Modelling the drivers of natural fire activity: the bias created by cropland fires

Ismail Bekar\textsuperscript{A,B} and Çağatay Tavşanoğlu\textsuperscript{A,C}

\textsuperscript{A}Fire Ecology and Seed Research Laboratory, Department of Biology, Hacettepe University, Beytepe TR-06800, Ankara, Turkey.
\textsuperscript{B}Present address: Forest Ecology, Institute of Terrestrial Ecosystems, ETH Zurich, Universitätsstrasse 16, CH-8092 Zürich, Switzerland.
\textsuperscript{C}Corresponding author. Email: ctavsan@hacettepe.edu.tr

Abstract. Wildland and cropland fires, which differ considerably in fire regime characteristics, have often been evaluated jointly to estimate regional or global fire regimes using satellite-based fire activity data. We hypothesised that excluding cropland fires will change the output of the models regarding the drivers of natural fire activity. We modelled MODIS fire activity data of western and southern Turkey for the years 2000–2015 using binomial generalised linear models in which many climatic, anthropogenic and geographic factors were included as predictor variables. For modelling, we used different datasets created by the exclusion of various cropland and vegetation land cover classes. More fire activity was observed as the number of cropland-dominated cells increased in a dataset. The explained deviance (%) of the binomial GLM differed substantially in the separate datasets for most of the variables. Moreover, excluding croplands gradually from the overall dataset resulted in a substantial decrease in the explained deviance (%) in the models for all variables. The results suggest that cropland fires have a significant effect on the output of fire regime models. Therefore, a clear distinction should be drawn between wildland and cropland fires in such models for a better understanding of natural fire activity.

Additional keywords: agricultural fire, climate, land cover, Mediterranean Basin, Turkey.

Received 9 October 2016, accepted 18 July 2017, published online 20 September 2017

Introduction

Croplands constitute 11% of Earth’s terrestrial surface, corresponding to nearly 40% of the Earth’s surface suitable for crop production (Bruinsma 2003). The increasing demand for food supply is expected to impose an increase in crop production in the near future (Godfray et al. 2010; Tilman et al. 2011). Fire has been widely used in agricultural management during harvesting, post-harvesting or pre-planting periods since the Neolithic Revolution (Turner et al. 2010). Such agricultural fires are widespread, comprising ~10% of total fires on Earth (Korontzi et al. 2006), and are among the principal causes of wildland fires (Leone et al. 2009; Ganteaume et al. 2013). Despite increasing research on fire ecology and fire regimes during the last two decades, the patterns of agricultural fires have remained unclear (Tulbure et al. 2011). The most studied properties of agricultural fires are their contribution to air pollution (Stohl et al. 2006; Li et al. 2010) and atmospheric gas emissions (Andreac and Merlet 2001; van der Werf et al. 2010; McCarty 2011; McCarty et al. 2012).

Owing to development in satellite fire detection (Justice et al. 2002; Giglio et al. 2003), we have been able to observe fires anywhere on Earth since 2000. Using this technology, we have been able to document and understand fire regimes on a regional and global scale (Chuvieco et al. 2008; Archibald et al. 2013; Murphy et al. 2013; Pausas and Ribeiro 2013). Accordingly, a large number of cropland fires documented by satellite data have also been acknowledged in many regions (Korontzi et al. 2006; Chuvieco et al. 2008).

However, cropland fires have often been ignored in studies determining the drivers of fire activity and regimes on a regional or global basis, especially if satellite fire activity data are used (Murphy et al. 2013; Pausas and Ribeiro 2013; Curt et al. 2015). Consequently, in most cases, cropland and wildland fires have been evaluated jointly. However, wildland and cropland fires clearly differ in fire regime characteristics, such as season, frequency and severity of fire (Le Page et al. 2010; Amraoui et al. 2013; Benali et al. 2017). Wildland fires are mostly related to the weather and climatic factors, and have a relatively shorter fire season, whereas cropland fires are set by humans and have mostly longer fire season, depending on the type of the agriculture (Chuvieco et al. 2008). Since such differences exist in many regions, lack of differentiation between cropland and wildland fires in models of fire regimes may lead to an under-estimation or over-estimation of natural fire regimes. This drawback would also create a problem in predicting future fire regimes under different climate change scenarios.
To date, several approaches have been used to detect cropland fires from satellite fire activity data. One approach considers the characteristics of fires or fire regimes, such as the frequency of fires, the duration of individual fires, the length of fire season and inter-annual variability in fire seasons. This approach has been used to define fire activity dominated by cropland fires (Chuvieco et al. 2008; Archibald et al. 2013) or to classify individual fires as cropland fires (Amraoui et al. 2013). Using this approach, for example, Chuvieco et al. (2008) were able to reveal the relative importance of cropland fires in different biomes and ecoregions. It is also possible to monitor and detect cropland fires by coupling satellite fire activity data and a land cover dataset (Korontzi et al. 2006; Le Page et al. 2010; van der Werf et al. 2010; Tulbure et al. 2011; Magi et al. 2012; Murphy et al. 2013; Curt et al. 2015; Rabin et al. 2015; Hall et al. 2016; Xie et al. 2016). The latter approach allows us to detect cropland fires from satellite data by considering fire activity on croplands. Korontzi et al. (2006) showed that global agricultural fire activity data revealed with this methodology is consistent with regional agricultural practices and crop production data, and therefore, is reliable.

The aim of our study is to understand the effect of considering cropland fires on the results of modelling of drivers of natural fire activity using satellite fire data. Considering the difference of cropland fires in frequency and season in comparison with wildland fires, we hypothesised that excluding cropland fires will change model output significantly. To test this hypothesis, we selected a region comprising various fire regimes, including wildland and cropland fires in the Mediterranean Basin, and used a modelling approach based on the exclusion of various vegetation and cropland land cover classes.

**Methods**

**Study area**

The study area is located in western and southern regions of Turkey and covers 199,935 km² in the eastern Mediterranean Basin (between 35.8 and 40.4°N and 26.2 and 36.8°E). The climate is typically Mediterranean, with dry summers and mild winters. Moreover, it shows substantial variability in temperature and precipitation within the study area (Bekar 2016). Mediterranean Turkey includes various vegetation types, such as sclerophyllous shrublands, coniferous forests (Pinus brutia, P. nigra, Cedrus libani, Abies cilicica) and sub-alpine grasslands (Atalay 1994). Moreover, the study area has been subjected to intense human activity for millennia, resulting in a significant impact on the landscape (Kaniewski et al. 2007; Şekerçioğlu et al. 2011), including human-induced fires for transforming naturally vegetated areas to pastures and croplands (Kaniewski et al. 2008).

**Data**

We used an 8-day summary of fire activity data for cells ~1 km² in size (MOD14A2) obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) instruments on the Terra satellite for the years 2000–2015 (Fig. 1). MODIS fire activity data have been proved a good proxy of active fires on Earth (Csiszar et al. 2006; Pausas and Ribeiro 2013).

We used several variables to explain the distribution of fire activity data throughout the study area, including climatic, anthropogenic and topographic variables (Table 1; Fig. 2). We obtained the agricultural land area size, livestock population size and human population size from the Turkish Statistical Institute at the district level (minimum, mean, and maximum district sizes were 5524, 11761, and 20177 km²) for the year 2015 (see http://www.tuik.gov.tr/, accessed 31 July 2016), and OpenStreetMap road network data for 2016 (Haklay and Weber 2008), as indicators of the human impact. To transform district-level data into GIS layers, data were first imported into ArcGIS (ver. 10.3.1, ESRI, Redlands, CA, USA). Subsequently, an interpolation technique was applied using the Inverse Distance Weighting tool in ArcGIS. GlobCover land cover data was transformed into ~1 km² (30 arc seconds) using ArcGIS.
resample tool. For climate data, we used 19 bioclimatic variables from the WorldClim database (BIOCLIM; Hijmans et al. 2005), and monthly potential evapotranspiration (PET) variables from the Global Aridity and PET databases (Consortium for Spatial Information 2016) for the period between 1960 and 1990. The Normalized Difference Vegetation Index (NDVI) dataset for 2015 (MOD13A3; National Aeronautics and Space Administration 2016) was used as a proxy of productivity and vegetation cover. Geographical variables, such as elevation, slope, solar radiation and terrain ruggedness index, were derived from the digital elevation model from the Consortium for Spatial Information database (Jarvis et al. 2008).

GlobCover land cover data (Bontemps et al. 2011) were used to create separate datasets by including or excluding cells (pixels) dominated by croplands or vegetation (based on United Nations Land Cover Classification System; Fig. 3). We used a simple exclusion method to create datasets, similar to the common technique of excluding cloud pixels from remotely sensed data (e.g. Zhu and Woodcock 2012). To create dataset 2, for example, 1 × 1-km cells classified as ‘cropland’ and ‘vegetation–cropland mosaic’ were excluded from dataset 1, which is the dataset that includes all cells in the study area. The same procedure was applied for creating the other datasets (Table 2). Dataset 2 and dataset 3 were created by excluding cropland-dominated cells in dataset 1, whereas dataset 4 and dataset 5 were created by excluding vegetation-dominated cells. As a consequence of the exclusion process, we obtained five datasets that differ in the land cover type that was included (Table 2). That is, the overall dataset (dataset 1), vegetation-dominated datasets (dataset 2 and dataset 3) and cropland-dominated datasets (dataset 4 and dataset 5).

Modelling

We modelled the contribution of anthropogenic, climatic, geographic factors and productivity to explain fire activity in the study area for each dataset. We considered two types of fire activity variables: the number of fires (count data with discrete structure) and the presence of fire (presence and absence data with binary structure) in the entire study period (years between 2000 and 2015). We used generalised linear models (GLMs) to elucidate the relative contribution of different variables in explaining fire activity in the study area. GLMs have many advantages over standard linear models or regression approaches; in particular, they are more flexible by assuming various families of probability distributions (Guisan et al. 2002; O’Hara and Kotze 2010). As the fire activity data had excess zeros, we used a zero-inflated negative binomial generalised linear model (ZINB-GLM) to handle zeros and over-dispersion in the count data (Zeileis et al. 2008; Brown et al. 2015). For presence and absence data, we used a binomial GLM. However, as preliminary analyses showed that ZINB-GLMs and binomial GLMs yielded similar results (Bekar 2016), we presented and interpreted only the results of the binomial GLM, as a simpler modelling approach (ZINB-GLM results are also given in Table S1). We considered the percentage of the explained deviance as the main model output when comparing different datasets for each variable, because all models yielded significant results (P < 0.0001 in most cases), even though the effect size is small due to the very large number of cells in the datasets.

Using the exclusion procedure summarised above, we compared model outputs for each variable in different datasets to reveal the effect of cropland fires on model results and the relative contribution of each variable to fire activity. All analyses were conducted using the R statistical software (ver. 3.2.0; R Foundation for Statistical Computing, Vienna, Austria; see https://www.R-project.org/, accessed 31 July 2016), and ZINB-GLMs were performed using the pscl package in R (Jackman et al. 2015).

Results

In the study area, fire activity was observed in 7.4% of the cells (based on dataset 1). The percentage of cells in which fire activity

---

### Table 1. Variables used in the models

All data were transformed to GIS layers at ~1 km² pixel resolution using ArcGIS. BIOCLIM, bioclimatic variables from the WorldClim database; PET, potential evapotranspiration; NDVI, Normalised Difference Vegetation Index

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Original resolution</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Climatic variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIOCLIM</td>
<td>Includes 19 bioclimatic variables (°C for temperature, mm for precipitation)</td>
<td>1 × 1 km</td>
<td>Raster</td>
</tr>
<tr>
<td>PET</td>
<td>Includes 12 monthly PET variables (mm)</td>
<td>1 × 1 km</td>
<td>Raster</td>
</tr>
<tr>
<td><strong>Anthropogenic variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human population size</td>
<td>Number of inhabitants</td>
<td>District</td>
<td>Excel</td>
</tr>
<tr>
<td>Livestock population size</td>
<td>Total number of sheep, goats, and cattle</td>
<td>District</td>
<td>Excel</td>
</tr>
<tr>
<td>Agricultural area size</td>
<td>Agricultural area (decare)</td>
<td>District</td>
<td>Excel</td>
</tr>
<tr>
<td>Road network</td>
<td>Total length of roads (km)</td>
<td>–</td>
<td>Polyline</td>
</tr>
<tr>
<td><strong>Geographic variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>Slope (°)</td>
<td>90 m</td>
<td>Raster</td>
</tr>
<tr>
<td>Aspect</td>
<td>Aspect relative to north</td>
<td>90 m</td>
<td>Raster</td>
</tr>
<tr>
<td>Elevation</td>
<td>Altitude from sea level (m)</td>
<td>1 × 1 km</td>
<td>Raster</td>
</tr>
<tr>
<td>Ruggedness Index</td>
<td>Terrain Ruggedness Index (m)</td>
<td>90 m</td>
<td>Raster</td>
</tr>
<tr>
<td>Solar radiation</td>
<td>Annual insolation (Wh m⁻²)</td>
<td>90 m</td>
<td>Raster</td>
</tr>
<tr>
<td><strong>Productivity and vegetation variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI</td>
<td>Includes 12 monthly NDVI variables</td>
<td>1 × 1 km</td>
<td>Raster</td>
</tr>
<tr>
<td>GlobCover</td>
<td>Categorical land cover data</td>
<td>300 m</td>
<td>Raster</td>
</tr>
</tbody>
</table>
was recorded decreased dramatically both in dataset 2 (3.6%) and dataset 3 (3.7%), whereas it increased slightly in dataset 4 (8.8%) and markedly in dataset 5 (12.0%) (Table 3). Consequently, both cropland and wildland fires were present in the study area; however, relatively more fire activity was observed as the number of cropland-dominated cells increased in a dataset.

The explained deviance (%) of the binomial GLM substantially differed in different datasets for most of the variables (Table 3). In dataset 2 and dataset 3, mean temperature of the wettest quarter (BIO 8) and the annual mean temperature (BIO 1) were the most explanatory climatic variables, whereas in dataset 4 and dataset 5, the most explanatory variables were mean temperature of the driest quarter (BIO 9) and mean temperature of the warmest quarter (BIO 10). The mean temperature of the coldest quarter (BIO 11) also explained a significant proportion of deviance in many datasets. The explained deviance (%) of the model of the human population size was twice as high in dataset 4 as in dataset 2 and dataset 3. Similarly, the explained deviance of the slope variable was also twice as high in dataset 4 and dataset 5 as in dataset 2 and dataset 3. In all datasets, the most explanatory productivity variable was the NDVI of December. NDVI values of December in dataset 4 and dataset 5 were comparable with those in dataset 1; however, NDVI values of July in the same datasets were
remarkably higher than those in dataset 1. Moreover, the explained deviance (%) values of NDVI variables clearly differed between cropland-dominated and vegetation-dominated datasets (Table 3). Elevation was among the variables that have the most explained deviance (%) in all datasets except dataset 5, as croplands are concentrated at lower elevations (Pekin 2016), and fire activity in natural vegetation is related to elevation (Bekar 2016) in the study region.

Excluding croplands gradually from the overall dataset resulted in a substantial decrease of the explained deviance (%) in the models for all variables. The decrease reached up to 8% in the mean temperature of the driest quarter (BIO9) and the mean temperature of the warmest quarter (BIO10), and 7% in the annual mean temperature (BIO1) and slope. On the contrary, excluding vegetation-dominated cells resulted in an increase of the explained deviance (%) in most variables in dataset 4. However, dataset 5, which includes only croplands, had comparable or lower explained deviance (%) values than dataset 1.

### Discussion

Our study showed that the explanatory power of predictor variables of fire activity data vary to a considerable extent in different datasets. The most apparent differences were between datasets including vegetation-dominated and cropland-dominated cells. This implies that cropland fires substantially affect the output of models including satellite fire activity data.

The possible problems that can be created by human-induced cropland fires in fire-regime modelling have already been

---

**Table 2. GlobCover land cover classes included in different datasets**

‘Cropland’ refers to the cells (1 km²) classified as irrigated or rain-fed cropland (without any vegetation), and ‘vegetation’ indicates the cells that are composed completely of vegetation and do not include any cropland. ‘Cropland–vegetation mosaic’ includes 50–70% cropland and 20–50% vegetation. ‘Vegetation–cropland mosaic’ includes 50–70% vegetation and 20–50% cropland. ‘+’ and ‘−’ indicate that a land cover class in a dataset was included or excluded.

<table>
<thead>
<tr>
<th>Land cover class</th>
<th>Cropland</th>
<th>Cropland–vegetation mosaic</th>
<th>Vegetation–cropland mosaic</th>
<th>Vegetation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The overall dataset</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Vegetation-dominated datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset 1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Cropland-dominated datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset 4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Dataset 5</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

**Table 3. Explained deviance (%) values (based on binomial generalised linear models, GLMs) for the best predictors of fire activity for each variable group in the study area for each dataset**

\( T_{\text{mean}} \) refers to the mean temperature. All models are statistically significant \( (P < 0.0001) \) except for NDVI-July in dataset 2, where \( P = 0.001 \). Total fire activity is the number of cells where fire activity was observed for years 2000–2015. Number of cells is the total number of cells within the study area in the corresponding dataset.
acknowledged (Archibald et al. 2013). However, the exclusion of agricultural fires in large-scale studies is problematic for fire modelling studies, because it is difficult to identify every agricultural practice on Earth by means of characteristics of fires. Characteristics of agricultural fires, such as season and duration, vary regionally, and a great deal of effort is required to distinguish agricultural fires from wildland fires globally using this methodology. Even though agricultural fires can be easily excluded from analyses when official fire statistics from national inventories are used, such a procedure cannot be applied in all cases, especially with fire activity data obtained from remote sensing. For example, Giglio et al. (2006) found that low fire radiative power is associated not only with areas of extensive croplands but also with heavily forested areas in some parts of the world.

Using land cover maps obtained from satellites presents a good solution to this problem. Recently, the increasing reliability of land cover databases has made them available for differentiating agricultural fires from wildland fires (Tulbure et al. 2011; Amraoui et al. 2013; Knorr et al. 2014). Similar approaches have successfully been used to estimate the percentage of burned areas in land cover classes (van der Werf et al. 2010) and to estimate post-fire regeneration of natural landscapes using satellite remote sensing data (Abdul Malak et al. 2015). Another approach may be the inclusion of agricultural areas within vegetation variables as a separate land cover class in order to understand the fire regimes of anthropogenic and natural origin in a region (e.g. Curt et al. 2015). However, such an approach will leave unanswered the question of what the drivers of natural fire regimes are, because this cannot be inferred from models including both agricultural lands and vegetation-covered areas. The issue of cropland fires may also affect the conclusions of future studies on fire regimes in relation to climate change, even if these studies are conducted at a lower resolution (e.g. Krawchuk et al. 2009). Since wildland and agricultural fires clearly differ in fire regime characteristics, such as intensity and season (Korontzi et al. 2006; Le Page et al. 2010; Benali et al. 2017), the differentiation of wildland and cropland fires in the evaluation of the drivers of fire regimes has the potential to improve fire regime models. Our study proved that land cover databases could be used for the purpose of differentiating agricultural fires from wildland fires when satellite-based fire activity data are used. The observed differences in the explained deviances of predictor variables in our models dominated by different land cover classes suggest that drivers of natural and agricultural fire activity can be distinguished by incorporating land cover.

In conclusion, our results suggest that cropland fires have a significant effect on the output of fire regime models. Therefore, a clear distinction should be drawn between wildland and cropland fires in such models for a better understanding of natural fire activity.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

We thank Alper Ertürk and Çağışan Karacagölü for their advice on data processing using GIS at the beginning of the work. We also thank two anonymous reviewers for their suggestions improving the manuscript. This study was supported by Hacettepe University Scientific Research Coordination Unit (project number FBB-2016-10110), and was part of the requirements for the M.Sc. thesis of İ. Bekar submitted to Hacettepe University.

References


Bekar I (2016) The role of anthropogenic and natural factors in shaping recent fire regimes in Mediterranean ecosystems. MSc thesis, Hacettepe University. [In Turkish with English summary]


National Aeronautics and Space Administration (2016) Vegetation Index. Available at http://go.nasa.gov/1UMSkgb [Verified 31 July 2016]


