Linking Craterellus Occurrences & Climate Change, an Attempt

How we sort, visualize and Analyze this data will unlock insights

Louis Leboa and Phong To
Department of Biology, University of Washington

Abstract:
Macrofungi, commonly referred to as mushrooms, play a key role in ecosystems by recycling forest floor detritus. Limited research has been done to study the effects of climate change on autumn-fruiting mushroom populations and distributions. Prior studies have established that increasing temperatures have caused earlier fruiting in spring-fruiting mushrooms. In our paper, we utilize historical data from Global Biodiversity and Information Facility (GBIF) on Craterellus prevalence in Sweden from 1950 until 2018 to determine whether or not we can observe the same effects on this economically impactful, autumn-fruiting mushroom given daily temperature data from the National Oceanic and Atmospheric Administration (NOAA. However, there was heavy sampling bias towards recent dates given recent rising interest in mushrooms. Attempting to correcting for this, we were unable to establish a significant correlation between year and amount of mushrooms collected per individual due to there not being an apparent change in temperature.

Introduction/Background:
Sustainable, equitable solutions to hunger, ecosystem management, and health are three issues that are pressing humanity at this moment that will persist into the foreseeable future. Macrofungi, commonly known as mushrooms, are organisms that are primed to assist humanity in addressing all three of those issues [1]. While acting as a food source that is capable of processing forest waste, certain studies have found that certain species of macrofungi can also act as key indicators of ecosystem health [2]. However, these problems that mushrooms pose to help solve are further exacerbated by climate change. In a study conducted by Kauserad et al., they concluded that global warming was contributing to the earlier fruiting that they observed in spring-fruiting macrofungi populations in Norway and the UK [3].

Chanterelles is a name given to a variety of delicious and edible mushroom species that has spawned a multi-million dollar industry surrounding the harvesting of them. Chanterelles generally start to fruit in the autumn. Given the economic weight that Chanterelles hold, the consensus that weather has a strong impact on their growth [4], and the fact that one of us will be living and harvesting mushrooms in Sweden within the year, we began our research by asking how has climate change affected the Craterellus family, a family know for chanterelles, occurence in Sweden? Sweden is one of the few mushroom capital of the world, and thanks to GBIF and NOAA, there are two rich databases filled with mushroom and climate data, respectively. While climate change is not temperature change alone, mushrooms are sensitive to temperatures and mycology studies often take temperature into account (Pilz et al. 1996). Few studies exists that directly examine the effects of changing surface temperatures on the growth patterns of of Craterellus, such as their location, seasonality, and prevalence over decades. In this study, we intend to examine if there has been an effect over the last 50 years.

Methods:

Data

The Global Biodiversity Information Facility provided our data pertaining to the Cantralles family of fungi in Sweden. The data was pulled from the database through there graphical search feature and was superficially limited to data from Sweden and of the Craterellus family, between the years 1950 and 2018 (for full specifics see the references API link in references). Initially we had thought to map all of the different fungi families but that was too much data to get through in the constricted timeline. The we limited our data to after 1950 because the data was not well reported before. The data from GBIF downloaded in a 20MB tsv file. The Data was 21,126 rows (recorded occurrences) by 45 columns.

The weather data this project uses comes from the National Ocean and Atomospheric Association (NOAA). Based in the United States, NOAA is known for their accurate surface temperature data collection from around the world. The dataset was pulled from Jonköping weather station located in Sweden between the years of 1977 and 2018. The data was downloaded as a .csv file and was 3462 rows long by 23 columns.

Visualizations

Once the data was parsed into a pandas dataframe in jupyter notebook we were able to begin visualizations. We first began visualizing using Plotly, a pay-for-use software (free for the first 25 graphs), that integrates with jupyter notebooks to make interactive graphs with widgets. This allowed us to quickly get a feel for our data. We created graphs with sliders to interact with out graphs. However, we soon realized that it would make more practical sense to use matplotlib, scikit learn, numpy, pandas and basemap to produce our visualization.

The fungi data had two purposes one was to visualize occurences in the database and the other use was to do analysis on Cantrellus in relation to temperature data over the 50 year period. To visualize the data, a function was set up to allow us to pull data by the year and then any other column we wished. We created visualization based on this function pulling data by the year for occurences per the month to find the month with highest occurrences and tracked this throughout the years to find the month with the highest number of occurence over the past 50 years (fig 1). We used the function again to pull all occurence for a given year and plotted the lat and long in a mapping library called Basemap. This allowed us to visualize the exact coordinates provided in the data set on a map of Sweden. The function that we built used a for-loop to interpolate between the years 1950 and 2018 to plot the data and for each year downloaded a copy of the image which were then merged into an animation via gifmaker.
For this figure we also were able to match (fig 2).

Analyses

For all of our statistical analyses, we used Python 3.7, utilizing the packages ‘matplotlib,’ ‘sklearn,’ ‘pandas,’ and ‘numpy.’ Much of our data is continuous in both axes, so for analyzing relations between the two we initially did a visual inspection of the trend, and then ran an appropriate regression depending on that inspection. Most of our data was analyzed using linear Least Mean Squares (LMS) regression for explanation of how related the compared variables are , though we also ran a non-linear exponential LMS regression for the data showing mushroom interest. We did these regressions because wanted to know if the two continuous variables were related, and the r-squared value should tell us if the regression line is able to predict the behaviours presented.

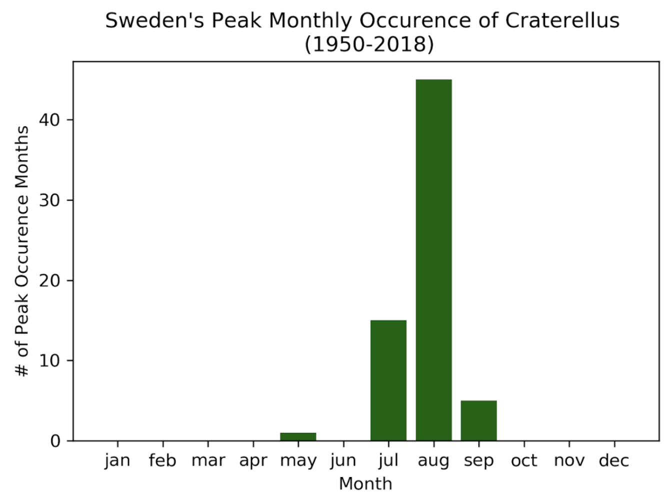
Results:

Initial Exploration: General Trends in Mushroom Growth

When initially examining our data, we wanted to see what information our dataset could tell us, and if this information was accurate. First, we wanted to know about our mushroom's seasonality; we knew from prior studies that this mushroom tended to fruit in the Autumn, so if we plotted a histogram of when these mushrooms were found, by visual inspection of we should expect to see the most mushrooms being found in the autumn months. We found that Chaterellus, based on our data, fruited the most in August (Figure 1).

However, there is exists some variance.

Figure 1. *Shows the number of times a given month had peak mushroom occurrence. August has the highest occurrences of Craterellus between the years 1950 and 2018, which is consistent with our known information on the growth behaviour of this mushroom.*



Our next question then concerned whether or not this trend was changing over time, and whether the data was consistent throughout time. We decided to visualize this data using the given coordinate data for location of mushroom catalogued, and repeated the histogram presentation for number of occurrences throughout the years for a visual inspection of growth trends (Figure 2). By visual inspection, found that most mushrooms were being recorded in the southern region of Sweden. However we also found that the number of occurrences of mushrooms recorded over the years varied dramatically; this would prevent us from making any substantial claim about trends over time if the difference was significant, which would lead us to our next experiment.

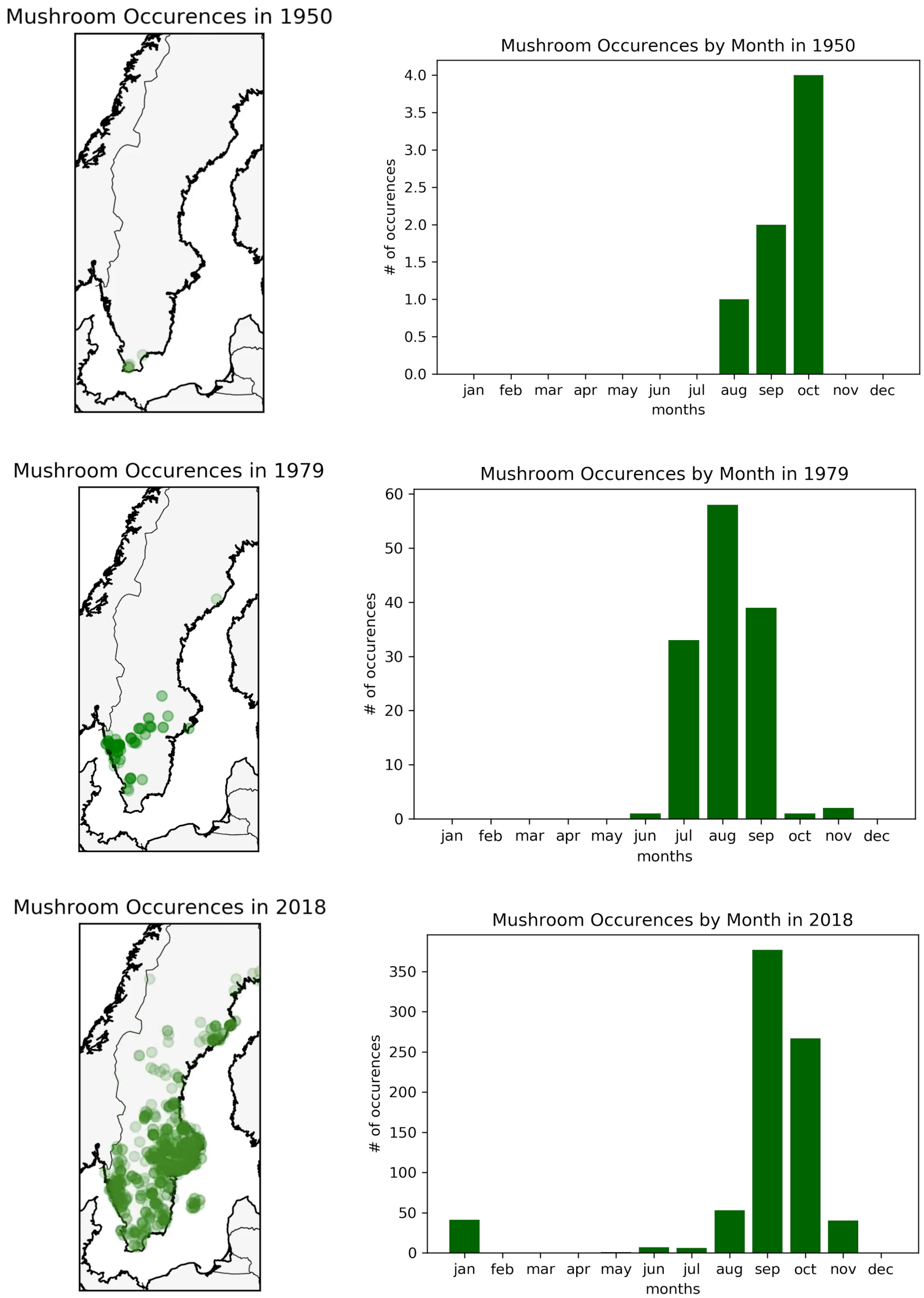
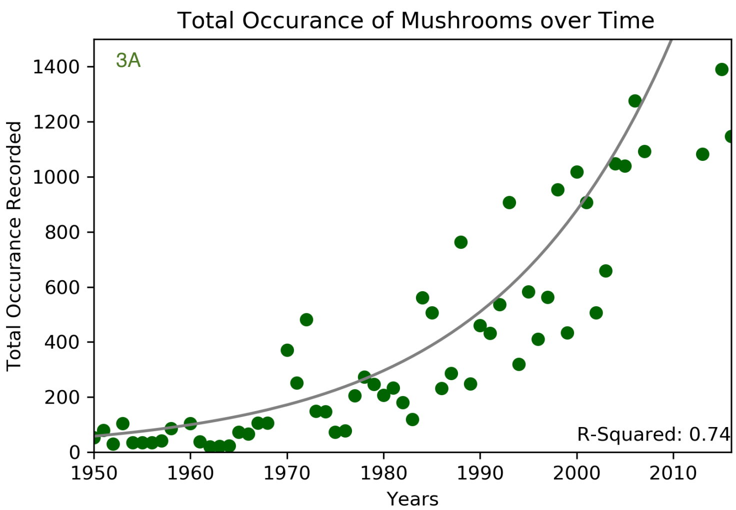


***Figure 2**. Two different representations of our data at three time points from our series of 68 years: the beginning (1950), the middle (1979) and the last in the series (2018). The left illustrate where the Cratterelus samples were collected via their gps coordinates. The green dots represent individual occurrences. The figures on the right are histograms showing how many mushrooms were found within each month for the given year. For an animated version that interpolates through all 68 years please see the movie in our folder.*



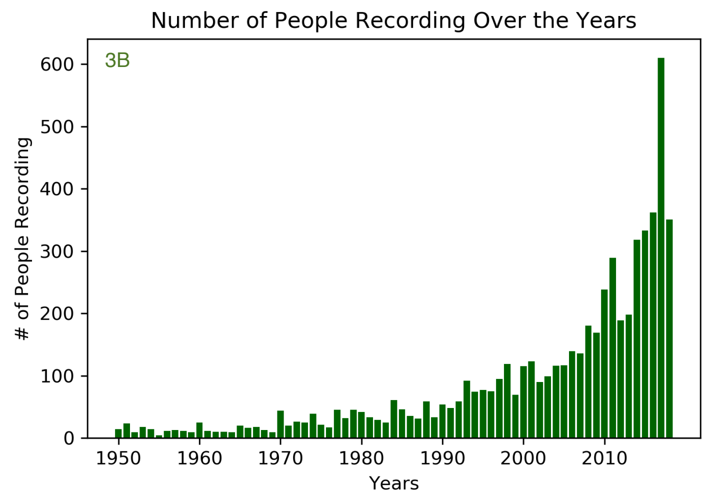
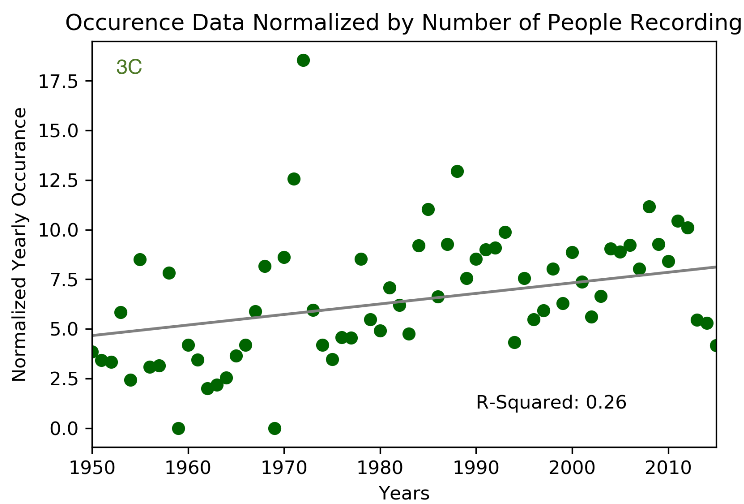


Figure 3. *These figures graph trends we found about our data. The graph on the left shows that there is a significant increase in the amount of data that is being recorded over the years (3A). The figure on the top right shows that, similarly, there is also an increase in the number of people collecting these mushrooms (3B). The figure on the bottom right is normalization of occurrence data by number of people recording (3C). The low R-squared value suggests there is not a strong correlation between year and number of mushrooms collected per person.*



Growing Interest in Mycology and Normalization of Data

Our next experiment then was to then confirm or deny the trend that more mushroom data was being collected over time, which we could analyze since our dataset includes the year found in entry of mushroom data (Figure 3). We ran a non-linear exponential LMS regression, and there was clear evidence of an increase in amount of data over time, with our R-Squared value for this comparison being .74. However, this would render our dataset unusable for further analysis since that introduces a heavy sampling bias towards periods of time with higher mushroom findings. So, we then tried to normalize our data by dividing the number of mushrooms found per year by the number of people who went out to collect them, hoping that it would normalize the data accurately under the assumption that each individual went out and tried to find as many mushrooms as they could; if abundances are different throughout the year, we would expect a clear trend showing differences in collection data over time. The results showed that while the regression line was slightly positive, the R-squared value was below .50 at 0.26, indicating that there was no a significant link between year and normalized monthly occurrence of mushrooms, meaning people weren't necessarily finding more mushrooms per trip (figure 3C). While this method doesn't account for every variable that could affect this factor, this was good news for us, because it means we can estimate the number of mushrooms and have more confidence about it than before . We could apply this normalization to a month, and observe for changes in the findings of that month. All that was left was to match it to the temperature data (figure 4).

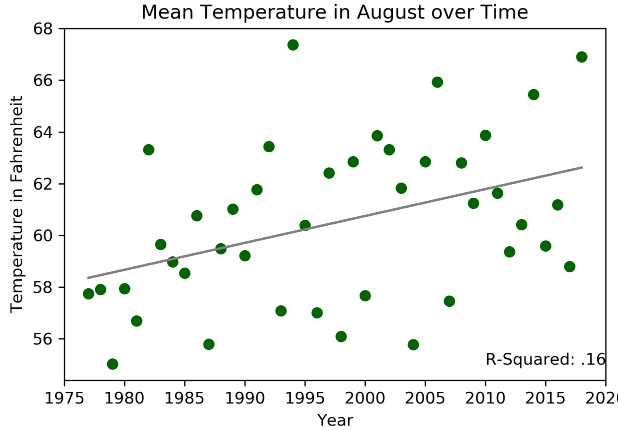


Figure 4. *Mean temperature of the month of August between 1977 and 2018 for Jonköpping. There is no clear trend in the data*

Temperature Change in Jonköpping, Sweden

From our prior experiments, we established two key pieces of information: August was when Craterellus was generally fruiting the most, and that the location that the mushrooms was growing was generally in southern Sweden. Thus, if we wanted to check for a relationship between changes in growth behaviour and changes in temperature, we could run a regression between those two variables. However, after analyzing our data, we found that there was not a significant difference in temperature change within the month of August over time, with the R-squared value for the linear regression being only .16. Because there was no significant change in temperature over time, there's no reason to graph the mushroom data against temperature.

Conclusions and Future Work:

We learned that from our chosen datasets, we cannot confidently answer our original question of, "how has temperature affected the prevalence of the craterellus family fruiting bodies?" The temperature data we have simply does not strongly suggest a change in temperature over time, and our normalization method is not the cleanest. However, we were able to draw some notable conclusions along the way: First, August seems to be the prime month for mushrooming, as our data points to August being the most fruitful month for growth of this family. Secondly, we learned that as time has passed, interest in mushrooms within the scientific community has increased, as our mushroom data is compiled from scientific research, and we note a clear exponential trend. Thirdly, we found that our normalization method seems to work, though there are still many variables that we can consider that may affect it, such as available budget for research on mushrooms and availability of students under to go out into the field and collect samples.

Future work that may be done on this subject will require more extensive and uniform dataset over time on mushroom occurrence. Also, average temperature data should be compiled and averaged across a wider range of stations in the correct regions for a more accurate and statistically correct model of temperature change. Considerations should also be given to trophic effects that may affect mushroom growth, such as climate changes effects on the trees that the mushroom grows on. If these trees are decaying at a higher rate because of climate change, the increase in available detritus for the mushrooms to clear may offset any effects from the change in temperature; Perhaps an in lab experiment could be warranted for collection of variably isolated data.

Contributions/Acknowledgements :

It is with great honor that Louis Leboa could co-author this paper with the esteemed Phong To. Mr. To was greatly responsible for running all of our Regression models and for understanding how our data could be worked with. He further helped with downloading and finding climate data. Mr. Leboa on the other hand handled harvesting mushroom data and visualizations of data. We would like to thank the data scientists that helped come up with ways to normalize our data, Kameron Harris and Eleanor Lutz.

Outside Software:

Plotly - initially used to create interactive plots that were later removed from the project as we found better ways to visualize the data. Download: Plot.ly
Basecamp library- used to visualize the geospatial coordinates over time. Download: available in Anaconda
For gif creation- We made 69 images of the map throughout years and uploaded. gifmaker.me.

References:

Data:
Data download files are attached in the Final Project Code and Data zip file we uploaded.
NOAA's data query link **-**
https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv&georegionabbv
GBIF's data query link -
http://api.gbif.org/v1/occurrence/download/request/0000381-190306134739812.zip
Background Refrences:
[1] Importance of mushrooms:
Chang, S. T. (1999). Global impact of edible and medicinal mushrooms on human welfare in the 21st century: nongreen revolution. *International Journal of Medicinal Mushrooms*, *1*(1).
[2] Temperature importance:
Pilz, D.; Molina, R., eds. 1996. Managing forest ecosystems to conserve fungus diversity and sustain wild mushroom harvests. Gen. Tech. Rep. PNW-GTR-371. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 104 p
[3]Mushroom occurrence study in spring flowering mushrooms -
Kauserud, H., Heegaard, E., Semenov, M. A., Boddy, L., Halvorsen, R., Stige, L. C., ... & Stenseth, N. C. (2009). Climate change and spring-fruiting fungi. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1685), 1169-1177.
http://www.els.net/WileyCDA/ElsArticle/refId-a0000369.html
[4] Pilz, D., Norvell, L., Danell, E., & Molina, R. (2003). *Ecology and management of commercially harvested chanterelle mushrooms*. Gen. Tech. Rep PNW-GTR-576. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
Coding References:
Mapping Visualization -
Least Means Squares, regression general -
https://towardsdatascience.com/polynomial-regression-bbe8b9d97491
Formatting graphs -
https://stackoverflow.com/questions/43757820/how-to-add-a-variable-to-python-plt-title
Exponential curve fitting/r2 -
http://pyhyd.blogspot.com/2017/06/exponential-curve-fit-in-numpy.htmlhttp://pyhyd.blogspot.com/2017/06/exponential-curve-fit-in-numpy.html