

Chytrý M., Axmanová I., Divíšek J., Friesová K., Holubová D., Palpurina S., Řezníčková M., Večeřa M. & Knollová I. (2026) Czech Vegetation Database: an open source of vegetation-plot data for research and applications. – *Preslia* 98: 187–207.

## Supplementary Data S2. Czech Vegetation Database data description.

Following the structure of the Turboveg 2 database management software, vegetation-plot data in the Czech Vegetation Database are divided into three parts:

1. Species data
2. Header data
3. Remarks

### 1 Species data

Species data include unique identifiers for vegetation-plot observations and corresponding species, along with information about the vegetation layer and estimated abundance.

**Relevé number (RELEVÉ\_NR).** Unique numerical code for the vegetation-plot observation (relevé) in the database.

**Species number (SPECIES\_NR).** Unique numerical code referring to the taxon in the Czechia-Slovakia Turboveg species checklist. Translations of this code to different taxon concepts and names, as used in the various sources listed below, are provided in a [crosswalk table](#).

**Original taxon name in the Turboveg 2 database (TURBOVEG\_NAME).** This name corresponds to the Czechia-Slovakia species checklist used in the Turboveg 2 database.

**Taxon name in the currently accepted nomenclature (KAPLAN\_PLUS\_NAME).** For vascular plants, this name corresponds to the [Pladias Database of Czech Flora and Vegetation](#), which adopts the taxonomy and nomenclature accepted in the second edition of the Key to the Flora of the Czech Republic (Kaplan et al. 2019), with a few additions. For bryophytes and lichens, it follows the DaLiBor database. Where not available, the concept is harmonized with the GBIF taxonomy.

- Kaplan Z., Danihelka J., Chrtek J., Kirschner J., Kubát K., Štech M. & Štěpánek J. (eds) (2019) *Klíč ke květeně České republiky [Key to the Flora of the Czech Republic]*. Ed. 2. Academia, Praha.
- Man M., Malíček J., Kalčík V., Novotný P., Chobot K. & Wild J. (2022) DaLiBor: Database of Lichens and Bryophytes of the Czech Republic. *Preslia* 94: 579–605. <https://doi.org/10.23855/preslia.2022.579>

### Taxon name in the currently accepted nomenclature, analysis-friendly version

**(KAPLAN\_PLUS\_NAME\_SIMPLE).** The same as KAPLAN\_PLUS\_NAME but with x instead of hybrid marks and ä, ě written as a, e for easier coding (e.g., *Salix x fragilis*, *Hierochloa australis*).

**Taxon author citation (AUTHORSHIP).** Information taken primarily from the Pladias database for vascular plants and from the DaLiBor database for bryophytes and lichens. If not available in these sources, it was taken from the GBIF backbone taxonomy (GBIF Secretariat 2023).

**Taxon name used in the CzechVeg-ESy expert system (CZECHVEG\_ESY\_NAME).** This taxonomy groups some taxonomically challenging taxa into higher units, such as aggregates. It is used in the expert system CzechVeg-ESy (Chytrý et al. 2020) for automatic vegetation classification according to Vegetation of the Czech Republic (Chytrý 2007–2013). Note that subspecies are abbreviated as “ssp.” and genus-level taxa are supplemented by “species” as in the original file.

- Chytrý M. (ed.) (2007–2013) *Vegetace České republiky 1–4. Vegetation of the Czech Republic 1–4*. Academia, Praha. <https://botzool.cz/vegsci/vegetationCR>
- Chytrý M., Tichý L., Boublík K., Černý T., Douda J., Hájek M., Hájková P., Hédli R., Kočí M., Krahulec F., Kučera T., Landucci F., Láníková D., Lososová Z., Navrátilová J., Petřík P., Preislerová Z., Řezníčková M., Roleček J., Sádlo J., Šumberová K., Vítková M. & Zelený D. (2020) *CzechVeg-ESy: Expert system for automatic classification of vegetation plots from the Czech Republic*. Zenodo. <https://doi.org/10.5281/zenodo.3605561>

**Taxon name following the Czech Expert system (CZECHVEG\_ESY\_NAME\_SUBSP).** This is the same as above but uses “subsp.” and does not add “species” to taxa determined at the genus level.

**Broad plant group (PLANT\_GROUP).** Each plant is assigned to one of the following categories for easier filtering and analysis of the dataset:

- vascular
- bryophyte
- lichen
- alga
- cyanobacterium

**Vegetation layer (LAYER).** If available in the original data, information on the vegetation layer where the species was recorded is provided:

- 0 no layer (no layer information is available, or records from different layers were merged)
- 1 tree layer (high; also used as the tree layer if not divided into sub-layers)
- 2 tree layer (middle)
- 3 tree layer (low)
- 4 shrub layer (high; also used as the shrub layer if not divided into sub-layers)
- 5 shrub layer (low)
- 6 herb layer
- 7 juvenile (used for young trees and shrubs in the herb layer)
- 8 seedling
- 9 moss layer (includes bryophytes and lichens)

**Cover-abundance scale (COVER\_SCALE).** The cover-abundance scale is indicated for each plot observation and specifies how the different cover codes assigned to each taxon should be translated to percentages. For details, see the description of individual cover-abundance scales in the header data below.

**Cover-abundance grade for each species record (COVER\_CODE).** Species cover-abundance, estimated as grades at the specified scale. For more details, see the description of individual cover-abundance scales in the header data below.

**Percentage cover (COVER\_PERC).** Cover-abundances are converted to percentage covers to enable direct comparisons of abundances estimated at different scales. In this approach, cover-abundance is considered as a percentage of the total plot area. The conversion is done using a [conversion table](#) and includes both the conversions suggested in EVA exports, using only integers, based on Turboveg 2 standards (COVER\_PERC\_EVA), and an adjusted version that includes decimal numbers (COVER\_PERC). The latter version converts Braun-Blanquet scale values r and + to 0.1% and 0.5%, respectively, which is more realistic than the conversion to the integer values of 1% and 2% used in the former version. Note that the covers of individual species may overlap in space, and the total sum often exceeds 100%. Handling cover overlaps when merging species covers is explained in our [tutorial](#).

## 2 Header data

So-called “header data” contain information on vegetation plots and plot observations other than species and their cover-abundances. They include details of sampling methods, location, vegetation structure, vegetation type, and environmental information.

In addition to basic variables, the Czech Vegetation Database contains variables with outdated content (replaced by new variables with up-to-date content), variables that are rarely or never used, and one variable with non-unified content. Individual private databases incorporated into the Czech Vegetation Database may also have contained other variables specific to those databases.

### 2.1 Basic header-data variables

**Relevé number (RELEVÉ\_NR).** Unique numerical code of the vegetation-plot observation (relevé) in the database.

**Country code (COUNTRY).** Two-letter country code according to ISO 3166-1 alpha-2. All plots within the territory of the Czech Republic have the value “CZ”.

**Biblioreference (REFERENCE).** Bibliographic reference for vegetation-plot observations published in a citable source (article, book, student thesis, research report). The dataset in the Turboveg 2 format (DBF) includes a reference numeric code that can be linked to references via a [crosswalk table](#). The CSV format also includes the full reference in the fields

**REFERENCE\_AUTHOR**, **REFERENCE\_YEAR**,

**REFERENCE\_TITLE** and **REFERENCE\_JOURNAL**. A complete database of biblioreferences is available here: <https://www.sci.muni.cz/botany/tvref>.

**Number of table in publication (TABLE\_NR).** Number of the table in which the vegetation-plot observation was published in the cited source. If the plot observatoin was published separately rather than included in a table, the page number is given after the abbreviation “p.” (e.g., “p. 15”).

**Number of relevé in table (NR\_IN\_TAB).** Number of the vegetation-plot observation in the cited source.

**Cover-abundance scale (COVER\_SCALE).** Species cover was recorded in vegetation plots either as percentages or using various scales derived from percentage cover intervals. Some scales are supplemented for species with low cover by grades indicating the number of individuals. The database contains both general scales used by multiple researchers and purpose-specific scales. In the database, grades of all scales are converted to the midpoint of the cover percentage interval they represent, as used in Turboveg 2. The dataset in the Turboveg 2 format (DBF) contains the scale numeric code. The CSV format also includes the full name of the scale in the

**COVER\_SCALE\_NAME** field. General scales and their conversions to percentages are as follows. Unused codes refer to purpose-specific scales (all scales can be found in the [crosswalk table](#)):

- 00 – Percentage (%)
- 01 – Braun-Blanquet (old), with seven grades: r = 1%, + = 2%, 1 = 3%, 2 = 13%, 3 = 38%, 4 = 63%, 5 = 88% (note that the COVER\_PERC field in the species data file also contains an alternative, more realistic conversion using decimal numbers: 0.1%, 0.5%, 3%, 13%, 37.5%, 62.5%, 87.5%)
- 02 – Braun-Blanquet (new), with nine grades: r = 1%, + = 2%, 1 = 3%, 2m = 4%, 2a = 8%, 2b = 18%, 3 = 38%, 4 = 63%, 5 = 88% (note that the COVER\_PERC field in the species data file also contains an alternative, more realistic conversion using decimal numbers: 0.1%, 0.5%, 3%, 4%, 8%, 18%, 37.5%, 62.5%, 87.5%)
- 03 – Londo: r1 = 1%, p1 = 1%, a1 = 1%, m1 = 1%, r2 = 2%, p2 = 2%, a2 = 2%, m2 = 2%, r4 = 4%, p4 = 4%, a4 = 4%, m4 = 4%, 1- = 7%, 1 = 10%, 1+ = 12%, 2- = 17%, 2 = 20%, 2+ = 22%, 3- = 27%, 3 = 30%, 3+ = 32%, 4- = 37%, 4 = 40%, 4+ = 42%, 5- = 47%, 5 = 50%, 5+ = 52%, 6- = 57%, 6 = 60%, 6+ = 62%, 7- = 67%, 7 = 70%, 7+ = 72%, 8- = 77%, 8 = 80%, 8+ = 82%, 9- = 87%, 9 = 90%, 9+ = 92%, 10 = 97%
- 04 – Presence/Absence: x, 1
- 05 – Ordinal scale (1–9), corresponding to the “Braun-Blanquet (new)” scale, but with mixed character-numeric symbols replaced by integers: 1 = 1%, 2 = 2%, 3 = 3%, 4 = 4%, 5 = 8%, 6 = 18%, 7 = 38%, 8 = 68%, 9 = 88%
- 06 – Barkman, Doing & Segal: r = 1%, +r = 1%, +p = 1%, +a = 1%, +b = 2%, 1p = 1%, 1a = 2%, 1b = 3%, 2m = 4%, 2a = 8%, 2b = 18%, 3a = 31%, 3b = 43%, 4a = 56%, 4b = 68%, 5a = 81%, 5b = 93%
- 07 – Doing: r = 1%, p = 1%, a = 2%, m = 4%, 1 = 10%, 2 = 20%, 3 = 30%, 4 = 40%, 5 = 50%, 6 = 60%, 7 = 70%, 8 = 80%, 9 = 90%, 10 = 97%
- 08 – Domin: + = 1%, 1 = 2%, 2 = 3%, 3 = 4%, 4 = 13%, 5 = 23%, 6 = 29%, 7 = 42%, 8 = 63%, 9 = 88%, 10 = 99%
- 10 – Zlatník: - = 1%, + = 2%, 1 = 3%, -2 = 10%, +2 = 20%, -3 = 31%, +3 = 44%, -4 = 56%, +4 = 69%, -5 = 81%, +5 = 94%
- 14 – Domin-Hadač: + = 0.5%, 1 = 1%, 2 = 2%, 3 = 4%, 4 = 10%, 5 = 20%, 6 = 29%, 7 = 42%, 8 = 63%, 9 = 85%, 10 = 98%
- 15 – Percentual (r, +): r = 0.1%, + = 0.5%, and integer values corresponding to percentages

- 18 – Hult-Sernader-DuRietz:  $r = 0.1\%$ ,  $+ = 0.5\%$ ,  $1 = 3.5\%$ ,  $2 = 8.3\%$ ,  $3 = 18.3\%$ ,  $4 = 37.5\%$ ,  $5 = 62.5\%$ ,  $6 = 87.5\%$

**Author code (AUTHOR).** Author of the vegetation-plot observation. The dataset in Turboveg 2 format (DBF) contains the author's numeric code. The CSV format also includes the full author name in the field **AUTHOR\_NAME**. Translation of codes to author names is available in the [crosswalk table](#).

**Date (DATE).** Date of vegetation-plot sampling, recorded in the format YYYYMMDD. The CSV format also includes **YEAR**.

**Syntaxon (SYNTAXON).** Vegetation unit (syntaxon) to which the vegetation-plot observation was subjectively assigned by its author or a database administrator. The dataset in Turboveg 2 format (DBF) contains the syntaxon code. The CSV format also includes the full syntaxon name in the field **SYNTAXON\_NAME**. Translation of codes to names is available in the [crosswalk table](#). The vegetation classification follows Chytrý (2007–2013).

- Chytrý M. (ed.) (2007–2013) *Vegetace České republiky 1–4. Vegetation of the Czech Republic 1–4*. Academia, Praha. <https://botzool.cz/vegsci/vegetationCR>

**CzechVeg-ESy code (ESY\_CODE).** Code of the vegetation unit according to the Vegetation of the Czech Republic (Chytrý 2007–2013), assigned based on vegetation classification by the expert system CzechVeg-Esy, version v1-2020-01-12 (<https://doi.org/10.5281/zenodo.3605562>). This field is empty for plot observations that were not unambiguously classified by the expert system.

**CzechVeg-ESy name (ESY\_NAME).** Name of the vegetation unit according to the Vegetation of the Czech Republic (Chytrý 2007–2013), assigned based on vegetation classification by the expert system CzechVeg-Esy, version v1-2020-01-12 (<https://doi.org/10.5281/zenodo.3605562>). This field is empty for plot observations that were not unambiguously classified by the expert system.

**Syntaxon code (ESY\_K\_CODE).** Code of the vegetation unit according to the expert system (taken from the CzechVeg-ESy code variable), supplemented by subjective classification from the Syntaxon field for plot observations not classified by the expert system.

**Syntaxon name (ESY\_K\_NAME).** Name of the vegetation unit according to the expert system (taken from the CzechVeg-ESy code variable), supplemented by subjective classification from the Syntaxon field for plot observations not classified by the expert system.

**Plot area (SURF\_AREA).** Area of the vegetation plot in m<sup>2</sup>. A value of -1 in the Turboveg files represents missing data.

**Altitude (ALTITUDE).** Altitude of the vegetation plot in meters. A value of -1 in the Turboveg files represents missing data.

**Aspect (EXPOSITION).** Slope aspect in azimuth degrees from 1 to 360°.

**Slope (INCLINATIO).** Slope inclination in degrees. A value of -1 in the Turboveg files represents missing data.

**Cover total (COV\_TOTAL).** Total percentage cover of all vegetation layers. A value of -1 in the Turboveg files represents missing data.

**Cover tree layer (COV\_TREES).** Percentage cover of the tree layer. A value of -1 in the Turboveg files represents missing data.

**Cover shrub layer (COV\_SHRUBS).** Percentage cover of the shrub layer. A value of -1 in the Turboveg files represents missing data.

**Cover herb layer (COV\_HERBS).** Percentage cover of the herb layer. A value of -1 in the Turboveg files represents missing data.

**Cover moss layer (COV\_MOSSES).** Percentage cover of the moss layer, including bryophytes and lichens. A value of -1 in the Turboveg files represents missing data.

**Height (high) tree layer (TREE\_HIGH).** Height of the tree layer in meters. If the heights of the higher and lower sublayers were estimated separately, this field contains the higher sublayer's height.

**Height (high) shrub layer (SHRUB\_HIGH).** Height of the shrub layer in meters. If the heights of the higher and lower sublayers were estimated separately, this field contains the higher sublayer's height.

**Height (high) herb layer (HERB\_HIGH).** Height of the herb layer in centimeters. If the heights of the higher and lower sublayers were estimated separately, this field contains the higher sublayer's height.

**Maximum height herb layer (HERB\_MAX).** Height of the tallest herbs exceeding the general herb layer, in centimeters.

**Mosses identified (MOSS\_IDENT).** Information on whether bryophytes were recorded and identified in the vegetation plot.

**Lichens identified (LICH\_IDENT).** Information on whether lichens were recorded and identified in the vegetation plot.

**Coordinate source (COORD\_CODE).** Method used to determine the coordinates of the vegetation plot, either in the field using a Global Navigation Satellite System or later using electronic or paper maps. The dataset in Turboveg 2 format (DBF) contains the coordinate numeric code as listed below. The CSV format also includes full information in the fields **COORD\_SOURCE**, **COORD\_SYST**, and **SCALE**.

- 01 – Global Navigation Satellite System (e.g., GPS or Galileo)
- 02 – Geobáze 100 (electronic map 1 : 100,000)
- 03 – Geobáze 50 (electronic map 1 : 50,000)
- 04 – Vojenské mapy 50 (paper maps 1 : 50,000)
- 05 – ZM 5 (paper maps 1 : 5,000)
- 06 – ZM 10 (paper maps 1 : 10,000)
- 07 – ZM 25 (paper maps 1 : 25,000)
- 08 – ZM 50 (paper maps 1 : 50,000)
- 09 – GIS (electronic map)
- 10 – Military maps 25 (paper maps 1 : 25,000)
- 11 – www.mapy.cz (electronic map or orthophoto)
- 12 – www.atlas.cz (electronic map or orthophoto)
- 13 – Google Earth (orthophoto)

**Locality (LOCALITY).** Verbal description of the locality of the vegetation plot.

**Latitude (LATITUDE).** Latitude in the format DDMMSS.SS in the WGS-84 system (e.g., 492215.12 corresponds to 49° 22' 15.12'' N).

**Longitude (LONGITUDE).** Longitude in the format DDMMSS.SS in the WGS-84 system (e.g., 170322.55 corresponds to 17° 3' 22.55'' E).

**Decimal-degree latitude (DEG\_LAT).** Latitude in decimal degrees (DD.DDDDDD) in the WGS-84 system.

**Decimal-degree longitude (DEG\_LON).** Longitude in decimal degrees (DD.DDDDDD) in the WGS-84 system.

**Precision (PRECISION).** Estimated uncertainty of the given geographic coordinates in meters.

**Field number (FIELD\_NR).** Field number or code of the vegetation-plot observation used by the author.

**Habitat (HABITAT).** Textual description of the abiotic or biotic environment in the vegetation-plot observation and its immediate surroundings.

**Geology (GEOLOGY).** Textual description of the geological bedrock in the vegetation plot.

**Soil (SOIL).** Textual description of the soil in the vegetation plot.

## 2.2 Header-data variables with outdated content replaced by new variables

**Syntaxon old (SYNTAX\_OLD).** Vegetation unit (syntaxon) to which the vegetation-plot observation was subjectively assigned by its author or database administrator according to the outdated classification system in the publication:

- Moravec J., Balátová-Tuláčková E., Blažková D., Hadač E., Hejný S., Husák Š., Jeník J., Kolbek J., Krahulec F., Kropáč Z., Neuhausl R., Rybníček K., Řehořek V. & Vicherek J. (1995): Rostlinná společenstva České republiky a jejich ohrožení [Plant communities of the Czech Republic and their threats]. Ed. 2. *Severočeskou přírodou*, Příloha 1995: 1–206.

The dataset in Turboveg 2 format (DBF) contains the syntaxon code. The CSV format also includes the full syntaxon name in the field **SYNTAXON\_OLD\_NAME**. The translation is available in the [crosswalk table](#).

### 2.3 Rarely or never used header-data variables

**UTM grid system code (UTM)**. Code of geographic position in the Universal Transverse Mercator (UTM) system.

**Cover lichen layer (COV\_LICHEN)**. Percentage cover of lichens. Usually left empty, as lichens are included in the moss layer.

**Cover algae layer (COV\_ALGAE)**. Percentage cover of algae. A value of -1 in the Turboveg files represents missing data.

**Cover litter layer (COV\_LITTER)**. Percentage cover of litter (dead plant material). A value of -1 in the Turboveg files represents missing data.

**Cover open water (COV\_WATER)**. Percentage cover of open water surface. A value of -1 in the Turboveg files represents missing data.

**Cover bare rock (COV\_ROCK)**. Percentage cover of rock outcrops and stones. A value of -1 in the Turboveg files represents missing data.

**Height low tree layer (TREE\_LOW)**. Height of the lower tree sublayer in meters.

**Height low shrub layer (SHRUB\_LOW)**. Height of the lower shrub sublayer in meters.

**Height low herb layer (HERB\_LOW)**. Height of the lower herb sublayer in centimeters.

**Maximum height cryptogam layer (CRYPT\_HIGH)**. Height of the tallest bryophytes or lichens in centimeters.

### 2.4 Inconsistently used header data variables with non-unified content

**Project code (PROJECT)**. Project codes were entered by some data contributors for their internal purposes and were not unified across the entire Czech Vegetation Database. Consequently, the same project code used by different data contributors can refer to different projects.

### 2.5 Environmental and biogeographical strata for data resampling

The following variables contain environmental or biogeographical strata (landscape types or regions) that can be used for stratified resampling of the Czech Vegetation Database.

**Env\_strata (ENV\_STRATA)**. Environmental strata used in the resampling procedure to create [CzechVeg-OpenStrat](#). For details, see the tutorial on [Data stratification](#).

**SKALICKY1988**. Phytogeographical regions according to Skalický (1988).

- Skalický V. (1988) Regionálně fytogeografické členění [Regional phytogeographic division]. In: Hejný S., Slavík B., Chrtek J., Tomšovic P. & Kovanda M. (eds), *Květena České socialistické republiky [Flora of the Czech Socialist Republic]* 1: 103–121, Academia, Praha.

**CULEK1996**. Biogeographical regions according to Culek (1996).

- Culek M. (ed.) (1996) *Biogeografické členění České republiky [Biogeographical division of the Czech Republic]*. Enigma, Praha.

**CULEK2005**. Landscape types (so-called biochores) delineated by Culek (2005) based on a combination of altitudinal vegetation zones, relief types, soil types, and their moisture.

- Culek M. (ed.) (2005) *Biogeografické členění České republiky II. díl. [Biogeographical division of the Czech Republic Part II.]* AOPK ČR, Praha.

**CHUMAN\_ROMPORTL2010**. Landscape types delineated by Chuman & Romportl (2010) using a modified TWINSpan algorithm and data on elevation, aspect, slope, soils, reconstructed natural vegetation, mean annual temperature, mean annual precipitation, and land cover for grid cells of 2 × 2 km.

- Chuman T. & Romportl D. (2010) Multivariate classification analysis of cultural landscapes: An example from the Czech Republic. *Landscape and Urban Planning*, 98(3–4), 200–209.

**DIVISEK2014\_UNCONST**. Landscape types delineated by Divíšek et al. (2014) using unconstrained hierarchical clustering of data on the distribution of natural habitats in grid cells of 5' longitude × 3' latitude.

- Divíšek J., Chytrý M., Grulich V. & Poláková L. (2014) Landscape classification of the Czech Republic based on the distribution of natural habitats. *Preslia*, 86(3), 209–231.

**DIVISEK2014\_CONST.** Regions delineated by Divíšek et al. (2014) using spatially constrained hierarchical clustering of data on the distribution of natural habitats in grid cells of 5' longitude × 3' latitude.

- Divíšek J., Chytrý M., Grulich, V. & Poláková L. (2014) Landscape classification of the Czech Republic based on the distribution of natural habitats. *Preslia*, 86(3), 209–231.

### **3 Remarks**

The **REMARKS** field contains non-standardized textual comments, such as details of taxonomic interpretation for some taxa reported in the plot, identification uncertainty for specific taxa, and sampling methods. There is one REMARKS field for each vegetation-plot.