Methods of data preparation

Selection of vegetation-plot data from European Vegetation Archive

The initial selection of fen vegetation-plot records from the European Vegetation Archive and our unpublished data sets (Table 1: Data sources) was conducted following the formal definition of a fen, reproducing the rules described in Peterka et al. (2017). The resulting data set of 37.273 records still contained some records transient between fens and other habitats, usually aquatic vegetation, mesic to wet grasslands, or moist rocks, or represented only bryophyte communities (e.g., bryophyte patches in fen woodlands). In some cases, their assignment to fens was highly questionable. Further cleaning was hence applied. We kept only the records with at least one bryophyte and at least one vascular-plant species from the list of mire and spring specialist species (compiled from Jiménez-Alfaro et al., 2014; Mucina et al., 2016; Peterka et al., 2017; Horsáková et al., 2018; see below). We further deleted three records from Russia, where the number of mire and spring species was exceptionally outlying, suggesting that the plot area was actually much larger than the indicated area of 100 m^2. The number of mire and spring specialist species in the resulting data set hence varies between 2 and 45. In order to remove aquatic, wet grassland and moist rock vegetation, we further deleted records with mean indicator value for moisture (unweighted mean; conventional Ellenberg-type values for all European mire species; Hájek, M., Dítě, D., Horsáková, et al., 2020) lower than 6 or higher than 10, or with mean indicator value for drought intolerance (weighted by both niche amplitude and cover, for details and justification see Hájek, M., Dítě, D., Horsáková, et al., 2020) lower than 5 or higher than 9. These restrictions effectively removed the non-fen records. All of the removed records were indeed those which were unclassified by the expert system for the recognition of the Eunis units (Chytrý et al., 2020).

A geographical stratification was applied - maximum five relevés were randomly selected per grid of 0.75 degree of latitude and 1.25 of longitude. The resulting data set contained 35,984 relevés.

Imputation of adjusted pH

Using the calibration data set (6,299 records for pH; 5,073 records for conductivity), we imputed the variable combining pH a calcium into a single variable, called adjusted pH (Plesková et al., 2016; Horsáková et al., 2018), into the entire European data set, hereinafter called "imputation data set".

In order to compare the precision of the two imputation methods described below, the calibration data subsets with both pH and conductivity measurements (5073 records) was randomly divided into two groups: 1/3 was training data set (1691 records) and 2/3 was testing data set (3382 records). We tested the efficiency of the two calibration methods: the imputation by weighted averaging and the Imputation by the compositional similarity (the MOSS method of Tichý et al., 2010).

For imputation by weighted averaging, medians (optimum) and inverse standard deviations (1/sd; i.e., amplitude) of pH and log-conductivity were calculated for each species in the calibration data set. Subsequently, optimum values were averaged for each vegetation-plot record in the imputation data sets using weighting by the amplitude. The method mirrors the averaging of ecological (e.g., Ellenberg) indicator values. Zero weight was set to rare species with extremely low amplitude and the resulting weight therefore higher than 10. Log-conductivity was then transformed back to conductivity.

For imputation by compositional similarity, we use the MOSS method (Tichý et al., 2010). For each record of the imputation data set, the method searched for five compositionally the most similar relevés in the calibration data set (the Bray-Curtis distance of the presence-absence data was used) and calculated means of their pH or conductivity, respectively. Maximum allowed distance was 0.6.

The MOSS method of imputation yielded Spearman correlation between the predicted and measured values 0.85 (pH) and 0.87 (conductivity), respectively, while weighted averaging method yielded values of 0.83 (pH) and 0.85 (conductivity) and a rather quadratic relationship with only few predicted values above 7. Root mean squared error of prediction (RMSP) was clearly lower for the MOSS method than for weighted averaging (0.31 versus 0.55 for log-conductivity; 0.52 versus 0.74 for pH).

Based on these results, we finally calibrated the imputation data set (35,984 vegetation-plot records). Initially, we assigned the original measurements of pH and conductivity (4,012 original measurements of pH, 3,993 original measurements of conductivity). Note that the calibration data set was larger (6,299 and 5,073 measurements, respectively) because it also contained springs, bogs and transitions to wet grasslands. The missing values were then imputed in the two steps. First, the imputation resulting from the MOSS method was applied. After it, 16% of vegetation-plot records still showed missing values of pH (6.115 records) and 21% still showed missing value of conductivity (8,187 records), because of the threshold of maximum allowed distance for the MOSS method. These missing values were imputed by the weighted averaging method that was available for all vegetation-plot records.

Calibration with water table depth

Each vegetation-plot record in the data set was calibrated by an indicator value for water-table depth, using mean indicator value for drought intolerance (minimum value of an ecological tolerance expressed at the scale 1-12; Hájek, Dítě, Horsáková, et al., 2020). The average value per each vegetation-plot record was weighted by both the niche amplitude (using tabelled values presented in Hájek, Dítě, Horsáková, et al., 2020) and the percentage cover of individual species. Hájek, Dítě, Horsáková, et al. (2020) found that weighting with niche amplitude and cover causes highest correlation with measured water table data in the datasets containing different habitats.

The nomenclature follows Euro+Med checklist for vascular plants, Hill et al. (2016) for mosses and Frey et al. (2006) for hepatics. For the content of aggregates see Peterka et al. (2017).

References

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J.S., … Schaminée J.H.J. (2020). EUNIS Habitat Classification: expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, *23*(4), 648-675. https://doi.org/10.1111/avsc.12519

Frey, W., Frahm, J.-P., Fischer, E. & Lobin, W. (2006). The liverworts, mosses and ferns of Europe. Harley Books.

Hájek, M., Dítě, D., Horsáková, V., Mikulášková, E., Peterka, T., Navrátilová, J., ... & Horsák, M. (2020). Towards the pan-European bioindication system: Assessing and testing updated hydrological indicator values for vascular plants and bryophytes in mires. *Ecological Indicators, 116,* 106527. https://doi.org/10.1016/j.ecolind.2020.106527

Hill, M. O., Bell, N., Bruggeman-Nannenga, M. A., Brugués, M., Cano, M. J., Enroth, J., ... & Söderström, L. (2006). An annotated checklist of the mosses of Europe and

Macaronesia. *Journal of bryology*, *28(3),* 198-267. https://doi.org/10.1179/174328206X119998

Horsáková, V., Hájek, M., Hájková, P., Dítě, D., & Horsák, M. (2018). Principal factors controlling the species richness of European fens differ between habitat specialists and matrix-derived species. *Diversity and Distributions*, *24*(6), 742-754. https://doi.org/10.1111/ddi.12718

Jiménez-Alfaro, B., Hájek, M., Ejrnaes, R., Rodwell, J., Pawlikowski, P., Weeda, E.J., Laitinen, J., Moen, A., Bergamini, A., Aunina, L., Sekulová, L., Tahvanainen, T., Gillet, F., Jandt, U., Dítě, D., Hájková, P., Corriol, G., Kondelin, H. and Díaz, T.E. (2014). Biogeographic patterns of base-rich fen vegetation across Europe. *Applied Vegetation Science*, 17, 367-380. https://doi.org/10.1111/avsc.12065

Mucina, L., Bültmann, H., Dierßen, K., Theurillat, J.-P., Raus, T., Čarni, A., ... Tichý, L. (2016), Vegetation of Europe: hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities. *Applied Vegetation Science,* 19: 3-264. https://doi.org/10.1111/avsc.12257

Peterka, T., Hájek, M., Jiroušek, M., Jiménez-Alfaro, B., Aunina, L., Bergamini, A., … & Chytrý, M. (2017). Formalized classification of European fen vegetation at the alliance level. *Applied Vegetation Science, 20(1),* 124-142. https://doi.org/10.1111/avsc.12271

Plesková, Z., Jiroušek, M., Peterka, T., Hájek, T., Dítě, D., Hájková, P., Navrátilová, J., Šimová, A., Syrovátka, V., & Hájek, M. (2016). Testing inter-regional variation in pH niches of fen mosses. *Journal of Vegetation Science, 27(2),* 352-364. https://doi.org/10.1111/jvs.12348

Tichý, L., Hájek, M. and Zelený, D. (2010), Imputation of environmental variables for vegetation plots based on compositional similarity. *Journal of Vegetation Science*, 21: 88-95. https://doi.org/10.1111/j.1654-1103.2009.01126.x