## Introduction

Fungi are made of one-cell thick hyphae, which are prone to evaporative water loss[6,12,13]. The majority of a fungus' lifespan is spent as mycelium, but when conditions are met, the mycelium forms a visible fruiting body, often with high water content[6,12]. Significant moisture is required to form the fruiting body[13]. Sudden low temperatures trigger mushroom growth[9.] High temperatures with low humidity may contribute to evaporation of moisture. Fires may burn fungi and substrate, and alter soil composition and carbon bioavailability; post-fire habitats are best exploited by pioneer species[16]. It is clear that precipitation, temperature and fire will impact fungal diversity. We hypothesize that high precipitation increases diversity, high temperatures, and fires decrease diversity. In this study, we use participatory science data from Mushroom Observer to quantify the validity of community datasets.

## Methods

Mushroom Observer is a participatory science website that collects public-reported mushroom sightings across the world.[7] The data contains species, locations, and dates for each observation.*

### *Fire: Data Description*

Wildfire data was retrieved from the *CAL FIRE incident database*; a repository of California fires >10 acres reported by local fire departments.[5] County, boundary, and location data was retrieved from *Opendatasoft*; a large public repository with a wide range of datasets.[17,19]

### *Fire: Data Analysis*

The California county boundary data was merged with the fire dataset to provide location data for each county. The area burned and the number of mushroom observations per year were totalled for each county.* Due to processing limitations, 15 counties were randomly chosen. Using a linear mixed model,[3] the total number of mushroom observations was regressed on the total area burned by fire. County was used as a random effect; the number of mushroom observations was more likely to be similar within counties due to the environmental conditions or the number of people making observations. To meet the assumptions of the linear mixed model, a fourth root transformation was done on the response variable (Figure 1).

* See appendix for more details

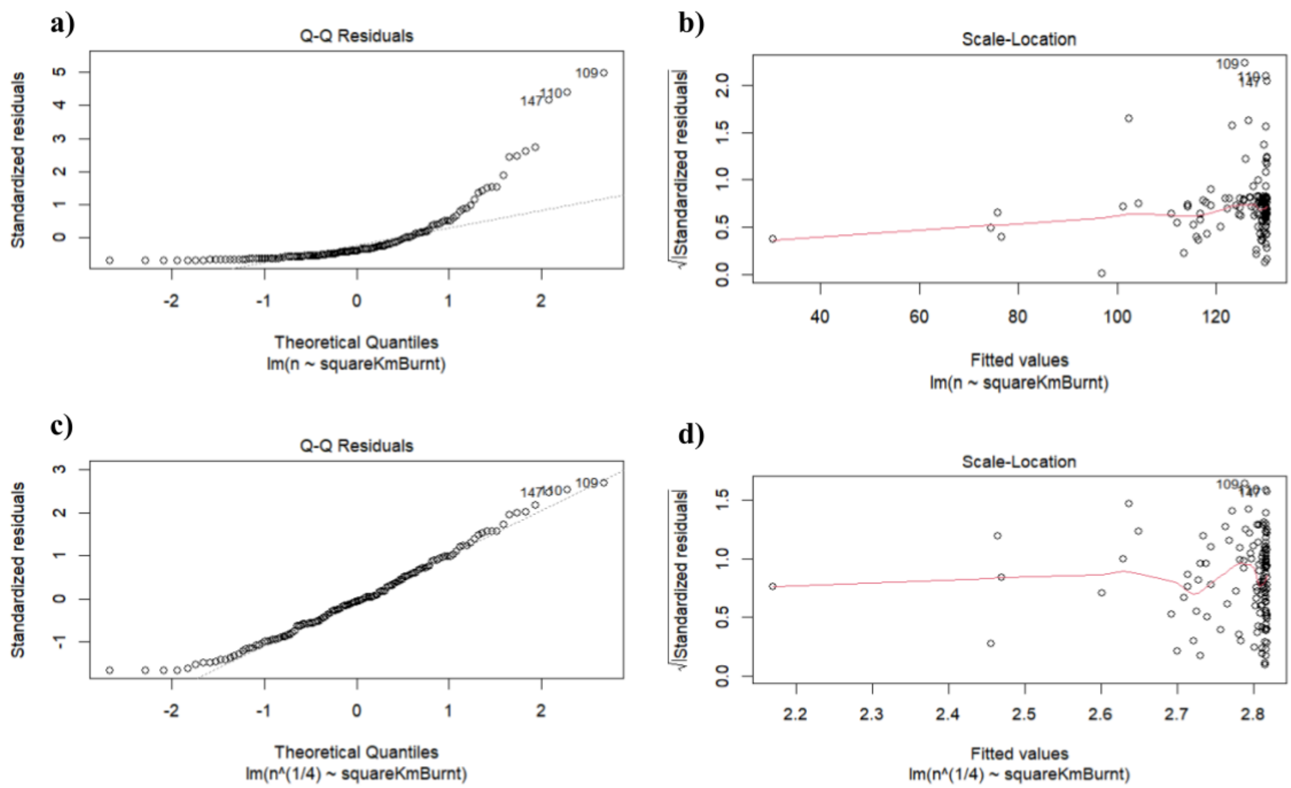


**Figure 1 .** Graphs showing the assumptions of the linear mixed model. **a)** Q-Q plot showing the violation of normality **b)** Scale-location plot showing the violation of homoscedasticity. Notice how the residuals are distributed in a funnel shape **c)** Q-Q plot showing normality being better met after the transformation **d)** Scale-location plot showing homoscedasticity being better met after the transformation

*Weather: Data Description*

All data comes from publicly available weather databases: the San Francisco weather data (2007-2020) comes from the Remote Automated Weather Stations network,[8] Ontario data from the ClimateData.ca project (2010-2022),[15] and Colombia data from the Climatic Research Unit gridded Time Series data set (2013-2022)[2,4]. All these databases use weather station data.

*Weather: Data Analysis*

Weather and mushroom data were analyzed on a monthly scale to smooth out temporal variation.* Monthly diversity indices were calculated using the Shannon Index,[9] as its calculations place more weight on rarer individuals. This offsets some bias in crowd-sourced observations, where popular species are reported more frequently than inconspicuous species. To test the effects of precipitation and temperature, generalized linear models (GLMs) were fitted over the abundance and diversity values using Poisson and gamma distributions, respectively.*

* See appendix for more details

## Results

### *Fire*

The LMM returns p = 0.855, failing to reject the null hypothesis that there is no relationship between the area burned and the number of mushroom observations (Figure 2).

Note that the conditional $R^2$ value was 0.67 while the marginal $R^2$ was only 0.0001, meaning that most of the variance in the data can be explained by county (Figure 2).

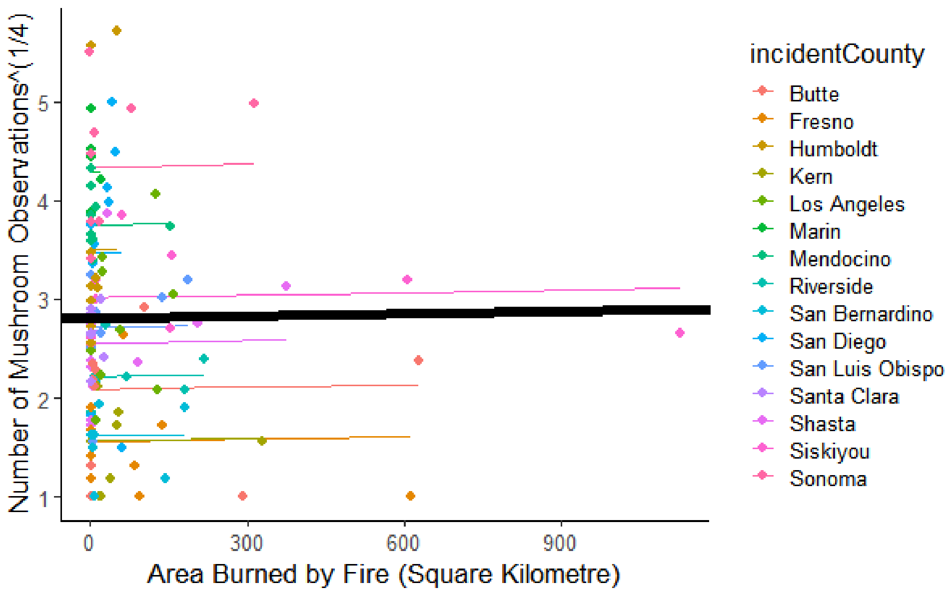


**Figure 2.** Regression of the number of mushroom observations on the area burned by fire, with county as a random effect. Notice the lack of trend and the widespread of colored lines around the black line that is only associated with the fixed effect of area burned by fire.*

### *Weather*

In the San Francisco Bay Area, observations are positively correlated with precipitation and negatively correlated with temperature (Figures 3a, 4a). In Ontario, however, temperature appears to be positively correlated with abundance, with little effect on precipitation. In Colombia, a clear pattern is not apparent from the time-series (Figures 3b,c, 4b,c).

* See appendix for more details

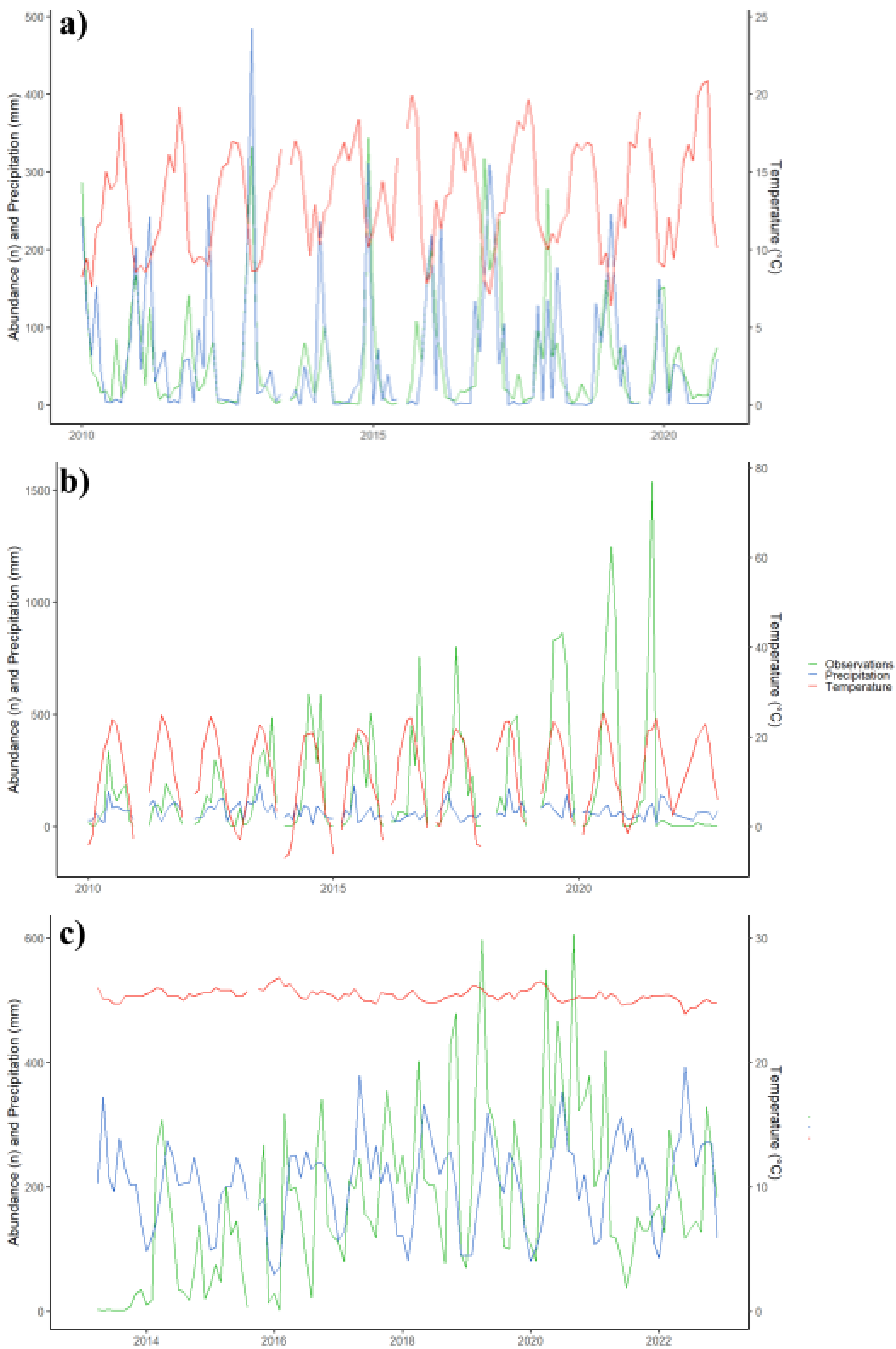


**Figure 3.** Mushroom Observer observation abundance and climate time series in **a)** the San Francisco bay area, **b)** Ontario, and **c)** Colombia. Note the different time scales

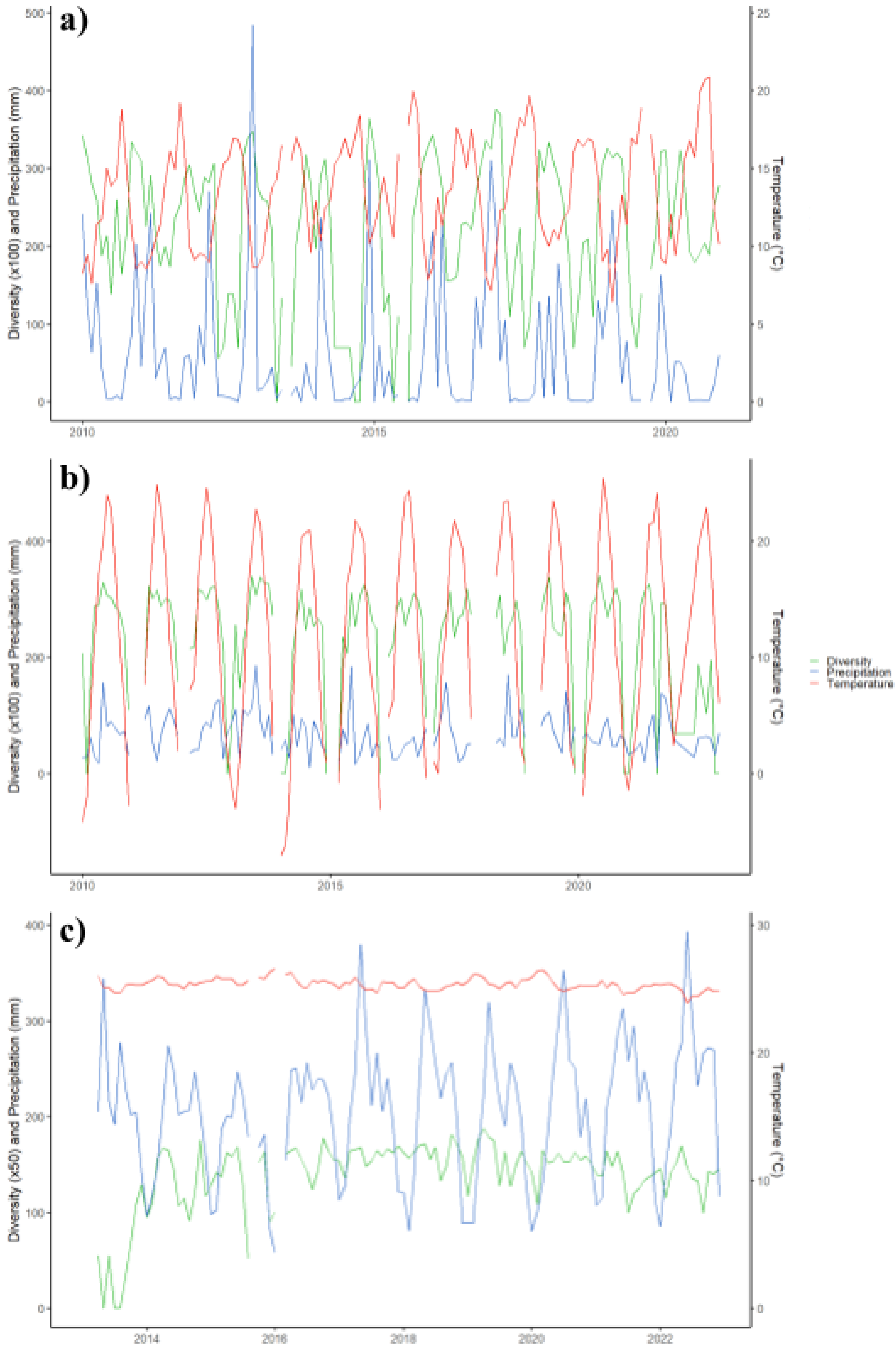


**Figure 4.** MushroomObserver observation diversity and climate time series in **a)** the San Francisco bay area, **b)** Ontario, and **c)** Colombia. Note the different time scales

GLM testing corroborates these observations, where temperature, precipitation, and their interaction show significant effects on abundance in particular (Table 1). San Francisco diversity showed similar results to abundance, but notably, only temperature was observed to have a significant effect in Ontario, and Colombia showed a weaker significance of temperature and precipitation. Note that the negative inverse link function of gamma causes opposite effect signs.

**Table 1.** GLM effect terms and associated p values. Model with lowest AIC values chosen. Blank spaces indicate that the best model did not include those terms.

| Location | Variable | Abundance | | Diversity | |
|---|---|---|---|---|---|
| | | β | p | β | p |
| San Francisco | Precipitation | -0.004 | 2.57e-16 | 0.0016 | 0.0092 |
| | Temperature | -0.15 | < 2e-16 | 0.021 | 4.8e-05 |
| | Precip. * Temp. | 0.00096 | < 2e-16 | -0.00022 | 0.00033 |
| Ontario | Precipitation | 0.0017 | < 2e-16 | | |
| | Temperature | 0.085 | < 2e-16 | -0.0072 | 2.0e-08 |
| | Precip. * Temp. | | | | |
| Colombia | Precipitation | -0.12 | < 2e-16 | -0.00034 | 0.043 |
| | Temperature | -0.49 | < 2e-16 | -0.054 | 0.031 |
| | Precip. * Temp. | 0.0049 | < 2e-16 | | |

**Discussion**

These results are inherently biased due to the nature of citizen science. A key assumption when forming any linear model is that observations are independent from one another, which does not occur within Mushroom Observer data. The majority of the data used was reported by a handful of people; the data also tracks these individuals' habits and schedules. A major goal was to see if meaningful results could be extracted from the large database despite these flaws.

*Fire*

Contrary to our predictions, no relationship was found between fire and mushroom observations. This may be due to large variations in the number of mushroom observations between counties, possibly masking the effect of fire. Results may be affected by the number of active Mushroom Observer users in each county. Some counties may not have enough observations for sufficient statistical power. Another explanation may be that some fungi may be resistant to fire. Fungi produce heat-shock proteins to survive heat stress[14] and therefore it is possible that mushroom abundance may not decrease as drastically as we predicted

*Weather*

We observe significant effects on abundance and diversity. Contrary to the predictions, however, analysis at multiple locations reveals that neither factor has a definite effect size or direction: the effect of temperature is opposite in San Francisco and Ontario. This is also seen in the interaction terms, which suggest that both variables adjust each others' effects.* While findings support that weather affects mushroom growth, outside factors like latitude, species, interactions, etc. also influence observed trends. Overall, it was found that participatory science was capable of observing significant effects.

* See appendix for more details

## References


1. Cheng J, Schloerke B, Karambelkar B, & Xie Y. 2023. _leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library_. R package version 2.2.0 https://CRAN.R-project.org/package=leaflet>
2. CRU CY [database]. 1901-. v.4.07. Climate Research Unit (University of East Anglia) and NCAS. [updated 2023 Apr 19; accessed 2023 Dec 3]. https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/crucy.2304181636.v4.07/countries/
3. Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
4. Harris I, Osborn TJ, Jones P, & Lister D. 2020. Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset. Sci Data. 7, 109. https://doi.org/10.1038/s41597-020-0453-3
5. Incident Data. 2009-. State of California: CAL FIRE. [updated 2023 Oct 30; accessed 2023 Oct 30]. https://incidents.fire.ca.gov/imapdata/mapdataall.csv
6. Lendzian KJ, Beck A. 2021. Barrier properties of fungal fruit body skins, pileipelles, contribute to protection against water loss. Scientific Reports. 11(1). doi:10.1038/s41598-021-88148-0. https://doi.org/10.1038/s41598-021-88148-0.
7. Mushroom Observer CSV files [database]. 2006-. Mushroom Observer, Inc. [updated 2023 Oct 3; accessed 2023 Oct 3]. https://github.com/MushroomObserver/mushroom-observer/blob/main/README_API.md#csv-files
8. Oakland North California RAWS USA Climate Archive [database]. 1992-. Western Regional Climate Center. [accessed 2023 Oct 30]. https://wrcc.dri.edu/cgi-bin/rawMAIN.pl?caCOKN
9. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, & Weedon J. 2022. _vegan: Community Ecology Package_. R package version 2.6-4, <https://CRAN.R-project.org/package=vegan>
10. Pinna S, Gévry M -f., Côté M, Sirois L. 2010. Factors influencing fructification phenology of edible mushrooms in a boreal mixed forest of Eastern Canada. Forest Ecology and Management. 260(3):294–301. doi:10.1016/j.foreco.2010.04.024. https://doi.org/10.1016/j.foreco.2010.04.024.
11. R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
12. Stojek K, Gillerot L, Jaroszewicz B. 2022. Predictors of mushroom production in the European temperate mixed deciduous forest. Forest Ecology and Management. 522:120451. doi:10.1016/j.foreco.2022.120451. https://doi.org/10.1016/j.foreco.2022.120451.

13. Straatsma G, Ayer F, Egli S. 2001. Species richness, abundance, and phenology of fungal fruit bodies over 21 years in a Swiss forest plot. Mycological Research. 105(5):515–523. https://doi.org/10.1017/s0953756201004154.
14. Tiwari S, Thakur R, Shankar J. 2015. Role of heat-shock proteins in cellular functions and in the biology of fungi. Biotechnology Research International. 2015:1-11.
15. TORONTO CITY [database]. 2002-. Computer Research Institute of Montréal: Climatedata.ca. [accessed 2023 Dec 3]. https://climatedata.ca/download/#station-download
16. Treseder KK, Mack MC, Cross A. 2004. RELATIONSHIPS AMONG FIRES, FUNGI, AND SOIL DYNAMICS IN ALASKAN BOREAL FORESTS. Ecological Applications. 14(6):1826–1838. doi:https://doi.org/10.1890/03-5133)
17. US County Boundaries. 2017. U.S. Census Bureau. [accessed 2023 Oct 30]. https://public.opendatasoft.com/explore/dataset/us-county-boundaries/export/?disjunctive.statefp&disjunctive.countyfp&disjunctive.name&disjunctive.namelsad&disjunctive.stusab&disjunctive.state_name&refine.stusab=CA
18. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, & Yutani H. 2019. “Welcome to the tidyverse.” _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686
19. Wickham H, Bryan J. 2023. _readxl: Read Excel Files_. R package version 1.4.3, <https://CRAN.R-project.org/package=readxl>

**Appendix - More detailed description of the data preparation for analysis**

1. The Mushroom Observer GitHub provides several different .csv files. There is a main observations file which only provides the date of observations. The complementary names and locations files are used to match taxonomic and location information to each observation. Numbers link these values, and so code was required to join these datasets. All data manipulation, exploration, and analysis was completed in R, with a great deal of tidyverse.[11,18]

2. A function to search for specific fires based on longitude/latitude, or fire name was made to aid in specificity. In addition, a visual map made using the Leaflet package[1] in R was created with the latitudinal and longitudinal bounds with an error threshold to include fires in close proximity to the counties for further visualization and comprehension for interpretation.

3. The observation data was subsetted according to the time frame and region of the weather data; latitudinal and longitudinal bounds for each region were roughly estimated by mapping the observations using Leaflet[1]. A rough area was chosen which enclosed all relevant observations while excluding observations not in the region.

4. While count data has a known poisson distribution, the distribution of diversity index values is unknown. Thus, in order to fit a GLM to the diversity calculations, maximum likelihood estimations were obtained for gamma, inverse-Gaussian, and exponential distributions fitted on top of the data. Akaike information criterion (AIC) scores were calculated from these likelihood estimates and the distribution corresponding to the lowest AIC value was utilized in the GLMs for diversity. For all locations, results showed that the gamma distribution was the best-fitting distribution (Figure 5). As a consequence, diversity values of 0 must be taken out as the gamma distribution does not contain zeros.

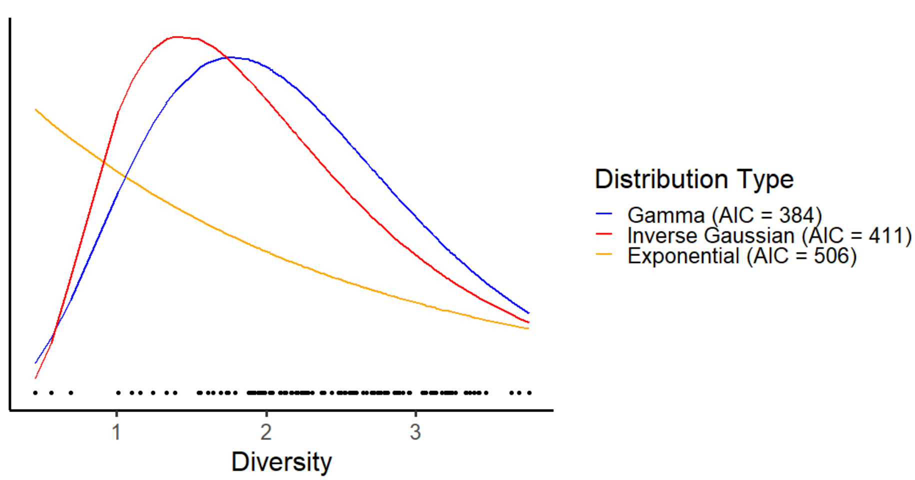


**Figure 5.** Distributions fit over San Francisco diversity values. Parameterized using MLEs.

5. Random intercept model had an AIC value of 326 while the random intercept-slope model had an AIC value of 330. Therefore, the random intercept model was chosen due to its better fit of the data.

6. https://www.desmos.com/calculator/1uqko70hmc - If interested, a quick put-together of the GLMs with interaction terms to see how the interaction terms change how temperature and precipitation affect abundance/diversity. Overall, the effect direction is as indicated by the individual coefficients, but the interaction term makes it so that large precipitation values cause temperatures to increase diversity/abundance, while low temperatures cause precipitation to decrease diversity/abundance.

**Supplemental**

*Global.Rmd*

- Central R notebook file containing all of the code used to manipulate, explore, and analyze data.

*observations.csv*

- The central Mushroom Observer datafile
- Each row represents one separate mushroom observation into the database
- Columns used:
    - name_id: number referring to taxonomic classification contained within the names.csv and name_classifications.csv files
    - when: date of the observation (In yyyy-mm-dd)
    - location_id: number referring to place of observation contained within the locations.csv file
    - is_collection_location: a binary value saying if the observed mushroom was found at the specified location or if it was imported from elsewhere (1 = collected on-site, 0 = brought in)

*names.csv*

- Additional Mushroom Observer file used to give mushroom ID information
- Each row represents a distinct species, genus, family, or etc.
- Columns used:
    - id: matches the name_id value from the observations.csv file
    - text_name: The species name matching the name id, or a less specific classification if species wasn't specified
    - rank: A number referring to the taxonomic level of the id (4 = species, 9 = genus, > 9 = family and beyond)

*name_classifications.csv*

- Further ID file which gives the whole taxonomic classification of each name id, from domain to species
- Columns used:
    - name_id: the same name_id as the observations file
    - family: the taxonomic family classification for the given ID

*locations.csv*

- Mushroom Observer file used to give location data for each observation
- Each row is a distinct location where a mushroom has previously been reported
- Columns used:
    - id: matches the location_id value from the observations.csv file

        - name: gives the location name, giving at least a city and country
        - north, south: the upper and lower latitudes for the location given
        - east, west: the eastern and western longitudes for the location given

*mapdataall.csv*

- Fire data including quantitative and qualitative data for California counties.
- Columns used:
    - incident_name
    - incident_date_created
    - incident_county
    - incident_acres_burned

*us-county-boundaries.xlsx*

- California county border data to segment and differentiate border thresholds.
- Columns used:
    - NAMELSAD (county name)
    - INTPTLAT (latitude)
    - INTPTLON (longitude)

*MonthClim.txt*

- Monthly climate data for San Francisco
- Each row is the average/total climate for each reported month
- Columns used:
    - date (in mm/yyyy)
    - mean temperature (in °C)
    - total precipitation (in mm)

*ontarioClimate.csv*

- Monthly climate data for Toronto (used as a proxy for south Ontario)
- Each row is the average/total climate for each reported month
- Columns used:
    - TOTAL_PRECIPITATION: total precipitation (in mm)
    - LOCAL_DATE: date (in yyyy/dd/mm)
    - MEAN_TEMPERATURE: mean temperature (in °C)

*colombiaPrecip.per*

- Monthly cumulative precipitation for Colombia, each row being one reported year
- Columns used:
    - YEAR: in yyyy
    - JAN, to DEC: total precipitation in the given month (in mm)

*colombiaTemp.per*

- Monthly mean temperature for Colombia, each row being one reported year
- Columns used:
    - YEAR: in yyyy
    - JAN, to DEC: mean temperature in the given month (in °C)