Check for updates

Editorial: Floristic and vegetation studies in the era of big data: challenges, trends and applications

Mauro Fois [1]*, Corrado Marcenò [2] and Scott Brian Franklin [3]

[1]Dipartimento di Scienze della Vita e dell'Ambiente, Centro Conservazione Biodiversità, Università degli Studi di Cagliari, Cagliari, Italy, [2]Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy, [3]Department of Biological Sciences, University of Northern Colorado, Greeley, CO, United States

Editorial on the Research Topic
Floristic and vegetation studies in the era of big data: challenges, trends and applications

Monitoring spatial and temporal dynamics in plant species and community diversity is one of the major challenges in ecology and conservation. Long-term and accurate data from systematic survey programmes, with standardized sampling designs and rigorous protocols, are unfortunately scarce and concentrated in a few biomes such as GLORIA (alpine, Pauli et al., 2001) or Grassplot (Palaearctic grasslands, Dengler et al., 2018). Recent trends show that the implementation of traditional field and natural history museum collections are not keeping pace; that is, the number of species occurrences and habitat data seem inadequate to examine the changes occurring (Beck et al., 2014). On the other hand, the increasing availability of big datasets, often derived from modern digital technology, is promisingly supplementing information to monitor changes.

In plant science, species data occurrences (Robertson et al., 2014) and vegetation-plot databases (Chytrý et al., 2016; Bruelheide et al., 2019) are two of the most common and powerful tools to supplement existing research and provide new perspectives on more complex and geographically broader questions. Floristic and vegetation data were historically almost exclusively retrieved by experts, while, more recently, an increasing contribution from citizen science programmes (Nugent, 2018) and naturalist community platforms (Marcenò et al., 2021) offer an interesting opportunity to retrieve data. Online platforms can amass large and updated data with relatively little effort, while monitoring programs based on expert knowledge are still crucial to avoid biases in species identification and unbalanced efforts on particular species and habitats. The scientific and economic value of expert data providers might, for instance, be recognized in order to re-evaluate the important roles of the field botanist (Crisci et al., 2020) and to maintain high quality and reliable data. Moreover, the formulation of more inclusive data-sharing agreements might allow the growth of stable cross-political and cross-biomes collaborations, and enhance the interest and reliability of the research products (Bruelheide et al., 2018).

In addition, merging big datasets to tackle research questions can be fruitful. Utilizing trait datasets (Kattge et al., 2020) in combination with vegetation plots, for example, may help interpret trends in geographic vegetation shifts, especially in relation to climate and disturbance datasets. The addition of several satellites that provide increased spatial and temporal coverage and resolution that can be linked to plot data and individual tree and shrub species offers an extrapolation of geographic coverage and ecological questions (e.g., Lake et al., 2022; Rocchini et al., 2022).

With the rapid development of implementing, managing, and processing big data in plant science, this collection of articles aims to show the opportunities offered by the use of big data in different fields of plant science and discuss the main gaps and future challenges to better use.

Catarino et al. used the species data occurrences amassed in GBIF and in different online herbaria to study the species diversity of the Leguminosae family in Angola. They identified 953 taxa, of which 165 are endemics, giving information on their biogeographical and conservation status, life form, and main traditional use. The importance of having harmonized data on online platforms to conduct further floristic studies in little-explored areas is fundamental. Nevertheless, the authors concluded that one of the main issues of this study was the absence of recent data. Although in Angola, the abandonment of field research for several years was caused by the war, the same trend occurs worldwide. Big data offer also the opportunity to predict changes in the distribution ranges of endemic flora. The high value of worldwide repositories was confirmed by Lannuzel et al., who collected data on 87,733 plant occurrences from a hundred different original datasets, which allowed to double the number of known narrow endemic taxa and elucidate 68 putative new species in New Caledonia. Despite the promising reliability of automated data filtering, a vast amount of work for taxonomic analysis by local and international taxonomists was recognized as the most powerful way to reduce data biases and loss.

Some important sources to support floristic and vegetation studies are CHELSA and WorldClim which contain different models on past and future climatic predictions to be correlated with species occurrence data. Peyre performed species distribution models for 664 species in the Andean Páramo. The models were able to predict the extinction rate of the species pool analyzed and future gains and losses areas in the Páramo. Despite the interesting results obtained, the author confirms the necessity to complement these results with fine-scale studies. Knowing the past to understand the future was the approach used by Almeida et al. to reconstruct the spatial distribution trends of the rockrose (*Cistus ladanifer* L.). In this case, data from 10 different available sources were used for a single, relatively abundant and widely distributed species. A large amount of data required a considerable effort to reduce bias, redundancy and reduce spatial autocorrelation.

The final set of 2,833 revised records was satisfactorily used to elucidate historical, present and future population retractions and expansions through species distribution modeling.

The increasing amount of freely available data offers valuable opportunities for studies in plant ecology, conservation and environmental restoration. This Research Topic shows implications for different fields, with the main general aim to fill gaps and strengthen links between Linnean (i.e., lacking knowledge on total number of species), Darwinian (i.e., lacking species evolutionary relationship knowledge), and Wallacean (i.e., lacking species distribution knowledge) shortfalls (Diniz-Filho et al., 2013, 2023). Overall, this work supports the need to improve the quality of such a large quantity of resources, despite their already current usefulness. Automated data checking can facilitate this process, although the continuous training of experts and the strengthening of collaborations between them remains crucial. We hope that this Research Topic collection will provide some background in this interesting and burgeoning field.

Author contributions

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Beck, J., Böller, M., Erhardt, A., and Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* 19, 10–15. doi: 10.1016/j.ecoinf.2013.11.002

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., et al. (2019). sPlot–a new tool for global vegetation analyses. *J.Veg. Sci.* 30, 161–186. doi: 10.1111/jvs.12710

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S. M., et al. (2018). Global trait–environment relationships of plant communities. *Nat. Ecol. Evol.* 2, 1906–1917. doi: 10.1038/s41559-018-0699-8

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., et al. (2016). European vegetation archive (EVA): an integrated database of European vegetation plots. *Appl. Veg. Sci.* 19, 173–180. doi: 10.1111/avsc.12191

Crisci, J. V., Katinas, L., Apodaca, M. J., and Hoch, P. C. (2020). The end of botany. *Trends Plant Sci.* 25, 1173–1176. doi: 10.1016/j.tplants.2020.09.012

Dengler, J., Wagner, V., Dembicz, I., García-Mijangos, I., Naqinezhad, A., Boch, S., et al. (2018). GrassPlot – a database of multiscale plant diversity in Palaearctic grasslands. *Phytocoenologia* 48, 331–347. doi: 10.1127/phyto/2018/0267

Diniz-Filho, J. A. F., Jardim, L., Guedes, J. J. M., Meyer, L., Stropp, J., Frateles, L. E. F., et al. (2023). Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Front. Biogeogr.* 15, e59566. doi: 10.21425/F5FBG59566

Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A.O., and Bini, L. 0M. (2013). Darwinian shortfalls in biodiversity conservation. *Trends Ecol. Evol.* 28, 689–695. doi: 10.1016/j.tree.2013.09.003

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., et al. (2020). TRY plant trait database–enhanced coverage and open access. *Glob. Chang. Biol.* 26, 119–188. doi: 10.1111/gcb.14904

Lake, T. A., Briscoe Runquist, R. D., and Moeller, D. A. (2022). Deep learning detects invasive plant species across complex landscapes using Worldview-2 and Planetscope satellite imagery. *Remote Sens. Ecol. Conserv.* 8, 875–889. doi: 10.1002/rse2.288

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., et al. (2021). Facebook groups as citizen science tools for plant species monitoring. *J. Appl. Ecol.* 58, 2018–2028. doi: 10.1111/1365-2664.13896

Nugent, J. (2018). iNaturalist: citizen science for 21st-century naturalists. *Sci. Scope* 41, 12. Available online at: https://my.nsta.org/resource/?id=10.2505/4/ss18_041_07_12

Pauli, H., Gottfried, M., Reiter, K., and Grabherr, G. (2001). "High mountain summits as sensitive indicators of climate change effects on vegetation patterns: the "multi summit-approach" of GLORIA (Global Observation Research Initiative in Alpine Environments)," in *Global Change and Protected Areas. Advances in Global Change Research,* eds G. Visconti, M. Beniston, E. D. Iannorelli, and D. Barba (Dordrecht: Springer), 45–51. doi: 10.1007/0-306-48051-4_6

Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., et al. (2014). The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. *PLoS ONE* 9, e102623. doi: 10.1371/journal.pone.0102623

Rocchini, D., Santos, M. J., Ustin, S. L., Féret, J.-B., Asner, G. P., Beierkuhnlein, C., et al. (2022). The spectral species concept in living color. *J. Geophys. Res. Biogeosci.* 127, e2022JG007026. doi: 10.1029/2022JG007026