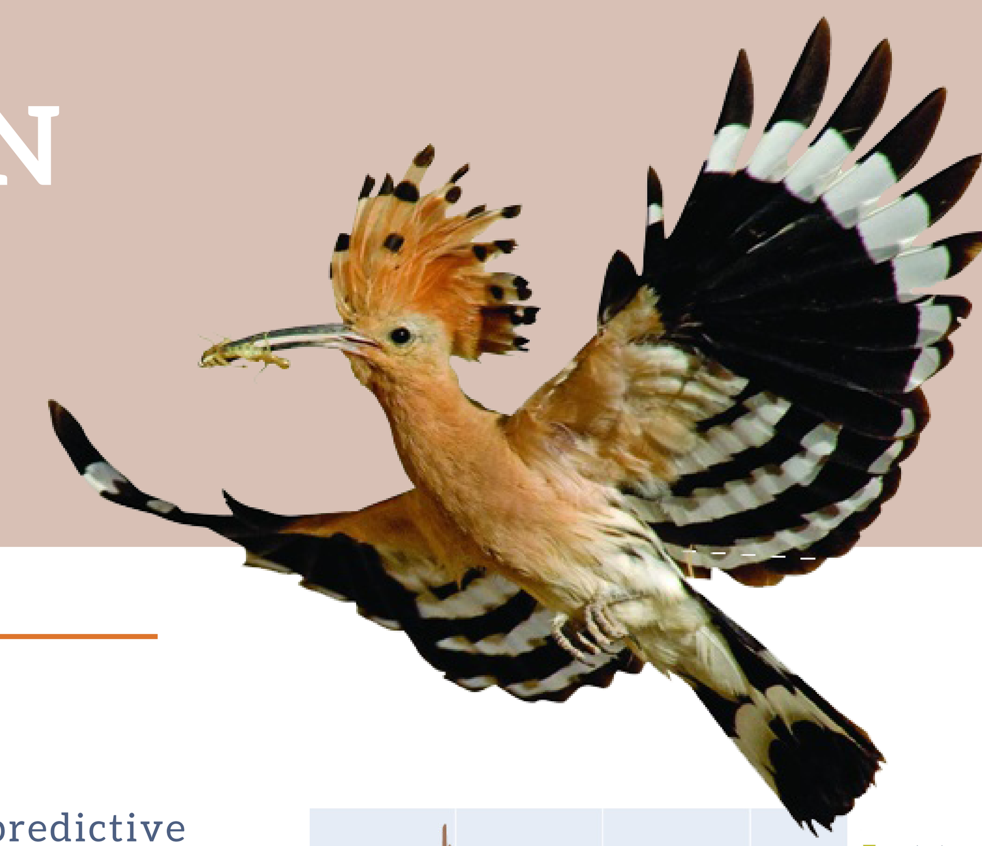
FEATURE GENERATION FOR ANOMALOUS BIRD SIGHTING DETECTION

HAW HAMBURG | MARS GROUP Multi-Agent Research & Simulation | DDA | VOGELWARTE.CH

Simon Birker
simon.birker@haw-hamburg.de



MOTIVATION & PROBLEM STATEMENT

The integration of additional environmental features derived from the spatial information into ecological citizen science datasets can offer advancements in validating the authenticity of the submitted records. Traditional validation methods relying solely on date, location, and species may not capture the full context of a sighting.

In collaboration with DDA and Vogelwarte Switzerland, Outlier Detection models capable of identifying implausible bird sightings were built. By integrating advanced environmental factors, such as land cover, altitudes, and weather conditions, we expect to enhance the model's ability to assess sighting plausibility.

RELATED WORK

- Research on the validation of citizen science data using Machine Learning is limited [1].
- Hence, in this field, we see a notable lack of research on feature generation in the context of ecological citizen science data validation.
- Using similar data, Teng et al. enriched bird sighting data with land cover, weather, and pedologic data to enhance species distribution modelling [2]. A similar endeavour was undertaken by Arenas-Castro et al. in [3].

DATA ANALYSIS

Each recorded bird sighting includes the species, date, and location represented by latitude and longitude coordinates (see **Figure 1**).

While latitude and longitude alone may lack substantial information for modelling, the extraction of more detailed information could provide valuable insights for detecting implausible data. For instance, birds often exhibit habitat preferences, such as specific land cover types and preferred altitude ranges. Aquatic birds, for example, can be assumed to be observed in proximity to water bodies, as illustrated in **Figure 2**.

Species: Great Crested Grebe
Date: 10th July '23
Latitude: 47.143118540352
Longitude: 8.152418532701

Figure 1. Features of a sample record. Images are not used in this project.

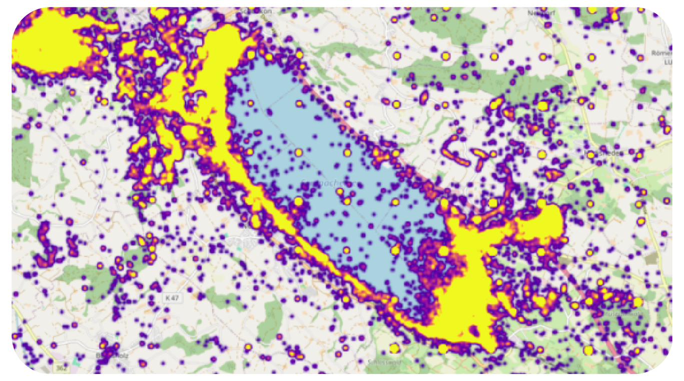


Figure 2: Habitat distribution of Great Crested Grebes at Sempacher See (CH).

GENERATED FEATURES

When creating the land cover, altitude, and weather features, the existing data is used, in particular the temporal and spatial parameters date, latitude, and longitude recorded at the time of the bird observation. The development process of each feature involved the application of several methods, whereby the optimal approach was selected for the final creation of the feature.

1 LAND COVER

Intuition: Birds observed in land cover surroundings atypical for their species can be regarded as less plausible.

Implementation: The latitude and longitude data were merged with the Corine Land Cover (CLC) dataset. This results in either the Land Cover at the sighting location or a percentage of the surrounding land covers. The percentage is calculated within a 1 km square around the sighting. **Figure 3** shows the percentages of land cover per species.

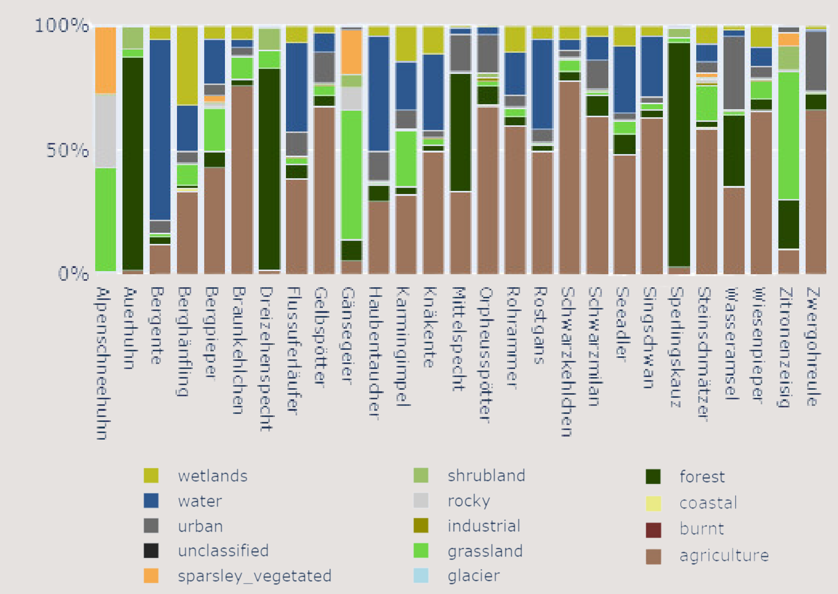


Figure 3: Generated Land Cover distribution per species.

2 ELEVATION

Intuition: Bird species are typically associated with specific altitude ranges; therefore, altitude data could serve as a valuable feature for validating the authenticity of a bird sighting.

Implementation: Elevation was computed from coordinates using API-based approaches, including Google Earth Engine, and local computations with the European Digital Elevation Model (DEM). The latter was selected for its rapid computation of altitude values.

3 WEATHER

Intuition: Weather conditions can affect bird behaviour, such as migration patterns or local habitat preferences.

Implementation: Weather features, including temperature, humidity, and wind, are fetched based on Earth Engine API and the Meteostat API. The latter was chosen due to a superior fetch speed. **Figure 4** shows different preferred temperature ranges between bird species.

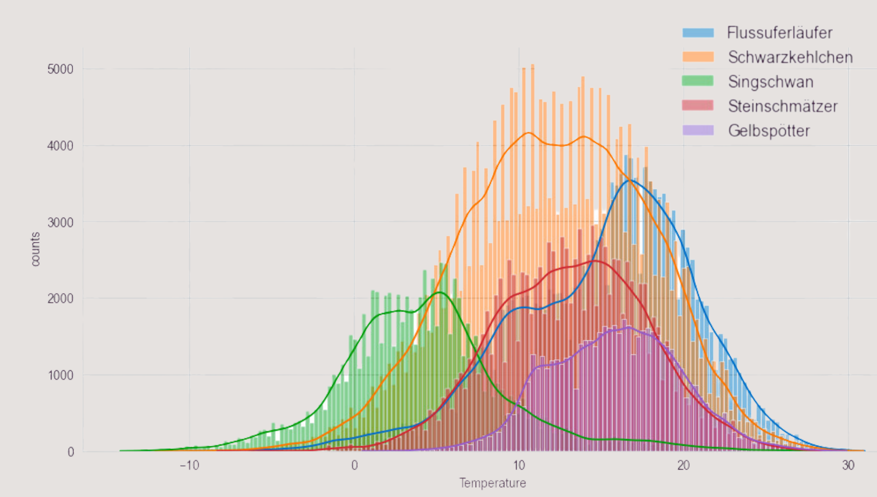
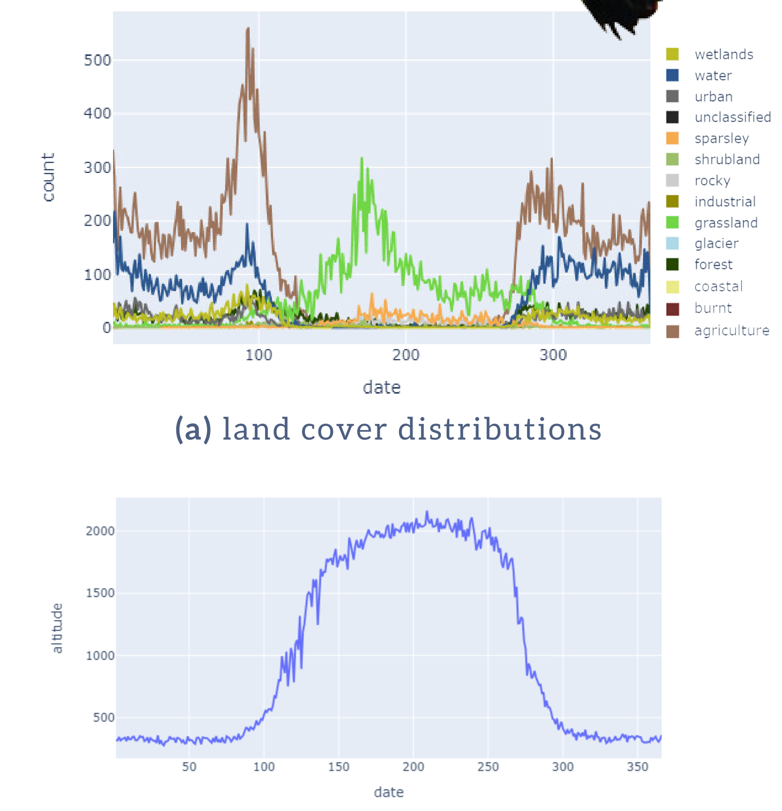


Figure 4: Temperature distribution for five specific bird species.

RESULTS

As a preliminary assessment of the predictive value of the generated features, the feature distribution of the water pipit was assessed. This migratory bird exhibits notable changes in altitude and land cover throughout the year. During the summer months, the water pipit moves to higher altitudes, which are predominantly characterized by grassland land cover. While this behaviour is not evident from location and date alone, the seasonal movement is observable in the land cover and altitude features, as depicted in **Figure 5**. This indicates that these features have the potential to significantly assist models in capturing implausible sightings.



(a) land cover distributions

(b) altitude distributions

Figure 5: Seasonal feature distributions of water pipit sightings.

FUTURE DIRECTIONS

The integration of land cover, elevation, and weather data into the analysis of bird sightings presents a promising approach for enhancing the accuracy of models designed to validate bird sightings. The newly generated features encapsulate critical environmental factors that could significantly influence bird presence and behaviour.

However, the utility of these features within predictive models remains to be properly tested. Future work will involve a detailed evaluation of the features' predictive value and their impact on model performance. This will ensure that the models not only leverage the environmental context of bird sightings but also improve validation accuracy.

REFERENCES

[1] Baker, Emily, et al. "The verification of ecological citizen science data: current approaches and future possibilities." Citizen Science: Theory and Practice 6.1 (2021).

[2] Teng, Mélisande, et al. "Bird distribution modelling using remote sensing and citizen science data." arXiv preprint arXiv:2305.01079 (2023).

[3] Arenas-Castro, Salvador, et al. "Effects of input data sources on species distribution model predictions across species with different distributional ranges." Journal of Biogeography 49.7 (2022): 1299-1312.