Long-term Burned Area Reconstruction through Deep Learning

Seppe Lampe 1  Bertrand Le Saux 2  Inne Vanderkelen 1  Wim Thiery 1

Abstract
Wildfire impact studies are significantly hampered by the absence of a global long-term burned area dataset. This prevents conclusive statements on the role of anthropogenic activity on wildfire impacts over the last century. Here, we propose a workflow to construct a 1901-2014 reanalysis of monthly global burned area at a 0.5° by 0.5° scale. A neural network will be trained with weather-related, vegetational, societal and economic input parameters, and burned area as output label for the 1982-2014 time period. This model can then be applied to the whole 1901-2014 time period to create a data-driven, long-term burned area reanalysis. This reconstruction will allow to investigate the long-term effect of anthropogenic activity on wildfire impacts, will be used as basis for detection and attribution studies and could help to reduce the uncertainties in future predictions.

1. Introduction
In recent years, there has been an unusually extensive wildfire activity all over the world. Forest fires raged across California in 2017, 2018 and 2020, Australia faced unprecedented bushfires in 2019-2020, and even Siberia was hit by wildfires in 2019 and 2020. Events like these cause a direct loss of life, with for instance 100 fatalities during the 2018 California wildfires and wildfire-induced respiratory problems causing premature deaths in large parts of the world (Reid et al., 2016; Porter et al., 2019; Matz et al., 2020). In addition, wildfires lead to significant economic damages and costs for fire suppression (Strader, 2018; Goss et al., 2020). While regular-sized wildfires sustain biodiversity and ecosystem health, megafires have clear adverse effects on ecosystems and biodiversity (Driscoll et al., 2010; North et al., 2015; Doerr & Santín, 2016; Andela et al., 2017). During the 2019-2020 Australian bushfires, an estimated one billion animals were killed, while hundreds of Australian plant and animal species now face extinction (DAWE, 2020; Filkov et al., 2020; Wintle et al., 2020).

Evaluating the imprint of human activity on climatic variables and impacts is done via the application of detection and attribution methodologies (Field et al., 2014). Detection refers to the process of demonstrating that climate or a system affected by climate has changes in a defined statistical sense, whereas attribution implies the evaluation of relative contributions of multiple causal factors to this change given a specific statistical confidence (Field et al., 2014). Instead of looking at wildfire impacts, like burned area, most wildfire detection and attribution studies have traditionally been limited to atmospheric variables, whereby most build on a version of the Fire Weather Index (FWI) (Bindoff et al., 2013; Gudmundsson et al., 2014; Krikken et al., 2019; Kirchmeier-Young et al., 2019; Abatzoglou et al., 2019; Kirchmeier-Young et al., 2019; van Oldenborgh et al., 2020).

The motivation behind the choice of atmospheric variables is that wildfire activity is partly determined by local weather i.e., prolonged periods of dry, hot weather increase the frequency and severity of wildfire activity. However, wildfire activity is actually influenced by a wide range of drivers, including, but not limited to, weather, topography, vegetation type and density, and firefighting measures (Turco et al., 2014; Abatzoglou & Williams, 2016; Goss et al., 2020; Podschwit & Cullen, 2020). Therefore, a more appropriate tool for capturing on-the-ground impacts of wildfires and for investigating the changes in their activity has been recently proposed: the measure of burned area (Abatzoglou & Williams, 2016; Andela et al., 2017).

Despite its relevance for representing wildfire impacts, burned area is much more complex to model compared to fire weather indices, due to all the confounding factors influencing burned area. As a consequence, burned area is typically only poorly represented in current-generation climate models that serve as the default input for detection and attribution studies (van Oldenborgh et al., 2020).

In addition to the challenges of modelling burned area, the absence of a long-term global burned area record hampers
Long-term Burned Area Reconstruction through Deep Learning

2. Data

The project relies on atmospheric (GSWP3-W5E5), vegetational (LUH2) and socioeconomic (ISIMIP3b simulations) data available through ISIMIP, and on burned area data from the FireCCILT11 dataset. An overview of these datasets is given in Figure 1 and in the following paragraphs.

Firstly, GSWP3-W5E5 contains daily reanalyses of the atmospheric climate from two separate datasets i.e., Global Soil Wetness Project Phase 3 (GSWP3) and Watch Forcing Data for ERA5 (W5E5), both of which represent daily global meteorological data on a 0.5° by 0.5° resolution, combined they span from 1901 to 2016 (Dirmeyer et al., 2006; Lange, 2019; Cucchi et al., 2020; Lange, 2020). Each pixel is represented by ten fields i.e., specific humidity, relative humidity, daily maximum, minimum and mean temperature, short and long wave downwelling radiation, surface air pressure and wind magnitude, and total precipitation.

Land use, land cover and land management information is provided in ISIMIP3a by Land Use Harmonization 2 (LUH2), an annual gridded (0.5° by 0.5°) dataset for the years 1850-2018 (Goldewijk et al., 2017; Hurtt et al., 2020). Gross Domestic Product (GDP) and annual population are represented in the ISIMIP3b simulations as an annual country-wide value for the years 1850-2014 and 1850-2020, respectively (Lange, 2020). Population density and wildfire activity are positively correlated, an increase in population density will generally lead to an increase in the number of fires (Krause et al., 2014; Flannigan et al., 2016; Read et al., 2018). There is an anti-correlation between GDP and wildfire activity due to increased fire management (Aldersley et al., 2011), while land use is closely linked to fuel availability (Westerling et al., 2006; Balch et al., 2017).

Lastly, the FireCCILT11 dataset contains global estimates of monthly burned area and is available in two spatial resolutions i.e., 0.05° and 0.25° (Fig. 2). FireCCILT11 is based on the Advanced Very High Resolution Radiometer Land Long Term Data Record (AVHRR-LTDR) and covers the period 1982-2018 with the exception of 1994 (Otón, 2020).

3. Methodology

In this section, we present the reconstruction task with its corresponding inputs and outputs, evaluate the feasibility of training such a model, and propose several architectures to implement in order to capture the complexity of the precursors.

Wildfire activity is governed by a multitude of processes and parameters, most of them related to weather, land use/cover and human activity. Although many of the parameters influencing wildfire activity are known, their exact mathematical relationship to wildfire activity is often not entirely clear.
Thus, to build a well-functioning prediction model, the most important of these parameters need to be considered by the system. Firstly, the characteristics of the local vegetation are most vital factor i.e., if there is no vegetation, there cannot be a wildfire. Secondly, the local weather pattern plays a significant role for wildfire activity through aridity and fuel availability. Lastly, as discussed earlier, the socio-economic development should also be included. Therefore, a neural network with as input (i) LUH2 land cover, land use and land management, (ii) GSWP3-W5E5 atmospheric reanalysis and (iii) ISIMIP3b GDP and population, and burned area (FireCCILT11) as prediction label will be trained. This network will thus consider vegetation-related parameters, the preceding weather pattern and socio-economic factors.

The label dataset (FireCCILT11) spans ~35 years (1982-2016 but 1994 is excluded) with monthly temporal resolution, resulting in 420 (35*12) data samples, where each data sample represents a global map of monthly burned area (Fig. 2). This number is too little for adequate training of a neural network. Therefore, at each time step, each pixel will be considered as a separate data point. As the GSWP3-W5E5 product only spans to 2016 and the ISIMIP3b GDP dataset to 2014, we cannot use the period 2015-2018 of the FireCCILT11 dataset. This results in 31 years of applicable, available data. This will be applied at a 0.5 by 0.5 resolution, resulting in 360 * 180 pixels per month. However, ~2/3 of those pixels represent oceans and seas and will therefore not be included in the model. A rough estimation of the total amount of data samples (31 * 12 * 360 * 180 = 2.7 * 10^6) indicates that there should be sufficient data to train a neural network with the aforementioned parameters.

Several considerations will have to be made during the construction and training of the network. Firstly, the model should optimally consider more than one month of atmospheric data. The dryness of vegetation and soil have a large impact on the occurrence and size of wildfires. This dryness is the result of local weather over the preceding months. Thus, the network should probably consider ~three months of antecedent atmospheric reanalysis data (GSWP3-W5E5). Furthermore, the amount of input data will need to be optimised. Without any changes, the neural network takes as input: ~one value each for land cover, land use, land management, GDP and population density but ~900 values of antecedent daily atmospheric reanalyses (three months * ten parameters per day). Several options are available to reduce the size of these atmospheric reanalyses, and thus reduce the total number of weights in the network e.g., manual selection, temporal upscaling, principle component analysis, etc. Even if the most suitable implementation for this project will need to be determined empirically, likely candidates for modelling such data will be recurrent neural networks such as gated recurrent units (Chung et al., 2014).

By dividing the pixels into separate data points, the network cannot learn geospatial unique information which might have an effect on wildfire activity e.g., topography. If the model does not reach the desired performance, it might be improved by adding a topography-related value for each pixel or analyse further which parameters might be missing. Furthermore, the training period could potentially be expanded to 2018 if suitable replacements can be found for the 2017-2018 period of GSWP3-W5E5 and 2015-2018 period of the ISIMIP3b GDP dataset. Given the recent time period of these data gaps, it is highly probable that there are alternatives for these periods. However, the assessment strategy of these alternatives might slightly differ from the GSWP3-W5E5 and ISIMIP3b GDP datasets. Therefore, we will only consider including these extensions if it is deemed needed. In that case, we will investigate fully-convolutional versions of recurrent networks, such as ConvLSTM (Shi et al., 2015).

If a network can be trained, which is sufficiently accurate and generalizing, it can be applied on the whole 1901-2014 time span to generate a new long-term burned area dataset, spanning 114 years at 0.5° by 0.5° spatial resolution and monthly time step. The reconstruction will be evaluated against existing long-term regional burned area data (e.g. available for California and selected European countries).

4. Conclusion & Discussion

Disentangling the intricate anthropogenic impact on wildfire activity is complex and still under debate. In addition, the lack of a long-term burned area dataset hampers trend detection and attribution in the field of wildfire impact studies. Here, we propose a workflow to train neural networks...
with ISIMIP data as input and the recently published FireC-CILT11 dataset as label to create a 114 year long burned area dataset. This dataset will, for the first time, allow to investigate the long-term effect of anthropogenic activity on wildfire impacts. It will also be the basis of further detection and attribution studies and could potentially reduce the uncertainties in future wildfire activity predictions. Furthermore, this dataset could become an essential asset for applying machine learning in wildfire research and in future applications in the wildfire management field.

5. Acknowledgements

The authors would like to thank the two reviewers for their constructive feedback and the AI4CC organisation for bringing the authors in contact via the mentorship program.

References


Long-term Burned Area Reconstruction through Deep Learning


Lange, S. Wfde5 over land merged with era5 over the ocean (w5e5). v. 1.0. 2019.

Lange, S. ISIMIP3BASD v2. 4.1, 2020.


