

Case study to determine the required minimum sampling density

Authors: Hendrik Pehlke, Dr. Alexa Wrede, Paul Kloss, Dr. Jennifer Dannheim

(Technical report of the project ANsätze zur Kostenreduzierung bei der ERhebung von Monitoringdaten für Offshore Vorhaben (ANKER), Arbeitspaket BENTHOS; FKZ 0325921)

Last update: 29. May 2020

Citation: Pehlke H, Wrede A, Kloss P, Dannheim J (2020). Case study to determine the required minimum sampling density. Technical Report of the project ANKER FKZ 0325921, pp. 22.



Alfred-Wegener-Institut, Helmholtz-Zentrum
für Polar- und Meeresforschung (AWI)
Am Handelshafen 12
27570 Bremerhaven

(1) Objective

The aim of the case study is to provide a scientific recommendation on sampling density based on real biological benthic data from the North Sea and the analysis settings chosen by the user (AP9.2).

(2) Introduction

The North Sea region is one of the most heavily used marine areas in the world. It is subject to increasing anthropogenic activities, such as shipping, marine food production and the construction of coastal protection and offshore windfarms (European Commission 2017). In the year 2012 the North Sea's blue economy alone represented 150 billion Euro, which amounts to approximately 30% of the total European Union budget, and employed at least 850.000 people (European Commission 2012). This growing human exploitation of marine resources by the different industrial sectors (fisheries, energy, minerals, transport, and recreation) increases the risks of long-term and possibly permanent changes in marine ecosystem functioning (Kenny 2018, Houde et al. 2014).

To counter this, the European Commission (inter alia) has implemented a sustainable "Blue Growth" strategy, which aims to ensure sustainable economic growth across marine sectors by recognizing the limits of naturally healthy, biodiverse, and biologically productive ecosystems (European Commission, 2012).

Accordingly, the German Federal Maritime and Hydrographic Agency ('Bundesamt für Seeschifffahrt und Hydrographie', BSH) has been commissioned to develop maritime spatial development plans for the German Exclusive Economic Zone (EEZ) of the North and Baltic Sea. These plans shall allow sustainable development of the maritime space and permit economic use, where they do not conflict with nature conservation and environmental protection, while ensuring the safety and ease of navigation (BSH, 2013).

Among the objects of environmental protection are benthic and epibenthic species (i.e. animals living in and on the sea floor). The benthos holds a key position in ecosystem functioning by providing food to higher trophic levels, sustaining fish populations and mediating biogeochemical cycles. In order to assess possible impacts on the benthos and other important functional groups, they are studied using standardized methods and monitoring programs (BSH 2013).

Monitoring of human activities and ecosystem components is continuously improving. This largely due to the development and application of autonomous and remote sensing technologies (Kenny and Sothoran, 2013). Understanding the causal pathways linking human activities, pressures and their actual impacts on ecosystem components, however, remains a challenging task (Korpinen and Andersen, 2016). This task is further complicated by the fact that many ecosystem components are themselves subject to natural variability. As this natural variability is an often-unknown quantity, it may lead to misinterpretations of the actual impacts of human activities on these ecosystem processes and functioning.

Numerous studies have already investigated the large-scale variability of the benthos in the German Bight. The aim of these studies was, however, to describe benthic communities based on different species compositions (Reiss et al. 2011, Pesch et al. 2005, Pehlke 2005, Rachor & Nehmer 2003, Salzwedel et al. 1985, Hagmeier 1925). Consequentially we, thus far, lack knowledge about the spatial patterns of benthos on smaller geographical scales, i.e. within benthic communities. In order to assess whether an observed variability is natural or anthropogenic, it is however necessary to be able to estimate this intra-habitat variability.

In answer to this knowledge gap Stieler (2016) demonstrated that a combination of a moving-window approach and the Bray-Curtis index are suitable to estimate the spatial intra-habitat variability (IHV) of two benthic communities (i.e. the *Amphiura filiformis* (Af) and *Bathyporeia-Tellina* (BT) community) in the Exclusive Economic Zone (EEZ) of the German North Sea.

This approach further allows determining whether the sampling effort and size of the chosen sampling area are appropriate to capture the variability within the habitat or community of interest.

In response to the expansion of wind energy at sea monitoring programs of each offshore wind farm aim to document the environmental status of ecosystems, monitor the effects of anthropogenic disturbances and detect potential environmental changes. To facilitate this, the monitoring of offshore wind farms consists of a sampling in the potential wind farm area and an unaffected area of comparable size and environmental conditions as a reference (BSH 2013).

The question arises, however, whether the size of the reference area and the number and spatial distribution of the sampling points within the monitoring areas allows to map the natural variability and thus to entangle the anthropogenic impact from the natural one. If this is the case, it is of great interest for cost reduction how many sampling points are minimally necessary and how they should be distributed in space. The tool "moving window approach" was developed to answer these questions through statistical analysis.

(3) Underlying background

The recommendation of the sampling density is calculated by the tool "moving window approach" that allows the user to load (benthic) data and modify the analysis data.

The recommendation of the minimum sample size is calculated on the basis of the Bray-Curtis index, which is calculated for a large number of different combinations of parts of the total data set (here: the data the user loaded into the tool). The number and design of these sub-data sets can be influenced by the user by changing various settings in the tool. The creation of these partial data sets is essentially based on the moving-window approach (Stieler, 2016).

Then the Bray-Curtis indices of all subsets (windows) are compared with the mean Bray-Curtis index of the total data. The subsets whose Bray-Curtis index is most similar to that of the overall data are selected as recommendations. For these subsets, the number of data points, minimum area and distribution of data points in space are analyzed. Thus a recommendation can be made regarding minimum number of data points, minimum size of the sample area and optimal arrangement of the sample points in space.

The "tool" was programmed using "R" and the package "Shiny" as a graphical user interface (GUI) and implemented as an interactive R Shiny web app.

The case study on the "moving window approach" for cost reduction (recommendation on sampling density) was developed within the project (AP8, AP9.2). The manageability of this "tool" was to be demonstrated by using case studies (here: benthos data) from a database. This proof is initially provided by two case studies: These consist of different benthos sample data from the *Amphiura filiformis* community and the *Bathyporeia-Fabulina* community (Rachor & Nehmer, 2003). An overview of the scope of these case study data is given in appendix table 2. In the final version, however, the data should be directly retrievable from the MARLIN database.

To illustrate the calculation of the Bray-Curtis index, here are some sample data in Table 1:

Table 1: Example data for calculating the Bray-Curtis Index

sample	species A	species B	species C	species D	species E	total	
1	10	0	4	7	0	21	
2	19	36	0	22	3	80	

To calculate Bray-Curtis dissimilarity, the absolute differences between the counts are added and divided by the sum of the frequencies in the two samples (here with the values from Table 1):

$$BC_{1,2} = \frac{|10 - 19| + |0 - 36| + |4 - 0| + |7 - 22| + |0 - 3|}{21 + 80} = \frac{67}{101} = 0.663$$

The Bray-Curtis dissimilarity, determines the difference, based on the Manhattan metric function, and lies in the range of values between 0 (similar) and 1 (dissimilar) and is usually multiplied by 100 and interpreted as a percentage.

The general formula for Bray-Curtis dissimilarity is as follows:

$$BC_{i,i'} = \frac{\sum_{j=1}^J |n_{ij} - n_{i'j}|}{n_{i+} + n_{i'+}}$$

where i, i' = samples,
 n_{ij} = counts (frequencies)
 n_{i+} = samples (rows) totals

Subtracting the Bray-Curtis dissimilarity from 100 gives a measure of similarity, the Bray-Curtis Index.

The app allows the presentation of the statistical analysis (as a box plot as well as a table), the cartographic presentation of the selected data sets and finally an analysis regarding the optimal sampling planning (number of sampling points, size of the study area, optimal distance of the sampling points from each other etc ...).

(4) Technical implementation

The programming was done with the software "R" (Version 3.5.3, R Core Team, 2014), but could also be implemented with other programming languages in the future. The software "R" itself is available as open source software under the GPL-2 or GPL-3 license and can be used free of charge privately and commercially.

"R" consists of several basic packages, which already provide a certain basic functionality that can be extended in a many ways by additional packages. The R scripts were programmed in "R-Studio" (RStudio Team, 2016), an integrated development environment and graphical user interface for "R" offered by RStudio, Inc.

The R package "shiny" (shiny: Web Application Framework for R, Chang et al. 2018) was used for programming the GUI. In addition, a number of other packages were used (see Appendix, Appendix Table 1).

In the context of this report it cannot be explained in detail how web applications are programmed using the R package "shiny" or which possibilities are available to the user using this package. It will only deal with the functionality and structure of shiny applications, as far as they contribute to the understanding of the underlying structure.

Shiny applications can be separated into two interacting parts by their structure (see figure 1). One part describes the structure of the application and is usually abbreviated as UI. In this section of the code, the entire layout of the application with all possible interactions, i.e. widgets, is defined. Widgets are web elements with which the user can interact, e.g. for entering data (see also Figure 5).

The second part is the server: Here, all outputs created in the UI are defined with R-Code, which can access all input values (Fritz, 2016).

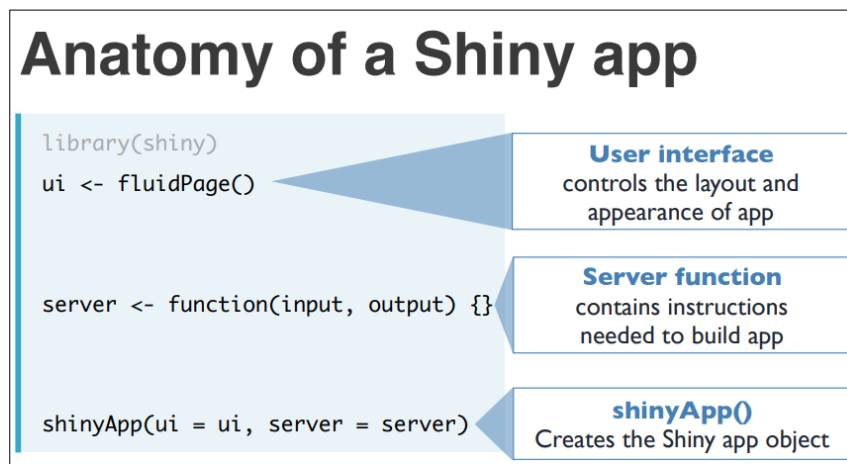


Figure 1: Structure of a Shiny app. Source: https://s3.amazonaws.com/assets.datacamp.com/production/course_4850/slides/Chapter1.pdf

The User-Interface (UI)

One can imagine the user interface (UI) of a web-app as the description of all visual elements. The code of the UI defines the structure of a digital document and is written in the text-based markup language Hypertext Markup Language (HTML).

Shiny uses the "fluidPage" function to create a display that automatically adapts to the dimensions of the user's browser window. One design the user interface of the app by placing elements in the fluidPage function.

For example, the simple UI function below creates a user interface with a title and sidebar layout that includes a sidebar panel and a main panel (see Figure 2).

```
ui <- fluidPage(
  titlePanel("title panel"),
  sidebarLayout(
    sidebarPanel("sidebar panel"),
    mainPanel("main panel")
  )
)
```

Figure 2: Basic UI function

The server part

Once the user interface is defined, all outputs created in this way can be defined in the server script using conventional R-code.

Conceptually, the server section of an app is a function whose parameters are all possible inputs and the labels of the created outputs.

All inputs defined by the user are passed to the server function as list objects. Here, one can imagine each individual input widget as a list entry (Fritz, 2016). A widget is a control element or a control element of a graphical user interface (see Figure 5). The app will be started with the following command (see Figure 4)

```
server <- function(input, output){}
```

Figure 3: Basic server function

Executing the application

Once the user interface and the server function have been defined, it remains to be clarified how the application can be executed. Up to Shiny version 0.10.1 it was necessary to define both parts of the application in separate scripts, which are stored with the names ui.R and server.R in a certain folder. Then, the app will be started with the following command (see Figure 4)

```
shinyApp(ui = ui, server = server)
```

Figure 4: Command used to start the Shiny app

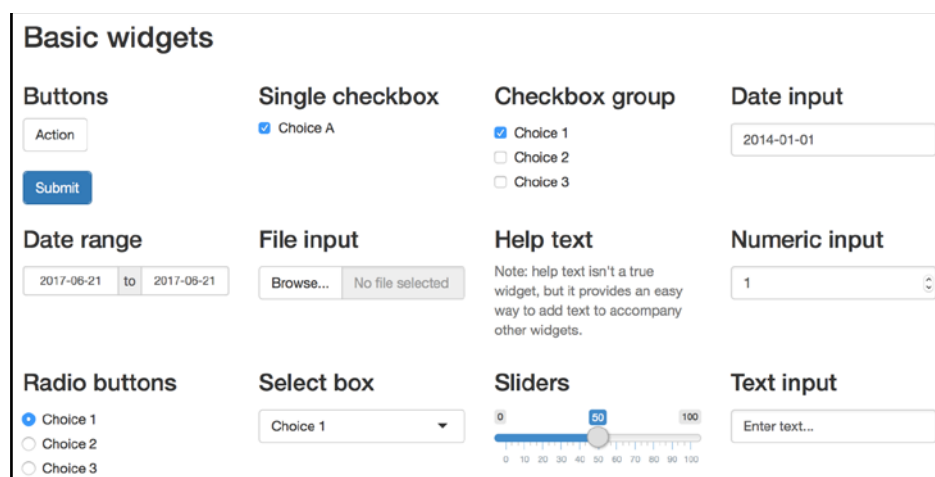


Figure 5: Basic widgets

(5) Description of the results

Structure of the website

The structure of the website is as follows: When you call up the application, the start screen ("Homepage", see Figure 6) opens. This screen is divided into two parts: In the side window on the left the navigation bar appears, in the main window widgets are displayed, with which the user can load and select data.

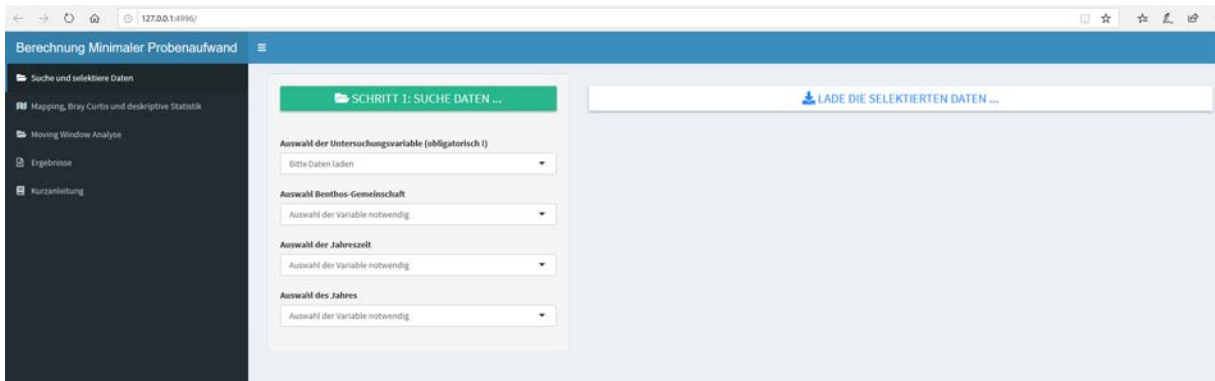


Figure 6: Start screen of the app

The main window of the app contains various widgets (see Figure 5) with which the user can define his data selection (see Figure 7).



Figure 7: Main panel of the homepage. (1) allows the data to be loaded and should later allow a connection to the database, (2) is a mandatory selection between abundance or biomass data, (3) allows a further selection to a benthic community, (4) allows the selection of a season, (5) here the selection can be limited to one year and with (6) the selected files are loaded.

The main window contains six widgets (see Figure 7): Widget 1 currently opens the path to the data and is intended to create a connection to the MARLIN database in the future. The single steps of the data selection are shown in figure 8.

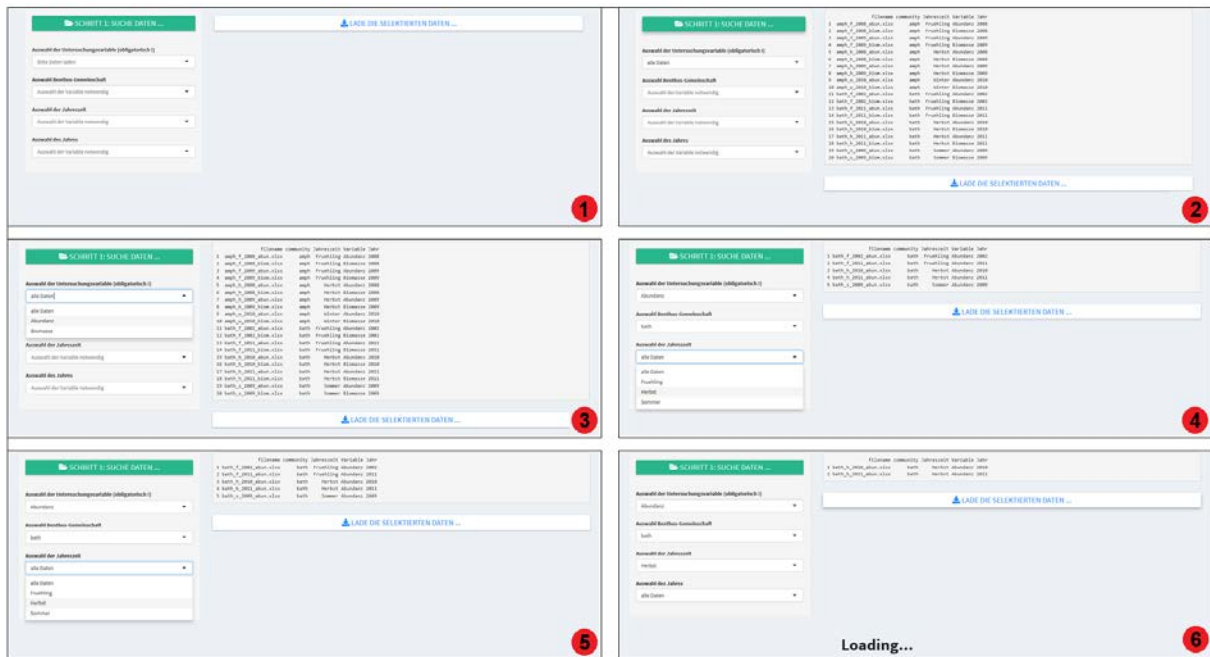


Figure 8: Individual steps of data selection in the main window of the homepage.

Image 1 in Figure 8 shows the start screen after calling the app, in image 2 the button "STEP 1: SEARCH DATA" was clicked and all available data will be displayed. Image 3 shows the selection options in the mandatory selection menu "Selection of the examination variable (mandatory!)". Here the user can currently choose between abundance and biomass. The third option here ("all data") serves to delete previous selections. The selection window "Selection Benthic Community" below allows the user to select the benthic community to be investigated, in our case there are three options, either:

1. "all data": a data selection without restriction of the benthic community
2. "amph": a reduction to data of the *Amphiura filiformis* community; or
3. 'bath': a reduction of the *Bathyporeia-Fabulina* community

Image 5 in Figure 8 shows the season selection options. Depending on the selections made before in the other widgets, the still possible choices are displayed here. Image 6 shows the abundance data of the *Bathyporeia-Fabulina* community, which were collected in autumn. A restriction to one year was not made in this example (image 8). After clicking the button "LOAD SELECTED DATA ..." the selected data sets (here 2 data sets) are loaded. After the data have been loaded completely, the next window "Mapping, Bray Curtis and Descriptive Statistics" will be opened automatically (see Figure 10).

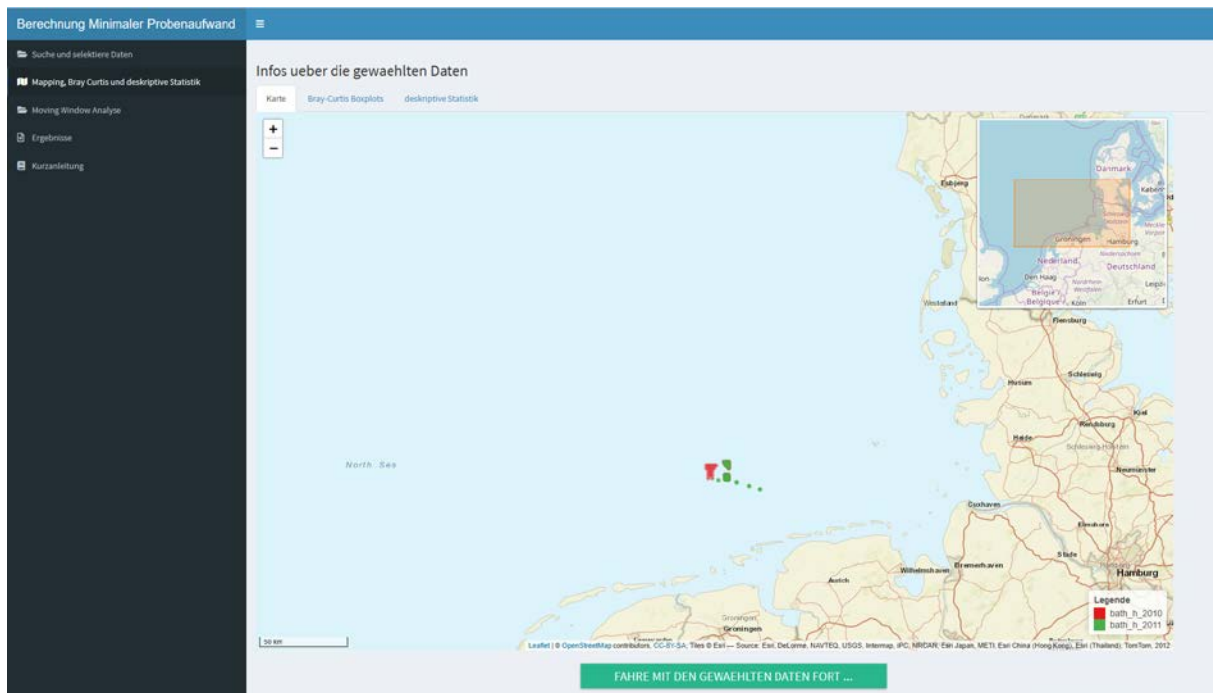


Figure 9: In the first screen, of the second tab "Mapping, Bray Curtis and Descriptive Statistics" the loaded data (see Figure 8) are displayed on a map. In addition, the main window also contain the tabs "Bray-Curtis Boxplots" and "Descriptive Statistics".

The main window of the second tab "Mapping, Bray Curtis and Descriptive Statistics" (see Figure 9) contains three additional tabs. The first these tabs named "Map" displays the loaded data (see Figure 8) in a map. It is possible to zoom in and out on the map and to move the map section. Furthermore, the main window also contains the tabs "Bray-Curtis Boxplots" and "Descriptive Statistics".



Figure 10: Screen of the "Bray-Curtis Boxplots" tab

Switching to the second tab "Bray-Curtis Boxplots", one can see the boxplots of the Bray-Curtis analyses for each individually loaded data set as well as the analysis of the summarized data. The box plot of the summarized data is displayed in a yellow color for easy identification. The mean value of the Bray-Curtis data is displayed as a red dot (see Figure 10 and 11).

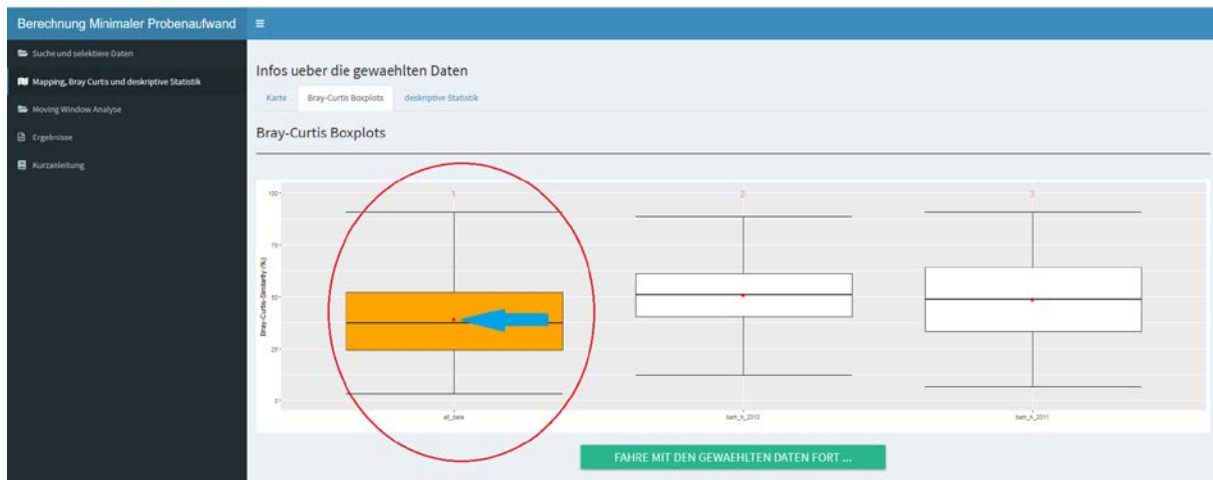


Figure 11: Screen of the "Bray-Curtis Boxplots" tab, the boxplot of merged data is highlighted (yellow colour).

Switching to the third tab "Descriptive Statistics", one will see the relevant statistical parameters of the individual selected files and those of the merged files ("all_data", see Figure 12).

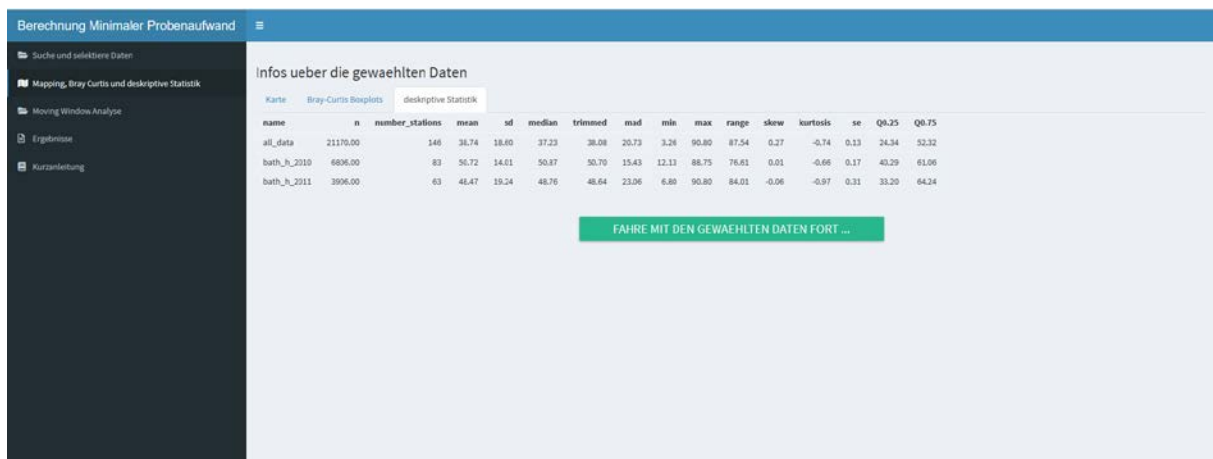


Figure 12: Screen of the "Descriptive Statistics" tab

In Figure 13, the statistical parameters of the merged data ("all_data") are highlighted with a red circle. The statistical parameters also include the mean value ("mean"), which has been framed in blue in this text for better visibility.

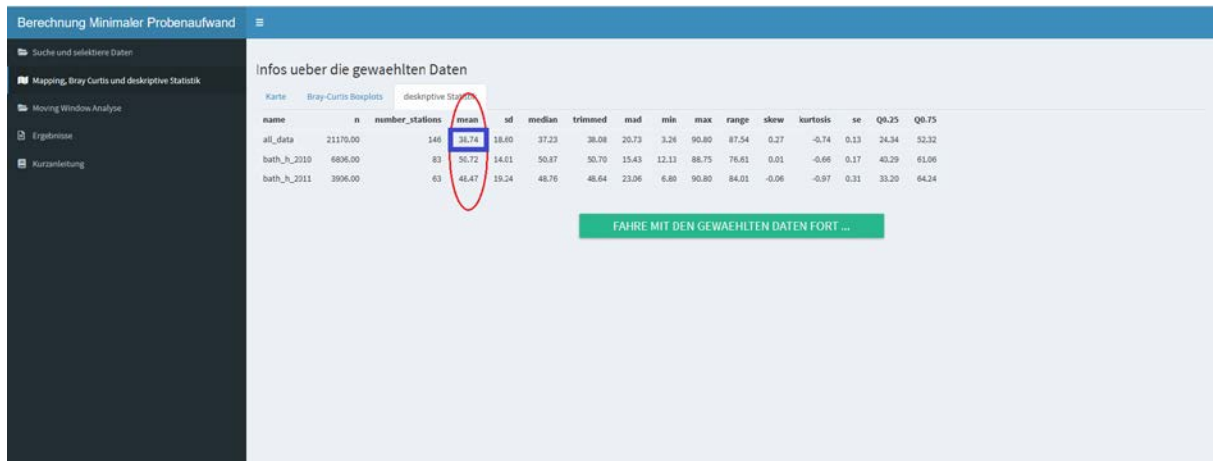


Figure 13: Screen of the tab "Descriptive Statistics". The merged data as well as the relevant mean value are highlighted here in this text (this is not the case in the app).

In this example, the mean value of the Bray-Curtis data for the summarized data is thus 38.74 (see Figure 13). This value is the value for the following moving-window analysis.

A click on the button "CONTINUE WITH THE SELECTED DATA ...", opens the next tab, in which the settings for the moving window analysis can be defined.

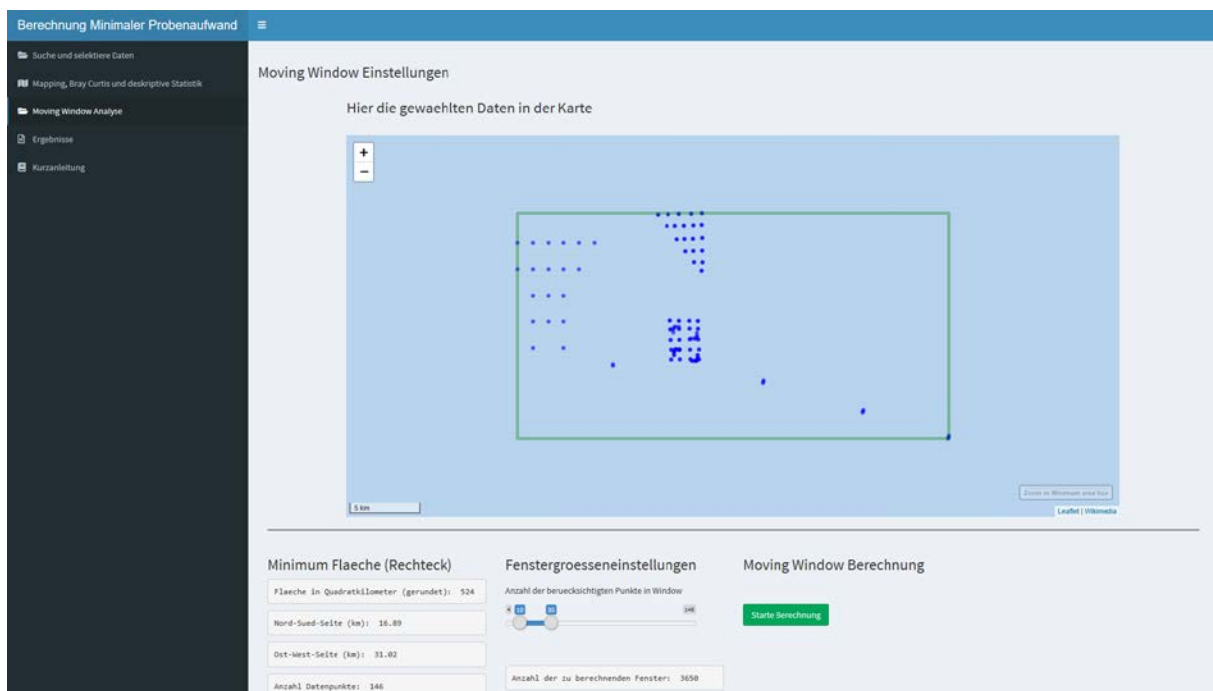


Figure 14: Screen of the tab "moving-window analysis".

In the main window of the third tab you can see the selected data on the map again. The map also shows the calculated minimum rectangular area containing all selected sampling points (see Figure 14). Below the map the following calculated information is displayed: The area of the minimum rectangular area in square kilometers, the North-South extension, the East-West extension (each in km) and the number of selected sampling points (see Figure 14). In addition to this information, one will also find the possibility to define the settings of the moving-window analysis. The slider allows to define both the minimum and

the maximum number of points a moving window may contain. Using these user setting inputs, the app immediately calculates how many windows have to be calculated.

Figure 15: Section of the main window of the tab "moving-window analysis" (see also Figure 14)

A click on the "Start calculation" button starts the analysis of the moving windows. In this example with the settings selected in Figure 15 (minimum = 10, maximum = 35). In the analysis, each of the 146 selected data points is defined as a starting point. Now the first of these starting points is selected. For this first starting point the next 9 neighboring data points are now selected (based on the Euclidean distance). These 10 data points (9 + start point = 10 = minimum, see Figure 15) simultaneously define a rectangle that represents the minimum (rectangular) area in which these 10 points lie.

Now a Bray-Curtis analysis is performed for these 10 points. The mean value of this Bray-Curtis calculation will be compared with the target value later (target value here = 38.74, see Figure 13). At the same time, the area of the calculated minimum rectangle is determined and a neighborhood analysis of the 10 points is carried out (minimum and maximum distance to the neighbors, as well as the mean value). It is analyzed whether the points are rather randomly distributed in space, whether the points are regularly distributed or whether clustering is present (this calculation is statistically only valid from 30 points upwards, but if the number of points is lower, the calculations can at least be used as an indication).

In the next step, the number of neighboring points around the first starting point is increased by one, so that there are now 11 points in the window (10 neighboring points + start point). Again, the calculation of the minimum rectangular area, the Bray-Curtis analysis and the neighborhood analysis are carried out. Gradually, the number of neighboring points is increased until finally the selected maximum value of points in the window is reached. (here 35 points, see Figure 15).

Then the next starting point is selected and the calculations start again with nine neighboring points. This is continued until the calculations for the last starting point (here number 146) have been completed. This procedure was developed and modified according to the moving-windows procedure presented in the master thesis of Stieler (2016).

A randomizer is to be added to this procedure, but this has not yet been implemented.

Berechnung Minimaler Probenaufwand

Suche und selektiere Daten

Mapping, Bray-Curtis und deskriptive Statistik

Moving Window Analyse

Ergebnisse

Kurzanleitung

Results

Moving Window Tabelle | Bray-Curtis Scatter-Plots | Statistische Auswertung

[1] "Anzahl der analysierten Nachbarpunkte (von.. bis): 10 bis 35"

Show 15 entries

Search:

	window_size	area_sqkm	pts_in_win	BC_mean	BC_sd
1	3338m*3786m	12.638	10	68.5761757080341	9.96575165011624
2	4420m*3786m	16.767	11	68.5171440889506	9.89988919800864
3	4420m*3786m	16.767	12	68.0271446261344	9.82726363653157
4	4420m*3825m	16.94	13	67.215580364625	10.154381622072
5	5518m*3825m	21.108	14	66.9595136533202	10.2308459566815
6	5518m*5601m	30.908	15	60.80194844297	17.8142442203024
7	5518m*5640m	31.124	16	57.4715419723674	19.0576794116731
8	5518m*5679m	31.134	17	57.2638534014601	18.2444889898866
9	5518m*7455m	41.139	18	55.2601470318861	18.2500931950758
10	5518m*7534m	41.572	19	53.1357061491366	18.66838720105
11	10052m*9152m	91.989	20	51.8123731642159	18.2754834099087
12	10052m*9152m	91.989	21	48.886405633453	19.6044481421215
13	10659m*9152m	97.547	22	48.0629626723196	19.2205558664108
14	10659m*9152m	97.547	23	45.142165670637	20.8993896138115
15	10765m*9152m	98.519	24	44.2997348633704	20.5392536278099

Showing 1 to 15 of 1,538 entries

Previous 1 2 3 4 5 ... 103 Next

Figure 16: First tab ("moving window Table") of the tab "Results".

The "Results" tab (see Figure 16) then displays the table of all calculated moving windows in the main window of the "moving window Table" tab. Below the column names there are filter options so that one can perform a filtering of the data (= subset of the result). A download option for the table is to be added, but this has not yet been implemented.

In the second tab "Bray-Curtis Scatter Plots" of the tab "Results" (see Figure 17) the results of the moving window analysis can be viewed as a scatter plot. It is possible to select freely the variables of both the x- and the y-axis. In addition, it is possible to zoom in and out if a closer look at certain areas is needed.

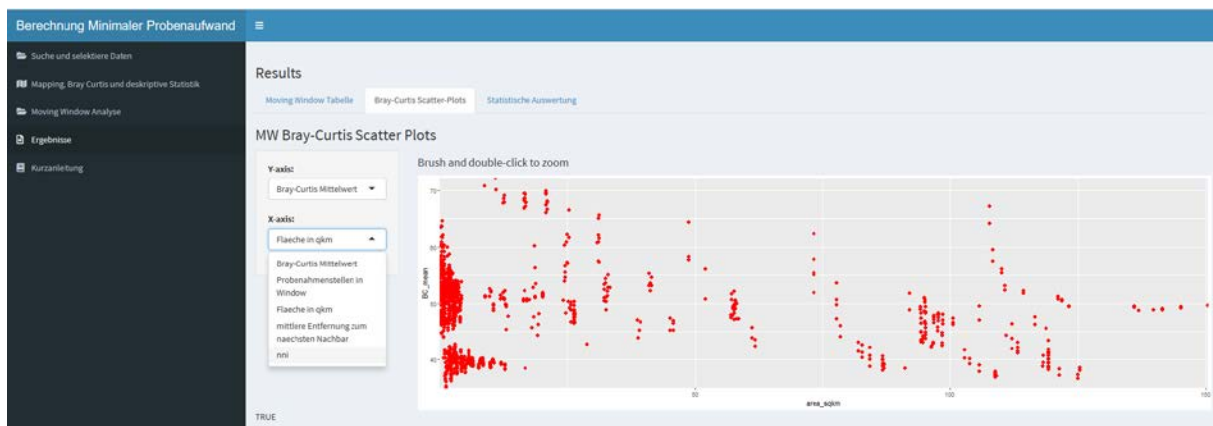


Figure 17: Second tab ("Bray-Curtis-Splatterplots") of the "Results" tab

The third tab ("Statistical Evaluation") of the tab "Results" allows a statistical evaluation of the moving-window analysis (see Figure 18).

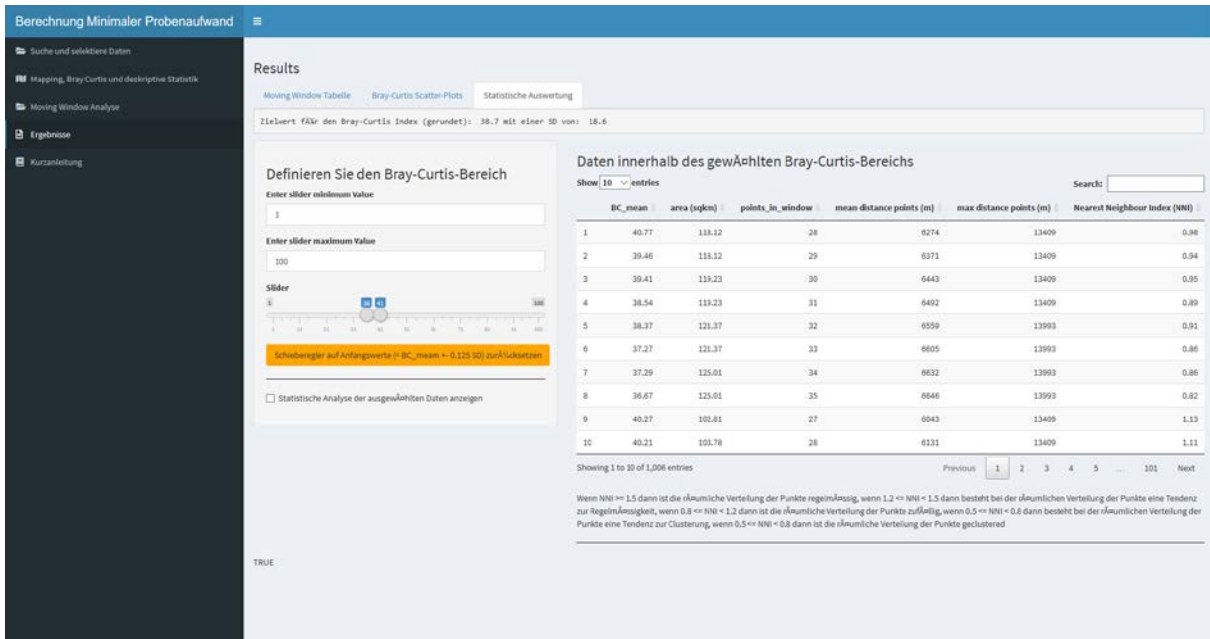


Figure 18: Third tab ("Statistical Evaluation") of the "Results" tab.

The main window contains a variety of information and setting options (see the following illustrations). At the top of the main window, the Bray-Curtis target value (in this example 38.74) is displayed again, to make the analyses easier for the user (see Figure 19, point 1). Underneath there is a slider that can be used to limit the amount of data in the moving windows table. The target variable is table column 1 (BC_mean) of this table. The user can define the minimum and maximum value of BC_mean with the slider now. The aim should be to bring the selected minimum and maximum as close as possible to the target value (38.74 in this example) while maintaining a statistically meaningful number of data in the data table. By changing the settings in the slider, the number of data contained in the displayed table is automatically adjusted. In order to get closer to the BC_mean target value (see Figure 19, point 1), it is possible to limit or refine the slider range. As can be seen in Figure 20, the minimum for table column 1 "BC_mean" was set to 37.7 (originally 0) and the maximum to 39.7 (originally 100, see Figure 19). These settings reduced the data set from originally 1538 to 1006.

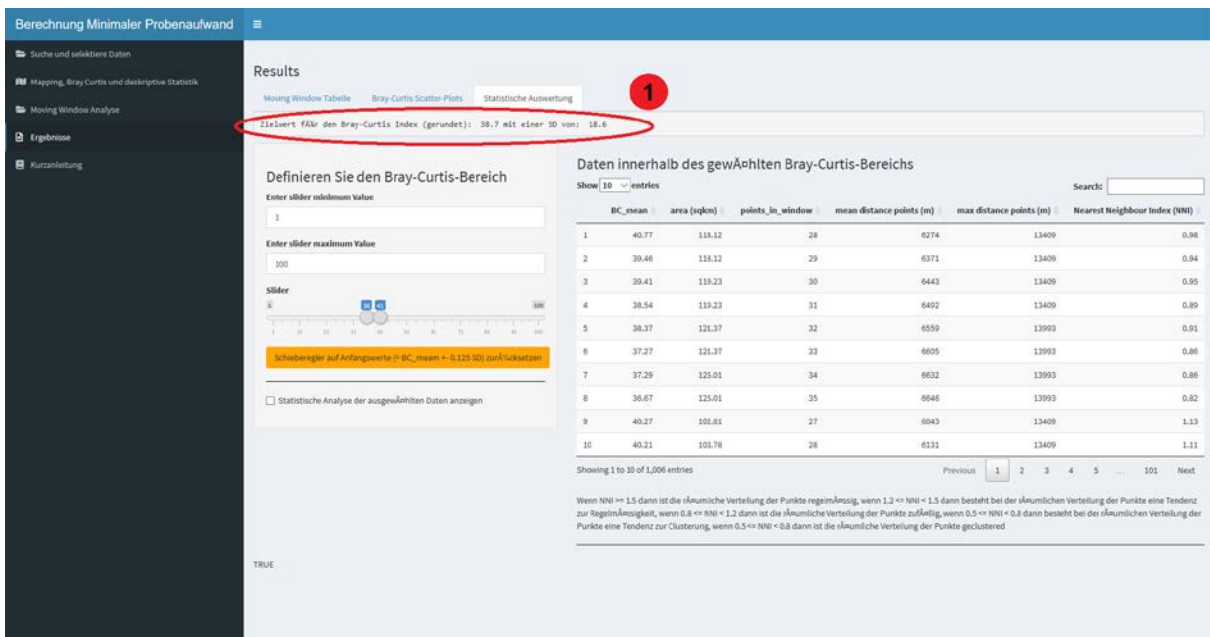


Figure 19: Third tab ("Statistical analysis") of the "Results" tab. Display of the rounded Bray-Curtis target value (in this example 38.7)

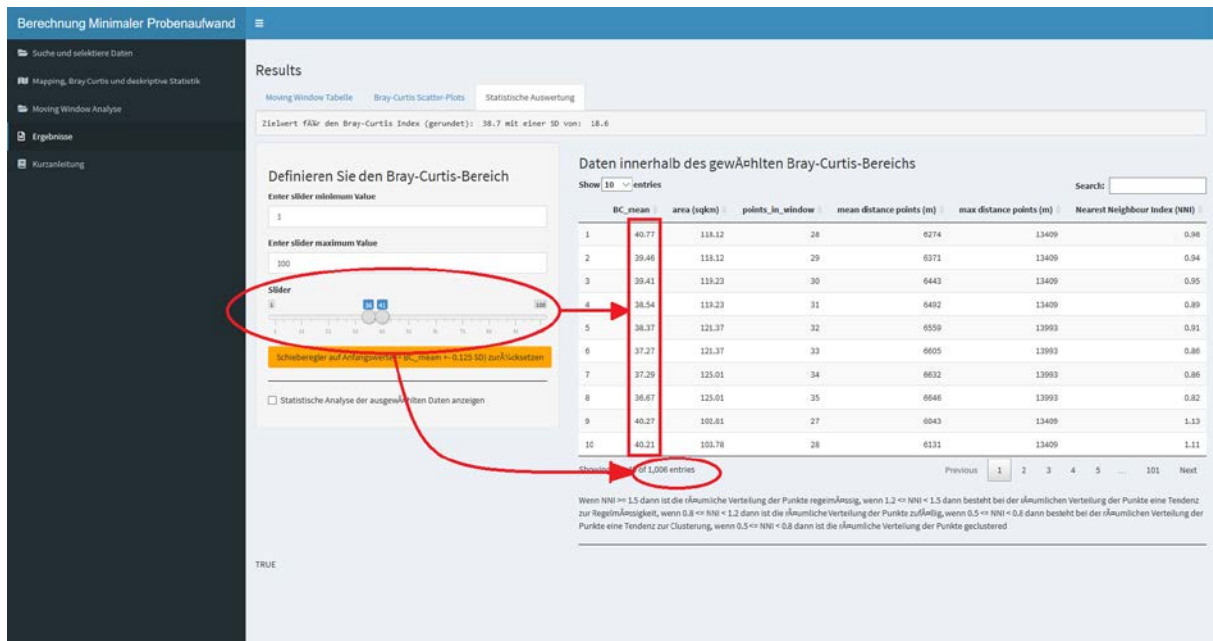


Figure 20: Third tab ("Statistical analysis") of the tab "Results". Changed slider values to narrow down the results.

It is also possible to change the minimum and maximum of the slider itself (see Figure 21). This is done by manually entering new values in the fields "Enter slider minimum value" and/or "Enter slider maximum value" (Figure 21, point 1). This changes the appearance of the slider and these values now define the start and end point of the slider (Figure 21, point 2). This allows an even more precise delimitation of the search range by the slider. In the slider, the range has now been limited to values between 38.5 and 39. As result, now the table only contains data whose "BC_MEAN" lies between these values (see Figure 21, points 2 and 3) and the number of data in the table has been further reduced to now 133 data sets (see Figure 21, point 4).



Figure 21: Third tab ("Statistical analysis") of the "Results" tab. Further setting options of the slider.

In the next step, these 133 data records are statistically evaluated. By selecting the button "Show statistical analysis of selected data" (see Figure 22, point 1) a summary statistics of the selected 133 data is shown.

The screenshot shows the 'Results' tab of a software application. The main content area is divided into several sections:

- Definieren Sie den Bray-Curtis-Bereich:** A section for defining the Bray-Curtis range, featuring a slider and input fields for minimum and maximum values. A checkbox labeled "Statistische Analyse der ausgewählten Daten anzeigen" (point 1) is highlighted with a red circle.
- Daten innerhalb des gewählten Bray-Curtis-Bereichs:** A table displaying 10 data entries with columns: BC_mean, area (sqkm), points_in_window, mean distance points (m), max distance points (m), and Nearest Neighbour Index (NNI). A green circle highlights the target value "Zielwert für den Bray-Curtis Index (gerundet) 30.7 m" (point 4).
- Statistik:** A summary statistics table for the selected data. A red arrow points to the "Mean" value in the "BC_mean" column (point 2), which is circled in red (point 3).

BC_mean	area (sqkm)	points_in_window	mean distance points (m)	max distance points (m)	Nearest Neighbour Index (NNI)
38.54	119.23	31	6492	13409	0.89
38.86	118.1	30	6357	13409	1.05
38.54	119.23	31	6492	13409	0.89
38.83	84.25	29	6026	12812	0.97
38.8	86.78	30	6057	12812	0.97
38.56	119.23	33	6680	13409	0.81
38.81	119.23	34	6760	13409	0.82
38.56	119.23	33	6680	13409	0.81
38.81	119.23	34	6760	13409	0.82
38.56	119.23	33	6680	13409	0.81

BC_mean	area (sqkm)	points_in_window	mean distance points (m)	max distance points (m)	Nearest Neighbour Index (NNI)
Min.: 38.50	Min.: 1.38	Min.: 30.00	Min.: 920	Min.: 1781	Min.: 0.8000
1st Qu.: 38.62	1st Qu.: 4.28	1st Qu.: 23.00	1st Qu.: 1380	1st Qu.: 2548	1st Qu.: 0.8000
Median: 38.72	Median: 6.65	Median: 30.00	Median: 1586	Median: 3587	Median: 0.1180
Mean: 38.74	Mean: 18.63	Mean: 26.71	Mean: 2068	Mean: 4491	Mean: 0.1789
3rd Qu.: 38.88	3rd Qu.: 18.38	3rd Qu.: 33.00	3rd Qu.: 1778	3rd Qu.: 4297	3rd Qu.: 0.1300
Max.: 38.99	Max.: 129.33	Max.: 35.00	Max.: 8768	Max.: 13409	Max.: 1.0500

Figure 22: Third tab ("Statistical analysis") of the tab "Results". Statistical evaluation of the selected data.

In the statistics table the following variables are displayed: "BC_mean", "area (sqkm)", "points_in_window", "mean distance points (m)", "max distance points (m)", and "Nearest Neighbor Index (NNI)" (Figure 22, point 2).

For each of these variables the minimum, 1st quartile, median, mean, 3rd quartile and maximum are displayed (see Figure 22, point 2).

To estimate the quality of the selected data sets, one should look at the first column ("BC_mean") of the statistics (Figure 22, point 3). The value "Mean" in this table column should come very close to the target value (Figure 22, point 4).

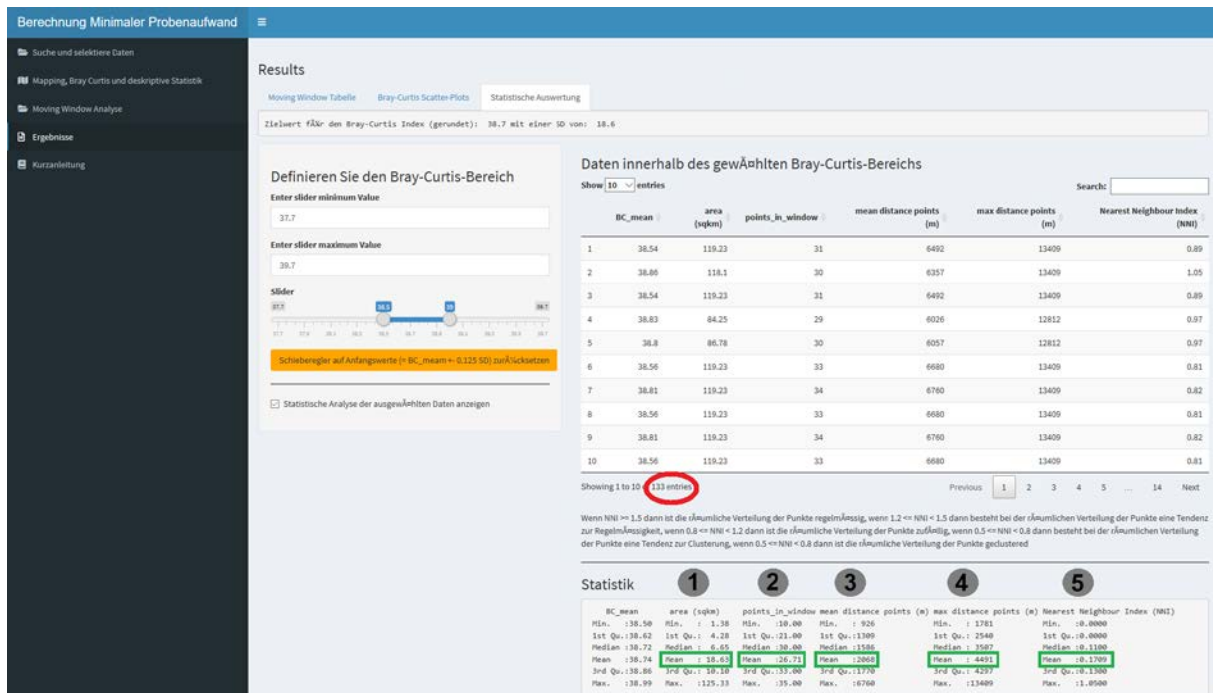


Figure 23: Third tab ("Statistical analysis") of the "Results" tab. More information about the statistics.

If this comparison between the mean value of "BC_mean" and the target value is satisfactory, it is time for a look at the other columns of the statistics (Figure 23, points 2-5). The column "area_sqkm" shows the descriptive statistics of the area sizes of the selected 133 data sets. The average value in our example is 18.63 km² (see Figure 23, point 1). The average number of sample points (see Figure 23, point 2) is 26.71. The minimum number (10) and the maximum number (35) of the sample points was previously defined by the user (see Figure 14 and 15). Point 4 and point 5 (Figure 23) analyze the average and maximum distance of the individual points from each other (in each of the 133 data sets = moving windows). This information can be useful for planning a future sampling pattern. The last column (point 5, Figure 23) shows the NNI index. This index calculates whether the point data are randomly distributed in space, whether the points are regularly distributed, or whether clustering is present. This information can also be helpful when planning sampling.

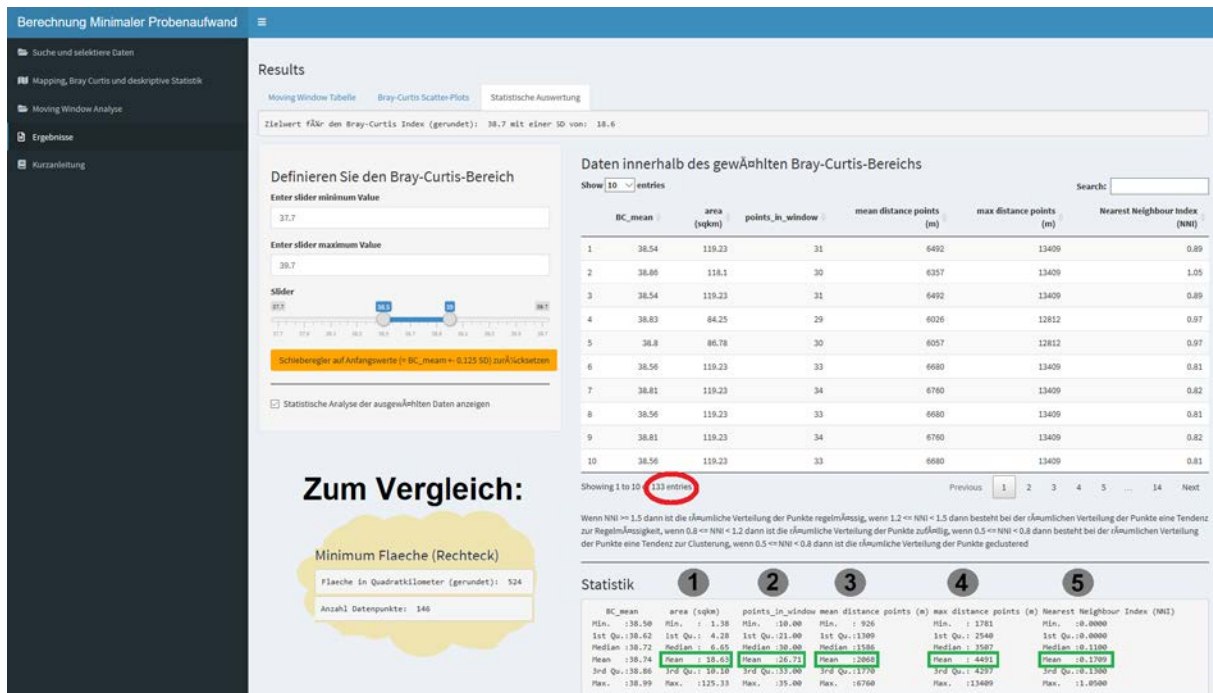


Figure 24: Third tab ("Statistical analysis") of the "Results" tab. Comparison of the achieved values with those of the original data loaded by the user

Comparing the values of the loaded original data with those of the selected 133 moving windows (see Figure 24), the benefits of moving window analysis are easier to understand. The original data contains 146 data points within an area of 524 km². The original data had an average Bray-Curtis value of 38.7.

For our example, the moving window analysis showed that a comparable Bray-Curtis index can be achieved with 27 sampling points on average, and that this can be achieved in an area that covers only 18.63 km² on average. This is less than 1/5 of the sampling points of the original data and the area to be sampled (mean value) is only 3% of the area of the original data.

(6) Appendix and literature

Digital Appendix

- Digital_annex_MovingWindow_AP9-R-script: Fully commented script for the calculation of the required minimum sampling density

Literature

Allaire, J. J. Application layout guide. Last Updated: 24 Jan 2014. URL : <http://shiny.rstudio.com/articles/layout-guide.htm>

AnalytixWare (2014). `_shinysky`: A set of Shiny UI components/widgets. R packageVersion 0.1.2, <URL: <https://github.com/AnalytixWare/ShinySky>>.

Aphalo PJ (2016). `_Learn R ...as you learnt your mother tongue_`. Leanpub. <URL: <https://leanpub.com/learnr>>.

Appelhans T, Detsch F, Reudenbach C, Woellauer S (2018). `_mapview`: Interactive Viewing of Spatial Data in R. R packageVersion 2.6.3, <URL: <https://CRAN.R-project.org/package=mapview>>.

Attali D (2018). `_shinyjs`: Easily Improve the User Experience of Your Shiny Apps in Seconds. R packageVersion 1.0, <URL: <https://CRAN.R-project.org/package=shinyjs>>.

Bailey E (2015). `_shinyBS`: Twitter Bootstrap Components for Shiny. R packageVersion 0.61, <URL: <https://CRAN.R-project.org/package=shinyBS>>.

Bivand R, Keitt T, Rowlingson B (2019). `_rgdal`: Bindings for the 'Geospatial' Data Abstraction Library. R packageVersion 1.4-3, <URL: <https://CRAN.R-project.org/package=rgdal>>.

Bivand R, Rundel C (2018). `_rgeos`: Interface to Geometry Engine - Open Source ('GEOS'). R packageVersion 0.4-2, <URL: <https://CRAN.R-project.org/package=rgeos>>.

Bivand RS, Pebesma E, Gomez-Rubio V (2013). `_Applied spatial data analysis with R, Second edition_`. Springer, NY. <URL: <http://www.asdar-book.org/>>.

Bundesamt für Seeschifffahrt und Hydrographie (BSH) (Hg.) (2013): Standard. Untersuchung der Auswirkungen von Offshore-Windenergieanlagen auf die Meeresumwelt (StUK4). Hamburg

Chang W, Borges Ribeiro B (2018). `_shinydashboard`: Create Dashboards with 'Shiny'. R packageVersion 0.7.1, <URL: <https://CRAN.R-project.org/package=shinydashboard>>.

Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2018). `_shiny`: Web Application Framework for R. R packageVersion 1.2.0, <URL: <https://CRAN.R-project.org/package=shiny>>.

Cheng J, Karambelkar B, Xie Y (2018). `_leaflet`: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R packageVersion 2.0.2, <URL: <https://CRAN.R-project.org/package=leaflet>>.

- de Mendiburu F (2019). `_agricolae`: Statistical Procedures for Agricultural Research_. R packageVersion 1.3-0, <URL: <https://CRAN.R-project.org/package=agricolae>>.
- Dowle M, Srinivasan A (2019). `_data.table`: Extension of `data.frame`_. R packageVersion 1.12.0, <URL: <https://CRAN.R-project.org/package=data.table>>.
- European Commission (2017): Report on the Blue Growth Strategy Towards more sustainable growth and jobs in the blue economy. European Commission, SWD (2017) 128 final. 61 pp.
- Firke S (2018). `_janitor`: Simple Tools for Examining and Cleaning Dirty Data_. R packageVersion 1.1.1, <URL: <https://CRAN.R-project.org/package=janitor>>.
- Fritz, C (2016). Explorative Datenvisualisierung mit Shiny in R. Bachelorarbeit. Ludwig-Maximilian-Universität München. Institut für Statistik.
- Graves S, Piepho H, Dorai-Raj LSwhfS (2015). `_multcompView`: Visualizations of Paired Comparisons_. R packageVersion 0.1-7, <URL: <https://CRAN.R-project.org/package=multcompView>>.
- Grosjean P (2019). `_SciViews-R`. UMONS, MONS, Belgium. <URL: <http://www.sciviews.org/SciViews-R>>.
- Hagmeier, A. (1925): Vorläufiger Bericht über die vorbereitenden Untersuchungen der Bodenfauna der Deutschen Bucht mit dem Petersen-Bodengreifer. In: Berichte der Deutschen Wissenschaftlichen Kommission für Meeresforschung 1: 249-272
- Hijmans R (2019). `_raster`: Geographic Data Analysis and Modeling_. R packageVersion 2.8-19, <URL: <https://CRAN.R-project.org/package=raster>>.
- Houde E. , Kenny A. J. & Zhou S. (2014): EBM in highly impacted coasts and estuaries. In *The Sea, Marine Ecosystem-Based Management*, 16: 475–522. Ed. by, Fogarty M. J., McCarthy J. J.. Harvard University Press, Boston, MA.
- Iannone R, Allaire J, Borges B (2018). `_flexdashboard`: R Markdown Format for Flexible Dashboards_. R packageVersion 0.5.1.1, <URL: <https://CRAN.R-project.org/package=flexdashboard>>.
- Judd A. D. , Backhaus T. & Goodsir F. (2015): An effective set of principles for practical implementation of marine cumulative effects assessment. *Environmental Science and Policy*, 54: 254–262.
- Kenny A. J. & Sotheran I. (2013): Characterising the physical properties of seabed habitats. In *Methods for the Study of Marine Benthos*, 4th edn, pp. 47–95. Ed. by Eleftheriou A.. Wiley Blackwell, Oxford.
- Kenny, A.J., Jenkins, C, Wood, D, Bolam, S.G., Mitchell, P., Scougal, C. & Judd, A (2018): Assessing cumulative human activities, pressures, and impacts on North Sea benthic habitats using a biological traits approach, *ICES Journal of Marine Science*, Volume 75, Issue 3, May-June 2018, Pages 1080–1092, <https://doi.org/10.1093/icesjms/fsx205>
- Korpinen S. & Andersen J. H. (2016): A global review of cumulative pressure and impact assessments in marine environments. *Frontiers in Marine Science*, 3: 153–164.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2019). `_vegan`:

- Community Ecology Package_. R packageVersion 2.5-4, <URL: <https://CRAN.R-project.org/package=vegan>>.
- Pebesma E (2018). "Simple Features for R: Standardized Support for Spatial Vector Data." *_The R Journal_*. <URL: <https://journal.r-project.org/archive/2018/RJ-2018-009/index.html>>.
- Pebesma EJ, Bivand RS (2005). "Classes and methods for spatial data in R." *_R News_*, *5*(2), 9-13. <URL: <https://CRAN.R-project.org/doc/Rnews/>>.
- Pehlke, H. (2005): Prädiktive Habitatkartierung für die Ausschließliche Wirtschaftszone (AWZ) der Nordsee. Vechta
- Perrier V, Meyer F, Granjon D (2019). *_shinyWidgets: Custom Inputs Widgets for Shiny_*. R packageVersion 0.4.8, <URL: <https://CRAN.R-project.org/package=shinyWidgets>>.
- Pesch, R., Pehlke, H. & Schroeder, W., (2005): Predictive Benthic Habitat Mapping in the North Sea using GIS and Statistical Methods. In: Hřebíček, J. & Ráček, J. (Hrsg.), *Informatics for Environmental Protection - Networking Environmental Information*. Brno: Masaryk University Brno.
- R Core Team (2019). *_R: A Language and Environment for Statistical Computing_*. R Foundation for Statistical Computing, Vienna, Austria. <URL: <https://www.R-project.org/>>.
- Rachor, E. & Nehmer, P. (2003): Erfassung und Bewertung ökologisch wertvoller Lebensräume in der Nordsee. Abschlußbericht für das F & E-Vorhaben FKZ 899 85 310 (Bundesamt für Naturschutz). Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven.
- Rachor, E.; Nehmer, P. (2003): Erfassung und Bewertung ökologisch wertvoller Lebensräume in der Nordsee. BfN-Projekt-Bericht. Bremerhaven
- Reiss H., Cunze S., König K., Neumann H. & Kröncke I. (2011): Species distribution modelling of marine benthos: a North Sea case study. *Mar Ecol Prog Ser* 442:71-86. <https://doi.org/10.3354/meps09391>
- Revelle W (2018). *_psych: Procedures for Psychological, Psychometric, and Personality Research_*. Northwestern University, Evanston, Illinois. R packageVersion 1.8.12, <URL: <https://CRAN.R-project.org/package=psych>>.
- RStudio Team (2016). *_RStudio: Integrated Development Environment for R_*. RStudio, Inc., Boston, MA. <URL: <http://www.rstudio.com/>>.
- Sali A (2017). *_shinycssloaders: Add CSS Loading Animations to 'shiny' Outputs_*. R packageVersion 0.2.0, <URL: <https://CRAN.R-project.org/package=shinycssloaders>>.
- Salzwedel, H.; Rachor, E. & Gerdes, D. (1985): Benthic macrofauna communities in the German Bight. In: Veröffentlichung des Institutes für Meeresforschung Bremerhaven 20: 199-267
- Sarkar D (2008). *_Lattice: Multivariate Data Visualization with R_*. Springer, New York. ISBN 978-0-387-75968-5, <URL: <http://lmdvr.r-forge.r-project.org>>.
- Simpson G (2019). *_permute: Functions for Generating Restricted Permutations of Data_*. R packageVersion 0.9-5, <URL: <https://CRAN.R-project.org/package=permute>>.

- Stieler, K. (2016): Intra-Habitat-Variabilität von Benthosgemeinschaften in der Ausschließlichen Wirtschaftszone der deutschen Nordsee. Thesis (Master). pp. 111. hdl:10013/epic.48875
- Temple Lang D, Wallace J (2018). `_RJSONIO: Serialize R Objects to JSON, JavaScript Object Notation_`. R packageVersion 1.3-1.1, <URL: <https://CRAN.R-project.org/package=RJSONIO>>.
- Walker A (2018). `_openxlsx: Read, Write and Edit XLSX Files_`. R packageVersion 4.1.0, <URL: <https://CRAN.R-project.org/package=openxlsx>>.
- Wickham H (2011). "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software*, *40*(1), 1-29. <URL: <http://www.jstatsoft.org/v40/i01/>>.
- Wickham H (2016). `_ggplot2: Elegant Graphics for Data Analysis_`. Springer-Verlag New York. ISBN 978-3-319-24277-4, <URL: <http://ggplot2.org>>.
- Wickham H, François R, Henry L, Müller K (2019). `_dplyr: A Grammar of Data Manipulation_`. R packageVersion 0.8.0.1, <URL: <https://CRAN.R-project.org/package=dplyr>>.
- Xie Y, Cheng J, Tan X (2018). `_DT: A Wrapper of the JavaScript Library 'DataTables'_`. R packageVersion 0.5, <URL: <https://CRAN.R-project.org/package=DT>>.