An Evaluation of the Impact of Ignition Location Uncertainty on Forest Fire Ignition Prediction Using Bayesian Logistic Regression

David Röbl
Institute of Geodesy, Graz University of Technology, Graz, Austria

Rizwan Bulbul
Institute of Geodesy, Graz University of Technology, Graz, Austria

Johannes Scholz
Institute of Geodesy, Graz University of Technology, Graz, Austria

Mortimer M. Müller
Institute of Silviculture, University of Natural Resources and Life Sciences, Vienna, Austria

Harald Vacik
Institute of Silviculture, University of Natural Resources and Life Sciences, Vienna, Austria

Abstract

This study investigates the impact of location uncertainty on the predictive performance of Bayesian Logistic Regression (BLR) for forest fire ignition prediction in Austria. Historical forest fire ignitions are used to create a dataset for training models with the capability to assess the general forest fire ignition susceptibility. Each recorded fire ignition contains a timestamp, the estimated location of the ignition and a radius defining the area within which the unknown true location of the ignition point is located. As the values of the predictive features are calculated based on the assumed location, and not the unknown true location, the training data is biased due to input uncertainties. This study is set to assess the impact of input data uncertainty on the predictive performance of the model. For this we use a data binning approach that splits the input data into groups based on their location uncertainty and use them later for training multiple BLR models. The predictive performance of the models is then compared based on their accuracy, area under the receiver operating characteristic curve (AUC) scores and brier scores. The study revealed that higher location uncertainty leads to decreased accuracy and AUC score, accompanied by an increase in the brier score, while demonstrating that the BLR model trained on a smaller high-quality dataset outperforms the model trained on the full dataset, despite its smaller size. The study's contribution is to provide insights into the practical implications of location uncertainty on the quality of forest fire susceptibility predictions, with potential implications for forest risk management and forest fire documentation.

2012 ACM Subject Classification Theory of computation → Bayesian analysis

Keywords and phrases Forest Fire Prediction, Ignition Location Uncertainty, Bayesian Logistic Regression, Bayesian Inference, Probabilistic Programming

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.62

Category Short Paper

Funding The research was carried out as part of the IGNITE – Improving the Assessment of Forest Fire Susceptibility project, which is funded by the Austrian Forest Fund (Federal Ministry of Agriculture, Forestry, Environment and Water Management).

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 62; pp. 62:1–62:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The impact of forest fires in Europe has been increasingly severe due to climate change, leading to longer fire seasons, expansion of affected areas, and unprecedented conditions for fire-fighting services [12]. In countries such as Austria, forest fire prediction models, which form the backbone of early warning systems, use manually collected incident reports to predict the outbreak and behaviour of forest fires. However, uncertainty in the input data, due to human involvement makes the data susceptible to various uncertainties. In order to create reliable predictive models for forest fires, it is essential to understand how input uncertainty impacts the accuracy of predictions. This study specifically investigates the impact of uncertainty surrounding the initial fire ignition point location on the accuracy of forest fire ignition predictions. Bayesian Logistic Regression (BLR) is a flexible approach for predictive modeling, particularly with input data uncertainties. It provides a robust mathematical model to quantify uncertainty, incorporate prior knowledge, and improve the model's generalization. Unlike the point estimates provided by traditional Logistic Regression (LR), the Bayesian method provides a full predictive posterior distributions, that quantifies input data and model uncertainty [4]. The primary objective of this study is to analyze the sensitivity of BLR models to forest fire ignition location uncertainty by training multiple models using training datasets with different levels of associated uncertainty. For this purpose this study utilizes the Austrian forest fire database, which stores the locations of past fire ignition points. Each point is associated with a positional uncertainty in the form of a distance radius, which determines the area where the forest fire may have started, as shown in Figure 1. The paper is organized as follows. In section 2 we elaborate on the related work, followed by the methodology described in section 3. This section covers data preparation, model training and evaluation. Section 4 covers the results and section 5 discusses the results achieved.

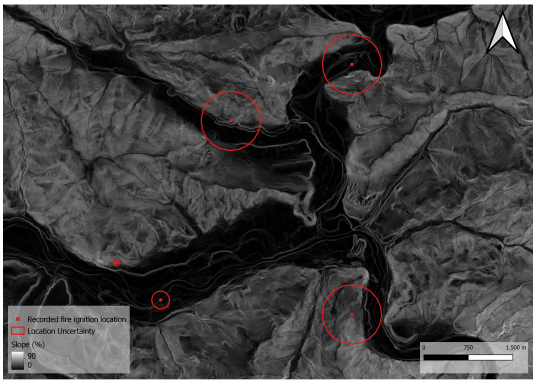

Figure 1 This map displays recorded fire ignition locations and their associated buffers, indicating the uncertainty of each ignition position. The slope raster underneath provides further insight into the terrain, showcasing strong variations within the uncertainty regions.

2 Related Work

Logistic Regression (LR) has been used extensively in wildfire science and management, according to [10], who provided a comprehensive review of Machine Learning (ML) applications in this area. BLR, on the other hand, has seen limited use in wildfire prediction. [5] applied BLR with uninformed priors to estimate the probability of large fires based on weather components, while [8] trained hierarchical BLR models with different priors to estimate the probability of fire occurrence based on forest vulnerability and climatic conditions. While previous studies have investigated the impact of weather conditions, land cover, and human activities on the predictive performance of wildfire fire ignition models using LR and other complex ML methods, few have examined the effect of location uncertainty on predictive models. [1] conducted a study to analyze the impact of fire ignition location uncertainty on kernel density estimates by systematically displacing ignition points and comparing the resulting density surfaces. In their study on wildfire prediction in Portugal, [7] utilized LR models. They found that the recorded ignition locations used for model training had a margin of error of up to 500 meters. However, they argued that the impact of this positional error on predictions could be considered negligible due to the large sample size and the small scale of the geospatial data used in their study. To the best of our knowledge, no study has yet investigated the impact of location uncertainty on the predictive performance using BLR models.

3 Methodology

3.1 Data Sources

In this study, the primary data source used was the Austrian forest fire database, which was established within the activities of European and nationally funded projects (AFRI and ALP FIIRS) [13]. This database covers forest fire incidents beginning in the 16th century with an almost complete documentation of forest fires events since the beginning of the 21st century and provides valuable information such as the coordinates of the assumed ignition point location, the location uncertainty radius, the cause of the fire, and the size of the affected area. The scope of this study was limited to human-caused fire incidents that occurred between 2001 and 2018 and have a location uncertainty of no more than 500 meters. A total of 955 fire events were considered in the analysis. To generate predictive features we used additional data sources covering a digital elevation model (data.gv.at; 10x10m), a building and population raster (100x100m), the street network (gip.gv.at) and a vegetation type raster (bfw.gv.at; 10x10m). All data layers were projected to the Austria Lambert reference system.

3.2 Data Preparation

To get an evenly balanced data set, we randomly sampled 1085 points within the forest domain, which we used as non-fire events. The study encompasses several features, namely: distance to buildings, population density, distance to roads, road type, distance to bicycle and pedestrian pathways, vegetation type, elevation, slope and aspect. These specific features were chosen, drawing upon the research conducted by [2] and [3]. The values associated with these features are calculated based on the incident point location. Finally, the recorded fire incidents are divided into four groups based on their associated distance radius, representing the uncertainty of the fire ignition location. The first group, serving as the validation set, includes all samples with an uncertainty smaller or equal to 100 meters. The other three groups, serving as training datasets, are created based on uncertainty thresholds that ensures a roughly equal distribution of samples across the groups. Furthermore, the training data from all groups are combined into a single additional training set. Table 1 provides an overview of the four groups and their corresponding uncertainty ranges and sample sizes.

Table 1 Overview of training and validation data groups.

Bin	Uncertainty Range (meter)	Size	Distribution (non-fire, fire)
Validation	[0, 100]	429	228, 201
Training 1	(100, 250]	580	319, 261
Training 2	(250, 400]	527	275, 252
Training 3	(400, 500]	504	263, 241
Training full	(100, 500]	1611	857, 754

3.3 Model Training

Each training dataset is used to fit both a traditional LR and a BLR model. LR is a statistical method that is well-suited for modeling binary outcomes, such as the presence or absence of forest fires. Unlike traditional LR, BLR assigns a prior probability distribution to the regression coefficients, which reflects prior beliefs about the relationship between the features and the outcome. By using Bayesian inference, the prior is combined with the likelihood of the observed data to obtain the posterior probability distribution of the coefficients. Our choice of prior distribution was a Student-T distribution with a mean of 0, a scale of 2.5, and 1 degree of freedom, resulting in a Cauchy distribution. This prior distribution is known to allow for robust inference and has been recommended for weakly informative priors in Bayesian analysis [9]. Before fitting the data to the model parameters, the numerical input features were standardized to improve model convergence. We utilized scikit-learn (scikit-learn.org) for traditional LR and the probabilistic programming library PyMC (pymc.io) for BLR, which leverages the Markov Chain Monte Carlo (MCMC) algorithm for Bayesian inference.

3.4 Model Evaluation

To assess the predictive performance of the various models on the validation set, we employ two common metrics: accuracy and area under the receiver operating characteristic curve (AUC). Accuracy is defined as the proportion of correctly classified incidents (i.e., whether a fire occurred or not) based on a threshold of 0.5 for the predicted probability values. AUC, on the other hand, measures the ability of the model to distinguish between fire and non-fire cases across all possible threshold values. Both accuracy and AUC have a scale from 0 to 1, where values above 0.5 suggest performance that exceeds random guessing. When evaluating the danger of forest fires, it's important to consider the probability values provided by the model, rather than just the binary classification. These values represent the model's uncertainty in identifying potential fires and indicate the danger of a fire starting under the observed conditions. Therefore, we additionally assess the quality of the probability estimates using the brier score. The brier score measures the average difference between the predicted probability and the actual outcome. A higher score indicates that the model's probability estimates are less reliable, while a lower score indicates greater reliability. The brier score ranges from 0 to 1 and was first introduced in [6].

4 Results

The reported accuracy, AUC and brier scores for the BLR models are mean values of 10 runs. Since the variation among the different outcomes is low, we do not report all model runs in this short paper. Figure 2 and Figure 3 depict accuracy and AUC scores for the LR and BLR models trained on the different datasets. The results clearly show that the

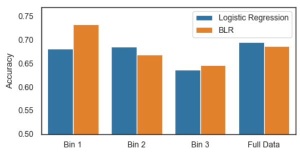

Figure 2 Accuracy (with threshold = 0.5) of LR and BLR models.

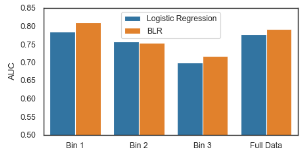

Figure 3 AUC score of LR and BLR models.

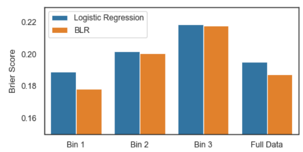

Figure 4 Brier score of LR and BLR models.

model performance decreases with increasing ignition location uncertainty. For the BLR models, there is a +8,6% accuracy, a +9,3% AUC and a -4% brier score (as shown in Figure 4) difference between the model trained on the high quality dataset (100-250 meter location uncertainty) and the model trained on the poor quality dataset (400-500 meter). The BLR trained on the high quality dataset even outperformed the BLR model trained on the full dataset (+4,6% accuracy, +1,8% auc and -1% brier score). When comparing the BLR and LR models, it can be seen that the BLR model trained on the high quality dataset performs significantly better than the LR model trained on the same data (+5,2% accuracy, +2,5% AUC). However, this observation does not apply to the models trained on the other datasets, except for the brier score (Figure 4), where BLR consistently outperforms LR by a small margin.

5 Discussion

The findings of this study highlight the impact of location uncertainty on the predictive performance of fire ignition models. The bias resulting from uncertainty about the true location of the fire ignition has a significant effect on the models' accuracy, with a clear decrease in performance as the location uncertainty increased in the training data. This phenomenon is attributed to location bias affecting all spatial features, especially those with high spatial variability, such as slope. Given the relatively small number of data samples available for forest fire ignitions in Austria, a critical question arises about whether using high-quality data (in terms of location uncertainty) is more advantageous than employing all available data with mixed quality for training purposes. Our study indicates that BLR is a suitable method for dealing with small data sets. It achieves better results when trained on a small high-quality dataset than when trained on a mixed-quality dataset containing roughly three times as many samples. In contrast, the traditional LR model trained on the high-quality data only achieves similar results as the one trained on the full dataset. The reason behind this is, that BLR allows prior knowledge to be incorporated regarding the relationship between the predictors and outcome variable. This incorporation works as a regularizer, constraining overfitting or underfitting in small datasets by reducing the parameter estimates towards the prior distribution. However, an extensive analysis of different prior distributions in our BLR model was not conducted, neglecting the fact that different features may require different sets of priors.Furthermore, there is an additional point that requires discussion. The interpretation of the probability values generated by the forest fire ignition prediction models can be somewhat ambiguous. While the probability score can be an indicator of the level of danger, it can also be viewed as a measure of uncertainty in the model's prediction. However, [11] argue that these two concepts, the predicted level of danger and the prediction uncertainty, should be treated separately. This suggests the need to investigate how we can use Bayesian inference, which provides additional information about the prediction uncertainty, to communicate both the predicted probability and the model's uncertainty to decision-makers in forest fire management.

6 Conclusion

In summary, this study highlights the importance of considering location uncertainty in fire ignition models, and the potential benefits of using BLR for dealing with small datasets. The findings of this study can have significant implications for forest fire management and documentation, as they suggest that investing in a high-quality dataset and utilizing BLR with weakly informed priors may help overcome the limitations posed by a small training dataset.

References

1 Giuseppe Amatulli, Fernando Peréz-Cabello, and Juan de la Riva. Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty. *Ecological modelling*, 200(3-4):321–333, 2007.

2 Natalie Arndt, Harald Vacik, Valerie Koch, Alexander Arpaci, and Hartnut Gossow. Modeling human-caused forest fire ignition for assessing forest fire danger in austria. *iForest-Biogeosciences and Forestry*, 6(6):315, 2013.

3 Alexander Arpaci, Bodo Malowerschnig, Oliver Sass, and Harald Vacik. Using multi variate data mining techniques for estimating fire susceptibility of tyrolean forests. *Applied Geography*, 53:258–270, 2014.

4 Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, and others. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 401–413, 2021.

5 Ross A Bradstock, JS Cohn, A Malcolm Gill, Michael Bedward, and C Lucas. Prediction of the probability of large fires in the sydney region of south-eastern australia using fire weather. *International Journal of Wildland Fire*, 18(8):932–943, 2009.

6 Glenn W Brier and others. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

7 Filipe X Catry, Francisco C Rego, Fernando L Bação, and Francisco Moreira. Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire*, 18(8):921–931, 2009.

8 Georgios Charizanos and Haydar Demirhan. Bayesian prediction of wildfire event probability using normalized difference vegetation index data from an Australian forest. *Ecological Informatics*, 73:101899, 2023.

9 Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013. doi:10.1201/b16018.

10 Piyush Jain, Sean C P Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505, 2020.

11 Meelis Kull and Peter A Flach. Reliability maps: a tool to enhance probability estimates and improve classification accuracy. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part II 14*, pages 18–33. Springer, 2014.

12 J. San-Miguel-Ayanz, T. Durrant, R. Boca, P. Maianti, G. Libertá, T. Artés-Vivancos, D. Oom, A. Branco, D. de Rigo, D. Ferrari, H. Pfeiffer, R. Grecchi, M. Onida, and P. Löffler. Forest Fires in Europe, Middle East and North Africa 2021. Technical report, Publications Office of the European Union, Luxembourg, 2022. doi:10.2760/34094.

13 Harald Vacik, Natalie Arndt, Alexander Arpaci, Valerie Koch, Mortimer Mueller, and Hartmut Gossow. Characterisation of forest fires in Austria. *Austrian Journal of Forest Science*, 128(1):1–31, 2011.