JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS002995

Key Points:

- Combining dynamical model prediction and deep learning (DL) can boost predictability of wildfire risk up to 14 days
- DL efficiently reduces the internal bias of dynamical model and enhances feature depiction up to 4 km resolution
- Predictions, especially associated with extreme fire weather, are well grounded on the characteristics of wind behavior

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J.-H. Yoon,
yjinho@gist.ac.kr

Citation:

Son, R., Ma, P.-L., Wang, H., Rasch, P. J., (Simon) Wang, S.-Y., Kim, H., et al. (2022). Deep learning provides substantial improvements to county-level fire weather forecasting over the western United States. *Journal of Advances in Modeling Earth Systems*, *14*, e2022MS002995. https://doi.org/10.1029/2022MS002995

Received 14 JAN 2022
Accepted 19 SEP 2022

Author Contributions:

Conceptualization: Rackhun Son, Shih-Yu (Simon) Wang, Jin-Ho Yoon
Data curation: Rackhun Son
Formal analysis: Rackhun Son, Shih-Yu (Simon) Wang, Jin-Ho Yoon
Investigation: Rackhun Son, Shih-Yu (Simon) Wang, Hyungjun Kim, Jee-Hoon Jeong, Jin-Ho Yoon

Deep Learning Provides Substantial Improvements to County-Level Fire Weather Forecasting Over the Western United States

Rackhun Son[1,2], Po-Lun Ma[3], Hailong Wang[3], Philp J. Rasch[3,4], Shih-Yu (Simon) Wang[5], Hyungjun Kim[6,7,8], Jee-Hoon Jeong[9], Kyo-Sun Sunny Lim[10], and Jin-Ho Yoon[2]

[1]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany, [2]School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju, Korea, [3]Pacific Northwest National Laboratory, Richland, WA, USA, [4]Department of Atmospheric Science, University of Washington, Seattle, WA, USA, [5]Department of Plants, Soils, and Climate, Utah State University, Logan, UT, USA, [6]Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology, Daejeon, Korea, [7]Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea, [8]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan, [9]Faculty of Earth and Environmental Sciences, Chonnam National University, Gwangju, Korea, [10]Department of Atmospheric Sciences, Center for Atmospheric REmote sensing (CARE), Kyungpook National University, Daegu, Korea

Abstract The recent wildfires in the western United States during 2018 and 2020 caused record-breaking fire damage and casualties. Despite remarkable advances in fire modeling and weather forecasting, it remains challenging to anticipate catastrophic wildfire events and associated damage. One key missing component is a fire weather prediction system with sufficiently long lead time capable of providing useful regional details. Here, we develop a hybrid prediction model of wildfire danger called CFS with super resolution (CFS-SR) as a proof of concept to fill that void. The CFS-SR model is constructed by integrating the Climate Forecast System version 2 with a deep learning (DL) technique from Single Image Super Resolution, a method widely used in enhancing image resolution. We show that for the 2018–2019 fire season, the CFS-SR model significantly improves accuracy in forecasting fire weather at lead times of up to 7 days with an enhanced spatial resolution up to 4 km. This level of high resolution provides county-level fire weather forecast, making it more practical for allocating resources to mitigate wildfire danger. Our study demonstrates that a proper combination of ensemble climate predictions with DL techniques can boost predictability at finer spatial scales, increasing the utility of fire weather forecasts for practical applications.

Plain Language Summary Fire weather prediction is significantly improved by a hybrid approach-a weather forecast model and machine learning algorithm. We train a deep learning (DL) technique with model simulations to directly produce a fire hazard estimate based solely on predicted weather conditions. The predictability of extreme danger fire conditions is enhanced which can be useful in reducing fire pre-suppression activities. The applied DL technique also successfully produces a higher spatial resolution (4 km) product for county-level fire forecasting. This approach shows that DL can be integrated in weather/climate models to produce better quality.

1. Introduction

Forecasting wildfire danger is a challenging task due to its complexity involving climate system, interactions with vegetation and socio-economic components (Hantson et al., 2016). As result of the remarkable progress in numerical weather forecasting (Bauer et al., 2015), its application for fire forecasting has also advanced, and some government agencies are now providing fire weather forecasting services. For instance, the National Interagency Coordination Center (NICC, https://www.nifc.gov/nicc/predictive/outlooks/outlooks.htm) issues public service forecasts of fire hazard in the United States for up to 7 days (called the 7-Day Significant Fire Potential Outlook). These forecasts categorize the fire potential into nine levels, considering fuel dryness, fire weather, ignition triggers, and resource capability. The European Forest Fire Information System (EFFIS, https://effis.jrc.ec.europa.eu) or the Global Wildfire Information System (GWIS, https://gwis.jrc.ec.europa.eu/) also provide fire danger forecast using four different indices based on the Fire Weather Index (FWI) system.

Methodology: Rackhun Son, Po-Lun Ma, Hailong Wang, Philp J. Rasch, Hyungjun Kim
Software: Rackhun Son, Jee-Hoon Jeong
Validation: Rackhun Son, Jee-Hoon Jeong, Jin-Ho Yoon
Visualization: Rackhun Son
Writing – original draft: Rackhun Son
Writing – review & editing: Po-Lun Ma, Hailong Wang, Philp J. Rasch, Shih-Yu (Simon) Wang, Hyungjun Kim, Jee-Hoon Jeong, Jin-Ho Yoon

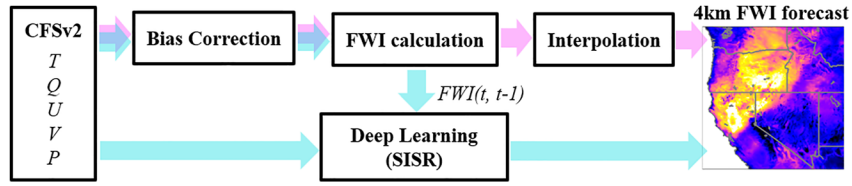

Figure 1. The schematic diagram for pipeline of CFS with super resolution (CFS-SR). For each 6-hourly (00Z, 06Z, 12Z, and 18Z) CFS forecasting, five weather components are used as input variables; temperature (*T*), specific humidity (*Q*), U-wind (*U*), and V-wind (*V*) at every 18Z UTC and daily averaged precipitation (*P*). Also, these five components are used to calculate Fire Weather Index (FWI) after correcting biases. As the light blue arrows indicate, FWI(*FWI(t)*) and the day prior to FWI (*FWI(t*-1)) is added on the input to train CFS-SR. The light red arrows indicate the interpolated FWI forecast from Climate Forecast System version 2 (CFSv2).

Many dynamical weather forecast models have been applied as a component of fire danger warning systems (Coen et al., 2020; Di Giuseppe et al., 2016; Mölders, 2010). This theory driven approach provides a rational basis for forecasts and offers a scientifically based inference process compared to previous empirical approaches. However, dynamical fire weather forecasting can require considerable computational resources, extra processing, and the coupling of regional numerical models to generate regional forecasts with reasonable spatial resolution (Roads et al., 2010). Yet, its utility is limited by multiple factors including accuracy of the underlying weather forecast model as well as accuracy of the additional model(s) used to forecast fire characteristics.

With a vast amount of data accumulated over decades, statistical approaches have also been widely used for efficient forecasting of various climate variables (Graff et al., 2020; Kouassi et al., 2020; Preisler & Westerling, 2007). Statistical methods are versatile in detecting internal features and anomalous patterns regardless of the data domain. Recently, machine learning (ML) methods have led to innovative achievements in various fields (Lecun et al., 2015) and are being actively studied in weather and climate domains (Ham et al., 2019; Kim et al., 2019), including wildfire prediction (Sayad et al., 2019; Song & Wang, 2020). Nevertheless, there is a concern that ML based modeling alone misses the understanding of the dynamical processes governing the occurrence of wildfire and its prediction, and this limitation motivates the present study that uses a hybrid approach combining physics-based modeling with ML (Pathak et al., 2018; Reichstein et al., 2019).

To exploit the physics-based modeling and ML approaches while maximizing predictability, we present a "proof-of-concept" hybrid prediction model, using a ML technique we call CFS with super resolution (CFS-SR), that blends the Climate Forecast System version 2 (CFSv2, see Methods) 6-hourly weather forecasts with a deep learning (DL) algorithm borrowed from the Single Image Super-Resolution (SISR) (see Methods) methodology, which has been widely applied to downscale climate data set (Cheng et al., 2021; Jiang et al., 2021; Vandal et al., 2018). CFS-SR is developed to produce weather driven wildfire danger forecasts at 4 km resolution, which matches the spatial scale of the Parameter Elevation Regression on Independent Slopes Model (PRISM) (Daly et al., 2008) data that estimates the FWI (see Methods) for the western United States. CFS-SR is expected not only to learn the nonlinear characteristics in the FWI calculation, but also to correct biases coming from the CFSv2 outputs, while enhancing the spatial resolution from the T126 Gaussian grid (approximately 90 km) to 4 km. This application aims not only to provide accurate fire weather forecasts at an individual county level over the western United States, but also to prove that the hybrid concept bridging physic-based model and ML can be an effective approach to boost predictability. Figure 1 summarizes the procedure (light blue arrows) for the application of the CFS-SR model.

2. Data and Methods

2.1. The Climate Forecast System Version 2

The CFSv2 (Saha et al., 2014) is a fully coupled forecast system model developed at the Environmental Modeling Center in the National Centers for Environmental Prediction. The CFSv2 offers a set of complete and consistent hindcasts over a 29-year period (1982–2010) along with operational forecasts from 2011 to the present. In this study, we use 6-hourly surface and radiative fluxes forecast outputs at T126 Gaussian resolution. The meteorological fields of the FWI system forecasted by CFSv2, including temperature (*T*), specific humidity (*Q*), horizontal wind components (*U*, *V*) at 18Z (UTC, the closest to noon local standard time on California) and daily averaged

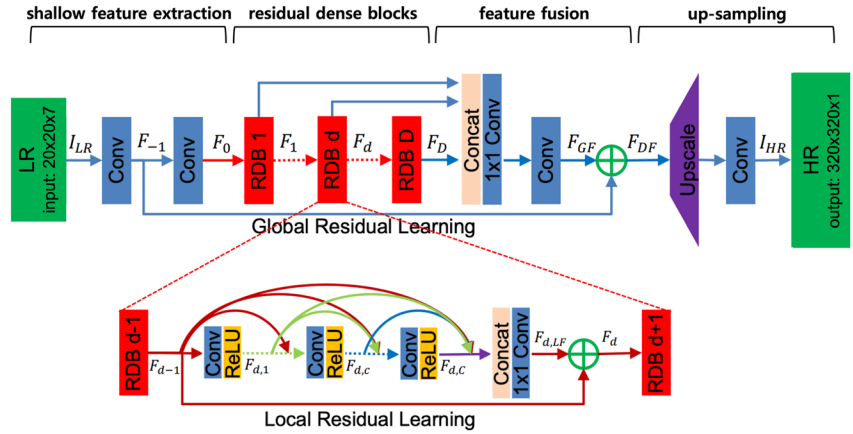

Figure 2. Model architecture diagram. The architecture of the Residual Dense Network (Figure 2 from Zhang et al., 2018) is on the top and that of the residual dense block (RDB, Figure 3 from Zhang et al., 2018) is on the bottom. Greenbox with low-resolution (LR) (high-resolution (HR)) indicates input (output) layer for 20 × 20 (320 × 320) spatial resolution with seven input variables: *T*, *Q*, *U*, *V*, *P*, and *FWI*(*t*-1, *t*) in Figure 1 (with only one output variable: Fire Weather Index). The blue box is for convolution layer, the red is for the RDB block, the apricot is for the concatenation layer, the purple is for the upscaling layer, and the yellow is for the non-linear activation function layer.

precipitation rate (*P*), as input variables for DL model. The FWI on the day of forecast and its preceding day are calculated using the CFSv2 outputs after bias correction to be fed to the DL model. The forecasts are initialized 4 times every day (00Z, 06Z, 12Z, and 18Z UTC), and each of them are evaluated for 9 months each year. With 6-hourly updating CFSv2 forecasts, four independent DL models are trained on a daily basis, and all of them are merged to produce a final forecast product. The forecasts are split into two periods, using 7 years (2011–2017) of the training period and the most recent 2 years (2018–2019) for the testing period.

2.2. Deep Learning Model and Training Strategy

In this study, we adopt DL architecture based on SISR. The SISR was developed as a challenging task in computer vision aiming to reconstruct a high-resolution (HR) image from a degraded low-resolution (LR) measurement. The SISR is an ill-posed inverse problem without a unique solution in that numerous HR images can be converted into the same LR image. The SISR has been used on various range of image processing tasks contributing to commercial, government and scientific applications (Malczewski & Stasiński, 2009). Climate research communities have also applied SISR to improve downscaling techniques (Cheng et al., 2021; Rodrigues et al., 2018; Sekiyama, 2020).

The first DL based SISR model, so called Super-Resolution Convolutional Neural Network, was introduced to provide significantly improved performance compared to earlier attempts to produce high resolution images (Dong et al., 2016). Starting from this, there has been remarkable progress in DL approaches with deeper and wider network structures. For instance, Enhanced Deep Residual Network integrated modified ResNet architecture (He et al., 2016) with scaling factor to stably expand the size of network (Lim et al., 2017). Deep Back-Projection Network adopted iterative an up and down projection procedure to correct prediction biases based on stacked upsampling results (Haris et al., 2018). Also, Information Distillation Network introduced distillation block to efficiently extract abundant features operating with simple bicubic interpolation (Hui et al., 2018).

At the part of DL algorithm in CFS-SR, we implement Residual Dense Network (RDN) (Zhang et al., 2018). The structure of RDN is divided into four functional parts: shallow feature extraction, residual dense blocks (RDBs), feature fusion and upsampling (Figure 2). The first two convolutional layers extract shallow features from LR inputs. The outputs are feed-forwarded into a set of RDBs to extract hierarchical features. Each RDB consists of convolutional layers with rectified linear units and the layers are densely connected by concatenation. At the last step of each block, proceeding features are combined through residual connection to improve local feature representation. All the outputs from RDBs are concatenated and adaptively fused with a 1 × 1 convolutional layer. Global feature fusion is then applied by a direct connection with the first layer. Finally, HR output is produced after upsampling spatial scale and adjusting depth through the final convolutional layer.

As the configuration of RDN, we stack 20 residual blocks (D), and each block consists of six convolutional layers (C) with 32 filters, which is named as growth rate in (Zhang et al., 2018). The first convolutional layer has seven filters to adopt input variables from CFSv2 (temperature, humidity, precipitation, U and V-wind, calculated FWI, and preceding-day FWI). The other convolutional layers of RDN are set to have 32 filters. Because the patch size of seven input variables is $20 \times 20 \times 7$ (33.5–51.5 N, 125.6–107.8 W) and the forecast target, which is PRISM-level FWI over the western United State, is $320 \times 320 \times 1$ (34.9–48.2 N, 125.0–111.7 W), the upsampling factor is set to 16. For each of the different forecasting lead time from 1 to 7, 14, and 21 days, we train an individual model with the same architecture, nine models in total are used in this study. Figure 2 summarizes the flow of CFS-SR operation and the details of model structure.

Considering that wildfire rarely occurs in the boreal winter in the area, only data sets from June to November are used in this study providing 1281 FWI images for training (2011–2017) and 366 for testing (2018–2019). For robust training five-fold cross validation is used and each model is trained with the Adaptive Moment Estimation 2 optimizer (Kingma & Ba, 2015) by setting the learning rate to 0.0002 with batch size to 16.

It is important to point out that our approach is not consistent with the typical concept of SISR. Instead of using a degraded PRISM-level FWI, we use coarse resolution forecasting products from CFSv2 as input. This is therefore not a classical SISR problem to reconstruct resolution, but rather a regression mapping task to generate high resolution output. However, a significant advantage of adopting the advanced SISR DL architecture, in comparison with previous approaches (Huang, 2020; Sun & Tang, 2020), is that interpolation is no longer required for the refining process to enhance resolution of an input data set.

2.3. Fire Weather Index

The FWI is a numerical rating for fire danger based on the Canadian Forest Service Fire Weather Index Rating System (Stocks et al., 1989). Six indices including FWI constitute the FWI system indicating fire intensity and behavior. They are calculated by using daily temperature, relative humidity, wind speed and 24-hr mean precipitation observed at mid-day local time. Among the six indices, Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and Drought Code (DC) are indicators of the moisture content in fuels with various range of time lag from hours to weeks. These codes characterize the relative mass of moisture in the different level of organic fuel layers: FFMC is for the top litter layer (less than 1 and 2 cm deep), DMC is for layers in moderate depth (5–10 cm deep) and DC is for 10–20 cm depth. The other three indices (Initial Spread Index (ISI), Build Up Index (BUI) and Fire Weather Index (FWI)) indicate fire behavior and intensity. ISI estimates the rate of fire spread by combining FFMC and wind speed, and BUI indicates the total amount of fuel available for combustion by combining DMC and DC. FWI merges ISI and BUI as a final product of the FWI system, estimating general information for potential fire danger. The values of FWI can be classified into six danger states (Table S1 in Supporting Information S1). Based on the systematic equations of FWI (van Wagner & Pickett, 1985), we use PRISM data set (Daly et al., 2008) to calculate high resolution FWI, except that wind components are from ERA5 (Hersbach et al., 2020) with a horizontal resolution of 0.25° (approximately 31 km) and interpolated into equivalent resolution of PRISM.

2.4. Bias Correction

When FWI is calculated with the CFSv2 outputs, the climatology of the Japanese 55-year Reanalysis (JRA55) (Kobayashi et al., 2015) is used to correct biases in the CFSv2. Based on the delta change approach (Maraun, 2016), the 6-hourly mean of the CFSv2 outputs, except for precipitation, at each lead time is replaced with the corresponding JRA55, using Equation 1:

$$\text{Corrected}_{h,l} = \text{CFS}v2_{h,l} - \overline{\text{CFV}v2_{h,l}} + \overline{\text{JRA55}_h}, \quad (1)$$

where, h is hour (00Z, 06Z, 12Z, and 18Z) and l is forecast lead time. For precipitation showing only positive values with distribution skewed to zero, a log transformation is executed before applying Equation 1, hence simplifying the computation by Equation 2:

$$\text{Corrected}_{h,l} = \text{CFS}v2_{h,l} \div \overline{\text{CFV}v2_{h,l}} \times \overline{\text{JRA55}_h}, \quad (2)$$

2.5. c-RMSE

The forecasting errors can be decomposed into biased and unbiased components (Taylor, 2001). The biased are mainly associated with systematic errors possibly caused by inaccuracy in initial/boundary conditions, errors in process parameterizations and misrepresentation of topography or coastline complexity. While unbiased error estimates can provide insights related to intrinsic predictive skill of a forecast model, the centered Root Mean Square Error (c-RMSE) is a metric to characterize errors associated with errors in weather features (patterns) by calculating the root mean square difference between two fields after removing the mean of each field. The c-RMSE is computed as Equation 3:

$$c-\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(f_i-\overline{f}\right)-\left(o_i-\overline{o}\right)\right]^2}, \tag{3}$$

where, f is forecasting output and o is PRISM observation as target in model training. To facilitate representation of multiple variables with different units and scales of variability when we evaluate model with c-RMSE, we normalize using the standard deviation of each field at each location.

2.6. Heidke Skill Score

For dichotomous categorical forecasts, there are only two possible outcomes (Yes or No) and all the possible cases can be summarized in a contingency table (Table S2 in Supporting Information S1). "A" is the number of the events forecasted to occur that correspond to the observation, called "hits." "B" is the number of the events forecasted to occur, but not in the observation, called "false alarms." "C" is the number of the events failed to forecast, called "misses." "D" is the number of events that didn't occur in the observation and correctly forecasted, called "correct rejections." The Heidke Skill Score (HSS) (Heidke, 1926) is one of the common verification measures using the contingency table, as expressed by Equation 4. It represents the fractional improvement by comparing difference with a randomly based/generated forecast, which is expected proportion to correct forecasts by chance. The improvement of model is normalized by the best possible improvement, enabling safely compare on different data sets. This can be simplified as Equation 5 based on the contingency table. A perfect score is 1 when a model correctly forecasts all of the "hits" and "correct rejections," while a score less than 0 means invalid forecasting capability.

$$\mathrm{HSS}=\frac{\text{model score}-\text{expected proportion}}{\text{perfect score}-\text{expected proportion}}, \tag{4}$$

$$\mathrm{HSS}=\frac{2\times(A\times D-B\times C)}{(A+C)\times(C+D)+(A+B)\times(B+D)}, \tag{5}$$

2.7. Sensitivity Test

An occlusion sensitivity test is one of the common methods used to evaluate ML image classification models (Zeiler & Fergus, 2014). Parts of an image are systematically occluded and its change in performance is spatially visualized to validate whether the target object can still be genuinely identified. In this study, we modified the method to quantify the sensitivity to each input variable, instead of spatial information of image, to identify its role in the FWI forecasting. The selected occluding variable is replaced by its daily climatological values, instead of zero (gray scale in RGB image), and the changes in c-RMSE are compared. The evaluation is repeated with each input variable individually occluded. In addition, we conduct the sensitivity tests in two ways, by applying the test: (a) for the entire test period and (b) only for days in the test period when FWI was observed higher than 50, defined as the extreme wildfire danger condition.

3. Results

3.1. Enhanced Fire Weather Forecast With Higher Resolution

Figures 3 and 4 display normalized c-RMSE (see Methods) and correlation score maps of FWI forecasts for the test period (2018–2019) over the western United States as a function of the forecasting lead time (a forecast

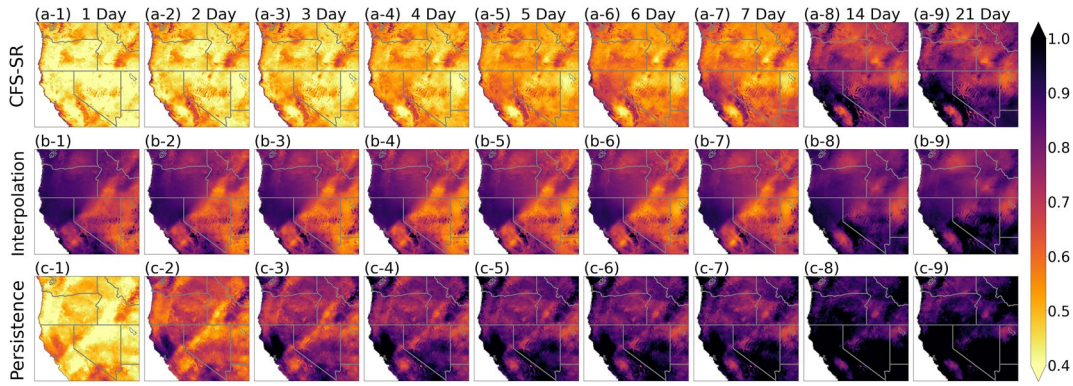

Figure 3. Normalized centered Root Mean Square Error (c-RMSE) map of forecasted Fire Weather Index (FWI) for the test data set. The c-RMSE is calculated between Prism FWI(target) and FWI predicted by (a) CFS with super resolution (CFS-SR), (b) the interpolated FWI from Climate Forecast System version 2 and (c) persistence for forecasting from lead time 1–21 day (1–9). The c-RMSE is normalized by standard deviation grid by grid during test data set period (01/06/2018–30/11/2019).

started at time *t-n* is evaluated at verification time *n*, where *n* varies between 1 and 21 days, for all times *t* in the test time period). The CFS-SR (Figures 3a and 4a) is compared to two baseline forecasting methods: (a) the bilinearly interpolated FWI forecast from CFSv2 (Figures 3b and 4b), which is described in Figure 1 with the light red arrows; and (b) a persistence forecast (Figures 3c and 4c), a measure commonly used in weather forecasting method that assumes that the predicted state is unchanged compared to its initial state. Compared to the directly interpolated results, CFS-SR has much reduced errors and higher correlations for the entire 7 days of lead time. The CFS-SR displays c-RMSE scores of less than 0.5 with higher than 0.9 correlation score over most of the domain for 1–3 days of lead time, and maintains less than 0.7 of the normalized c-RMSE with correlation coefficients higher than 0.7 in 5-day forecasts (Figure 3a1-5 and 4a1-5). However, the scores in CFS-SR significantly decline after 14 days, particularly in northern California (Figures 3a-8, 3a-9 and 4a-8, 4a-9), presenting apparent limits on subseasonal forecasting. The persistence forecast shows a fairly high scores in the 1-day forecast (Figures 3c-1 and 4c-1). However, it does not outperform CFS-SR, and its skill scores drop drastically and become the least skillful forecast after 3 days.

The improvement in the performance of CFS-SR is also evident at the state-level skill score comparisons. The c-RMSE (Figure 5) and correlation score (Figure 6) are calculated in the same ways with the previous results, but use spatially averaged FWI at five states in the targeting domain: California, Oregon, Idaho, Washington, and

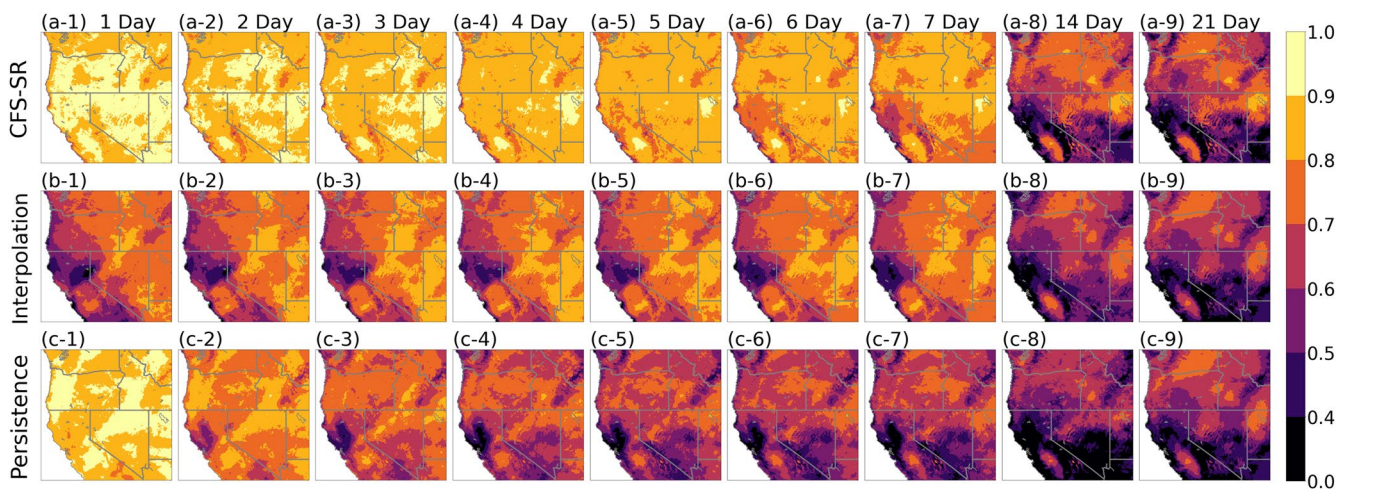

Figure 4. Correlation map of forecasted Fire Weather Index (FWI) for the test data set. Correlation between Prism FWI(target) and FWI predicted by (a) CFS with super resolution (CFS-SR), (b) the interpolated FWI from Climate Forecast System version 2 and (c) persistence forecasting from lead time 1–21 day (1–9). Correlation is calculated grid by grid during test data set period (01/06/2018–30/11/2019).

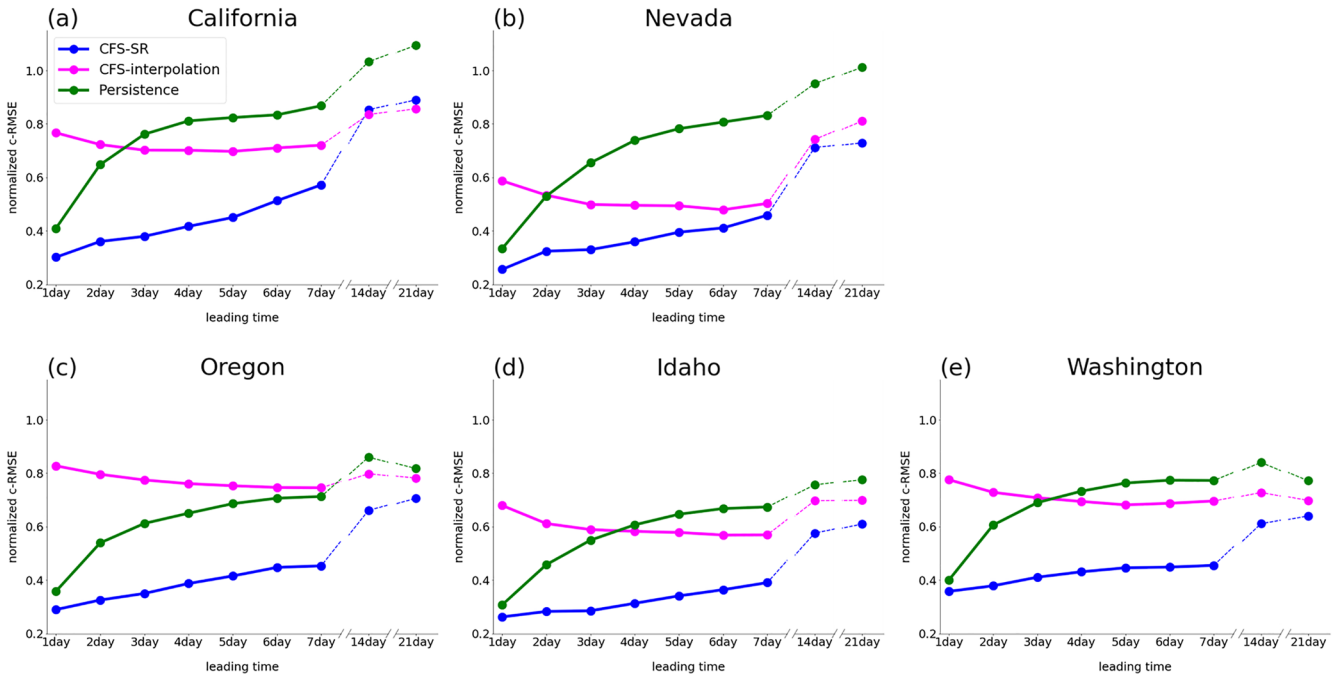

Figure 5. The normalized centered Root Mean Square Error (c-RMSE) comparison for states. The normalized c-RMSE between Prism Fire Weather Index (FWI) (target) and FWI predicted by CFS with super resolution (CFS-SR) (blue), the interpolated FWI from Climate Forecast System version 2 (magenta) and persistence forecasting (green) from lead time 1–21 day (1–9). The score is calculated during test period (01/06/2018–30/11/2019) for area averaged over (a) California, (b) Oregon, (c) Idaho, (d) Washington, and (e) Nevada.

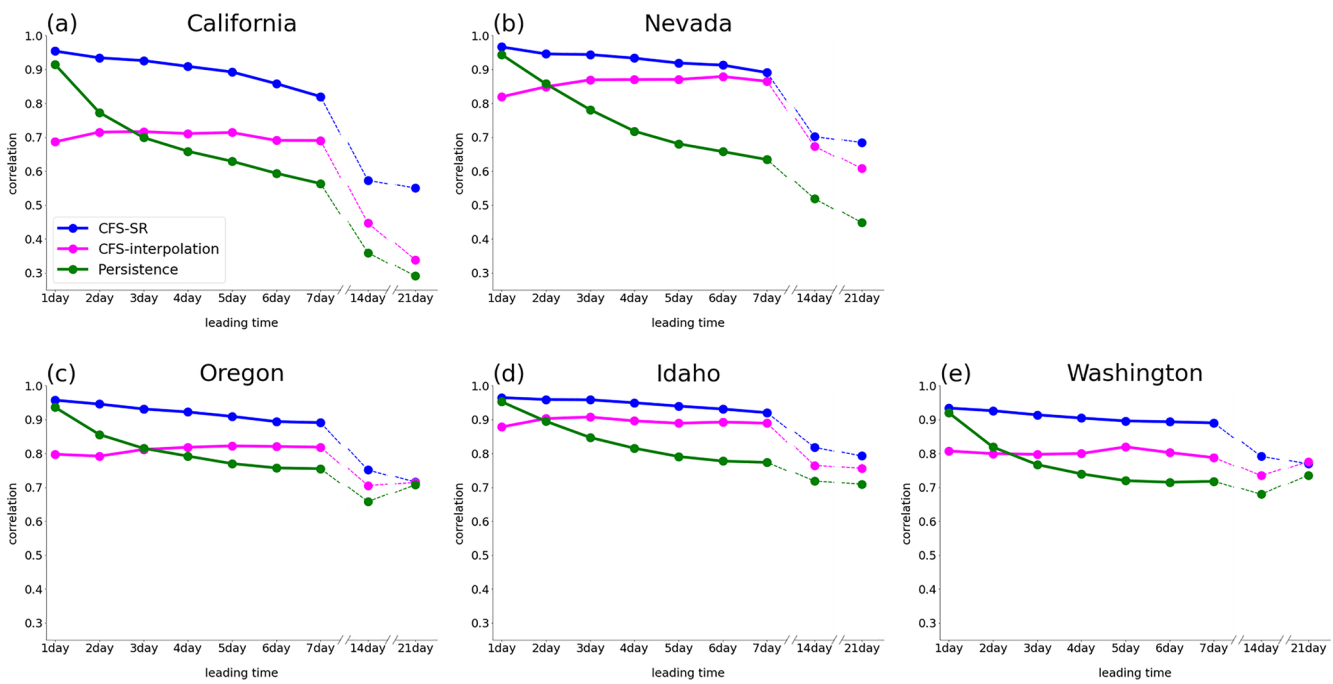

Figure 6. Correlation skill comparison in states. Correlation between Prism Fire Weather Index (FWI) (target) and FWI predicted by CFS with super resolution (CFS-SR) (blue), the interpolated FWI from Climate Forecast System version 2 (magenta) and persistence forecasting (green) from lead time 1–21 day (1–9). The score is calculated during test period (01/606/2018–30/11/2019) for area averaged over (a) California, (b) Oregon, (c) Idaho, (d) Washington, and (e) Nevada.

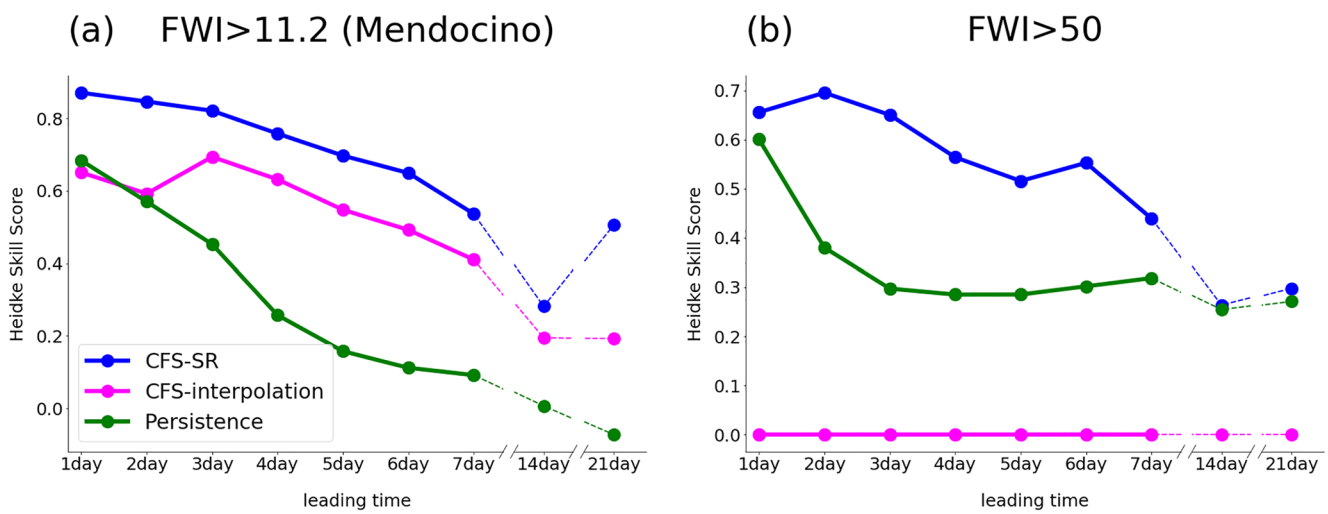

Figure 7. Forecasting skill comparison in Mendocino county. Heidke Skill Score (HSS) is compared for the dichotomous forecasting performance in (a) the moderate danger state (FWI > 11.2) and (b) the extreme danger state (FWI > 50.0).

Nevada. For each state, CFS-SR outperforms other baselines for the 1–7 days forecasting. The high-performance of CFS-SR becomes more apparent in California, showing the largest improvement in c-RMSE of 0.3 and the correlation of 0.2 in the 3 days of lead time (Figures 5a and 6a). Comparing to the interpolated FWI from CFSv2, the northern states (Oregon, Idaho, and Washington), also show significant improvements in the c-RMSE (Figures 5c–5e) compared to the other forecasts. Although the changes of the correlation are relatively small in the northern states (Figures 6c–6e), the absolute values of correlation are all generally above 0.9 until the 7 days of forecast.

In order to evaluate CFS-SR for extreme fire weather conditions, we investigate forecasting performance for Mendocino and Butte counties in California, where the most destructive fires occurred in the last 5 years. Figure 7 compares forecast performance in Mendocino county, where fire damage of historic proportion occurred in 2018 (the Mendocino Complex fire, July 27–18 September 2018). It is crucial to also skillfully forecast extremely high FWI to provide an alert for the most susceptible conditions of large wildfires. To depict moderate and extreme dangers of wildfire, FWI values of 11.2 and 50 are chosen respectively (Table S1 in Supporting Information S1) to be evaluated using the HSS metric. In terms of moderate danger of wildfire (Figure 7a), CFS-SR generally shows higher HSS in comparison to the other two forecasts (baselines). The performance gap between CFS-SR and the interpolated CFSv2 gradually becomes smaller with the increase in lead time, but the gap widens again after 2 weeks. The persistence forecast shows a fairly high HSS in the 1-day forecast, but it decreases drastically and even becomes sub-zero after day 14, meaning it does not hold any value as a forecast. The superiority of CFS-SR

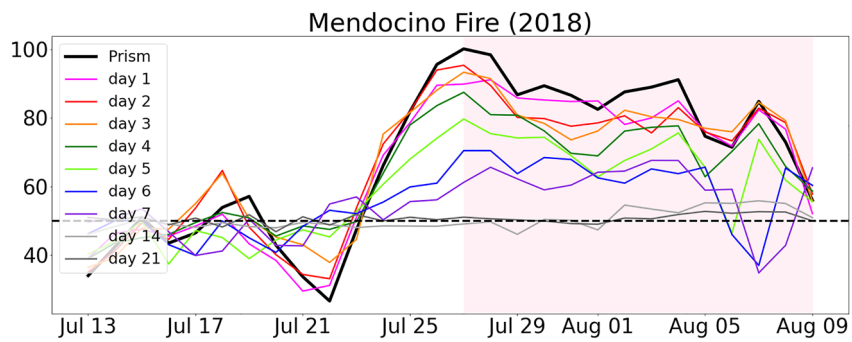

Figure 8. Time-series of forecasted Fire Weather Index (FWI). The time-series of Parameter Elevation Regression on Independent Slopes Model (PRISM) FWI and forecasted FWI from 1 to 21 lead time averaged over Mendocino county for the outbreak of Mendocino Complex Fire (light red shade). The values above the black dash line (FWI = 50) indicates an extreme danger state of wildfire.

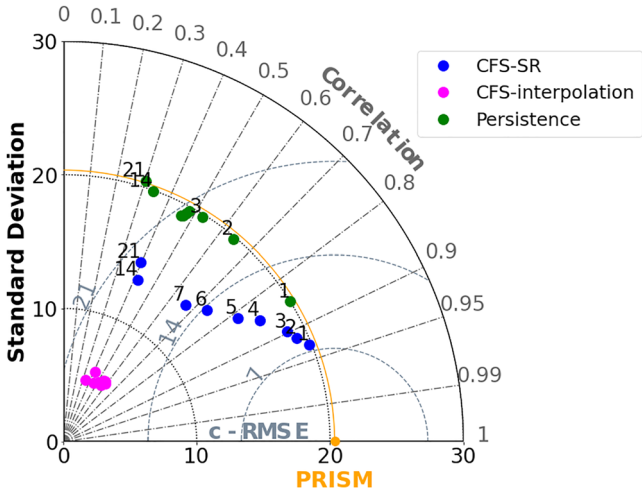

Figure 9. Talyor diagram of forecasting performance in Mendocino county. Three different forecasting results for Mendocino county, which are CFS with super resolution (CFS-SR) (blue), the interpolated Fire Weather Index (FWI) from Climate Forecast System version 2 (magenta) and persistence forecast (green), are compared for leading time 1–21 days (annotated with numbers). Prism (target) is marked by orange point.

is profound in forecasting the extreme danger of wildfire (Figure 7b), while the interpolated CFS completely fails to capture it during the test period. This implies that the underestimated FWI of CFSv2 is significantly improved by the hybridized ML procedure.

These results are also evident in forecasted FWI time-series (Figure 8) showing that abrupt changes in fire risk are accurately forecast for more days. For instance, the sudden rise of FWI before the occurrence of Mendocino complex fire on July 27th is successfully predicted by CFS-SR only for 5 days (Figures S1a–S1e in Supporting Information S1); however, it is still much better than the interpolated FWI from CFSv2. A Taylor diagram (Taylor, 2001) summarizing the performance comparison of CFS-SR indicates that c-RMSE becomes larger as the lead time increases, with a diminishing correlation and standard deviation (Figure 9).

In terms of the Camp Fire that started on November 8 in Butte County, which is another historical catastrophe in California, CFS-SR also provides accurate FWI forecasts up to 5 days earlier, although the value of FWI reached the highest degree a few days ahead of the fire outbreak on November 8th (Figures S2 and S3 in Supporting Information S1). The HSS and the Taylor diagram show almost similar results with the Mendocino, except that lower HSS is observed in CFS-SR than the interpolated CFSv2 after 14 days (Figure S4 in Supporting Information S1). Spatial maps of forecasted FWI for both events show consistent results with the county-averaged time series, with almost equivalent magnitude of FWI around the location where the fires occurred (Figure 10 and Figure S5 in Supporting Information S1).

3.2. Climate Variable Sensitivity

In order to explain the physical mechanism leading to the forecast improvement, we test the sensitivity of the CFS-SR forecasts by occluding each input variable and identify the most sensitive variable on each grid in Figure 11. For the testing period, the preceding-day FWI (*FWI*(*t*-1), yellow) appears to be the most critical variable for all lead times (Figures 11a–11g). For the first day of forecasting, these results can be expected because the preceding-day FWI is obtained from observations, while the other variables are from numerical model products.

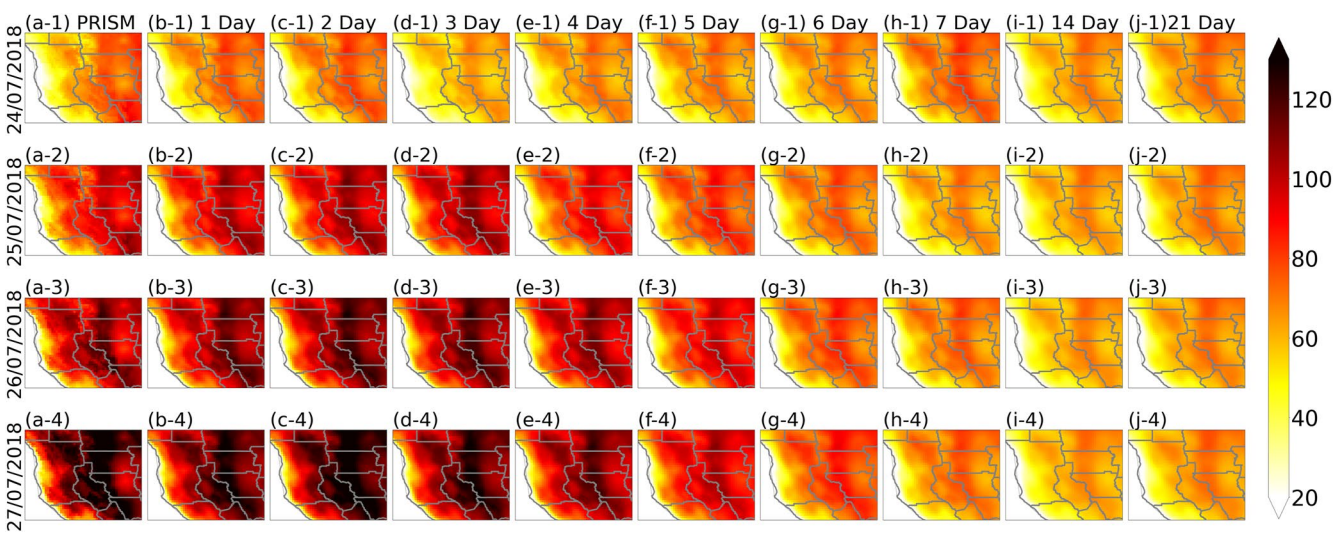

Figure 10. Fire Weather Index map before the Mendocino Complex Fire occurrence. (a) Parameter Elevation Regression on Independent Slopes Model, b–j. CFS with super resolution forecasting results leading time from 1 to 21 days. The mapping domain is set (124-121.8 W, 38.5–40.2 N) to include fire damage area; Mendocino, Lake, Colusa and Glenn counties.

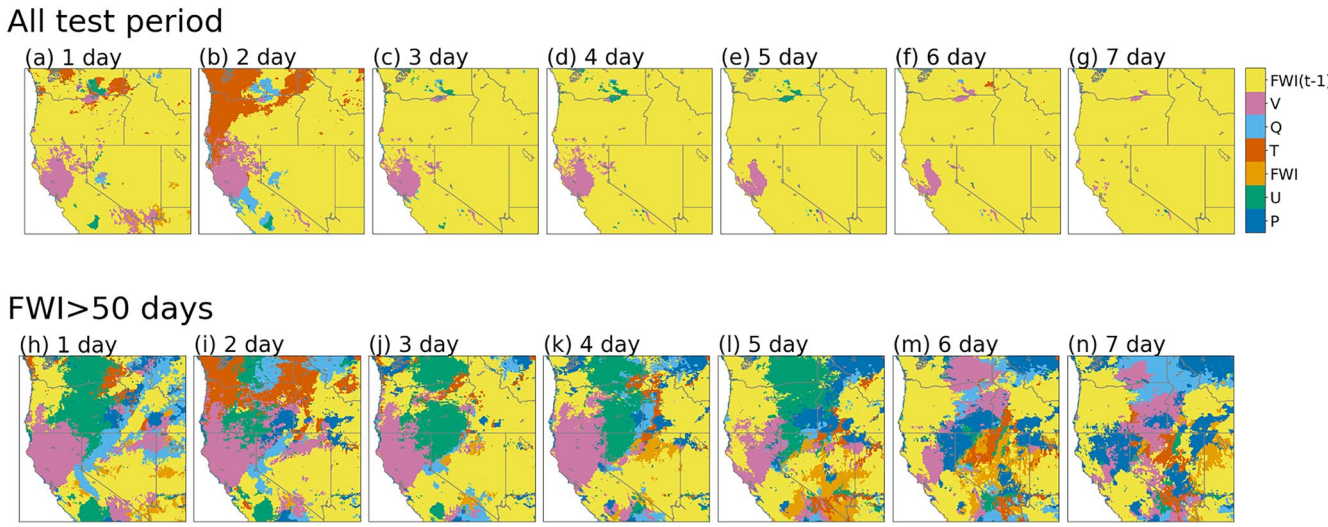

Figure 11. The spatial map for the most sensitive variables from occlusion sensitivity test. The most sensitive variables from the occlusion sensitivity test are visualized on each grid a–g. in all test period (01/06/2018–30/11/2019) and h–n. in days where Fire Weather Index recorded higher than 50. Each column is sorted from 1 to 7 day of leading time.

Interestingly, however, the sensitivity for the preceding-day FWI still remains as the most influential for other consecutive leading time, where the preceding-day FWI is also from the numerical model. Although some parts of California reveal higher sensitivity contributed from other variables, such as the meridional wind component (*V*) and temperature (*T*), the preceding-day FWI generally shows the greatest increase of c-RMSE for California (Figure 12). This can be interpreted in two ways: first, FWI has high persistence for a 1-day span (Figures 3c-1 and 4c-1) and second, the intrinsic numerical model errors tend to grow with lead time (Saha et al., 2006).

The forecasts are also sensitive to wind, and they are more sensitive for the extreme wildfire danger conditions in California where and when FWI is higher than 50 (Figures 11h–11l). The high sensitivity to the meridional wind component (*V*, pink) is mostly concentrated over northern California, where extreme wildfires occur frequently, and the horizontal wind component (*U*, green) appears more eastward in the southwestern part of Oregon, which echoes the September 2020 fire (Stuivenvolt Allen et al., 2021). In the extreme danger conditions, the sensitivity to the preceding-day FWI reduces while the sensitivity to the wind components intensifies, particularly in northern California (Figures S7a–S7c in Supporting Information S1), as is evident in Figure 13. Given the fact that the FWI calculation uses wind speed to influence the fire spread potential, it is understandable that CFS-SR identifies winds as the most influential factor for the extreme fire danger. Furthermore, the geographical distribution of the wind component makes sense, because the northeasterly wind across the Rock Mountains delivers dry air mass as it descends along the western slope of the mountains. The winds also tend to intensify through the valleys in

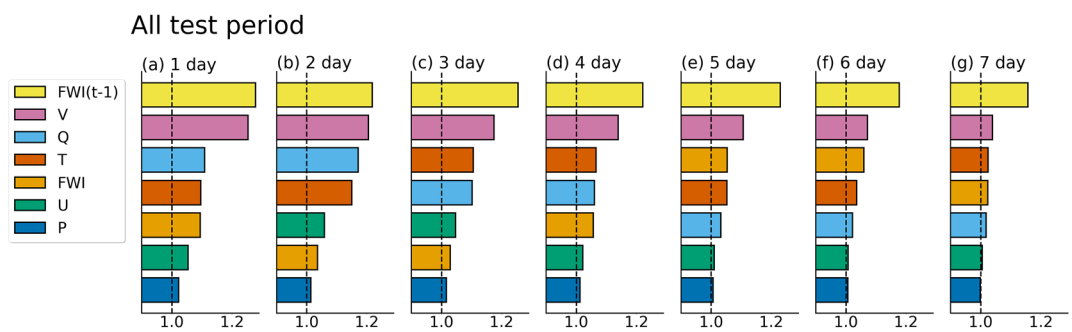

Figure 12. The results of occlusion sensitivity test in California for all test period. The rate of changes in forecasting errors are compared a–g. from first to seventh day of leading time by occluding each input variable in all test period (01/06/2018–30/11/2019). The black dash indicates the ratio value of 1.0 meaning that centered Root Mean Square Error (c-RMSE) is exactly equivalent to CFS with super resolution with all input information.

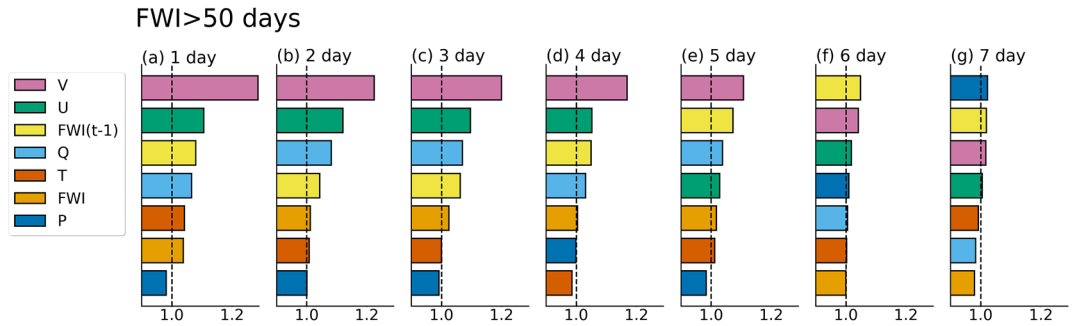

Figure 13. The results of occlusion sensitivity test in California for extreme wildfire risk days. The rate of changes in forecasting errors are compared a–g. from first to seventh day of leading time by occluding each input variable in days where Fire Weather Index (FWI) recorded higher than 50 during the test period (01/06/2018–30/11/2019). The black dash indicates the ratio value of 1.0 meaning that centered Root Mean Square Error is exactly equivalent to CFS with super resolution with all input information.

California (Gedalof, 2011), such as the so-called Diablo, sundowner or Santa Ana winds, which are associated with extreme wildfires in California (Brewer & Clements, 2020; Nauslar et al., 2018).

Temperature and humidity also have a noticeable influence in northern California, with a stronger temperature (humidity) effect in the north (south) (Figures S6d, S6e and S7d, S7e in Supporting Information S1). Precipitation rate, on the other hand, has the least influence among other variables (Figure S6g in Supporting Information S1), even though it plays an important role in the FWI system. In the extreme danger condition, its influence appears higher at sixth and seventh day of leading time, but its magnitude seems neglectable (Figure S7g in Supporting Information S1). The results suggest that CFS-SR is trained to extract relevant information from other weather variables associated with precipitation but not from precipitation itself.

4. Discussion

In this study, we have developed a hybrid prediction model combining the CFSv2 forecast and a DL algorithm based upon SISR, called CFS-SR, to forecast weather-driven wildfire danger over the western United States. The forecasted FWI shows not only higher accuracy over the entire fire season but also demonstrates a significant improvement in identifying the likelihood of large fire occurrences. Furthermore, once training of DL is completed, an enhanced resolution (4 km) was made possible without the need to conduct expensive computations to provide more local estimates of wildfire danger. These advantages are most apparent when consid nowcasting and near-term forecast. For instance, CFS-SR merged with a real-time operational dynamic model, successfully forecasts the significant increase in FWI from 1 to 5 days prior to the occurrence of the August complex fire over Mendocino county (Figure 14) and other neighbor damaged areas (Figure S8 in Supporting Information S1).

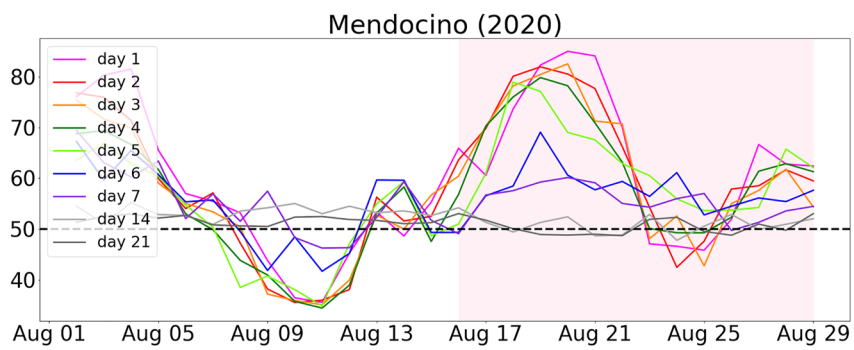

Figure 14. Time-series of forecasted Fire Weather Index (FWI). Time-series of FWI averaged over Mendocino county for outbreak of the August Complex Fire (16/08/2020~, light red shade). The values above the black dash line (FWI = 50) indicates an extreme danger state of wildfire.

Table 1
Comparison of Bias Correction and Fire Weather Index (FWI) Existence

	FWI(x)	FWI(o)
BC(x)	9.20	7.96
BC(o)	9.21	8.33

Note. BC denotes Bias Correction. Each performance is compared by c-RMSE for 3-day forecasting.

It is also interesting to examine whether the improved FWI prediction might be a result only of bias correction of the meteorological forecasts (Maraun, 2016). Table 1 shows the impact of FWI estimates using bias corrected and uncorrected CFSv2, separately, as input data set of CFS-SR (i.e., bias correction produced using the DL methodology). The results are not significantly different when the FWI products are excluded in the input data set, although the uncorrected CFSv2 outputs show slightly better performance (Table 1). When we provide FWI based on bias-corrected CFSv2 outputs as well as the raw CFSv2 outputs, the forecasting errors are significantly improved, and yet the best score is from the uncorrected CFSv2 outputs. This implies that some intrinsic properties of climate model biases useful for forecasting wildfire can be obtained from the uncorrected CFSv2 outputs. We conclude that CFS-SR still has some limitations in identifying all non-linear physical characteristics of FWI from meteorological fields alone.

With the explosive growth in DL applications, human-interpretable methods of DL models, a.k.a. explainable artificial intelligence (XAI), have shown remarkable development in recent years (Samek et al., 2021). Some of these DL applications, such as activation mapping (Zhou et al., 2016) and Layer Relevance Propagation (Bach et al., 2015), have been used in climate research and they provide a deeper level of scientific understanding from successfully trained DL models (Toms et al., 2020) that otherwise would not be revealed. To explore the limitations of DL associated with the complexity of multiple layers, we applied an occlusion sensitivity test (Zeiler & Fergus, 2014) to sort out the importance of input variables for the DL model in CFS-SR. It appears that the preceding-day FWI and the wind components play a more important role in forecasting FWI than other variables. Future work will focus on eliminating other deficiencies and exploring the non-linear relationship between the variables with respect to the climate driven wildfire danger.

In terms of the general application of our current method, 9 years of data availability has limited the number of fire events, hindering model training for extreme fire weather, and interannual and decadal variations in fire weather (Son et al., 2021; Westerling & Swetnam, 2003). However, more than 11,000 fire events annually were reported, and the fire damage was estimated to exceed 2.4 million Acres per year during the model training period in four states: California, Idaho, Nevada, and Oregon according to the National Interagency Fire Center (https://www.nifc.gov/fire-information/statistics). In comparison to the past 9 years from 2002 to 2010, the number of fire occurrence increased by 0.5%, and fire damage increased more than 30%. We also expect that meteorological input components indirectly reflect the climate driven long-term variabilities of fire weather. Nevertheless, a longer data period for model training is desirable, and longer training and testing periods should be used in future work.

In terms of spatial resolution, CFS-SR can be further improved to better satisfy needs of stakeholders and communities. For instance, the High-Resolution Rapid Refresh data product publicly provides 3 km horizontal resolution forecasting with hourly updates (Dowell et al., 2022) and the North American Mesoscale Forecast System is also accessible for 1.5–3 km resolution forecasting for 36–84 hr (Rogers et al., 2005). CFS-SR should be evaluated against these products, and they might be improved by application of similar techniques. CFS-SR is also able to produce forecasts for a much longer time scale than those products, although its accuracy beyond a week is low. Nevertheless, the results of CFS-SR in this study present "proof-of-concept" of the potential of the hybrid approach between physic-based dynamic model and DL. The spatial resolution and skill of the final target forecasts will be sensitive to the details of the dynamical model and DL choices. The rapid growth of ML/DL with advanced architectures and flexible cost constraints continuously provide new candidates in the hybrid approach. Thus, we expect that additional exploration in the hybrid approach could provide improved spatial resolution and the accuracy of subseasonal or even longer term forecasting compared to the current CFS-SR algorithm. It is noted here that stability of model performance and availability of rather longer period model output could be essential to achieve better performance of the hybrid approach presented here.

Table 2
Comparison Between Different Single Image Super-Resolution Algorithms

	RDN	IDN	EDSR	D-DBPN
c-RMSE	7.96	7.94	7.89	7.92
Number of parameters	1.0 M	1.4 M	1.8 M	20.3 M

Note. Each performance is compared by c-RMSE for 3-days forecasting. DBPN, Deep Back-Projection Network; EDSR, Enhanced Deep Residual Network; IDN, Information Distillation Network; RDN, Residual Dense Network.

We did explore variants of the SISR DL algorithm with current input models, and optimization choices and found they made only small differences in

prediction skill. Slightly improvements were achieved with more trainable parameters, which generally requires more computational resources and time spent in model training, but there is also a risk of overfitting (Table 2). While the cost of training the DL is not insignificant, its subsequent application is very inexpensive, and it only takes a few minutes to produce a timely high-resolution forecast. Furthermore, a striking advantage is its flexibility, which allows for an integration with any regional or global numerical model outputs leading to the possibility of using other DL algorithms, or the algorithm could be trained to produce other weather-related products tailored to the needs of other stakeholders or regions.

Data Availability Statement

All datasets used in this study are publicly available. CFSv2 (Saha et al., 2014) is available at https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00877. PRISM (Daly et al., 2008) is available at https://www.prism.oregonstate.edu/. ERA5 reanalysis (Hersbach et al., 2020) is available at https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset. The source code of RDN (Zhang et al., 2018) is available at https://github.com/yulunzhang/RDN. Also, the outputs of CFS-SR are openly available in Zenodo at https://doi.org/10.5281/zenodo.5652227.

Acknowledgments
This work was supported by the GIST Research Institute (GRI) grant funded by the GIST in 2022 and the Korean Meteorological Agency under the grant KMI2018-07010. HW and PJR acknowledge support from the U.S. Department of Energy (DOE) Office of Science Regional and Global Model Analysis (RGMA) Program area. The Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. P.-L.M. was supported by the "Enabling Aerosol-cloud interactions at GLobal convection-permitting scalES (EAGLES)" project (74358) funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Earth System Model Development program area. S.Y.W. is supported by the DOE RGMA program under Award Number DE-SC0016605. H.K. acknowledges Grant-in-Aid for Scientific Research (18KK0117) from Japan Society for the Promotion of Science (JSPS). Computational resource in this research was partially supported under the "HPC Support" Project, by the "Ministry of Science and ICT" and NIPA.

References

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*(7), e0130140. https://doi.org/10.1371/journal.pone.0130140

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Brewer, M. J., & Clements, C. B. (2020). The 2018 Camp Fire: Meteorological analysis using in situ observations and numerical simulations. *Atmosphere*, *11*(1), 47. https://doi.org/10.3390/atmos11010047

Cheng, J., Liu, J., Kuang, Q., Xu, Z., Shen, C., Liu, W., & Zhou, K. (2021). DeepDT: Generative adversarial network for high-resolution climate prediction. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. https://doi.org/10.1109/lgrs.2020.3041760

Coen, J. L., Schroeder, W., Conway, S., & Tarnay, L. (2020). Computational modeling of extreme wildland fire events: A synthesis of scientific understanding with applications to forecasting, land management, and firefighter safety. *Journal of Computational Science*, *45*, 101152. https://doi.org/10.1016/j.jocs.2020.101152

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Di Giuseppe, F., Pappenberger, F., Wetterhall, F., Krzeminski, B., Camia, A., Libertá, G., & Miguel, J. S. (2016). The potential predictability of fire danger provided by numerical weather prediction. *Journal of Applied Meteorology and Climatology*, *55*(11), 2469–2491. https://doi.org/10.1175/JAMC-D-15-0297.1

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., et al. (2022). The high-resolution rapid refresh (HRRR): An hourly updating convection-allowing forecast model. Part 1: Motivation and system description. *Weather and Forecasting*.

Gedalof, Z. (2011). Climate and spatial patterns of wildfire in North America. In *The landscape ecology of fire* (pp. 89–115). Springer.

Graff, C. A., Coffield, S. R., Chen, Y., Foufoula-Georgiou, E., Foufoula-Georgiou, E., Randerson, J. T., & Smyth, P. (2020). Forecasting daily wildfire activity using poisson regression. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(7), 4837–4851. https://doi.org/10.1109/TGRS.2020.2968029

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., et al. (2016). The status and challenge of global fire modelling. *Biogeosciences*, *13*(11), 3359–3375. https://doi.org/10.5194/bg-13-3359-2016

Haris, M., Shakhnarovich, G., & Ukita, N. (2018). Deep back-projection networks for super-resolution. *Proceedings—IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1664–1673.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).

Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler*, *8*(4), 301–349. https://doi.org/10.2307/519729

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Huang, X. (2020). Deep-learning based climate downscaling using the super-resolution method: A case study over the western US. *Geoscientific Model Development Discussions*, 1–18.

Hui, Z., Wang, X., & Gao, X. (2018). Fast and accurate single image super-resolution via information distillation network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 723–731.

Jiang, Y., Yang, K., Shao, C., Zhou, X., Zhao, L., Chen, Y., & Wu, H. (2021). A downscaling approach for constructing high-resolution precipitation dataset over the Tibetan Plateau from ERA5 reanalysis. *Atmospheric Research*, *256*, 105574. https://doi.org/10.1016/j.atmosres.2021.105574

Kim, S., Kim, H., Lee, J., Yoon, S., Kahou, S. E., Kashinath, K., & Prabhat, M. (2019). Deep-hurricane-tracker: Tracking and forecasting extreme climate events. *Proceedings—2019 IEEE Winter Conference on Applications of Computer Vision, WACV*, 1761–1769. https://doi.org/10.1109/WACV.2019.00192

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*. Paper presented at.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kouassi, J.-L., Wandan, N., & Mbow, C. (2020). Predictive modeling of wildfire occurrence and damage in a Tropical Savanna ecosystem of West Africa. *Fire*, *3*(3), 42. https://doi.org/10.3390/fire3030042

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1132–1140).

Malczewski, K., & Stasiński, R. (2009). Super resolution for multimedia, image, and video processing applications. In *Recent Advances in Multimedia Signal Processing and Communications* (pp. 171–208). Springer.

Maraun, D. (2016). Bias correcting climate change simulations—A critical review. *Current Climate Change Reports*, *2*(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Mölders, N. (2010). Comparison of Canadian forest fire danger rating system and national fire danger rating system fire indices derived from weather research and forecasting (WRF) model data for the June 2005 Interior Alaska wildfires. *Atmospheric Research*, *95*(2–3), 290–306. https://doi.org/10.1016/j.atmosres.2009.03.010

Nauslar, N. J., Abatzoglou, J. T., & Marsh, P. T. (2018). The 2017 North Bay and Southern California fires: A case study. *Fire*, *1*(1), 18. https://doi.org/10.3390/fire1010018

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., & Ott, E. (2018). Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos*, *28*(4), 041101. https://doi.org/10.1063/1.5028373

Preisler, H. K., & Westerling, A. L. (2007). Statistical model for forecasting monthly large wildfire events in Western United States. *Journal of Applied Meteorology and Climatology*, *46*(7), 1020–1030. https://doi.org/10.1175/JAM2513.1

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Roads, J., Tripp, P., Juang, H., Wang, J., Fujioka, F., & Chen, S. (2010). NCEP—ECPC monthly to seasonal US fire danger forecasts. *International Journal of Wildland Fire*, *19*(4), 399. https://doi.org/10.1071/WF07079

Rodrigues, E. R., Oliveira, I., Cunha, R., & Netto, M. (2018). DeepDownscale: A deep learning strategy for high-resolution weather forecast. In *2018 IEEE 14th International Conference on E-Science (e-Science)* (pp. 415–422).

Rogers, E., Lin, Y., Mitchell, K., Wu, W., Ferrier, B., Gayno, G., et al. (2005). The NCEP North American mesoscale modeling system: Final Eta model/analysis changes and preliminary experiments using the WRF-NMM. In *Preprints, 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction* (Vol. 4). American Meteor Society, CD-ROM B.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., et al. (2006). The NCEP climate forecast system. *Journal of Climate*, *19*(15), 3483–3517. https://doi.org/10.1175/jcli3812.1

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, *109*(3), 247–278. https://doi.org/10.1109/jproc.2021.3060483

Sayad, Y. O., Mousannif, H., & Al Moatassime, H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, *104*, 130–146. https://doi.org/10.1016/j.firesaf.2019.01.006

Sekiyama, T. T. (2020). Statistical downscaling of temperature distributions from the synoptic scale to the mesoscale using deep convolutional neural networks. ArXiv Preprint ArXiv:2007.10839.

Son, R., Wang, S. Y. S., Kim, S. H., Kim, H., Jeong, J.-H., & Yoon, J.-H. (2021). Recurrent pattern of extreme fire weather in California. *Environmental Research Letters*, *16*(9), 094031. https://doi.org/10.1088/1748-9326/ac1f44

Song, Y., & Wang, Y. (2020). Global wildfire outlook forecast with neural networks. *Remote Sensing*, *12*(14), 2246. https://doi.org/10.3390/rs12142246

Stocks, B. J., Lynham, T. J., Lawson, B. D., Alexander, M. E., Van Wagner, C. E., McAlpine, R. S., & Dubé, D. E. (1989). Canadian forest fire danger rating system: An overview. *The Forestry Chronicle*, *65*(4), 258–265. https://doi.org/10.5558/tfc65258-4

Stuivenvolt Allen, J., Simon Wang, S., LaPlante, M. D., & Yoon, J. (2021). Three western Pacific typhoons strengthened fire weather in the recent northwest US conflagration. *Geophysical Research Letters*, *48*(3), e2020GL091430. https://doi.org/10.1029/2020gl091430

Sun, A. Y., & Tang, G. (2020). Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets. *Frontiers in Water*, *56*. https://doi.org/10.3389/frwa.2020.536743

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002002. https://doi.org/10.1029/2019ms002002

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2018). *Generating high resolution climate change projections through single image super-resolution: An abridged version*. International Joint Conferences on Artificial Intelligence Organization.

van Wagner, C. E., & Pickett, T. L. (1985). In *Equations and FORTRAN program for the Canadian forest fire weather index system* (Vol. *33*).

Westerling, A. L., & Swetnam, T. W. (2003). Interannual to decadal drought and wildfire in the western United States. *EOS, Transactions American Geophysical Union*, *84*(49), 545–555. https://doi.org/10.1029/2003eo490001

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*(pp. 818–833).

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929.