

EGU22-5632

<https://doi.org/10.5194/egusphere-egu22-5632>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Data-Driven Sentinel-2 Based Deep Feature Extraction to Improve Insect Species Distribution Models

Joe Phillips¹, Ce Zhang^{1,3}, Bryan Williams², and Susan Jarvis³

¹Lancaster Environment Centre, Lancaster University, LA1 4YQ, U.K

²School of Computing and Communications, Lancaster University, LA1 4WA, U.K.

³UK Centre for Ecology & Hydrology, Library Avenue, Lancaster, LA1 4AP, U.K

Despite being a vital part of ecosystems, insects are dying out at unprecedented rates across the globe. To help address this in the UK, UK Centre for Ecology & Hydrology (UKCEH) are creating a tool to utilise insect species distribution models (SDMs) for better facilitating future conservation efforts via volunteer-led insect tracking procedures. Based on these SDM models, we explored the inclusion of additional covariate information via 10-20m² bands of temporally-aggregated Sentinel-2 data taken over the North of England in 2017 to improve the predictive performance. Here, we matched the 10-20m² resolution of the satellite data to the coarse 100² insect observation data via four methodologies of increasing complexity. First, we considered standard pixel-based approaches, performing aggregation by taking both the mean and standard deviation over the 10m² pixels. Second, we explored object-based approaches to address the modifiable areal unit problem by applying the SNIC superpixels algorithm over the extent, with the mean and standard deviation of the pixels taken within each segment. The resulting dataset was then re-projected to a resolution of 100m² by taking the modal values of the 10m² pixels, which were provided with the aggregated values of their parent segment. Third, we took the UKCEH-created 2017 Land Cover Map (LCM) dataset and sampled 42,000, random 100m² areas, evenly distributed about their modal land cover classes. We trained the U-Net Deep Learning model using the Sentinel-2 satellite images and LCM classes, by which data-driven features were extracted from the network over each 100m² extent. Finally, as with the second approach, we used the superpixels segments instead as the units of analysis, sampling 21,000 segments, and taking the smallest bounding box around each of them. An attention-based U-Net was then adopted to mask each of the segments from their background and extract deep features. In a similar fashion to the second approach, we then re-projected the resulting dataset to a resolution of 100m², taking the modal segment values accordingly. Using cross-validated AUCs over various species of moths and butterflies, we found that the object-based deep learning approach achieved the best accuracy when used with the SDMs. As such, we conclude that the novel approach of spatially aggregating satellite data via object-based, deep feature extraction has the potential to benefit similar, model-based aggregation needs and catalyse a step-change in ecological and environmental applications in the future.