International Journal of Geo-Information

MDPI

Article

GIS and Machine Learning for Analysing Influencing Factors of Bushfires Using 40-Year Spatio-Temporal Bushfire Data

Wanqin He [1], Sara Shirowzhan [2,*] and Christopher James Pettit [2]

[1] School of Built Environment, Kensington Campus, The University of New South Wales, Sydney, NSW 2052, Australia; wanqin.he@unswalumni.com
[2] City Futures Research Centre, The University of New South Wales, Sydney, NSW 2052, Australia; c.pettit@unsw.edu.au
* Correspondence: s.shirowzhan@unsw.edu.au

Abstract: The causes of bushfires are extremely complex, and their scale of burning and probability of occurrence are influenced by the interaction of a variety of factors such as meteorological factors, topography, human activity and vegetation type. An in-depth understanding of the combined mechanisms of factors affecting the occurrence and spread of bushfires is needed to support the development of effective fire prevention plans and fire suppression measures and aid planning for geographic, ecological maintenance and urban emergency management. This study aimed to explore how bushfires, meteorological variability and other natural factors have interacted over the past 40 years in NSW Australia and how these influencing factors synergistically drive bushfires. The CSIRO's Spark toolkit has been used to simulate bushfire burning spread over 24 h. The study uses NSW wildfire data from 1981–2020, combined with meteorological factors (temperature, precipitation, wind speed), vegetation data (NDVI data, vegetation type) and topography (slope, soil moisture) data to analyse the relationship between bushfires and influencing factors quantitatively. Machine learning-random forest regression was then used to determine the differences in the influence of bushfire factors on the incidence and burn scale of bushfires. Finally, the data on each influence factor was imported into Spark, and the results of the random forest model were used to set different influence weights in Spark to visualise the spread of bushfires burning over 24 h in four hotspot regions of bushfire in NSW. Wind speed, air temperature and soil moisture were found to have the most significant influence on the spread of bushfires, with the combined contribution of these three factors exceeding 60%, determining the spread of bushfires and the scale of burning. Precipitation and vegetation showed a greater influence on the annual frequency of bushfires. In addition, burn simulations show that wind direction influences the main direction of fire spread, whereas the shape of the flame front is mainly due to the influence of land classification. Besides, the simulation results from Spark could predict the temporal and spatial spread of fire, which is a potential decision aid for fireproofing agencies. The results of this study can inform how fire agencies can better understand fire occurrence mechanisms and use bushfire prediction and simulation techniques to support both their operational (short-term) and strategic (long-term) fire management responses and policies.

Keywords: bushfire; GIS; random forest; machine learning algorithm; fire simulation; spatial analysis

Citation: He, W.; Shirowzhan, S.; Pettit, C.J. GIS and Machine Learning for Analysing Influencing Factors of Bushfires Using 40-Year Spatio-Temporal Bushfire Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 336. https://doi.org/10.3390/ijgi11060336

Academic Editors: Wolfgang Kainz, Matteo Gentilucci, Marco Materazzi, Margherita Bufalini and Gilberto Pambianchi

Received: 31 March 2022
Accepted: 1 June 2022
Published: 6 June 2022

1. Introduction

Between 1981 and 2017, the bushfires affected approximately 6.2 million people and caused more than 2000 injuries and deaths worldwide [1]. Australia is one of the world's highest bushfire-prone countries, with bushfires having killed more than 800 people and billions of animals in the country since 1851. The Australian bushfires from November 2019 to February 2020 were among the worst on record, burning 12.8 million hectares of land, dealing a fatal blow to Australia's biological environment and affecting 57% of the Australian population [2]. At the same time, bushfires are inevitable in Australia, where

forest species are dominated by eucalyptus trees that have adapted to bushfires, as the northern savannah burns annually [3]. The southeast of mainland Australia is one of the highest fire-risk regions in the world due to its unique climatic characteristics. The region is home to nine of Australia's ten largest cities and three-quarters of the country's population. The region is heavily forested, with red eucalyptus and Flindersia australis as the dominant tree species, and the forest floor is covered with a variety of shrubs and fungi [4]. Rapid fuel accumulation and a high balance of fuel loads characterise this plant community [5]. The climate of the region is Mediterranean, with hot, dry summers and mild, wet winters [6]. Rains in spring and winter promote fuel (vegetation) growth, whereas dry summers increase the risk of fire. Forest burning releases greenhouse gases, solid particles and other gases that pose serious risks to the atmosphere and human health, and rampant bushfires can be hazardous to ecosystems, habitats and human life [7–9]. Therefore, understanding the combined mechanisms of factors that influence the occurrence and spread of bushfires is beneficial for developing effective fire prevention plans and fire suppression measures, as well as assisting in the planning of geographic, ecological maintenance and urban emergency management.

The causes of bushfires are extremely complex, and their scale and probability of occurrence are influenced by the interaction of meteorological factors, topography, human activity, vegetation type and many other factors. The impact of these factors on the spread of bushfires also varies with the study area and spatial scale. In particular, the climate is widely recognised as an important factor influencing changes in bushfire dynamics [10–17]. The influence of climate on bushfires is mainly manifested at large spatial scales, where it affects the occurrence of bushfires in two ways: by increasing the concentration of carbon dioxide in the air and rising temperatures, and by changing the water content and distribution patterns of combustible materials in the environment [18]. At small scales, factors such as combustibles, topography and human activity are likely to be the main factors influencing bushfires. The type and water content of combustible material directly affects the fire capacity of the area [19]. Topography (e.g., elevation, slope and aspect) also influences the amount and spatial distribution structure of vegetation or combustible material in the area, and thus the rate and direction of bushfire spread. Secondly, human activities influence the spatial pattern and frequency of bushfires through actions such as building expansion, construction of transportation networks and outdoor activities [20]. These factors interact with each other and ultimately trigger changes in the dynamics of bushfires, and the study of these interacting mechanisms can help us understand bushfires in more detail.

Current studies on the drivers of bushfires and fires prediction are mainly based on mathematical and statistical methods. By extrapolating mathematical models of the relationship between bushfire conditions and various elements such as meteorology, topography and socio-humanities, current studies further predict the probability of bushfires occurring on the ground. Among these, multiple linear regression and logistic regression (LR) are widely used to model the probability of forest fire occurrence [21]. Bradstock used Bayesian logistic regression to explore the relative influence of the environmental and drought components of the Forest Fire Danger Index (FFDI) on the probability of fire ignition [22]. Another study used remotely sensed image data and constructed a binary logistic regression model to produce a probability map of fire occurrence in south eastern Australia [23]. Geographically weighted regression has also been introduced into logistic models (GWLR) to analyse the influence of human factors on wildfires [24]. In addition to LR and GWLR models, the random forest algorithm is also a method that has been used in bushfire simulation research. The random forest algorithm can handle a large number of independent variables and can automatically select the important ones. The testing performance of random forest does not decrease (due to overfitting) as the number of trees increase [25]. Therefore, it does not overfit when researchers use as many trees as they want. Random forest algorithms allow more flexibility than regular regression methods [25] in assessing complex interactions between variables. Several studies have

used random forest models to quantify the effects of climate, vegetation type, topography and human activity on the distribution patterns of bushfires. The results were compared with traditional multiple linear regression methods and showed that the random forest algorithm was better at prediction [26]. At present, studies on the influence of bushfire factors have mainly focused on a specific region, whereas analyses of the effect on bushfires at large, geographic spatial scales are still lacking. In this research, the scale of the study is within the whole of New South Wales, and the time scale is extended to 40 years to analyse historical bushfire data.

In addition, fires can change rapidly, and timely prediction of spatial and temporal changes in fire is one of the main challenges of bushfires management. Both limited fire-fighting resources and the need for rapid emergency response require managers to be able to accurately predict the location of fires and fire-spread trends in a short period [27]. Simplifying large fire occurrence patterns, mapping fires, identifying fire-spread mechanisms and modelling fire effects are the best measures for planning and mitigating fire effects [28]. With the increase in computer image processing power and the maturity of cloud computing technology, the acknowledgement of bushfire prediction has shifted from various forms of calculating mathematical models to computer-based two-dimensional bushfire simulation models [29]. Spark is a toolkit developed by CSIRO in Australia to simulate the spread of forest fires over terrain [30]. The toolkit contains several modules that allow reading and writing to geographical data, computational models to simulate the spread of flame fronts and visualisation and a range of resultant data analysis tools. It allows the user to apply any available fire-spread algorithms depending on the situation and preference [31]. Fire simulations performed by Spark require the input of many datasets, including maps of land classification or fuel types, terrain, fuel information and weather data. Spark also allows the user to autonomously input scripts to combine these datasets in different ways to calculate the spread of different areas. All calculations in Spark are parallelised on the GPU architecture, allowing simulations to run much faster than in real-time [31]. This rapid prediction of fire spread can provide management agencies with a clear emergency response strategy in a short time and mobilise rescue resources in time to avoid severe damage.

As previously mentioned, many aspects of bushfire research in Australia have been discussed in various ways, but some gaps remain. To date, there has been less documented research using long-term spatial-temporal data to analyse the influencing drivers and predict potential spatial patterns of bushfires. This study attempts to address this gap by using 40 years of historical fire data, climatic factors and other influences in NSW to analyse the effect of different factors on the frequency and scale of bushfires across NSW. Due to data limitations, previous studies on bushfire impact factors and forest fire prediction have focused on a specific region, whereas analyses where the study area covers an entire fire-prone state are lacking. To fill this gap, the objectives of this paper are:

1. To explore the interactions among historical bushfire data, meteorological data and other data over the past 40 years in a large area in Australia, i.e., NSW, with an area of 801,150 km^2 and a population of 8.166 million (by September 2020);
2. To predict the burning spread of bushfire for 24 h at four hotspot regions using the Spark toolkit.

2. Materials and Methods

To achieve the objectives of this study, as shown in Figure 1, the study first used NSW bushfire open data from 1981–2020, combined with meteorological data, vegetation data and topography data to statistically analyse the relationship between bushfire occurrence frequency, burn scale and influence factors. Machine learning-random forest regression was then used to determine the differences in the influence of bushfire influence factors on the incidence and burn scale of bushfires. Indeed, this algorithm helped us to investigate the relative importance of each driver on bushfire. Finally, the data on each influence factor was imported into Spark, and the results of the random forest model were used to select

the most important drivers for the flame-spread model in Spark to simulate the spread of bushfires burning within 24 h in four hotspot regions of NSW. The details of the materials and methods used in this work are outlined below.

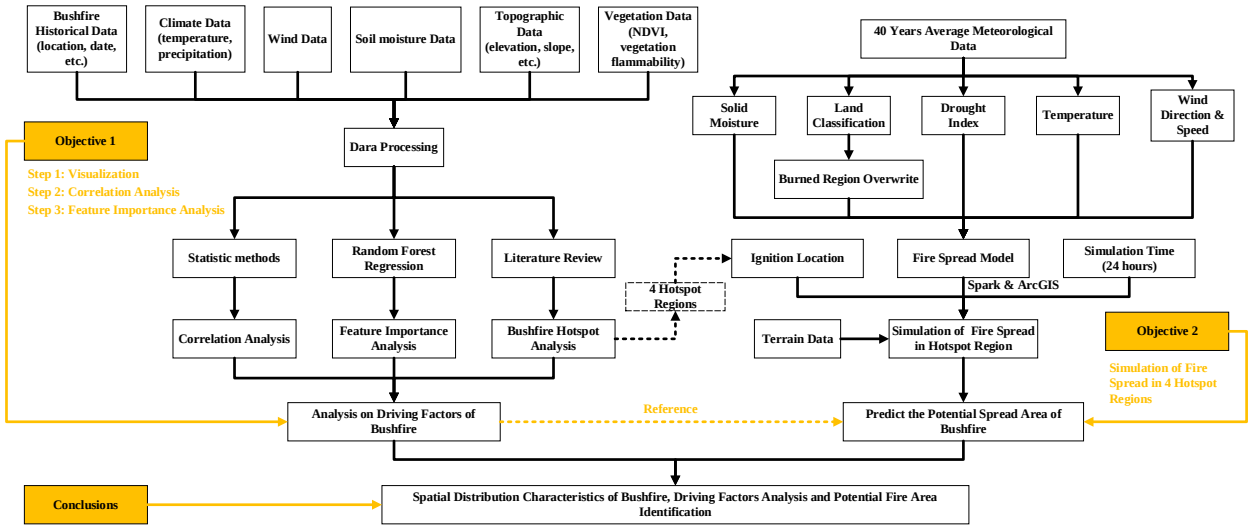

Figure 1. Research flowchart.

2.1. Study Area

New South Wales is located in the southeast of the Australian continent, and it borders Victoria and Queensland to the north and south and the Pacific Ocean to the east (see Figure 2). The state can be geographically divided into four regions: Sydney, Newcastle, Wollongong and the Blue Mountain. It represents one of the most populous states in Australia. The state currently has around eight million inhabitants, two-thirds of whom live in the Greater Sydney region [32].

Figure 2. Study area.

2.2. Data

In order to analyse the data of bushfire and its drivers, a range of data were collected in this study, as shown in Table 1.

Table 1. The data of bushfire and its drivers used in this study.

Data	Description of Data	Data Format	Source
Bushfire	Bushfire characteristics (date, area)	Shapefile	NSW Department of Planning, Industry and Environment
Climate	Air temperature	NetCDF	University of East Anglia Climate Research Unit (CRU)
Climate	Precipitation	NetCDF	University of East Anglia Climate Research Unit (CRU)
Soil moisture	Monthly soil moisture	NetCDF	NOAA Climate Prediction Centre (CPC)
Wind	10 m wind speed	NetCDF	the ERA5 dataset
Flammability	Vegetation flammability	Shapefile	NSW Department of Planning, Industry and Environment
Vegetation	NDVI	GeoTIFF	MODIS 1-km MYD13A3 NDVI, Collection 5
Topographic	Terrain slope	GeoTIFF	The Commonwealth of Australia (Geoscience Australia)

2.2.1. Bushfire Data

The primary source of bushfire data was the Shapefile entitled "NPWS Fire History—Wildfires and Prescribed Burns" published by the NSW Department of Planning, Industry and Environment in 2010. This dataset includes the final fire boundaries for each year from 1900 onwards. The fire types are divided into "Wildfire" and "Prescribed Burn". The fire boundaries are distinguished by the year of the fire, and the attribute table of the Shapefile contains information on the date of fire origin, type, area and perimeter.

2.2.2. Climate Data

Precipitation and temperature data were obtained from CRU TS4.04: Climatic Research Unit (CRU) Time-to-Month Variation in Climate published by the University of East Anglia Climatic Research Unit (CRU). NetCDF's data format contains global month-by-month variation in cloudiness, diurnal temperature range, frost-day frequency, wet-day frequency, potential evapotranspiration (PET), precipitation, daily mean temperature, monthly mean temperature, and daily maximum and minimum temperatures for the period 1901 to 2020 [33]. The spatial distribution of this data is high-resolution (0.5 × 0.5 degree) grids, generated using ADW interpolation.

2.2.3. Soil Moisture Data

Soil moisture data were obtained using the "CPC Soil Moisture" dataset published by the NOAA Climate Prediction Centre (CPC). This dataset contains monthly average soil moisture data from 1948 to May 2021 at a spatial resolution of 0.5 deg. The data was cleaned, statistically analysed and visualised using Python.

2.2.4. Wind Data

The study selected the wind speed data from the ERA5 dataset, the fifth generation of the ECMWF's global climate and weather reanalysis dataset for the last 40–70 years, which

consists of four main subgroups: hourly and monthly data, including barometric levels (upper airfields) and single levels (atmospheric, wave and land surface quantities) [34]. The dataset provides hourly estimates of atmospheric, wave and land surface volumes since 1950 and has been precomputed for ensemble means and distribution. The study used Python to process the dataset.

2.2.5. Vegetation Data

In this study, the Normalized Difference Vegetation Index (NDVI) was obtained from the Collection 5 MODIS Global Monthly Vegetation Index Product Series (MYD13A3) dataset. This dataset provides a gridded Level 3 product in the sinusoidal projection with a spatial resolution of 1 km (km) per month, including the MODIS NDVI and Enhanced Vegetation Index (EVI) [35]. The Vegetation Index is used for global vegetation status monitoring and for products showing land cover and land cover change. These data can be used as inputs to simulate global biogeochemical and hydrological processes and global and regional climate [35]. It was used as the index of fuel load in this study. According to USGS remote sensing phenology (2018), NDVI values range from +1.0~−1.0. Barren rocky, sandy or snowy areas typically show low NDVI values (e.g., 0.1 or less). Sparse vegetation such as shrubs and grass or aged crops may show moderate NDVI values (about 0.2 to 0.5). High NDVI values (about 0.6 to 0.9) correspond to dense vegetation in temperate and tropical forests or crops at peak growth [36].

2.3. Processing and Analysis of Data

2.3.1. Processing of Data

As presented in Table 1, there were three data formats in this work: Shapefile (.shp), NetCDF (.nc) and GeoTIFF (.tif). In this work, these data were read, extracted, processed and visualised using Python. The historical bushfire data were in the Shapefile format. GeoPandas in Python is an open-source library for working and manipulating geographic data, which was used here to read bushfire data. The geometry type of this geographic data was polygons, showing the spread regions of bushfire, which can be replaced by its centroid for further analysis. The format of topographic data and vegetation data were GeoTIFF, which can be converted into Shapefile using ArcGIS. Therefore, they also can be obtained by GeoPandas. The climate, soil moisture and wind data were NetCDF files, which can be read by another Python library, Xarray. The data read by Xarray looked like an n-dimensional array that included the geographic information, whereas the data read by GeoPandas, which were GeoDataFrame, looked like a data sheet with one column of geometry information. To keep the data format consistent in Python, the meshgrid was utilized to convert Xarray data to GeoDataFrame, of which the geometry type was polygons. After obtaining all the data in Python, the projection between points and polygons was utilized to establish a connection between bushfire data and other data. Finally, they can be analysed using statistical methods.

2.3.2. Correlation Analysis

After obtaining bushfire data on seven key influencing factors (air temperature, precipitation, soil moisture, wind speed, slope, NDVI and vegetation flammability), we firstly wanted to investigate their correlation. In this work, the Pearson correlation coefficient was used to measure the degree of correlation (linear correlation) between two variables (e.g., X and Y), with a value between −1 and 1. The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between these two variables [37]:

$$\rho X, Y = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

The above equation defines the overall correlation coefficient and is often represented by the lowercase Greek letter ρ as the representative symbol. Estimating the covariance and

standard deviation of the sample gives the Pearson correlation coefficient, often represented by the lowercase letter r [37]:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (2)$$

r can also be obtained from the mean of the standard scores at the sample points to obtain an expression equivalent to the above equation:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma_X}\right)\left(\frac{Y_i - \overline{Y}}{\sigma_Y}\right) \quad (3)$$

In $\frac{X_i - \overline{X}}{\sigma_X}$, the $\overline{X}$ and σ_X denote the standard score, and the sample mean and sample standard deviation of the X_i, respectively (Rodgers and Nicewander, 1988). A negative value of r represents a negative correlation, whereas a positive value represents a positive correlation. In addition, $0 \leq |r| < 0.3$ represents a low correlation; $0.3 \leq |r| < 0.8$, moderate correlation; and $0.8 \leq |r| <= 1$, high correlation. It should be noted that correlation is not the same as causation. A correlation indicates that the variables on either side of the equal sign will change simultaneously, whereas a causal relationship is due to one variable causing a change in another variable [38].

2.4. Random Forest Machine Learning Algorithm

Due to the complexity of bushfire data, the correlation analysis was used to obtain the interrelationship between bushfires and their seven influencing drivers. However, causation of these factors was not investigated. In order to estimate the importance of each driver on bushfire occurrence, a random forest regression method was applied to model bushfire data and other influencing drivers' data. Before conducting the random forest regression, the multicollinearity between model variables needed to be checked.

2.4.1. Multicollinearity Test for Model Factors

Multicollinearity is the existence of a complete or near-complete linear relationship between the explanatory variables in an equation; for example, one explanatory variable can be represented by a linear combination of other explanatory variables [39]. It is an important issue that commonly arises in current multiple regression analyses. If there is typical multicollinearity between these independent variables, then the regression fit generates biased covariate estimates, bringing about significant standard errors of the regression coefficients. The stability of the model is then reduced, and statistical inference is invalidated as a result. It is not easy to achieve complete independence between the different independent variables in specific regression analysis sessions. Therefore, before regression analysis is carried out, the independent variables need to be diagnosed for multicollinearity, and variables with multicollinearity need to be removed to improve the model's accuracy. In this study, the variance inflation factor (VIF) is used for the diagnosis of multicollinearity, and the expression of the variance inflation factor is:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (4)$$

R_i^2 is the coefficient of determination in linear regression and reflects the percentage of the change in the dependent variable explained by the regression equation [40]. It can be obtained from the square of the complex correlation coefficient between the dependent and independent variables or from the ratio of the squared residuals to the total sum of squares of the regression equation. When $0 < VIF < 10$, there is no multicollinearity between the variables; when $10 \leq VIF < 100$, there is strong multicollinearity between the variables; when $100 \leq VIF$, there is severe multicollinearity between the variables [39]. In the actual

test, the factors with a high VIF were removed from the factors, and those with a low VIF were retained, and so on, until a combination of factors with a low correlation was obtained to enhance the explanatory power of the model.

2.4.2. Theory and Methods of the Random Forest Algorithm

Random forest is a statistical theory that Breiman and Cutler proposed in 2001 [41]. It is a combinatorial classifier that can solve both classification problems and regression problems [41]. The random forest algorithm uses the bootstrap method to select x samples from the original data with a randomised put-back sampling, and a total of x sampling sessions are performed to generate x training sets. For each of these x training sets, x decision tree models are trained. Then, y variables are randomly selected at the nodes of each decision tree, and the one with the best information gain ratio (Gini index) among the y variables is selected for splitting. According to Fotheringham and Brunsdon (2010), the variable with the best classification power should be selected for branching at this point, and each tree is left free to grow to its maximum without any pruning, resulting in x results [42]. The result of the random forest is based on the combination of these x results. For classification problems, the final classification result is determined by classifier voting; for regression problems, the final prediction is determined by the mean of the predicted values. The advantages of the random forest algorithm are that it can handle a large number of input variables, assess the importance of variables, and is generally free from overfitting [26]. The random forest algorithm can also calculate importance scores for the respective variables to evaluate the importance of each fire-risk factor in the model.

Therefore, the random forest algorithm was used in this study to calculate the weighting of various natural environmental factors on the area burned by bushfires. The higher the importance score of a factor, the greater the influence of that factor on the area burned by bushfires. According to Cutler et al. (2007), there are two general measures of the importance of a factor: one is the reduction in the mean Gini index, where the greater the value, the greater the importance of the factor, and the other is the decrease in prediction accuracy when noise is added, where the more significant the decrease, the greater the importance of the factor [43].

2.5. Spark-Based Prediction for Future Spatial Pattern of Bushfires

After completing the analysis of importance of the seven drivers of bushfire, the predictions using those drivers could be realized. In this work, Spark was used as the platform to predict the bushfire occurrence and spread. The schematic is presented in Figure 3.

Based on Vinser et al. [44], the analysis of the hot and cold spots of bushfire occurrence in NSW over the past 40 years and the analysis of bushfires drivers, and the frequency and severity of fire occurrence in different areas are affected by different drivers (temperature, precipitation, soil moisture, wind speed, slope, NDVI and vegetation flammability). To further investigate the fire-spread trends, this study conducted a 24-h bushfire-spread simulation based on the Spark platform for the new hotspot areas in NSW identified in the Visner et al. study [44].

The Spark platform developed by CSIRO was used in this study for bushfire-spread simulation. Spark is a toolkit for simulating the spread of bushfires over terrain [45]. This toolkit contains many modules that allow reading and writing to geographical data, computational models to simulate the spread of flame fronts, and visualisation and a range of resultant data analysis tools. Combustion is a complex reaction flow problem that is difficult to resolve at large temporal and spatial scales. Therefore, to carry out bushfire burning simulations on large, spatial time scales, several empirical models have been developed to describe the spreading behaviour of bushfires [46]. These models calculate the rate of spread of bushfire fronts through several driving factors (temperature, wind speed, soil moisture, topography, land type, etc.). The CSIRO team currently offers dozens of flame-spread models for different land classifications, which are freely available [45] in

this platform. Calculations in Spark occur in vertically stacked layers, as shown in Figure 4. The top layer of data is the land classification, which has different flame-speed models for different land configurations, and the other layers are the other driver layers (temperature, wind speed, wind direction, elevation, relative humidity, dryness, etc.). Thus, for each grid, Spark uses the corresponding computational model and driver data to calculate the flame propagation velocity for that grid.

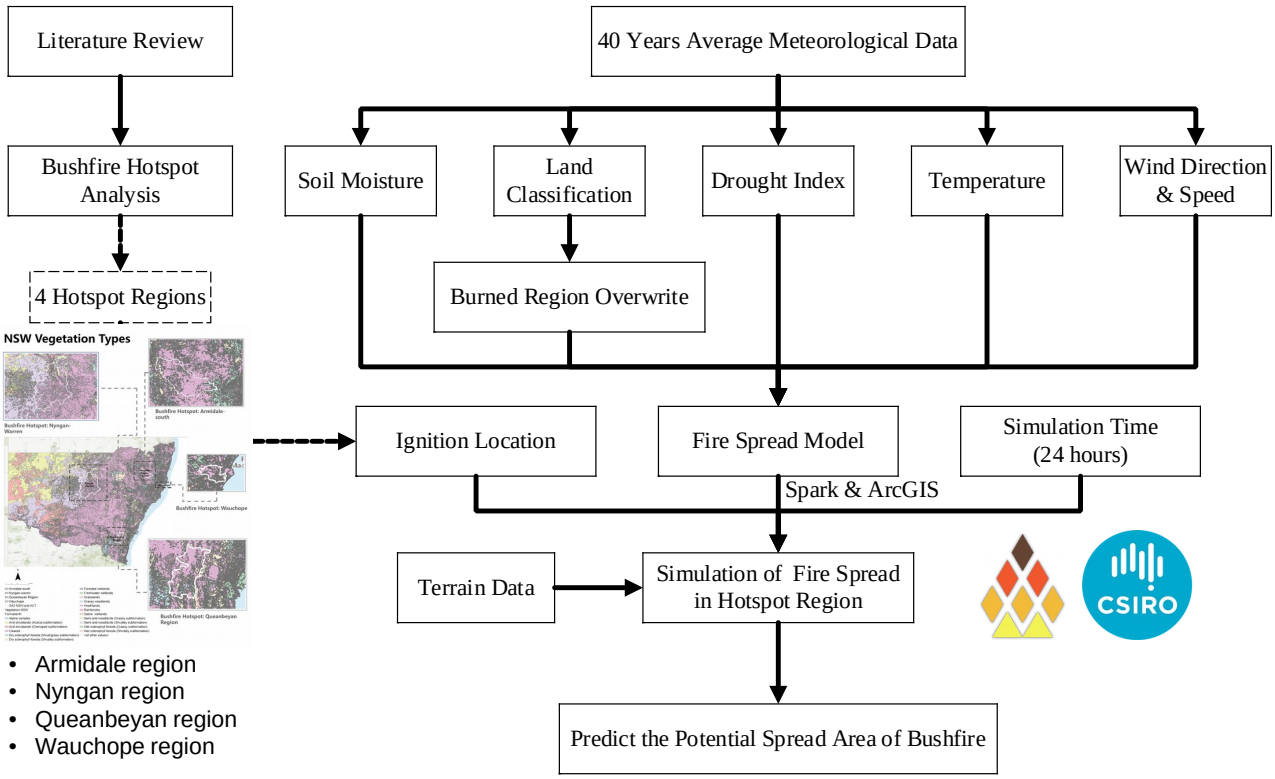

Figure 3. The schematic of bushfire predictions based on Spark.

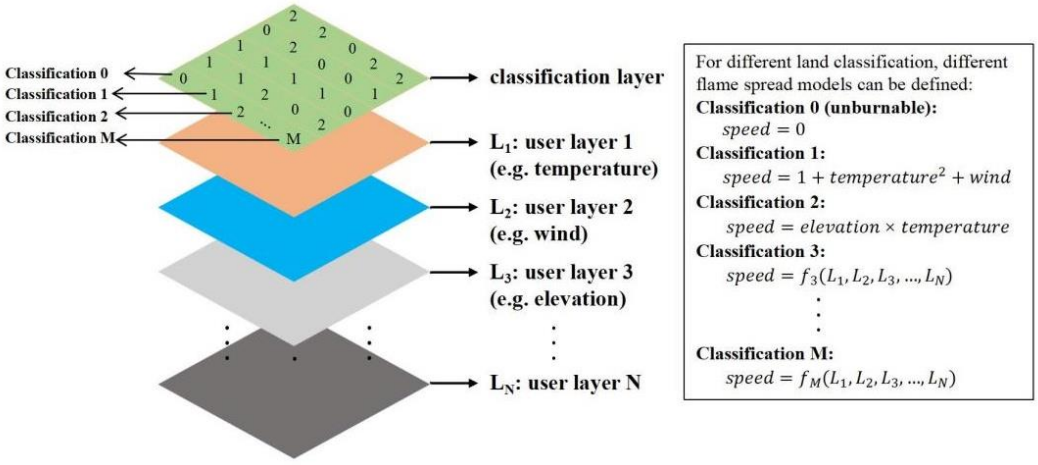

Figure 4. Schematic of Spark layers.

According to the Spark GUI setup, the specific calculation process is as follows:

1. Configure the calculation parameters. The basic parameters of the simulation are first set to be start/end time, total simulation time, simulation resolution and simulation projection. The fire area is then set by entering the coordinates of the fire point, the

fire radius and the fire time. Finally, the output file format is set, either as a GeoTIFF file or as a Shapefile, as required.

2. Data input. Depending on the needs of the calculation, this part needs to import meteorological data (wind speed, wind direction, temperature, relative humidity, dryness index), and the data format is not limited to the NetCDF format. The more critical topographic data to be imported here are land classification and elevation data, which are of the GeoTIFF file type, and burned areas that can be set to override the land classification data. Spark provides custom inputs to import additional data for specific input parts of the computational model.
3. Series input. The input data can also be series input data and gridded input data, which can be achieved by importing a CSV file and providing a Python script file to read it.
4. The initialisation model is a preprocessing model that allows the data imported earlier to be preprocessed and then imported into the computational model, which is a handy model.
5. Postprocessing models can be used to set up postprocessing models if you need to output layers of customised data after the calculation has been completed.

Based on the Spark platform and its computational process described above, this study set out four different fire zones in the bushfire hotspot areas in NSW over 40 years. This study used Spark to simulate the spread of fires in the high-incidence areas during the critical rescue time, which is 24 h after the fire has occurred, to reference bushfire safety in NSW. According to Spark's calculation process, the simulation time was set to 24 h, and the coordinates of the fire-starting points were in the Armidale region (−30.546511, 151.685837), Nyngan region (−31.684909, 147.060736), Queanbeyan region (−35.352470, 149.230795) and Wauchope region (−31.465117, 152.745067). The radius of the ignition area was set at 1200 m according to our prior simulations, which facilitated the observation of the spread of bushfires. The meteorological parameters (temperature, wind speed, relative humidity, etc.) were calculated using 40-year annual averages for the site, and the wind direction was used in the northwest division for uniform comparison. In consideration of the economy and accuracy of the calculation and the use of the bushfire-spread model, these 35 types were grouped into four classes, class 0 being water areas, i.e., areas considered unburnable; class 1 being grassland areas, which are easier to burn; class 2 being forest areas, which are relatively less easy to burn compared to grassland areas; and class 3 being urban areas, whose burning rate is mainly influenced by wind speed. The corresponding bushfire-spread model was then chosen according to the classification. By setting up bushfire models for four different high-bushfire areas in this way, a reference for fire safety in NSW is provided.

3. Results

3.1. The Relationship between Bushfires and Their Different Influencing Factors

To visualise the relationship between bushfire variability and its drivers, this paper examined the extent to which seven categories of drivers (temperature, precipitation, soil moisture, wind speed, slope, NDVI and vegetation flammability) influence bushfires, using the bushfire-spread area and bushfire annual frequency as the dependent variables. These seven influencing factors are usually considered to impact bushfires significantly. However, with changes in geographical extent and time scale, the forces of the influencing factors may also change [18,19]. Therefore, the study first processed the bushfire influencing factors' data so that they correspond to the bushfire data at the same point in time. Data visualisation was then used to illustrate the linear relationship and correlation between the bushfire data and the seven influencing factors' data. This was conducted to explore the relationship between the bushfires that occurred in NSW and these influencing factors. The data visualisation results are shown below (Figures 5–8).

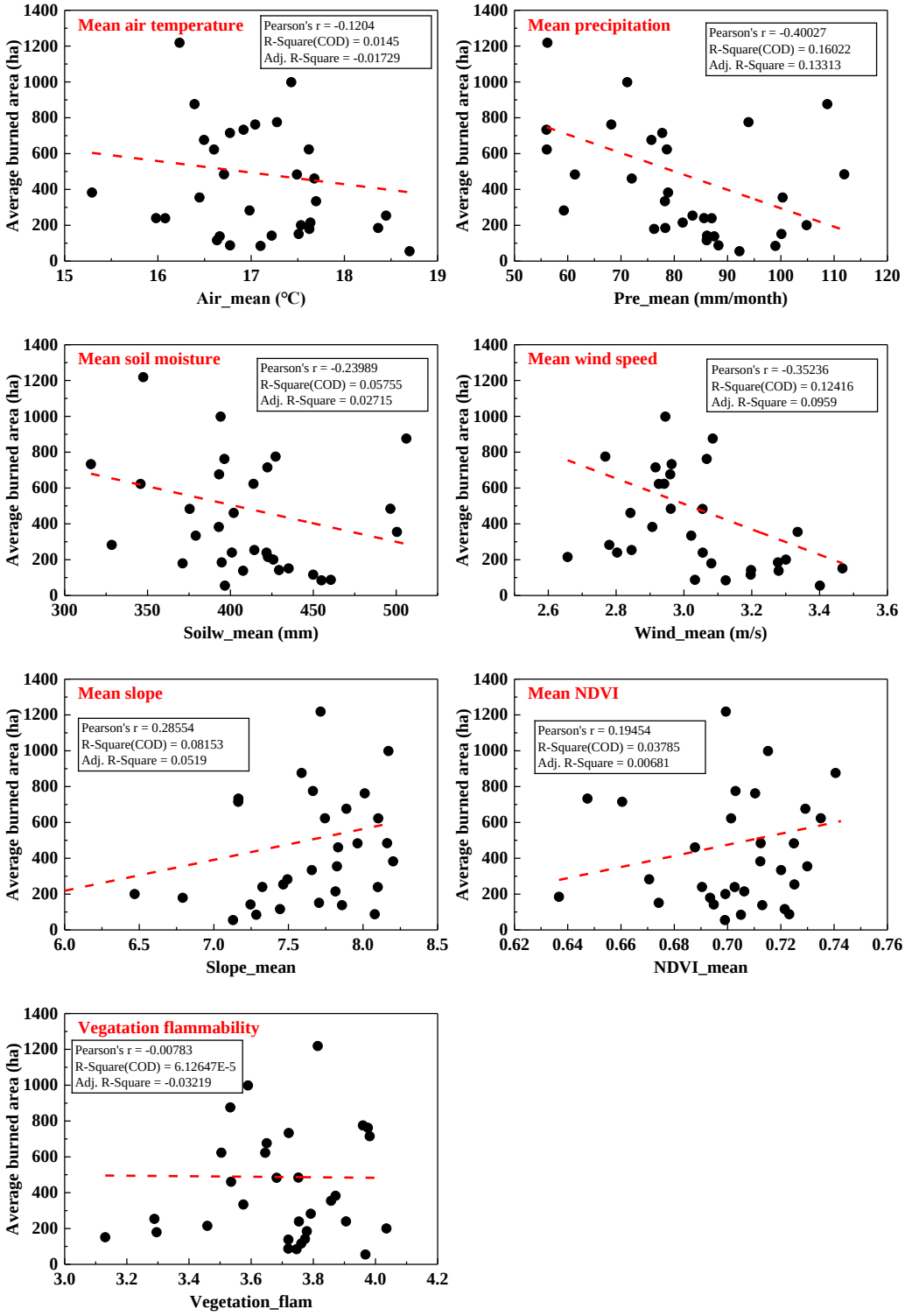

Figure 5. Scatter plot of the area burned by bushfires as a function of seven categories of influencing factors' variations: mean air temperature (denoted as Air_mean), mean precipitation (denoted as Pre_mean), mean soil moisture (denoted as Soilw_mean), mean wind speed (denoted as Wind_mean), mean terrain slope (denoted as Slope_mean), mean NDVI (denoted as NDVI_mean) and vegetation flammability (denoted as Vegetation_flam).

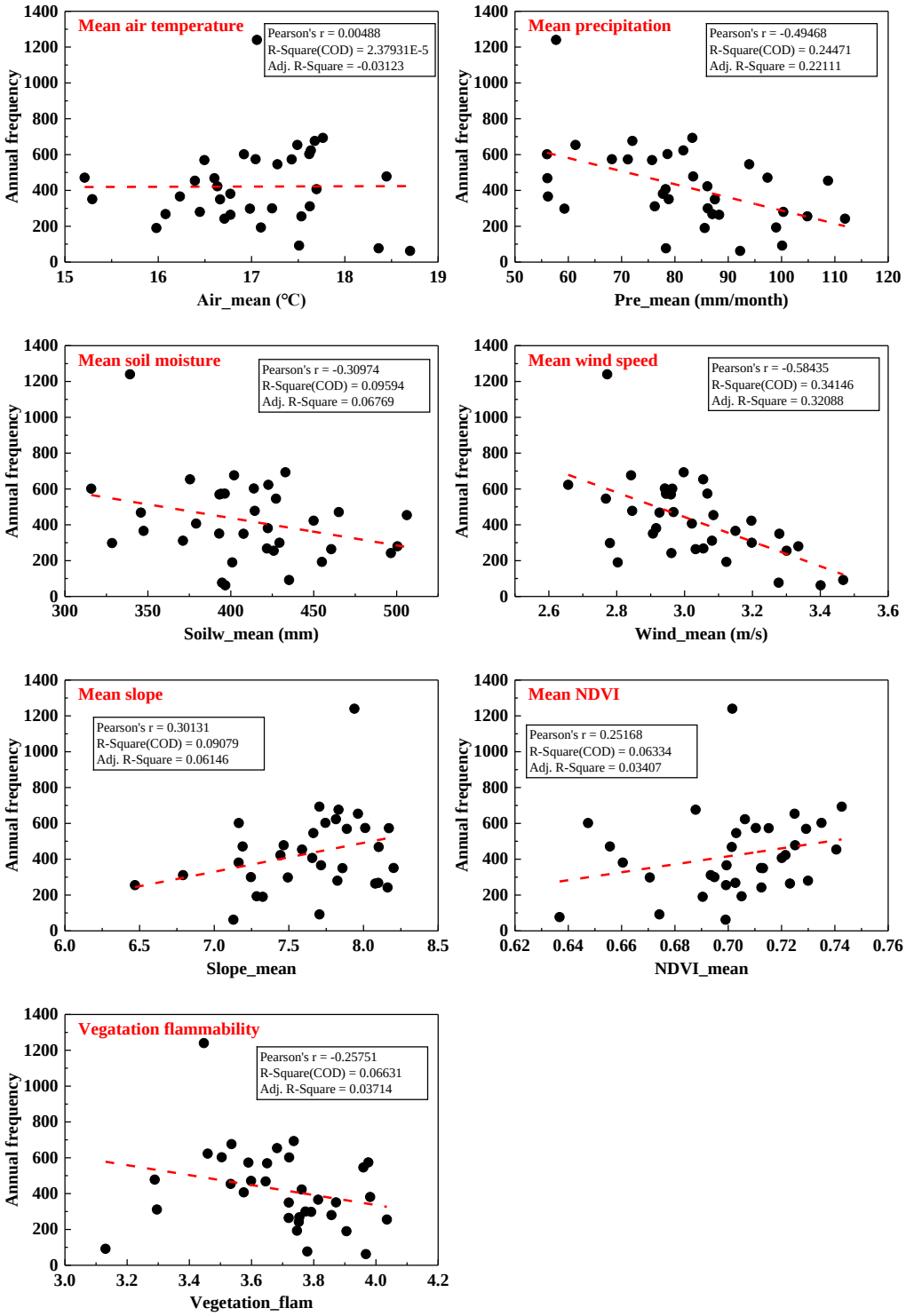

Figure 6. Scatter plot of the annual frequency of bushfires as a function of seven categories of influencing factors' variations (temperature, precipitation, soil moisture, wind speed, terrain slope, NDVI and vegetation flammability).

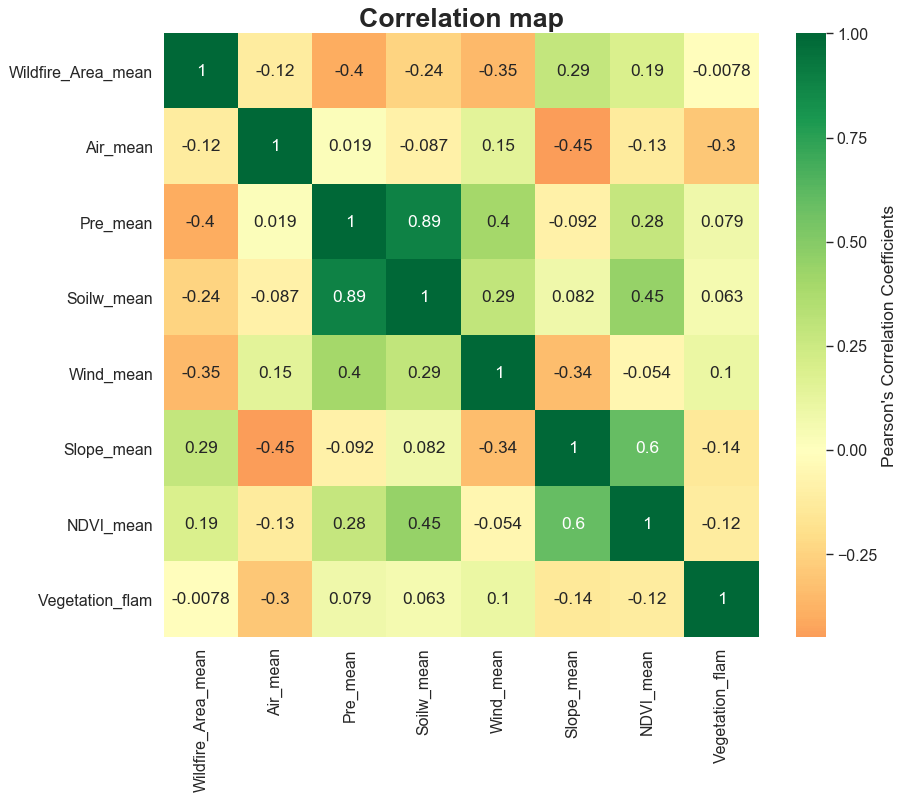

Figure 7. Pearson correlation coefficients of seven categories of influencing factors on the bushfire burned area.

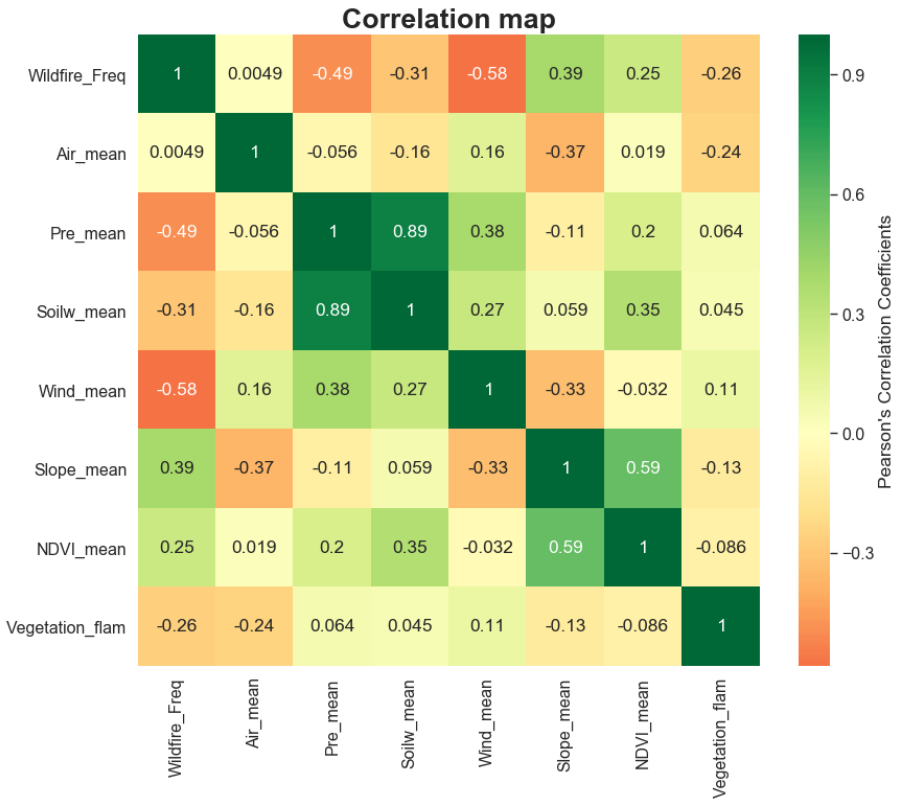

Figure 8. Pearson correlation coefficients of seven categories of influencing factors on the bushfire annual frequency.

3.1.1. Linear Regression Analysis between Bushfires and Influencing Factors

Due to the high number of bushfires in New South Wales over four decades, multiple bushfires can occur at the same point in time. Therefore, this study used the average area spread of bushfires per year for this analysis. Figure 5 presents the variation and linear fit of the average burned area of bushfires over 40 years in New South Wales with seven categories of influencing factors (temperature, precipitation, soil moisture, wind speed, topographic slope, NDVI and vegetation flammability).

In the present work, the linear regression method was used to preliminarily analyse the relationships between bushfires and their seven influencing factors. Compared with other factors (with an R-square much less than 0.1), the linear fit of the fire-spread area to precipitation and wind speed was relatively significant, with R-squares of 0.16 and 0.12, respectively. In general, the linear relationship was not strong. However, based on the positive and negative slope of the linear fit, it was possible to roughly determine the trend between fire spread and these seven influencing factors. The average burned area of bushfire decreased with increasing mean annual temperature, precipitation, soil moisture and wind speed; the average burned area of bushfire increased with increasing slope and NDVI values. Besides, the correlation between bushfire-spread area and the classification of vegetation combustibles was very slight.

Similarly, the study considered the linear relationship between the annual frequency of bushfires and the influencing factors over four decades in New South Wales. Figure 6 illustrates the variation in the annual frequency of bushfires with the seven influencing factors. A linearly fitted curve was also performed for each influencing factor to show the trend. Mean annual temperature, soil moisture and NDVI showed very weak linear fits to bushfire frequency. Similar to analysis of the average burned area of bushfires, precipitation and wind speed showed relatively high linear fits to bushfire frequency compared with other factors. However, their R-square was below 0.3, which is still very weak. Overall, there was a weak linear relationship between the seven driver categories and bushfire frequency per year. Bushfires are all caused by complex interactions of different drivers at different spatial and temporal scales [47,48]. This study, therefore, proceeds to calculate Pearson's correlation coefficients between all the influencing factors, as well as between them and the spread and frequency of bushfires in the same spatial and temporal context.

3.1.2. Pearson Correlation Analysis between Bushfires and Influencing Factors

The correlation coefficients between the bushfire burned area and the seven influencing factors are given in Figure 7. The trend between them can be judged according to the positive and negative correlation coefficients. Depending on the absolute value of the correlation coefficient, the correlation strength can be judged. The average precipitation, wind speed and slope correlated with fire spread, whereas the other four influencing factors showed a very low correlation with fire spread. There were different interactions between each of the seven influencing factors.

Although the relationships between bushfires and their factors are not linear, the linear fitting results could preliminarily provide their gross variation tendency. As shown in Figure 7, the area burned by bushfires decreased with air temperature when considering the average change over the year. However, according to Figure 6, as the temperature increased, the annual frequency of bushfires increased. This phenomenon may be because when the temperature is higher, the water content of combustible materials is lower, and the more likely it is that combustible materials ignite. Frequent fires consumed much combustible material, and therefore, bushfire burn areas became small due to a loss of combustible material. This means that when the year's average temperature is higher, the areas of the fires may tend to be small, but the fire frequency could be high throughout the year.

Furthermore, bushfire burned areas showed a strong negative correlation with precipitation and soil moisture, suggesting that years with more precipitation or soil moisture were less likely to have large fires. Adequate precipitation and moist soils ensure that the

forest area is at a good level of moisture and that the water content of combustible material is high, thus inhibiting the spread of fires.

There was a negative correlation between bushfire burned area and wind speed. When bushfires occur, microclimates are created, which have a greater influence on bushfires' occurrence and behavioural characteristics than macroclimates as the mechanisms of bushfire internal interactions are highly complex. According to Figure 6, as the wind speed increased, the annual frequency of bushfires decreased. High wind speed would make the small flame blow off [49], leading to difficulty in igniting the bushfire. Therefore, the annual frequency of bushfires decreases and the area burned by bushfires also decreases. In addition to wind speed, fire spread is also influenced by wind direction and topography. Windward fires spread faster than upwind fires, with sidewind fires in between [50]. According to Figure 7, the area burned by bushfires was positively correlated with slope. Fire fronts advance faster on uphill slopes and slower on downhill slopes, thus influencing the fire's extent [1].

In addition, NDVI was positively correlated with the area burned by bushfires, whereas there was a very weak correlation between vegetation flammability and bushfire area burned. The study used NDVI values to represent the growth of vegetation, with higher values indicating greater plant growth. Vegetation flammability was classified according to vegetation type, with higher values being more flammable [51]. Vegetation is the essential condition for the burning of forests. Areas rich in the accumulation of combustible material are conducive to the spread of fire. Different shapes, structures, sizes and water contents of combustible materials lead to different scales of fire spread [3].

The analysis results are shown in Figure 8, using annual bushfire frequency as the study parameter. In terms of climatic factors, there was a weak positive correlation between the average annual temperature and the annually occurring frequency of bushfires. Temperature changes affect the water content of forest combustibles. The higher the temperature, the lower the water content of combustible material and the more likely it is to ignite, especially dead leaves and fine combustible material in ditches and pond meadows, barren slopes, etc. [52]. In addition, bushfires usually burn due to natural causes or anthropogenic fires. In the case of anthropogenic ignition, the effects of climatic and geological factors are attenuated.

Mean annual precipitation and soil moisture were negatively correlated with the frequency of bushfires, whereas there was a strong positive correlation between precipitation and soil moisture. Precipitation can increase the water content of forest combustibles, the relative humidity of the air and soil moisture. Although suitable precipitation and soil moisture promote vigorous plant growth, they also accumulate a certain amount of combustible material for forest fires. Therefore, air and soil moisture levels can drop and dehydrate plants into flammable material without a stable amount of precipitation, thus increasing the probability of bushfires.

Wind interacts with other influences to affect the occurrence of bushfires [50]. The data used in the study were monthly average wind speed data, which were negatively correlated with the annual frequency of bushfires. High wind speed would easily blow out a small flame [49]. It is challenging for a bush to be ignited in a high wind environment, which limits the outbreak of bushfires.

The slope was positively correlated with the annual frequency of bushfires. Slope size directly affects the change of moisture in combustible materials [53]. With high or steep slopes, moisture residence times are short, and combustibles tend to dry out, thus making them more susceptible to bushfires. Conversely, a gentle slope with a long moisture residence time, moist forest floor and high moisture content of combustibles reduce the probability of fire.

In addition, as shown in Figure 8, NDVI was positively correlated with annual bushfire frequency. Vegetation flammability was negatively correlated with both bushfire frequency and NDVI. Since NDVI can be used to estimate the density of vegetation on an area of land [54], the negative correlation between vegetation flammability and NDVI shows that

the higher the vegetation flammability, the lower the density of vegetation. This may imply that areas with high vegetation flammability are affected by fire and exhibit poorer vegetation growth. In addition, due to the deterioration of vegetation growth, the amount of combustible fire material is reduced, thus reducing the frequency of bushfires. This trend may occur in a cyclical cycle, whereby when fire frequency stabilises, vegetation would grow luxuriantly, and the area may again enter a flammable state.

Overall, the above analysis confirms the relevance of the influencing factors to bushfires and their relationship with each other. Seven influencing factors interact with each other and influence the occurrence and spread of a bushfire at the same spatial and temporal scale. This complicates the identification of the relative importance of each influencing factor on bushfires. In order to analyse the extent to which they each influence bushfire development and understand the importance of these influences on bushfires, the study gives further results in Section 3.2 through a stochastic forest regression model.

3.2. Random Forest Regression Model—Analysis of Relative Importance of Influencing Factors

3.2.1. Multicollinearity Test Result

Figures 9 and 10 show the results of the seven bushfire influencing factors on bushfire burned area and annual frequency, respectively. The two bar charts were very similar. Most of the VIF values were close to 1, indicating no complete or near complete linear relationship between most bushfire drivers. In contrast, they had slightly higher VIF values than the other drivers due to the strong correlation between precipitation and soil moisture. Overall, all VIF values were less than 10, which means no multicollinearity between these seven bushfire influencing factors [39].

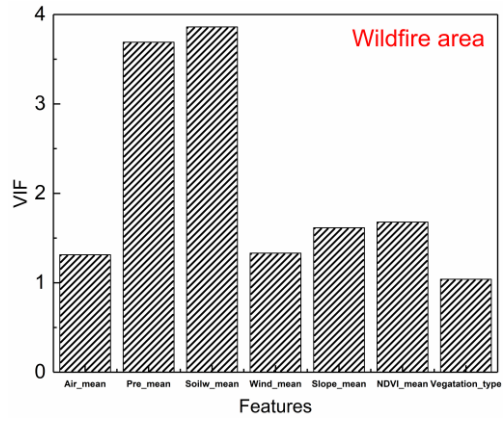

Figure 9. VIF values for the seven drivers on the area of bushfire spread.

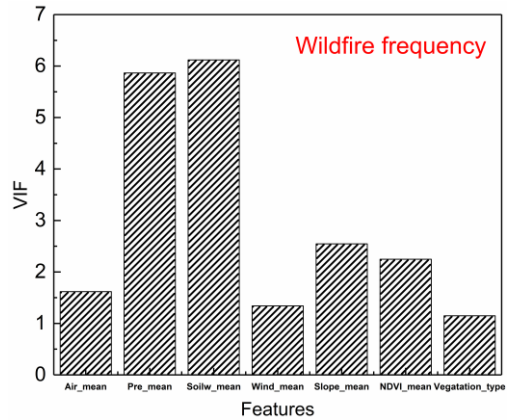

Figure 10. VIF values for the seven drivers on the annual frequency of bushfires.

3.2.2. Relative Importance of Influencing Drivers

After testing for multicollinearity in the independent variables, this study used random forest regression to model the area's annual frequency of bushfire based on 40 years of bushfire occurrence in NSW with the help of the random forest class library in scikit-learn [55]. The scikit-learn is a free machine learning library for the Python programming language, which features various classification, regression and clustering algorithms. The number of learners (n_estimators) was set to 400, the maximum depth of the decision tree (max_depth) was set to 10 and the minimum number of internal node subdivision samples (min_samples_split) was set to 5. The regression model of bushfire-spread area was obtained by setting max_leaf_nodes to 60. In order to determine the importance of each driver in the area of bushfire spread, the importance values can be checked by using "feature_importance_" in scikit-learn, which is a permutation test to find the importance of the feature points.

The random forest regression model can quantify the relative importance of each bushfire influencing driver. Relative importance for bushfire area and annual frequency are shown in Figures 11 and 12, respectively. It can be seen that the relative importance of the seven drivers for bushfire area and annual frequency were different. The top three important drivers for bushfire area were wind speed, air temperature and soil moisture, which together account for more than 60% of the importance, indicating that they play important roles in determining the spread of bushfires (Figure 11). As for the annual frequency of bushfire, its top three important drivers were wind speed, precipitation and vegetation flammability, which was slightly different from those for bushfire area. In general, wind speed was the most important influencing factor for both bushfire area and frequency.

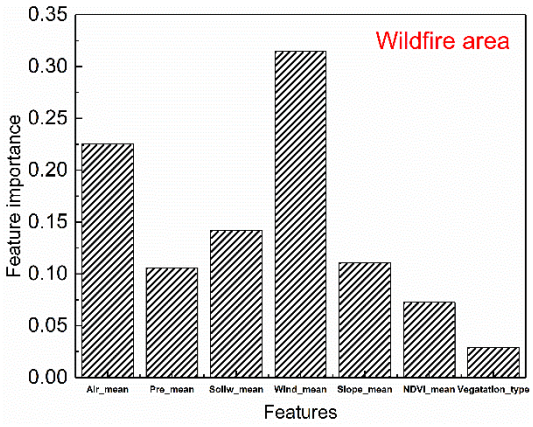

Figure 11. Importance of each driver relative to the area of bushfire spread.

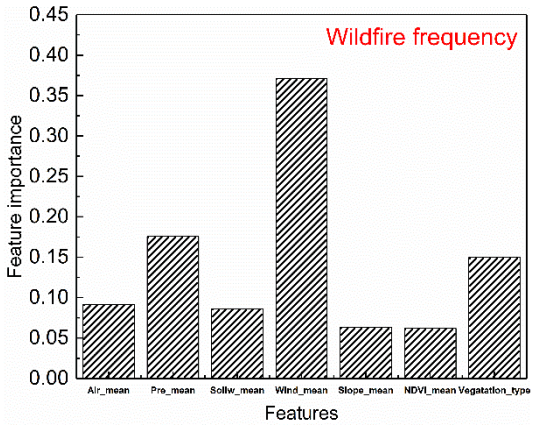

Figure 12. Importance of each driver relative to the annual frequency of bushfire.

3.3. Spark-Based Predictions for Future Spatial Pattern of Bushfires

The above analysis showed the relationship between bushfires and their influencing drivers. Next, the predictions of bushfire spread for four hotspot regions are demonstrated in this subsection.

3.3.1. Armidale Region

The ignition in this area was located to the southeast of the main town of Armidale, and the simulation results of the spread of the fire within 24 h in this area is shown in Figure 13. The direction of fire spread was mainly influenced by the direction of the wind and tended to spread from northwest to southeast. As shown in Figure 13b,c, this fire area's land type and topography are relatively uniform, with the majority being grassland and mountainous types; thus, the fire spread outwards with a relatively uniform flame front. The spread rate of the fire into the urban area, although also present, was relatively small. Calculations estimated that in 24 h, the fire will be about to spread into the forested area to the southeast. In order to avoid a larger bushfire, it is, therefore, necessary to contain this type of fire within 24 h.

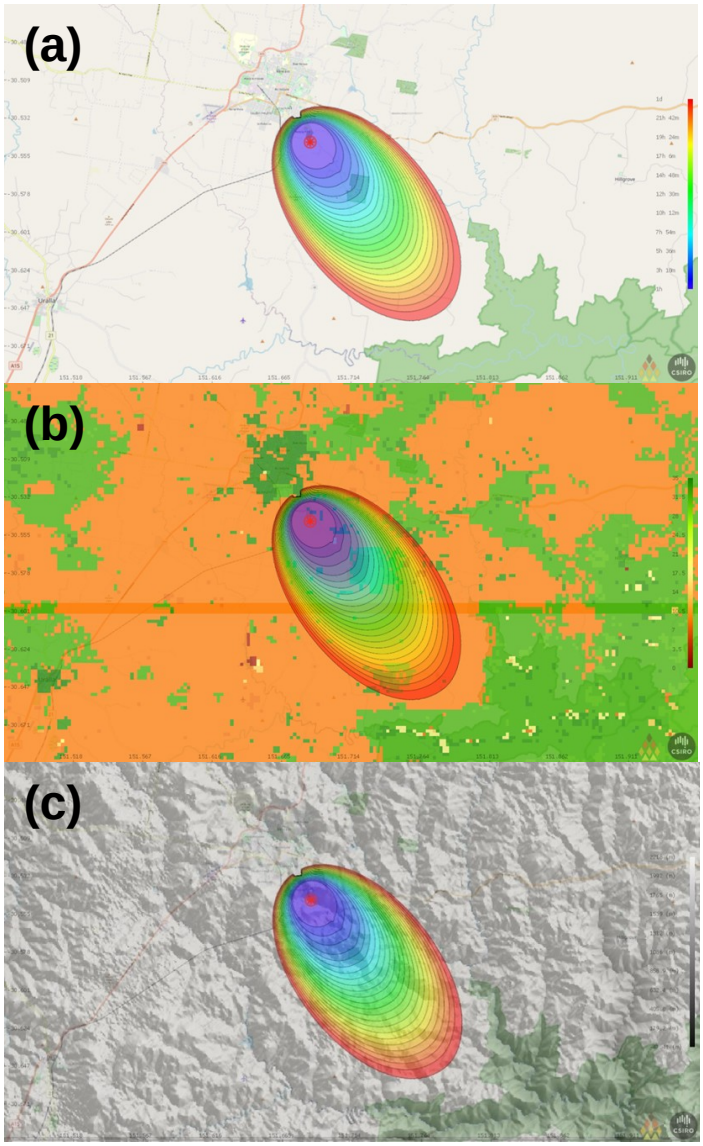

Figure 13. Simulation results of fire spread in the Armidale region over a 24 h period: (**a**) basemap; (**b**) with vegetation-type basemap; (**c**) with terrain basemap.

3.3.2. Nyngan Region

The location of the fire was set to the southwest of Nyngan airport, and Figure 14 shows the spread of the fire within 24 h of its appearance in the region. Similarly, the main direction of the fire spread was consistent with the direction of the wind, from northwest to southeast. Unlike the Armidale region, the flame front in this region was not smooth, mainly due to the uneven distribution of vegetation in the region, as shown in Figure 14b. Figure 14c shows the simulation results of fire spread over topographic layers in the Nyngan region over 24 h. The topography in this region is relatively uniform, and its effect on the flame front was relatively consistent. As can be seen from Figure 14a, there are two traffic roads within the 24 h flame area: Pangee Road and Tottenham Road. Based on the fire spread, it is clear that Pangee Road will be affected by this fire 3–4 h after the fire has occurred; Tottenham will also be on fire 13–14 h after the fire has occurred. It is, therefore, necessary to control the corresponding roads in time to ensure the safety of life in the event of a fire.

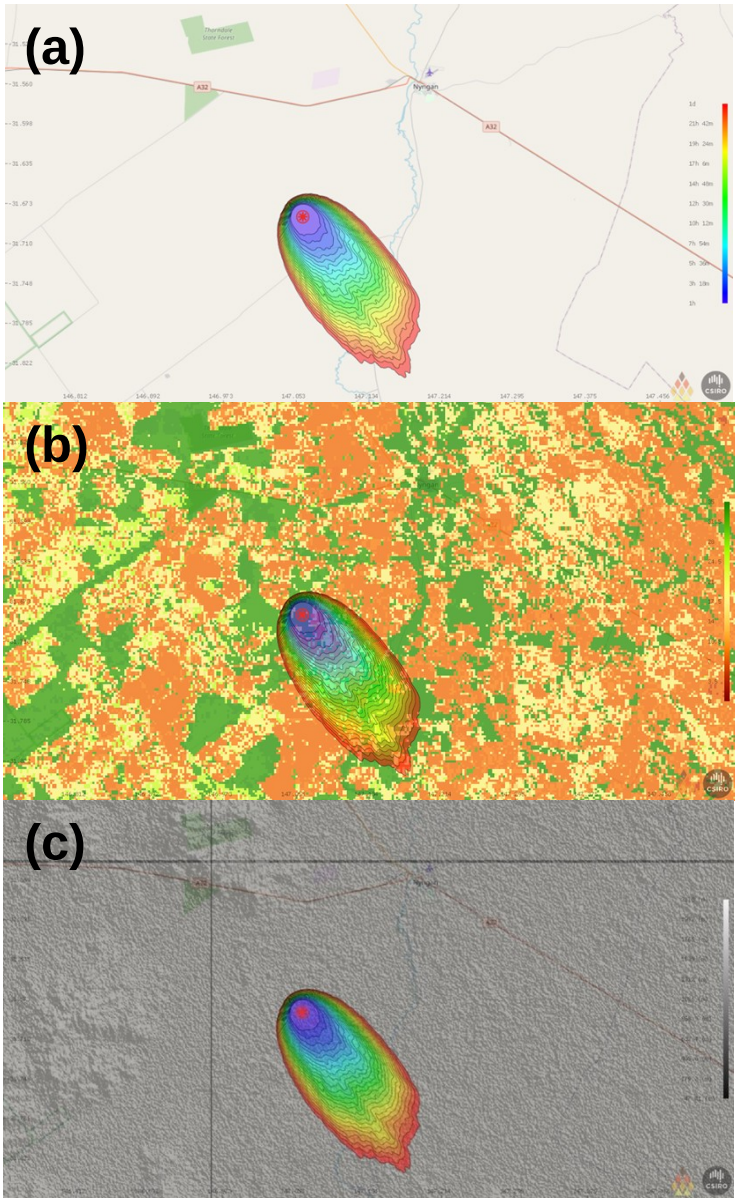

Figure 14. Simulation results of fire spread in the Nyngan region over a 24 h period: (**a**) basemap; (**b**) with vegetation-type basemap; (**c**) with terrain basemap.

3.3.3. Queanbeyan Region

The ignition area was located to the southeast of the main urban area of Queanbeyan, with the wind direction set to the northwest. In the vicinity of this ignition area, there are various types of land classification, including towns, grasslands, and waters. Figure 15 shows the 24 h spread of the fire. To the northwest of the ignition site, the fire spread upwind and towards the city, with a prolonged rate of spread and essentially no tendency to spread northwest. In contrast, the fire spread more rapidly to the southeast of the fire site. As shown in Figure 15b,c, the fire was held back in the southwest direction due to the obstruction by the water. The southeast direction is where the main residential areas within Queanbeyan Region are gathered. Therefore, it is imperative to contain the fire and evacuate the inhabitants promptly within 24 h of the fire.

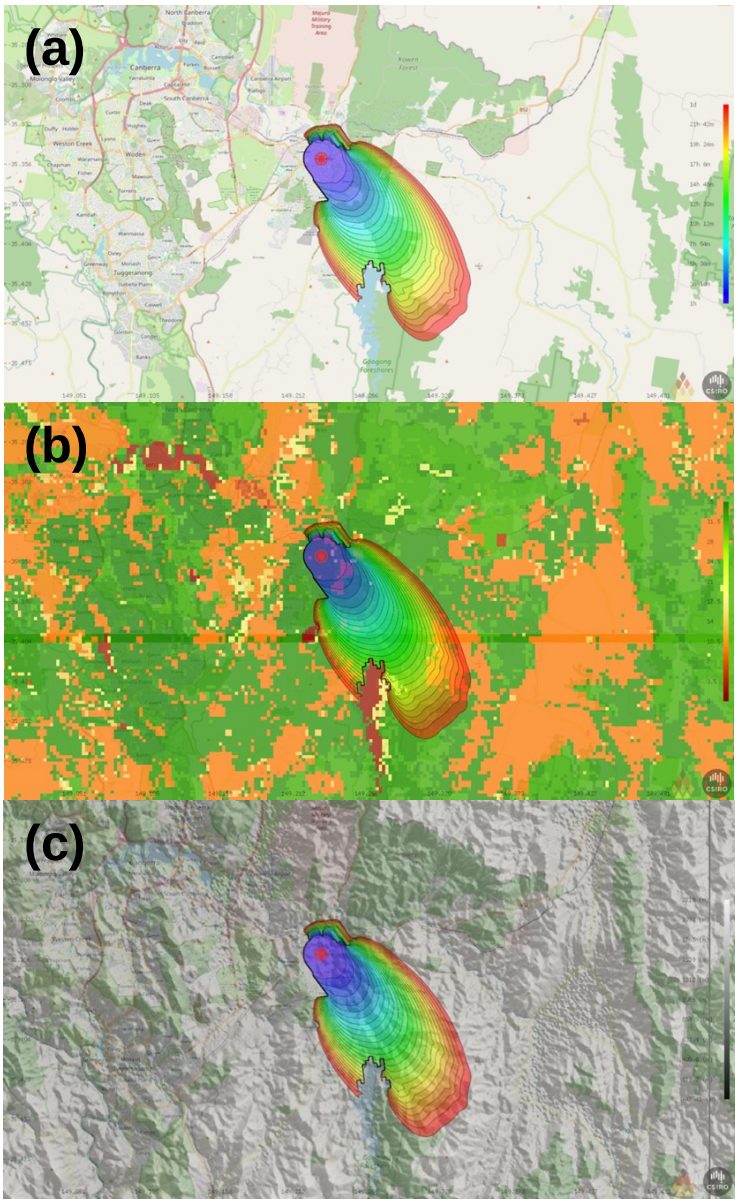

Figure 15. Simulation results of fire spread in the Queanbeyan region over a 24 h period: (**a**) basemap; (**b**) with vegetation-type basemap; (**c**) with terrain basemap.

3.3.4. Wauchope Region

The Wauchope region is located on the coast and is also a high-risk area for bushfires. Similarly, fires in the Wauchope region gradually spread towards the sea in a northwesterly

wind direction setting. Figure 16 shows the simulation of fire spread in the Wauchope region, with the fire front spreading right up to the sea after 24 h. Similar to the Queanbeyan region, the fire spread very slowly in the northwest direction due to the northwest wind, whereas the fire spread faster in the southeast direction. Figure 16b shows a layer of land classification, which shows that the land classification in this burning area is relatively uniform and, therefore, resulted in a more rounded fire front in the simulation. Four to five hours after the fire started at this fire site, the flames spread to the vicinity of the water. Although water is not combustible, fires can spread around water. Figure 16c shows the results of fire spread over the topography for the Wauchope region, from which it can be seen that topography had little effect on fire spread.

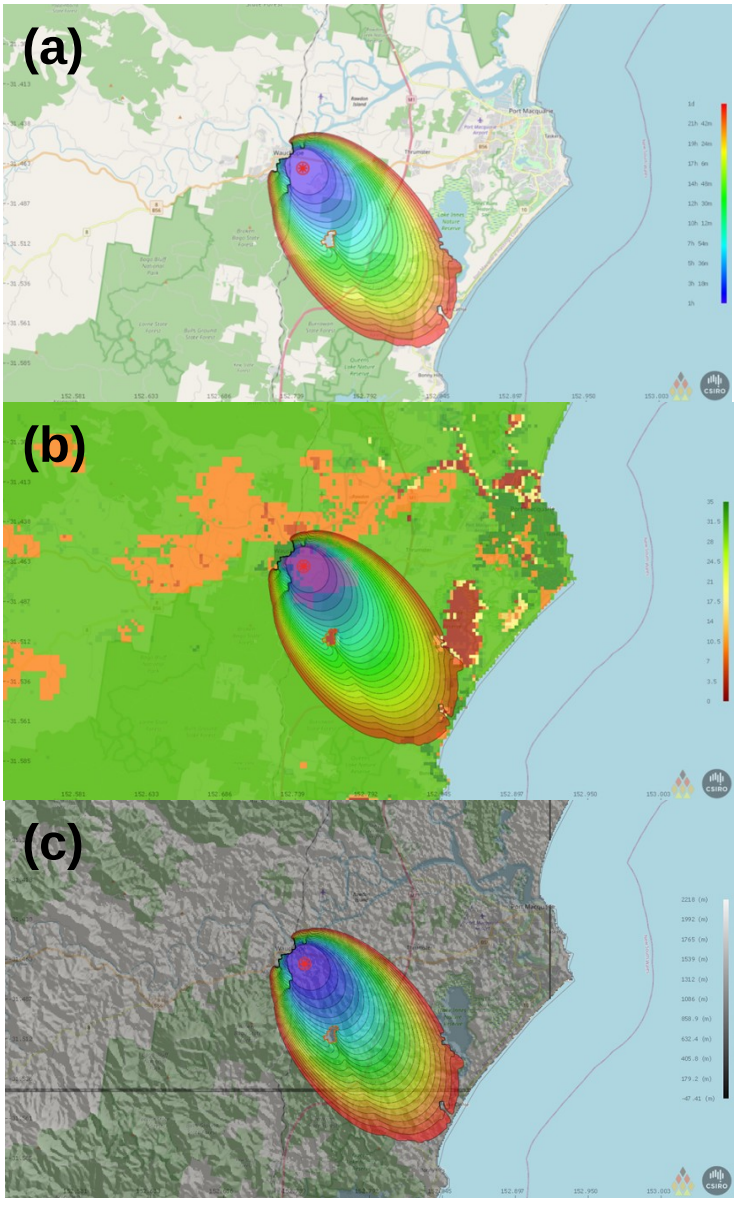

Figure 16. Simulation results of fire spread in the Wauchope region over a 24 h period: (**a**) basemap; (**b**) with vegetation-type basemap; (**c**) with terrain basemap.

4. Discussion

4.1. The Relationship between Bushfires and Their Different Influencing Factors

The study utilised Pearson correlation coefficients to further explore the mechanisms of interaction between the bushfire drivers and how they cooperate to drive bushfires. As

for the area burned by bushfires, the analysis found that the variations of bushfire spread along its drivers are non-linear, and the trends are complex, making it difficult to describe them quantitatively. However, it is clear that the probability of occurrence of large-scale bushfires is higher at higher average temperatures (>14 °C), lower average precipitation (<75 mm/month), lower soil moisture (<400 mm), lower wind speeds (<4 m/s) and higher NDVI (>0.5).

Besides, the correlation map of bushfire annual frequency and influencing drivers shows that there was only a weak positive correlation with air temperature and NDVI, a moderate positive correlation with slope, and a moderate negative correlation with precipitation, soil moisture and wind speed. The air temperature and NDVI provide dry combustibles for bushfire, which helps with the occurrence and spread of bushfire. Besides, fire fronts advance faster on uphill slopes and slow down on downhill slopes; for every 10-degree increase in slope, the speed of the fire front doubles [1]. Therefore, changes in slope and slope direction will affect the spread of bushfires. Moreover, there are the negative correlations between soil moisture, precipitation and the annual frequency of bushfires, which keeps the previous hypothesis consistent that the higher the soil moisture, the lower the bushfire occurrence. Based on the analysis in Section 3.1, it is clear that the physical properties of the soil are reduced between bushfire burns, resulting in a reduction in infiltration rates and vegetation growth rates. This affects the accumulation of fuel. Soil moisture is also relatively low in areas that have been burned, and there is a lack of combustible material for burning. The wind is also a relatively strong factor for bushfire. On the one hand, it can increase the possibilities of bushfire occurrence by removing water vapour from the forest, reducing the humidity of the air within the forest and replenishing oxygen to the fire. But on the other hand, high wind speed can largely reduce the temperature of forest combustibles due to convective heat transfer and evaporation heat loss, which reduces the possibility of bushfire occurrence. In this work, it was found that the general effect of wind speed on bushfire annual frequency is negative. In general, the combined relevance of all drivers was calculated here in the same scenario so that the influence of other drivers must be taken into account when analysing individual parameters.

We recommend that future studies consider obtaining more accurate meteorological and natural factors' data (i.e., monthly or quarterly data) for specific regions and incorporate human factors (i.e., population density, building density) to investigate in greater depth the mechanisms of interaction between influences and how they synergistically drive bushfires.

4.2. Random Forest Regression Model—Analysis of Relative Importance of Influencing Factors

The spatial pattern of bushfire areas is driven by many factors such as climate, topography and vegetation, but there are differences in the contribution of different factors. A random forest regression model was used to calculate the relative importance of each driver concerning the area and annual frequency of bushfire spread. The analysis revealed that the top three drivers of importance on bushfire area were wind speed, air temperature and soil moisture, with the sum of their importance exceeding 60%, indicating that they play a decisive role in the spread of bushfire trends. As for bushfire annual frequency, its top three important drivers were wind speed, precipitation and vegetation flammability, which was slightly different from those for bushfire area. For both of them, the wind speed was the most important driver. Wind works to start fires by blowing flames into fresh fuel, bringing it to the ignition point and providing a continuous supply of oxygen. Wind also ignites new fires by dissipating burning embers into the air. At the same time, wind can alter thermal convection and increase the horizontal flow of heat, significantly increasing the heat in front of the fire head and accelerating its spread. However, high wind speed can largely reduce the temperature of forest combustibles due to convective heat transfer and evaporation heat loss, which reduces the possibility of bushfire occurrence.

Besides wind speed, the other two important drivers for bushfire area are air temperature and soil moisture, which is different from that for bushfire annual frequency. The

difference between bushfire area and bushfire annual frequency is that the bushfire area is affected by the fire spreading, whereas bushfire annual frequency is more impacted by the fire ignition. Air temperature changes affect the water content of forest combustibles. The higher the temperature, the lower the water content of combustible material and the drier it becomes, and therefore, the easier it is for combustible material to ignite and the easier it is for bushfires to occur once an ignition source is present. Soil moisture is an environmental factor sensitive to climate change and is influenced by temperature and rainfall. Analysis of soil moisture provides a clear picture of regional climate trends. Suitable soil moisture also means good vegetation growth, as well as combustible material accumulation. Therefore, these two drivers are more important for fire spreading (bushfire area). Precipitation directly extinguishes small fires and lowers down the frequency of fire occurrence. Vegetation flammability represents the flammability of burn materials, which determines whether the fire can be ignited. Therefore, this shows that precipitation and vegetation flammability are two drivers that have relatively large impacts on bushfire annual frequency.

4.3. Spark-Based Predictions for Future Spatial Pattern of Bushfires

These predictions presented a 24 h simulation of bushfire spread in each of the four hot spots of bushfire occurrence in NSW based on four decades of bushfire hotspot analysis with the Spark platform. Although the wind speed, temperature and land classification were different, the direction of fire spread was relatively consistent, which indicates that the wind direction directly influences the main direction of fire spread; if we compare the flame fronts of bushfires in the four hot spots, we can find that their flame fronts have different shapes, which is mainly due to the influence of land classification. Different land classifications correspond to different rates of fire spread, and the difference in spread rate affects the smoothness of the flame front. When the type of land classification is relatively homogeneous, the overall shape of the fire front is relatively smooth (e.g., the Armidale Region); when the type of land classification is complex, and the rate of flame spread varies significantly from region to region, the shape of the fire front is relatively complex and unsmooth (e.g., the Queanbeyan Region).

In addition, the 24 h flame-spread range of the four hotspot areas can be used as an early warning or evacuation area for emergency response. This allows for the evacuation of residents in potential bushfire-spread areas within the first 24 h of a bushfire to achieve minimal damage to people and property.

For future studies based on this research, we would suggest including specific flora and fauna species and their habitat into the fire-spread simulations. This would enable fire response and environmental agencies to understand the likely impact on species and to put in place procedures and strategies to manage this and mitigate this impact.

5. Conclusions

In this work, the historical bushfire data, meteorological data and terrain data were analysed to figure out the interactions between bushfires and their seven drivers. To predict the burning spread of bushfire within 24 h at four hotspot regions, the simulations were carried out using the Spark toolkit. The following are detailed procedures and conclusions.

Firstly, we used correlation analysis methods to explore the relationships between bushfires and their influencing drivers with 40 years of historical bushfire data, climate data, wind data, soil moisture data, topographic data and vegetation data from NSW. The data collected (see Appendix A) included two characteristics of bushfire (area and annual frequency of bushfire) and seven bushfire drivers (air temperature, precipitation, soil moisture, wind speed, slope, NDVI and vegetation flammability). It was found that the probability of occurrence of large-scale bushfires is higher at higher average temperatures (>14 °C), lower average precipitation (<75 mm/month), lower soil moisture (<400 mm), lower wind speeds (<4 m/s) and higher NDVI (>0.5).

Due to the complex relationships between bushfires and their drivers, the random forest regression method was applied in this work to figure out the relative importance

of these drivers. It showed that the top three drivers for bushfire area are wind speed, air temperature and soil moisture. For bushfire annual frequency, the top three drivers are wind speed, precipitation and vegetation flammability. The difference between bushfire area and bushfire annual frequency is that the bushfire area is affected by the fire spreading, whereas bushfire annual frequency is more impacted by the fire ignition. The wind speed is the most important driver for both bushfire area and annual frequency, because wind has important effects both on fire spreading and fire ignition. The air temperature and soil moisture are two factors that have relatively long durations. Therefore, they have less effects on fire ignition but have important effects on bushfire area (fire spreading). However, precipitation and vegetation flammability are two factors that determine the flammability of burn material, which has relatively large impacts on bushfire annual frequency (fire ignition).

Furthermore, after quantifying the contributions of the key influencing drivers, predictions of fire spreading in four hot spots in NSW were simulated using the CSIRO's Spark toolkit. We found that the main direction of bushfire spread was directly influenced by the wind direction, whereas the land classification mainly influenced the shape of the bushfire front. Different land classifications corresponded to different rates of fire spread, and differences in spread rates affected the smoothness of the fire front. The 24 h flame-spread range of this hotspot area can be used as an early warning or evacuation area for emergency response and can be used to warn and evacuate people within the simulated spread area of a bushfire within 24 h of its occurrence in order to minimise damage.

This study provides a reference for fireproofing agencies to use the bushfire simulation model to more accurately delineate fire-sensitive areas to help them better develop fire management policies and measures for bushfires.

Author Contributions: Conceptualization, Wanqin He and Sara Shirowzhan; methodology, Wanqin He and Sara Shirowzhan; software, Wanqin He; formal analysis, Wanqin He and Sara Shirowzhan; investigation, Wanqin He, Sara Shirowzhan and Christopher James Pettit; resources, Christopher James Pettit; data curation, Wanqin He; writing—original draft preparation, Wanqin He and Sara Shirowzhan; writing—review and editing, Wanqin He, Sara Shirowzhan and Christopher James Pettit; visualization, Wanqin He and Sara Shirowzhan; supervision, Sara Shirowzhan and Christopher James Pettit. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used for analysis in the study were drawn from publicly accessible datasets published by various research institutions as well as official sources, which are indicated in the manuscript. If you have any queries about the data, you can request the data for this study from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The processed bushfire data and Jupyter codes used in this work are provided at the author's GitHub website https://github.com/ZoeHE36/Bushfire-analysis, accessed on: 31 May 2022 including the linear fitting analysis, correlation analysis and random forest regression (Sections 2.3 and 2.4).

References

1. The United Nations Office for Disaster Risk Reduction, the Centre for Research on the Epidemiology of Disasters, the Institute of Health and Society. *Economic Losses, Poverty & Disasters 1998–2017*; World Health Organization: Brussels, Belgium, 2017.
2. Burton, J. It Was a Line of Fire Coming at Us': South West Firefighters Return Home. Available online: https://www.busseltonmail.com.au/story/6620313/it-was-a-line-of-fire-coming-at-us-firefighters-return-home/ (accessed on 17 June 2021).
3. Commonwealth Scientific and Industrial Research Organisation (CSIRO). Bushfires in Australia- Prepared for the 2009 Senate Inquiry into Bushfires in Australia. Available online: https://www.aph.gov.au/DocumentStore.ashx?id=3d4e5dd5-9374-48e9-b3f4-4e6e96da27f5 (accessed on 17 June 2021).
4. Maiden, J.H. *The Forest Flora of New South Wales*; W. A. Gullick: Sydney, Australia, 1904.

5. Morrison, D.A.; Buckney, R.T.; Bewick, B.J.; Cary, C.J. Conservation conflicts over burning bush in south-eastern Australia. *Biol. Conserv.* **1996**, *76*, 167–175. [CrossRef]
6. Lucas, C.H.K.; Mills, G.; Bathols, J. *Bushfire Weather in Southeast Australia: Recent Trends and Projected Climate Change Impacts*; Bureau of Meteorology Research Centre: Melbourne, Australia, 2007.
7. Aldersley, A.; Murray, S.J.; Cornell, S.E. Global and regional analysis of climate and human drivers of wildfire. *Sci. Total Environ.* **2011**, *409*, 3472–3481. [CrossRef] [PubMed]
8. Chen, Z. Effects of fire on major forest ecosystem processes. *Chin. J. Appl. Ecol.* **2006**, *17*, 1726–1732.
9. Attiwill, P.M. The disturbance of forest ecosystems: The ecological basis for conservative management. *For. Ecol. Manag.* **1994**, *63*, 247–300. [CrossRef]
10. Flannigan, M.D.; Harrington, J.B. A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). *J. Appl. Meteorol.* **1988**, *27*, 441–451. [CrossRef]
11. Carrega, P. A meteorological index of forest fire hazard in Mediterranean France. *Int. J. Wildland Fire* **1991**, *1*, 79–86. [CrossRef]
12. Davis, F.W.; Michaelsen, J. Sensitivity of fire regime in chaparral ecosystems to climate change. *Glob. Chang. Mediterr.-Type Ecosyst.* **1995**, *117*, 435–456.
13. Gill, A.M.; Moore, P.H.R. Regional and historical fire weather patterns pertinent to the January 1994 Sydney bushfires. *Proc. Linn. Soc. N. S. W.* **1996**, *116*, 27–35.
14. Mensing, S.A.; Michaelsen, J.; Byrne, R. A 560-Year Record of Santa Ana Fires Reconstructed from Charcoal Deposited in the Santa Barbara Basin, California. *Quat. Res.* **1999**, *51*, 295–305. [CrossRef]
15. Moritz, M.A. Spatiotemporal analysis of controls on shrubland fire regimes: Age dependency and fire hazard. *Ecology* **2003**, *84*, 351–361. [CrossRef]
16. Keeley, J.E. Impact of antecedent climate on fire regimes in coastal California. *Int. J. Wildland Fire* **2004**, *13*, 173–182. [CrossRef]
17. Peters, D.P.; Pielke, R.A.; Bestelmeyer, B.T.; Allen, C.D.; Munson-McGee, S.; Havstad, K.M. Cross-scale interactions, non-linearities, and forecasting catastrophic events. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15130–15135. [CrossRef]
18. Yue, C.; Luo, C.F.; Shu, L.F.; Shen, Z.H. Advances in wildfire research under the background of global change. *J. Resour. Ecol.* **2020**, *40*, 385–401.
19. Minnich, R.A.; Bahre, C.J. Wildland Fire and Chaparral Succession Along the California Baja-California Boundary. *Int. J. Wildland Fire* **1995**, *5*, 13–24. [CrossRef]
20. Chas-Amil, M.L.; Prestemon, J.P.; Mcclean, C.J.; Touza, J. Human-ignited wildfire patterns and responses to policy shifts. *Appl. Geogr.* **2015**, *56*, 164–176. [CrossRef]
21. Rhman, S.; Chang, H.C.; Magill, C.; Tomkins, K.; Hehir, W. *Forest Fire Occurrence and Modelling in Southeastern Australia*, 1st ed.; Szmyt, J., Ed.; Intechopen: London, UK, 2018.
22. Bradstock, R.A.; Cohn, J.S.; Gill, A.M.; Bedward, M.; Lucas, C. Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather. *Int. J. Wildland Fire* **2009**, *18*, 932–943. [CrossRef]
23. Zhang, Y.; Lim, S.; Sharples, J.J. Modelling spatial patterns of wildfire occurrence in South-Eastern Australia. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1800–1815. [CrossRef]
24. Rodrigues, M.; Riva, J.; Fotheringham, S. Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Appl. Geogr.* **2014**, *48*, 52–63. [CrossRef]
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
26. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [CrossRef]
27. Andrews, P.; Finney, M.; Fischetti, M. Predicting Wildfires. *Sci. Am.* **2007**, *297*, 46–51. [CrossRef] [PubMed]
28. Gibos, K.; Slijepcevic, A.; Wells, T.; Fogarty, L. Building fire behavior analyst (FBAN) capability and capacity: Lessons learned from Victoria, Australia's bushfire behavior predictive services strategy. In Proceedings of the Large Wildland Fires Conference, Missoula, MT, USA, 29 September 2015.
29. Queensland Fire and Emergency Services. 2015–16 Annual Report. Available online: https://documents.parliament.qld.gov.au/tableOffice/TabledPapers/2016/5516T1613.pdf (accessed on 17 June 2021).
30. Neale, T.; May, D. Bushfire simulators and analysis in Australia: Insights into an emerging sociotechnical practice. *Environ. Hazards* **2018**, *17*, 200–218. [CrossRef]
31. Garg, S.; Forbes-Smith, N.; Hilton, J.; Prakash, M. SparkCloud: A Cloud-Based Elastic Bushfire Simulation Service. *Remote Sens.* **2018**, *10*, 74. [CrossRef]
32. Australian Bureau of Statistics. National, State and Territory Population—September 2020. Available online: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release (accessed on 17 June 2021).
33. CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) Version 4.04 of High-Resolution Gridded Data of Month-By-Month Variation in Climate (Jan. 1901–Dec. 2019), Centre for Environmental Data Analysis, Date of Citation. Available online: https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9 (accessed on 26 April 2021).
34. ERA5-Land Monthly Averaged Data from 1950 to Present, European Centre for Medium-Range Weather Forecasts. Available online: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=overview (accessed on 26 April 2021).

35. MODIS/Aqua Vegetation Indices Monthly L3 Global 1 km SIN Grid, USUG. Available online: https://lpdaac.usgs.gov/products/myd13a3v061/ (accessed on 26 April 2021).
36. NDVI, the Foundation for Remote Sensing Phenology, USUG. Available online: https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science_center_objects=0#qt-science_center_objects (accessed on 26 April 2021).
37. Rodgers, J.L.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [CrossRef]
38. Stigler, S.M. Francis Galton's Account of the Invention of Correlation. *Stat. Sci.* **1989**, *4*, 73–79. [CrossRef]
39. Fotheringham, A.S.; Harris, P.; Charlton, M.; Lu, B. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; John Wiley and Sons: New York, NY, USA, 2002.
40. Cressie, N.A.C. *Statistics for Spatial Data, Revised ed.*; John Wiley & Sons: Toronto, ON, Canada, 1993.
41. Breiman, L.; Cutler, R.A. Random Forests Machine Learning. *J. Clin. Microbiol.* **2001**, *2*, 199–228.
42. Fotheringham, A.S.; Brunsdom, C. Local Forms of Spatial analysis. *Geogr. Anal.* **2010**, *31*, 340–358. [CrossRef]
43. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef]
44. Visner, M.; Shirowzhan, S.; Pettit, C. Spatial Analysis, Interactive Visualisation and GIS-Based Dashboard for Monitoring Spatio-Temporal Changes of Hotspots of Bushfires over 100 Years in New South Wales, Australia. *Buildings* **2021**, *11*, 37. [CrossRef]
45. Hilton, J.; Swedosh, W.; Hetherton, L.; Sullivan, A.; Prakash, M. Spark User Guide 1.1.2. CSIRO, Australia. 2019. Available online: https://research.csiro.au/static/spark/Spark_applications_user_guide_v112.pdf (accessed on 20 March 2021).
46. Sullivan, A.L. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *Int. J. Wildland Fire* **2006**, *18*, 369–386. [CrossRef]
47. Krawchuk, M.A.; Moritz, M.A.; Parisien, M.A.; Van Dorn, J.; Hayhoe, K. Global Pyrogeography: The Current and Future Distribution of Wildfire. *PLoS ONE* **2009**, *4*, e5102. [CrossRef]
48. Archibald, S.; Lehmann, C.E.R.; Gómez-Dans, J.L.; Bradstock, R.A. Defining pyromes and global syndromes of fire regimes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6442–6447. [CrossRef]
49. Turns, S.R. *An Introduction to Combustion: Concepts and Applications*; McGraw-Hill Education: New York, NY, USA, 2012.
50. Jin, X.Z.; Han, S.T.; Cheng, B.Y. Overview of the research on the rate and pattern of forest fire spread. *For. Technol. Newsl.* **1998**, *12*, 1–3.
51. Keith, D. *Ocean Shores to Desert Dunes: The Native Vegetation of New South Wales and the ACT*; Department of Environment and Conservation (NSW): Parramatta, Australia, 2004.
52. Hoffmann, W.A. Regional feedbacks among fire, climate, and tropical deforestation. *J. Geophys. Res. Planets* **2003**, *108*, 4721. [CrossRef]
53. Ju, E.D. Forest fires spread. *For. Fire Prev.* **1986**, *2*, 31–33.
54. Schinasi, L.H.; Benmarhnia, T.; Roos, A.J.D. Modification of the association between high ambient temperature and health by urban microclimate indicators: A systematic review and meta-analysis. *Environ. Res.* **2018**, *161*, 168–180. [CrossRef]
55. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.