

Article

Effect of Socioeconomic Variables in Predicting Global Fire Ignition Occurrence

Tichaona Mukunga ^{1,*}, Matthias Forkel ², Matthew Forrest ³ , Ruxandra-Maria Zotta ¹, Nirlipta Pande ¹, Stefan Schlaffer ¹ and Wouter Dorigo ^{1,*} 

¹ Climate and Environmental Remote Sensing Research Unit, Department of Geodesy and Geoinformation, Technische Universität Wien (TU Wien), Wiedner Hauptstraße 8-10, 1040 Vienna, Austria

² Institute for Photogrammetry and Remote Sensing, Technische Universität Dresden (TU Dresden), 01062 Dresden, Germany

³ Senckenberg Biodiversity and Climate Research Centre (SBiK-F), 60325 Frankfurt am Main, Germany

* Correspondence: tichaona.mukunga@tuwien.ac.at (T.M.); wouter.dorigo@tuwien.ac.at (W.D.)

Abstract: Fires are a pervasive feature of the terrestrial biosphere and contribute large carbon emissions within the earth system. Humans are responsible for the majority of fire ignitions. Physical and empirical models are used to estimate the future effects of fires on vegetation dynamics and the Earth's system. However, there is no consensus on how human-caused fire ignitions should be represented in such models. This study aimed to identify which globally available predictors of human activity explain global fire ignitions as observed by satellites. We applied a random forest machine learning framework to state-of-the-art global climate, vegetation, and land cover datasets to establish a baseline against which influences of socioeconomic data (cropland fraction, gross domestic product (GDP), road density, livestock density, grazed lands) on fire ignition occurrence were evaluated. Our results showed that a baseline random forest without human predictors captured the spatial patterns of fire ignitions globally, with hotspots over Sub-Saharan Africa and South East Asia. Adding single human predictors to the baseline model revealed that human variables vary in their effects on fire ignitions and that of the variables considered GDP is the most vital driver of fire ignitions. A combined model with all human predictors showed that the human variables improve the ignition predictions in most regions of the world, with some regions exhibiting worse predictions than the baseline model. We concluded that an ensemble of human predictors can add value to physical and empirical models. There are complex relationships between the variables, as evidenced by the improvement in bias in the combined model compared to the individual models. Furthermore, the variables tested have complex relationships that random forests may struggle to disentangle. Further work is required to detangle the complex regional relationships between these variables. These variables, e.g., population density, are well documented to have substantial effects on fire at local and regional scales; we determined that these variables may provide more insight at more continental scales.



Citation: Mukunga, T.; Forkel, M.; Forrest, M.; Zotta, R.-M.; Pande, N.; Schlaffer, S.; Dorigo, W. Effect of Socioeconomic Variables in Predicting Global Fire Ignition Occurrence. *Fire* **2023**, *6*, 197. <https://doi.org/10.3390/fire6050197>

Academic Editor: Grant Williamson

Received: 24 February 2023

Revised: 26 April 2023

Accepted: 27 April 2023

Published: 10 May 2023

Keywords: fire; machine learning; socioeconomic drivers



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fire has critical effects on the Earth's system, including land surface processes through vegetation dynamics, the carbon cycle, and local climate through bio-geophysical feedback [1]. Large fires in tropical and boreal forests have the ability to release terrestrial carbon stores, which can amplify climate change. Fire emissions also increase atmospheric aerosols, affecting the radiation budget and albedo through atmospheric scattering and increasing the cloud condensation nuclei, changing cloud cover and precipitation. The impact of fires on carbon cycling, atmospheric chemistry, and human respiratory health cannot be underestimated, regardless of their origin, including cropland burning or other

land management practices. These fires lead to the release of substantial amounts of carbon dioxide, which further exacerbates climate change. Moreover, releasing toxic gases and particulate matter from fires severely affects human health, particularly respiratory diseases. Additionally, these pollutants adversely affect atmospheric chemistry, forming ozone and other air pollutants that can affect human health and contribute to climate change. As such, it is crucial to understand the complex interactions between fires, carbon cycling, human health, and atmospheric chemistry to manage and mitigate these events' impacts effectively.

Naturally occurring fires are most frequently caused by lightning [2]. Depending on the circumstance, there are also volcanic, meteor, and coal seam fires. Lightning-caused fires are more likely to burn over larger areas and at hotter temperatures in hotter and drier conditions [3]. Forests degraded by disease and fragmented by deforestation are also more susceptible to fire [4].

While fire activity over millennial time scales is mainly driven by climate, evidence from archaeological, historical, and contemporary sources suggest that human societies have likely modified fire regimes since the Holocene [5]. Human behaviors that alter ignition frequency and seasonal timing or change landscape flammability can modulate the strength and amplitude of the relationship between a fire regime and weather conditions. Humans can alter various aspects of fire, including ignition occurrence (through starting fires), speed of propagation (through landscape management), size, spread, burned area, and duration (through controlling fuel loads, fragmentation, and fire-fighting). Since the late 18th century, anthropogenic influences on fire activity have increased, reflecting the impact of human industrialization and population growth on the environment [6]. Such anthropogenic impacts on fires are related to economic circumstances [7].

Knowing the source of ignition is crucial to understanding fire dynamics due to its initial position within the development of a fire. Understanding the physical and non-physical states that may lead to ignition events is critical in reducing fire occurrence and damage potential. Furthermore, refs. [8,9] it is fire ignition occurrence at different temporal and spatial scales, as these effects are not well understood, let alone quantified [8,9].

Fire modeling has been developed as part of the evolution of vegetation and Earth system modeling efforts [10,11], but these physics-based/empirical simulators heavily rely on complex parametrization [12]. They typically render low accuracies of fire variables, e.g., burned area, ignition density, rate of fire spread, and fire size [1,13–15], and have prediction bias when compared to observations.

Data mining and machine learning (ML) techniques have become essential aids in shedding light on the underlying relationships that influence fire within the Earth's system. Linear regression models were the first empirical fire ignition models [16] and were used to model the number of natural and human-caused fires together. In the 1980s, generalized logistic models were introduced to model the human-caused fires, binary and Poisson logistic regression models were used for predicting the number of human-caused fires [17]. These methods are easy to use and to interpret [18]. Unlike traditional physics-based and regression models, ML algorithms directly learn mappings between parametric rules from the data. They do not require physical parameterization [19], which is particularly beneficial when the number of considered variables is extensive. Random forest (RF) ML models are highly suitable for investigating emergent fire relationships, mainly because of their ability to evaluate non-linear relationships [7,20–23].

Human activities and their associated environmental influence on global fire ignitions can be represented through various indicators, e.g., infrastructure and agricultural land use. These variables may represent aspects of human behavior that can produce multiple, and even conflicting, effects on fire ressential to understand how the human dimension affects egimes, while different human indicators can have conflicting influences depending on the region. Selecting human indicators for model development is challenging due to correlations between available socioeconomic indicators. Since most fires are human-caused, it is imperative to understand which human dimensions have the most considerable effect on fire ignition [24]. Such information can help us to construct more accurate process

representations in physics-based models and, hence, to build better models to predict future climatic feedback mechanisms.

This study aimed to determine which indicators might help improve the modeling of human ignitions on a global scale. Ref. [21] introduced a novel machine learning approach to explain functional relationships between predictors and burned area. In this study, we adopted and applied this method to model the ignition occurrence of available global observational datasets, focusing on all fire starts, including agricultural burning. Based on this approach, we identified the most significant socioeconomic variables that explain human-fire ignition relations for the present day and how the effects of these variables vary regionally.

We first describe the observational datasets and the derived variables to develop RF models, and we describe the modeling approach (Section 2). In Section 3, we present our RF models' global performance (Section 3.1) and how the selected socioeconomic predictor variables contribute to model performance (Sections 3.2–3.8). Next, we discuss the interpretation of our results (Section 4) and the importance of certain predictor variables for global ignition occurrence modeling. In Section 5, we conclude the study and offer directions for further research.

2. Materials and Methods

We used global monthly ignition occurrences from the Global Fire Atlas (GFA) [25] as our response variable and several datasets on vegetation states, climate, land cover, soil moisture, and socioeconomic variables as predictor variables in the model development (Table 1).

Table 1. Full inventory of datasets used in the study.

Dataset	Derived Variables	Description	Native Spatial Resolution	Period	Temporal Resolution
Global Fire Atlas	[25]				
	Ignition occurrence	Fire ignition occurrences per pixel and month	500 m	January 2003–December 2016	monthly
Predictor variables					
Land cover	ESA land cover_cci version 2.0.7, http://maps.elie.ucl.ac.be/CCI/viewer/index.php (accessed on 5 August 2020)	Land cover classes were translated to fractional coverages of plant functional types (PFTs) in 0.25° grid cells [26]	300 m	January 1992–December 2015	annual
Climate and soil moisture					
CRU	CRU TS4.04 climate data [27]				
	Tmx (degrees Celsius)	Max temperature (degrees Celsius)	0.50°	January 1901–December 2019	monthly
	Dtr	Diurnal temperature range (degrees Celsius)	0.50°	January 1901–December 2019	monthly
	Wet (days)	Number of wet days	0.50°	January 1901–December 2019	monthly
	Pet (mm)	Potential evapotranspiration (millimeters)	0.50°	January 1901–December 2019	monthly
GPCC	Global Precipitation Climatology Centre (GPCC) [28]				
	Precip (mm/month)	Daily precipitation (mm/month)	0.25°	January 1891–December 2019	monthly
Soil Moisture	ESA soil moisture_cci version 6.1, http://cci.esa.int/data (accessed on 18 July 2022)				
	sm (m ³ m ⁻³)	Mean monthly soil moisture	0.25°	January 1978–December 2020	monthly

Table 1. Cont.

Dataset	Derived Variables	Description	Native Spatial Resolution	Period	Temporal Resolution
Vegetation state					
FAPAR/LAI	[29]				
	FPAR (unitless)	Fraction of absorbed Photosynthetic Active Radiation	500 m	January 2000–December 2019	8-day average
	LAI ($\text{m}^3 \text{m}^{-3}$)	Leaf Area Index	0.25°	January 2000–December 2019	8-day average
VOD					
	The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA) [30]				
	VOD_K lag (unitless)	Ku-band anomalies in Vegetation Optical Depth	0.25°	July 1987–July 2017	monthly
Biomass					
	ESA Biomass Climate Change Initiative: Global datasets of forest above-ground biomass for the year 2017, v1 [31]				
	agb (Mg ha^{-1})	Above Ground Biomass	0.25°	2017–2017	static
Socioeconomics					
Population density					
	Anthropogenic land-use estimates for the Holocene–HYDE 3.2 [32]				
	Popdens	Number of heads per square km	0.083333°	2000–2017	annual
Road density					
	Global Roads Inventory Project (GRIP) [33]				
	road_density (m/km^2)	Global patterns of current and future road infrastructure	n/a	static	static
GDP					
	[34] Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015				
	GDP_PPP (constant 2011 international US dollar)	Gross Domestic Product Purchasing Power Parity	Country level	1990–2015	annual
Livestock density					
	Global Livestock distribution	Livestock distribution per pixel; [35]	5 min	2010	static
Grazed land fraction					
	Pasture fraction	Area fraction of managed pasture; [36]	0.83333°	850–2015	annual

2.1. Ignition Density

Global ignition density data from the GFA [25] provide a satellite-derived estimate of monthly global ignition density for the period 2003–2016. GFA provides ignition density aggregated to a global grid with a cell size of 0.25° . The methodology and validation are presented in [25], while details on the underlying 500 m resolution daily burned area product (MCD64A1 collection 6) are described in [37].

2.2. Landcover

Land cover data were taken from the ESA CCI land cover CCI v 2.0.7 product, which provides annual global land cover maps at 300 m spatial resolution, covering the epoch 1992–2015. We translated the land cover classes into plant functional types (PFTs) to be comparable with the classification used in global vegetation models [38]. The following nine PFTs were derived: broadleaved evergreen tree and shrub (TreeBE, ShrubBE), broadleaved deciduous tree and shrub (TreeBD, ShrubBD), needle-leaved evergreen tree and shrub (TreeNE, ShrubNE), needle-leaved deciduous tree (TreeND), natural grass or herbaceous vegetation (Herb), and managed grasslands or crops (Crop). The land cover maps provide vegetation types and were spatially aggregated and expressed as the fractional coverage of PFTs within a 0.25° grid cell. The fractional coverage was further used to determine the 30% threshold of all PFTs combined, below which pixels were excluded from model analyses; this was to define a proxy by which we could determine if a pixel would have sufficient fuel to sustain a large fire enough to be detected in the Global Fire Atlas algorithm.

2.3. Climate

We used monthly data of maximum air temperature, diurnal temperature range (DTR), and the monthly number of wet days from the Climate Research Unit (CRU) TS4.04 dataset [27]. DTR has long been used to predict fire weather conditions because it is sensitive to stable weather conditions usually associated with low humidity and supportive of fire activity [39,40]. These datasets provide monthly climate time series at a resolution of 0.5° based on spatially interpolated weather station observations. Precipitation data were acquired from the Global Precipitation Climatology Centre (GPCC) version 7 dataset [28] at the native spatial and monthly temporal resolution of 0.5°. These variables were used to represent the various climate states in the models.

2.4. Soil Moisture

Surface soil moisture was taken from the ESA CCI soil moisture dataset (version 6.1 COMBINED), which is based on merging soil moisture products from various active and passive satellite sensors [41,42]. The dataset represents moisture conditions in the upper soil layer (~2 cm) and is available at a spatial resolution of 0.25°, and it shows the daily time step for 1979–2020. Winter gaps in the dataset were forward and backfilled using linear interpolation. These filling approaches assume that once the soil is frozen, water content within the soil does not change until the next measurement; therefore, no combustion occurs until warmer temperatures remove the water content in the soil.

2.5. Vegetation State

To account for the influence of fuel availability on fire ignitions, we used the MOD15A2H version 6 Moderate Resolution Imagine Spectroradiometer (MODIS) combined Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), an 8-day composite dataset with a 500 m resolution. LAI is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and one-half the total needle surface area per unit ground area in coniferous canopies [43] and adds foliage information to the RF models. FAPAR is the fraction of incident photosynthetically active radiation between 400 and 700 nm, absorbed by the green elements of a vegetation canopy.

Vegetation optical depth (VOD) accounts for the attenuation of microwaves through vegetation as a function of water content and vegetation structure, and thus, it can be used as an indicator of fuel moisture condition. We included the Ku-band of VOD from the Vegetation Optical Depth Climate Archive (VODCA) dataset [30], which combines multiple VOD datasets derived from various sensors (SSM/I, TMI, AMSR-E, Windsat, and AMSR-2), retrieved using the Land Parameter Retrieval Model (LPRM) [44]. We computed the anomaly by deducting the long-term climatology from each instance to describe climate variability accurately.

Above-ground biomass (AGB; Mg ha⁻¹) was acquired from the ESA CCI BIOMASS v1 dataset [45]. Biomass is the amount of living biomass organic matter stored in vegetation above the soil and was used here as another indicator of fuel availability. It is available as a static dataset for the year 2017.

2.6. Socio-Economic Variables

2.6.1. Population Density

Population density was taken from the Historical Database of the Global Environment (HYDE version 3.2) [32], a combination of gridded historical population and land-use estimates. Historical records were used to model population density at the provincial and national levels. Algorithms were then used to spatially distribute the total population and land-use areas to a spatial resolution of 0.083333°.

2.6.2. Gross Domestic Product

The GDP (per capita) used in this study is described by [34]. The data represent the average GDP of each grid cell and are given in 2011 international US dollars. The

information is derived from GDP per capita multiplied by gridded population data from Global Human Settlement (GHS) [46]. The GHS processing framework uses assorted data, including global archives of fine-scale satellite imagery, census data, and volunteered geographic information. The GDP product has a global extent at a resolution of 0.083333° for annual time steps from 1990 to 2015.

2.6.3. Road Density

We used the Global Roads Intercomparison Project (GRIP) [33] dataset to represent road access as a potential cause of ignitions. The GRIP dataset consists of global and regional vector datasets in ESRI file geodatabase and shapefile format and global raster datasets of road density at a five arcminutes resolution ($\sim 8 \times 8 \text{ km}^2$). The effect of road density was tested to assess the influence of landscape fragmentation on ignition occurrence predictions.

2.6.4. Gridded Livestock

The gridded livestock of the world (GLW v3) is a peer-reviewed spatial dataset on livestock distribution for the reference year. GLW version 3 was compiled from the datasets presented in Table 1 representing the following species: cattle, sheep, goats, buffaloes, and horses. These data were normalized to livestock units to represent the gridded livestock density. The individual species datasets are available at the global extent and 5 min of arc resolution. Livestock was included to test the influence of land-use management across different regions on fire occurrence.

2.6.5. Grazed Lands

Land Use Harmonization (LUH2) [36] contains historical reconstructions of land-use and land-use transitions, including managed pastures and rangelands; we combined these two to create a “grazed lands” variable, which more accurately describes the land that is grazed by domestic and wild herbivores. Similar to the baseline historical scenario, datasets include annual gridded fractions of land-use states, all transitions between those states, and associated management layers for the period 850–2015. The “high” and “low” scenarios are based on tentative high and low data-driven land-use reconstructions from HYDE and accompanying wood harvest and were aggregated to produce an average state of the grazed lands variable.

2.7. Data Preparation

A significant limitation of ignition occurrence data from the GFA is that the algorithm can only detect fires over 21 ha [25]. All other datasets used in this study are aggregated to monthly averages to predict these ignition occurrence data. Next, all datasets were rasterized (where necessary) and resampled to a 0.25° grid using conservative remapping, described in [47], to avoid smoothing anomalies through alternative approaches, such as bilinear interpolation [21]. The values of annual datasets were repeated 12 times each year, and static datasets were replicated to match the total number of months corresponding to the ignitions target. Socioeconomic variables tend to be annual or static because economic development does not change quickly, but the vegetation and climate variables need to be monthly to capture seasonal changes.

2.8. Random Forest Model Setup

We use RF regression to model the relationships between ignition occurrence and the potential drivers (predictors). RF is an ensemble learning approach in which multiple decision trees are built using a randomly sampled subset of the training observations [48]. The final model constitutes the average result of the decision trees. In averaging, RF also mitigates (though it does not entirely resolve) the overfitting problem inherent in decision tree modeling [49]. We used scikit-learn version 24.1 [50] to determine hyperparameter settings of the baseline model using Grid Search to create a parameter grid based on the results of a random search. Grid search is a model hyperparameter optimization technique,

and in Python, it is provided in the GridSearchCV class. GridSearchCV was combined with k-fold cross-validation ($k = 5$) to construct and evaluate one model for each combination of parameters. From the Grid Search, we determined the following parameters: the number of estimators ($n_estimators = 500$) determines the number of trees whose predictions are averaged; the maximum depth ($max_depth = 7$) limits the number of split levels; setting $min_samples_leaf = 3$ requires that all final splits contain at least three samples; $Max_features = 8$ refers to the number of features to be considered when looking for the best split; and the minimum number of samples to perform a split ($min_samples_split = 2$).

2.9. Variable Selection

To detect multicollinearity in the climate and vegetation variables, we used the Variance Inflation Factor (VIF) [51]. VIF is calculated by regressing a predictor against a model with all other predictors. This results in R^2 values that can be used to compute:

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

where R_i^2 is the unadjusted coefficient of determination of regressing each variable on the remaining variables, and these remaining variables may contribute little or no additional information to the model [52]. VIF ranges from 1 upwards; the VIF decimal form tells what percentage variance is inflated for each coefficient, e.g., a VIF of 1.5 means the variance of a particular coefficient is 50% larger than expected if there were no correlation with other predictors. We employed VIF to iteratively filter heavily correlated variables to identify and eliminate them from model evaluation. We used a threshold of <10 for the VIF to consider variables to use for our experiments. The following variables were found to be heavily correlated with other features and were thus removed due to the VIF process: LAI, dtr, and tmx.

Table 2 shows the variables we used in our baseline model and the VIF, each before and after iteratively removing heavily correlated features.

Table 2. Baseline model features after filtering and VIF pre- and post-filtering.

Variable	Pre-VIF	Post-VIF
agb	1.38	1.31
sm	5.01	4.85
pftTreeBE	81.69	2.44
pftHerb	102.36	4.05
pftShrubBD	48.86	3.43
pftShrubNE	15.59	3.25
pftTreeBD	57.88	2.12
pftTreeNE	15.17	1.46
fAPAR	212.55	9.03
vod_K_anomalies	8.92	1.03
pet	22.66	5.45
wet	4.10	3.90
precip	2.68	2.54
pftTreeND	6.47	1.13
tmx	27.90	-
dtr	16.14	-
LAI	4052.46	-

2.10. Model Training Iterations

We trained several RF models to test hypotheses about the influence of human-related drivers on global fire ignitions. The baseline model without human predictors was used to establish a reference for improving the model, which only includes the climate, vegetation, and land cover variables shown in Table 2. We selected this approach because it enables us to assess the model's ability to predict spatial patterns over individual months. We split the combined dataset with all the climate, vegetation, land cover, and human variables into training (2003–2011) and testing (2012–2016), where each of the subsets of the original data contains global monthly values, thereby ensuring that no information from the test data subset is included in the model training.

Next, we added a single socioeconomic variable to the baseline model and trained the model to predict ignitions over the testing period. We replaced the socioeconomic variable with another for the third model before training the RF model again. We repeated this process for all subsequent variables, as shown in Table 3. Finally, we trained a model with all human variables added to the baseline model.

Table 3. Performance of models against the Global Fire Atlas testing subset. MAE and NMSE are provided monthly.

Predictor Variables	R ²	MAE	NMSE
Baseline model (BL): WET + Precip + PET + SM + fPAR + VOD_K' + Herb + TreeBD + ShrubBD + TreeBE + ShrubBE + TreeNE + ShrubNE + TreeND	0.53	0.33	0.04
BL + pftCrop (CF)	0.54	0.28	0.04
BL + Livestock density (LD)	0.56	0.32	0.03
BL + Grazed lands fraction (GLF)	0.55	0.30	0.04
BL +GDP	0.52	0.32	0.05
BL + Road density (RD)	0.53	0.33	0.04
BL + Population density (PD)	0.57	0.31	0.03
Full model: (FM): BL + pftCrop + GDP + grazed lands fraction + livestock density + road density	0.63	0.81	0.03

With monthly predictions and observations of ignition occurrence per pixel, we assessed the change for each model pixel induced by adding each human predictor to the baseline model using:

$$\Delta = V - B \quad (2)$$

V is the temporal mean of the predicted ignitions from the model with a single added human variable, and B is the temporal mean of predicted ignitions from the baseline model. We also computed model improvement for the full model, defined as:

$$\text{Improvement} = |F - O| - |B - O| \quad (3)$$

where F is the mean ignitions from the full model, B is the mean baseline predictions, and O is the temporal mean of the observed dataset (GFA).

Next, we computed, for each iteration, the mean absolute error (MAE) and normalized mean square error (NMSE), two commonly used evaluation metrics for a random forest model. MAE measures the average magnitude of the errors in a set of predictions without considering their direction. It is calculated as the sum of the absolute differences between the actual and predicted values, divided by the number of predictions. On the other hand, NMSE is a variant of mean square error (MSE) normalized by the variance of the actual values. It is calculated as the MSE divided by the variance of the actual values. Both MAE

and NMSE can be useful for evaluating the results of a random forest model, but MAE is generally preferred when there are outliers in the data since it is less sensitive to them.

Lastly, we used accumulated local effects (ALEs) to examine the coupled relationships fitted by the RF models. ALE is a robust alternative to partial dependence plots (PDPs) or individual conditional expectation (ICE) [53]. To compute the effect of a feature on prediction, the ALE technique uses all instances of the variable with about the same value as the first instance and obtains the model predictions. This gives the pure feature effect of the variable without considering the effect of the correlated features. The instances of each feature are first divided into intervals. For the data points in each interval, the difference between the predictions for each instance within the interval is replaced with the interval limits. These differences are later accumulated and centered, resulting in the ALE curve. We assessed the impact of each predictor variable on ignition occurrence in isolation using 1D ALEs, taking into account the effect of all other predictor variables. ALEs were computed both for the individual human predictor model runs and for the full model including all human predictors.

2.11. Data Sampling

Since fire ignitions do not occur in large areas of space and time, this presents the need to balance the ignition occurrence data better; a lower ratio of fires to non-fire conditions results in lower predictive capability. In addition, using all available data points is not practical, as many of them will never meet the threshold necessary for combustion. Therefore, we oversampled our data to boost ignition representation. We filtered all data into two subsets: one where ignitions are recorded and another where no ignitions are recorded. We oversampled ignitions by randomly selecting the same number of data points from the non-ignition data subset and by combining it with the ignition-only data, thus raising ignition representation from approximately 5% to 50%. Finally, we intermixed the oversampled dataset. This was performed for the time series of each pixel. These intermixed data were split into training and testing sets.

3. Results

3.1. Performance of the Climate/Vegetation Baseline Model

The baseline model (BL) captured the general ignition patterns of GFA, particularly in the high ignition regions (Figure 1a,b). However, it consistently overpredicted the average number of monthly ignition occurrences and overpredicted fire ignitions across most regions (Figure 1c), a limitation previously identified in regression models [54,55]. The areas with largest overpredictions were Central America, Northeastern Brazil, sub-Saharan Africa, Eastern Europe, India, East and Southeast Asia, and the coastal regions of Australia. Underpredictions notably occurred in Angola, Sierra Leone, and Cambodia.

Table 3 shows the predictive skill for all configurations tested. The baseline model achieved an out-of-bag R^2 of 0.95 for the training dataset and 0.53 for the test dataset. Differences in MAE between the individual models were minimal but are likely significant due to the large sample sizes used. Only the full model (FM) showed substantially deviating skill from the baseline model and from the baseline model in combination with any single socioeconomic predictor. The results of the different models are discussed in the following.

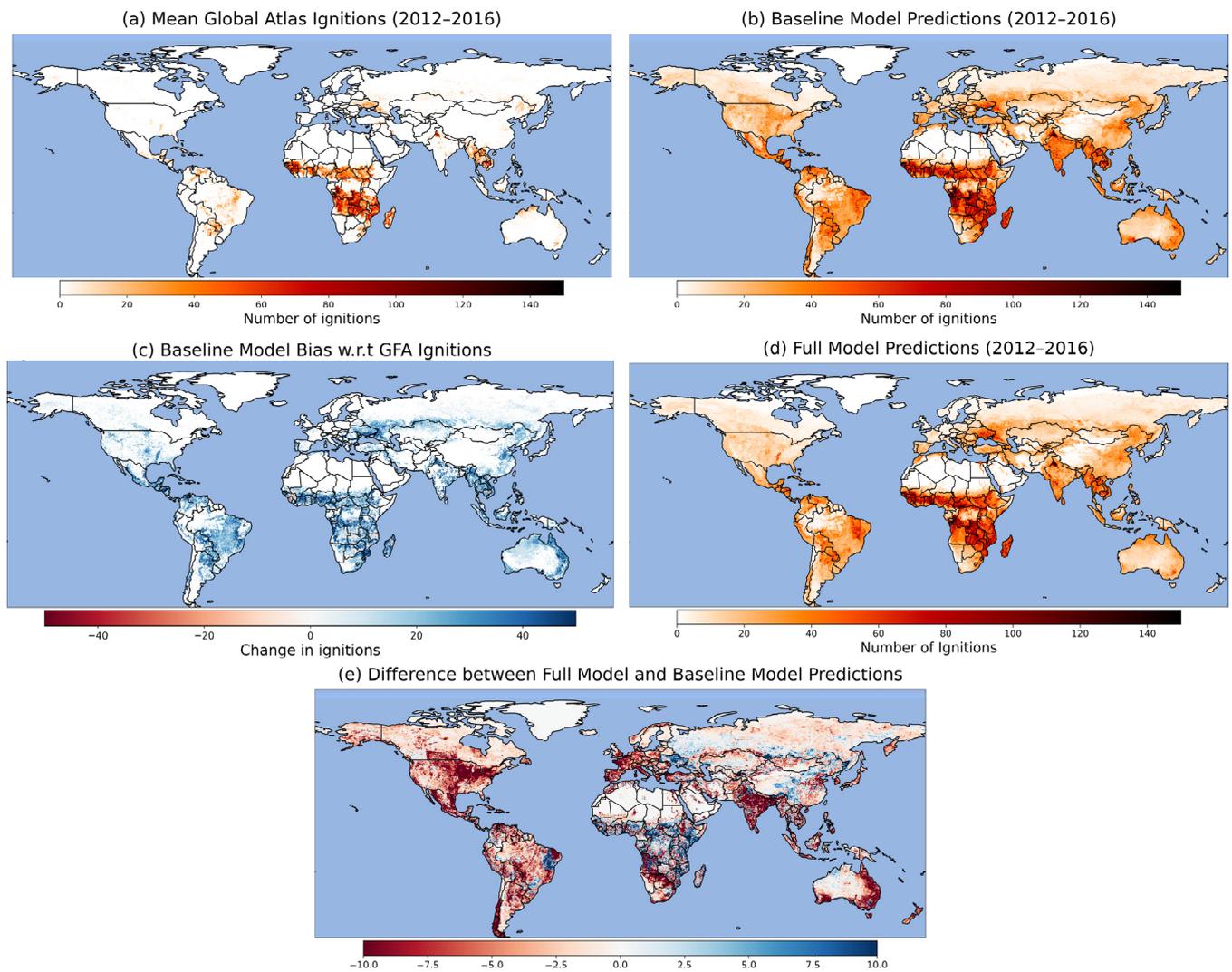


Figure 1. (a) The average number of observed ignitions per year from the Global Fire Atlas dataset, (b) random forest predictions of global fire ignitions from the baseline model (BL), (c) bias between BL and GFA test data, (d) full model (FM) predictions with all baseline and human variables, and (e) skill difference between FM and BL computed from Equation (3).

3.2. Effect of GDP

The impact of adding GDP to the baseline model varied across regions (Figure 2a,b); regions where adding GDP increased the number of predicted ignitions mostly had lower GDP, whereas regions where it reduced the number of ignitions were predominantly represented by higher-income countries. Across North, Central, and much of South America, Europe, and Australia, GDP influenced the model to suppress fires. In Africa, the model showed two patterns in ignition predictions, decreases in west and southern Africa (e.g., in Angola, Botswana, South Africa) and an increase in fire ignitions in East Africa (e.g., in Kenya, Tanzania, Somalia). The other regions where GDP acted to increase fires are northeastern Brazil, Ukraine, Southern and Central India, Central Asia, and most of China.

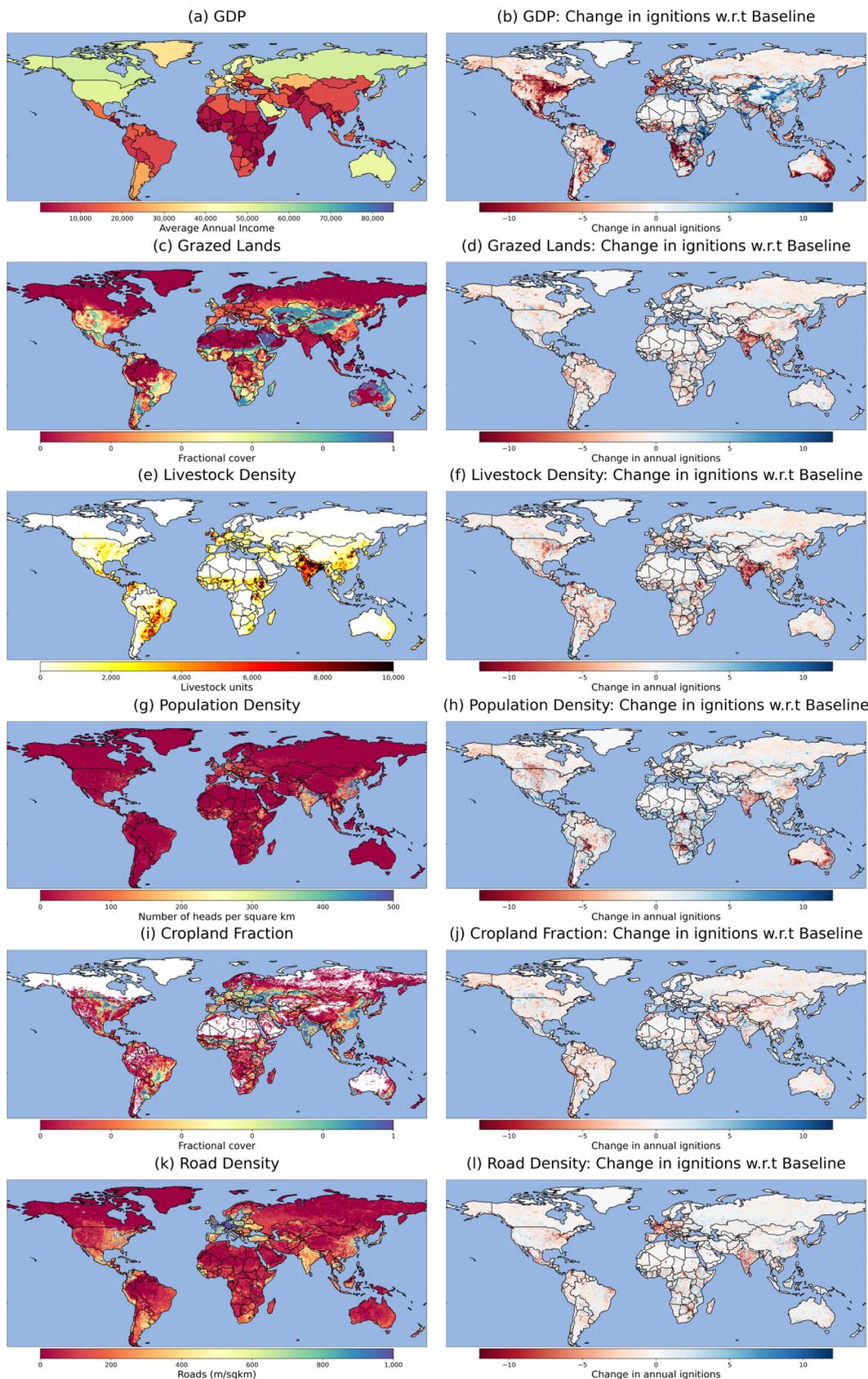


Figure 2. Left column: pixel-wise means for each human variable. White colors indicate zero. **Right** column: difference between each model run with human variables and the baseline model. Negative values imply reduced ignitions w.r.t BL.

3.3. Effect of Grazed Land Fraction (GLF) and Livestock Density (LD)

While spatial distributions of GLF and LD diverged (Figure 2c,e), they had very similar impact patterns on fire ignition (Figure 2d,f). Adding them to the model globally reduced ignition occurrence; although, changes were relatively small. Only in South Asia did it substantially reduce the number of ignitions. This region has a very high livestock density but a low fraction of grazed lands, which suggests that livestock is held on farms and not on pastures. Most regions with a high grazed-land fraction are extensively managed barren lands with little livestock and hence have little effect on ignition occurrence (e.g., central Asia and the Arabian peninsula).

3.4. Effect of Population Density (PD)

The effect of population depends upon the region and socioeconomic development (Figure 2g,h). In developing and newly industrialized countries (e.g., India and East China), we observed a decrease in ignitions across regions with high PD, while less-populated areas in these regions showed increased ignitions, e.g., much of Sub-Saharan Africa (with the exception of Botswana), Mexico, and Peru. On the contrary, PD model predictions in regions with high economic development and low PD were generally lower than the baseline, e.g., in the interior of Australia outside of the large metropolitan areas, southern Chile, most of Canada, and Spain.

3.5. Effect of Cropland Fraction (CF)

Generally, an increase in ignitions was observed in regions with a high cropland fraction, most evidently in Argentina, Northern India, and Eastern China (Figure 2i,j). This is in line with the expectation that CF tends to increase ignitions due to the burning of cropland residue following the growing season. However, the observed impacts are relatively weak.

3.6. Effect of Road Density (RD)

There is a clear relationship between road density and fire ignition predictions. In regions with higher RD, e.g., western Europe, India, the Eastern United States, and Mexico, predictions were generally lower than the baseline, while regions with low RD had the opposite effect (Figure 2k,l).

3.7. Combined Effect of All Human Variables

The full model (FM; Figure 1e) generally showed prediction improvements, i.e., a decrease of the positive bias of the baseline model (Figure 1b) with respect to the GFA reference (Figure 1a), even in regions where some of the individual human variable models showed a poorer performance than the baseline model. Areas with the most notable improvement when adding the ensemble of human predictors were India, southern Africa, Western Europe, the Australian east and southwest coasts, and across the Americas. The most notable regions where the FM model performed worse than the baseline, i.e., an increase of predictions with respect to the baseline model, were Eastern Brazil, Angola, and Ukraine.

3.8. Accumulated Local Effect Analysis

The ALE curves of the individual human predictor models and of the FM with all human predictors gave very similar results, which is a sign of the robustness of the ALE method. For this reason, in the following, we only show and discuss the ALE diagrams based on the models with single human predictors.

We found GDP and cropland models to have the most substantial effect on the ability of our models to predict ignition occurrence (Figure 3). While there are artifacts in the ALE due to necessarily using country-level data in a gridded analysis, we saw a strong tendency for ignitions to decrease with increasing GDP. The relationship between GDP and changes in prediction remained constant beyond approximately USD 7000 per year.

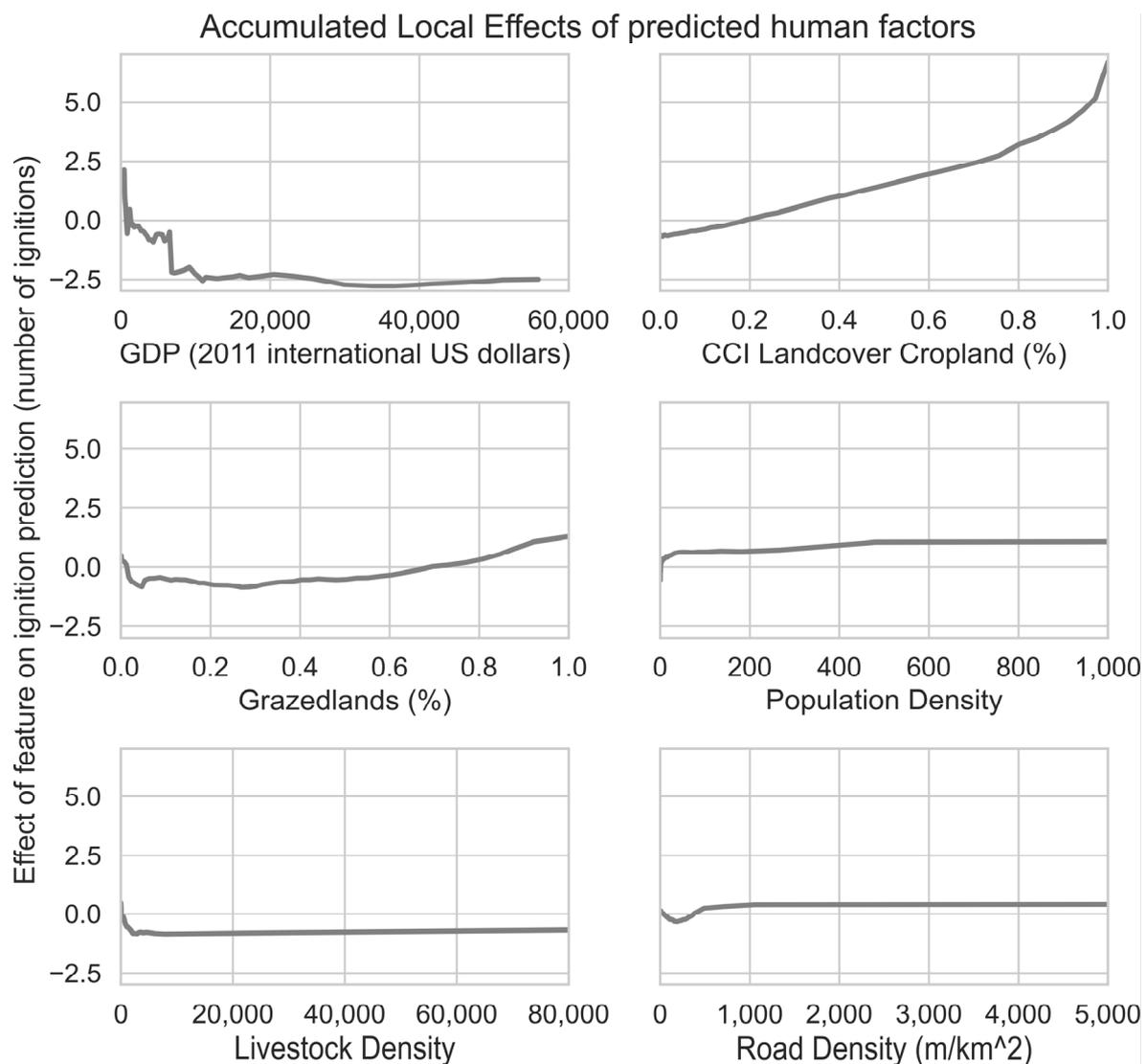


Figure 3. ALE analysis based on separate models for each of the individual socioeconomic variables.

For cropland fraction, the spatial patterns of the impacts were not as strong as for GDP. Its ALE plot (Figure 3b) showed that it even more strongly affected the average prediction, increasing up to six ignitions per year. This can in part be attributed to substantial effects in regions with high cropland fractions, e.g., Northern India. A similar pattern, but with a lower magnitude, was observed for grazed lands.

For RD, there was a reduction of fire ignitions in regions where the road density was less than 200 m/km², followed by an increase in ignition predictions up till ~500 m/km², beyond which there was no impact (Figure 3f). For PD and LD, there was only some response for very low values.

4. Discussion

4.1. Overall Performance

The difference in performance of the baseline model between the training and testing data (R^2 of 0.95 vs. 0.53, respectively) suggests that the model struggles to generalize the results due to overfitting. Hence, some caution is needed when interpreting the addition of socioeconomic predictors to the baseline model, which was shown to have limited global impact for most variables (Table 3, Figure 3). This could be due to the seasonality in the multiple correlated features from the baseline model, which can result in the models weighing more on the correlated than the static or annual socioeconomic features. Hence,

the effects of the different variables tested in this study should be interpreted as a general indication of the spatial variation in predictor effects on ignition occurrence rather than in absolute terms.

We noted that the substantial increase in R^2 (from 0.53 to 0.63) from the baseline model to the full model concurred with an increase in MAE (Table 3). This, at first sight contradictory, result can occur when data have a lot of outliers or extreme values, to which the MAE is more sensitive than the R^2 value. In general, R^2 indicates the proportion of the variance in the dependent variable explained by the model, while the MAE measures the average magnitude of the error in the predictions made by the model. Thus, a model with a high R^2 (good average skill) may still have a high MAE if it is not making accurate predictions for the outliers.

4.2. GDP

GDP was shown to influence the most robust patterns in ignition changes with respect to the baseline model; however, interpreting the effect of GDP was not straightforward because GDP is a proxy for economic development, which entails a broad spectrum of activities. Ref. [56] expressed concern at the assumption that GDP is a measure of progress and that increasing GDP is positively correlated with increasing quality of life. They argued that increasing GDP may indeed be indicative of increasing societal inequality. In this context, a GDP increase may indicate societal imbalance, making the overall argument linking higher GDP in more developed countries difficult to explain, as GDP may be high in regions where the majority of the population is poor. However, GDP has been used as an index of the cultural influences on the use of fire [40] and remains a pivotal index for evaluating the influence of human activity on fire.

Across the densely populated regions of Australia, western Europe, and the northeastern United States, GDP was shown to decrease modeled ignition occurrences. GDP in these regions was approximately between USD 45,000–USD 60,000. GDP in such high-income nations is usually not directly tied to environmental degradation, as the main economic sectors (e.g., the service industry) do not rely on the environmental extraction of resources. The model with GDP also captured the unique effects of Ethiopia's topography, as it reflected land-use practices. The topography of arid flatlands to the east and mountainous, precipitation-rich regions to the west resulted in significant differences in land-use practices. Ref. [57] described Ethiopia's complex ecosystems' micro and macro-ecological states; the country heavily relies on strict land-use practices to maximize GDP output. Ref. [58] found an increased frequency of fires in recent decades attributed to expanding human populations in the southwestern regions and national parks; we showed similar trends in GDP across the country, not just in the regions they specified. Across China, where GDP was relatively low, GDP influence on fire ignitions was positive. This is consistent with [59], who identified a positive relationship between GDP and fire occurrence in Southeast China.

4.3. Grazed Lands and Livestock

Figure 3c shows that adding GLF to the baseline model caused the most change in fire ignition predictions at very low and high grazed land fractions but little change elsewhere. Ref. [60] found that coupling fire and grazing reduced fuel accumulation and, hence, ignition potential on rangelands across all weather patterns. This is in line with the sharp decrease in ignition occurrence at low grazed lands fractional cover (Figure 3c), owing to low fuel availability, which was followed by a gradual increase in fire occurrence when GLF, and hence fuel availability, increased. Livestock density (Figure 3e) showed a sharp decrease in ignitions caused by shallow livestock density ranges (up to 1000 heads); this suppression stabilized at this value and did not change as livestock density increased. This is because livestock continually removes fuel from the landscape, which reduces the chances of ignitions and increases landscape fragmentation, leading to smaller patches of fuel, which may not easily ignite, or if ignition occurs, the fires remain too small to be detected by the MODIS.

4.4. Cropland

Figure 3b shows that the impact of cropland fraction on ignitions is strongly positive. The model likely learned this behavior from the regions in Africa, India, China, and Argentina, where the annual crop residues are burnt after harvesting. Maize, rice, and cotton are major crops in India [61]. Maize, rice, and wheat account for 90% of China's total food production [62], maize is a major crop in Sub-Saharan countries, and soybean is a major crop in Argentina [63]. Despite the absence of many large-scale farmlands, cropland as a variable primarily acted to increase ignitions in Africa. Botswana, Namibia, and the eastern half of South Africa showed a sizeable increase in ignition occurrence despite low fractional cropland cover. Furthermore, these patterns are visible in the Delta map of the FM (Figure 1e). The ALE plot for cropland (Figure 3b) showed a robust positive relationship between cropland cover and model predictive performance over the entire cropland fraction domain, including the lower ranges. Thus, the spatial prediction patterns met our expectations over this region.

4.5. Road and Population Density

Road density in the United States (Figure 2i) is much higher across the eastern than in the western half. Correspondingly, we found that high road density leads to decreased ignitions over the eastern regions of the United States, while leading to slightly increased fire occurrences over the western regions. More roads should mean more people access the landscape and start fires; however, regions with high road density also tend to be more urbanized (cf. Figure 2k); therefore, there are mechanisms in place preventing fires in order to protect infrastructure. Thus, in the event of a fire, swift action is often taken to suppress the fire. Road density suppresses fires mainly in economically active regions, e.g., western Europe, eastern China, and much of India. The relationship between RD and the other BL variables is complex. However, low RD (under 200 m/km²) leads to a slight decrease in fires (Figure 3f); this may be a consequence of sparsely populated and vegetated areas, which could indicate low fuel availability because they are located in desert regions. As mentioned earlier, we also noted that the effects of road density and livestock density on ignitions are the lowest according to the ALE plots because they are static datasets. A key consideration with these two datasets is that the road density will remain mostly constant, but livestock density will change over periods.

4.6. Full Model

In general, the differences between the FM and BL models (Figure 1e) are broadly dominated by GDP (Figure 2b), as illustrated by the similarities between these maps. Across much of North, Central, and South America, ignition patterns in the FM were very similar to those in the GDP model; although, in most of South America, the FM predicted slightly fewer ignitions than GDP. This is likely a translated effect from grazed lands, population density, cropland fraction, and road density, which were shown earlier to reduce ignitions with respect to the BL model. This is in line with [64], who found reduced fire occurrences in Brazilian-managed pasture lands compared to grasslands when studying fire regimes.

Across East to West Sub-Saharan Africa, FM predicted higher ignitions. This mainly followed the spatial patterns of GDP, but FM predicted more fires over West Africa's shrubland, forest, and herbaceous vegetation. FM predictions over South and East Asia were primarily driven by grazed lands, livestock, population density, and road density. These variables showed FM predicting lower ignitions than BL, overpowering the effect of cropland and GDP in the region, which as individual predictors increase the number of ignitions. In the Middle East, combining all human variables in FM reduced ignitions with respect to BL, driven mainly by cropland fraction. Over Western Europe, both road density and GDP drove FM model predictions, overriding the effects of population density, grazed lands, and cropland fraction.

Regarding the ability of FM to reduce the model bias compared to the GFA reference dataset, the largest bias reduction was over the northeastern United States, Northeastern

Brazil, Sub-Saharan Africa, South Asia, the Australian coasts, and western Europe. For the latter, there was a clear contrast with the former Soviet Union countries, where adding socio-economic predictors increased the bias with respect to the observed ignitions.

4.7. Data and Model Limitations

All baseline model variables were based on climate or climate-driven vegetation states; thus, they were heavily correlated. While highly correlated values do not cause multicollinearity issues in RF models per se, most of these variables have monthly temporal resolution, while the land cover and human variables are available only on an annual basis or even static. This makes the RF models biased towards the climate-driven variables, especially because ignition occurrences also have seasonal patterns. Another limitation is that country-aggregated GDP, while it is the strongest indicator of human-induced fire occurrence, is challenging to understand fully. There can be significant variation in GDP within a country's borders, and with countries such as the United States, Brazil, and Australia, the changes in economic output can be tied to variations in climate, vegetation types, local regulations, and land cover regimes. A single value for a country's GDP dilutes the subtle nuances in the data across various regions within a country's borders. Furthermore, oil-rich countries skew the country-level GDP data (e.g., Russia), leading to high GDP, which is not representative of the prosperity of the average population. We also noted that the random forests used in this study could be better optimized to reduce overfitting and improve predictions. More sophisticated (deep learning) methods could also be employed to gain more insight into the relationships between fire occurrence and human activity.

The ESA-CCI Landcover classes we used induced discrepancies in fire prediction over Central and Eastern Brazil. The classification scheme of the ESA CCI Landcover classes is a potential limitation for predicting fires at regional scales. The ESA CCI Landcover classes use a general classification scheme that may not accurately represent the unique landcover types and management practices in specific regions.

For example, in Brazil, the classification scheme may not adequately capture the distinction between different types of savanna vegetation, such as cerrado and campo limpo, which have different fire regimes and responses to fire. Similarly, the classification scheme may not capture the distinction between different types of forest, such as rainforest and seasonal forest, which also have different fire regimes and responses to fire. Furthermore, the classification scheme may not capture the influence of human activities on landcover and fire regimes. For instance, the classification scheme may not differentiate between different types of agricultural land use, such as soybean and sugarcane plantations, which have different fire management practices and fire risks.

These limitations can result in inaccurate predictions of fire and can lead to misunderstandings in regional interpretations of the results. To overcome this limitation, alternative landcover datasets that incorporate more-detailed and region-specific classification schemes could be used. These datasets could be derived from ground-based surveys or higher resolution satellite imagery that can capture more nuanced landcover features and management practices.

5. Conclusions

We identified GDP as the most significant socioeconomic parameter that explains human-fire ignition relations for the present day. The effects of population density, grazing land fraction, and livestock density improved the model performance in regions such as India. This indicates that the presence of people and animals affects the likelihood of ignitions occurring, but the simulated relationships are not monotonic, implying that further research is needed to understand the impact of these variables in conjunction with the variability of other human, climate, and land cover variables. For example, increasing the cropland fraction can increase detected fire ignitions (due to people entering the landscape and setting fires) or decrease them (due to fragmentation and suppression measures),

but our models cannot show these relationships. We also noted that some fires start in agricultural land, but then can (both intentionally and accidentally) burn into natural vegetation. Higher road density can increase access to the landscape, allowing for more fires to start, but it can also introduce fragmentation, resulting in smaller fires that remain undetected by the satellite; we are unable to show this from our model results, suggesting that the models have some room for improvement in showing these nuances clearly. In addition, higher road densities not only amplify fragmentation, reducing the likelihood of detecting small fires, but they also increase the potential for the improved reporting and control of fires before they increase in size. However, our results highlight that a unique global response of fire ignitions to socioeconomic variables does not exist and that human effects on fire ignitions vary regionally. This implies the need to explore regional drivers of fire ignitions and to build hybrid models by representing the complex effects of humans on fire ignitions with, e.g., machine learning models in global process-oriented fire and Earth system models.

Author Contributions: Conceptualization, M.F. (Matthias Forkel) and W.D.; methodology, M.F. (Matthias Forkel), M.F. (Matthew Forrest), W.D. and T.M.; software, T.M. and N.P.; validation, T.M.; formal analysis, T.M.; investigation, T.M.; resources, T.M. and W.D.; data curation, M.F. (Matthias Forkel), M.F. (Matthew Forrest) and T.M.; writing—original draft preparation, T.M.; writing—review and editing, T.M., W.D., S.S., R.-M.Z. and M.F. (Matthew Forrest); visualization, T.M.; supervision, M.F. (Matthias Forkel) and W.D.; project administration, W.D.; funding acquisition, W.D. and M.F. (Matthias Forkel). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Austrian Science Fund (FWF), grant number I 4271-N29.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Data are publicly available and referenced in the article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Lasslop, G.; Coppola, A.I.; Voulgarakis, A.; Yue, C.; Veraverbeke, S. Influence of Fire on the Carbon Cycle and Climate. *Curr. Clim. Chang. Rep.* **2019**, *5*, 112–123. [[CrossRef](#)]
2. Li, Y.; Mickley, L.J.; Liu, P.; Kaplan, J.O. Trends and spatial shifts in lightning fires and smoke concentrations in response to 21st century climate over the national forests and parks of the western United States. *Atmos. Chem. Phys.* **2020**, *20*, 8827–8838. [[CrossRef](#)]
3. Burrows, N.D.; Burbidge, A.A.; Fuller, P.J.; Behn, G. Evidence of altered fire regimes in the Western Desert region of Australia. *Conserv. Sci. West. Aust.* **2006**, *5*, 14–26.
4. Cawley, K.M.; Hohner, A.K.; McKee, G.A.; Borch, T.; Omur-Ozbek, P.; Oropeza, J.; Rosario-Ortiz, F.L. Characterization and spatial distribution of particulate and soluble carbon and nitrogen from wildfire-impacted sediments. *J. Soils Sediments* **2018**, *18*, 1314–1326. [[CrossRef](#)]
5. Dietze, E.; Theuerkauf, M.; Bloom, K.; Brauer, A.; Dörfler, W.; Feeser, I.; Feurdean, A.; Gedminienė, L.; Giesecke, T.; Jahns, S.; et al. Holocene fire activity during low-natural flammability periods reveals scale-dependent cultural human-fire relationships in Europe. *Quat. Sci. Rev.* **2018**, *201*, 44–56. [[CrossRef](#)]
6. Bowman, D.M.J.S.; Kolden, C.A.; Abatzoglou, J.T.; Johnston, F.H.; van der Werf, G.R.; Flannigan, M. Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* **2020**, *1*, 500–515. [[CrossRef](#)]
7. Chuvieco, E.; Giglio, L.; Justice, C. Global characterization of fire activity: Toward defining fire regimes from Earth observation data. *Glob. Chang. Biol.* **2008**, *14*, 1488–1502. [[CrossRef](#)]
8. Balch, J.K.; Bradley, B.A.; Abatzoglou, J.T.; Chelsea Nagy, R.; Fusco, E.J.; Mahood, A.L. Human-started wildfires expand the fire niche across the United States. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2946–2951. [[CrossRef](#)]
9. Cattau, M.E.; Wessman, C.; Mahood, A.; Balch, J.K. Anthropogenic and lightning-started fires are becoming larger and more frequent over a longer season length in the U.S.A. *Glob. Ecol. Biogeogr.* **2020**, *29*, 668–681. [[CrossRef](#)]
10. Hantson, S.; Arnet, A.; Harrison, S.P.; Kelley, D.I.; Colin Prentice, I.; Rabin, S.S.; Archibald, S.; Mouillot, F.; Arnold, S.R.; Artaxo, P.; et al. The status and challenge of global fire modelling. *Biogeosciences* **2016**, *13*, 3359–3375. [[CrossRef](#)]
11. Head, L. Transformative change requires resisting a new normal. *Nat. Clim. Chang.* **2020**, *10*, 173–174. [[CrossRef](#)]

12. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [[CrossRef](#)]
13. Kim, S.J.; Lim, C.H.; Kim, G.S.; Lee, J.; Geiger, T.; Rahmati, O.; Son, Y.; Lee, W.K. Multi-temporal analysis of forest fire probability using socio-economic and environmental variables. *Remote Sens.* **2019**, *11*, 86. [[CrossRef](#)]
14. Lee, S.W.; Davidson, R.A. Physics-based simulation model of post-earthquake fire spread. *J. Earthq. Eng.* **2010**, *14*, 670–687. [[CrossRef](#)]
15. Hoffman, C.M.; Canfield, J.; Linn, R.R.; Mell, W.; Sieg, C.H.; Pimont, F.; Ziegler, J. Evaluating Crown Fire Rate of Spread Predictions from Physics-Based Models. *Fire Technol.* **2016**, *52*, 221–237. [[CrossRef](#)]
16. Haines, D.A. *Relation between the National Fire Danger Spread Component and Fire Activity in the Lake States*; North Central Forest Experiment Station, Forest Service, U.S. Department of Agriculture: St. Paul, MN, USA, 1970.
17. Martell, D.L.; Stocks, B.J. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Can. J. For. Res.* **1987**, *17*, 394–401. [[CrossRef](#)]
18. Levi, M.R.; Bestelmeyer, B.T. Biophysical influences on the spatial distribution of fire in the desert grassland region of the southwestern USA. *Landsc. Ecol.* **2016**, *31*, 2079–2095. [[CrossRef](#)]
19. Mohajane, M.; Costache, R.; Karimi, F.; Bao Pham, Q.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecol. Indic.* **2021**, *129*, 107869. [[CrossRef](#)]
20. Archibald, S.; Roy, D.P.; van Wilgen, B.W.; Scholes, R.J. What limits fire? An examination of drivers of burnt area in Southern Africa. *Glob. Chang. Biol.* **2009**, *15*, 613–630. [[CrossRef](#)]
21. Forkel, M.; Dorigo, W.; Lasslop, G.; Teubner, I.; Chuvieco, E.; Thonicke, K. A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). *Geosci. Model Dev.* **2017**, *10*, 4443–4476. [[CrossRef](#)]
22. Forkel, M.; Andela, N.; P Harrison, S.; Lasslop, G.; Van Marle, M.; Chuvieco, E.; Dorigo, W.; Forrest, M.; Hantson, S.; Heil, A.; et al. Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models. *Biogeosciences* **2019**, *16*, 57–76. [[CrossRef](#)]
23. Kuhn-Régnier, A.; Voulgarakis, A.; Nowack, P.; Forkel, M.; Prentice, I.C.; Harrison, S. Quantifying the Importance of Antecedent Fuel-Related Vegetation Properties for Burnt Area using Random Forests. *Biogeosciences* **2020**, 1–24. [[CrossRef](#)]
24. Maingi, J.K.; Henry, M.C. Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. *Int. J. Wildl. Fire* **2007**, *16*, 23. [[CrossRef](#)]
25. Andela, N.; Morton, D.C.; Giglio, L.; Paugam, R.; Chen, Y.; Hantson, S.; van der Werf, G.R.; Randerson, J.T. The Global Fire Atlas of individual fire size, duration, speed, and direction. *Earth Syst. Sci. Data* **2019**, *11*, 529–552. [[CrossRef](#)]
26. Defourny, P. ESA Land Cover Climate Change Initiative (Land_Cover_cci): Land Cover Maps, v2.0.7. Centre for Environmental Data Analysis. 2019. Available online: <https://catalogue.ceda.ac.uk/uuid/b382ebe6679d44b8b0e68ea4ef4b701c> (accessed on 15 May 2021).
27. Harris, I.; Osborn, T.J.; Jones, P.; Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **2020**, *7*, 109. [[CrossRef](#)]
28. Rudolf, B.; Beck, C.; Grieser, J.; Schneider, U. *Global Precipitation Analysis Products of the GPCC*; Global Precipitation Climatology Centre (GPCC): Offenbach, Germany, 2015.
29. Myneni, R.; Knyazikhin, Y.; Park, T. MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006. 2015. *NASA EOSDIS Land Process. DAAC* **2015**. [[CrossRef](#)]
30. Moesinger, L.; Dorigo, W.; De Jeu, R.; Van Der Schalie, R.; Scanlon, T.; Teubner, I.; Forkel, M. The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA). *Earth Syst. Sci. Data* **2020**, *12*, 177–196. [[CrossRef](#)]
31. Santoro, M.; Cartus, O. ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the year 2017, v1. *Cent. Environ. Data Anal.* **2019**. [[CrossRef](#)]
32. Goldewijk, K.K.; Beusen, A.; Doelman, J.; Stehfest, E. Anthropogenic land use estimates for the Holocene—HYDE 3.2. *Earth Syst. Sci. Data* **2017**, *9*, 927–953. [[CrossRef](#)]
33. Meijer, J.R.; Huijbregts, M.A.J.; Schotten, K.C.G.J.; Schipper, A.M. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* **2018**, *13*, 064006. [[CrossRef](#)]
34. Kummu, M.; Taka, M.; Guillaume, J.H.A. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Sci. Data* **2018**, *5*, 10–13. [[CrossRef](#)] [[PubMed](#)]
35. Gilbert, M.; Nicolas, G.; Cinardi, G.; Van Boeckel, T.P.; Vanwambeke, S.O.; Wint, G.R.W.; Robinson, T.P. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **2018**, *5*, 180227. [[CrossRef](#)] [[PubMed](#)]
36. Popp, A.; Calvin, K.; Fujimori, S.; Havlik, P.; Humpenöder, F.; Stehfest, E.; Bodirsky, B.L.; Dietrich, J.P.; Doelmann, J.C.; Gusti, M.; et al. Land-use futures in the shared socio-economic pathways. *Glob. Environ. Chang.* **2017**, *42*, 331–345. [[CrossRef](#)]
37. Schroeder, W.; Giglio, L. NASA VIIRS Land Science Investigator Processing System (SIPS) Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m & 750 m Active Fire Products: Product User’s Guide Version 1.4. *Nasa* **2018**, *1*, 23.
38. Poulter, B.; Ciais, P.; Hodson, E.; Lischke, H.; Maignan, F.; Plummer, S.; Zimmermann, N.E. Plant functional type mapping for earth system models. *Geosci. Model Dev.* **2011**, *4*, 993–1010. [[CrossRef](#)]
39. Venevsky, S.; Thonicke, K.; Sitch, S.; Cramer, W. Simulating fire regimes in human-dominated ecosystems: Iberian Peninsula case study. *Glob. Chang. Biol.* **2002**, *8*, 984–998. [[CrossRef](#)]

40. Bistinas, I.; Harrison, S.P.; Prentice, I.C.; Pereira, J.M.C. Causal relationships versus emergent patterns in the global controls of fire frequency. *Biogeosciences* **2014**, *11*, 5087–5101. [[CrossRef](#)]
41. Gruber, A.; Scanlon, T.; Van Der Schalie, R.; Wagner, W.; Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **2019**, *11*, 717–739. [[CrossRef](#)]
42. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [[CrossRef](#)]
43. Jiang, C.; Ryu, Y.; Fang, H.; Myneni, R.; Claverie, M.; Zhu, Z. Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. *Glob. Chang. Biol.* **2017**, *23*, 4133–4146. [[CrossRef](#)]
44. Van Der Schalie, R.; Kerr, Y.H.; Wigneron, J.P.; Rodríguez-Fernández, N.J.; Al-Yaari, A.; Jeu, R.A.M.d. Global SMOS Soil Moisture Retrievals from The Land Parameter Retrieval Model. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *45*, 125–134. [[CrossRef](#)]
45. Santoro, M.; Cartus, O.; Carvalhais, N.; Rozendaal, D.M.A.; Avitabile, V.; Araza, A.; De Bruin, S.; Herold, M.; Quegan, S.; Rodríguez-Veiga, P.; et al. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth Syst. Sci. Data* **2021**, *13*, 3927–3950. [[CrossRef](#)]
46. Pesaresi, M.; Ehrlich, D.; Florczyk, A.J.; Freire, S.; Julea, A.; Kemper, T.; Soille, P.; Syrris, V. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs*; Publications Office of the European Union: Luxembourg, Luxembourg, 2016; ISBN 9789279550126.
47. Valcke, S.; Piacentini, A.; Jonville, G. Benchmarking Regridding Libraries Used in Earth System Modelling. *Math. Comput. Appl.* **2022**, *27*, 31. [[CrossRef](#)]
48. Pavlov, Y.L. Random forests. In *Text Mining with Machine Learning*; CRC Press: Boca Raton, FL, USA, 2019; pp. 1–122. [[CrossRef](#)]
49. Ghogh, B.; Crowley, M. The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial. *arXiv* **2019**, arXiv:1905.12787.
50. Bisong, E. Introduction to Scikit-learn. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Apress: Berkeley, CA, USA, 2019; pp. 215–229.
51. García, C.B.; García, J.; López Martín, M.M.; Salmerón, R. Collinearity: Revisiting the variance inflation factor in ridge regression. *J. Appl. Stat.* **2015**, *42*, 648–661. [[CrossRef](#)]
52. Salmerón, R.; García, C.B.; García, J. Variance Inflation Factor and Condition Number in multiple linear regression. *J. Stat. Comput. Simul.* **2018**, *88*, 2365–2384. [[CrossRef](#)]
53. Apley, D.W.; Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1059–1086. [[CrossRef](#)]
54. Shearman, T.M.; Varner, J.M.; Hood, S.M.; Cansler, C.A.; Hiers, J.K. Modelling post-fire tree mortality: Can random forest improve discrimination of imbalanced data? *Ecol. Modell.* **2019**, *414*, 108855. [[CrossRef](#)]
55. Barker, J.S.; Gray, A.N.; Fried, J.S. The Effects of Crown Scorch on Post-fire Delayed Mortality Are Modified by Drought Exposure in California (USA). *Fire* **2022**, *5*, 21. [[CrossRef](#)]
56. Almunia, J. Measuring progress, true wealth, and the well-being of nations. *Brussels* **2007**, *4*, 551–594.
57. Asefa, M.; Cao, M.; He, Y.; Mekonnen, E.; Song, X.; Yang, J. Ethiopian vegetation types, climate and topography. *Plant Divers.* **2020**, *42*, 302–311. [[CrossRef](#)] [[PubMed](#)]
58. Gil-Romera, G.; Adolf, C.; Benito, B.M.; Bittner, L.; Johansson, M.U.; Grady, D.A.; Lamb, H.F.; Lemma, B.; Fekadu, M.; Glaser, B.; et al. Long-term fire resilience of the Ericaceous Belt, Bale Mountains, Ethiopia. *Biol. Lett.* **2019**, *15*, 20190357. [[CrossRef](#)] [[PubMed](#)]
59. Guo, F.; Zhang, L.; Jin, S.; Tigabu, M.; Su, Z.; Wang, W. Modeling anthropogenic fire occurrence in the boreal forest of China using logistic regression and random forests. *Forests* **2016**, *7*, 250. [[CrossRef](#)]
60. Starns, H.D.; Fuhlendorf, S.D.; Elmore, R.D.; Twidwell, D.; Thacker, E.T.; Hovick, T.J.; Luttbeg, B. Recoupling fire and grazing reduces wildland fuel loads on rangelands. *Ecosphere* **2019**, *10*, e02578. [[CrossRef](#)]
61. Priya, R.; Ramesh, D.; Khosla, E. Naïve Bayes MapReduce Precision Agricultural Model. In Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; pp. 99–104.
62. Ying, H.; Yin, Y.; Zheng, H.; Wang, Y.; Zhang, Q.; Xue, Y.; Stefanovski, D.; Cui, Z.; Dou, Z. Newer and select maize, wheat, and rice varieties can help mitigate N footprint while producing more grain. *Glob. Chang. Biol.* **2019**, *25*, 4273–4281. [[CrossRef](#)]
63. Di Mauro, G.; Cipriotti, P.A.; Gallo, S.; Rotundo, J.L. Environmental and management variables explain soybean yield gap variability in Central Argentina. *Eur. J. Agron.* **2018**, *99*, 186–194. [[CrossRef](#)]
64. Alvarado, S.T.; Silva, T.S.F.; Archibald, S. Management impacts on fire occurrence: A comparison of fire regimes of African and South American tropical savannas in different protected areas. *J. Environ. Manag.* **2018**, *218*, 79–87. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.