When Specimens Become Environmental Traits: Automating data extraction in aerobiology

Mónica González-Alonso[‡], Jaime Ulayar-Arroyo[§], Arturo H. Ariño[‡,|]

‡ University of Navarra, Institute of Biodiversity and Environment, Pamplona, Spain
§ University of Navarra, Pamplona, Spain
| Institute of Data Science and Artificial Intelligence (DATAI), Pamplona, Spain

Corresponding author: Arturo H. Ariño (artarip@unav.es)

Abstract

The identity of a biological specimen can be considered the most basic type of trait: the species (or higher taxon) the specimen belongs to, one of the three most fundamental data elements (along with location and time) that configure the primary biodiversity data record whose quantification may underpin biological monitoring programs. A general image containing specimens (for example, a photo from a surveillance camera), where individual organisms are to be identified, is akin to an image of a single organism where traits are to be recognized: the individual, often distinct organisms in the image, become the traits of the image.

In many cases, automated segmentation and classification/identification and counting of individuals captured within general images containing multiple organisms is most desirable —provided precision and recall exceed a reasonable threshold: that which minimizes the false-positive error rate, as compared to human identification in images. When this occurs, quantitative monitoring data are amenable to good automation, saving much time and human effort.

One particularly well-suited case is that of airborne pollen and spore monitoring, hitherto done manually in most of the world. However, systems are being developed that can recognize atmospheric pollen content from images. Two main approaches are possible: either fully automated systems, or recognition from captured field images. In both cases, some instance of artificial intelligence, most often a convolutional neural network, is tasked with the recognition of pollen and spore types, after proper training, from microscopic images. Several problems exist for the wider adoption of such approaches. Among them, the workflow order of segmentation and identification, the portability of the neural network training for different locations, the differences in performance for distinct particle types, and the generation and archival of images: field images vs. individual (segmented) detected specimens, which call for wholly different approaches to metadata.

In this presentation we review the general outlook of automated airborne pollen and spore monitoring data, and compare performance data in some specific deep learning settings and workflows, based on actual monitoring work, training and tests with real images.

Keywords

pollen, spores, convolutional neural networks, segmentation, artificial intelligence, monitoring

Presenting author

Arturo H. Ariño, Mónica González-Alonso

Presented at

TDWG 2022

Funding program

MGA is funded by the Asociación de Amigos de la Universidad de Navarra

Conflicts of interest

The authors declare no CoI.