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; 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

[1]Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium [2]Φ-lab, European Space Agency, Frascati, Italy. Correspondence to: Seppe Lampe <Seppe.Lampe@vub.be>, Bertrand Le Saux <Bertrand.Le.Saux@esa.int>, Inne Vanderkelen <Inne.Vanderkelen@vub.be>, Wim Thiery <Wim.Thiery@vub.be>.

Tackling Climate Change with Machine Learning Workshop at ICML 2021.

the detection and attribution of wildfire activity (Randerson et al., 2015; Andela et al., 2017; Forkel et al., 2019). Currently, the most widely-used wildfire dataset, Global Fire Emissions Database version 4 (GFED4), provides satellite-derived burned area observations at 0.25° horizontal resolution (Randerson et al., 2015). However, this dataset contains only ~20 years of data, and its validity for long-term trends is highly uncertain, especially on a global scale. Recently (Dec 2020), a new global burned area dataset, Fire Climate Change Initiative Long-Term v1.1 (FireCCILT11), has been released. FireCCILT11 was developed as part of the Fire project of the European Space Agency Climate Change initiative and spans ~36 years thanks to the harmonisation of measurements originating from a range of satellites (Otón, 2020). This dataset provides new opportunities to investigate the long-term changes in wildfire activity impacts. However, despite its significant increase in time span, FireCCILT11 only goes back to 1982, preventing any conclusive statements on wildfire impacts before this period.

Novel machine learning techniques demonstrate a large potential to gap-fill and back-extend climate impact datasets (Humphrey et al., 2017; 2018; Padrón et al., 2020; Ghiggi et al., 2019; Lange, 2020). For instance, data-driven statistical modelling has been used to back-extend satellite-based terrestrial water storage estimates (Humphrey et al., 2017; 2018), whereas a random forest approach enabled the reconstruction of monthly runoff rates (Ghiggi et al., 2019) and renewable freshwater resources (Padrón et al., 2020). The combination of this new dataset (FireCCILT11) with these novel methodological approaches generates momentum to push the boundaries of current wildfire research. Here, we propose a workflow to reconstruct a new long-term (1901-2014) burned area dataset. This dataset will allow to investigate the long-term effect of anthropogenic activity on wildifre impacts, will be part of further detection and attribution studies and can potentially reduce the uncertainties in future wildfire predictions. Furthermore, this dataset can be an essential asset for applying machine learning in wildfire research and in future applications in the wildfire management sector.

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.

Firestly, 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, com-

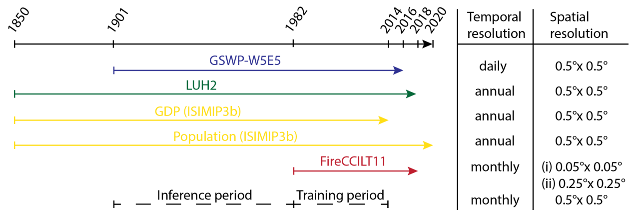

Figure 1. The time span, temporal resolution and spatial resolution of the datasets included in this proposal.

bined 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-LLTDR) 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.

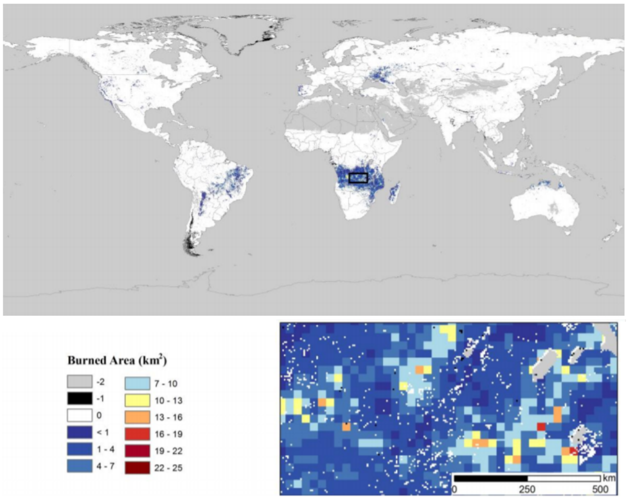

Figure 2. The global burned area of August 1982 according to the FireCCILT11 product (Otón, 2020).

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, $\sim\frac{2}{3}$ 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 ($31 * 12 * 360 * 180 * \frac{1}{3} \approx 8.0$ $* 10^6$) 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 FireCCILT11 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

Abatzoglou, J. T. and Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42):11770–11775, 2016.

Abatzoglou, J. T., Williams, A. P., and Barbero, R. Global Emergence of Anthropogenic Climate Change in Fire Weather Indices. *Geophysical Research Letters*, 46(1): 326–336, 1 2019. ISSN 0094-8276.

Aldersley, A., Murray, S. J., and Cornell, S. E. Global and regional analysis of climate and human drivers of wildfire. *Science of the Total Environment*, 409(18):3472–3481, 2011.

Andela, N., Morton, D. C., Giglio, L., Chen, Y., Van Der Werf, G. R., Kasibhatla, P. S., DeFries, R. S., Collatz, G. J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Yue, C., and Randerson, J. T. A human-driven decline in global burned area. *Science*, 356(6345):1356–1362, 6 2017. ISSN 10959203.

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., and Mahood, A. L. Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, 114(11): 2946–2951, 2017.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., and Others. Detection and attribution of climate change: from global to regional. In *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2013.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, 12(3):2097–2120, 2020.

DAWE. Wildlife and threatened species bushfire recovery research and resources — Department of Agriculture, Water and the Environment, 2020.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N. GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, 87(10):1381–1398, 2006.

Doerr, S. H. and Santín, C. Global trends in wildfire and its impacts: Perceptions versus realities in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696), 6 2016. ISSN 14712970.

Driscoll, D. A., Lindenmayer, D. B., Bennett, A. F., Bode, M., Bradstock, R. A., Cary, G. J., Clarke, M. F., Dexter, N., Fensham, R., Friend, G., Gill, M., James, S., Kay, G., Keith, D. A., MacGregor, C., Russell-Smith, J., Salt, D., Watson James, J. E., Williams Richard J., R. J., and York, A. Fire management for biodiversity conservation: Key research questions and our capacity to answer them, 9 2010. ISSN 00063207.

Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L. *Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects*. Cambridge University Press, 2014.

Filkov, A. I., Ngo, T., Matthews, S., Telfer, S., and Penman, T. D. Impact of Australia's catastrophic 2019/20 bushfire season on communities and environment. Retrospective analysis and current trends. *Journal of Safety Science and Resilience*, 1(1):44–56, 9 2020. ISSN 26664496.

Flannigan, M. D., Wotton, B. M., Marshall, G. A., de Groot, W. J., Johnston, J., Jurko, N., and Cantin, A. S. Fuel moisture sensitivity to temperature and precipitation: climate change implications. *Climatic Change*, 134(1-2):59–71, 1 2016. ISSN 15731480.

Forkel, M., Dorigo, W., Lasslop, G., Chuvieco, E., Hantson, S., Heil, A., Teubner, I., Thonicke, K., and Harrison, S. P. Recent global and regional trends in burned area and their compensating environmental controls. *Environmental Research Communications*, 1(5):051005, 6 2019.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L. GRUN: An observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, 11(4):1655–1674, 2019. ISSN 18663516.

Goldewijk, K. K., Beusen, A., Doelman, J., and Stehfest, E. Anthropogenic land use estimates for the Holocene - HYDE 3.2. *Earth System Science Data*, 9(2):927–953, 12 2017. ISSN 18663516.

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., and Diffenbaugh, N. S. Climate change is increasing the likelihood of extreme autumn wildfire conditions across california. *Environmental Research Letters*, 15(9):094016, 9 2020. ISSN 17489326.

Gudmundsson, L., Rego, F. C., Rocha, M., and Seneviratne, S. I. Predicting above normal wildfire activity in southern Europe as a function of meteorological drought. *Environmental Research Letters*, 9(8):084008, 8 2014. ISSN 17489326.

Humphrey, V., Gudmundsson, L., and Seneviratne, S. I. A global reconstruction of climate-driven subdecadal water storage variability. *Geophysical Research Letters*, 44(5): 2300–2309, 3 2017. ISSN 0094-8276.

Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., and Seneviratne, S. I. Sensitivity of atmospheric CO 2 growth rate to observed changes in terrestrial water storage. *Nature*, 560(7720):628–631, 2018.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., and Others. Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. *Geoscientific Model Development*, 13(11):5425–5464, 2020.

Kirchmeier-Young, M. C., Gillett, N. P., Zwiers, F. W., Cannon, A. J., and Anslow, F. S. Attribution of the Influence of Human-Induced Climate Change on an Extreme Fire Season. *Earth's Future*, 7(1):2–10, 1 2019. ISSN 2328-4277.

Krause, A., Kloster, S., Wilkenskjeld, S., and Paeth, H. The sensitivity of global wildfires to simulated past, present, and future lightning frequency. *Journal of Geophysical Research: Biogeosciences*, 119(3):312–322, 3 2014. ISSN 21698953.

Krikken, F., Lehner, F., Haustein, K., Drobyshev, I., and van Oldenborgh, G. J. Attribution of the role of climate change in the forest fires in Sweden 2018. *Natural Hazards and Earth System Sciences*, pp. 1–24, 2019. ISSN 1561-8633.

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

Lange, S. ISIMIP3BASD v2. 4.1, 2020.

Matz, C. J., Egyed, M., Xi, G., Racine, J., Pavlovic, R., Rittmaster, R., Henderson, S. B., and Stieb, D. M. Health impact analysis of PM2.5 from wildfire smoke in Canada (2013–2015, 2017–2018). *Science of the Total Environment*, 725:138506, 7 2020. ISSN 18791026.

North, M. P., Stephens, S. L., Collins, B. M., Agee, J. K., Aplet, G., Franklin, J. F., and Fule, P. Z. Reform forest fire management. *Science*, 349(6254):1280–1281, 2015.

Otón, G. Esa climate change initiative–fire_cci d4. 2.2 product user guide-avhrr-long term data record (pug-ltdr). 2020.

Padrón, R. S., Gudmundsson, L., Decharme, B., Ducharne, A., Lawrence, D. M., Mao, J., Peano, D., Krinner, G., Kim, H., and Seneviratne, S. I. Observed changes in dry-season water availability attributed to human-induced climate change. *Nature Geoscience*, 13(7):477–481, 7 2020. ISSN 17520908.

Podschwit, H. and Cullen, A. Patterns and trends in simultaneous wildfire activity in the United States from 1984 to 2015. *International Journal of Wildland Fire*, 29(12): 1057, 12 2020. ISSN 1049-8001.

Porter, T. W., Crowfoot, W., and Newsom, G. 2018 Wildfire Activity Statistics. Technical report, 2019.

Randerson, J. T., Van Der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S. Global Fire Emissions Database, Version 4.1 (GFEDv4). *ORNL DAAC*, 2015.

Read, N., Duff, T. J., and Taylor, P. G. A lightning-caused wildfire ignition forecasting model for operational use. *Agricultural and Forest Meteorology*, 253-254:233–246, 5 2018. ISSN 01681923.

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., and Elliott, C. T. Critical Review of Health Impacts of Wildfire Smoke Exposure. *Environmental Health Perspectives*, 124(9):1334–1343, 9 2016. ISSN 0091-6765.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214*, 2015.

Strader, S. M. Spatiotemporal changes in conterminous us wildfire exposure from 1940 to 2010. *Natural hazards*, 92(1):543–565, 2018.

Turco, M., Llasat, M. C., von Hardenberg, J., and Provenzale, A. Climate change impacts on wildfires in a Mediterranean environment. *Climatic Change*, 125(3-4):369–380, 7 2014. ISSN 01650009.

van Oldenborgh, G. J., Krikken, F., Lewis, S., Leach, N., Lehner, F., Saunders, K., van Weele, M., Haustein, K., Li, S., Wallom, D., Sparrow, S., Arrighi, J., Singh, R., van Aalst, M., Philip, S., Vautard, R., and Otto, F. Attribution of the Australian bushfire risk to anthropogenic climate change. *Natural Hazards and Earth System Sciences*, pp. 1–46, 2020. ISSN 1561-8633.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., and Swetnam, T. W. Warming and earlier spring increase Western U.S. forest wildfire activity. *Science*, 313(5789):940–943, 8 2006. ISSN 00368075.

Wintle, B. A., Legge, S., and Woinarski, J. C. After the Megafires: What Next for Australian Wildlife?, 9 2020. ISSN 01695347.