

bcDiscriminator

Version 1.14 (20090510)

Runkel, Marckmann & Schuster GbR

ecoObs

<http://www.ecoobs.com>

Legal

Warranties

The application bcDiscriminator was developed under strict rules, nevertheless errors may be existing with the implementation. Please report such directly to us listing in addition your computer system, OS version and a detailed instruction on how to reproduce the error.

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA

In no event can we be held liable for any direct, indirect, incidental or any consequential damages risen by the use of the software.

System prerequisites

The application bcDiscriminator was developed for Apple computers running 10.4 or newer OS versions. It runs on PPC as well as Intel Macs. We recommend main memory of at least 512 MB. You will also need the statistical tool R (versions 2.6 through 2.7 are compatible) as well as two statistical R packages (randomForest and kernlab).

Contact

Volker Runkel
Runkel, Marckmann & Schuster GbR
ecoObs
Reindelstrasse 2
90402 Nuernberg
Germany

Contents

1. Overview	4
2. Installation	4
2.1. „Binary-Distribution“	4
2.2. „Source-Distribution“	4
2.3. Required third party applications	4
2.3.1. R	4
2.3.2. Required R-packages	5
3. Using bcDiscriminator	6
3.1. Start	6
3.2. The UI	6
3.3. Preferences	6
3.4. Interaction with bcAdmin	7
4. The inner working of bcDiscriminator	7
4.1. Steps in the analysis process	7
4.1.1. Identification and outlier detection	7
4.1.2. Determination of the results	10
4.2 Measurement files (csv) and call parameters	11

1. Overview

bcDiscriminator identifies bat calls using measurements to group and species level by statistical algorithms. The measurements have to be stored in csv files and can be analyzed in single file or batch mode. Results are displayed on screen as well as in a result file (.res) at the same location as the according csv file.

The application upon start loads all necessary R packets as well as the identification objects („training calls“). It expects an existing installation of R and the necessary packets.

2. Installation

2.1. „Binary-Distribution“

Most users download the binary distribution, that means, the version that can be run without compiling it first. To be able to use the binary version you need to follow the instructions listed in chapter 2.3. . Please copy the downloaded application to your /Applications folder for installation.

2.2. „Source-Distribution“

For all of you who are interested in the inner workings of bcDiscriminator or who want to participate in development of the application, we offer the sources for download. You'll need Xcode 3, which can be freely downloaded from Apple's developer homepage.

2.3. Required third party applications

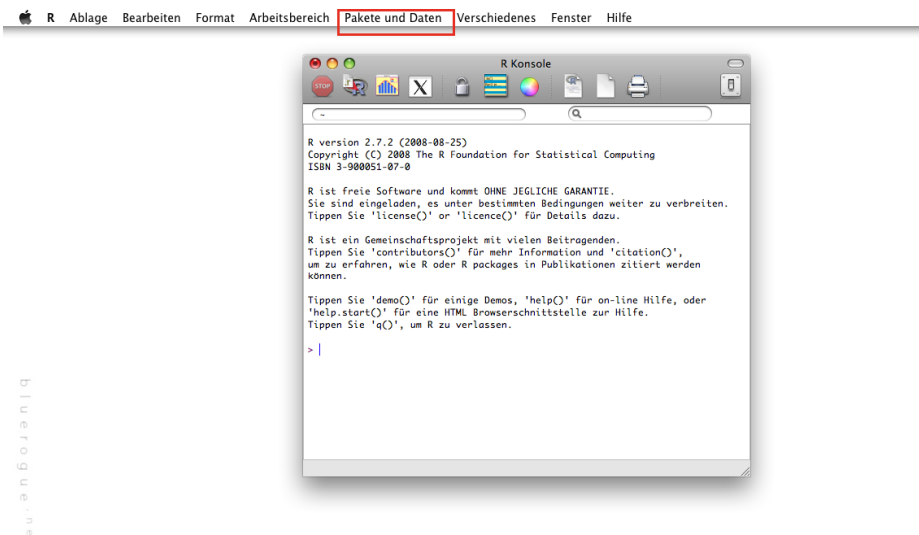
2.3.1. R

The identification of bat calls is possible using algorithms implemented in the free statistical tool R. Thus, you need to install R prior to running bcDiscriminator. If not otherwise stated, currently the versions 2.6 through 2.7 of R are compatible to bcDiscriminator. You can download R at:

[R for Mac OS X](#)

2.3.2. Required R-packages

Within R you will need to install two packages that bcDiscriminator relies on. These two are the packages randomForest and kernlab. To do so you have to start R. The packages can be installed via „Packages & Data“ -> „Package Installer“ . Make sure the first popup button lists „CRAN (binaries)“. After clicking „Get List“ a dialog opens and asks which server to be used. Austria provides to be a fast and stable server and should be chosen. From the list you have to choose randomForest and kernlab and install them via the button „Install selected“. Quit R after the installation is completed successfully. bcDiscriminator is now usable.



3. Using bcDiscriminator

3.1. Start

After starting bcDiscriminator the R environment is started up and all necessary packages are loaded. The startup takes between 4 to 20 seconds depending on your machine specs.

3.2. The UI

In the upper half of the application window an - in the beginning empty - table holds the results of all calls of a single csv file. Left of the table a button labeled „Choose Input File“ resides. This button is used to select csv files from the hard disc. In the lower half a text view prints log messages.

After loading all packages the „Choose Input File“ button can be used to locate and analyze existing csv files. If a single file is selected, the results of each call stay visible in the table. If multiple files were selected all of these are processed and the result display in the table is overwritten with each new file. That way only the last files call results are displayed. If the checkbox „omit result files“ is checked, no .res files are written.

3.3. Preferences

Beginning with version 1.13 the behaviour of bcDiscriminator and result extraction can be influenced.

Decimal separator: to allow using system settings with „.“ as decimal separator we added a preference. Please set this accordingly to your system settings. Otherwise csv files produced with bcAdmin can not be read properly.

Species selection: The criteria for accepting a species and writing it to the result file are now set-able. Standard values are 3 calls and an averaged probability of 0.6 (=60%).

Outlier detection: Optimal outlier detection thresholds are used on a per group/species basis if not set otherwise. You can also choose to use a

fixed outlier threshold for all groups/species, but it is not recommended.

New averaging: A new averaging method was implemented to get an average probability of all calls of the same identification level. The new algorithm takes the number of calls into account. You can choose to use this new method or the old from versions prior 1.13.

3.4. Interaction with bcAdmin

The application bcAdmin creates measurements compatible to bcDiscriminator. after analyzing files with bcAdmin, bcDiscriminator can use the resulting csv files. When all csv files were run through the id process, result files can be imported in bcAdmin again.

4. The inner working of bcDiscriminator

4.1. Steps in the analysis process

4.1.1. Identification and outlier detection

Identification is done for each single call of a sequence in a stepwise process, to be able to detect more than one species (which would be the case of using averages of all calls in a sequence). The possible species or groups that can be detected are shown in figure 1. Currently the genera *Rhinolophus*, *Tadarida* and *Miniopterus* are missing. The species *Myotis dasycneme*, *Myotis emarginatus*, *Myotis cappaccini* as well as *Nyctalus lasiopterus* are missing.

The following abbreviations are used:

Abb.	Species	Abb.	Species
Nnoc	<i>Nyctalus noctula</i>	Mdau	<i>Myotis daubentonii</i>
Nlei	<i>Nyctalus leisleri</i>	Mbec	<i>Myotis bechsteinii</i>
Enil	<i>Eptesicus nilssonii</i>	Ppyg	<i>Pipistrellus pygmaeus</i>
Eser	<i>Eptesicus serotinus</i>	Ppip	<i>Pipistrellus pipistrellus</i>
Vmur	<i>Vespertilio murinus</i>	Pnat	<i>Pipistrellus nathusii</i>

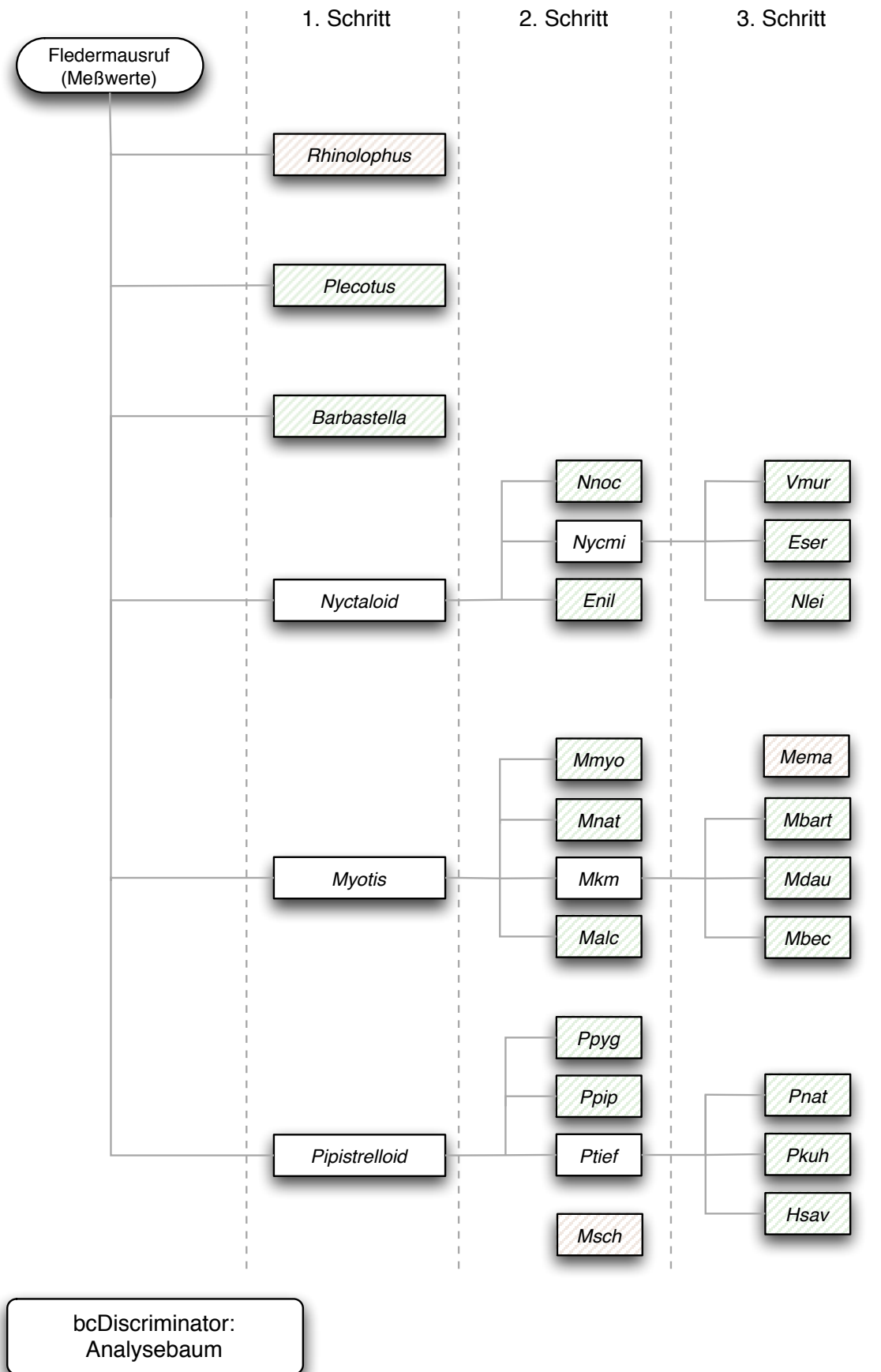


figure 1: groups and analysis steps

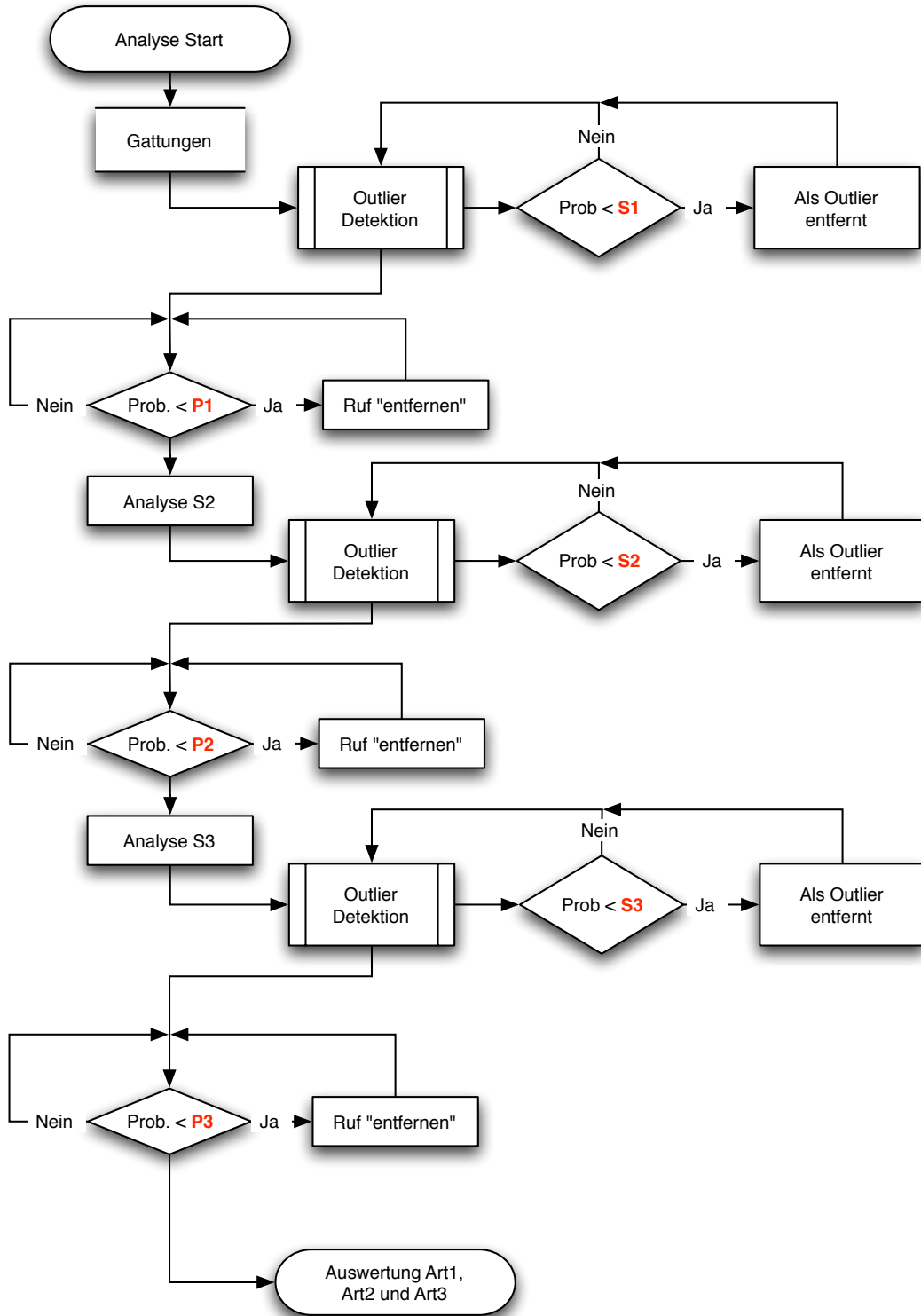


figure 2: decision tree

Abb.	Species	Abb.	Species
Mmyo	<i>Myotis myotis</i>	Pkuh	<i>Pipistrellus kuhlii</i>
Mnat	<i>Myotis nattereri</i>	Hsav	<i>Hypsugo savii</i>
Malc	<i>Myotis alcaethoe</i>		
Mbart	<i>Myotis brandtii/mystacinus</i>		

Analysis is started on genus level always. After identification via random-Forest every call is tested for outlier-ness via svm. If a call is detected to be an outlier, it is not examined further, but the result of the step before is saved. In the next step the result of the previous step is used to determine which identification objects are to be used for further analysis. When a call reaches the end of the tree (a species) or when the probability falls below the threshold, no further steps are taken, but the result for the call is saved. Thus, a maximum of three analysis steps are made (see figure 2 as well).

4.1.2. Determination of the results

The aim of bcDiscriminator is to use the measurements from bcAdmin to identify the recorded species. Since not always only one species is calling in a single recording, bcDiscriminator can extract up to three species. For the reliable extraction of a species, all calls ending in the same group/species in the stepwise analysis (see above) are collected. Their count as well as averaged probabilities are then computed. A certain number of calls (defaults to three) and a certain probability (defaults to 60%) has to be reached that a result is extracted. The applications then sorts the possible results based on call count and probability. Up to three results are extracted from this sorted list. If fewer than three calls were recorded all together, this criterium is ignored and still a species extracted based on probability alone. If no species was valid for extraction, „Spec.“ is returned.

After a successful analysis the results are displayed in the table and a result file is written (if not omitted by purpose). Starting with version 1.14

bcDiscriminator additionally adds the species and probability of each call to the .data file (generated by bcAdmin). bcAdmin can use this information in its call display. If no data file is available or a difference of call counts in data and csv file were found, these single call results are not written out to the data file.

4.2 Measurement files (csv) and call parameters

bcDiscriminator is designed to read csv files with measurements. Appropriate csv files are currently only written by bcAdmin and bcAnalyze. The csv files needs to be formatted as described in the following. Column headers are expected in the first line, the second to last line holds the actual measurements. Each column is separated by a semicolon (;). decimal numbers either have „.“ or „,“ as decimal separator. The column headers are:

Column	Description
Datei	filename, not required
Art	species, not required
Ruf	call number, not required
Dur	duration in ms
Sfreq	start frequency (kHz)
Efreq	end frequency (kHz)
Stime	start of call in file (ms), not required
NMod	description after this table
FMod	description after this table
FRmin	description after this table
Rmin	description after this table
tRmin	description after this table
Rlastms	description after this table

Column	Description
Flastms	description after this table
X10, X11 ... X60	description after this table
X62, X64 ... X148	description after this table

„FRmin“, „Rmin“ and „tRmin“ are calculated from the location within the call with the lowest slope. Computed are the frequency at that location (FRmin), the slope (Rmin) and the distance to the call end (tRmin) measured in ms. Within the last millisecond of the call again the lowest slope is determined and frequency (Flastms) as well as slope (Rlastms) are measured.

The columns X10 to X148 are computed as follows: Along the call each 100µs the frequency is measured. These measurements are combined in a histogram with bin widths of 1 kHz (10 to 60 kHz) and 2 kHz (60 to 150 kHz). A moving average algorithm is additionally used to build the histogram. The values NMod and FMod are directly determined from within the histogram. The first is the count of measurements of the histogram bin with the maximum number and the second its frequency.

From a call of *Pipistrellus pipistrellus* for example the following results:

Feld	Wert		Feld	Wert
Dur	4,5		X49	48
Sfreq	92,361313		X50	31
Efreq	47,414757		X51	23,75
Stime	493,668640		X52	19
NMod	78		X53	16,25
FMod	48		X54	13,75
FRmin	47,598934		X55	11,75
Rmin	0,175629		X56	10,5

Feld	Wert		Feld	Wert
tRmin	-0,90		X57	9,5
Rlastms	0,031929		X58	8
Flastms	47,685329		X59	6
X10-X45	0		X60	8
X46	26,5		X62	6,5
X47	74		X64	5,5
X48	78	