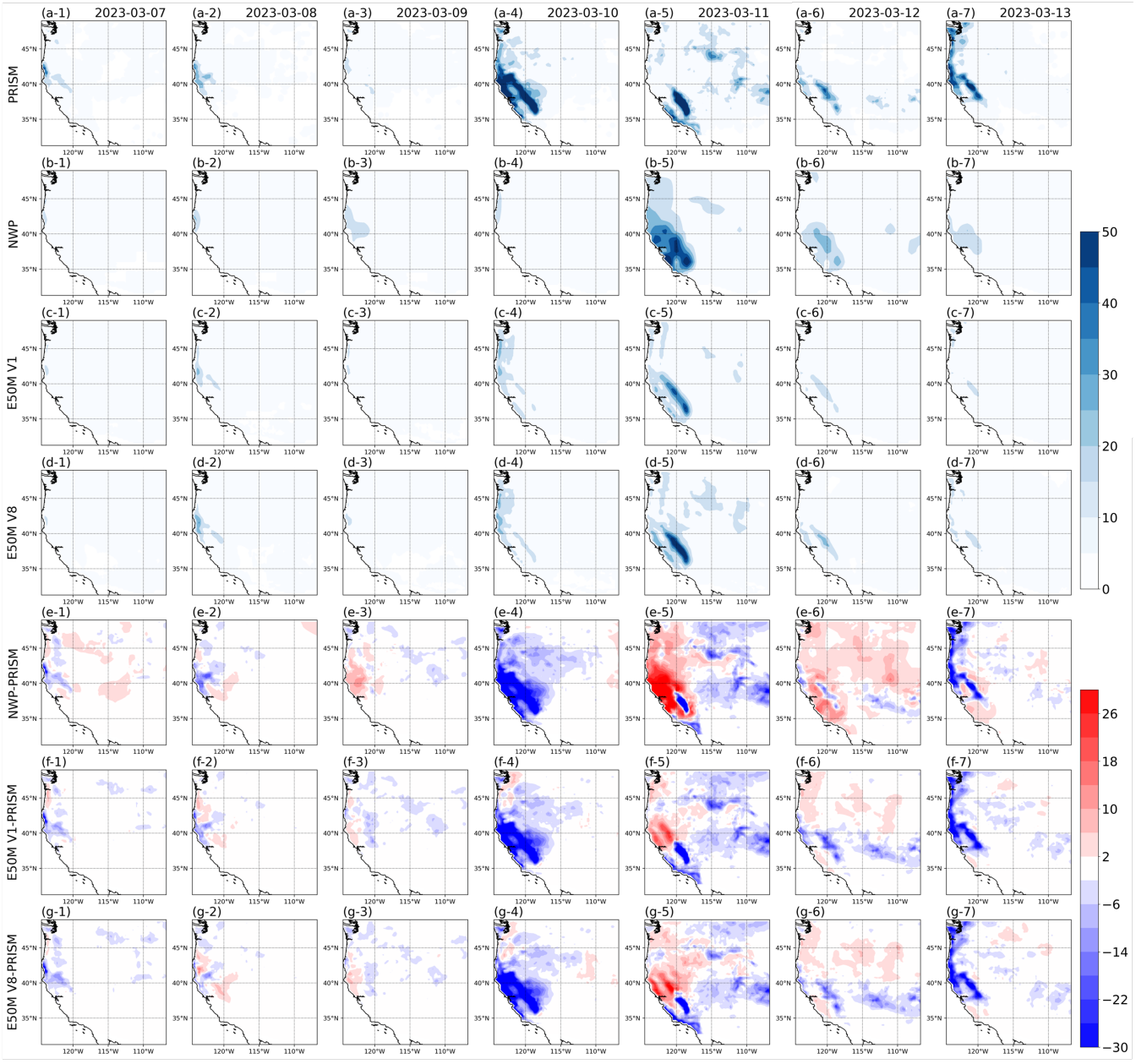

Figure 6. Daily precipitation forecasts for the Western U.S. from March 7 to March 13, 2023, with initial condition on March 6. In other words, March 7 (13) is the forecast with lead day 1 (7). The layout is the same as Figure 4.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

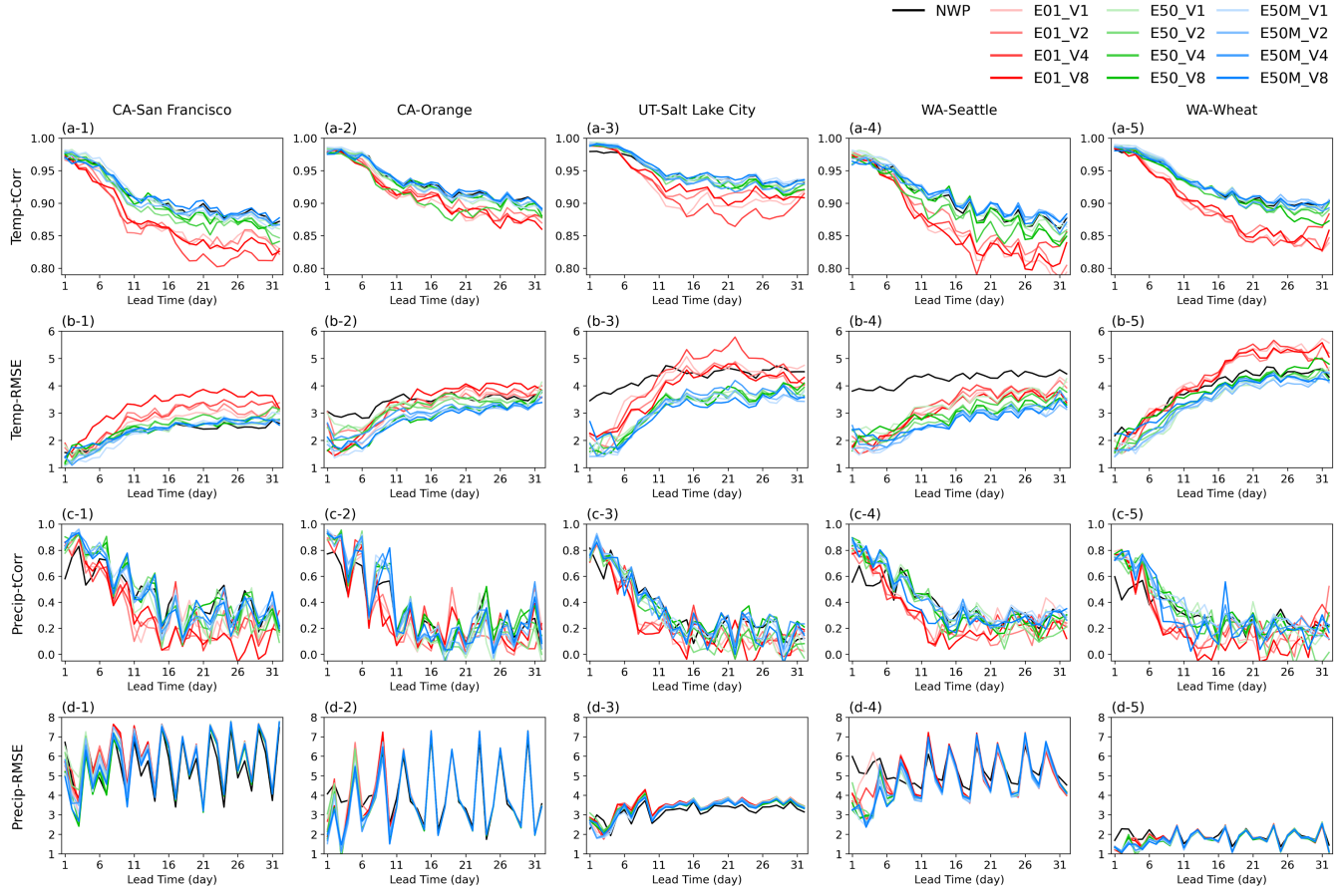

Figure 7. Performance metrics for temperature (Temp) and precipitation (Precip) forecasts across five county-level regions in the Western U.S., comparing NWP and various 3D U-Net model configurations (E01, E50, E50M with V1-V8 input variables) over a 32-day lead time. Metrics include $tCorr$ and RMSE. (a) Temperature correlation, (b) Temperature RMSE, (c) Precipitation correlation, (d) Precipitation RMSE. Columns represent different regions: (1) San Francisco, CA (2) Orange farm, CA, (3) Salt Lake City, UT, (4) Seattle, WA, and (5) Wheat farming area, WA.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

or comparable performance relative to NWP while the degree of enhancement varies significantly across regions. Along with the result of Section 3.1, E50M surpasses the other ensemble configuration's scores, and no conspicuous performance difference between varying the number of variables yields. Some areas, such as Seattle, show more pronounced enhancements in predictability, possibly due to the region's more uniform maritime climate. In contrast, areas with more complex terrain or microclimates show more modest improvements, highlighting the persistent challenges in downscaling to highly localized conditions.

For precipitation forecasts, 3D U-Net models enhance correlation on the weather scale but not on the sub-seasonal scale especially in two regions in Washington (Figure 7(c-4,5)). Moreover, correlation exhibits higher variability in performance across different 3D U-Net configurations compared to temperature forecasts. Figure 7(d) reveals a complex pattern. In most regions, 3D U-Net models and NWP show comparable RMSE values, with neither consistently outperforming the other across all lead times.

Note that the performance differences among 3D U-Net configurations in both targets are generally small at this county scale, consistent with the findings at larger spatial scales. This implies that the benefits of increased model complexity may diminish at very fine spatial resolutions, where local factors become increasingly dominant.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

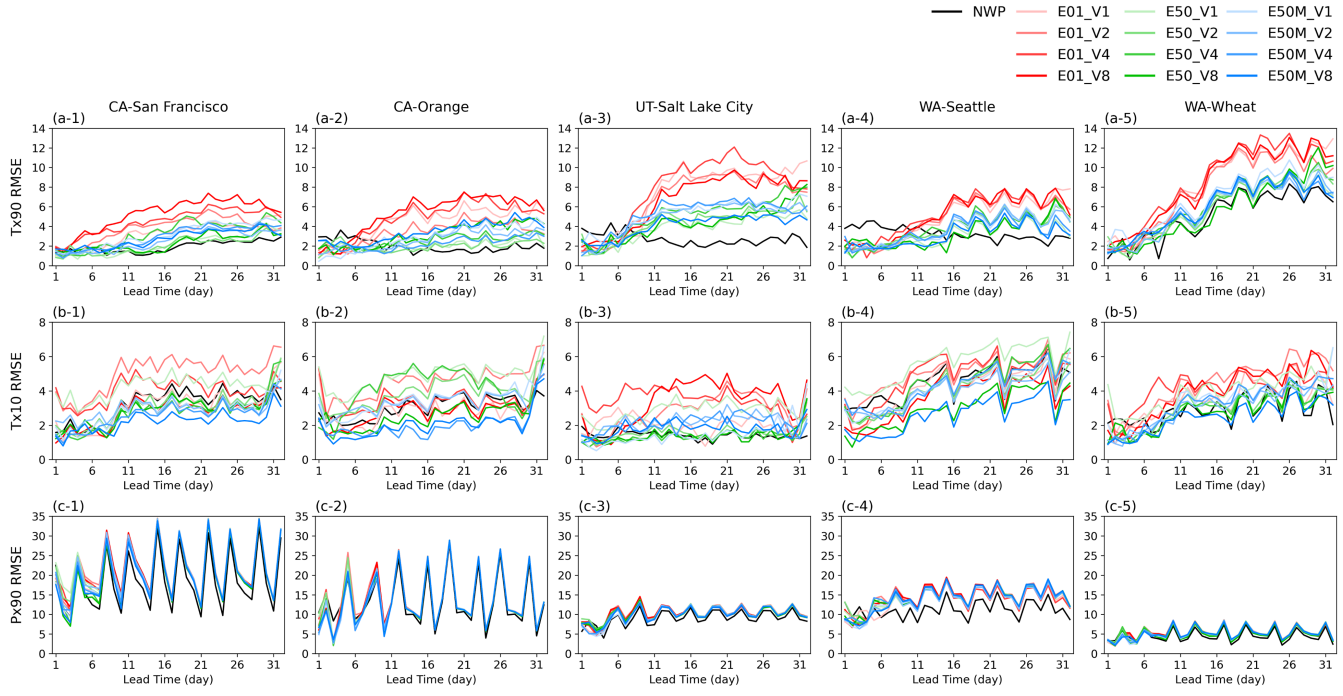

Figure 8. RMSE comparison across five U.S. counties for extreme temperature and precipitation forecasts. Results show NWP and various 3D U-Net configurations (E01, E50, E50M) with different input variables (V1-V8). Temperature metrics include 90th percentile (tx90) and 10th percentile (tx10). Precipitation uses 90th percentile (px90). (a) Temperature tx90 RMSE, (b) Temperature tx10 RMSE, (c) Precipitation px90 RMSE. Columns represent the same regions as in Figure 7.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

Figure 8 elaborates the RMSE of each five regions regarding heat, cold, and precipitation extremes. Even though predicting heat extremes in Salt Lake City and Seattle in weather-scale is improved (Figure 8(a-3,4)), 3D U-Net models don't outperform NWP in these extreme cases. Still, performance varies considerably on location, period, and extreme type. These can be attributed to several factors: Deep learning models are predisposed to yield results in which extreme values are smooth-out, called blurring effect (Lam et al., 2023), and tend to converge to the mean state (Bonavita, 2024). Furthermore, Olivetti and Messori highlights the similar result of Figure 8 that global scale deep learning models often struggle with capturing the full range of variability in extreme events, especially in long-term prediction.

4 Conclusions

The findings of this study highlight the dual benefits of using the 3D U-Net architecture for sub-seasonal forecasting, namely enhanced accuracy and improved spatial resolution. By applying 3D U-Net-based post-processing to NWP models, the study demonstrated significant improvements in predicting both temperature and precipitation, especially in complex terrains and localized regions. The model's ability to downscale forecasts to higher spatial resolutions provided finer details, which are crucial for decision-making in regional disaster management. Nonetheless, some possible drawbacks remain evident. First and foremost, there was a spatial pattern improvement in precipitation, but the underestimation of precipitation in coastal and mountainous areas persisted. The added diversity in data could not resolve these limitations. Second, predicting extreme precipitation events with high accuracy is a challenging task. While the 3D U-Net could capture general patterns and improve spatial details, it still struggled to fully enhance extreme forecasts' accuracy.

The 3D U-Net model showed mixed performance for both temperature and precipitation forecasts at the county level. While 3D U-Net outperformed NWP models in predicting temperature such as in Seattle, its performance in precipitation forecasting was less consistent. The model was able to enhance spatial resolution and predictability for temperature at finer scales but struggled to deliver comparable improvements for precipitation. While the 3D U-Net model is effective for downscaling temperature forecasts at the county level, further refinement is needed to improve its ability to capture precipitation patterns, particularly in regions with complex weather dynamics.

In conclusion, 3D U-Net's integration into sub-seasonal forecasting models offers substantial improvements such as capturing fine-scale weather patterns over traditional NWPs while maintaining computational efficiency. This model's ability makes it a promising tool for a wide range of atmospheric science applications, from short-term weather to sub-seasonal predictions. Still, further refinements are necessary to address the underperformance in extreme event predictions and to explore the optimal balance between model complexity and forecast skill. Future work could focus on enhancing the model's ability to handle extreme weather conditions and exploring new approaches for post-processing sub-seasonal predictions in diverse and complex terrains.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

Code and data availability. The ECWMF perturbed forecast can be downloaded from https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/. The PRISM dataset can be downloaded from https://prism.oregonstate.edu/. The model code is archived on Zenodo(Ryu et al., 2025), and at https://zenodo.org/records/14776781

Author contributions. The study was conceptualized by Jihun Ryu, and Jin-Ho Yoon. Jihun Ryu has done the data analysis, visualization, and writing the original draft. Hisu Kim has done the data analysis and writing the original draft. Reviewing and editing the manuscript is done by Shih-Yu (Simon) Wang and Jin-Ho Yoon.

Competing interests. The authors declare no competing interests.

Acknowledgements. This research is funded by the National Research Foundation of Korea. Shih-Yu (Simon) Wang is supported by the U.S. Department of Energy/Office of Science under Award DE-SC0016605 and the U.S. SERDP Project RC20-3056.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

References

Aich, M., Hess, P., Pan, B., Bathiany, S., Huang, Y., and Boers, N.: Conditional diffusion models for downscaling & bias correction of Earth system model precipitation, arXiv preprint arXiv:2404.14416, 2024.

Ardilouze, C., Batté, L., and Déqué, M.: Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave, Advances in Science and Research, 14, 115–121, 2017.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, 2023.

Bonavita, M.: On some limitations of current machine learning weather prediction models, Geophysical Research Letters, 51, e2023GL107 377, 2024.

Chen, L., Zhong, X., Li, H., Wu, J., Lu, B., Chen, D., Xie, S.-P., Wu, L., Chao, Q., Lin, C., et al.: A machine learning model that outperforms conventional global subseasonal forecast models, Nature Communications, 15, 6425, 2024.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, International Journal of Climatology: a Journal of the Royal Meteorological Society, 28, 2031–2064, 2008.

Deng, Q., Lu, P., Zhao, S., and Yuan, N.: U-Net: A deep-learning method for improving summer precipitation forecasts in China, Atmospheric and Oceanic Science Letters, 16, 100 322, 2023.

Faijaroenmongkol, T., Sarinnapakorn, K., and Vateekul, P.: Sub-Seasonal Precipitation Bias-Correction in Thailand Using Attention U-Net With Seasonal and Meteorological Effects, IEEE Access, 11, 135 463–135 475, 2023.

Höhlein, K., Schulz, B., Westermann, R., and Lerch, S.: Postprocessing of ensemble weather forecasts using permutation-invariant neural networks, Artificial Intelligence for the Earth Systems, 3, e230 070, 2024.

Horat, N. and Lerch, S.: Deep Learning for Postprocessing Global Probabilistic Forecasts on Subseasonal Time Scales, Monthly Weather Review, 152, 667–687, 2024.

Hu, W., Ghazvinian, M., Chapman, W. E., Sengupta, A., Ralph, F. M., and Delle Monache, L.: Deep learning forecast uncertainty for precipitation over the Western United States, Monthly Weather Review, 151, 1367–1385, 2023.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, 2023.

Olivetti, L. and Messori, G.: Do data-driven models beat numerical models in forecasting weather extremes? A comparison of IFS HRES, Pangu-Weather and GraphCast, EGUsphere, 2024, 1–35, https://doi.org/10.5194/egusphere-2024-1042, 2024.

Rasp, S. and Lerch, S.: Neural networks for postprocessing ensemble weather forecasts, Monthly Weather Review, 146, 3885–3900, 2018.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P.: Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP, Geoscientific model development, 11, 3681–3712, 2018.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., pp. 234–241, Springer International Publishing, Cham, ISBN 978-3-319-24574-4, 2015.

Ryu, J., Wang, S.-Y., Jeong, J.-H., and Yoon, J.-H.: Sub-seasonal prediction skill: is the mean state a good model evaluation metric?, Climate Dynamics, pp. 1–16, 2024.

https://doi.org/10.5194/egusphere-2025-308
Preprint. Discussion started: 26 March 2025

Ryu, J., Kim, H., Wang, S.-Y. S., and Yoon, J.-H.: W-US-BCSR/IRnAF: 1.0 Initial Release (v1.0), https://doi.org/10.5281/zenodo.14776781, 2025.

Schulz, B. and Lerch, S.: Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison, Monthly Weather Review, 150, 235–257, 2022.

Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., Nico, P. S., Feldman, D. R., Jones, A. D., Collins, W. D., et al.: A low-to-no snow future and its impacts on water resources in the western United States, Nature Reviews Earth & Environment, 2, 800–819, 2021.

Son, R., Ma, P.-L., Wang, H., Rasch, P. J., Wang, S.-Y., Kim, H., Jeong, J.-H., Lim, K.-S. S., and Yoon, J.-H.: Deep learning provides substantial improvements to county-level fire weather forecasting over the western united states, Journal of Advances in Modeling Earth Systems, 14, e2022MS002 995, 2022.

Vitart, F. and Robertson, A. W.: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events, npj climate and atmospheric science, 1, 3, 2018.

Wang, L., Qian, Y., Leung, L. R., Chen, X., Sarangi, C., Lu, J., Song, F., Gao, Y., Lin, G., and Zhang, Y.: Multiple metrics informed projections of future precipitation in China, Geophysical Research Letters, 48, e2021GL093 810, 2021.

Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N.: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 502, 2021.

Woolnough, S., Vitart, F., Robertson, A. W., Coelho, C. A., Lee, R., Lin, H., Kumar, A., Stan, C., Balmaseda, M., Caltabiano, N., et al.: Celebrating 10 Years of the Subseasonal to Seasonal Prediction Project and Looking to the Future, Bulletin of the American Meteorological Society, 105, E521–E526, 2024.

Xin, F., Shen, Y., and Lu, C.: Application of a weighted ensemble forecasting method based on online learning in subseasonal forecast in the South China, Geoscience Letters, 11, 5, 2024.

Yang, B., Qian, Y., Lin, G., Leung, L. R., Rasch, P. J., Zhang, G. J., McFarlane, S. A., Zhao, C., Zhang, Y., Wang, H., et al.: Uncertainty quantification and parameter tuning in the CAM5 Zhang-McFarlane convection scheme and impact of improved convection on the global circulation and climate, Journal of Geophysical Research: Atmospheres, 118, 395–415, 2013.