Fire in the Earth System Abstracts
Vol. 1 FES-Artificial Intelligence (AI) and Machine Learning (ML) for wildfires-82
Valencia, Spain, 3-7 November, 2021

Evaluating and comparing statistical and machine learning methods for fire occurrence prediction

[1]Phelps, Nathan and [2]Woolford, Douglas G.

[1]*Department of Statistical and Actuarial Sciences and Department of Computer Science, University of Western Ontario, London N6A 3K7, Canada*
[2]*Department of Statistical and Actuarial Sciences, University of Western Ontario, London N6A 3K7, Canada*

Abstract

Some wildland fire management agencies use fine-scale, spatio-temporal wildland fire occurrence prediction (FOP) models to predict fire occurrence potential and to create spatially-explicit maps highlighting areas that are more likely to have fires. Such FOP models can support fire management decisions such as organized detection and the positioning of suppression resources. FOP modelling techniques are typically empirically-based, including both statistical modelling and machine learning approaches. Since fire occurrences are rare events in space-time, subsampling is commonly employed to create the dataset used to train (i.e., fit) fine-scale FOP models. However, failing to properly calibrate to account for such subsampling can lead to issues, including non-optimal model selection as well as models that can be highly biased towards overprediction. Through a case-study of human-caused FOP in a provincial fire control zone in the Lac La Biche region of Alberta, Canada for 1996 through 2016, we illustrate such issues and demonstrate the need to properly calibrate fine-scale FOP models so that they output true fire occurrence probabilities, discussing methods for proper calibration and providing a set of guidelines for the effective evaluation and comparison of different candidate models. Following these methods, we compare a variety of properly calibrated FOP models for our study region, including bagged classification trees, random forests, neural networks, logistic regression models, and logistic generalized additive models. Although previous studies have suggested that machine learning approaches may outperform logistic regression models, our results show that the more flexible logistic generalized additive modelling approach can be highly competitive with machine learning methods. Moreover, since logistic generalized additive models are commonly viewed as much more interpretable than “black box” machine learning models, this statistical modelling approach may be preferred for FOP. Consequently, we advocate that the pros and cons of statistical and machine learning approaches should be discussed with fire management agencies when collaborating to develop FOP models for operational use. Additionally, future studies of FOP models should compare logistic generalized models with machine learning approaches in other wildland fire ecosystems to determine if logistic generalized additive models are consistently competitive with alternative approaches.

Keywords: artificial intelligence, calibration, forest fire occurrence, wildfire occurrence

Acknowledgments: We acknowledge the support of the Government of Alberta, the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2015–04221 and its Undergraduate Student Research Award program), and the Institute for Catastrophic Loss Reduction.