Supplementary material to

Dengler, J., Jansen., F., … & Gillet, F. (2023) Ecological Indicator Values for Europe (EIVE) 1.0. *Vegetation Classification and Survey*.

Supplementary material 4. Methodological details of the calculations of the three variants of niche position (median, mean, weighted mean) and the three variants of niche width (nw1, nw2, nw3).

Creating a consensus system of niche positions

To derive European values of niche position, we applied three different approaches to combine the scaled EIVs of all systems in which the respective taxon was included: (i) median; (ii) mean and (iii) weighted mean. With our "niche positions" we aim to approximate the positions on ecological gradients where half of the occurrences of a plant species are below and the other half above. Therefore, niche position differs from realised niche optimum (the environmental conditions under which a species is most frequent and/or reaches the highest cover values), particularly in case of skewed or bimodal distributions. We based the weighting on the $\log_{10}$-transformed area (in km^2), while counting (partial) areas that were covered by more than one EIV system only partially. For example, each of the three Hungarian EIV systems was assigned only one third of the area of Hungary. The motivation behind this weighting was that EIV systems that cover larger and geographically distinct areas add more information than those that cover smaller or overlapping areas, but the information does not increase linearly. For the EIVs that covered only a habitat-specific subset of the flora of the respective region (e.g. grasslands or mires), we did not apply the rule of overlapping territories as this would have made the calculation very complicated with negligible effects on the final weighting factors. We then subtracted a constant so that the lowest weighting factor ($EIVweight_j$) was 1.0 ("Georgia") and the largest 5.0 ("European_Mires") (see Supplementary material 1).

Accordingly, the initial indicator value of a taxon *i* of the European consensus system (*EIVEini*) was derived as follows from the scaled values in the individual EIV systems (*EIVini*) in the three calculation variants (Fig. 2, step 2):

$$EIVEini.med_i = \mathrm{median_j}(EIVini_{i,j})$$

$$EIVEini.m_i = \mathrm{mean_j}(EIVini_{i,j})$$

$$EIVEini.wm_i = \frac{\sum_j \ (EIVini_{i,j} \cdot EIV.weight_j)}{\sum_j \ EIV.weight_j}$$

We evaluated the results via linear regression and correlation coefficients of *EIVEini* against all expert-scaled EIV systems (*EIVini*) for each of the five indicators (Supplementary material 10). If the expert-based scaling was perfect, all regression lines would be on the 1:1 line. While many came close, most had a shallower slope (Supplementary material 9), meaning that the range of realised environmental conditions in a region was smaller than assumed in the expert-based scaling. Only in two cases (Ukraine and USSR_Tsyganov for M) the opposite was true (Supplementary material 10). We thus tried to remove the remaining major discrepancies in the concepts of the different EIV systems by an automated linear optimisation (Fig. 2, step 3) with a_j and b_j being intercept and slope, respectively, of the regression *EIVEini* vs. *EIVini*, separately for the three calculation variants, by adjusting the values of both *EIVini* and *EIVEini* in an iteration to get *EIVadj* and *EIVEadj*, respectively:

$$EIVadj.med_{i,j} = a.med_j + b.med_j \cdot EIVini_{i,j}$$

$$EIVadj.m_{i,j} = a.m_j + b.m_j \cdot EIVini_{i,j}$$

$$EIVadj.wm_{i,j} = a.wm_j + b.wm_j \cdot EIVini_{i,j}$$

This numerical procedure standardised all regression lines for *EIVEini* vs. *EIVadj* to lie exactly on the 1:1 line. Subsequently, we created a new consensus system *EIVEadj* from the *EIVadj* values with the three variants as before (Fig. 2, step 4):

$$EIVEadj.med_i = \mathrm{median_j}(EIVadj.med_{i,j})$$

$$EIVEadj.m_i = \mathrm{mean_j}(EIVadj.m_{i,j})$$

$$EIVEadj.wm_i = \frac{\sum_j \, (EIVadj.wm_{i,j} \cdot EIV.weight_j)}{\sum_j \, EIV.weight_j}$$

The resulting fit between *EIVEadj* and *EIVadj* was on average better (i.e. the slope was closer to 1) than between *EIVEini* and *EIVini* (Supplementary material 10). The mean slope based on the mean variant for all five indicators (see Supplementary material 10), but when we tried another round of iteration, this did not or only marginally lead to further improvement. Thus, we stuck with *EIVEadj*. However, the iteration generally caused a contraction, or very rarely an expansion, of the value range, so that *EIVEadj* did not cover the full intended range of 0 to 10 anymore. Thus, a final step of rescaling (Fig. 2, step 5) was applied to the three variants of *EIVEadj* to get *EIVEres* as the European indicator values of niche position:

$$EIVEres.med_i = 10\,\frac{EIVEadj.med_i - \min(EIVEadj.med)}{\max(EIVEadj.med) \ - \min(EIVEadj.med)}$$

$$EIVEres.m_i = 10\,\frac{EIVEadj.m_i - \min(EIVEadj.m)}{\max(EIVEadj.m) \ - \min(EIVEadj.m)}$$

$$EIVEres.wm_i = 10\frac{EIVEadj.wm_i - \min(EIVEadj.wm)}{\max(EIVEadj.wm)\ - \min(EIVEadj.wm)}$$

The exactly same rescaling was applied to the corresponding variants of *EIVadj* to get *EIVres* (Fig. 2, step 6):

$$EIVres.med_{i,j} = 10\frac{EIVEadj.med_i - \min(EIVEadj.med)}{\max(EIVEadj.med)\ - \min(EIVEadj.med)}$$

$$EIVres.m_{i,j} = 10\frac{EIVEadj.m_i - \min(EIVEadj.m)}{\max(EIVEadj.m)\ - \min(EIVEadj.m)}$$

$$EIVres.wm_{i,j} = 10\frac{EIVEadj.wm_i - \min(EIVEadj.wm)}{\max(EIVEadj.wm)\ - \min(EIVEadj.wm)}$$

Deriving European niche width indicators

To establish a European indicator of niche width for each taxon on each of the five niche dimensions, we developed a separate workflow for the heterogeneous information in the various EIV systems. While some provided only a niche position, others provided niche width information as a range (minimum and maximum values) or as amplitude classes with two to four levels. If a source EIV system contained categorical niche amplitude information, we harmonised the coding. In Supplementary material 2, amplitude classes are stored as "#" for particularly narrow amplitude, "I" for normal amplitude, "II" for wide amplitude, but not indifferent, and "x" for "indifferent". We considered uncertain information (coded by smaller font in Ellenberg et al. 1991) for the purpose of calculating mean indicator values of a plot as equivalent to a wide amplitude (II).

For the further calculations, we chose the final outcomes of the EIVE niche position calculation, i.e. the rescaled values (*EIVEres*) of the best variant (see above). In EIV systems *j* with range-based niche width coding, we derived the amplitude of taxon *j* ($EIV.a_{i,j}$) as follows (Fig. 2, step 7):

$$EIV.a_{i,j} = (EIVori.max_{i,j} - EIVori.min_{i,j})\frac{\max\left(EIVres_j\right)\ -\ \min\left(EIVres_j\right)}{\max\left(EIVori_j\right)\ -\ \min\left(EIVori_j\right)}$$

If for a certain taxon in a range-based system, minimum and maximum were the same ($EIVori.max_{i,j} = EIVori.min_{i,j}$), we assigned to $EIV.a_{i,j}$ half of the minimum non-zero amplitude that occurred for other taxa in this system to account for the fact that a niche width of zero does not exist. In case of EIV systems with categorical niche width coding, we assumed standard widths *w* for each of the four categories on the scale of 0 to 10, namely # → 1.25, I → 2.5, II → 5 and x → 7.5 (Fig. 2, step 8). In absence of precise definitions (which was the case for most of the sources), we assume that these assignments generally should reflect more or less the intended meaning of the authors, at least their relative relationships. The final amplitude of taxon *i* in EIV system *j* ($EIV.a_{i,j}$) was calculated as follows (note that here *EIVini* and not *EIVori* had to be used as starting point):

$$EIV.a_{i,j} = w\frac{\max\left(EIVres_j\right) - \min\left(EIVres_j\right)}{\max\left(EIVini_j\right) - min\left(EIVini_j\right)}$$

To derive European indicators for niche width, we applied three different approaches to combine the rescaled niche position and niche width indicators of all EIV systems (Fig. 2, step 9). They have been constructed to meet the idea that the niche width at European level is composed of intraregional and interregional variability in the niches.

The first niche width index is based on the total range across all EIV systems, considering the amplitude and disregarding minimum values below 0 and maximum values above 10 (Fig. 2, step 7):

$$EIVE.nw1_i = \min\left(10, EIVres_i + \max(1, \frac{EIV.a_i}{2})\right) - \max\left(0, EIVres_i - \max(1, \frac{EIV.a_i}{2})\right)$$

Note that a rescaled value of 1 was used as a default for half of the niche width when the species *i* was not included in any EIV system with niche width information.

The second niche width index was calculated as the sum of the average amplitude of taxon *i* across EIV systems (intra-regional variation) and of the position range (inter-regional variation), which is the difference between the extreme values of niche position for taxon *i* across all EIV systems, bounded to a maximum of 10:

$$EIVE.nw2_i = \min\left(10, \overline{EIV.a_i} + \max(EIVres_i) - \min(EIVres_i)\right)$$

The average amplitude was calculated only from the EIV systems with explicit niche width information. For taxa for which no information about amplitude was available in any system, we assigned to the average amplitude the mean rescaled amplitude of all taxa with available information.

The third niche width index was calculated as the sum of the average amplitude of taxon *i* across EIV systems (intra-regional variation) and of twice the population standard deviation (σ) of the niche position (inter-regional variation), bounded to a maximum of 10:

$$EIVE.nw3_i = \min\left(10, \overline{EIV.a_i} + 2\,\sigma(EIVres_i)\right)$$