User’s guide for program CoDeK (Cocktail Determination Key)

Authors: David Zelený (zeleny.david@gmail.com) & Ching-Feng Li (chingfeng.li@gmail.com)

Related website: http://www.davidzeleny.net/wiki/doku.php/codek:start

Note: This user's guide is updated version of Appendix S2 in Li et al. (2013).

Introduction

CoDeK is a software application, based on R program (www.r-project.org), which allows automatic classification of vegetation samples into vegetation types defined by Cocktail determination key. To run this application, you need the table of one or more vegetation plots (samples, relevés), file with the Cocktail determination key and optionally also species checklist, which allows checking for correct use of species nomenclature in both files of vegetation plots and Cocktail determination key.

Software Requirements

CoDeK is basically an R script, which must be run in R program. In any case, you need to install R program on your computer; R can be downloaded for free at www.r-project.org (click here to download the most recent version of R).

The CoDeK application has been tested in R version 2.15, however, it should run also in newer R version (and probably also in older ones).

CoDeK can be launched in two ways:

1. using batch file CoDeK.bat, or
2. from JUICE program (Tichý 2002); JUICE is a program for editing and analysis of vegetation data and it can be downloaded for free at http://www.sci.muni.cz/botany/juice/. JUICE version 7.0.67 has been tested. Note, that even if you launch CoDeK from JUICE, you still need to have R program installed on your computer.

Operation system requirements

Windows XP, Vista, 7 (other OS which can run R program can be probably also used, but the application has to be launched manually by copying the script in CoDeK_vX.X.r into R console). If using also JUICE, consult the http://www.sci.muni.cz/botany/juice/ website for system requirements of JUICE program.

Download

Option 1: Running CoDeK directly using batch file CoDeK.bat

Unzip the supplementary materials in *CoDeK_vX.X.zip* into your computer; directory *CoDeK* will be automatically created. This directory contains the following files:

1. CoDeK program: *CoDeK_bat*, *CoDeK_vX.X.r*;
2. Cocktail determination key for Taiwan forest vegetation (Li et al. 2013): *Det_key_Taiwan_forest_vegetation.txt*;
3. Checklist for Taiwan flora: *Checklist_Taiwan_flora_20111214.txt*;
4. Example vegetation data in txt format (Li 1997): *example_data.txt*;
5. Example vegetation data in JUICE format (Li 1997): *example_data_JUICE.wct*, *example_data_JUICE.exp*, *example_data_JUICE.str*.

Important: you need to change the file name *CoDeK_bat* into *CoDeK.bat* (i.e., replace "_" by ".", using file manager).

Example vegetation data represent succession stages from secondary to primary subtropical lowland forest (Li 1997). Example data in txt format and JUICE format are identical (JUICE format additionally contains header data such as geographical coordinates and few environmental factors).

Option 2: Running CoDeK from JUICE program

You need to download only the *CoDeK_vX.X.r* file - this is part of the archive *CoDeK_vX.X.zip*, or can be downloaded separately.

Program settings

Option 1: Running CoDeK directly using batch file CoDeK.bat

If you will run CoDeK directly using the batch file *CoDeK.bat*, to set it up you have two options:

1. keep the files *CoDeK.bat* and *CoDeK_vX.X.r* in the directory CoDeK outside of the Program Files folder (e.g. in your Documents folder), and to change the directory name in the file *CoDeK.bat*. This can be done in the following way:
    1. open the file *CoDeK.bat* using notepad (in Windows, right click on the file *CoDeK.bat* and select *Edit*);
    2. add the directory name (surrounded by quotation marks) in front of the script (Fig. 1). If you have e.g. R version 2.12.2, it will be probably installed in the directory c:\Program Files\R\R-2.12.2\. Ad this directory in a format `c:\Program Files\R\R-2.12.2\bin\` (don't forget to add `bin\` at the end of the directory, as the *R.exe* file is inside the folder bin). Note the differences between a forward slash (`/`) and a backslash (`\`) – the latter need to be used!
    3. save the edit and close notepad.

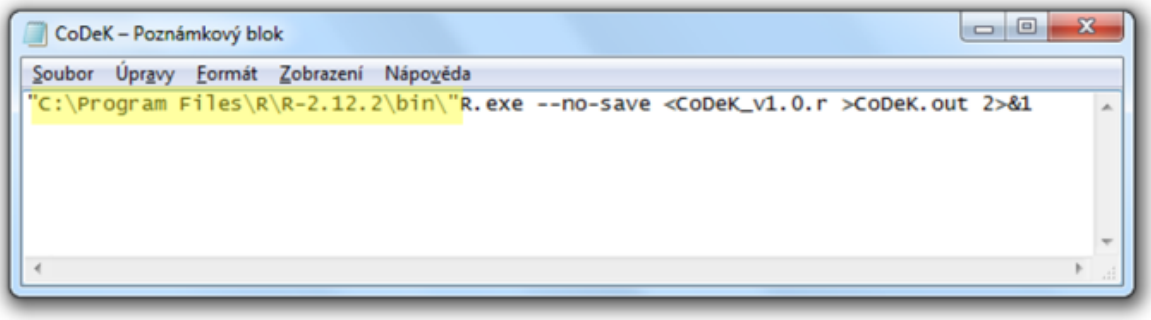


Figure 1

2. or, save the files *CoDeK.bat* and *CoDeK_vX.X.r* into the R directory, where is located the file *R.exe*. In standard situation, this should be somewhere like "c:\Program Files\R\R-2.12.2\bin" (R-2.12.2 is the version of R; if you use different version, the numbers will be different). Note, however, that this may not work in newer Windows versions (8, 10) with restricted permission for using Program Files directory - in

that case, previous option (to keep files out of Program Files folder) is preferred.

Option 2: Running CoDeK from JUICE program

To launch the CoDeK application from JUICE program, you need to install the JUICE program first (JUICE can be operated under Windows OS (optionally other OS using Windows emulator such as Wine for Linux) and it can be downloaded at http://www.sci.muni.cz/botany/juice/?idm=3). Consult JUICE website and JUICE manual if you encounter problems while installing the program.

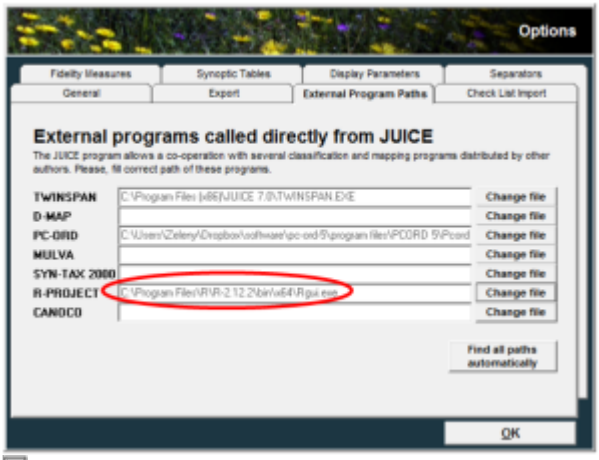


Figure 2

In JUICE, you need to setup the path to R program. In JUICE menu, go to *File* > *Options*, and choose the folder *External Program Paths* (Fig. 2). At the line of *R Project*, click on *Change Path button* and navigate to the directory of *Rgui.exe* file (if you installed R version 2.12.2, you will probably find it in the directory *c:\Program Files\R\R-2.12.2\bin*, either in subfolder *i386* (if you are using 32 bit version of R) or *x64* (if you use 64 bit version). If you use R version older than 2.9.0, you will find *Rgui.exe* file directly in the folder *bin*, which does not contain subfolder mentioned above. Click on *OK* button to leave the *Options*. Note: in JUICE options, you MUST specify the file *Rgui.exe* file, not *Rterm.exe* or *R.exe*!

Required format of vegetation data file

Option 1: Running CoDeK directly using batch file CoDeK.bat

The table of vegetation plots contains percentage data about species relative IVI or cover; species are in rows, plots are in columns. First row contains sample names (without spaces), first column contains species names. Species names must be identical as those listed in Checklist, otherwise the Determination key may not identify the plots appropriately. CoDeK offers simple function to check compatibility between the species names in the table of vegetation plots and the Checklist. All plots stored in the file will be used for determination of vegetation types; if you wish to determine only part of the data, remove them from this file. Vegetation data should be stored in a plain text format, with cells separated by tabulator. This format can be created e.g. using Excel spreadsheet program (use the function *Save as* and choose *Text (separated by tabulators)*). For details of the format, see *example_data.txt* file directly (it can be opened in Notepad program or imported into Excel).

Option 2: Running CoDeK from JUICE program

To run CoDeK from JUICE, you need to import vegetation data to JUICE program first. Spreadsheet format data (e.g. from Excel) can be imported into JUICE directly; however, note that you need to add manually three rows at the beginning of the file (for details about this, see JUICE manual, part 1, section 1.4.4 Spreadsheet Format

(Microsoft® Excel® Table), which is available at JUICE website http://www.sci.muni.cz/botany/juice/ at the section Manuals). Optionally, vegetation data can be imported into JUICE from Turboveg program (Hennekens & Schaminée 2001; see JUICE manual for more details).

Running the program using example datasets

Option 1: Running CoDeK directly using batch file CoDeK.bat

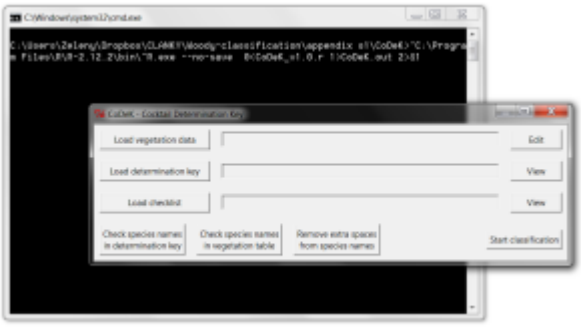
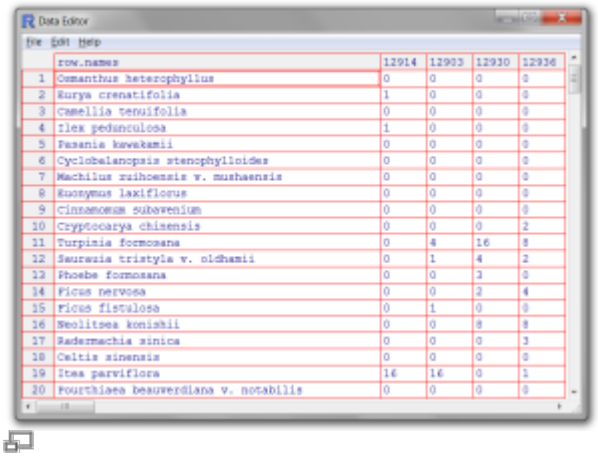


Figure 3

1. Launch the program by executing *CoDeK.bat file*. The Command Prompt window will open, and after that will pop up the graphical user interface of the CoDeK program (Fig. 3). If the Command Prompt window disappears immediately and the user interface does not occur, you have probably set up wrong directory for the *R.exe* program and the *CoDeK.bat* application cannot find it. Check it again following the steps described in the section Program settings, Option 1.



	row.names	12914	12933	12930	12936
1	Cosmanthus heterophyllus	0	0	0	0
2	Eurya crenatifolia	1	0	0	0
3	Camellia tenuifolia	0	0	0	0
4	Ilex pedunculosa	1	0	0	0
5	Pasania kawakamii	0	0	0	0
6	Cyclobalanopsis stenophylloides	0	0	0	0
7	Machilus zuihoensis v. mushaensis	0	0	0	0
8	Euonymus laxiflorus	0	0	0	0
9	Cinnamomum subavenium	0	0	0	0
10	Cryptocarya chinensis	0	0	0	2
11	Turpinia formosana	0	4	16	8
12	Saurauia tristyla v. oldhamii	0	1	4	2
13	Phoebe formosana	0	0	1	0
14	Ficus nervosa	0	0	2	4
15	Ficus fistulosa	0	1	0	0
16	Neolitsea konishii	0	0	8	8
17	Radermachia sinica	0	0	0	3
18	Celtis sinensis	0	0	0	0
19	Itea parviflora	16	16	0	1
20	Fourthiaea beauverdiana v. notabilis	0	0	0	0

Figure 4

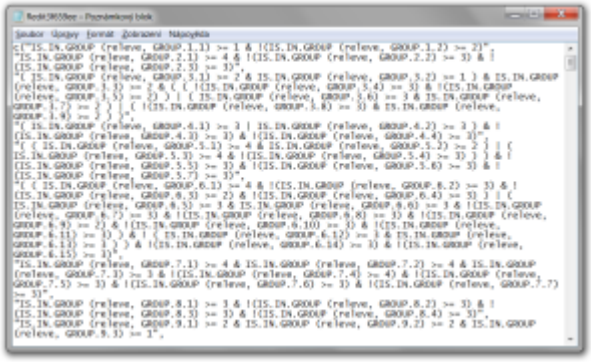


Figure 5

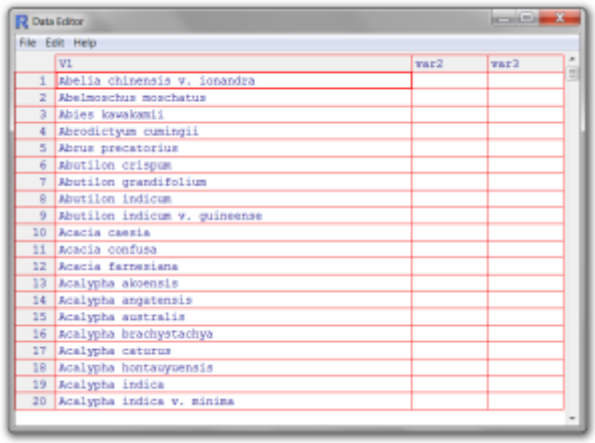


Figure 6

2. Load appropriate files into CoDeK. Click on *Load vegetation data* and select *example_data.txt* file, click on *Load determination key* and select *Det_key_Taiwan_forest_vegetation.txt*, and finally click on *Load checklist* and select *Checklist_Taiwan_flora_20111214.txt*. The buttons *Edit* or *View* are intended for viewing the files in order to check if the files were loaded correctly (see Fig. 4 for vegetation data, Fig. 5 for determination key and Fig. 6 for checklist). In case of vegetation table, the data can be also edited (changes will be saved into the data), in case of determination key and checklist the files can be only viewed. Note: changes in the vegetation data made in CoDeK application will not be saved into the original file (e.g. *example_data.txt*) – they will influence only currently loaded data and their determination.
3. Check the nomenclature of species names used in the determination key (click on *Check species names in determination key*) and vegetation data (*Check species names in vegetation table*). If there is a wrong name, it will be copied into clipboard – to see it, open some text editor (Word, notepad) and past the content of the clipboard (CTRL+V). Optionally, if the species names in the vegetation table contain extra spaces after the name (e.g. „Pasania kawakamii " instead of „Pasania kawakamii"), you can remove these spaces by clicking on button *Remove extra spaces from species names*. Note: only spaces at the end of the species names are removed, not the spaces between genus and species name; only spaces in the species of vegetation data will be considered, not the spaces in species of determination key or checklist. You can see the changes in vegetation data by clicking *Edit* button at the top right corner.
4. If all the species names have been treated, click on *Start classification* to run the determination key. Progress bar will appear, and after the determination is finished, the button Open result file will appear (Fig. 7).



Figure 7

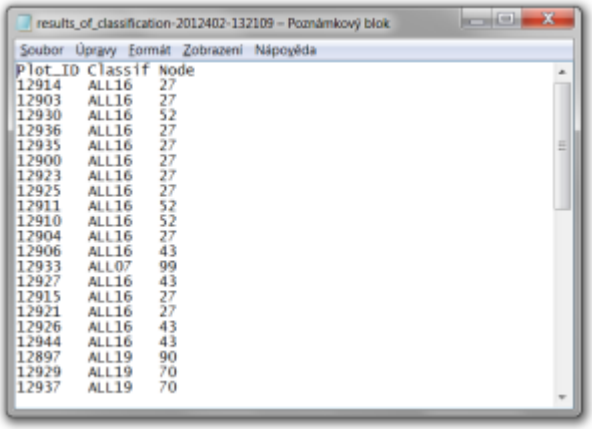


Figure 8

1. Click *Open result file* button. The result file will open in notepad or other associated program (Fig. 8). It contains three columns:
    1. Plot_ID – codes of vegetation plots;
    2. Classif – code of resulting assignment into vegetation type;
    3. Node – the row in the key, which lead to the assignment of the vegetation plot (for reference, to know which formula in the determination key is responsible for assignment of particular plot).

Option 2: Running CoDeK from JUICE program

1. Open JUICE program and open the file with example data (in JUICE menu, go to *File* > *Open* and find the file *example_data_JUICE.wct*). Vegetation table in JUICE looks like those on Fig. 9.

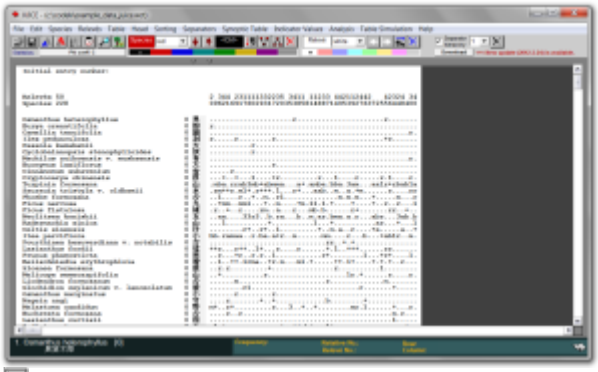


Figure 9

2. In JUICE, launch the export into R (so called JUICE-R function) using the keyboard shortcut CTRL+W (or, in JUICE menu, go to *Analyses* > *JUICE-R functions*). The export wizard will launch (Fig. 10).



Figure 10

3. In the JUICE-R wizard, you need to select two things:
    1. color of plots used in analysis (option *Relevés Used in Analysis*, yellow marking in Fig. 10) – if you have not specify any color in JUICE, choose option All;
    2. Append R script (red marking in Fig. 10) – choose *CoDeK_vX.X.r* file, which contains the R script running the CoDeK application.
    3. Click on Run button. The wizard will close, and data will be exported from JUICE. In a while, graphical user interface identical to the one on Fig. 3 should appear.

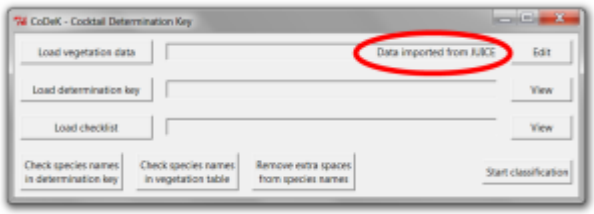


Figure 11

4. Vegetation data are already imported from JUICE (Fig. 11). Load the determination key and the checklist. Also, you can check species names in determination key and vegetation table. There should not be any extra spaces in species names in vegetation data, as these are automatically removed while the data are imported from JUICE.
5. Click on *Start classification* button. After running determination key, *Open result file* button will appear (Fig. 7) and clicking it will open notepad with the results (Fig. 8).
6. If you wish to import the result of determination back to JUICE, you can copy the table in the result file (Fig. 8) onto clipboard (CTRL+A will select the whole table, CTRL+C will copy it to the clipboard) and go to JUICE program menu *Edit > Paste Clipboard To White Short Header* (if you have changed the color, you need to paste it as a selected color – to select appropriate color, in JUICE menu click the color under the heading Relevé). The codes of vegetation types will be copied into the short header in JUICE (Fig. 12).

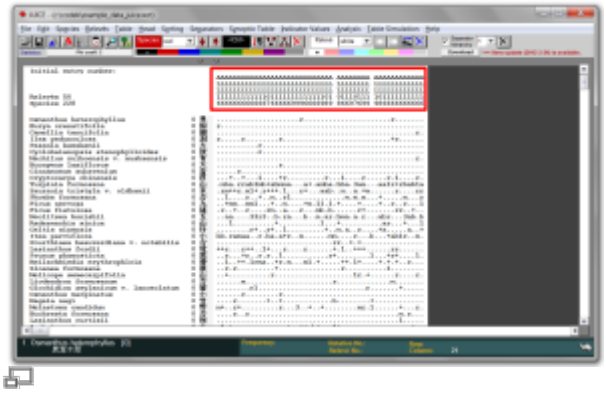


Figure 12

Result files

After running the determination key, the **.txt* file with the results of classification is created, either in CoDeK directory or the R program directory (if you use *CoDeK.bat* option, it will be at the same folder as *CoDeK.bat* file; if you use JUICE to run the key, the file will be in R program directory). The name of the file is *results_of_classification-YYYYMMDD-hhmmss.txt*, where *YYYYMMDD-hhmmss* is a time stamp (e.g. 20120402-104338, which means 2nd April 2012 at 10:43:38). Each run of determination key will produce new result file with relevant time stamp. The file can be opened in notepad, Excel or any other editing program.

Error handling

After running CoDeK application, new file *CoDeK.out* is created (either in CoDeK directory or R program directory). This file contains technical report about the process of calculation. If error occurs, this file can serve as a reference for searching the potential bugs (follow Error or Warning messages in the file). If error occurs, first check the website http://www.davidzeleny.net/wiki/doku.php/codek:start for availability of updates of the CoDeK program or trouble shooting. In case you still experience troubles in using the CoDeK program, do not hesitate to contact us (zeleny.david@gmail.com, chingfeng.li@gmail.com). Providing the *CoDeK.out* file (in case you run CoDeK directly) or *basic.out* file (in case you run CoDeK as JUICE-R function) and optionally also example of your data will greatly improve our ability to determine where the error happened.

References

- Hennekens, S.M. & Schaminée, J.H.J. 2001. TURBOVEG, a comprehensive data base management system for vegetation data. Journal of Vegetation Science 12: 589–591.
- Li, C.-F. 1997. Vegetation Succession of the Machilus-Castanopsis Zone in the Northwestern Region of Taiwan. Master thesis. Department of Forestry, National Taiwan University, Taipei, TW.
- Tichý, L. 2002. JUICE, software for vegetation classification. Journal of Vegetation Science 13: 451–453.