Investigating the affect of forest fragmentation on wildfires

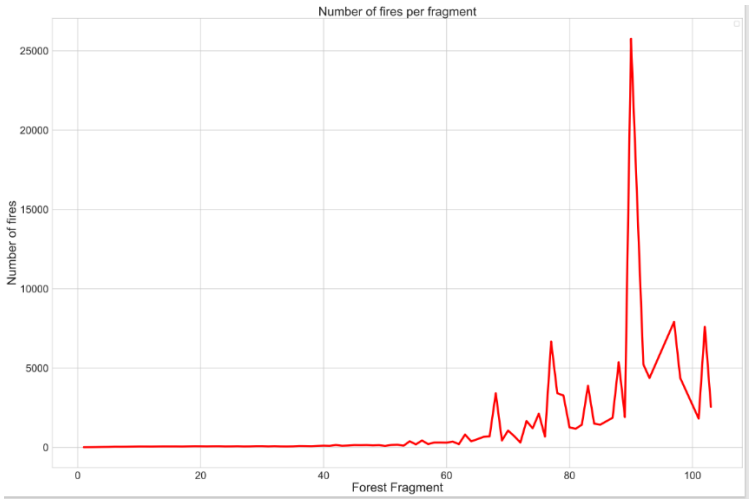
Forest fragments under research in the Biological Dynamics of Forest Fragments Project. Photo by: Richard Bierregaard.

After being accepted to the research program at Imperial College London I was very glad since one the main reasons I wanted to take part was due to my interest in machine learning. I was very interested to see how machine learning was in the real world applied to a scientific setting. I never was able to understand how scientists were able to use machine learning techniques in inter disciplinary settings to enhance their research. I was stuck on only being able to apply machine learning techniques to toy datasets that were already nice and set up. However, this is the first time I see what real ML like which is a lot of it is gathering data and cleaning it, not just training the model.

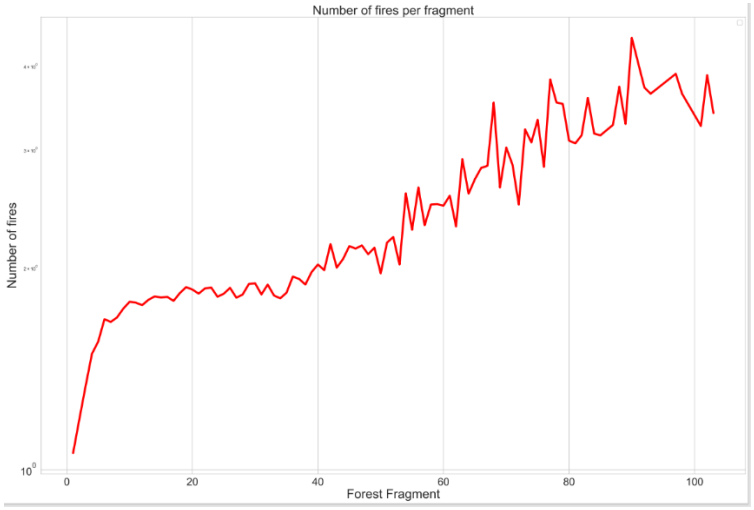
Initially I did not know much about wildfires and the features that could influence a wildfire. I assumed wildfires had a few natural causes and that the reason for an increase in wildfires was due direct human activity(such as campfires or electric powerlines). After doing my initial reading I was amazed on how different biomes adapt to fires, for instance seeds that would only sprout after a certain temperature was reached. It never did occur to me that wildfires in some places were beneficial.

My first task was to analyse the global fire atlas dataset and see if there is any relationship between fire burnt area and the forest fragmentation dataset that was provided. All this was done in the South American region, there was a focus on Mato Grosso(Brazil). I had to match the resolution of the datasets for consistency. However, one issue with the forest fragmentation dataset was that the values that were given were “classes” and not measurable data such as km^2 for instance. This makes it difficult to see relationships between number of fires and the forest fragment class. To get past this we normalized the data and wanted to see if there is a relationship between forest fragment class and number of fires in that class. There was a positive correlation between number of fires and forest fragment class, even when we considered the size of each forest fragment class. This was possible since QGIS software allows us to calculate the area of polygons (km^2). This now made the forest fragmentation dataset measurable which we could also use as input to machine learning

models since you cannot use “classes” as well as measurable data(e.g. area, speed, pressure etc) in the algorithm. We replaced the class with the area instead.



How many fires per forest fragment, considering the area of each forest fragment. This is because a larger area fragment will of course will always have more fires, but we want to see if forest fragment affects number of fires. It would not be fair to not normalize the data



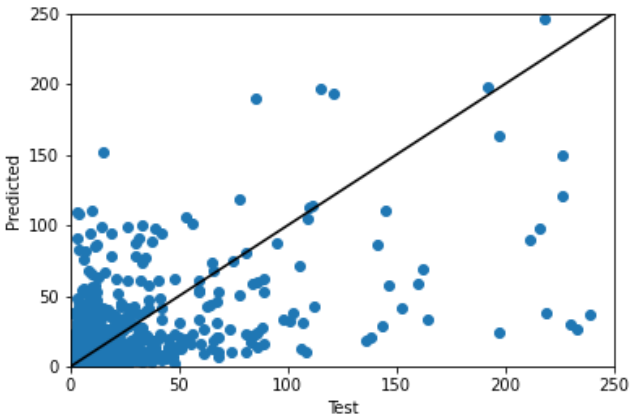
Log version of graph 1

We chose random forest and a neural network as the machine learning algorithms. As stated before, we had to instead of including the forest fragment class into the algorithms, include the area of the forest fragment that each fire was associated with. Other input features we chose were biomass, slope, tree cover, grass cover, wind u, wind v, mean air temperature, surface pressure and duration from global fire atlas dataset.

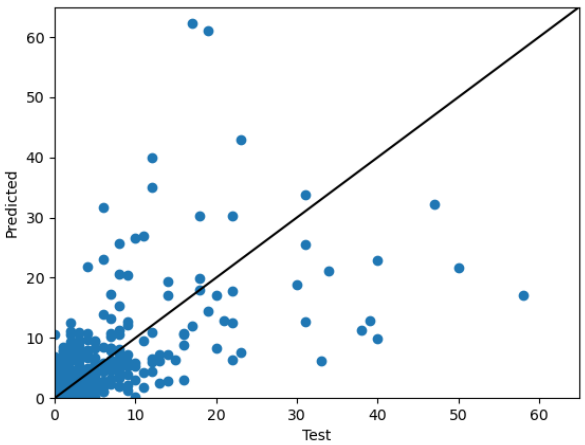
At first, we had to reduce the number of fires of duration =1 since there was a disproportionate number of durations = 1 that could influence the predictions. Initially there were 37k data points however duration = 1 covered 16k of these datapoints. So, we randomly chose 2k data points of duration = 1 and placed it in. Everything else was kept the same, where the max was 2.2k data points for a given duration and the minimum was 1(for the largest duration fire). Note that we did try other ways to sort the data such as 1k per duration and minimum 250, and other configurations but found

that there wasn't really an improvement, so we stuck to the original plan. Also, another important note that only 2003-2010 years were chosen to form this dataset due to time constraints. Furthermore, when calculating the data from google earth engine, we calculated the mean for example biomass with a 10km radius from a given ignition point. We also did the same test for Central Kalimantan but from years 2003-2016 as there are less fires so there was enough time to process them.

After gathering the values of all the data, we stored it into an attribute table to input it to the random forest algorithm. We tried various ways of predicting wildfires such as, could we predict wildfire size from just the features from google earth engine without duration. Ultimately after all this testing we found out that without duration, you cannot predict the size of a fire accurately, this was shown to us by the very poor score. We decided to take duration and features to predict the fire size. We wanted to see the affect of forest fragmentation and so tried the machine learning models with and without area of forest fragment. Using a neural network didn't provide good results, most likely due to lack of inputs since Neural networks are usually appropriate with lots of inputs. However, using random forest generated positive results showing an increase of fire size prediction score of about 3% when including area of forest fragment of the fire ignition point. This score would increase when the ratio of training and test score where different ratios.



Random forest results with area of forest fragment to predict fire size. Mato Grosso 2003-2010



Random forest results with area of forest fragment to predict fire size. Central Kalimantan 2003-2016

This research was done as an undergraduate research placement in Imperial College London and the Leverhulme Centre for Wildfires, Environment and society who provided me with funding. It has been a great experience to learn a lot about data analysis and machine learning in the context of climate science. My supervisors Dr Ramesh Ningthoujam and Dr Sibo Cheng have been very patient with me and extremely helpful in doing this project, thus making it enjoyable and useful for my goals in the future. I thank them immensely for granting me this great opportunity which has provided me the skills that are very transferrable to other interest of mine in the future.