Solar activity Impact on Firefighter Interventions: Factors Analysis

Naoufal Sirri*[1] [0000-0002-9900-337X] and Christophe Guyeux[1] [0000-0003-0195-4378]

FEMTO-ST Institute, UMR 6174 CNRS, University of Franche-Comté, Belfort, France, {naoufal.sirri,christophe.guyeux}@univ-fcomte.fr

Abstract. This research is a natural continuation of studies exploring various categories of variables and their implications for firefighters. The motivation behind this study stems from the acknowledgment that solar activity, although it may affect health and the environment, typically is not an immediate priority for firefighting interventions. Firefighters are primarily summoned for emergency situations, prompting contemplation on how solar activity, despite its environmental impact, might exert influence on their operations. In this context, our investigation aims to comprehensively evaluate how solar activity influences firefighting interventions, with a specific focus on intervention frequency. The study's relevance is underscored by the significant impacts of solar risks on both health and the environment. Notably, the summer 2023 heatwave assessment in France recorded a number of heat-related deaths, reaching nearly 1500 during the four heatwave episodes and exceeding 5000 over the entire summer period, as documented by the Ministry of Ecology, emphasizes the crucial nature of our research. Over an eight-year period, from 2015 to 2023, our methodology encompasses data preparation, in-depth analysis, and the application of the XGBoost predictive model, known for its resilience to outliers. The iterative training pipeline selects features that enhance the RMSE score over a 24-hour horizon, highlighting the crucial importance of variables related to solar activity, especially over extended periods. The key conclusion drawn from this study is these variables exert a progressive impact on interventions, suggesting increased relevance in predicting outcomes over prolonged durations. This precision in understanding the models associated with the presence or absence of solar risk offers a practical approach to anticipate resource management, improve firefighter response times, and contribute to saving lives by addressing intervention failures during major incidents. This study initiates a comprehensive exploration of variable families to understand the factors influencing firefighting activities.

Keywords: Firefighters intervention · Feature selection · Solar activity impact · Intervention causes.

1 Introduction

Solar activity significantly influences the frequency and nature of firefighter interventions, especially in critical domains like public health and the environment. Exposure to ultraviolet (UV) radiation has detrimental health effects, leading to an upsurge in firefighter interventions. The Ministry of Ecology underscores that 75% of individuals experience sunburn during their initial sun exposure, and only 40% apply sunscreen every two hours [1]. In the summer of 2023 in France, the number of heat-related deaths amounted to nearly 1500 during the four heatwave episodes and exceeded 5000 over the entire summer period. Emergency healthcare utilization related to heat reached almost 20,000, showing a significant increase during heatwave episodes and highlighting the imperative for preventive measures and climate change adaptation [2]. However, leveraging solar activity monitoring data for specific event detection proves intricate. The intricacies of spatial and temporal variability, coupled with the establishment of detection thresholds, necessitate thorough analysis. This study aims to investigate the potential correlation between solar activity and firefighting interventions, thereby contributing to the formulation of effective monitoring strategies.

Previous research has made notable strides in predicting solar risk and firefighting interventions, leveraging advancements in artificial intelligence. [3] introduced the artificial neural network (ANN) method for solar energy forecasting based on historical power measurements and meteorological forecasts. This 24-hour horizon approach enhances precision for operational planning in transportation networks and photovoltaic systems within electricity markets. On the other hand, [4] investigated models for forecasting solar radiation in Turkey based on essential meteorological variables. They favored the artificial neural network (ANN) due to its superior accuracy compared to adaptive network fuzzy inference system (ANFIS) models, multiple linear regression (MLR), and traditional empirical equations. In contrast, [5] proposed a fuzzy inference system based on brain emotional learning (BELFIS), showing competitive performance in forecasting solar cycles 16 to 24 compared to other models, including the adaptive neuro-fuzzy inference system. Additionally, [6] addressed thirteen artificial intelligence models to estimate solar radiation in Kurdistan, with the two-step deep extreme learning machine (DELM) model and its variations combined with variational mode decomposition (VMD) standing out for significantly improved accuracy, reflected in an average reduction of RMSE ranging from 13.3% to 48.25% for Darbandikhan and from 5.22% to 40.33% for Dukan. Nevertheless, few studies have explored the prediction of firefighting interventions in the context of firefighting operations.[7] used long short-term memory to predict the number of firefighter deployments in France. Similarly, [8] compared XGBoost, Gradient Boosting, and AdaBoost, identifying XGBoost as the optimal predictive model. Finally, [9] demonstrated the feasibility of a continuously updated and optimized database using dedicated feature selection tools for XGBoost.

Therefore, our research team actively engages in this endeavor by delving into an innovative perspective: the examination of solar activity and its repercussions on firefighting interventions and assistance to individuals. This exploration is facilitated through comprehensive historical datasets, teeming with information regarding both solar activity and firefighting interventions. The focal inquiry of this article revolves around the extent to which solar activity impacts firefighting interventions, with a specific focus on intervention frequency. Within our research, we have crafted a hypothesis as well as its prediction to address this inquiry:

- Hypothesis: Higher levels of solar activity variables will show a correlation with a sustained rise in firefighting interventions.
  Prediction: The existence of solar flux on Earth is anticipated to influence health, weather condition, and fire risk, leading to interventions over an extended timeframe.

To conduct our research systematically, we meticulously devised a methodological plan, adhering rigorously to its details as outlined below. Section 2 provides a comprehensive overview of the methods and materials employed in the experiments, outlining our experimental approach. The research process unfolds in Section 3, where the attained results are presented with careful consideration. Moving to Section 4, the discussion chapter conducts a thorough analysis to address the initial research question. This section offers a critical reflection on the results, emphasizing relevant implications and interpretations. Finally, Section 5 concludes our study by synthesizing the key findings, highlighting significant contributions, and proposing potential avenues for future research. This methodological framework ensured a systematic exploration of the correlation between solar activity and firefighting interventions, maintaining a comprehensive and organized approach throughout our investigation.

2 Methods

2.1 Data Preparation

Data Acquisition This research relies on an extensive dataset provided by the Fire and Rescue Department of Doubs (SDIS 25), France, comprising 320,847 interventions recorded from January 1, 2015, 00:00:00, to December 30, 2023, 12:00:00. Each individual intervention has been fastidiously documented, including identification code, start and end dates and times, location, intervention type, as well as response and duration times. The analysis of temporal trends involved extracting specific dates and times for each incident. Several parameters were considered to comprehend the circumstances of incidents, such as weather conditions, water levels of major rivers in Doubs (French firefighters are responsible for a variety of rescues, including flooding events), epidemiological data, academic periods, holidays, sports events, lunar phases, and more. To predict the number of future interventions, a dictionary was initiated, incorporating data from firefighters' interventions and supplemented with information from various sources. The process is detailed as follows:

- Dictionary keys, generated in hourly blocks, span from "01/01/2015 00:00:00" to "30/12/2023 11:00:00" in the "YYYY-MM-DD hh:mm:ss" format.
- Initially, our reference for meteorological data was Météo France [10] (the french public meteorological service). However, challenges arose due to limitations in accessing distant major stations and the 3-hour sampling interval, impacting both geographical and temporal precision. Although Météo France's bulletins furnish data on diverse meteorological risks, the introduction of MeteoStat [11] was pivotal in addressing this limitation. MeteoStat provides future forecasts, complementing Météo France's capabilities.
  - Importation data from three meteorological stations situated in the Doubs department, encompassing parameters like atmospheric pressure, cloudiness, barometric trends, temperature, humidity, precipitation in the last hour, dew point, precipitation in the last three hours, gusts over a period, average wind speed every 10 minutes, horizontal visibility, average wind direction every 10 minutes, and finally, climate current, sourced from Météo France.
  - Gathering vigilance bulletins from MétéoFrance covering various meteorological risks (winds, rains, floods, storms, snow, freezing rain, heatwaves, extreme cold) with assigned color codes (green, orange, red, yellow), thereby enhancing meteorological data [12].
  - Utilizing the MeteoStat API to access climatic variables such as temperature, dew point, precipitation, snow, wind speed and direction, pressure, and humidity from open meteorological and climatic data. Temperature readings are obtained from a grid of 11x11 points that spans the entire department.
- Incorporation of information on the elevation of the top forty rivers in Doubs, sourced from the "Hydroreel french governmental service". This involved populating a dictionary with the most proximate average readings for each hourly block [13].
- Inclusion of calendar-related details encompassing the hour, day, day of the week, day of the year, month, and year.
- Incorporation of information on holidays and academic breaks [14].
- Integration of events like Ramadan, lockdowns, and curfews.
- Inclusion of weekly epidemiological data obtained from the Sentinelles Network, covering conditions such as chickenpox, flu, and acute diarrhea [15].
- Ongoing retrieval of diverse Earth images through NASA's VIIRS and MODIS satellites, utilizing varying wavelengths and resolutions for assessing fire spread in specific regions [16] and [17].
- Inclusion of data on fine particulate matter ($PM_{2.5}$, PM_{10}), ozone (O_3), and nitrogen oxides (NO_2) from various air quality monitoring stations in the vicinity [18].
- Utilization of the [19] libraries to calculate the distance between the Earth, the Moon, and the Sun. Leveraging Astral [20] for parameters related to the sun and moon, such as moon phase and moonrise, to analyze their impact on natural disasters. Additionally, sunrise and sunset information is employed to define a boolean variable indicating "night" or "day".

– Introduction of variables like radio flux at 10 cm, number of sunspots, sunspot area, and X-rays to evaluate the influence of solar activity [21].
– Integration of variables associated to French league and Champions League football matches, recognized as potential indicators influencing interventions [22].

The chosen variables, extracted from an analysis of firefighters' interventions, are intended to ascertain their potential impact. While this comprehensive approach carries risks of incorporating non-influential variables, it facilitates the comparison of correlations between variables related to solar activity and other parameters. Encompassing a range of risks such as accidents, fires, floods, etc., these variables are present in the compilation of firefighters' interventions.

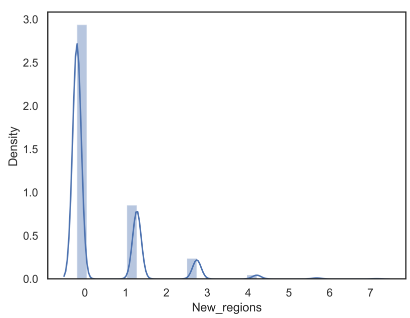
Data Pre-processing During the data processing phase, linear interpolation was employed to address the incompleteness of certain meteorological data. To adapt the data to our learning model, two methods from the Scikit-learn library [23] were utilized. The "StandardScaler" method was applied to normalize numerical variables, encompassing the year, hour, humidity, dew point, wind speed and direction, flu, cloudiness, precipitation, gusts, visibility, statistics on chickenpox, temperature, and acute diarrhea, river heights, and moon distance. This method rescales the distribution of values to have a mean of zero and a unit variance.

Furthermore, the "TargetEncoder" method [24] was applied to transform categorical variables, including day of the week, year, month, barometric trend, holidays and events, by the average of the corresponding target variable. The original target values (the number of interventions) were preserved, as discrete values more accurately represent the distribution of interventions.

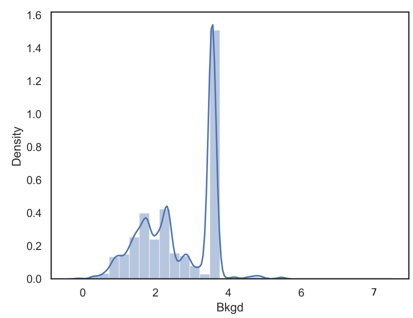
Data Mining The in-depth analysis of the dataset has proven pivotal for extracting relevant observations within the scope of our research. On average, around 30,000 interventions per year were observed. A noteworthy observation is the annual increase in the number of interventions. Concerning the variables associated with solar activity, Table 1 summarizes pivotal statistical measures pertaining to diverse solar activity variables. It provides insights into the central tendency, spread, and distribution patterns observed across various categories of variables in this family. Regarding variables such as New_regions, Xray and Optical have a significant frequency of zero values, which could potentially impact the characteristics of the dataset. The majority of these variables manifest as continuous variables, predominantly exhibiting a positively skewed distribution, except for two variables New_regions and Bkgd which show the presence of several modes in the distribution (see Figures 1 (a) and 1 (b)). In the context of temporal data, these modes could be attributable to solar cycles, spatial variability, exceptional solar events and complex interactions between different components, reflecting the dynamic complexity of solar activity.

Table 1: Data analysis of solar activity variables

Variable	**Mean**	**Std**	**IQR**	**Min**	**Max**	**Distinct values**
Radio_flux_10cm	94.1	31.4	39.0	64.0	343.0	144
SESC_Sunspot_number	39.1	43.6	63.0	0.0	227.0	176
Sunspot_area	273.5	381.3	420.0	0.0	2410.0	201
New_regions	0.4	0.7	1.0	0.0	5.0	6
Bkgd	1.75	1.00	1.77	0	5.5	175
XrayC	2.1	3.7	3.0	0.0	30.0	22
XrayM	0.2	0.8	0.0	0.0	11.0	11
XrayX	0.0	0.1	0.0	0.0	2.0	3
XrayS	2.7	5.1	3.0	0.0	55.0	38
Optical1	0.2	0.6	0.0	0.0	7.0	7
Optical2	0.0	0.2	0.0	0.0	2.0	3
Optical3	0.0	0.1	0.0	0.0	1.0	2



(a) Distribution of New_regions



(b) Distribution of Bkgd

Fig. 1: Analysis of the distribution of New_regions and Bkgd

Additionally, time series analysis of these variables reveals a similarity. To support this conjecture, we conducted a correlation analysis between these variables, yielding results that validate substantial correlations. For instance, the variables Radio_flux_10cm, SESC_Sunspot_number and Sunspot_area are highly correlated with each other, ranging from 0.85 to 0.92, meaning that a higher number of sunspots corresponds to a higher radio flux, or a larger sunspot area corresponds to a higher radio flux. This indicates that these three variables provide the same information for prediction, which is the impact of solar flux. The variables XrayC, XrayS and XrayM are correlated with each other, ranging from 0.51 to 0.7. Finally, in this analysis concerning the variable New_region, it is observed that an increase in new regions corresponds to a decrease in radio flux activity due to the lack of solar spot activity.

In conclusion of our analysis, we attempted to assess the impact of solar activity on the overall target variable through a correlation analysis. The results reveal a small, non-significant impact, particularly observed in the context of background radiation (Bkgd), showing a correlation of 0.14 with the target variable. These observations offer a preliminary insight into the limited influence of solar activity on intervention patterns. Although there is no correlation, the possibility of another, more complex link existing guides our understanding of the intricate relationships between solar activity and incidents requiring emergency services intervention.

2.2 Feature and Model Selection

Traditionally, it has been common practice to leverage all available features in the training dataset under the assumption that incorporating all information would lead to the construction of an optimal model. However, there are two primary reasons for advocating the necessity of limiting the number of considered features. Firstly, some variables may exhibit strong correlations, while others may contribute limited predictive power, potentially resulting in poor generalization or the introduction of redundant information. Secondly, the inclusion of a large number of features can significantly increase computation times without necessarily enhancing model performance [25]. Consequently, working with a more restricted set of features has the potential to yield more efficient results. In this study, the feature selection method employed, known as 'feature importance', assigns a score to each variable in the dataset, with higher scores indicating increased relevance [26]. The threshold was set to select the top 400 most significant features. Various selection techniques were applied, including:

1. High Variance: retaining features with a variance greater than 0.5.
2. Pearson Correlation Coefficient: filtering out correlations with the target variable whose absolute value is equal to or greater than 0.4.
3. Spearman Correlation Coefficient: applying the same principle as the Pearson coefficient.
4. Chi-Square Selector: using the chi-square test to evaluate the dependence between each feature and the target variable after normalizing features with the 'Min-Max Scaler' function.
5. Extreme Gradient Boosting (XGBoost) [27]: utilizing predefined hyperparameters (maximum depth = 7, number of estimators = 100000, early stopping after 10 rounds), conducting training, and subsequently determining or retrieving feature importance.
6. Light Gradient Boosting Model (LightGBM) [28]: employing specified hyperparameters (learning rate = 0.1, objective function = regression, metric = RMSE, number of leaves = 2^7, maximum depth = 7, number of estimators = 100000, early stopping after 10 rounds), and extracting feature importance as mentioned earlier.

The outcome of this selection process is a consolidated list of features, focusing on those that are consistently identified across multiple techniques. Out of the initial 3912 features, approximately 10% have been retained, resulting in a subset of 400 features for the training of our model.

Our model selection for this analysis centers around the implementation of Extreme Gradient Boosting (XGBoost) [27]. Renowned for its scalability, efficiency, flexibility, and speed, XGBoost excels in handling intricate datasets and achieving high levels of predictive accuracy. Its effectiveness stems from the integration of gradient boosting principles, enabling the algorithm to iteratively enhance the performance of weak learners. Widely applied in diverse domains, ranging from finance to healthcare, XGBoost's versatility extends to addressing regression, classification, and ranking tasks with exceptional efficiency. Its popularity is further underscored by its adeptness in managing missing data, feature selection, and accommodating a variety of data types.

2.3 Approaches Implemented for The Prediction Tool

To put our theoretical analysis into practice, we devised a robust training and optimization pipeline (see Figure 2).

1. We initiated multiple training iterations, spanning different time horizons from 1 to 24 hours, to assess solar activity impact up to a day ahead.
2. Following this, we conducted feature extraction as outlined in the "Feature Selection" section.
3. Iteratively, each chosen variable from the prior phase was added to the XGBoost model for training.
4. Each iteration involved splitting the dataset into training, testing, and validation sets, starting with 20% for testing and 80% for training-validation, with subsequent adjustments to 80% training and 20% validation.
5. We trained the XGBoost model using predefined hyperparameters (see Table 2) and incorporated early stopping with "early_stopping_rounds=15" to cease training if the validation set's performance did not improve for 15 consecutive rounds. This decision was guided by the fixed 100,000 estimators in our model.
6. Prediction scores were computed using the RMSE metric (Equation 1), with results compared to previous scores. Variables showing improvement were retained in the input list through a sequential feature selection approach, and prior scores were updated.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

7. Finally, we pursued an ultimate step to optimize prediction results by exploring diverse hyperparameter combinations (see Table 3). This involved using "early_stopping_rounds=25" in the context of a fixed parameter with 100,000 estimators. Our iterative approach enabled the selection of optimal hyperparameters, revealing a combination of values leading to optimized performance.

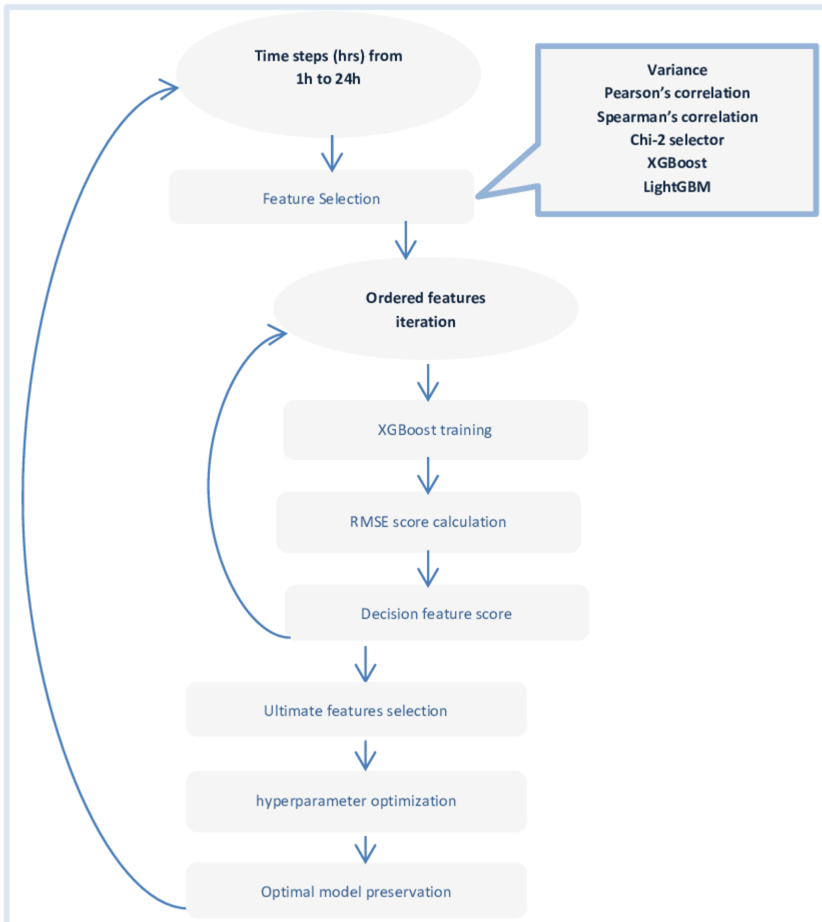


Fig. 2: Training pipeline representation model

Table 2: XGBoost default hyperparameters

Hyperparameter	**Values**
max_depth	7
min_child_weight	1
gamma	0
subsample	0.8
colsample_bytree	0.8
learning_rate	0.1

Table 3: Settings XGBoost hyperparameters

Hyperparameter	**XGBoost**
`max_depth`	[2, 14]
`min_child_weight`	[0, 14]
`gamma`	[0.0, 0.4]
`subsample`	[0.6, 0.9]
`colsample_bytree`	[0.6, 0.9]
`learning_rate`	[0.01, 0.009]

3 Results

Considerable efforts were dedicated to gathering, processing, and compiling a comprehensive dataset, with a focus on integrating information from diverse sources of past interventions. Particular emphasis was placed on assessing the significance of each variable within this context. As previously stated, the essence of this study lies in uncovering insightful observations regarding the influence of solar activity on intervention prediction. The preceding section detailed a methodology for selecting essential features. Subsequent to training our model across a temporal span covering an entire day, we present the outcomes of this feature selection, emphasizing variables linked to solar activity whose incorporation led to enhanced prediction performance (see Table 4).

Table 4: Exploring the impact of Feature Selection on solar activity analysis

Time horizon	**Solar activity**	**Feature Selection Technique**	**Rank**
9h	Sunspot_area	Variance and XGBoost	153
12h	Radio_flux_10cm	Variance and XGBoost	160
13h	Bkgd	Variance and LightGBM	103
	Sunspot_area	Variance and LightGBM	150
	Radio_flux_10cm	Variance and XGBoost	169
15h	Bkgd	Variance and LightGBM	83
	Sunspot_area		132
17h	Radio_flux_10cm	Variance and XGBoost	170
18h			181
19h			174
20h			178
22h			184
23h			173
24h			193

Moreover, we present the classification outcomes derived from our model, ready for interpretation in the subsequent section. Table 5 highlights the predictive performance of the XGBoost model concerning the intervention count, incorporating horizons that encompass solar activity variables, with the most superior performances emphasized in bold. Moreover, it provides RMSE outcomes both before and after the inclusion of solar activity variables, along with the percentage improvement. Figure 3 illustrates the results of 300 samples attempting to predict an atypical number of interventions at the 12th hour. Noteworthy is the substantial 2.67% improvement in RMSE achieved by selecting a single solar activity variable during this period compared to the selections of other variables in preceding horizons. Figures 4 (a) and 4 (b) visually represent the precise number of predictions, showcasing a maximum of 0 to 17 errors for the 12th hour before and after incorporating the solar activity variable, respectively. Given that the XGBoost model generates decimal predictions (e.g., 8.32 interventions), the results were rounded to the nearest integer (here, 8 interventions) for alignment with real-world applications.

Remarkably, XGBoost demonstrated its most favorable RMSE score of 1.6782 at the 2nd hour without considering any of the examined solar activity variables. In contrast, the optimum RMSE score, achieved by incorporating the solar activity variable Radio_flux_10cm, is 2.3451 at the 12th hour. Subsequently, both the XGBoost model

Table 5: Insights into classification outcomes

Time horizon	Solar activity	RMSE pre-sel	RMSE post-sel	Improvement
9h	Sunspot_area	3.6833	3.6395	1.20%
12h	Radio_flux_10cm	2.3473	**2.3451**	0.09%
13h	Bkgd	3.6993	3.6723	0.74%
	Sunspot_area	3.5642	3.5633	0.03%
	Radio_flux_10cm	3.5260	3.5101	0.45%
15h	Bkgd	3.7677	3.7449	0.61%
	Sunspot_area	3.5623	3.5594	0.08%
17h	Radio_flux_10cm	3.5846	3.5839	0.02%
18h		2.3980	2.3891	0.37%
19h		3.8072	3.7825	0.65%
20h		2.4549	2.4409	0.57%
22h		2.4113	2.3830	1.19%
23h		2.5803	2.5112	**2.67%**
24h		2.5179	2.4786	1.59%

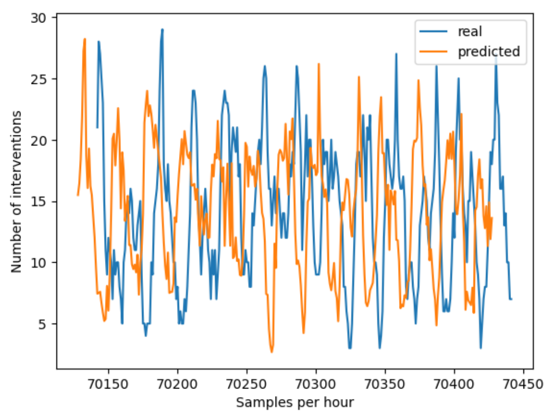


Fig. 3: Prediction for the 12th hour

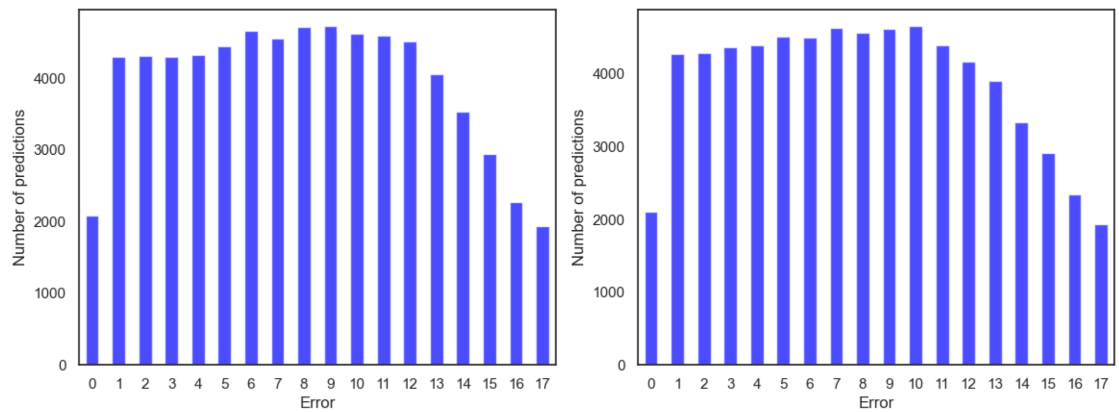
Table 6: Best hyperparameter for the 2nd hour

Hyperparameter	Values
max_depth	4
min_child_weight	4
gamma	0.2
subsample	0.9
colsample_bytree	0.8
learning_rate	0.1

Table 7: Best hyperparameter for the 12th hour

Hyperparameter	Values
max_depth	7
min_child_weight	3
gamma	0.2
subsample	0.7
colsample_bytree	0.7
learning_rate	0.005

trained at the second hour without solar activity and the model featuring the inclusion of at least one solar activity variable at the 12th hour underwent a meticulous grid search for parameter fine-tuning. This method systematically explores various combi-



(a) Prediction's error at the 12th hour pre-selection (b) Prediction's error at the 12th hour post-selection

Fig. 4: Insights from predictive modeling

nations of attributes to identify the most effective solution. Tables 6 and 7 provide details on the hyperparameters discovered and utilized, ensuring the model achieves its peak performance.

4 Discussion

The present study sought to evaluate how solar activity influences the prediction of firefighting interventions over an eight-year span from 2015 to 2023. Our methodological approach was carefully designed, commencing with meticulous data preparation and extensive analysis to inform our experimental decisions, encompassing feature selection through diverse statistical and machine learning techniques. XGBoost was selected as the predictive model due to its resilience in handling outliers, a quality particularly advantageous for continuous solar activity variables (see Table 1).

Our methodology included the implementation of an iterative training pipeline, where features enhancing the RMSE score over a 24-hour horizon were systematically selected (refer to Figure 2). The findings, outlined in Table 5, underscore the significance of variables associated with solar activity, particularly during extended timeframes. Notably, at the 12th hour, the presence of the radio flux contributed to a promising RMSE score of 2.345, and by the 23rd hour, utilizing the Radio_flux_10cm feature exhibited the most substantial RMSE improvement at 2.67%. Nevertheless, during the initial hours, spanning from the 1st to the 8th hour, the absence of solar activity-related variables resulted in the most favorable overall outcome for the day, yielding an RMSE of 1.678. This observation suggests a progressive impact of solar activity variables over extended timeframes, highlighting their increased utility for long-term predictions. Thus, our findings clearly indicate that variables related to solar phenomena Sunspot_Area and Bkgd exert a noticeable impact over medium-term horizons, specifically from the 9th to the 15th hour. On the contrary, the radio flux exhibits an effect on both medium and long-term horizons, stretching up to the 23rd hour, due to diverse temporal and atmospheric influences. Consequently, we can deduce from these outcomes that these variables may produce delayed effects on weather conditions or other factors influencing firefighter interventions. Solar phenomena can affect various aspects of the Earth's atmosphere, resulting in varied consequences at different hours of the day. This may elucidate why the impact becomes apparent later in the day. Despite the potential impact of solar risk on health and the environment, its effects are generally not immediate triggers for firefighter interventions, as they are primarily mobilized for immediate emergency situations.

The consideration revolves around understanding why there are not enough selected solar activity variables that could enhance the prediction. For instance, by taking into account the improved prediction result of 2.345 after selecting the radio flux in the 12th hour, it appears that other features within the same category simply provide redundant information and cannot be selected as they will not contribute to improving the prediction result. Additionally, as depicted in Figure 5, it becomes evident, after calculating the correlation of variables across different categories with the radio flux feature, that solar activity variables show a strong correlation, ranking high. Notably, variables like the number of sunspots (SEC_Sunspot_number), X-ray (XrayC, XrayS, XrayM), and new solar regions, while also observing correlation with other variables not belonging to the same family, such as overall climate, wind speed and acute respiratory infections. The logical explanation for this result is indirect causal correlation, meaning that solar activity might have an indirect influence on acute respiratory infections and wind speed. For example, variations in solar activity could impact the climate, thereby altering wind speed and, consequently, influencing the spread of respiratory infections. To delve further, a brief detailed analysis of the prediction improvement after selecting a solar activity variable, as depicted in Figure 6, reveals a consistently higher average improvement in predictions during the summer. This strongly advocates for the inclusion of solar activity variables during this season to achieve promising prediction outcomes.

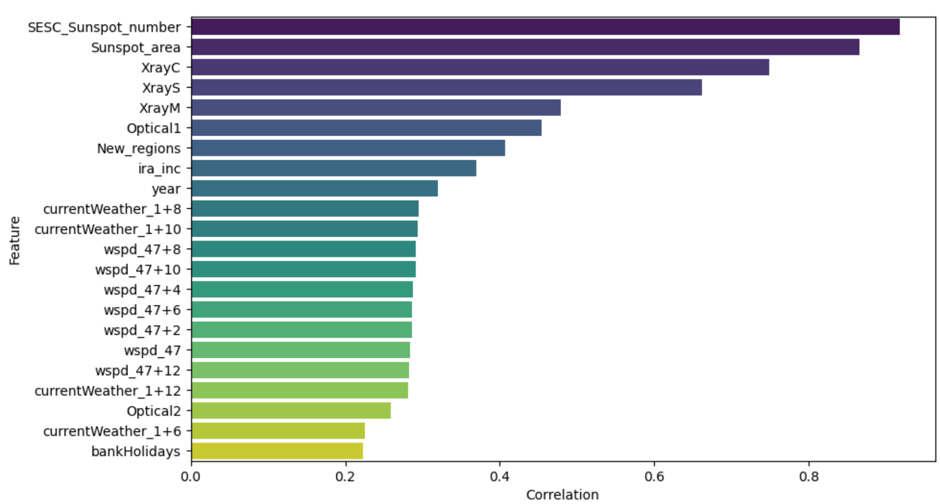


Fig. 5: Correlation with Radio_flux_10cm at the 12th hour

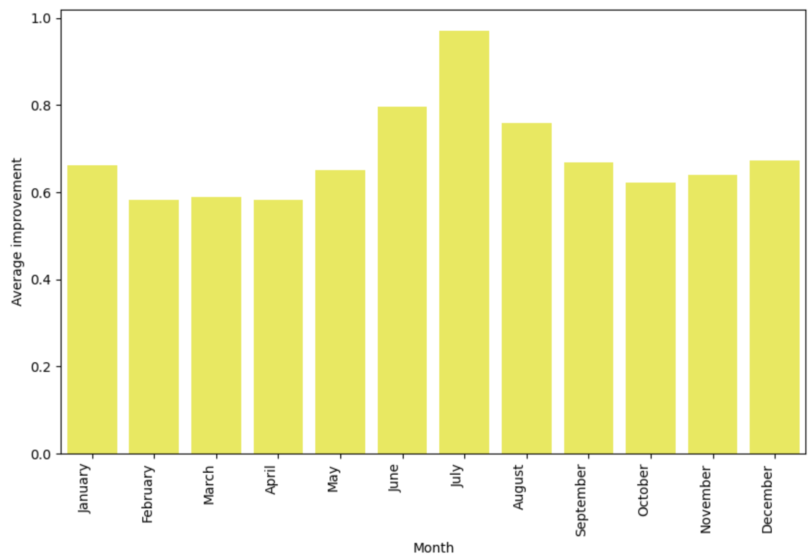


Fig. 6: Improvement statistics per month

A detailed examination, illustrated in Figure 4, uncovers a decrease in prediction errors, thereby validating our results. More specifically, the figure highlights a reduction in errors, particularly an increase in errors of 0 and 1, following the inclusion of the radio flux variable compared to the errors observed before its inclusion. This research deepens our comprehension of how solar activity influences firefighting interventions, presenting novel insights. However, some limitations, such as the implementation of other boosting models or machine learning methods for comparison purposes, and the exclusive use of grid search for hyperparameter optimization, have tempered our findings. Future avenues will involve exploring additional solar activity variables, investigating the combined impact of solar activity and meteorology, employing alternative models, and adopting more sophisticated sequential optimization approaches, such as Bayesian optimization.

5 Conclusion

This study constitutes a further exploration within a series of articles aiming to investigate diverse variable categories and their implications for firefighting interventions. Focusing on predicting future intervention numbers, a critical aspect for firefighting services globally, this article explores the influence of solar activity on emergency service forecasts. Utilizing an extensive eight-year dataset from the Fire and Rescue Department (SDIS 25) in the Doubs region, France, the research examines how the predictive model endeavors to identify significant trends associated with solar activity variables during distinct periods throughout the day. The findings suggest that variables in this category may have a cumulative impact, emphasizing the growing relevance of forecasts over extended timeframes. For instance, the consistent influence of radio flux from the 12th to the 24rd hour underscores the increased utility of predictions for later time intervals. While this approach proves valuable for resource management and response time optimization, ongoing improvements are imperative for seamless integration into firefighting decision-making processes.

In the continuation of our research, we will delve deeper into assessing the nuanced impact of variable categories (epidemiology, alert bulletins, events, etc.) on forecasts across diverse time horizons. We aim to extract valuable insights from these domains and pinpoint the most influential variables based on intervention types and prediction timelines. Leveraging these insights, we plan to refine our predictions, particularly for long-term scenarios, by selecting models and their hyperparameters that align with the heightened relevance of specific variables. Additionally, we will focus on incorporating geolocated variables to fine-tune predictions based on factors such as population density, forest area, etc. Our overarching objective is to refine our predictive capabilities, facilitating the implementation of operational strategies that proactively address the demand for firefighting services.

References

1. Public Health France. https://www.santepubliquefrance.fr/presse/2018/risques-solaires-et-cancers-les-habitudes-et-connaissances-des-francais-decryptees. Accessed on 21 February 2024.
2. Public Health France. https://www.santepubliquefrance.fr/presse/2024/bilan-canicule-et-sante-un-ete-2023-marque-par-4-episodes-de-canicule-avec-un-impact-sanitaire-important. Accessed on 21 February 2024.
3. Chen, Changsong and Duan, Shanxu and Cai, Tao and Liu, Bangyin, *Online 24-h solar power forecasting based on weather type classification using artificial neural network*, *Solar energy*, vol. 85, no. 11, pp. 2856–2870, 2011, Publisher: Elsevier.
4. Citakoglu, Hatice, *Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation*, *Computers and Electronics in Agriculture*, vol. 118, p. 28–37, 2015, Publisher: Elsevier.
5. Parsapoor, Mahboobeh and Bilstrup, Urban and Svensson, Bertil, *Forecasting solar activity with computational intelligence models*, *IEEE Access*, vol. 6, pp. 70902–70909, 2018, Publisher: IEEE.
6. Alizamir, Meysam and Othman Ahmed, Kaywan and Shiri, Jalal and Fakheri Fard, Ahmad and Kim, Sungwon and Heddam, Salim and Kisi, Ozgur, *A New Insight for Daily Solar Radiation Prediction by Meteorological Data Using an Advanced Artificial Intelligence Algorithm: Deep Extreme Learning Machine Integrated with Variational Mode Decomposition Technique*, *Sustainability*, vol. 15, pp. 11275, 2023, Publisher: MDPI.
7. Nahuis, Selene Leya Cerna and Guyeux, Christophe and Arcolezi, Héber Hwang and Couturier, Raphaël and Royer, Guillaume and Lotufo, Anna Diva Plasencia. *Long short-term memory for predicting firemen interventions.* In: *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, 2019, pp. 1132–1137.
8. Cerna, Selene and Guyeux, Christophe and Arcolezi, Héber H and Couturier, Raphaël and Royer, Guillaume. *A comparison of LSTM and XGBoost for predicting firemen interventions.* In: *World Conference on Information Systems and Technologies*, Springer, 2020, pp. 424–434.
9. Guyeux, Christophe and Makhoul, Abdallah and Bahi, Jacques M. *How to build an optimal and operational knowledge base to predict firefighters' interventions.* In: *Proceedings of SAI Intelligent Systems Conference*, Springer, 2022, pp. 558–572.

10. Météo-France. https://www.ecologie.gouv.fr/. Accessed on 21 February 2024.
11. Meteo-Stat. https://pypi.org/project/meteostat/. Accessed on 21 February 2024.
12. Vigilance-France. https://vigilance.meteofrance.fr/fr. Accessed on 21 February 2024.
13. Ministry of Ecological Transition. http://www.hydro.eaufrance.fr/. Accessed on 21 February 2024.
14. Ministry of National Education. http://www.education.gouv.fr/pid25058/le-calendrier-scolaire.html. Accessed on 21 February 2024.
15. The Sentinel Network. https://www.sentiweb.fr/?page=table. Accessed on 21 February 2024.
16. NASA. https://lance.modaps.eosdis.nasa.gov/viirs/. Accessed on 21 February 2024.
17. MODIS. https://lance.modaps.eosdis.nasa.gov/modis/. Accessed on 21 February 2024.
18. ATMO-BFC. https://www.atmo-bfc.org/accueil. Accessed on 21 February 2024.
19. Skyfield. https://github.com/skyfielders/python-skyfield. Accessed on 21 February 2024.
20. Astral. https://pypi.org/project/astral/0.5/. Accessed on 21 February 2024.
21. NASA. https://www.swpc.noaa.gov/. Accessed on 21 February 2024.
22. Soccer. https://www.footendirect.com/. Accessed on 21 February 2024.
23. Pedregosa, Fabian and Varoquaux, Gael and Gramfort, Alexandre and Michel, Vincent and Thirion, Bertrand and Grisel, Olivier and Blondel, Mathieu and Prettenhofer, Peter and Weiss, Ron and Dubourg, Vincent and others. *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, 12, 2011, pp. 2825–2830.
24. Target Encoder. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.TargetEncoder.html. Accessed on 21 February 2024.
25. Garreta, Raúl and Moncecchi, Guillermo. *Learning scikit-learn: machine learning in Python: experience the benefits of machine learning techniques by applying them to real-world problems using Python and the open source scikit-learn library.* (No Title), 2013.
26. Zien, Alexander and Krämer, Nicole and Sonnenburg, Sören and Rätsch, Gunnar. *The feature importance ranking measure.* In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II*, pp. 694–709, 2009, Springer.
27. Chen, Tianqi and Guestrin, Carlos. *Xgboost: A scalable tree boosting system.* In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
28. Ke, Guolin and Meng, Qi and Finley, Thomas and Wang, Taifeng and Chen, Wei and Ma, Weidong and Ye, Qiwei and Liu, Tie-Yan. *Lightgbm: A highly efficient gradient boosting decision tree.* In: *Advances in neural information processing systems*, volume 30, 2017.