Applied Computing and Geosciences 30 (2026) 100355

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

Convolutional neural networks for wildfire spread and intensity prediction

Maryam Moradpour *, Pankaj Kumar, Gholam Ali Hoshyaripour

Institute of Meteorology and Climate Research – Troposphere Research (IMK-TRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

ARTICLE INFO

Dataset link: https://www.frdr-dfdr.ca/repo/dataset/dd39c220-356d-4f06-a8e5-77016c648ca4, https://radar.kit.edu/radar/en/dataset/32n2dzbcds3y29qx?token=YcWXazVwcukmZJRXpLuF, https://radar.kit.edu/radar/en/dataset/nkhysv17ts8hw7r2?token=kDJOygZMKoxjMDhjEaSz

Keywords:
Wildfire dynamics
Convolutional neural network (CNN)
Deep learning
Fire behavior modeling
Atmospheric simulation

ABSTRACT

Wildfires significantly impact ecosystems, human health, infrastructure, and the climate, making accurate prediction of fire behavior and its effects critical. Traditional physics-based models simulate fire-atmosphere interactions in detail but are computationally expensive and limited in real-time applications. In addition, uncertainties in input parameters and simplified combustion representations can reduce their reliability in forecasting wildfire-driven emissions and plume dynamics. On the other hand, empirical and statistical models are computationally efficient but often lack the ability to capture the nonlinear and coupled processes that drive wildfire spread.

This study presents a deep learning approach using a convolutional neural network (CNN) to predict wildfire dynamics under varying environmental conditions of wind, fuel, and atmospheric stability. The model is trained on a high resolution Weather Research and Forecasting (WRF) model coupled with the SFIRE (WRF-SFIRE) simulation dataset and predicts the temporal evolution of wildfire spread, represented through ground-level heat flux (GHF) fields as an indicator for fire intensity and progression. Model performance is evaluated using root mean square error ($RMSE$) of 14.3 kW/m^2, mean absolute error (MAE) of 6.6 kW/m^2, correlation coefficient (R^2) of 84%, and the Structure–Amplitude–Location (SAL) method for spatial verification.

Results show that the CNN effectively reproduces the spatial and temporal evolution of wildfire dynamics, closely aligning with reference simulations across diverse conditions. By accurately capturing fire spread and intensity patterns at much lower computational cost, the proposed approach demonstrates the potential of deep learning to complement existing fire modeling frameworks and to support faster, scalable forecasting of wildfire behavior.

1. Introduction

Wildfires are a significant global concern, causing substantial damage to ecosystems, human lives, and property, while also contributing to climate change (Jain, 2022; Peterson et al., 2021). Accurate modeling of wildfire dynamics is crucial for effective fire management, risk assessment, and mitigation strategies. These models help land managers make decisions about preventing and fighting fires (Ackerman et al., 2000; Reddington et al., 2015).

Traditionally, simulation of wildfires has relied significantly on physics-based models that aim to capture the complex interactions between fire behavior and atmospheric processes. These models couple atmospheric and wildland fire models, simulating fire behavior by solving complex equations that govern processes such as fluid dynamics, combustion, and atmospheric interactions. The interaction between atmospheric processes and wildfires is complex and bidirectional. Weather conditions, such as temperature, wind, and humidity, strongly influence fire behavior and determine fire ignition, spread, and intensity. On the other hand, wildfires release large amounts of particulates and gases into the atmosphere, which impact air quality and regional climate (Liu et al., 2016; Filonchyk et al., 2022). Smoke plumes also affect atmospheric radiation and cloud formation. Therefore, accurately modeling these interactions is crucial (Coen and Riggan, 2014; Peace et al., 2015, 2022; Paugam et al., 2016). Despite the capability of coupled fire-atmosphere models, they are often computationally expensive and time-consuming, limiting their use in real-time forecasting for large-scale fire events. To reduce computational demands, many models employ subgrid-scale parameterizations and simplify combustion-related processes, but these simplifications can affect accuracy. Model performance is further constrained by uncertainties in fire input parameters and by the limited availability of fine-scale, near-surface meteorological data, both of which can reduce the reliability of wildfire predictions. Coen et al. (2022), Carter et al. (2019), Wiedinmyer et al. (2023). In parallel, empirical and statistical models have been widely used due to their computational efficiency.

* Corresponding author.
E-mail address: maryam.moradpour@kit.edu (M. Moradpour).

https://doi.org/10.1016/j.acags.2026.100355
Received 26 October 2025; Received in revised form 20 May 2026; Accepted 21 May 2026
Available online 1 June 2026

These approaches rely on simplified relationships, such as regression-based fire spread rates, fuel-based risk indices, or empirical plume-rise parameterizations (Xi et al., 2019; Sullivan, 2009). While they provide rapid estimates for large-scale or operational applications, they often cannot capture nonlinear feedbacks between fire, atmosphere, and fuels, limiting their predictive skill under extreme or rapidly evolving conditions (Sullivan, 2007; Taylor et al., 2013; Hernandez et al., 2015).

More recently, machine learning (ML) and deep learning (DL) have emerged as promising alternatives. By learning directly from large-scale simulations and observational datasets, ML and DL approaches can approximate complex nonlinear relationships at a fraction of the computational cost (Di Giuseppe et al., 2025; Kumar et al., 2024; Li et al., 2023; Sengupta and Woodford, 2025). Methods such as decision trees, regression models, and neural networks have been applied to wildfire occurrence prediction, risk mapping, and hotspot detection (Hodges and Lattimer, 2019; Aldersley et al., 2011; Burge et al., 2020; Zhang et al., 2023). However, most of these applications focus on static fire attributes (e.g., burned area or ignition probability), whereas the dynamic evolution of fire spread remains underexplored (Prapas et al., 2021; Samek et al., 2021; Son et al., 2022; Mambile et al., 2024; Küçükarslan, 2024; Jain et al., 2020) .

In atmospheric modeling, wildfire sources are often treated as static sources, meaning that fires are represented by fixed sources in space and time, often based on daily or monthly averages (Coen and Schroeder, 2013; Kaiser et al., 2012; Andela et al., 2015). This simplification neglects the strong variability of fire intensity throughout the day and the dynamic growth of the fire front, both of which are critical for accurately predicting plume rise and smoke dispersion. Walter et al. (2016) demonstrated that including a diurnal cycle of fire size and intensity, together with a plume rise parameterization, improved the agreement between simulated and observed aerosol optical depth (AOD). Similarly, Bruckert et al. (2022) showed in the context of volcanic eruptions that explicitly resolving the temporal evolution of emission phases and plume dynamics substantially improves the prediction of mass loading and dispersion. These findings highlight the importance of treating wildfire sources as dynamic. However, while machine learning and deep learning have been increasingly applied to wildfire prediction, their potential for representing the dynamic evolution of wildfires in space and time remains underexplored, which motivates the present study.

The aim of this study is to develop a data-driven model that can predict the dynamics of wildfire spread. Ground heat flux (GHF) from WRF-SFIRE simulations is used as an indicator of wildfire intensity and spread, enabling a convolutional neural network (CNN) model to capture the temporal and spatial variability. Specifically, the goal is to explore the potential of deep learning to learn the complex relationship between meteorological conditions, fuel characteristics, and wildfire behavior as simulated by a high-resolution physics-based coupled fire-atmosphere model. This paper is organized as follows. Section 2 describes the dataset, preprocessing steps, and the architecture and training of CNN model, as well as the methods used for model evaluation. Section 3 presents the results, including statistical performance metrics and spatial verification using the Structure–Amplitude–Location (SAL) method. Section 4 discusses the findings in the context of wildfire modeling and highlights limitations and directions for future research.

2. Methodology

2.1. Data description

In this study, a subset of the synthetic wildfire plume dataset developed by Moisseeva and Stull (2020) was used, consisting of 147 NetCDF files generated through numerical simulations using the WRF-SFIRE coupled modeling system. WRF-SFIRE combines the Weather Research and Forecasting (WRF) model with the SFIRE module, a semi-empirical algorithm for simulating wildland fire spread and its interactions with atmospheric processes. This dataset was designed to provide a controlled yet physically realistic framework for studying fire-atmosphere coupling, smoke dispersion, and plume dynamics (Moisseeva and Stull, 2021). Each NetCDF file corresponds to a unique wildfire scenario, characterized by variations in ambient wind speed (W), fuel type (F), atmospheric stability (R), and fireline length (L). Ambient wind speeds range from 3 to 12 m/s, covering a range of fire regimes from low-intensity to wind-driven fires. Atmospheric stability is represented through nine vertical profiles of potential temperature, which affect plume buoyancy and vertical transport. These profiles result in simulation domain heights ranging from approximately 3000 to 5000 m, depending on the stability structure required to support appropriate fire spread behavior.

Fuel types follow the Anderson category classification system (Anderson, 1982), which represents different vegetation characteristics relevant to fire modeling. The fireline lengths in the dataset include values of 1 km, 2 km and 4 km; however, only one simulation corresponds to a fireline length of 1 km and another to 4 km. Most simulations (105 out of 147) are configured with a fireline length of 2 km and a domain height of 3000 m. In order to ensure dataset homogeneity and maintain comparability across scenarios, only these 105 simulations were selected and analyzed in the present study. While this selection reduces variability and facilitates model training, it also constrains the diversity of fire configurations represented in the dataset. Consequently, the trained model is primarily applicable to wildfire scenarios with similar ignition extents and domain height, and its direct generalization to cases with different fireline lengths or vertical domain configurations may be limited.

Each dataset consists of a three-dimensional simulation domain that includes a range of meteorological and fire behavior variables. The meteorological fields comprise the three components of wind velocity (u, v, w) in m/s, temperature (T) in K, atmospheric pressure (P) in Pa, and water vapor mixing ratio (Q_{vapor}) in kg/kg. In addition to atmospheric variables, the dataset includes fire behavior parameters, most notably the ground-level heat flux from the fire (GHF) expressed in W/m^2, which serves as a key indicator of fire intensity and spread.

The full simulation domain is discretized into a grid of $250 \times 500 \times 50$ cells, corresponding to a physical extent of 10 km in the south-north direction, 20 km in the west-east direction, and 3 km in the vertical direction. Due to the high spatial resolution and large volume of data associated with each parameter, direct use of the entire domain in modeling is impractical. Therefore, in order to reduce computational cost while retaining the most relevant physical characteristics, this study focuses exclusively on a smaller subdomain that encompasses the active fire region. The selected subdomain includes the first 15 vertical levels, corresponding to the near-surface layers where most of the fire-atmosphere interactions occur. This spatial subset effectively captures the evolution of the fire plume and the associated meteorological responses, while significantly reducing the dimensionality of the data for subsequent analysis and model training.

2.2. Data pre-processing

In this study, meteorological and fire-related parameters at time step t are used as input features to predict the ground-level heat flux (GHF) at the subsequent time step $(t + 1)$. The proposed framework is therefore designed as a single-step prediction model rather than an autoregressive forecasting system. Specifically, the model predicts GHF at time step $(t + 1)$ using atmospheric, fuel-related, and fire-related conditions at time step t , while the predicted output is not recursively fed back into the model as input for future predictions. Among the input variables, the ground heat flux field at time step t is explicitly included as an input channel to provide the model with information about the current fire intensity and spatial heat distribution, thereby improving the prediction of the subsequent heat flux evolution.

Table 1
Fuel load mapping based on Anderson fuel classes.

Fuel Class	1	2	3	4	5	6	7	8	9	10	11	12	13
Fuel Load (kg/m^2)	0.166	0.896	0.674	3.591	0.784	1.344	1.091	1.120	0.780	2.692	2.582	7.749	13.024

Table 2
Stability profile mapping based on profile slope.

Stability Class	0	1	2	3	4	5	6	7	8
Profile Slope (m/K)	124.56	50.72	50.83	70.00	33.56	579.84	582.37	229.89	323.96

The input variables consist of six three-dimensional atmospheric fields: wind components (u, v, w), air temperature (T), atmospheric pressure (P), and water vapor mixing ratio (Q_{vapor}), each available at 15 vertical levels, along with three two-dimensional variables: fuel type, atmospheric stability profile, and ground heat flux at time step t. In order to incorporate the three-dimensional variables into the two-dimensional convolutional neural network (CNN), the vertical dimension is mapped onto the channel dimension rather than treated as an additional spatial axis. Specifically, each vertical level of each variable is considered an independent feature channel. As a result, the six three-dimensional variables contribute $6 \times 15 = 90$ channels. The three two-dimensional variables each contribute one channel, resulting in 3 additional channels. In total, the input to the CNN consists of 93 channels (90 from three-dimensional variables and 3 from two-dimensional variables), arranged as a tensor of size $(\text{south\_north}, \text{west\_east}, \text{channels})$. This representation preserves the horizontal spatial structure while enabling the model to implicitly capture vertical dependencies through inter-channel relationships.

As previously described, the fuel type is based on the Anderson category fuel model (Anderson, 1982), which characterizes different vegetation or fuel types commonly used in modeling fire behavior. Each category is associated with several physical properties, including fuel load, fuel bed depth, and fuel moisture content. Here, the fuel type is quantitatively described using fuel load, serving as a representative parameter aligned with Anderson fuel classification, shown in Table 1.

To numerically represent atmospheric stability, each potential temperature profile was linearly interpolated and the slope of the resulting line was used as a scalar characteristic, as can be observed in Table 2. This slope provides a simplified but meaningful measure of vertical stability, capturing the degree of atmospheric stratification relevant to plume rise and fire-induced convection.

Since the feature set includes variables with inherently different units and value ranges, direct use of raw data can lead to biased learning, where variables with larger magnitudes dominate the model training process (Zou et al., 2025; Lakshmanaswamy et al., 2024). To mitigate this issue and ensure balanced feature representation, Z-score standardization was applied to each input feature independently. This transformation rescales each variable to have zero mean and unit variance, allowing the model to treat all features on a comparable scale (Singh et al., 2025; Elshewey et al., 2025). The standardization process is defined as follows:

$$Z_i = \frac{x_i - \text{mean}_i}{\sigma_i} \tag{1}$$

Where Z_i is the standardized value, x_i is the original value of the variable, μ denotes the mean, and σ_i represents the standard deviation of the corresponding characteristic. This normalization step ensures numerical stability and improves the convergence behavior of the machine learning algorithm, while preserving the relative distribution of the input data.

2.3. Convolutional neural network modeling

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms specifically designed to handle data with grid-like topology (Khan et al., 2020). Their ability to extract spatial features through localized convolutional operations makes them well suited for analyzing the structured output of numerical weather-fire models, such as WRF-SFIRE. In recent years, CNNs have been increasingly applied to wildfire research, as they have shown good capability in tasks such as fire spread prediction, smoke dispersion, and fire risk assessment (Eddin et al., 2023; Zhang et al., 2021, 2022; Malloy, 2021; Allaire et al., 2021; Santopaolo et al., 2021). In this study, a CNN was also implemented to predict the ground heat flux (GHF) at the next time step from a multi-channel input consisting of meteorological and fire behavior variables. In this framework, the input to the CNN consists of multi-channel grids representing various meteorological and fire-related parameters, including wind components, temperature, pressure, water vapor mixing ratio, fuel load, atmospheric stability profile slope, and ground heat flux at the current time step. Each variable is treated as a separate input channel, allowing the model to jointly learn spatial patterns and cross-variable interactions.

The architecture used in this research is based on the work of Marjani and Mesgari (2023), who demonstrated the efficacy of multi-kernel convolutional blocks in capturing complex spatial patterns across varying scales. As shown in Fig. 1, the network consists of five sequential ConcatConvBlocks, each designed to extract features using convolutional filters of three different sizes: 3×3, 5×5, and 7×7. These kernels capture spatial information from the fine scale to the large scale, and their outputs are concatenated along the channel dimension. This structure enhances the model capacity to learn from multi-resolution features simultaneously. Each ConcatConvBlock is followed by a max-pooling operation with a kernel size of 2 and stride of 2, which progressively reduces the spatial dimensions of the feature maps while retaining the most important information. The first convolutional block receives 93 input channels, corresponding to the number of feature variables in the dataset. As the model progresses, the number of channels increases to enhance its representational power, following the sequence of 16, 32, 64, 128, and 256 filters in successive blocks as presented in Table 3. After the fifth block, the total number of concatenated feature channels reaches 768.

The output of the final convolutional block is flattened and passed to a fully connected (dense) network. This dense sub-network consists of three layers:

- A first layer with 128 neurons,
- A second layer with 256 neurons,
- A final output layer with 125,000 neurons (corresponding to a flattened 250×500 GHF field).

ReLU activation functions are used throughout the dense layers, and dropout layers with a rate of 0.3 are applied after the first two dense layers to prevent overfitting. The output of the final layer is reshaped into a 2D array with dimensions of 250 (south–north) by 500 (west–east), representing the predicted GHF distribution in the next time step.

In order to train the proposed CNN model, the dataset was divided into training, validation, and test subsets. After constructing the lagged input-target pairs $(t \rightarrow t+1)$, all samples from the wildfire simulations were combined into a single dataset and reshaped into independent samples for model training. The dataset was then randomly divided into

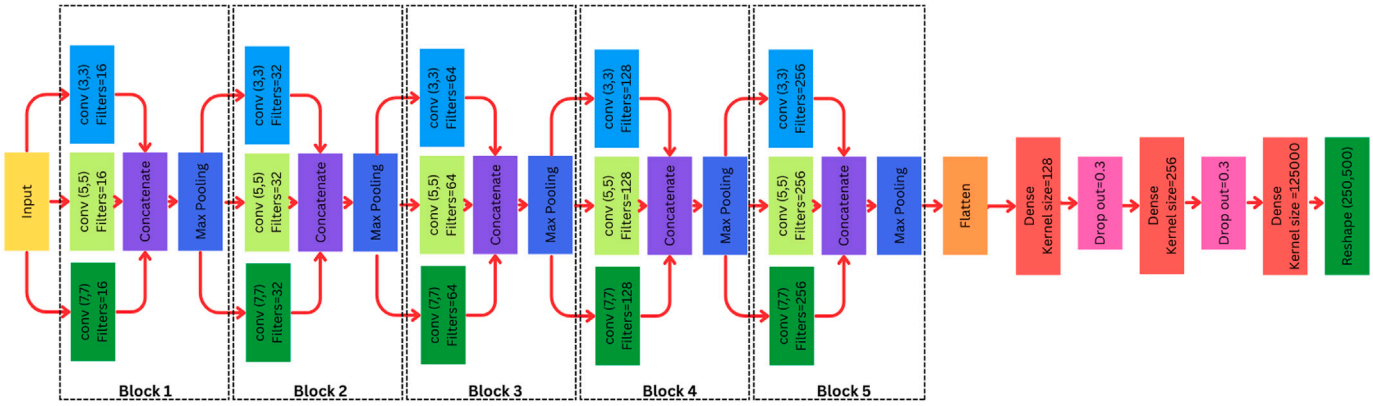

Fig. 1. CNN architecture used in this study.

Table 3
Structure of the multi-kernel CNN model.

Block	Layer type	Kernel size	Filters
1	Convolution	(3,3)	16
	Convolution	(5,5)	16
	Convolution	(7,7)	16
	Concatenation	–	–
	Max pooling	–	–
2	Convolution	(3,3)	32
	Convolution	(5,5)	32
	Convolution	(7,7)	32
	Concatenation	–	–
	Max pooling	–	–
3	Convolution	(3,3)	64
	Convolution	(5,5)	64
	Convolution	(7,7)	64
	Concatenation	–	–
	Max pooling	–	–
4	Convolution	(3,3)	128
	Convolution	(5,5)	128
	Convolution	(7,7)	128
	Concatenation	–	–
	Max pooling	–	–
5	Convolution	(3,3)	256
	Convolution	(5,5)	256
	Convolution	(7,7)	256
	Concatenation	–	–
	Max pooling	–	–
–	Flatten	–	–
–	Dense	128	–
–	Dropout	–	–
–	Dense	256	–
–	Dropout	–	–
–	Dense	125 000	–
–	Reshape	(250,500)	–

training (70%), validation (15%), and test (15%) subsets at the time-step level. As a result, time steps from the same wildfire simulation may appear in different subsets. Although this approach increases the number and diversity of training samples, the strong temporal correlation between consecutive wildfire states may introduce some temporal leakage between the training and evaluation datasets. Therefore, the reported model performance should be interpreted with this limitation in mind. Training was guided by the Mean Squared Error (MSE) loss function, which measures the average squared difference between predicted and actual values. The Adam optimizer was used with a learning rate and a weight decay to improve convergence and prevent overfitting. The validation dataset was used to monitor generalization and determine the best stopping point. The final model was selected based on the checkpoint with the lowest validation loss.

The hyperparameter tuning process was carried out to improve model performance. Several training runs were performed with different settings to identify effective combinations of hyperparameters such as the learning rate, batch size, dropout rate, number of filters, and activation functions. These hyperparameters were adjusted on the basis of model performance in the validation set, focusing on reducing prediction errors and avoiding overfitting. This process ensured that the CNN model was stable and capable of capturing the underlying patterns in the data. This model architecture effectively integrates spatial information across multiple scales and variable types, leveraging the structure of the input data and the physics-informed design of the simulation domain. By avoiding the use of recurrent connections and instead focusing only on spatial encoding through convolution and dense layers, the model achieves both computational efficiency and strong predictive performance.

2.3.1. Quantitative model evaluation metrics

After training, the performance of the model was quantitatively evaluated in the test dataset, which includes wildfire simulation cases that were not seen during training or validation. Three statistical metrics were used to assess the accuracy and quality of the predictions:

- Root Mean Square Error (RMSE): Indicates the square root of the average squared differences between predicted and actual values, representing the magnitude of prediction errors as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{GHF}_{\mathrm{predicted},i} - \mathrm{GHF}_{\mathrm{actual},i})^2} \tag{2}$$

  Where $GHF_{\mathrm{predicted}}$ is the ground heat flux values predicted by the CNN model, and GHF_{actual} is the corresponding values from the WRF-SFIRE dataset. N represents the total number of samples.
- Mean Absolute Error (MAE): Measures the average absolute differences between the predicted and actual values, reflecting the average deviation as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{GHF}_{\mathrm{predicted},i} - \mathrm{GHF}_{\mathrm{actual},i}\right| \tag{3}$$

- Coefficient of Determination (R^2): Evaluates how well the predicted values approximate the actual values, indicating the goodness of fit, and calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(\mathrm{GHF}_{\mathrm{actual},i} - \mathrm{GHF}_{\mathrm{predicted},i}\right)^2}{\sum_{i=1}^{N} \left(\mathrm{GHF}_{\mathrm{actual},i} - \overline{\mathrm{GHF}}_{\mathrm{actual}}\right)^2} \quad (4)$$

Here $\overline{\mathrm{GHF}}_{\mathrm{actual}}$ is the mean of the values.

2.3.2. Qualitative assessment through spatiotemporal contour visualization

To visually assess the spatial accuracy and temporal evolution of the CNN predictions, a dedicated Python-based post-processing framework was developed to generate two-dimensional contour plots of both predicted and reference ground heat flux (GHF) fields at each simulation time step. The visualization procedure was applied sequentially over the full temporal duration of each wildfire simulation case, enabling direct frame-by-frame comparison of the predicted and actual fire behavior patterns. The generated contour plots were subsequently compiled into animated GIFs to provide a continuous representation of the spatiotemporal evolution of wildfire spread and associated heat flux dynamics.

These visualizations allow qualitative evaluation of several important aspects of model performance, including the location and propagation of the fire front, the spatial structure of high-intensity burning regions, and the temporal development of heat flux distributions. By comparing the predicted and reference GHF fields over time, the animations help identify the model capability to reproduce complex fire spread characteristics such as front shape, propagation direction, expansion rate, and localized intensity variations.

In addition to quantitative statistical metrics, the animated visualizations provide an intuitive and physically interpretable assessment of prediction quality, particularly for evaluating spatial coherence and dynamic fire behavior that may not be fully captured by scalar error measures alone.

2.3.3. SAL (Structure-Amplitude-Location) analysis

The Structure-Amplitude-Location (SAL) method was used to perform a more comprehensive spatial evaluation (Wernli et al., 2008). SAL is a feature-based verification technique designed to compare 2D spatial fields such as precipitation (Wernli et al., 2008) or aerosol mass loading (Bruckert et al., 2022). In this study, the SAL (Structure–Amplitude–Location) verification framework is applied to evaluate the precision of the GHF predictions generated by a CNN model. CNN-based predictions are compared to reference data produced by the WRF-SFIRE wildfire simulation model. The SAL method provides a spatially meaningful assessment. It is a feature-based verification technique designed to compare 2D spatial fields by decomposing prediction errors into three interpretable components: Structure (S), Amplitude (A), and Location (L) (Wernli et al., 2009; Gofa et al., 2018).

Component A evaluates the difference in overall magnitude (mean intensity) of GHF between the CNN predictions and the WRF-SFIRE simulations. The A score ranges from −2 to 2, with 0 indicating perfect agreement. A positive value suggests overestimation by the CNN, while a negative value indicates underestimation.

$$A = \frac{\overline{GHF}_{predicted} - \overline{GHF}_{actual}}{0.5 \times (\overline{GHF}_{predicted} + \overline{GHF}_{actual})} \quad (5)$$

Where $GHF_{\mathrm{predicted}}$ is the GHF predicted by the CNN model, GHF_{actual} is GHF output from WRF-SFIRE, and $\overline{GHF}$ is the spatial average in the domain. The S component quantifies differences in the spatial distribution and shape of GHF patterns. It measures whether the predicted heat structures are more dispersed or more localized than those in the WRF-SFIRE output. It also ranges from −2 to 2, with 0 indicating perfect agreement. A positive S value indicates broader predicted patterns, while a negative value reflects more compact structures.

$$S = \frac{V_{\mathrm{predicted}} - V_{\mathrm{actual}}}{0.5 \times (V_{\mathrm{predicted}} + V_{\mathrm{actual}})} \quad (6)$$

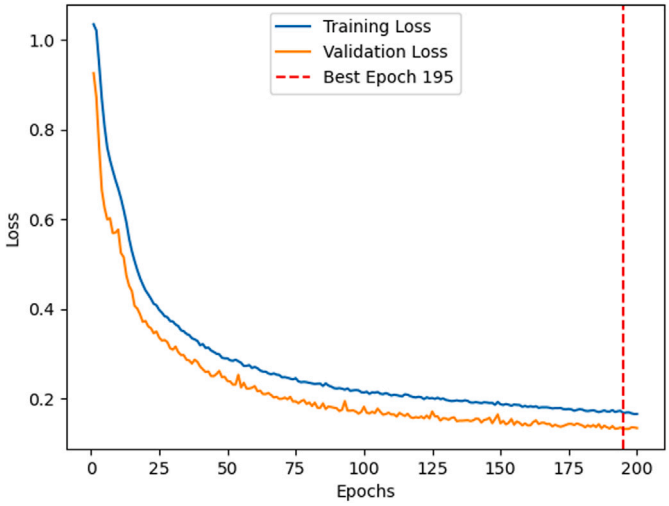

Fig. 2. Training and validation loss over epochs during CNN model training.

Here V is computed from the thresholded objects (contiguous regions above a chosen GHF threshold, typically corresponding to fire-active patches or hotspots):

$$V = \frac{\sum_i \left(\frac{GHF_i}{GHF_i^{\max}}\right)}{\sum_i GHF_i} \quad (7)$$

Where GHF_i is the total GHF in object i and GHF_i^{max} is the peak of GHF in object i. The L component measures the spatial alignment between the predicted and observed GHF distributions. It ranges from 0 (perfect match) to 2 (maximum displacement) and is composed of two sub-components:

$$L = L_1 + L_2 \quad (8)$$

L_1 measures the shift between the intensity-weighted centers of heat of the two full fields as:

$$L_1 = \frac{\left|\mathbf{x}(GHF_{\mathrm{predicted}}) - \mathbf{x}(GHF_{\mathrm{actual}})\right|}{d} \quad (9)$$

L_2 reflects differences in how the objects are spread around their field center of heat:

$$L_2 = \frac{2}{d}\left|\left(\frac{\sum_i GHF_i\,|\mathbf{x}_i - \mathbf{x}|}{\sum_i GHF_i}\right)_{\mathrm{predicted}} - \left(\frac{\sum_i GHF_i\,|\mathbf{x}_i - \mathbf{x}|}{\sum_i GHF_i}\right)_{\mathrm{actual}}\right| \quad (10)$$

Where x_i is the intensity-weighted center of heat of object, $x(GHF)$ is the center of heat of the whole field; and d is the maximum distance across the domain (used for normalization). Each sub-component $L_1, L_2 \in [0,1]$, so $L \in [0,2]$ with $L = 0$ indicating perfect location agreement.

3. Results and discussion

3.1. Model training and validation performance

The CNN model was trained using 70% of the dataset, while 15% was reserved for validation and 15% for final testing. Throughout the training process, the model and the validation loss function were monitored to assess convergence and generalization performance. Fig. 2 shows the evolution of training and validation losses over epochs, indicating that the model reached a stable minimum without significant overfitting. The optimal model checkpoint was selected on the basis of the lowest validation loss at epoch equal to 195.

Table 4
Summary of hyperparameter tuning and selected optimal values for the CNN model.

Hyperparameter	Range tested	Optimal value
Learning rate	10^{-3}, 5×10^{-4}, 10^{-4}, 5×10^{-5}, 10^{-5}	5×10^{-5}
Weight decay	10^{-6}, 10^{-5}, 10^{-4}	10^{-5}
Batch size	16, 32, 64, 128	64
Dropout rate	0.2, 0.3, 0.4, 0.5	0.3
Activation function	ReLU, LeakyReLU, GELU, Identity, Softplus, ELU	ReLU
Loss function	MSE, MAE, Huber, SmoothL1, peak-aware MSE[a], smoothness-constrained MSE[b]	MSE
Optimizer	Adam, RMSprop, SGD	Adam
Pooling method	Max pooling, Average pooling	Max pooling
Epochs (max)	100, 150, 200, 250, 300	200

[a] Custom loss emphasizing high-intensity pixels (peak-aware MSE).
[b] MSE with spatial smoothness penalty to reduce sharp gradients.

From a computational perspective, physics-based wildfire models such as WRF-SFIRE solve coupled atmospheric dynamics and fire behavior through time-stepping numerical integration, requiring the system state to be updated sequentially over many time steps. As a result, generating a full simulation is computationally intensive and typically involves substantial runtime on high-performance computing systems. In contrast, the CNN surrogate model is trained offline and subsequently performs inference through a single forward pass of the network. This allows direct prediction of the spatial ground heat flux field at a given time step without iterative time integration. This fundamental difference in computational approach leads to a substantial reduction in cost per prediction step, highlighting the potential of the surrogate model for rapid forecasting and time-sensitive applications. In order to quantitatively evaluate the computational efficiency of the proposed framework, an inference runtime benchmark was conducted on an NVIDIA GPU during the test phase. The trained CNN model processed 1659 test samples in 48 s, corresponding to an average inference time of 29 ms per sample and a throughput of 34 samples per second. These results demonstrate that the proposed deep learning framework can generate GHF predictions within milliseconds, providing a substantial computational advantage over conventional coupled fire–atmosphere simulations typically executed on HPC systems. The obtained runtime performance highlights the potential suitability of the proposed model for computationally efficient wildfire prediction applications

Before the final training, an extensive hyperparameter search was conducted to identify the most effective configuration for stable convergence and accurate prediction. As summarized in Table 4, the tuning covered a broad range of key parameters: learning rate, weight decay, batch size, dropout rate, activation functions, pooling strategies, optimizers, and the maximum number of training epochs. Based on this systematic exploration, the optimal model employed the Adam optimizer with a learning rate of 5×10^{-5}, a weight decay of 10^{-5}, a batch size of 64, a dropout rate of 0.3, max pooling, and ReLU activation, providing stable convergence and effective regularization.

Particular emphasis was placed on the choice of loss function, as it directly determines how the model balances errors across different fire intensities and spatial structures. Alongside standard loss functions including mean squared error (MSE), mean absolute error (MAE), SmoothL1, and Huber loss, two specialized formulations were tested. The peak-aware MSE enhances the standard MSE by assigning higher weights to high-intensity pixels, which represent regions of strong ground heat flux. This encourages the network to more accurately capture sharp gradients and the extreme values characteristic of active fire fronts. The smoothness constraint loss function is an MSE-based regularization that introduces a spatial regularization term that penalizes abrupt variations between neighboring grid cells, promoting physically coherent and spatially consistent predictions of ground heat flux. The mathematical formulations of these custom loss functions, including the weighting scheme and regularization coefficients used in the present implementation, are provided in the Supplementary Material. Although the custom losses enhanced sensitivity to extreme events and improved spatial consistency, the conventional MSE yielded the lowest validation loss and provided the best overall balance between intensity sensitivity and spatial smoothness, and was therefore adopted for final training. A quantitative comparison of these formulations is provided in Table 5. The results show that the standard MSE achieves the lowest error metrics and the highest coefficient of determination among the evaluated configurations, indicating the best overall predictive performance. Both the peak-aware and smoothness-constrained losses yield higher errors and lower R^2 values, with the smoothness-constrained formulation performing better than the peak-aware variant but still not exceeding the accuracy of the standard MSE. These findings indicate that the investigated loss modifications, under the current implementation and selected hyperparameter settings, did not improve the predictive capability of the model relative to the standard MSE formulation. However, since only two relatively simple and untuned custom loss formulations were evaluated, these results should not be interpreted as evidence that loss-function modifications are inherently ineffective for mitigating peak underestimation. Consequently, the conventional MSE was selected for the final training due to its superior overall performance in the present study.

Table 5
Quantitative comparison of selected loss functions including MSE, peak-aware MSE, and smoothness-constrained MSE during hyperparameter tuning on the test dataset computed over burning (non-zero GHF) grid cells.

Loss function	RMSE (kW m^{-2})	MAE (kW m^{-2})	R^2 (%)
MSE	14.3	6.6	84
Peak-aware MSE	21.7	12.9	70
Smoothness-constrained MSE	17.9	8.3	79

Table 6
Ablation study results evaluated over burning cells of the test dataset. RMSE and MAE are reported in kW m^{-2} and R^2 is expressed as percentage.

Model variant	RMSE (kW m^{-2})	MAE (kW m^{-2})	R^2 (%)
Full model (Dropout=0.0)	16.0	7.5	83
Full model (Dropout=0.1)	16.9	7.6	81
Full model (Dropout=0.3)	**14.3**	**6.6**	**84**
Full model (Dropout=0.5)	19.8	9.7	75
K3 only	18.0	8.5	79
K5 only	17.6	8.0	80
K7 only	17.4	8.1	80
K3+K5	17.3	7.8	80
K3+K7	16.8	7.6	82
Reduced-width model	17.7	8.0	80

To further evaluate the robustness of the proposed architecture and quantify its sensitivity to structural design choices, a comprehensive ablation study was conducted. While the hyperparameter search identified a configuration that ensured stable convergence, the ablation analysis isolates the contribution of three key architectural components: (i) dropout rate, (ii) convolutional kernel configuration (multi-scale feature extraction), and (iii) network width (number of filters).

All ablation variants were evaluated on the held-out test dataset using physically de-normalized ground heat flux (GHF) values. Performance metrics were computed exclusively over burning cells (non-zero GHF), ensuring that the evaluation reflects model behavior at active fire fronts, where predictive accuracy is most critical. RMSE and MAE are reported in kW m^{-2}, and R^2 values are expressed as percentages. The results are presented in Table 6.

The first four rows of Table 6 summarize the sensitivity of the full architecture (as defined in Table 3) to variations in dropout rate. The

results demonstrate a clear bias–variance trade-off. A moderate dropout rate of 0.3 yields the best performance, achieving the lowest RMSE (14.3 kW m^{-2}) and highest R^2 (84%). Removing dropout entirely (0.0) slightly degrades performance, suggesting mild overfitting. In contrast, excessive dropout (0.5) substantially increases RMSE (19.8 kW m^{-2}) and reduces R^2 to 75%, indicating underfitting due to overly aggressive regularization. These findings confirm that moderate stochastic regularization enhances generalization across diverse wildfire scenarios. Rows 5–9 of Table 6 present the impact of modifying the convolutional kernel configuration. The single-kernel models (3 × 3 (K3 only), 5 × 5 (K5 only), and 7 × 7 (K7 only)) are treated as architectural baselines, representing simplified network designs with a single receptive field. These baselines provide a reference for evaluating the effectiveness of the proposed multi-scale feature extraction strategy. Single-kernel models (3×3, 5×5, or 7 × 7 only) consistently exhibit higher errors and lower R^2 compared to the full multi-kernel configuration. Dual-kernel combinations (K3+K5 and K3+K7) improve performance relative to single-kernel variants but remain inferior to the full multi-scale design. The full architecture reduces RMSE by approximately 15%–20% relative to single-kernel models. This demonstrates that wildfire heat flux prediction benefits from simultaneous extraction of fine-scale gradients at the fire front and broader mesoscale spatial structures. A single receptive field size is insufficient to capture the multi-scale dynamics inherent in coupled atmosphere–fire interactions. The final row of Table 6 shows the performance of a reduced-width model in which the filter progression (16, 32, 64, 128, 256) was halved. The resulting degradation in performance (RMSE = 17.7 kW m^{-2}, R^2 = 80%) indicates that adequate representational capacity is necessary to model the complex nonlinear relationships between meteorological forcing and fire behavior. The observed performance decline confirms that the proposed architecture is not over-parameterized, but appropriately scaled to the physical complexity of the problem.

Collectively, the ablation results demonstrate that the adopted architecture is empirically justified rather than arbitrary. The combination of multi-scale convolutional blocks and moderate regularization provides the optimal balance between representational capacity and generalization. These findings strengthen the methodological rigor of the proposed surrogate modeling framework and confirm the robustness of the architectural design choices. To further assess the statistical robustness of the dropout ablation results, additional independent training runs using different random states were performed, and the corresponding mean ± standard deviation statistics are provided in the Supplementary Material. The repeated experiments confirmed that the relative ranking and performance trends reported in Table 6, particularly the superior performance of the dropout = 0.3 configuration, remain consistent across different random splits.

3.2. Quantitative evaluation

The predictive capacity of the model was evaluated on the test dataset using RMSE, MAE, and R^2. These metrics quantify the magnitude and consistency of the prediction errors in all test samples as described in Section 2.3.1. The results summarized in Table 7. It should be noted that the performance metrics reported in Table 7 are computed over burning areas only, excluding grid cells with zero GHF values. This ensures that the evaluation reflects model accuracy within active fire regions rather than being dominated by non-burning background pixels.

The results indicate that the model achieved low RMSE and MAE values, with an R^2 above 0.8, suggesting strong predictive performance in capturing the spatial–temporal evolution of ground heat flux within active fire zones.

Fig. 3 illustrates the temporal evolution of MAE and RMSE in predicting GHF across all wildfire scenarios, evaluated over 79 time steps. The solid lines represent the metrics computed on all grid cells, including regions with zero GHF, while the dashed lines exclude zero-flux areas to focus on active fire zones. When zero values are included, both MAE and RMSE remain consistently low, with an average MAE of approximately 0.5 kW m^{-2} and RMSE around 1 kW m^{-2}, due to the dominance of non-burning areas in the domain. In contrast, excluding zero values reveals substantially higher errors, with RMSE reaching up to 24 kW m^{-2} during the early stages of fire progression. Over time, the errors in the no-zero case stabilize, indicating improved model performance as the fire evolves and the heat flux distribution becomes more coherent. This contrast highlights the importance of evaluating model accuracy specifically within active fire zones.

Table 7
CNN model performance on the test dataset computed over burning (non-zero GHF) grid cells.

Metric	Value
RMSE	14.3 kW/m^2
MAE	6.6 kW/m^2
R^2	0.84

Fig. 4 presents a density scatter plot that compares the predicted and actual ground heat flux (GHF) values at all points in the spatiotemporal grid. The intensity of the color indicates the density of points per pixel, with yellow denoting the regions of the highest data concentration. The red dashed line represents the ideal prediction trend. The distribution reveals a strong concentration of data points along the ideal prediction trend (diagonal), indicating that the model accurately captures the majority of the GHF values. However, systematic deviations are observed in certain regimes. The model demonstrates strong agreement with actual values in the low-to-moderate GHF range, whereas deviations emerge at higher fluxes, where underprediction becomes more pronounced. Specifically, the model tends to underestimate the midrange GHF values, approximately 100–175 kW/m^2, as observed by the asymmetric density distribution located above the diagonal red line. In contrast, underprediction is apparent for high-intensity GHF events that are stronger than 175 kW/m^2, indicating that the model may struggle to extrapolate beyond the dominant range of the training data. Despite these challenges, the overall structure of the plot supports the general ability of the model to reproduce spatio-temporal variability in GHF with reasonable accuracy.

The distribution of prediction error, shown in Fig. 5, shows that most of the model predictions are close to the actual values, with the highest number of errors centered around zero. The error values range from −50 to 50 kW/m^2, and the use of the vertical axis scale helps to highlight even small differences. Although the model performs well overall, the distribution is slightly shifted to the left, suggesting a tendency to underestimate the actual values. This pattern shows that, although the model is generally accurate, there is a small bias in its predictions.

3.3. Visual assessment: Comparison of the ground heat flux contours

As described in Section 2.3.2, the spatiotemporal evaluation of wildfire spread is performed using 2D contour plots of ground heat flux (GHF) across all time steps and wildfire scenarios. These visualizations provide a comprehensive assessment of the capacity of the CNN model to replicate the temporal evolution and spatial distribution of the surface heat flux. The full set of contour and GIF outputs is publicly accessible through the link provided in the Data Availability section. In order to improve visual clarity, very low-magnitude background values in non-burning regions were suppressed in some contour plots to reduce visual noise and emphasize dominant fire structures. This operation was applied exclusively for visualization purposes and did not modify the raw model outputs used for quantitative evaluation or statistical analysis.

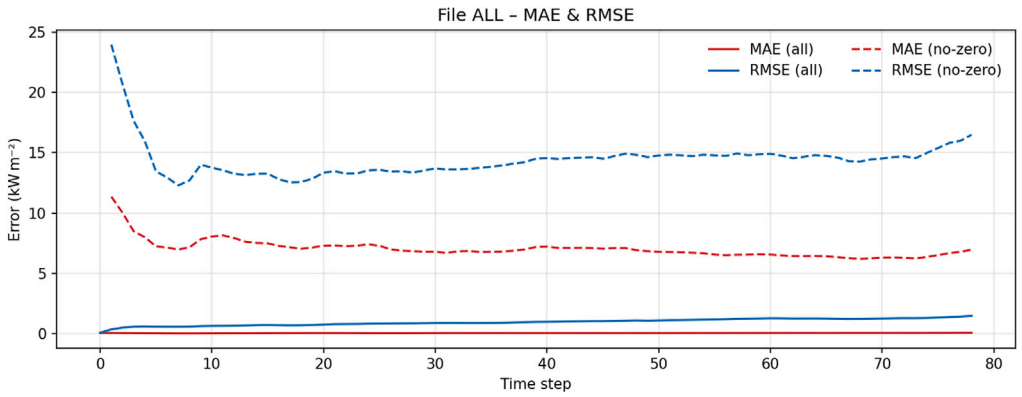

Fig. 3. Temporal variation of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in predicting ground-level heat flux (GHF) across all scenarios.

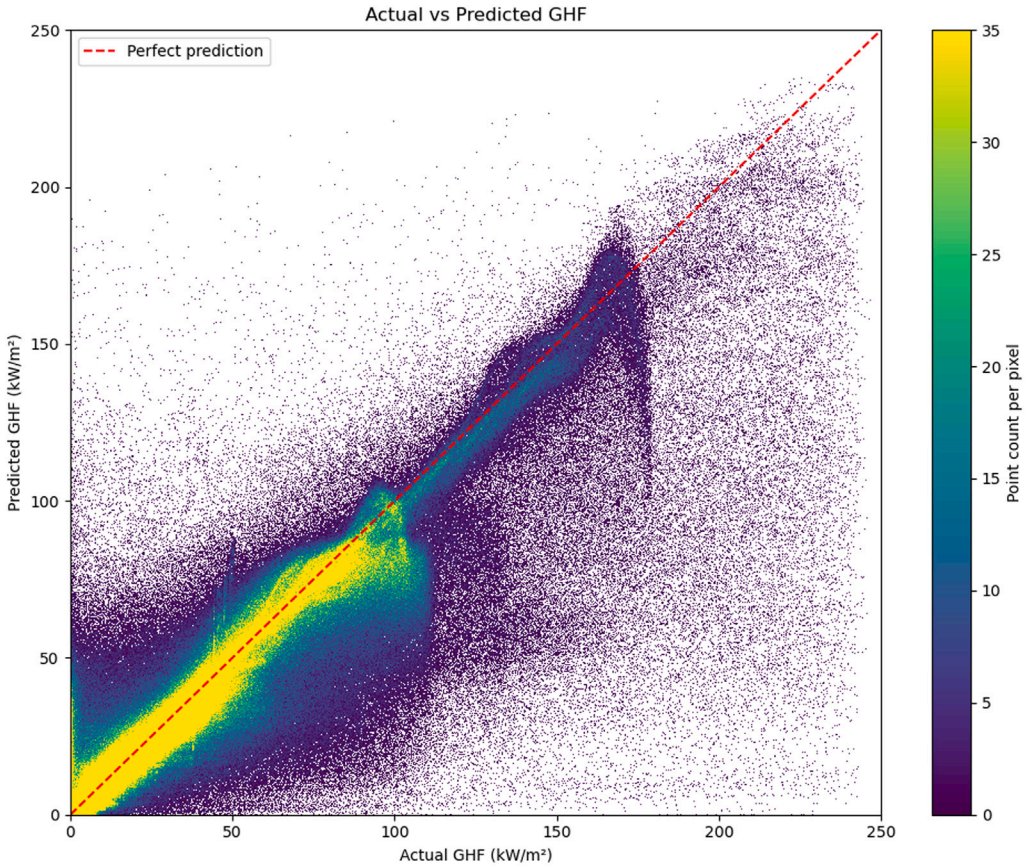

Fig. 4. Density scatter plot comparing predicted and actual ground heat flux (GHF) values across all spatiotemporal grid points.

Figs. 6–9 present representative contour plots in different time steps for four different wildfire scenarios, each characterized by different environmental conditions, including the ambient wind speed, fuel type, and atmospheric stability profile. For each case, the left contour shows the actual GHF derived from the WRF-SFIRE simulations, while the middle contour presents the corresponding prediction generated by the trained CNN model. In order to find model overestimation or underestimation, the signed error between actual GHF and predicted GHF is computed and visualized in the right contour of each figure as:

$$\text{Error} = \text{Predicted GHF} - \text{Actual GHF} \tag{11}$$

Visual evaluation reveals good agreement between the predicted and actual heat flux distributions in both shape and intensity. For example, in Fig. 6, corresponding to scenario W10F7R1, the fire exhibits an approximately parabolic shape due to the higher wind speed and dense fuel load. The CNN model accurately reproduces both the extent and intensity of the heat flux in this case. It is also observed that as time passes from Figs. 6(a) to 6(d), the fire spreads further. As observed in the right column, the model shows spatially varying errors: It tends to overestimate GHF along the lateral edges of the fire front, while underestimating the intensity in the central region. The strongest underestimate occurs at the advancing fire front, where the actual GHF peaks are not fully captured by the model.

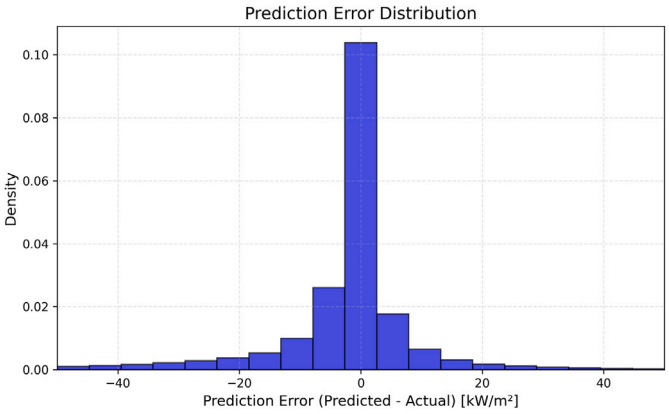

Fig. 5. Histogram of prediction error distribution (Predicted – Actual) in kW/m^2.

In contrast, Fig. 7 shows the scenario W5F1R3, where the spread of the fire is significantly weaker due to a limited type of fuel (F1), characterized by a low fuel load, as described in Table 1. Although the CNN model generally underestimates the intensity of GHF in this scenario, it successfully captures the overall shape and limited spread of the fire. At time step 30, the model predicts a higher GHF along the outer edges of the fire line as can be seen in Fig. 7(a). It is observed that a consistent pattern appears in later stages (time steps 45, 60, and 75) so that the model underestimates the intensity at the forward edge of the fire, where it is actively spreading, while overestimates values behind the front and along the sides. Overall, this leads to a smoother predicted fire front, with weaker hot spots in the advancing zones and values higher than the actual values in the burned regions.

Fig. 8 shows the scenario W5F4R1, where intense and aggressive fire behavior is observed due to the presence of a more flammable fuel type (F4) combined with favorable environmental conditions. The CNN model performs well in capturing both the intensity and spatial extent of the fire front. Meanwhile, Fig. 9, corresponding to scenario W5F4R3, shows a more moderate spread. This variation is attributable to differences in the atmospheric stability profile (R3 vs. R1), which influences vertical mixing and heat dispersion. In addition, the right column of Figs. 8 and 9 indicates that the model underestimates the GHF in fire front edges and overestimate it behind the front regions and sides, but overall it tends to underestimate the values of these wildfire scenarios.

In general, the model demonstrates consistent performance under a wide range of conditions, effectively capturing the spatial dynamics and temporal evolution of fire intensity. This indicates a strong ability to generalize across wildfire scenarios with varying input conditions. However, despite this overall robustness, a systematic underestimation of peak intensity at the advancing fire front is consistently observed in the predictions. In order to investigate the origin of this behavior, different loss functions were evaluated during model development, including MSE, MAE, Huber, SmoothL1, as well as custom formulations such as a peak-aware MSE and a smoothness-constrained loss, as summarized in Table 4. In particular, the custom loss formulations did not improve the representation of local maxima nor eliminate the bias, as quantitatively demonstrated in Table 5, suggesting that the investigated loss modifications were not sufficient to substantially reduce the underestimation of peak fire-front intensity under the current implementation and selected hyperparameter settings. To further separate the contributions of data imbalance and architectural smoothing to this bias, quantitative evidence from both the loss function comparison and the ablation study is examined. However, because only two relatively simple and untuned custom formulations were examined, the present results do not exclude the possibility that alternative loss designs, adaptive weighting strategies, or further hyperparameter optimization could improve peak-value prediction performance. This observation suggests that the underlying cause is more closely related to data characteristics and model behavior.

A more reasonable explanation lies in the characteristics of the training data and the model architecture. First, the composition of the dataset plays a critical role. High ground heat flux (GHF) values at the fire front occur over relatively small spatial regions and short time intervals, whereas most of the domain is dominated by low to moderate flux values. This imbalance results in limited exposure of the model to extreme events, reducing its ability to accurately learn and reproduce peak intensities. The peak-aware MSE loss was specifically designed to counteract this imbalance by assigning higher weights to high-intensity pixels. Despite this intent, it yielded substantially worse overall performance (RMSE = 21.7 kW m^{-2}, R^2 = 70%) compared to standard MSE (RMSE = 14.3 kW m^{-2}, R^2 = 84%) as indicated in Table 5, suggesting that the simple weighting scheme applied here was insufficient under the current configuration and may have introduced gradient instability during training. Nevertheless, the measurable change in model response when the loss function was modified to emphasize rare high-intensity events serves as indirect quantitative evidence that data imbalance contributes independently to peak underestimation. More sophisticated resampling strategies, focal loss formulations, or carefully tuned weighting schemes therefore remain a promising direction for future work. Second, the convolutional neural network architecture introduces an inherent smoothing effect. Operations such as strided convolutions, pooling, and upsampling tend to diffuse localized high-intensity features, thereby attenuating sharp gradients at the fire front. The contribution of architectural smoothing is directly evidenced by the ablation study results (Table 6). Reduced-capacity variants including the reduced-width model (RMSE = 17.7 kW m^2, R^2 = 80%) and single-kernel configurations (RMSE = 17.4–18.0 kW m^2, R^2 = 79%–80%) consistently exhibit higher errors and stronger underestimation of peak values compared to the full multi-scale architecture (RMSE = 14.3 kW m^2, R^2 = 84%). This confirms that insufficient representational capacity and a limited receptive field amplify spatial over-smoothing, suppressing the resolution of sharp intensity gradients at active fire fronts. Conversely, the full architecture combining multi-scale convolutional blocks (K3+K5+K7) with moderate dropout (0.3) substantially mitigates this source of bias by enabling simultaneous extraction of fine-scale gradients and broader spatial structures.

The combined influence of data imbalance and architectural smoothing leads to systematically smoother predictions, which in turn results in the underestimation of peak fire-front intensity, even though the overall spatial structure of the fire is well captured. These two mechanisms act through distinct pathways: architectural smoothing globally limits the spatial resolution of intensity gradients, while data imbalance selectively suppresses the accurate prediction of rare, extreme-value events. Disentangling these effects confirms that the adopted architecture already addresses the smoothing component effectively, and that future efforts should focus on the data imbalance component through more advanced loss formulations or training data augmentation strategies targeting high-intensity fire-front samples.

3.4. SAL score distribution

In order to assess spatial agreement between the predicted and actual GHF fields, the SAL metrics (structure-amplitude-location) were applied to the entire dataset in all time steps as in the formulation presented in Section 2.3.3. Here, the temporal evolution and convergence behavior of the SAL components throughout the fire event are validated.

Fig. 10 illustrates the temporal evolution of the SAL components in the structure–amplitude (S–A) plane, with color indicating the location (L) component. The Structure (S) and Amplitude (A) components are plotted on the horizontal and vertical axes, respectively, while the Location (L) component is represented by the color scale of the points.

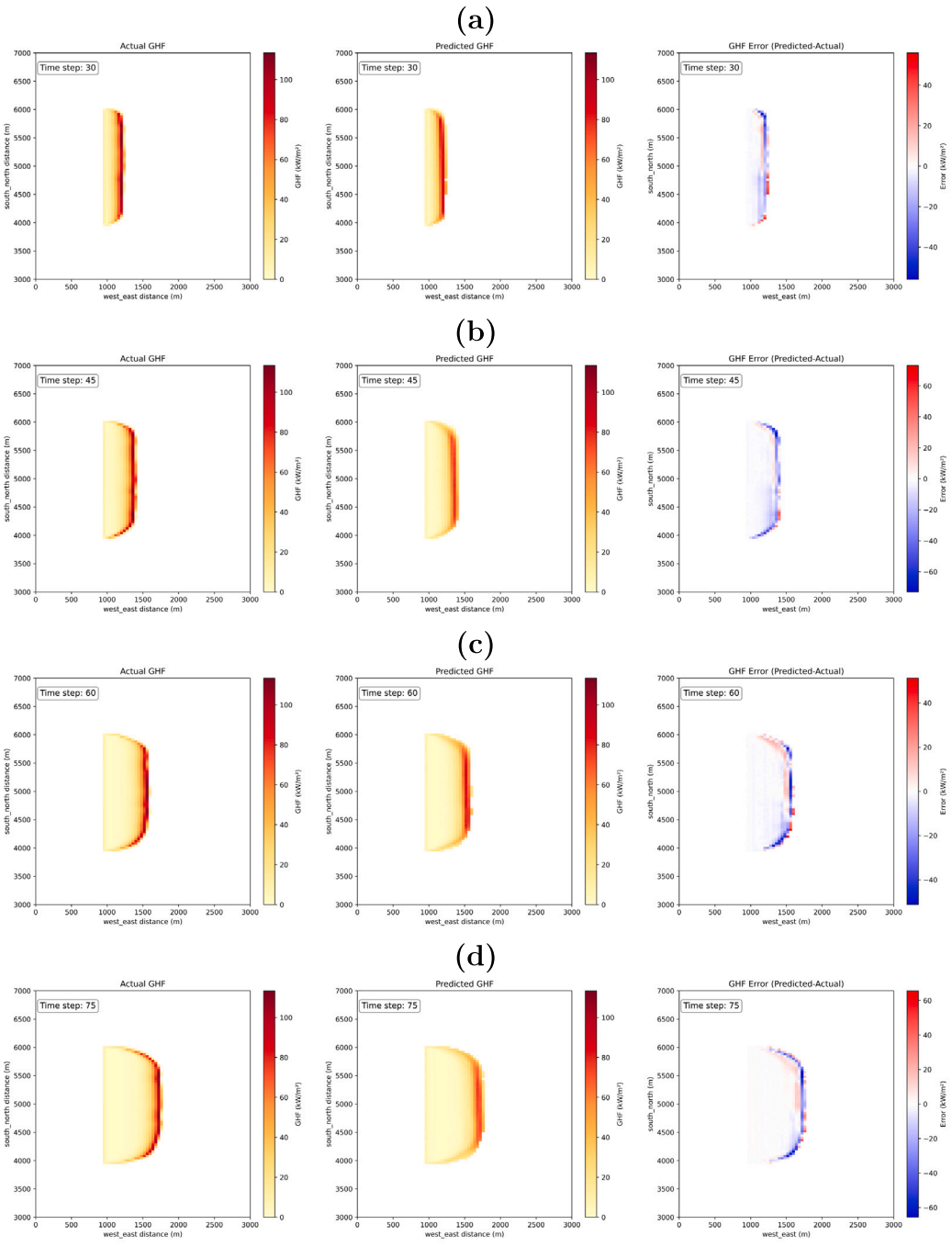

Fig. 6. Visual comparison of ground heat flux for different wildfire scenario W10F7R1 at four time steps: (a) 30, (b) 45, (c) 60, and (d) 75. Here, the left contour shows the Actual GHF from WRF-SFIRE simulations, the middle column shows the Predicted GHF from the CNN model, and the right column shows Error.

To improve readability and reduce visual density, the SAL trajectory is shown using temporally aggregated values. The first few active-fire time steps are displayed individually, while subsequent steps are averaged over fixed temporal windows. The temporal evolution is explicitly illustrated using an arrowed trajectory, where each marker represents a successive time step in the fire progression. The initial active-fire time step is highlighted with a red-edged circle, intermediate steps are shown with black-edged circles, and the final stage is marked with a pink-edged circle. This visual encoding clarifies the temporal development of the fire and facilitates interpretation of the convergence behavior of the simulations towards the optimal reference point (0,0,0).

It should be noted that each SAL diagram in Figs. 10(a)–(d) corresponds to the wildfire scenarios shown in Figs. 6–9. Comparing the different SAL diagrams in Figs. 10(a)–(d) with the GHF contours in Figs. 6–9 demonstrates the enhanced ability of the CNN model to predict extreme wildfire scenarios. For example, in Fig. 10(b), which represents the weakest wildfire scenario as shown in Fig. 7, the scatter is wider and mid-run fluctuations are larger, showing a period of stronger intensity and structure errors. . However, the model successfully captures the location parameter. The CNN steadily improves its spatial alignment (L) and partly stabilizes its structural agreement (S), but moderate underprediction of amplitude persists ($A < 0$), reflecting the challenge of reproducing small, fluctuating heat flux in a low-intensity fire.

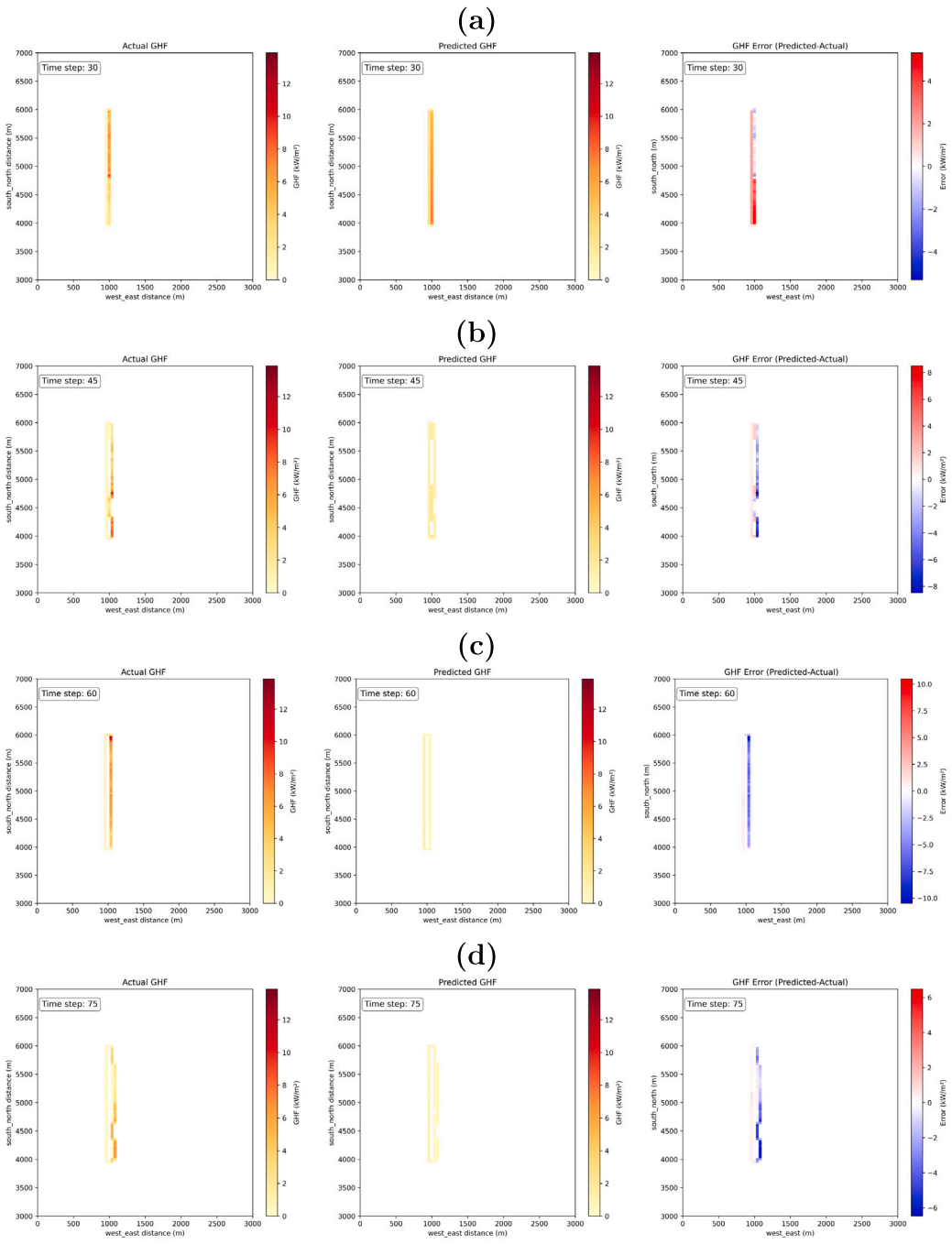

Fig. 7. Visual comparison of ground heat flux for different wildfire scenario W5F1R3 at four time steps: (a) 30, (b) 45, (c) 60, and (d) 75. Here, the left contour shows the Actual GHF from WRF-SFIRE simulations, the middle column shows the Predicted GHF from the CNN model, and the right column shows Error.

The results demonstrate that at the beginning of the simulation, the SAL points are moderately displaced from the origin, indicating initial discrepancies in both intensity and spatial structure. Across all cases, the first points start somewhat away from the origin and often have negative A, meaning the CNN slightly underestimates ground heat flux and shows minor shape differences at the beginning. However, in Figs. 10(a), (c), (d) as the simulation progresses, the points increasingly converge around the origin ($S = 0, A = 0$), with the L values decreasing in magnitude. This convergence indicates that the model progressively aligns better with the observed fire spread, both in structure and magnitude (S, A), and the spatial displacement (L) also reduces with time. This behavior occurs because, in the early stage, only a small portion of the domain shows strong heat flux, making the data very sparse and harder for the CNN to learn precisely. The SAL metric itself also exaggerates small location shifts when the affected area is tiny. Moreover, the training dataset contains more samples of later fire stages, so the network naturally performs better as the fire grows and the spatial patterns become larger and more frequent.

In general, the distribution of the SAL scores confirms that the model achieves stable and reliable spatial predictions, particularly in the later stages of fire spread, where accurate prediction is crucial for fire impact assessments and response planning.

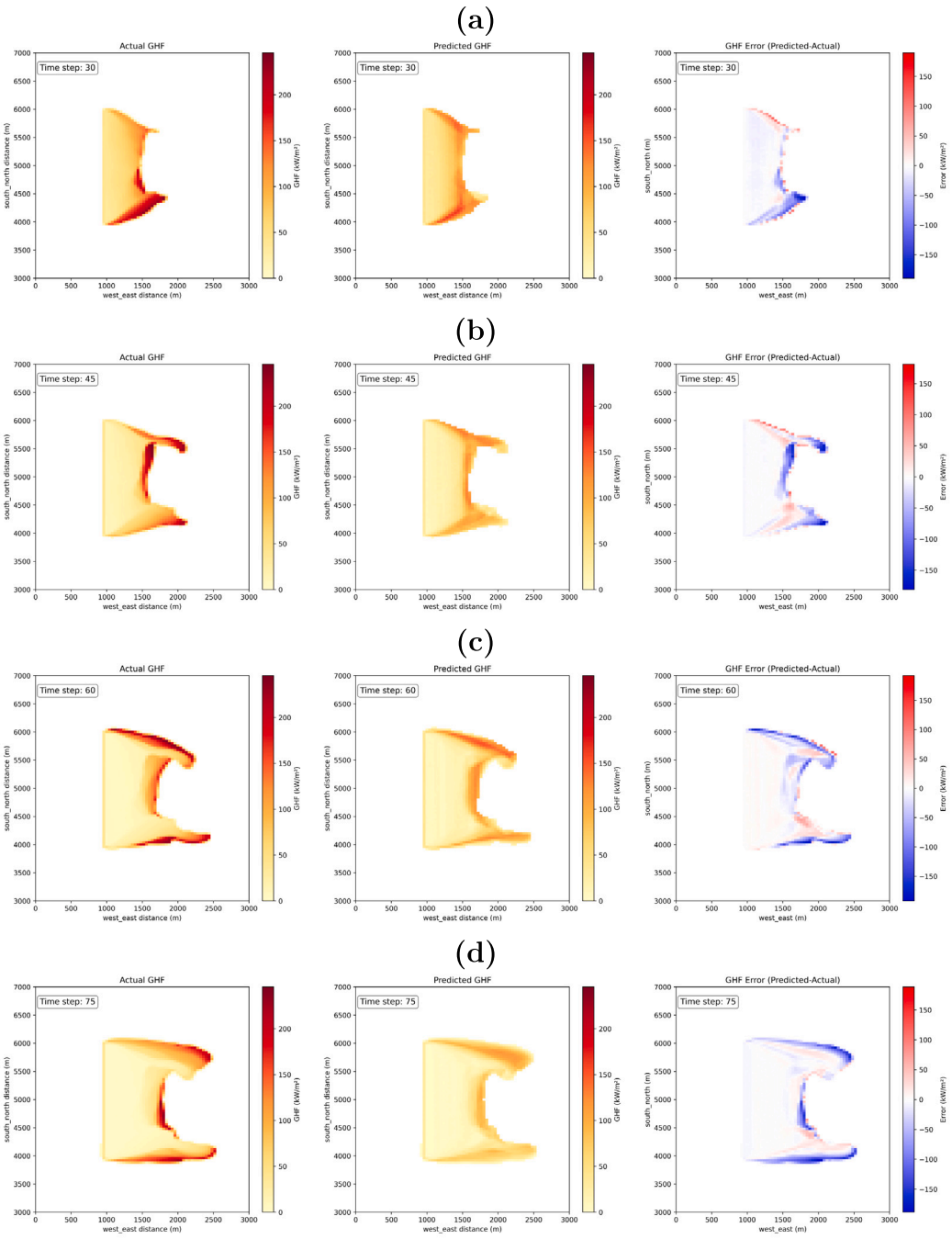

Fig. 8. Visual comparison of ground heat flux for different wildfire scenario W5F4R1 at four time steps: (a) 30, (b) 45, (c) 60, and (d) 75. Here, the left contour shows the Actual GHF from WRF-SFIRE simulations, the middle column shows the Predicted GHF from the CNN model, and the right column shows Error.

4. Conclusion

This study developed and evaluated a CNN model to predict GHF in wildfire simulations using meteorological and fuel data as input. The model architecture used a multiscale feature extraction strategy through ConcatConvBlocks, followed by fully connected layers, allowing the extraction of both fine and coarse spatial patterns critical for fire spread prediction.

Training and validation of the model demonstrated that the CNN model has a strong predictive performance, as indicated by quantitative metrics: RMSE of 14.3 kW/m^2, MAE of 6.6 kW/m^2, and R^2 of 0.84 in the test dataset. Further qualitative evaluation using 2D contour comparisons of actual versus predicted GHF confirmed that CNN accurately reproduced both the intensity and the spatial structure of the wildfire spread across different environmental scenarios. The SAL analysis provided a more detailed spatial evaluation. The results showed that most of the predictions had values of S, A, and L close to zero, suggesting a high degree of structural, amplitude, and positional agreement between the model predictions and the WRF-SFIRE simulation outputs over time. The temporal evolution of SAL scores indicated that CNN predictions converged towards more accurate representations of fire spread as the wildfire evolved, demonstrating the robustness of the model across different conditions of wind, fuel and atmospheric stability. Despite these promising results, some limitations were identified. In scenarios with low fuel load and weak fire behavior, CNN showed minor deviations in predicting fire intensity, although the overall shape of the fire was still correctly captured. This suggests that the model performance may be sensitive to data sparsity and to the representation of extreme

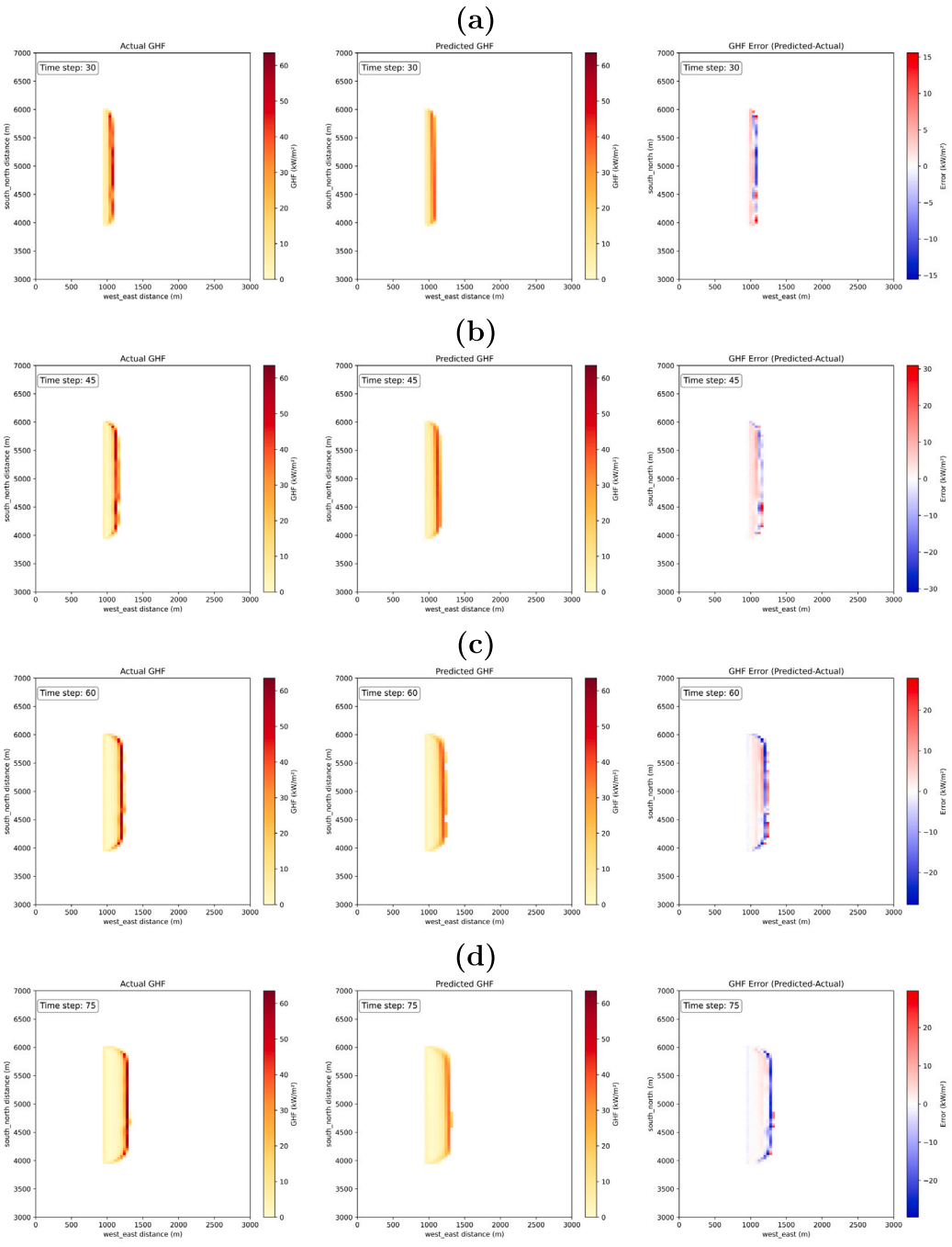

Fig. 9. Visual comparison of ground heat flux for different wildfire scenario W5F4R3 at four time steps: (a) 30, (b) 45, (c) 60, and (d) 75. Here, the left contour shows the Actual GHF from WRF-SFIRE simulations, the middle column shows the Predicted GHF from the CNN model, and the right column shows Error.

conditions within the training set. Furthermore, across most time steps negative ($A < 0$) is observed, indicating a persistent underestimation of fire-front intensity. This bias can be attributed to two interacting factors. On one hand, the limited number of extreme GHF values in the training data reduces the model ability to accurately learn high-intensity peaks. On the other hand, architectural smoothing effects from strided convolutions, pooling, and upsampling further dampen localized maxima. Together, these effects explain the systematic amplitude bias while maintaining good structural agreement, as reflected by near-zero S and L components.

In addition, a key limitation of this study is that the dataset consists of a homogeneous subset of wildfire simulations, all initialized with a fireline length of 2 km and a domain height of 3000 m. While this consistency supports stable model training and evaluation, it restricts the diversity of fire dynamics captured by the model. As a result, direct applicability of the model to wildfire scenarios with different ignition extents, fireline geometries, or domain configurations is limited. Furthermore, the dataset partitioning was performed at the time-step level rather than at the simulation level, meaning that temporally adjacent samples from the same wildfire simulation may be distributed across the training, validation, and test subsets. Due to the strong temporal correlation between consecutive wildfire states, this strategy may introduce temporal leakage and could partially influence the reported predictive performance. Expanding the dataset to include a broader range of initial fire conditions, particularly varying fireline lengths and domain heights, would improve the robustness and generalizability of

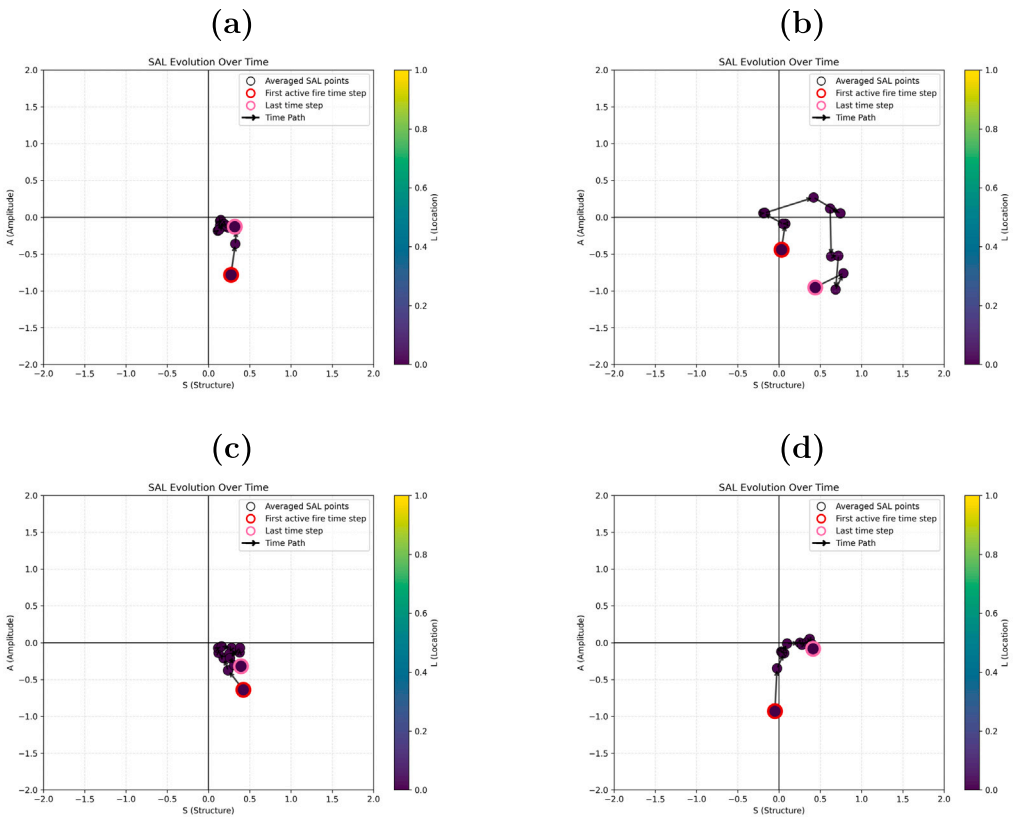

Fig. 10. Temporal evolution of SAL analysis for different wildfire scenarios of: (a) W10F7R1, (b) W5F1R3, (c) W5F4R1, (d) W5F4R3. The arrowed trajectory indicates the temporal progression of the fire, starting from the initial active-fire time step (red-edged circle), through intermediate stages (black-edged circles), and ending at the final stage (pink-edged circle). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the model and enable its application to more diverse wildfire scenarios. Moreover, the use of a single slope of the potential temperature profile as a measure of vertical stability is practical, but simplifies complex atmospheric stratification and introduces additional uncertainty, which should be addressed in the next steps.

Future work will focus on addressing these limitations by expanding the training dataset to include a wider range of fire behavior and environmental conditions. In particular, increasing the representation of extreme fire-front events through targeted sampling and data augmentation strategies will help mitigate the imbalance between high- and low-intensity regions. Simulation-level dataset partitioning should also be investigated, in which complete wildfire scenarios are assigned exclusively to either the training, validation, or test subsets, in order to provide a more rigorous assessment of model generalization to unseen wildfire events. In addition, more advanced loss formulations will be explored to better capture extreme values. For example, focal-type or dynamically weighted loss functions could be used to assign greater importance to high-intensity pixels, improving the learning of sharp gradients at the fire front. Architectural improvements will also be considered to reduce smoothing effects inherent to convolutional networks. In particular, encoder–decoder architectures with skip connections, such as U-Net, may help preserve fine-scale spatial features and improve the reconstruction of sharp fire-front structures. Furthermore, the integration of explicit physical constraints into the CNN architecture, such as enforcing energy conservation principles or embedding fire spread dynamics through physics-informed neural network formulations, could enhance predictive consistency beyond purely data-driven approaches. A comprehensive quantitative benchmarking of runtime, memory footprint, and scalability relative to physics-based wildfire models is planned to assess operational feasibility. Finally, incorporating uncertainty quantification methods would enable more robust and reliable applications in operational wildfire forecasting.

In conclusion, this study highlights the potential of deep learning, particularly CNN-based models, to complement traditional numerical simulations like WRF-SFIRE. The proposed model offers a faster and computationally efficient alternative for predicting wildfire dynamics, supporting the development of real-time decision-making tools for wildfire management.

CRediT authorship contribution statement

Maryam Moradpour: Conceptualization, Methodology, Software, Formal Analysis, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Pankaj Kumar:** Methodology, Software, Writing – Review & Editing. **Gholam Ali Hoshyaripour:** Project Initiation, Conceptualization, Supervision, Project Administration, Writing – Review & Editing.

Funding

Part of this research has been funded by the project *PermaStrom*, Germany (grant no. 03EI4010B) within the seventh Energieforschungsprogramm of the German Federal Ministry of Economic Affairs and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz, BMWK). Open Access funding enabled and organized by Project DEAL. This article has been supported by the use of ChatGPT, an AI language model developed by OpenAI, for grammar and phrasing corrections. The authors have critically reviewed and ensured the accuracy of all content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bb1070.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.acags.2026.100355.

Data availability

The wildfire dataset used in this study, originally developed by Moisseeva and Stull (2020), is available through the link presented as https://www.frdr-dfdr.ca/repo/dataset/dd39c220-356d-4f06-a8e5-77016c648ca4. The model outputs including ground heat flux contours at each time step, GIF animations for each wildfire scenario, and the SAL analysis diagrams are available at https://radar.kit.edu/radar/en/dataset/32n2dzbcds3y29qx?token=YcWXazVwcukmZJRXpLuF. The Python scripts and Jupyter notebooks used to develop the CNN model and to generate all figures presented in this paper are available at https://radar.kit.edu/radar/en/dataset/nkhysv17ts8hw7r2?token=kDJOygZMKoxjMDhjEaSz.

References

Ackerman, A.S., Toon, O., Stevens, D., Heymsfield, A., Ramanathan, V., Welton, E., 2000. Reduction of tropical cloudiness by soot. Science 288 (5468), 1042–1047.

Aldersley, A., Murray, S.J., Cornell, S.E., 2011. Global and regional analysis of climate and human drivers of wildfire. Sci. Total Environ. 409 (18), 3472–3481.

Allaire, F., Mallet, V., Filippi, J.-B., 2021. Emulation of wildland fire spread simulation using deep learning. Neural Netw. 141, 184–198.

Andela, N., Kaiser, J., Van der Werf, G., Wooster, M., 2015. New fire diurnal cycle characterizations to improve fire radiative energy assessments made from MODIS observations. Atmos. Chem. Phys. 15 (15), 8831–8846.

Anderson, H.E., 1982. Aids to determining fuel models for estimating fire behavior. Bark Beetles, Fuels, Fire Bibliogr. 143.

Bruckert, J., Hoshyaripour, G.A., Horváth, Á., Muser, L.O., Prata, F.J., Hoose, C., Vogel, B., 2022. Online treatment of eruption dynamics improves the volcanic ash and SO_2 dispersion forecast: case of the 2019 Raikoke eruption. Atmos. Chem. Phys. 22 (5), 3535–3552. http://dx.doi.org/10.5194/acp-22-3535-2022.

Burge, J., Bonanni, M., Ihme, M., Hu, L., 2020. Convolutional LSTM neural networks for modeling wildland fire dynamics. arXiv preprint arXiv:2012.06679.

Carter, T.S., Heald, C.L., Jimenez, J.L., Campuzano-Jost, P., Kondo, Y., Moteki, N., Schwarz, J.P., Wiedinmyer, C., Darmenov, A.S., da Silva, A.M., et al., 2019. How emissions uncertainty influences the distribution and radiative impacts of smoke from fires in north america. Atmos. Chem. Phys. Discuss. 2019, 1–50.

Coen, J., Cruz, M., Rosales-Giron, D., Speer, K., Goodrick, S., 2022. Coupled fire–atmosphere model evaluation and challenges. In: Speer, K., Goodrick, S. (Eds.), Wildland Fire Dynamics. pp. 209–249.

Coen, J.L., Riggan, P.J., 2014. Simulation and thermal imaging of the 2006 esperanza wildfire in southern california: application of a coupled weather–wildland fire model. Int. J. Wildland Fire 23 (6), 755–770.

Coen, J.L., Schroeder, W., 2013. Use of spatially refined satellite remote sensing fire detection data to initialize and evaluate coupled weather-wildfire growth model simulations. Geophys. Res. Lett. 40 (20), 5536–5541.

Di Giuseppe, F., McNorton, J., Lombardi, A., Wetterhall, F., 2025. Global data-driven prediction of fire activity. Nat. Commun. 16 (1), 2918.

Eddin, M.H.S., Roscher, R., Gall, J., 2023. Location-aware adaptive normalization: A deep learning approach for wildfire danger forecasting. IEEE Trans. Geosci. Remote Sens. 61, 1–18.

Elshewey, A.M., Osman, A.M., Youssef, R.Y., El-Bakry, H.M., Hassan, S.A., 2025. Enhancing forest fires classification using a hybrid convolutional and BiLSTM deep learning model. Model. Earth Syst. Environ. 11 (5), 379.

Filonchyk, M., Peterson, M.P., Sun, D., 2022. Deterioration of air quality associated with the 2020 US wildfires. Sci. Total Environ. 826, 154103.

Gofa, F., Boucouvala, D., Louka, P., Flocas, H., 2018. Spatial verification approaches as a tool to evaluate the performance of high resolution precipitation forecasts. Atmos. Res. 208, 78–87.

Hernandez, C., Keribin, C., Drobinski, P., Turquety, S., 2015. Statistical modelling of wildfire size and intensity: a step toward meteorological forecasting of summer extreme fire risk. In: Annales Geophysicae, vol. 33, (12), Copernicus GmbH Göttingen, Germany, pp. 1495–1506.

Hodges, J.L., Lattimer, B.Y., 2019. Wildland fire spread modeling using convolutional neural networks. Fire Technol. 55, 2115–2142.

Jain, T., 2022. White pine plantations in moist forests: Do they make sense with increased wildfire potential? In: Science You Can Use (in 5 Minutes), July 2022. Fort Collins, CO: US Department of Agriculture, Forest Service, xRocky Mountain Research Station, 2 P..

Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D., 2020. A review of machine learning applications in wildfire science and management. Environ. Rev. 28 (4), 478–505.

Kaiser, J., Heil, A., Andreae, M., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M., Suttie, M., et al., 2012. Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. Biogeosciences 9 (1), 527–554.

Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53 (8), 5455–5516.

Küçükarslan, A.B., 2024. An overview of machine learning (ML) techniques applied to forest fire studies. Eurasian J. For. Sci. 12 (1), 1–9.

Kumar, P., Vogel, H., Bruckert, J., Muth, L.J., Hoshyaripour, G.A., 2024. Mieai: a neural network for calculating optical properties of internally mixed aerosol in atmospheric models. Npj Clim. Atmos. Sci. 7 (1), 110.

Lakshmanaswamy, P., Sundaram, A., Sudanthiran, T., 2024. Prioritizing the right to environment: Enhancing forest fire detection and prevention through satellite data and machine learning algorithms for early warning systems. Remote. Sens. Earth Syst. Sci. 7 (4), 472–485.

Li, X., Wang, X., Sun, S., Wang, Y., Li, S., Li, D., 2023. Predicting the wildland fire spread using a mixed-input CNN model with both channel and spatial attention mechanisms. Fire Technol. 59 (5), 2683–2717.

Liu, J.C., Mickley, L.J., Sulprizio, M.P., Dominici, F., Yue, X., Ebisu, K., Anderson, G.B., Khan, R.F., Bravo, M.A., Bell, M.L., 2016. Particulate air pollution from wildfires in the western US under climate change. Clim. Change 138 (3), 655–666.

Malloy, K., 2021. Spatiotemporal wildfire forecasting in california using recurrent neural networks.

Mambile, C., Kaijage, S., Leo, J., 2024. Application of deep learning in forest fire prediction: a systematic review. IEEE Access.

Marjani, M., Mesgari, M., 2023. The large-scale wildfire spread prediction using a multi-kernel convolutional neural network. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. 10, 483–488.

Moisseeva, N., Stull, R., 2020. WRF-SFIRE LES synthetic wildfire plume dataset, federated research data repository.

Moisseeva, N., Stull, R., 2021. Wildfire smoke-plume rise: a simple energy balance parameterization. Atmos. Chem. Phys. 21 (3), 1407–1425, Publisher: Copernicus GmbH. URL https://acp.copernicus.org/articles/21/1407/2021/acp-21-1407-2021.html.

Paugam, R., Wooster, M., Freitas, S., Val Martin, M., 2016. A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. Atmos. Chem. Phys. 16 (2), 907–925.

Peace, M., Greenslade, J., Ye, H., Kepert, J.D., 2022. Simulations of the waroona fire using the coupled atmosphere–fire model ACCESS-fire. J. South. Hemisph. Earth Syst. Sci. 72 (2), 126–138.

Peace, M., Mattner, T., Mills, G., Kepert, J., McCaw, L., 2015. Fire-modified meteorology in a coupled fire–atmosphere model. J. Appl. Meteorol. Clim. 54 (3), 704–720.

Peterson, D.A., Fromm, M.D., McRae, R.H., Campbell, J.R., Hyer, E.J., Taha, G., Camacho, C.P., Kablick III, G.P., Schmidt, C.C., DeLand, M.T., 2021. Australia's black summer pyrocumulonimbus super outbreak reveals potential for increasingly extreme stratospheric smoke events. NPJ Clim. Atmos. Sci. 4 (1), 38.

Prapas, I., Kondylatos, S., Papoutsis, I., Camps-Valls, G., Ronco, M., Fernández-Torres, M.-Á., Guillem, M.P., Carvalhais, N., 2021. Deep learning methods for daily wildfire danger forecasting. arXiv preprint arXiv:2111.02736.

Reddington, C., Butt, E., Ridley, D., Artaxo, P., Morgan, W., Coe, H., Spracklen, D., 2015. Air quality and human health improvements from reductions in deforestation-related fire in Brazil. Nat. Geosci. 8 (10), 768–771.

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. Proc. IEEE 109 (3), 247–278.

Santopaolo, A., Saif, S.S., Pietrabissa, A., Giuseppi, A., 2021. Forest fire risk prediction from satellite data with convolutional neural networks. In: 2021 29th Mediterranean Conference on Control and Automation. MED, IEEE, pp. 360–367.

Sengupta, A., Woodford, B.J., 2025. Recent advances in explainable machine learning models for wildfire prediction. Appl. Comput. Geosci. 100266.

Singh, H., Ang, L.-M., Srivastava, S.K., 2025. Benchmarking artificial neural networks and U-net convolutional architectures for wildfire susceptibility prediction: Innovations in geospatial intelligence. IEEE Trans. Geosci. Remote Sens..

Son, R., Ma, P.-L., Wang, H., Rasch, P.J., Wang, S.-Y., Kim, H., Jeong, J.-H., Lim, K.-S.S., Yoon, J.-H., 2022. Deep learning provides substantial improvements to county-level fire weather forecasting over the western United States. J. Adv. Model. Earth Syst. 14 (10), e2022MS002995.

Sullivan, A., 2007. A review of wildland fire spread modelling, 1990-present 2: Empirical and quasi-empirical models. arXiv preprint arXiv:0706.4128.

Sullivan, A.L., 2009. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. Int. J. Wildland Fire 18 (4), 369–386.

Taylor, S.W., Woolford, D.G., Dean, C., Martell, D.L., 2013. Wildfire prediction to inform fire management: statistical science challenges.

Walter, C., Freitas, S.R., Kottmeier, C., Kraut, I., Rieger, D., Vogel, H., Vogel, B., 2016. The importance of plume rise on the concentrations and atmospheric impacts of biomass burning aerosol. Atmos. Chem. Phys. 16 (14), 9201–9219.

Wernli, H., Hofmann, C., Zimmer, M., 2009. Spatial forecast verification methods intercomparison project: Application of the SAL technique. Weather. Forecast. 24 (6), 1472–1484.

Wernli, H., Paulat, M., Hagen, M., Frei, C., 2008. SAL—A novel quality measure for the verification of quantitative precipitation forecasts. Mon. Weather Rev. 136 (11), 4470–4487.

Wiedinmyer, C., Kimura, Y., McDonald-Buller, E.C., Emmons, L.K., Buchholz, R.R., Tang, W., Seto, K., Joseph, M.B., Barsanti, K.C., Carlton, A.G., et al., 2023. The fire inventory from NCAR version 2.5: an updated global fire emissions model for climate and chemistry applications. Geosci. Model. Dev. 16 (13), 3873–3891.

Xi, D.D., Taylor, S.W., Woolford, D.G., Dean, C., 2019. Statistical models of key components of wildfire risk. Annu. Rev. Stat. Appl. 6 (1), 197–222.

Zhang, Y., Mao, J., Ricciuto, D.M., Jin, M., Yu, Y., Shi, X., Wullschleger, S., Tang, R., Liu, J., 2023. Global fire modelling and control attributions based on the ensemble machine learning and satellite observations. Sci. Remote. Sens. 7, 100088.

Zhang, G., Wang, M., Liu, K., 2021. Deep neural networks for global wildfire susceptibility modelling. Ecol. Indic. 127, 107735.

Zhang, G., Wang, M., Liu, K., 2022. Dynamic prediction of global monthly burned area with hybrid deep neural networks. Ecol. Appl. 32 (5), e2610.

Zou, L., Shao, H., Liu, Y., Ren, C., Chen, Q., Bai, H., Huang, Z., Gu, Y., 2025. A study on the susceptibility of wildfire disasters in wuzhou city based on interpretability model. IEEE Access.