Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-time Fire Evolution Forecasting

Joe R. McNorton

joe.mcnorton@ecmwf.int

European Centre for Medium-Range Weather Forecasts

Francesca Giuseppe

European Centre for Medium-Range Weather Forecasts

Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11553556/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Real-time Fire Evolution Forecasting

Joe McNorton[1*], Francesca Di Giuseppe[1]

[1] European Centre for Medium-Range Weather Forecasts, Reading, RG2 9AX, UK

Correspondence to: Joe R. McNorton (joe.mcnorton@ecmwf.int)

Abstract. What an active wildfire does in the hours after detection determines suppression demands, risks to communities and emissions, yet wide-scale operational forecasting at this timescale remains limited. Mechanistic spread models need detailed inputs and substantial computing resources, while fire danger systems work at daily or longer timescales and do not describe individual fires. We present Sparky-Evolve, a probabilistic machine-learning model that forecasts the persistence and intensity of active fires within ~1 km grid cells at 15-minute resolution up to 6 h ahead. It is trained on 1.25 million fire events observed by the Flexible Combined Imager aboard Meteosat Third Generation (MTG), using fire radiative power (FRP) with weather, vegetation and human predictors. An XGBoost classifier estimates whether FRP remains detectable, a Temporal Fusion Transformer estimates FRP quantiles conditional on persistence, and the two are coupled in an autoregressive ensemble. On held-out events, the ensemble reduced conditional FRP error by 30 % relative to persistence, with well-calibrated uncertainty up to 2 h, and estimated when detectable activity would end with a mean absolute error of 39 min. The most intense fires were underestimated. For three 2026 fires in Angola, Spain and Ukraine, the ensemble captured the post-peak decline in FRP that persistence misses, reducing errors by 68–87 %, and kept survival probabilities well above climatology. Forecasts can be issued about 1 h after first detection and are inexpensive enough to run for every active fire in the MTG domain. Sparky-Evolve can therefore help identify which newly detected fires are likely to persist, supporting the prioritisation of suppression resources and near-real-time air-quality forecasting.

Introduction

Wildfires threaten lives, property and ecosystems, and are a major source of trace gases and aerosols that degrade air quality and harm human health [1–3]. Many of these impacts are determined within hours, depending on whether a fire persists, intensifies or dies down. This short-term behaviour shapes both the danger to communities and firefighters and the amount of pollutants released into the atmosphere [4,5]. Extreme wildfire events are becoming more frequent and intense in many regions under climate change [6–8], and Europe recorded its largest burned area of the satellite era in 2025 [9]. Forecasting how active fires evolve over the coming hours is therefore increasingly relevant to fire management and public health.

Current tools leave a gap at this timescale and fall broadly into two categories. Fire-behaviour models, including mechanistic fire-behaviour and coupled atmosphere–fire models, simulate fire spread in detail [10–15]. However, they need high-resolution

fuel, weather and topography data and substantial computing resources, which limits their routine use where many fires burn at once. They are therefore mainly used for scenario analysis of individual fires to support firefighting operations [4,11], rather than for comparing and prioritising fires across a region. Machine-learning models reduce this cost, either by emulating physics-based simulators [16,17] or by learning spread from observations [18,19]. However, they predict binary fire presence one day ahead, without intensity or calibrated uncertainty. A second category, fire-danger models, addresses the likelihood of fire rather than its behaviour. Fire-danger indices and machine-learning probability of fire models estimate the likelihood of fire occurrence, or the burned area, from days to months ahead[20–23]. They describe the conditions under which fires may start and spread, not how a detected fire will evolve. Fire managers deciding where to deploy suppression resources, issue warnings or plan evacuations therefore have no routine, domain-wide information on which active fires are likely to persist or intensify [4,5,19]. Emissions systems face a similar limitation. Satellite-based biomass-burning emission estimates commonly integrate fire radiative power (FRP) over time to obtain fire radiative energy (FRE), which is then converted into fuel consumption or emissions [24]. Because satellite observations sample fire activity intermittently, this integration requires assumptions about how fire activity evolves between observations, such as persistence or a prescribed diurnal cycle [25–28]. Beyond the latest observation, the same assumptions are used to extend emissions into the forecast period, even where observations are frequent. They misrepresent fires that intensify, decline or go out, introducing errors into FRE and emission estimates that propagate into air-quality forecasts [29].

Geostationary sensors can now resolve the sub-daily evolution of fire activity at km-scale [30]. This reduces the need for gap-filling assumptions and provides the observational record needed to forecast fire evolution directly. The Flexible Combined Imager (FCI) aboard Meteosat Third Generation (MTG) provides pixel-level FRP at approximately 1-km resolution every 10 minutes over Europe, Africa and western Asia [31], which provides sufficient data to train data-driven models. Forecasting FRP is not, however, a standard regression problem. An FRP time series combines two processes. One is the discrete transition from detection to non-detection, and the other is continuous variation in intensity while the fire remains active. A single regression model tends to blend the two, predicting small positive FRP where a fire has gone out and underestimating intensity where it persists. Uncertainty also grows quickly with lead time, so useful forecasts must be probabilistic or ensemble based.

Here we present Sparky-Evolve, a probabilistic machine-learning model that forecasts active-fire persistence and intensity at approximately 1-km and 15-minute resolution from MTG FRP observations. The model separates whether detectable fire activity persists from how intense the fire is, if it does, and couples the two in an autoregressive ensemble. It produces three products: the probability of continued fire activity, unconditional expected FRP for near-real-time emissions estimation, and conditional FRP for monitoring persistent fires and prioritising resources. We evaluate the model on held-out fire events and illustrate its behaviour with three case studies from 2026 in Angola, Spain and Ukraine. Because the model is computationally light, it can be run continuously across all active fires in the MTG domain shortly after satellite detection. Between May and July 2026, there was an average of about 1,500 active ~1 km grid cells per 15-minute timestep.

Results

Characteristics of Satellite Observations

We defined a fire event as a sequence of MTG detections in a ~1 km grid cell at 15-minute resolution with no more than one missing timestep, giving 1.25 million events from May to July 2026 (Methods). Fire events are concentrated in the savannas of southern-central Africa, consistent with their dominance of global fire activity [32,33]. Southern Africa accounts for about 95% of observed active timesteps across the four regions in Figure 1d (7.78 million), whereas Europe contributes comparatively few (0.18 million) (Figure 1a). Fire intensity follows a contrasting pattern. Median event peak FRP is generally lower across the African savannas and higher across the Mediterranean, eastern Europe and the Middle East (Figure 1b) because of the underlying vegetation composition. These regions therefore contribute disproportionately to total fire radiative energy (FRE) relative to their number of events (Figure 1c).

The FRP distribution is strongly skewed. The median event peak is 26 MW, the 99th percentile is 340 MW, and only a few events exceed 1,000 MW (Figure 1f). Event durations are similarly skewed, and most events end within a few hours (Figure 1e). The diurnal cycle of detections varies by region (Figure 1d). Southern African detections peak sharply in the mid-afternoon near the time of maximum daily temperature, around 15:00 local solar time, consistent with earlier geostationary studies [34]. Detections in northern Africa and Mediterranean Europe peak at a similar time but are spread over more hours caused by longer burn events, in part due to local vegetation and winds sustaining burning. Detections in northern Europe peak later, around 18:00, and continue into the evening. A single prescribed diurnal cycle, as often used to extrapolate fire activity from sparse polar-orbiting overpasses [35,36], may misrepresent sub-daily variability in some regions.

The evolution of fire activity after detection depends on event magnitude (Figure 1g). In all classes, mean FRP peaks within the first hour after first detection and then declines, although individual events evolve very differently [37]. Low-FRE events become almost entirely inactive within about 2 h, whereas high-FRE events decline more slowly, and about a third remain active 3 h after first detection. This highlights the importance of timeliness in the issuing of any real-time forecasts. Fire persistence and intensity therefore vary systematically between events, which motivates modelling them separately.

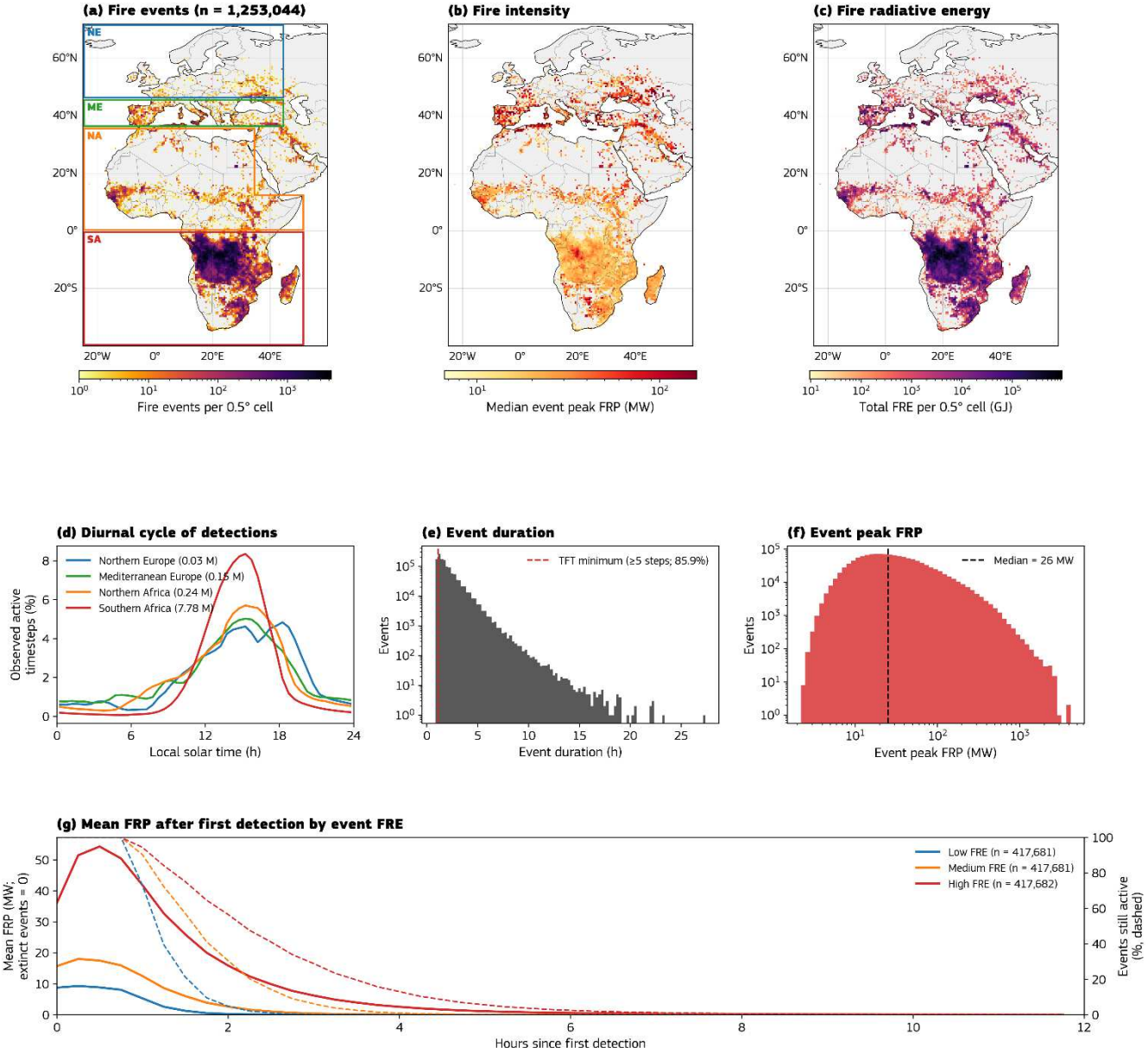


Figure 1 Overview of the MTG fire event dataset for May–July 2026. (a) Number of fire events, (b) median event peak fire radiative power (FRP) and (c) total fire radiative energy (FRE) aggregated per 0.5° cell. Coloured boxes in (a) outline the four regions used in (d): Northern Europe (NE; ≥46°N, 25°W–45°E), Mediterranean Europe (ME; 36–46°N, 25°W–45°E), Northern Africa (NA; 0–36°N, 25°W–52°E, excluding areas east of 35°E and north of 12°N) and Southern Africa (SA; south of the equator, 25°W–52°E). (d) Diurnal cycle of observed active detections in local solar time for each region, as a percentage of regional totals. (e) Distribution of event duration. The dashed red line marks the minimum of five active timesteps required by the Temporal Fusion Transformer (TFT), which 85.9% of events meet. (f) Distribution of event peak FRP. The dashed line marks the median. (g) Mean FRP against time since first detection for events grouped into terciles of total event FRE: low (≤61.7 GJ), medium (61.7–140.5 GJ) and high (>140.5 GJ). Shading shows the 95% confidence interval. Dashed lines (right axis) show the percentage of events in each class that are still active.

Forecasting System

Sparky-Evolve forecasts each active grid cell independently (Figure 2). A survival component, an XGBoost classifier, estimates the probability that FRP remains detectable at the next 15-minute timestep. An intensity component, a Temporal Fusion Transformer (TFT), predicts FRP quantiles if it does. Forecasts are initialised after four active timesteps, 1 h after first detection, and run as a 50-member autoregressive ensemble up to 6 h ahead, with the sample FRP of each member fed back as input to the next step. This gives three products: the survival fraction, conditional FRP from active members, and unconditional FRP which is probability-weighted.

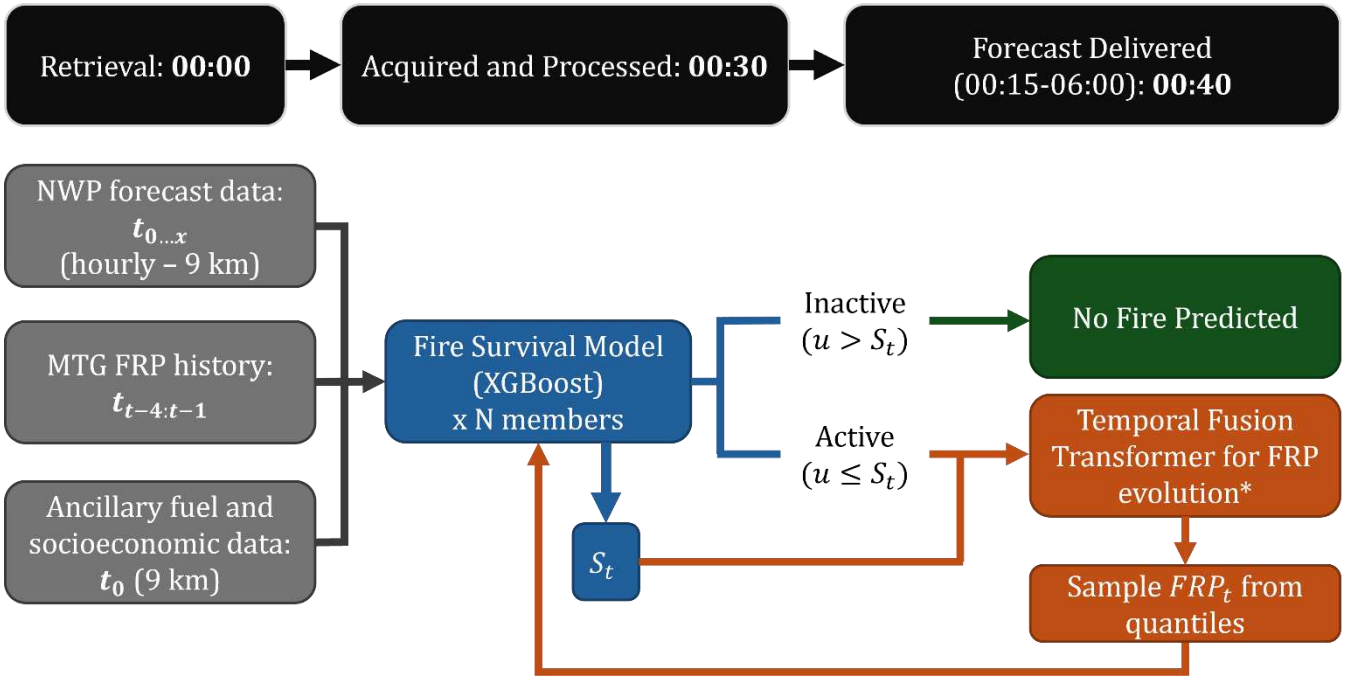


Figure 2 Schematic of the Sparky-Evolve system. Top: Theoretical operational timeline. Bottom: at each 15-minute timestep, each ensemble member remains active if $\boldsymbol{u \leq S_t}$, where $\boldsymbol{u}$ is a uniform random number and $\boldsymbol{S_t}$ the survival probability. For active members, FRP is sampled from TFT quantiles and fed back as input for the next timestep. Uses up to 12 preceding timesteps (3 h) of FRP history.

Survival and Extinction Timing

At a single 15-minute step, the survival component achieved a test AUC of 0.69 (validation 0.70). Discrimination is moderate because most fires persist from one step to the next, 83% of test samples remained active. Lagged FRP and its trend and acceleration terms accounted for about two-thirds of the total gain (Figure 3a). The remaining gain was spread across the environmental predictors, each contributing less than 3%. The most important were dewpoint and air temperature, followed by surface pressure, precipitation, wind gust, boundary-layer height and fuel loads. This is consistent with near-surface dryness, mixing and fuel availability influencing whether fires persist. Because class weighting lowers the raw probabilities, Platt

scaling was required. It reduced the test Brier score from 0.220 to 0.131 and the log loss from 0.631 to 0.424, without changing the ranking of predictions.

In the ensemble, the largest gains came within the first hour, when survival is most uncertain and most relevant operationally. Up to 1 h, the forecast Brier score was 0.204, compared with 0.212 for the climatological baseline (skill +3.8%). Skill was positive at every lead time in this window and peaked at +5.4% at 45 min (Figure 3c). The baseline already captures the typical decline in activity after detection, so this gain shows that the ensemble identifies which individual fires are likely to stop being detected. Beyond 2 h, the two forecasts converged, because most events had ended (observed active frequency 0.05 at 3 h).

A survival forecast is well calibrated if, among fires given a probability p of remaining active, a fraction close to p remains active. The ensemble met this standard for most forecasts (Figure 3b). At 15 min, p forecasts above 0.8 (60% of cases) differed from observed frequencies by less than 0.03. At 2–3 h, p forecasts below 0.2 (60–93% of cases) differed by less than 0.05. At intermediate probabilities (0.3–0.6), the ensemble overestimated survival by up to 0.13 from 1 h onwards.

Unlike the climatological baseline, the ensemble also estimates when each fire will end. Extinction time had a mean absolute error of 39 min and a bias of −10 min, and the ensemble 5–95% range contained the observed extinction for 95.6% of events (Figure 3d). Errors were asymmetric as for most events, extinction was predicted 15–30 min late, but a tail of early predictions for the longest-lived fires gave a slightly negative mean. Survival forecasts are therefore pulled towards typical event behaviour. Mean predicted duration (1.8 h) was close to the observed mean (1.9 h).

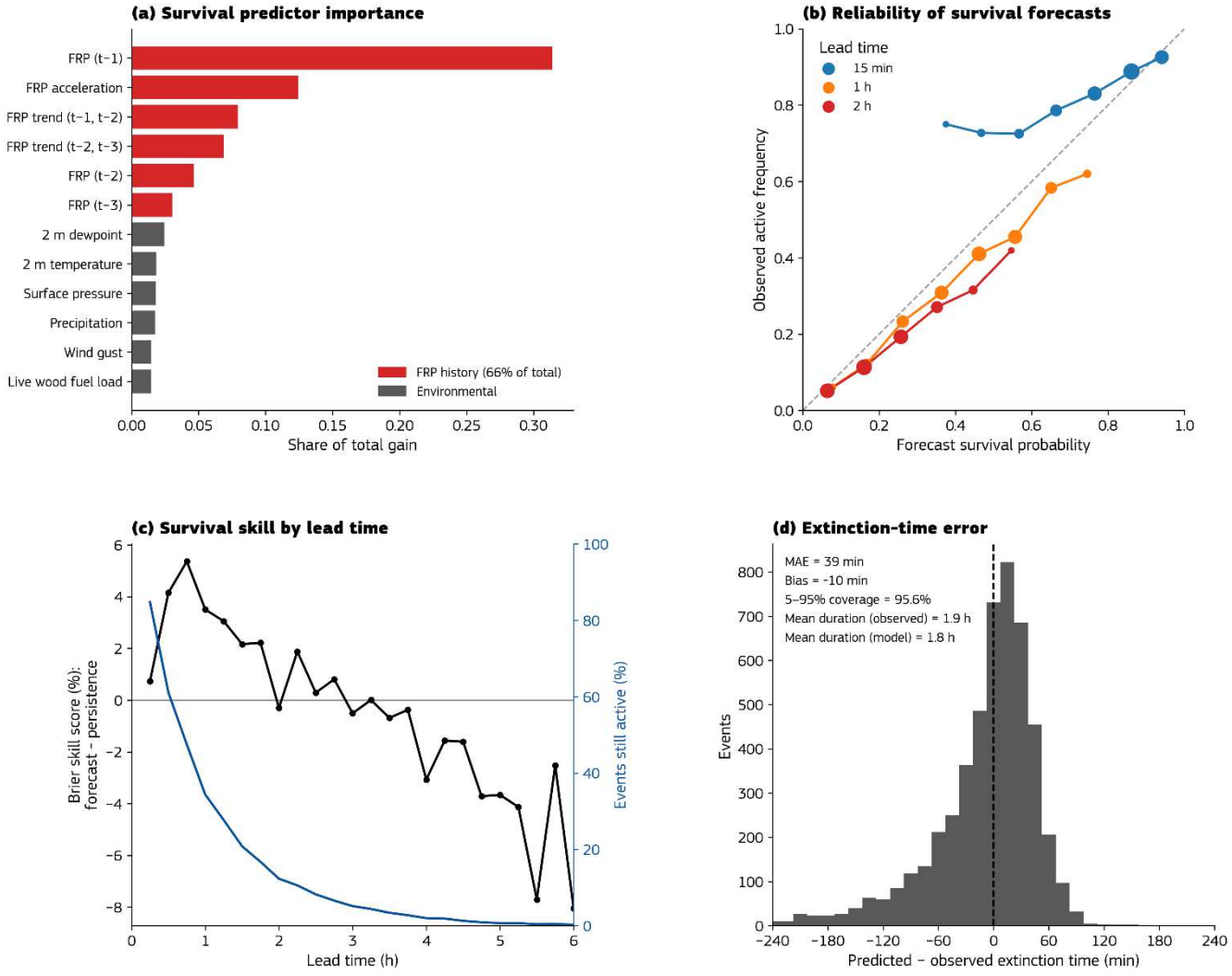


Figure 3 Evaluation of survival forecasts. (a) XGBoost gain for most important survival predictors. Red bars show FRP history terms, which account for 66% of total gain, and grey bars show other predictors. (b) Reliability of ensemble survival forecasts at lead times of 15 min, 1 h and 2 h. Each point shows the observed fraction of events still active against the mean forecast survival probability within a probability bin. Marker size indicates the number of forecasts in each bin, and the dashed line marks perfect reliability. (c) Brier skill score of the ensemble survival forecast relative to the climatological survival baseline, by lead time (black, left axis), with the percentage of events still active (blue, right axis). (d) Distribution of the ensemble-median predicted minus the observed extinction time. Negative values indicate extinction predicted too early.

Fire Intensity

For ongoing fire events, the median FRP of active ensemble members had a mean absolute error (MAE) of 12.7 MW, compared with 18.1 MW for persistence of the last observed FRP (skill +30%), and a mean bias of −4.2 MW. The ensemble outperformed persistence from the first lead time (+18% at 15 min), and skill increased to 40–48% between 1.75 and 2.25 h, as observed FRP moved away from its value at the forecast origin (Figure 4a). The 10–90% range of active members contained 77% of observations, close to the expected 80%. Coverage stayed between 75% and 83% up to 2 h, while the mean interval width

increased from 30 MW at 15 min to about 50 MW at 2 h (Figure 4b). Beyond 2.5 h, coverage fell below 70% and the bias became positive. Fewer than 100 active observations remained after 4 h, so scores at these lead times are not interpreted further. For 0.5% of active observations, no member was still active, so no conditional forecast was available.

When inactive members were counted as zero, the unconditional ensemble mean outperformed the baseline for the first 45 min (skill +5 to +9%). Over the full 6 h, it had a lower RMSE than the baseline (15.8 vs 16.6 MW) but a slightly higher MAE (4.5 vs 4.1 MW). Because the baseline combines the last observed FRP with the climatological survival curve, it already captures the typical decline in activity, and it decays towards zero more quickly. MAE therefore favours the baseline at longer lead times, when most observations are zero, whereas RMSE gives more weight to intense fires, where the ensemble performs better. The small positive bias (1.4 MW) reflects the overestimation of survival at intermediate probabilities. The 5–95% range contained 98% of observations, above the expected 90%, mainly because inactive observations fall inside any range that starts at zero.

Errors depended strongly on fire intensity (Figure 4c, d). For the lower half of ongoing fire observations (2.1–11.6 MW; median 6.9 MW), the ensemble overestimated FRP by 4.0 MW, and its median forecast was 1.4 times the observed value. Between the 50th and 90th percentiles (11.6–46.6 MW), the bias was small (−2.6 MW). Above the 90th percentile, FRP was increasingly underestimated, with a median forecast-to-observed ratio of 0.46 between the 90th and 95th percentiles (46.6–72.2 MW) and 0.39 between the 95th and 99th percentiles (72.2–166 MW). Above the 99th percentile of ongoing fire observations (166–1,870 MW; median 229 MW), the bias reached −217 MW and the ratio fell to 0.26. These 168 observations came from 101 events, with no single event dominating. The ensemble therefore pulls forecasts towards typical intensities, overestimating weak fires and substantially underestimating the most intense ones. This is consistent with the scarcity of extreme fires in the training data, despite tail oversampling (methods).

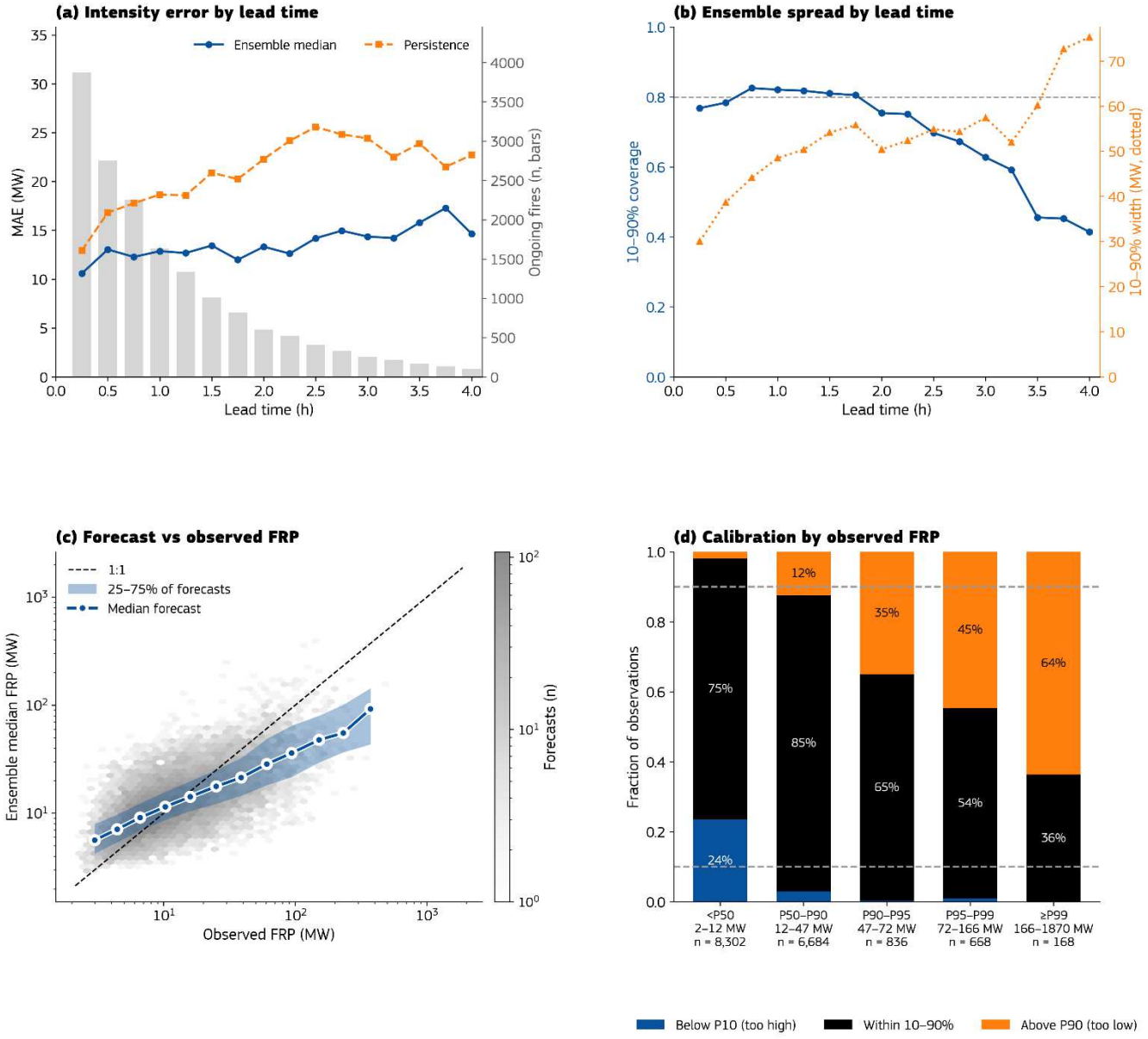


Figure 4 Evaluation of intensity forecasts. (a) Mean absolute error of the median FRP of active ensemble members (blue) and of persistence of the last observed FRP (orange, dashed), by lead time, relative to observations. Grey bars (right axis) show the number of ongoing fires. (b) Fraction of observations within the 10–90% range of active members (blue; the dashed line marks the expected 80%) and the mean width of that range (orange, dotted; right axis). (c) Ensemble median FRP against observed FRP. Grey shading shows the number of forecasts. The blue line and band show the median and 25–75% range of forecasts in log-spaced bins of observed FRP. (d) Fraction of observations below the ensemble 10th percentile (blue; forecast too high), within the 10–90% range (black) and above the 90th percentile (orange; forecast too low), by observed FRP band. Dashed grey lines mark the expected 10% and 90% boundaries.

Case Studies

Three test fires from contrasting environments illustrate how the ensemble forecasts individual events (Figure 5). Each case is the detection episode of a single MTG pixel within a larger fire. Cases were selected from intense, long-lived test fires that ended within the 6 h horizon.

The southern African case in Cuanza Sul, Angola on 24 July occurred during the dry season, when savanna and woodland fires are widespread across Angola [38]. (Figure 5a, b). Such fires dominate the dataset, and southern Africa accounts for 95% of observed active timesteps across the regions in Figure 1d. The pixel was first detected at 11:00 local solar time and became inactive at ~13:45, before the mid-afternoon peak in southern African detections (Figure 1d). Its peak FRP of 175 MW was well above the median event peak of 26 MW (Figure 1f). It remained active for 2.75 h, whereas low-FRE events are almost entirely inactive within 2 h (Figure 1g). FRP peaked at the forecast origin and fell to 35 MW within the next 45 min. The ensemble median captured this decline, the MAE was 80% lower than for persistence, and every observation fell within the 10–90% range. The ensemble survival probability declined more slowly than the climatological curve. At the observed extinction, 2 h after the forecast origin, it was 0.34, compared with 0.15 for the climatological curve. It had already fallen below 0.5 at 2.25 h, however, so the pixel stayed active ~30 min longer than the ensemble central estimate. This matches the tendency to predict the end of longer-lived fires too early.

The Iberian case in Guadalajara, Spain on 16 July 2026 was the most intense of the three (Figure 5c, d). It was part of the La Mierla wildfire, one of the largest recorded in Spain, which burned approximately 32,000 ha and led to the evacuation of numerous settlements [39]. The fire spread through a Mediterranean mountain landscape of forest, grassland and shrubland and affected several protected areas. For this pixel, FRP reached 1,200 MW at the forecast origin, above the 97th percentile for the region. It fell to 120 MW within 75 min and then stayed between 100 and 300 MW until the pixel became inactive, 4.25 h after first detection. The ensemble median declined faster than observed during the first 30 min, giving a short-lived underestimate (mean bias −93 MW). This is consistent with the underestimation of the most intense fires. Over the full episode, however, the ensemble reduced the MAE from 899 MW for persistence to 115 MW (skill +87%), and all observations fell within the 10–90% range. The ensemble survival probability stayed above 0.9 for the first 1.5 h after the forecast origin. At the observed extinction it was 0.40, whereas the climatological curve had fallen below 0.05. Extinction was predicted 15 min early.

The eastern European case in Kyiv Oblast, Ukraine, on 30 June 2026 occurred during a period of documented wildfire activity in the Chornobyl Exclusion Zone, where fires in radioactively contaminated forest prompted assessments of radionuclide emissions and atmospheric transport [40] (Figure 5e, f). FRP peaked at 239 MW at the forecast origin, halved within 15 min and then fluctuated between about 35 and 120 MW until extinction. The ensemble median followed this lower level and reduced

the MAE from 164 MW for persistence to 53 MW (skill +68%). 83% of observations fell within the 10–90% range. The ensemble predicted extinction at the observed time, when its survival probability was 0.48, compared with 0.15 for the climatological curve. As members became inactive, the 10–90% range widened substantially, and by 2.5 h few active members remained. Some members captured the final rise in observed FRP to 118 MW, but the ensemble median did not.

In all three cases, the ensemble captured the decline in FRP after the forecast origin, which persistence cannot represent, and kept survival probabilities well above the climatological curve for these long-lived fires. The cases also show the limitations found in the full evaluation suite. Survival probabilities at the observed extinction were only 0.34–0.48, so extinction timing remained uncertain, and for the most intense fire the decline in FRP was predicted too early. The ensemble is therefore most useful for indicating whether a fire is likely to weaken and roughly when it may end, rather than for predicting the exact intensity of the most intense fires.

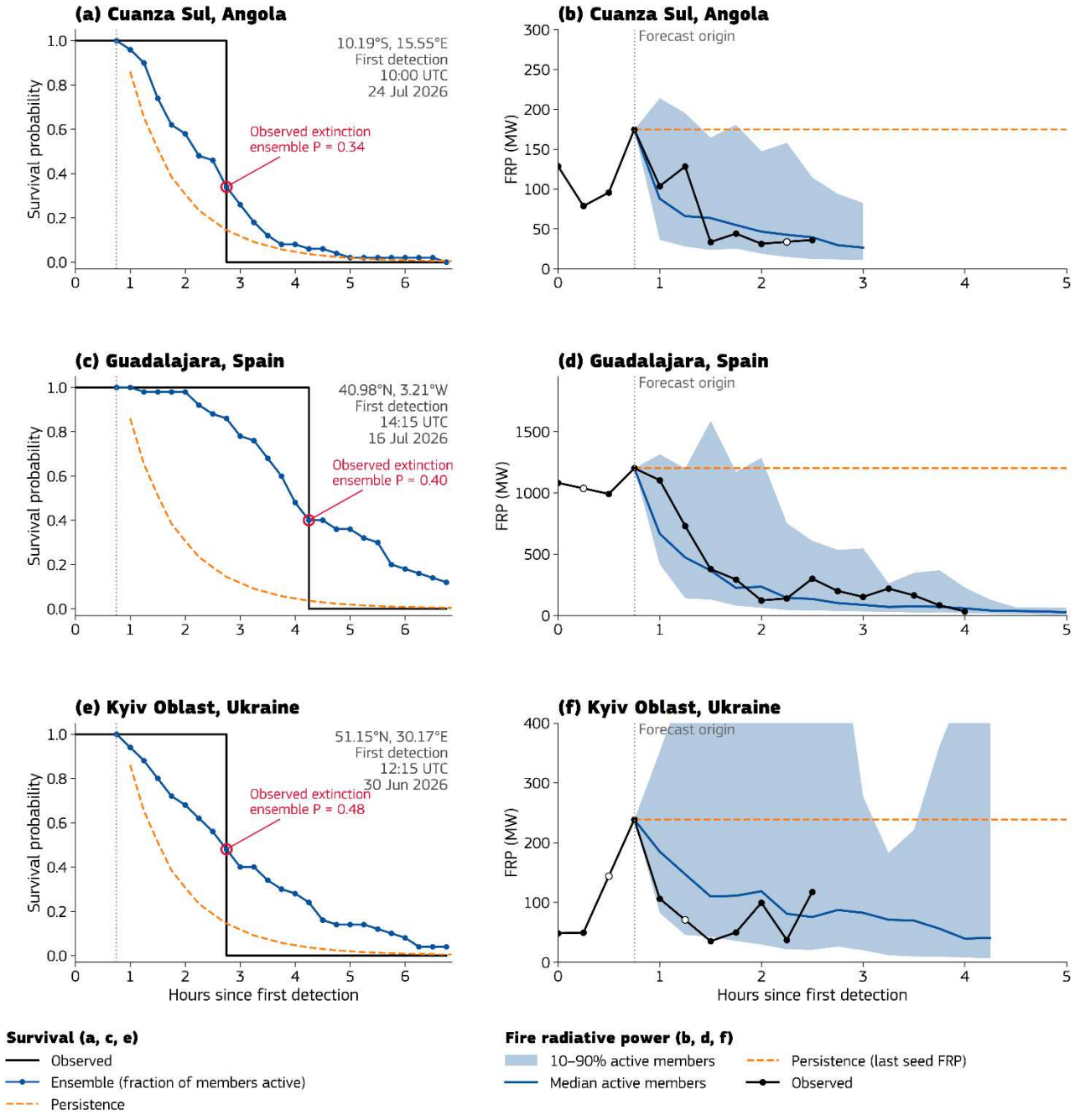


Figure 5 Ensemble forecasts for three test fire events. (a, b) Cuanza Sul, Angola; (c, d) Guadalajara, Spain; (e, f) Kyiv Oblast, Ukraine. Left panels show observed survival (black), the fraction of the 50 ensemble members still active (blue) and the climatological survival baseline (orange, dashed), which is the fraction of training events still active at each lead time. Red circles mark the ensemble survival probability at the first observed non-detection. Right panels show observed FRP (black), the median (blue line) and 10–90% range (shading) of active ensemble members, and persistence of the last observed FRP before the forecast origin (orange, dashed). Open circles mark interpolated observations, which are not scored. The first forecast timestep is 1 h after first detection.

Discussion

Sparky-Evolve combines geostationary FRP observations with meteorological, vegetation and socioeconomic predictors to produce probabilistic forecasts of sub-daily fire evolution. The key design choice is to model fire survival separately from intensity, treating extinction as a distinct outcome rather than a decline in intensity towards zero. This gives three complementary products. The survival fraction indicates whether a fire will persist and when it is likely to end. Conditional FRP indicates how intensely it will burn if it does persist. Unconditional FRP gives a probability-weighted estimate of fire activity for emissions. Together, these offer support for fire monitoring, response prioritisation and fire-emission forecasting.

Most of the forecast information comes from the recent state of the fire. Lagged FRP, trend and acceleration terms provided about two-thirds of the gain in the survival model, and no single environmental predictor contributed more than 3%, although gain is shared among correlated predictors. Over the next few hours, the fire trajectory is therefore the strongest guide to whether it persists, and weather, fuel availability and other variables adjust this signal. This does not imply that environmental conditions are unimportant. Fire weather and fuel moisture influence how individual fires grow and decline [7,37], so their effects are already partly reflected in the recent fire trajectory. The dominance of recent fire state also explains why skill is concentrated at short lead times. As forecast FRP replaces observed FRP in the autoregressive loop, the main source of information weakens. Strong fire–atmosphere feedback, such as that associated with pyrocumulonimbus development, may alter these relationships but were not assessed separately here.

This dependence on recent fire state shows the value of high-frequency observations. MTG observations show that the timing and spread of the diurnal cycle differ between southern Africa, the Mediterranean and northern Europe, in agreement with existing geostationary studies [30,34]. Fire evolution also depends on the fire itself. Low-FRE events become almost entirely inactive within about 2 h of detection, whereas about a third of high-FRE events remain active after 3 h (Figure 1g). How long a fire burns and how its intensity changes therefore depend on its initial size and intensity and on the fuel available to it, as well as on time of day and region. Neither persistence, which holds FRP fixed, nor a generalised prescribed diurnal cycle, as often used to extend sparse polar-orbiting observations [35,36], can represent this variation across fires and biomes. At sub-hourly resolution, MTG observes this evolution directly [31]. By learning from these observations together with fuel and meteorological predictors, Sparky-Evolve conditions each forecast on the recent trajectory of the fire and local environment instead of assuming a fixed temporal profile. The benefit of this appears in each of the three products.

The first product, the ensemble survival fraction, gives an event-specific forecast of the probability that a fire remains active. It is evaluated against a climatological survival curve averaged across all training events, which gives every fire the same forecast. The ensemble forecast improved on climatology at every lead time in the first hour (Brier skill up to +5.4%). This advantage is illustrated by the long-lived fires in the case studies, where the ensemble kept survival probabilities well above

climatology while the fires remained active. A useful probabilistic forecast should be reliable as well as skilful [41]. Survival forecasts were well calibrated at high and low probabilities, which account for most forecasts, but overestimated survival by up to 0.13 at intermediate probabilities. Unlike climatology, the ensemble also estimates when each fire will end. Extinction time had a mean absolute error of 39 min, and the ensemble 5–95% range contained the observed extinction for 95.6% of events. For most events extinction was predicted slightly late, but for the longest-lived fires it was predicted early, which pulls forecasts towards typical event behaviour.

The second product, unconditional FRP, averages over all members, including those that have become inactive, and so gives a probability-weighted estimate of near-term fire activity. It is not a realistic trajectory for any single fire. However, emissions systems such as GFAS [25] aggregate FRP over grid cells and time steps, so this unconditional FRP becomes a useful variable. In these systems, emissions are derived from fire radiative energy through its relationship with fuel consumption [24,42]. If Sparky-Evolve were integrated within emission systems such as the Global Fire Assimilation System (GFAS)[25], this consumption could be used to update the fuel available to each fire during an event, so that the forecast could account for fuel depletion as the fire burns. The unconditional product improved on the baseline for the first 45 min and reduced RMSE over 6 h, mainly through better forecasts for intense fires, which dominate emissions. It could therefore replace persistence or fixed diurnal-cycle assumptions when emissions are extended between observations or a few hours ahead, with potential benefits for smoke and air-quality forecasts during major fire episodes.

The third product, conditional FRP, describes how intense a fire could plausibly be if it persists. It showed the clearest skill, reducing error by 30% relative to persistence. The conditional FRP uncertainty was well calibrated, with the 10–90% range containing 75–83% of observations at lead times up to 2 h. This product is most relevant to monitoring ongoing fires, identifying those likely to weaken or persist, and prioritising suppression resources when capacity is limited [5].

The case studies show how these products can work together on individual fires in contrasting biomes. These studies include a savanna fire in Angola, the La Mierla wildfire in Spain and a fire in the Chornobyl Exclusion Zone. In all three cases, the ensemble captured the sharp decline in FRP after the forecast origin, which persistence cannot represent, and reduced the error by 68–87% relative to persistence. It also kept survival probabilities well above climatology for these long-lived fires and predicted extinction within 30 min, at the observed time in the Ukrainian case. The fires differed considerably in intensity (peak FRP of 175–1,200 MW) and duration (2.75–4.25 h), illustrating that the model can produce useful forecasts for events with differing behaviour.

The autoregressive structure allows forecasts to be reinitialised whenever new FRP observations or numerical weather prediction (NWP) fields become available, which limits the accumulation of uncertainty during an ongoing event. Since recent FRP carries most of the information, frequent reinitialisation is likely to matter more than extending the forecast horizon. The

15-minute model timestep does not imply 15-minute operational latency. Practical update frequency depends on satellite acquisition, Level-2 product generation, data delivery, preprocessing and the availability of predictor fields [31]. Because forecasts are initialised after four active timesteps, the first forecast for a fire could be available shortly after 1 h from first detection, with the delay set mainly by product latency. Within these constraints, the computational cost is small enough for the model to be run continuously across all active fires in the MTG domain, which is not feasible for mechanistic spread models [11,13,15].

Several limitations should be considered by future development. First, the most intense fires are systematically underestimated. Above the 99th percentile of ongoing fire observations, the median forecast was about a quarter of the observed value, and in the Guadalajara case the decline in FRP was predicted too early. This reflects the rarity of very high-FRP states in a three-month record dominated by African savanna fires [32]. Longer training periods, tail-weighted loss functions or additional training data from other geostationary sensors could address this. Second, performance varies by region. Among significant test fires, the ensemble beat persistence for 70% in southern Africa, 58% in Iberia and 49% in eastern Europe, and eastern Europe showed a gap between training and test performance. More European training data, or fine-tuning on European fires, may be needed before operational use over Europe. Third, treating extinction as an absorbing state excludes re-detection after cloud or smoke obscuration and rekindling, which would need a separate re-ignition component. Because survival is defined by detectability, the model also inherits the detection limits of the MTG product. These include omission of small or low-intensity fires and obscuration by cloud and thick smoke [43]. Fourth, each grid cell is modelled independently, so lateral spread, spotting and interactions between neighbouring cells are not represented. Spatially coupled extensions, for example graph-based or convolutional architectures, could link adjacent cells and forecast whole fires rather than individual pixels. Fifth, ensemble spread comes only from the survival and intensity components, driven by a single deterministic weather forecast. Ensemble NWP forcing could add meteorological uncertainty, particularly at longer lead times [44]. Finally, the model is limited to the MTG domain. Geostationary FRP products are also available over the Americas and, over East Asia and Australia. Together with MTG, these could support a near-global product. Sensors differ in pixel size, detection thresholds and retrieval algorithms, however, so the model would need to be retrained for each sensor or applied to harmonised FRP inputs. High latitudes, which geostationary satellites view poorly, would remain a gap.

Conclusion

Sparky-Evolve forecasts how active fires will evolve over the next few hours, a timescale that current operational tools largely miss. It separates whether a fire stays detectable from how intense it is and combines the two in an autoregressive ensemble. This gives calibrated, event-specific forecasts of persistence and intensity at approximately 1-km and 15-minute resolution. On held-out events, the system improved on climatological survival in the first hour and reduced conditional FRP errors by 30% relative to persistence.

The practical value of Sparky-Evolve lies in its timeliness. The framework is also straightforward to train and deploy, with limited preprocessing and modest computational demands. This makes retraining as new observations become available practical and allows forecasts to run continuously across the full domain. Because most of the forecast skill comes from the recent trajectory of the fire, the system depends on frequent observations, which geostationary sensors provide. MTG observes active fires across Europe, Africa and western Asia every 10 minutes, with short product latency. Forecasts can therefore start about 1 h after first detection, or sooner if fewer lagged timesteps are used, and be updated as each new observation arrives. Inference can run continuously for every active fire across the domain, which is not feasible with mechanistic spread models. Combining observations from other geostationary sensors could extend this to global coverage.

Two operational communities could benefit from the deployment of Sparky-Evolve. For air-quality forecasting, emissions could follow fires as they intensify, decline or go out, instead of assuming persistence or a fixed diurnal cycle. For fire management, forecasters would get routine, continental-scale information on which fires are likely to persist or stay intense in the hours that matter for allocating suppression resources and issuing warnings. Additional training data could improve the representation of the most extreme fires, and the system could be extended to include fire spread between neighbouring cells.

Methods

Active Fire Observations

Input data were derived from the MTG FRP-PIXEL product, generated from FCI observations aboard MTG-I1. The Level-2 product provides pixel-level active-fire detections, including acquisition time, location, FRP and detection confidence, at approximately 1-2 km resolution at the sub-satellite point and a 10-minute repeat cycle [31]. We used all detections across Europe, Africa and Western Asia from May to July of 2026, resulting in 22.9 million detections. Where a grid cell had multiple detections in the same timestep, their mean FRP was used. Cells with more than 300 detections over the analysis period were removed in full, to exclude persistent sources, such as gas flares (1.1 million). Data were aggregated to a regular 0.01° grid at 15-minute resolution.

Within each grid cell, detections separated by two or more missing timesteps were treated as distinct fire events, and single missing timesteps were linearly interpolated (11.9% of active timesteps). Events with fewer than four observed timesteps were excluded, since forecasts are initialised after four active timesteps. Four zero-FRP timesteps were then appended to each event to represent post-detection inactivity. Because events in the same cell can be separated by as few as two missing timesteps, the appended timesteps of 51,681 events (4.1%) overlapped renewed detections, so the survival target represents non-detection at the next timestep rather than definitive extinction. The final dataset contained 1,253,044 fire events and 14.9 million timesteps, of which 9.9 million were active (Figure 1). Event durations ranged from 4 to 109 active timesteps (1–27.25 h), with a median

of 7 timesteps (1.75 h). Of all events, 1,076,646 (85.9%) had at least five active timesteps, the minimum required for TFT training (methods).

Predictors

Each timestep was paired with predictors describing meteorology, fuel, vegetation, orography and human activity (Table S1). The predictor set is similar to those used in recent data-driven studies of fire danger [23] and climate scale burned area [45]. Each predictor was matched to the 0.01° fire grid by nearest neighbour.

Meteorological predictors were taken from short-range ECMWF Integrated Forecasting System (IFS) high-resolution forecasts at hourly resolution, using forecast steps 1–12 h from the 00 and 12 UTC runs. These were 2 m temperature, 2 m dewpoint temperature, skin temperature, 10 m wind speed, instantaneous 10 m wind gust, boundary-layer height, surface pressure, top-layer volumetric soil moisture and precipitation. Each 15-minute timestep was assigned the nearest hourly field. Training on forecasts rather than reanalysis means the model learns relationships under the same type of forcing it receives operationally, so it may partly learn to compensate for systematic forecast biases.

Fuel was represented using Sparky-Fuel [46]. Daily dead foliage, dead wood and live fuel moisture content were taken from the most recent field available at or before each timestep, so no future information was used. Monthly live leaf, live wood, dead foliage and dead wood fuel loads were taken a machine-learning fuel emulator [45], which is constrained by ESA-CCI above-ground biomass [47].

Vegetation was described by ECLand high- and low-vegetation types and cover fractions [48], together with monthly climatological leaf area index (LAI). Orography was represented by the standard deviation, slope, anisotropy and orientation of sub-grid orography from the ECMWF operational system. These describe terrain roughness and orientation, both of which can affect near-surface flow and fire behaviour. Human influence was represented by gridded population density [49] and road density from the Global Roads Inventory Project [50]. As in [45], these act as proxies for ignition pressure, accessibility and suppression capacity, not as direct causal controls.

Recent fire state was represented by FRP at up to four preceding timesteps, each with a flag indicating whether that lag was available within the event. Two first-order trend terms, $FRP_{t-1} - FRP_{t-2}$ and $FRP_{t-2} - FRP_{t-3}$, were also included, together with an acceleration term defined as the difference between the two trends. Lagged variables were calculated separately within each fire event, so FRP information was never transferred between unrelated fires. The same predictors were used in the autoregressive ensemble, with lagged FRP taken from the trajectory of each member.

Model Structure

Sparky-Evolve predicts fire evolution within each active grid cell. It does not simulate fire-front propagation or new ignitions. The model has two stages (Figure 2). The survival component estimates whether detectable FRP persists at the next timestep. The intensity component estimates the distribution of FRP conditional on that persistence. This separates the discrete process of detection or extinction from the continuous variation of intensity during active burning. Survival is defined as the persistence of detectable FRP in the MTG product. Non-detection is therefore not necessarily evidence that combustion has ceased, because it may also reflect cloud or smoke obscuration, or FRP below the detection threshold.

Twenty percent of events were held out for testing. The remainder were split into training (64%), calibration (20%) and validation (16%) partitions, which are 51.2%, 16.0% and 12.8% of all events. The calibration partition was used only to calibrate survival probabilities. The validation partition was used for TFT early stopping and for developing the ensemble system, including the selection of ρ. The test partition was used once, for the final evaluation. The survival component is an eXtreme Gradient Boosting (XGBoost) binary classifier [51] that estimates:

$$\boldsymbol{S_t = P(F_t = 1 \mid F_{t-1} = 1, x_t, p_{t-4:t-1})} \qquad \mathbf{(1)}$$

where $\boldsymbol{S_t}$ is the survival probability, $\boldsymbol{F_t}$ is the active fire state, $\boldsymbol{x_t}$ denotes the predictors at time t and $\boldsymbol{p_{t-4:t-1}}$ denotes the FRP lag of up to four timesteps, trend and acceleration terms, either observed or forecast. Training samples were all timesteps preceded by an active timestep, including the first appended zero-FRP timestep of each event, which provides the non-detection examples. Class imbalance was handled by reweighting the active class by the ratio of negative to positive training samples, which biases raw probabilities low. These were therefore recalibrated with Platt scaling [52], fitted on calibration fire events using only timesteps preceded by four active timesteps, which matches how forecasts are initialised.

The intensity component is a Temporal Fusion Transformer (TFT) [53] that predicts conditional quantiles of FRP given by:

$$\boldsymbol{Q_q(FRP_t \mid F_t = 1, x_{t-k:t}, p_{t-k:t-1}, S_t), 4 \leq k \leq 12} \qquad \mathbf{(2)}$$

where $\boldsymbol{Q_q}$ is the conditional quantile at probability level q, for q = 0.05, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.95, and k is the encoder length. Quantile regression was used instead of a single point forecast so that uncertainty in FRP could vary with fire state. The TFT was trained on active timesteps only, with a log-transformed FRP target (Figure 1f). Its inputs are the survival-component predictors plus $\boldsymbol{S_t}$. Because intense fire events are rare, training samples were oversampled according to the maximum FRP in their encoder history. Hyperparameters, oversampling weights and training details for both components are given in Table S2.

Autoregressive Ensemble Forecasting

Forecasts are initialised after four active timesteps, so the first forecast timestep is 1 h after first detection. Each forecast is a 50-member ensemble whose members share the observed starting sequence. At each subsequent 15-minute timestep, every active member recalculates its lagged FRP, trend and acceleration terms from its own trajectory, and the survival component gives a calibrated probability $\boldsymbol{S_t}$. The member stays active if $\boldsymbol{u \leq S_t}$, where u is a random number between 0 and 1. Otherwise it is assigned zero FRP from that timestep on.

For each active member, the TFT predicts conditional FRP quantiles from up to 12 preceding timesteps of observed and simulated history, the predictors and $\boldsymbol{S_t}$. One FRP value is then drawn by mapping a probability level onto the quantiles, back-transformed to MW, through piecewise-linear interpolation. Levels below 0.05 or above 0.95 are assigned the outermost quantiles. To represent persistently strong or weak fires, each member's probability level follows a latent first-order autoregressive process ($\rho = 0.5$; Supplementary Text), so a member that samples a high quantile tends to stay in the upper part of the distribution. On the validation partition, $\rho = 0.5$ improved interval coverage at 1–2 h relative to independent draws, without the long-lead bias of larger values (Table S3).

Sampled values are fed back as inputs to both components for 24 timesteps (6 h), progressively replacing observations with forecasts. Three products are derived at each lead time; the survival fraction, which is the proportion of members still active, the unconditional FRP from all members, with inactive members counted as zero, and conditional FRP from active members only.

Model Evaluation

The system was evaluated on a random sample of 5,000 of the ~250,000 eligible events in the test partition. Each event was scored at every lead time up to 6 h, excluding interpolated timesteps. The observed survival target is one until the first observed non-detection and zero afterwards. Events whose time series ended, or contained a gap, before 6 h were right-censored, so they were scored up to the censoring time and excluded from the extinction-timing metrics.

The survival fraction was assessed using the Brier score and reliability. The one-step probability $\boldsymbol{S_t}$, averaged over active members, was assessed separately at timesteps up to and including the first non-detection. Extinction timing was assessed by the error between the ensemble-median and observed first non-detection, and by the fraction of events whose observed time fell within the ensemble 5–95% range. Conditional FRP was scored only before the first non-detection. Unconditional FRP was scored against observed FRP up to that point and against zero thereafter. Both were assessed with MAE, RMSE, bias and interval coverage.

Skill, defined as 1 − (model score / reference score), is reported by lead time against a reference baseline. For survival, the baseline at lead k is the climatological fraction of training events whose first non-detection occurred after k. For conditional FRP, it is persistence of the last observed value before initialisation, and for unconditional FRP it is the product of the two. Beyond 3 h, fewer than 5% of events remain active, so FRP metrics at these lead times rest on few timesteps and are shown with their sample sizes.

Data Availability

The MTG FCI FRP-PIXEL product is distributed by the EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA SAF) and is available after free registration at https://datalsasaf.lsasvcs.ipma.pt/. ECMWF IFS high-resolution forecasts used for training are available from the ECMWF archive under ECMWF's data policy, and a subset of real-time IFS forecasts is available as ECMWF open data under a CC-BY-4.0 licence (https://doi.org/10.21957/open-data). Source data for all figures, together with the plotting scripts, are available on Zenodo – temporary link whilst in review (unpublished) (https://zenodo.org/records/23696660?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImJkNWJhODVkLTM1ZmMtNDY1My04ZDJmLTMyNTcxZmZiMjVmMCIsImRhdGEiOnt9LCJyYW5kb20iOiI5YTdiYWMyZjJhZDIxOTViNjFmM2JjYmY3ODBjMmQ4MiJ9.N_-oBl-pkjI8HLlQKIP-9zRnOaju9tAX_z9XmCX2IwOb0y3YBAGIbyeocgeD2WLMv-6FyidHdhGvmrwAj5cimg)

Acknowledgements

This work was funded by the Copernicus Emergency Management Service contract no. 942604 between the Joint Research Centre and ECMWF. JM is funded through The CoFlame European Union project, grant agreement no. 101296262. The authors recognise the efforts of all those involved in the MTG data production.

Author contributions

JM and FDG developed the concept for the study. JM developed the methodology, acquired the data, performed the analysis and drafted the manuscript. FDG reviewed manuscript text.

Competing interests

The authors declare that they have no conflict of interest.

Supplementary Materials

Table S1. Predictors used by the survival component (XGBoost) and intensity component (TFT) of Sparky-Evolve, grouped by type, with source, native spatial and temporal resolution. All predictors were gridded to ~1 km with 15 minute timesteps. The survival probability is an input to the TFT only.

Group	Variable	Source	Native resolution	Temporal resolution
Fire history	Lagged FRP	[31]	~1 km	10 min
Fire history	FRP lag-availability flags	Derived	~1 km	10 min
Fire history	FRP trends and acceleration	Derived	~1 km	10 min
Model	Survival probability	Survival model	~1 km	15 min
Meteorology	2 m temperature	IFS HRES	~9 km	Hourly
Meteorology	2 m dewpoint temperature	IFS HRES	~9 km	Hourly
Meteorology	Skin temperature	IFS HRES	~9 km	Hourly
Meteorology	10 m wind speed	IFS HRES	~9 km	Hourly
Meteorology	10 m wind gust	IFS HRES	~9 km	Hourly
Meteorology	Boundary-layer height	IFS HRES	~9 km	Hourly
Meteorology	Surface pressure	IFS HRES	~9 km	Hourly
Meteorology	Top-layer soil moisture	IFS HRES	~9 km	Hourly
Meteorology	Precipitation	IFS HRES	~9 km	Hourly
Fuel moisture	Dead foliage/wood moisture content	[46]	~9 km	Daily
Fuel moisture	Live fuel moisture content	[46]	~9 km	Daily
Fuel load	Live leaf/wood,	[46]	~9 km	Monthly
Fuel load	Dead foliage/wood	[46]	~9 km	Monthly
Vegetation	High- and low-vegetation type	[48]	~9 km	Static
Vegetation	High- and low-vegetation cover	[48]	~9 km	Static
Vegetation	LAI, high and low vegetation	[48]	~9 km	Monthly climatology
Orography	Standard deviation	[48]	~9 km	Static
Orography	Slope	[48]	~9 km	Static

Group	Variable	Source	Native resolution	Temporal resolution
Orography	Anisotropy	[48]	~9 km	Static
Orography	Angle	[48]	~9 km	Static
Human	Population density	[49]	~9 km	Annual
Human	Road density	[50]	~9 km	Static

Supplementary Text

Temporally correlated sampling of FRP

At each timestep, the FRP of each active member is drawn by mapping a probability level τ_t onto the seven quantiles predicted by the TFT. The quantiles are first back-transformed to MW, and τ_t is then mapped by piecewise-linear interpolation. Levels below 0.05 or above 0.95 are assigned the lowest or highest predicted quantile. To represent persistent intensity anomalies, τ_t follows a first-order autoregressive (AR(1)) process in standard-normal space:

$$z_t = \rho z_{t-1} + \sqrt{1-\rho^2}\,\varepsilon_t, \varepsilon_t \sim N(0,1), \tau_t = \Phi(z_t)$$

Here z_t is a latent anomaly for each member on a standard normal scale. Positive values place the member above the median predicted by the TFT, and negative values place it below. The first term, ρz_{t-1}, carries part of the previous anomaly forward, so ρ sets how strongly a member remembers whether it was running high or low. The second term adds new variability, ε_t is an independent standard normal draw at each timestep, scaled by $\sqrt{1-\rho^2}$. This scaling keeps the variance of z_t at one, so z_t remains standard normal at every timestep for any value of ρ, given that the initial value $z_0 \sim N(0,1)$ is drawn independently for each member. Φ, the standard normal cumulative distribution function, converts z_t into a probability level τ_t between 0 and 1. Because z_t is standard normal, τ_t is uniformly distributed. Each individual FRP draw therefore follows the distribution predicted by the TFT, except that the outer 5% at each end is clamped to the 0.05 and 0.95 quantiles.

Given the member-specific history, ρ does not change the spread of FRP at a single timestep, which comes entirely from the predicted quantiles. Instead, it controls how members evolve. The correlation between latent values k timesteps apart is ρ^k, so a member above its predicted median tends to stay above it for a few timesteps. This represents persistent anomalies, such as a fire that burns hotter than its predictors suggest. Because each draw feeds back as lagged FRP into both components, ρ also affects ensemble spread and bias at longer lead times. With ρ = 0, draws are independent and trajectories fluctuate between quantiles at each timestep. As ρ approaches 1, each member stays near the same quantile level throughout the forecast.

ρ was selected from 0, 0.5, 0.8, 0.9 and 0.95 using 1,000 validation events (Table S3). With independent draws, the 80% conditional interval was too narrow, with coverage of 0.72 at 1 h and 0.58 at 2 h. With ρ = 0.5, coverage was 0.82 and 0.76. Larger values gave over-coverage at 1 h (0.86–0.89) and an increasing high bias in conditional FRP at 3 h: +17, +35 and +80 MW for ρ = 0.8, 0.9 and 0.95, compared with +3 MW for ρ = 0.5. This bias arises because members in the upper quantiles both stay there and are more likely to survive, so the remaining active members are increasingly the high ones. We therefore selected ρ = 0.5.

Table S2. Configuration of the survival component (XGBoost), intensity component (TFT) and autoregressive ensemble. n_{neg} and n_{pos} are the numbers of non-detection and active training samples. Because active samples dominate, the active-class weight is below 1. S_t is the Platt-calibrated survival probability. Oversampling percentiles refer to the maximum FRP in each training events encoder history. The selected checkpoint had the lowest validation loss before early stopping. ρ is the lag-one correlation of the AR(1) process used to sample quantile levels (Supplementary Text, Table S3).

Component	Setting	Value
Survival Component (XGBoost)	Objective	Binary logistic
	Boosting iterations	800
	Max depth / learning rate	6 / 0.03
	Row / predictor subsampling	0.8 / 0.8
	Min child weight / min split loss / L2	5 / 0.1 / 1.0
	Class weighting	Active class weighted by n_{neg} / n_{pos} (< 1)
	Calibration	Platt scaling on calibration events, using timesteps preceded by four active timesteps
Intensity Component (TFT)	Quantiles	0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95
	Encoder length	4–12 timesteps
	Hidden size / attention heads / hidden continuous size	64 / 4 / 32
	Dropout / batch size / gradient clip	0.1 / 128 / 0.1
	Optimiser	Adam, learning rate 0.001
	S_t input during training	Out-of-fold predictions from 5-fold cross-validation, grouped by event
	Oversampling weights (percentile of encoder maximum FRP)	<90th: 1.0; 90–95th: 1.5; 95–99th: 2.5; ≥99th: 4.0 (training only)

Component	Setting	Value
	Training	≤50 epochs (31,986 batches each), early stopping on validation quantile loss (patience 5)
	Selected checkpoint	Epoch 3 of 8 (validation loss 0.170)
Ensemble	Members / horizon / ρ	50 / 24 steps (6 h) / 0.5
	FRP sampling	AR(1) quantile levels, interpolated in MW (Supplementary Text)

Table S3. Coverage of the 80% conditional FRP interval (nominal 0.80) and conditional FRP bias (MW) at 1, 2 and 3 h for each AR(1) correlation ρ, evaluated on 1,000 validation events. The asterisk marks the selected value (ρ = 0.5).

ρ	Cov 1 h	Cov 2 h	Cov 3 h	Bias 1 h (MW)	Bias 2 h (MW)	Bias 3 h (MW)
0	0.72	0.58	0.42	−4.2	−7.9	−4.4
0.5*	0.82	0.76	0.60	−4.7	−0.5	+3.4
0.8	0.86	0.78	0.60	−1.3	+4.3	+17.4
0.9	0.87	0.80	0.50	−1.5	+10.9	+35.3
0.95	0.89	0.77	0.50	−0.5	+16.6	+80.0

References

1. Bowman, D. M. J. S. *et al.* Fire in the Earth System. *Science* **324**, 481–484 (2009).
2. Johnston, F. H. *et al.* Unprecedented health costs of smoke-related PM2.5 from the 2019–20 Australian megafires. *Nat Sustain* **4**, 42–47 (2020).
3. Moritz, M. A. *et al.* Learning to coexist with wildfire. *Nature* **515**, 58–66 (2014).
4. Noonan-Wright, E. K. *et al.* Developing the US Wildland Fire Decision Support System. *Journal of Combustion* **2011**, 168473 (2011).
5. Thompson, M. P. & Calkin, D. E. Uncertainty and risk in wildland fire management: A review. *Journal of Environmental Management* **92**, 1895–1909 (2011).
6. Cunningham, C. X., Williamson, G. J. & Bowman, D. M. J. S. Increasing frequency and intensity of the most extreme wildfires on Earth. *Nat Ecol Evol* **8**, 1420–1425 (2024).

7. Jolly, W. M. *et al.* Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat Commun* **6**, 7537 (2015).
8. Jones, M. W. *et al.* Global and Regional Trends and Drivers of Fire Under Climate Change. *Reviews of Geophysics* **60**, e2020RG000726 (2022).
9. Burton, C. *et al.* State of Wildfires 2025–2026. Preprint at https://doi.org/10.5194/essd-2026-587 (2026).
10. Rothermel, R. C. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*. vol. INT-115 40 https://www.fs.usda.gov/rm/pubs_int/int_rp115.pdf (1972).
11. Finney, M. A. *FARSITE: Fire Area Simulator-Model Development and Evaluation*. RMRS-RP-4 https://research.fs.usda.gov/treesearch/4617 (1998) doi:10.2737/RMRS-RP-4.
12. Tymstra, C., Bryce, R. W., Wotton, B. M., Taylor, S. W. & Armitage, O. B. *Development and Structure of Prometheus: The Canadian Wildland Fire Growth Simulation Model*. vol. NOR-X-417 https://publications.gc.ca/collections/collection_2010/nrcan/Fo133-1-417-eng.pdf (2010).
13. Mandel, J., Beezley, J. D. & Kochanski, A. K. Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geosci. Model Dev.* **4**, 591–610 (2011).
14. Coen, J. L. *et al.* WRF-Fire: Coupled Weather–Wildland Fire Modeling with the Weather Research and Forecasting Model. *Journal of Applied Meteorology and Climatology* **52**, 16–38 (2013).
15. Linn, R. R. *et al.* QUIC-fire: A fast-running simulation tool for prescribed fire planning. *Environmental Modelling & Software* **125**, 104616 (2020).
16. Hodges, J. L. & Lattimer, B. Y. Wildland Fire Spread Modeling Using Convolutional Neural Networks. *Fire Technol* **55**, 2115–2142 (2019).
17. Allaire, F., Mallet, V. & Filippi, J.-B. Emulation of wildland fire spread simulation using deep learning. *Neural Networks* **141**, 184–198 (2021).
18. Radke, D., Hessler, A. & Ellsworth, D. FireCast: Leveraging Deep Learning to Predict Wildfire Spread. in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* 4575–4581 (International Joint Conferences on Artificial Intelligence Organization, Macao, China, 2019). doi:10.24963/ijcai.2019/636.

19. Huot, F. *et al.* Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading From Remote-Sensing Data. *IEEE Trans. Geosci. Remote Sensing* **60**, 1–13 (2022).

20. Van Wagner, C. E. *Development and Structure of the Canadian Forest Fire Weather Index System*. vol. 35 37 https://cfs.nrcan.gc.ca/pubwarehouse/pdfs/19927.pdf (1987).

21. Giuseppe, F. D. *et al.* The Potential Predictability of Fire Danger Provided by Numerical Weather Prediction. *Journal of Applied Meteorology and Climatology* **55**, 2469–2491 (2016).

22. Jain, P. *et al.* A review of machine learning applications in wildfire science and management. *Environ. Rev.* **28**, 478–505 (2020).

23. McNorton, J. R., Di Giuseppe, F., Pinnington, E., Chantry, M. & Barnard, C. A Global Probability-Of-Fire (PoF) Forecast. *Geophysical Research Letters* **51**, e2023GL107929 (2024).

24. Wooster, M. J., Roberts, G., Perry, G. L. W. & Kaufman, Y. J. Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *J. Geophys. Res.* **110**, 2005JD006318 (2005).

25. Kaiser, J. W. *et al.* Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences* **9**, 527–554 (2012).

26. Ichoku, C. & Ellison, L. Global top-down smoke-aerosol emissions estimation using satellite fire radiative power measurements. *Atmos. Chem. Phys.* **14**, 6643–6667 (2014).

27. Darmenov, A. S. & da Silva, A. M. *The Quick Fire Emissions Dataset (QFED): Documentation of Versions 2.1, 2.2 and 2.4*. https://ntrs.nasa.gov/citations/20180005253 (2015).

28. Li, F. *et al.* Hourly biomass burning emissions product from blended geostationary and polar-orbiting satellites for air quality forecasting applications. *Remote Sensing of Environment* **281**, 113237 (2022).

29. Di Giuseppe, F., Rémy, S., Pappenberger, F. & Wetterhall, F. Using the Fire Weather Index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations. *Atmos. Chem. Phys.* **18**, 5359–5370 (2018).

30. Roberts, G. J. & Wooster, M. J. Fire Detection and Fire Characterization Over Africa Using Meteosat SEVIRI. *IEEE Transactions on Geoscience and Remote Sensing* **46**, 1200–1218 (2008).

31. Xu, W. *et al.* Major improvements in spaceborne early fire detection and small-fire FRP retrieval with the meteosat third generation flexible combined imager. *Science of Remote Sensing* **13**, 100366 (2026).

32. Giglio, L., Randerson, J. T. & Van Der Werf, G. R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *JGR Biogeosciences* **118**, 317–328 (2013).

33. Van Der Werf, G. R. *et al.* Global fire emissions estimates during 1997–2016. *Earth Syst. Sci. Data* **9**, 697–720 (2017).

34. Roberts, G., Wooster, M. J. & Lagoudakis, E. Annual and diurnal african biomass burning temporal dynamics. *Biogeosciences* **6**, 849–866 (2009).

35. Giglio, L. Characterization of the tropical diurnal fire cycle using VIRS and MODIS observations. *Remote Sensing of Environment* **108**, 407–421 (2007).

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

37. Andela, N. *et al.* The Global Fire Atlas of individual fire size, duration, speed and direction. *Earth Syst. Sci. Data* **11**, 529–552 (2019).

38. van Wilgen, B. W., de Klerk, H. M., Stellmes, M. & Archibald, S. An analysis of the recent fire regimes in the Angolan catchment of the Okavango Delta, Central Africa. *Fire Ecology* **18**, 13 (2022).

39. MITECO. El MITECO habilita un informe interactivo con los datos de los incendios forestales. https://www.miteco.gob.es/content/dam/miteco/es/prensa/septiembre-2026/26_09_09_Ndp_El_MITECO_habilita_un_informe_interactivo_con_los_datos.pdf (2026).

40. NASU. Scientists of the Institute for Safety Problems of Nuclear Power Plants of the NAS of Ukraine assessed the consequences of wildland fires in the Chernobyl Exclusion Zone in June–July 2026. https://www.nas.gov.ua/en/news/naukovci-institutu-problem-bezpeki-atomnih-elektrostanciy-nan-ukrani-ocinili-naslidki-prirodnih-pozhezh-u-zoni-vidchuzhennya-chornobilsko-aes-u-chervni--lipni-2026-roku (2026).

41. Gneiting, T., Balabdaoui, F. & Raftery, A. E. Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **69**, 243–268 (2007).

42. Freeborn, P. H. *et al.* Relationships between energy release, fuel mass loss, and trace gas and aerosol emissions during laboratory biomass fires. *J. Geophys. Res.* **113**, 2007JD008679 (2008).

43. Roberts, G. *et al.* LSA SAF Meteosat FRP products – Part 2: Evaluation and demonstration for use in the Copernicus Atmosphere Monitoring Service (CAMS). *Atmos. Chem. Phys.* **15**, 13241–13267 (2015).

44. Leutbecher, M. & Palmer, T. N. Ensemble forecasting. *Journal of Computational Physics* **227**, 3515–3539 (2008).

45. McNorton, J., Comyn-Platt, E., Di Giuseppe, F., Cagnazzo, C. & Barnard, C. Global fuel and burned area reconstructions and projections for 1950–2100. Preprint at https://doi.org/10.5194/egusphere-2026-5374 (2026).

46. McNorton, J. R. & Di Giuseppe, F. A global fuel characteristic model and dataset for wildfire prediction. *Biogeosciences* **21**, 279–300 (2024).

47. Santoro, M. *et al.* Design and performance of the Climate Change Initiative Biomass global retrieval algorithm. *Science of Remote Sensing* **10**, 100169 (2024).

48. Boussetta, S. *et al.* ECLand: The ECMWF Land Surface Modelling System. *Atmosphere* **12**, 723 (2021).

49. Zhuang, H. *et al.* Mapping high-resolution global gridded population distribution from 1870 to 2100. *Science of The Total Environment* **955**, 176867 (2024).

50. Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* **13**, 064006 (2018).

51. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2939672.2939785.

52. Platt, J. C. Probabilities for SV Machines. in *Advances in Large-Margin Classifiers* (eds Smola, A. J., Bartlett, P., Schölkopf, B. & Schuurmans, D.) 61–74 (MIT Press, Cambridge, MA, 2000). doi:10.7551/mitpress/1113.003.0008.

53. Lim, B., Arik, S. Ö., Loeff, N. & Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **37**, 1748–1764 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SparkyEvolveMcNortonSupplementaryMaterials.docx