A visual guide to many of Humboldt's input parameters

https://github.com/jasonleebrown/humboldt.git
http://www.jasonleebrown.org

Index of Humboldt parameters covered

Input Parameters:

col.env

e.var

env1

env2

Input Files

Environmental Data: environmental variables for all sites of the study area 1 (**env1**) or study area 2 (**env1**).

Column names should be x, y, X1,X2,... Xn; with X1-Xn being any string label. If env1=env2, input the same file twice. I typical import tab delimited (.txt) or comma separated files (.csv).

Column numbers as R sees them:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

x	y	BIO_1	BIO_2	BIO_3	BIO_4	BIO_5	BIO_6	BIO_7	BIO_8	BIO_9	BIO_10	BIO_11	BIO_12	BIO_13	BIO_14	BIO_15	BIO_16	BIO_17	BIO_18	BIO_19
-38.12499	73.12501	-226	37	11	9508	-54	-368	314	-228	-118	-92	-331	1228	131	77	15	379	260	274	300
-37.70833	73.12501	-226	36	11	9480	-55	-367	312	-228	-119	-93	-331	1234	131	77	15	380	261	276	303
-37.29166	73.12501	-226	37	11	9440	-56	-367	311	-228	-119	-93	-330	1234	131	76	15	380	260	276	303
-38.95833	73.95834	-225	39	12	9846	-47	-371	324	-227	-113	-86	-334	1091	116	71	15	339	234	248	261
-38.54166	73.95834	-225	39	12	9802	-48	-370	322	-227	-114	-87	-333	1097	116	71	15	340	235	250	264
-38.54166	73.12501	-225	37	11	9562	-52	-368	316	-228	-117	-91	-331	1217	130	77	15	376	258	272	297
-37.70833	72.70834	-225	36	11	9357	-56	-364	308	-227	-119	-93	-328	1285	137	79	15	396	270	284	317
-37.29166	72.70834	-225	36	11	9318	-57	-364	307	-227	-120	-94	-328	1290	137	79	15	396	272	286	318
-39.37499	73.95834	-224	39	12	9864	-46	-371	325	-227	-113	-85	-334	1091	115	71	15	337	232	246	258
-37.70833	73.95834	-224	39	12	9732	-49	-370	321	-226	-114	-87	-332	1099	116	70	15	339	235	251	264
-38.54166	73.54167	-224	38	11	9689	-49	-368	319	-226	-114	-88	-331	1149	122	74	15	355	245	259	265
38.12499	73.54167	-224	38	11	9643	-50	-368	318	-227	-115	-89	-331	1154	123	73	15	358	245	260	27
37.70833	73.54167	-224	38	11	9611	-51	-368	317	-227	-115	-89	-331	1159	123	73	15	358	246	262	28
37.29166	73.54167	-224	38	12	9541	-52	-367	315	-225	-116	-90	-330	1159	122	73	15	357	246	263	28
38.95833	73.12501	-224	37	11	9628	-50	-367	317	-227	-114	-88	-330	1199	128	77	15	371	255	268	2

col.env Only if reductype= “STANDARD” (else ignore this parameter), this is the parameter specifies the number of columns to trim environmental space on. This can be any number of columns.

EXAMPLE

col.env= c(3:10)

e.var= c(3:21)

e.var The selection of variables to include in all of the analyses. Note: this is a separate, more generalized parameter than **col.env**, but must contain all variables specified in **col.env**

Input Parameters:

sp1

sp2

Input Files

Species or Population Data: occurrence sites for species/population 1 (**sp1**) at study area 1 (**env1**) or species/population 2 (**sp2**) at study area 2 (**env2**). Column names should be 'sp','x',’y’. I typical import tab delimited (.txt) or comma separated files (.csv).

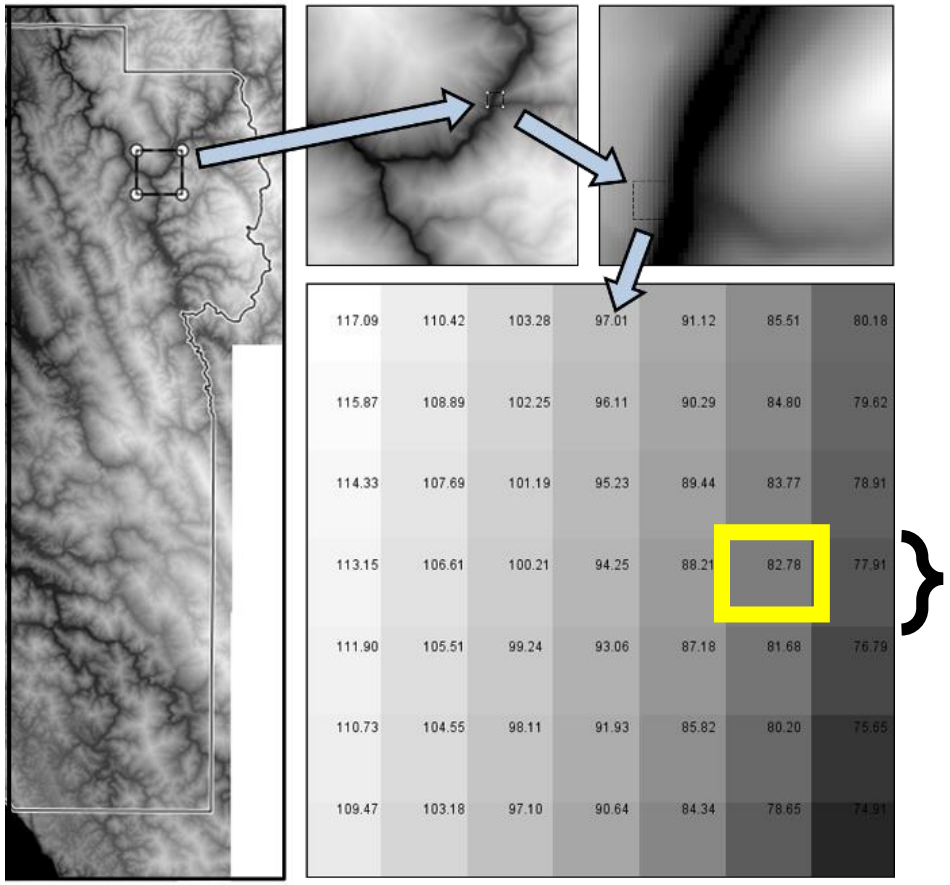
sp	x	y
Conium_maculatum_Native	-4.82	37.84
Conium_maculatum_Native	-4.84	38.74
Conium_maculatum_Native	-3.64944	40.22694
Conium_maculatum_Native	9.15	41.38333
Conium_maculatum_Native	-2.762	42.556
Conium_maculatum_Native	-2.626	42.916
Conium_maculatum_Native	3.62621	46.7645
Conium_maculatum_Native	3.46469	46.8137
Conium_maculatum_Native	3.27334	46.8646
Conium_maculatum_Native	3.09221	47.0176
Conium_maculatum_Native	1.64331	48.0519
Conium_maculatum_Native	3.72252	48.2556
Conium_maculatum_Native	2.39086	48.4391
Conium_maculatum_Native	1.96675	48.6847
Conium_maculatum_Native	9.96	49.77
Conium_maculatum_Native	-5.13957	49.992
Conium_maculatum_Native	-5.01148	50.17534

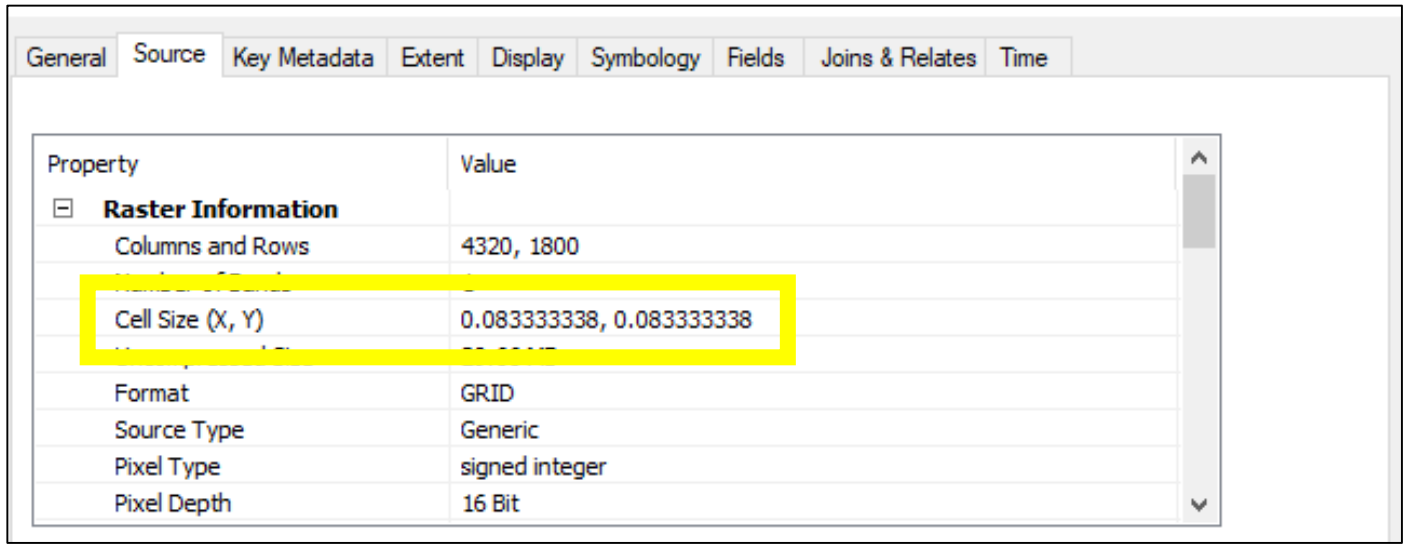
Input Parameter:

env.reso

Input Files

env.reso = the resolution of the input environmental data (**env1** and **env2**) in decimal degrees. This will be obtained from the raster GIS data used to create environmental data files (**env1** and **env2**)





0.083333338 decimal degrees

EXAMPLE

env.reso=0.41669

Note:

Here the raster data were clipped to the desired extent in ArcGIS. Then one of the clipped rasters was converted to a point dataset and latitude and longitude was added to this shapefile. Then all the raster layers were sampled using the point dataset. Lastly the final point shapefile was saved as a ".csv" or text file, cleaned up in Excel and the saved for importation into R.

Input Parameters:

rarefy.dist

rarefy.units

Preparing Occurrence Data

Why rarefy occurrence data? For optimal performance, this method requires input occurrence data to be spatially independent. It is common for researchers to introduce environmental biases into their analyses from spatially autocorrelated input occurrences. This causes niche estimation to be over-fit towards environmental bias resulting from sampling bias introduced from spatially clustered occurrences. The elimination of spatial clusters of localities is important for proper evaluation of a species niche. When spatial clusters of localities exist, species niches can be overly concentrated in certain E-space.

The **rarefy.dist** and **rarefy.units** parameters addresses this issue by spatially filtering locality data by a user input distance, reducing occurrence localities to a single point within the specified Euclidian distance.

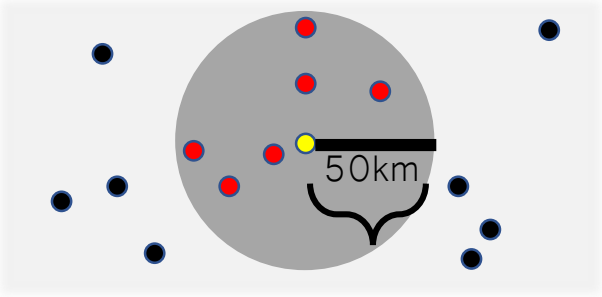
rarefy.dist Remove occurrences closer than a minimum distance to each other (this function uses the *humboldt.occ.rarefy* function). Values need to be in km[recommended] or decimal degrees. Note: rarefy.dist=0 will remove no occurrences

rarefy.units The units of **rarefy.dist** parameter, either "km" for kilometers or "dd" for decimal degrees

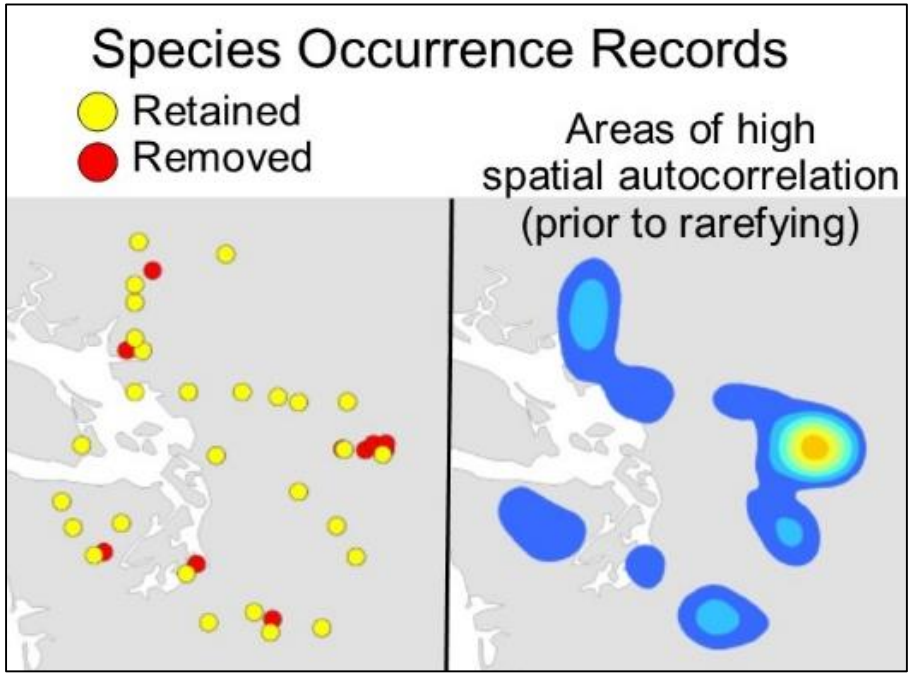
EXAMPLE

rarefy.dist= 50

rarefy.unts="km"



rarefy distance



Input Parameters:

env.trim

env.trim.type

trim.buffer.sp1

trim.buffer.sp2

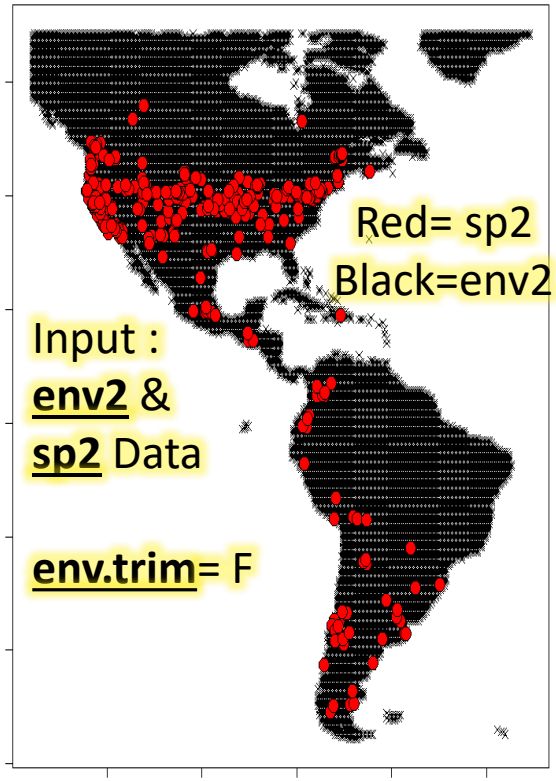
trim.mask1

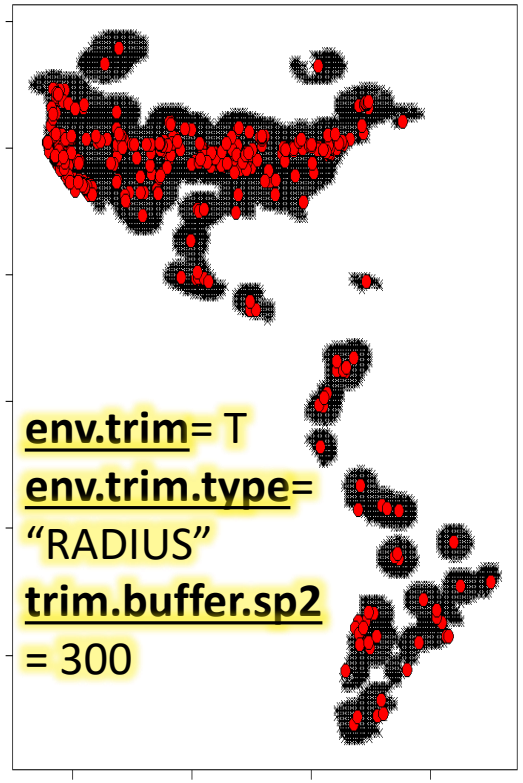
trim.mask2

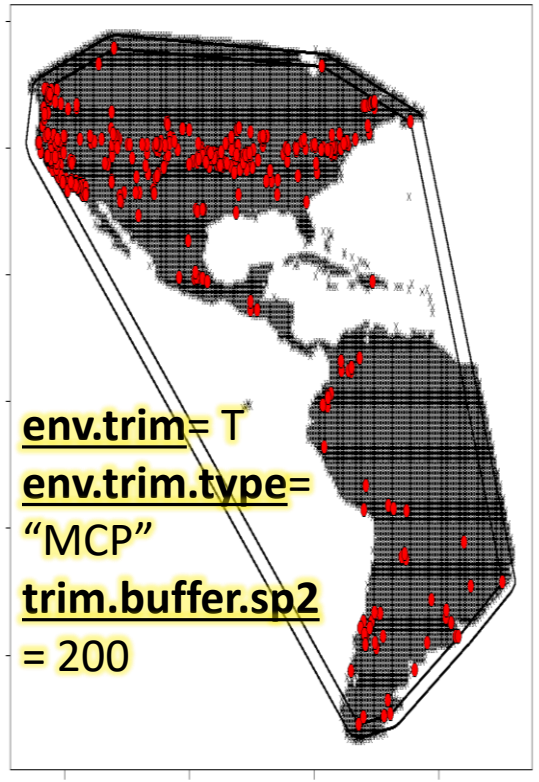
Accessible Environments

Why trim environments input? The niche divergence test requires input environments to be reduced to ONLY habitats that are accessible to the species in geography. In this test it is important to clarify the habitat a species currently exists in and the availability of adjacent habitats. Humboldt present three ways to do this: via a buffered minimum convex polygon of input occurrence localities, buffering input occurrence localities, & via a user input mask (an advanced method using an input polygon shapefile WGS1894, input as **trim.mask1** or **trim.mask2** = "R environment name").

env.trim.type = "MCP"
trim.buffer.sp1 = 300

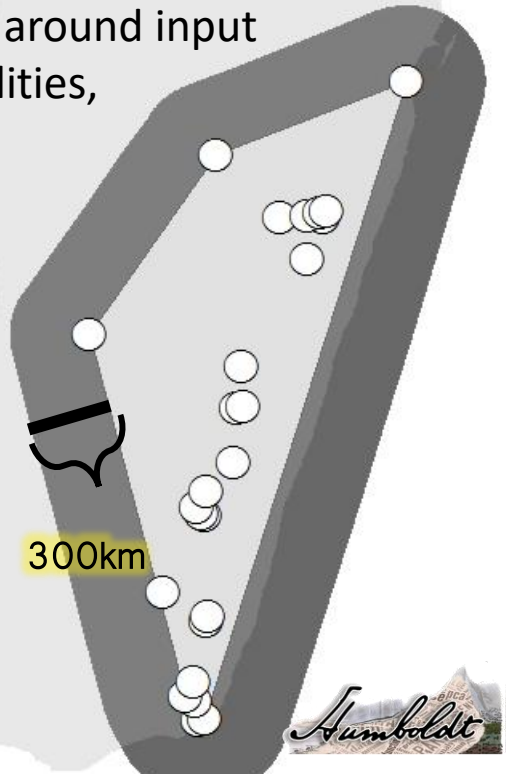






These parameters creates a minimum-convex-polygon around input occurrence localities, which is then buffered by the parameter: **trim.buffer.sp1**

Here the buffer Is 300km.



Input Parameters:

reduce.env

reductype

Analogous and Non-analogous E-space

Why do analogous and non-analogous environments matter? The niche divergence test requires input environments to be only analogous environments. Thus, environmental space that is **only shared** among habitats is used in the quantitative tests of niche similarity. Thereby differences between species are not solely due to different access to environments that can lead to the false appearance of different niches.

Step 1: Trim E-space Limits

reduce.env The format to trim environments input to analogous space. If reduce.env=1, the second input environment (env2) will be trim the match the first input (env1). If reduce.env=2, both input environments trimmed so that extents of both are identical (the lower maximum value observed in env1 and env2 and the higher minimum value observed in env1 and env2 will be used to trim environmental space for each PC/environmental variable) If reduce.env=0, you will skip trimming environmental space

reducetype Only used if **reduce.env**= 1 or 2. The **reducetype** parameter specifies the format for how to reduce environmental space ("PCA" or "STANDARD"). If **reductype**="PCA", the environmental space will be trimmed based on two principal components. If **reductype**="STANDARD", the environmental space will be trimmed by each included variable specified in **col.env**.

EXAMPLE

reduce.env= 2

reducetype="PCA"

Input Parameters:

reduce.env

reductype

Step 1: Trim E-space Limits

Below are examples of how the **reduce.env** and **reductype** parameters trim the range of environments. Note this only limits the max and min values of environments to similar ranges. Unique combinations of E-space not shared between environments (non-analogous) likely still exist.

	Post **env.trim** Environments		**reduce.env**=0 **reducetype**=NA		**reduce.env**=1 **reducetype**= "PCA"		**reduce.env**=1 **reducetype**= "STANDARD"		**reduce.env**=2 **reducetype**= "PCA"		**reduce.env**=2 **reducetype**= "STANDARD"	
	Env1 Input	Env2 Input	Env1 Output	Env2 Output	Env1 Output	Env2 Output	Env1 Output	Env2 Output	Env1 Output	Env2 Output	Env1 Output	Env2 Output
Bio1: max value	291	314	291	314	-	-	291	291	-	-	291	291
Bio1: min value	-226	-177	-226	-177	-	-	-226	-177	-	-	-177	-177
Bio12: max value	2729	8463	2729	8463	-	-	2729	2729	-	-	2729	2729
Bio13: min value	5	0	5	0	-	-	5	0	-	-	5	5
PC1: max value	5.12	3.41	5.12	3.41	5.12	3.41	-	-	3.41	3.41	-	-
PC1: min value	-3.35	-4.17	-3.35	-4.17	-3.35	-3.35	-	-	-3.35	-3.35	-	-
PC2: max value	7.54	2.17	7.54	2.17	7.54	2.17	-	-	2.17	2.17	-	-
PC2: min value	-1.24	-12.1	-1.24	-12.1	-1.24	-1.24	-	-	-1.24	-1.24	-	-

EXAMPLE `reduce.env= 2, reducetype="STANDARD", col.env=c(3,14), e.var= c(3:21)`

Step 2: Remove Non-Analogous E-space

Input Parameters:

reduce.env

reductype

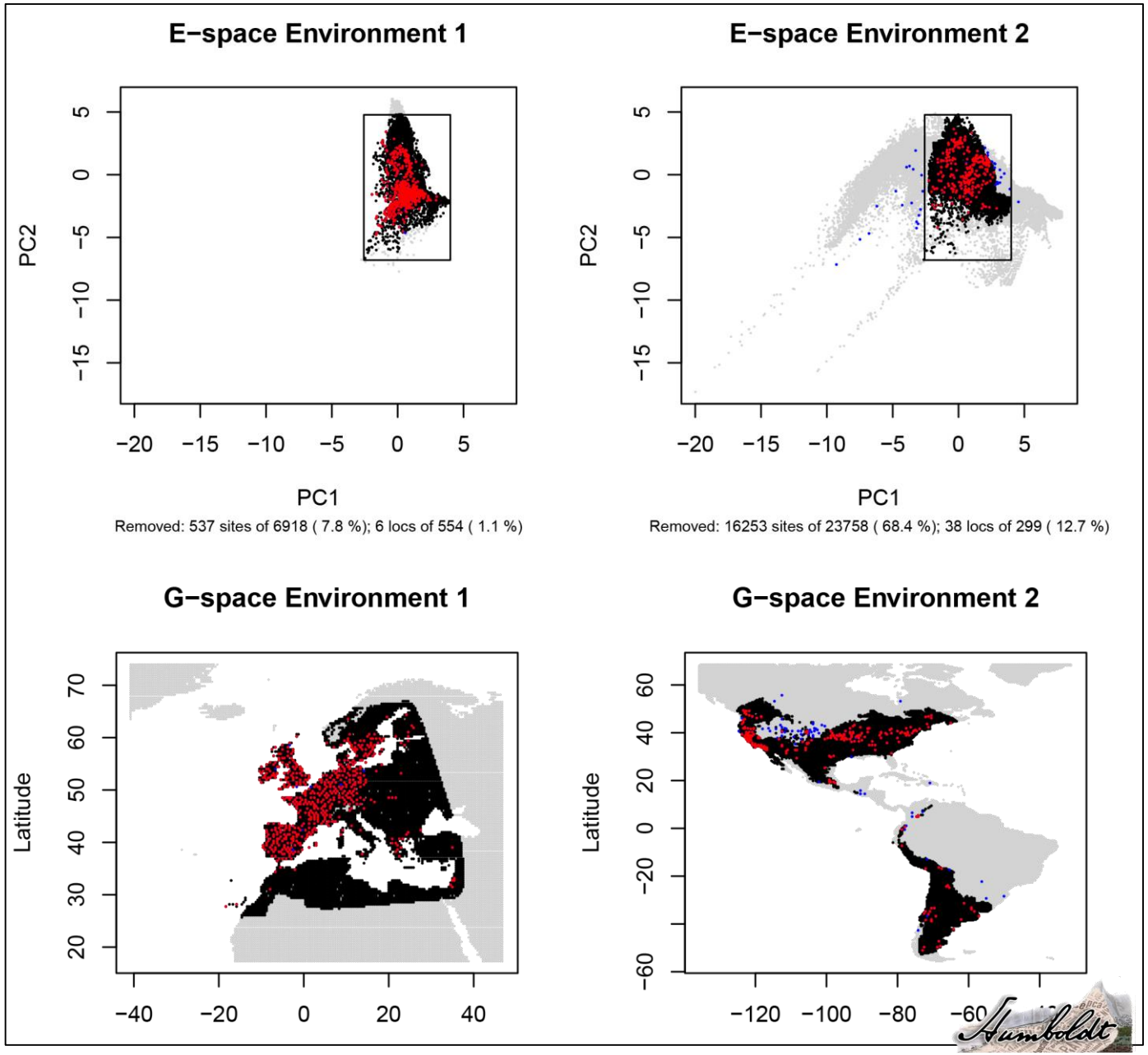
non.analogous.environments

Removing non-analogous environments

Step 1: Trim E-space Limits
Step 2: Remove non-analogous E-space

The second step is to remove E-space that is not identical among both habitats (option only available for **reductype**= "PCA"). The **reduce.env** parameter trims habitats to shared max and mins, however often, combinations of PC1 & PC2 values are unique to one habitat and cannot be occupied by species in the other habitat.

- Occurrence localities **retained** after E-space is reduced to analogous environments
- Occurrence localities **removed** after E-space is reduced to analogous environments
- E-space **retained** after after input environments are reduced to analogous environments
- E-space **removed** after input environments are reduced to analogous environments**



Analogous and Non-analogous E-space

Input Parameters:

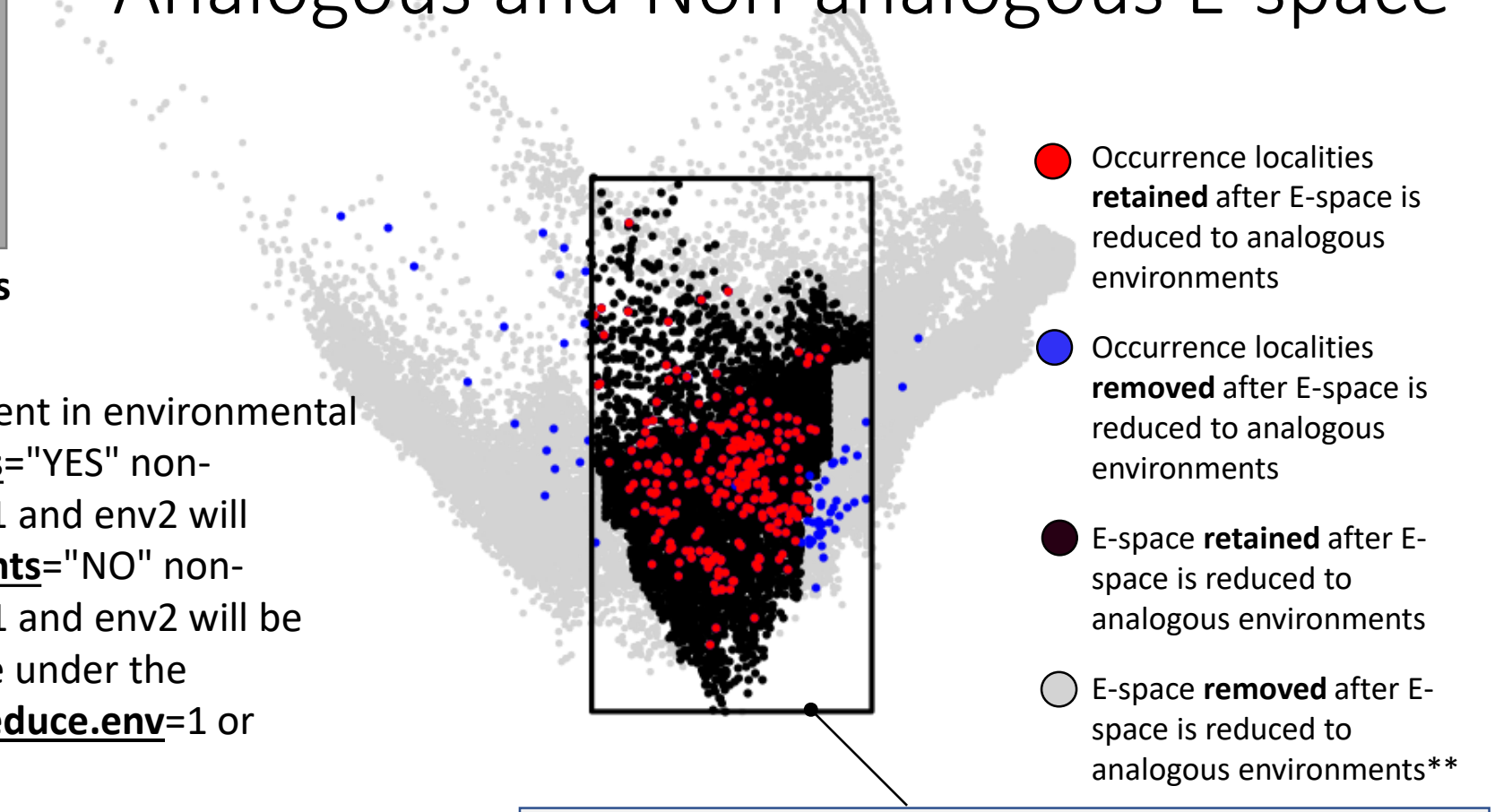
reduce.env

reductype

non.analogous.environments

Removing non-analogous environments

Do you allow non-analogous environment in environmental space? If **non.analogous.environments**="YES" non-analogous environments between env1 and env2 will retained. If **non.analogous.environments**="NO" non-analogous environments between env1 and env2 will be removed. This parameter is only usable under the combinations of **reductype**="PCA" & **reduce.env**=1 or **reductype**="PCA" & **reduce.env**=2



- Occurrence localities **retained** after E-space is reduced to analogous environments
- Occurrence localities **removed** after E-space is reduced to analogous environments
- E-space **retained** after E-space is reduced to analogous environments
- E-space **removed** after E-space is reduced to analogous environments**

Black box depicts shared E-space (based only on maximum and minimum values of each PCs). If **non.analogous.environments**= "YES", all values within this box will be retained** If **non.analogous.environments**= "NO", only areas shared between both environments are retained within this box.

EXAMPLE

reduce.env= 2

reducetype="PCA"

non.analogous.environments="YES"

***points can also be removed via the **env.trim** parameter, which trims input environments in geographic space to a buffered minimum convex polygon of the focal species occurrence localities. This reduces the input environmental space to accessible habitats to the focal species.*

Input Parameters:

nae.window

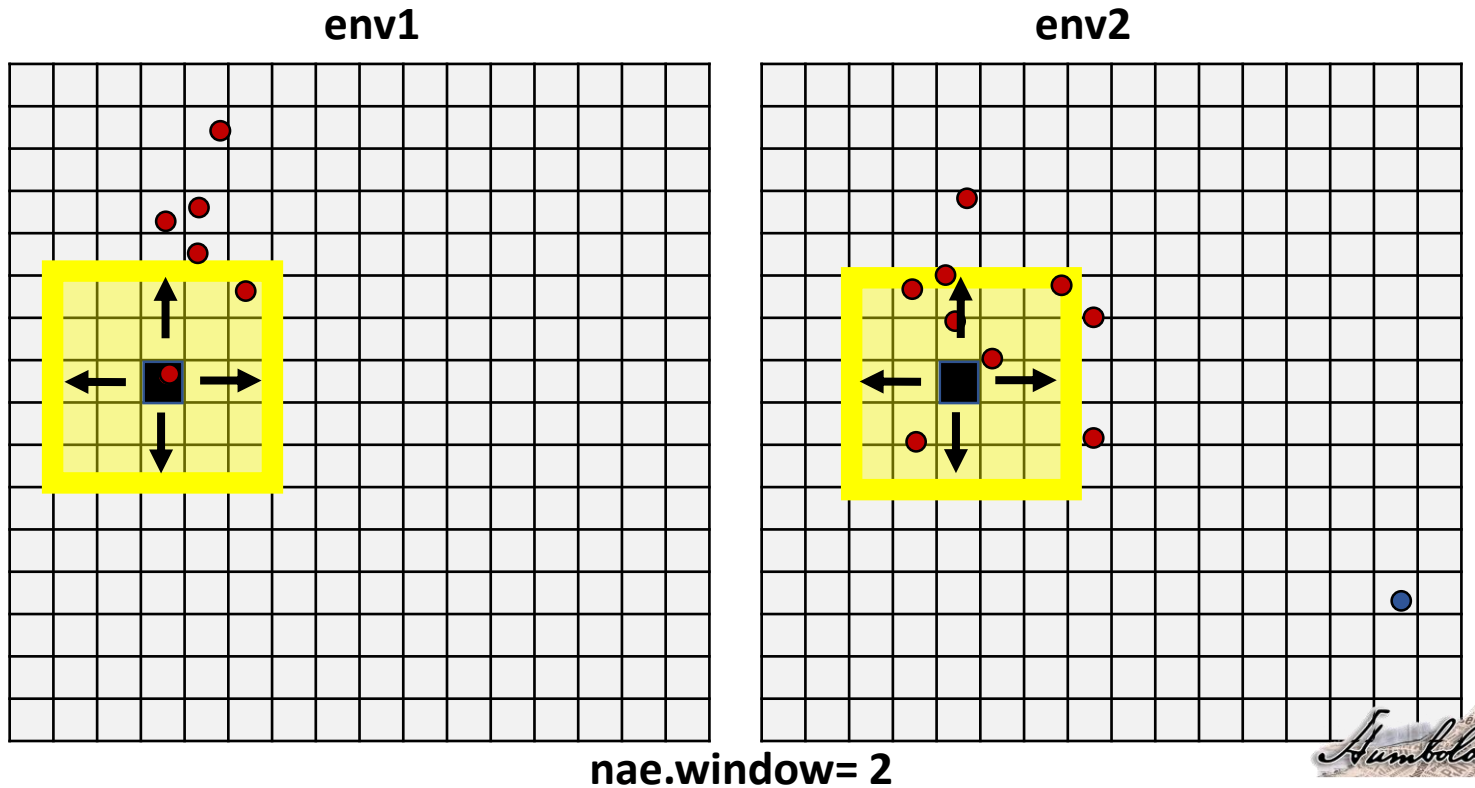
Analogous and Non-analogous E-space

Fine tuning the removing non-analogous environments

The trimming of E-space is a bit of an art. Do points need to be perfect matches for them to be analogous or do you allow for a little variation? The **nae.window** parameter depicts of how tight or generalized the overlap of data need to be to be considered analogous. The **nae.window** parameter depicts the spatial window from which non-analogous environments will be quantified. The non-analogous environments are characterized by gridding the E-space of **env1** and **env2** into a **R** x **R** grid (e.g. 100 x 100). If **nae.window** parameter= 0, points absent from a cell in one environment will be removed from the other. If **nae.window**>0, values absent from a window (or neighborhood of cells) in one environment will be trimmed from the other. The **nae.window** value characterizes the number of cells to search from the focal cell of environmental space values in the other environment. The larger the **nae.window** value, the fewer non-analogous environments removed. This parameter allows imperfect overlap of environments. If areas of environmental space are a little patchy between environments—but generally present– a larger **nae.window** value will retain more of the patch environments. The default value is a **nae.window**=5

EXAMPLE

nae.window= 2



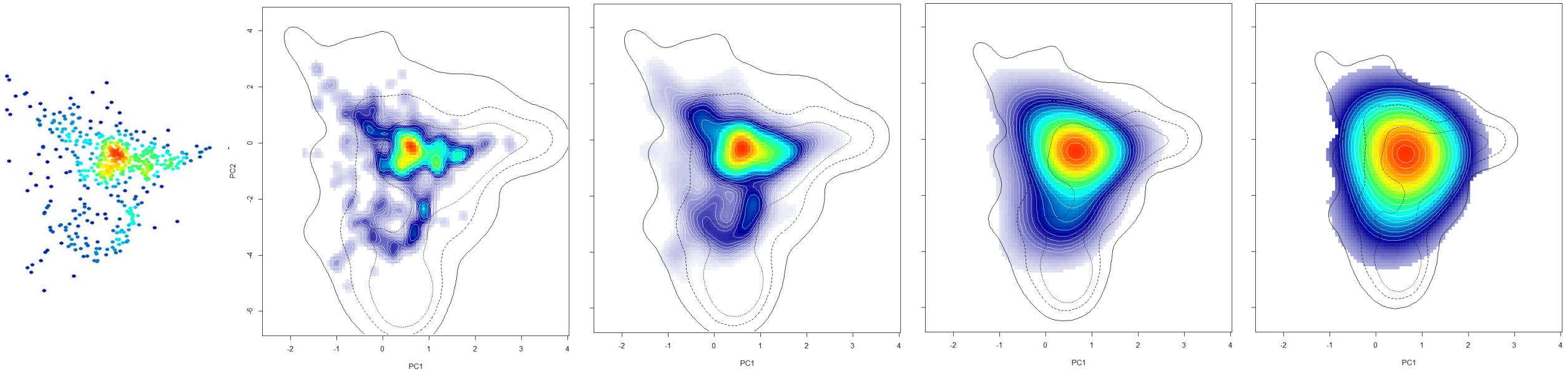
Input Parameters:

kern.smooth

Quantifying Niches

The point data need to be converted to a continuous surface to quantify the species' niches. There are three parameters that control: the shape of the surface (**kern.smooth**), the spatial resolution of the E-space (**R**), and the threshold at with low suitability habitats are considered part of the niche (**thresh.espace.z**).

The **kern.smooth** parameter dictates the scale at which kernel smoothing occurs on environmental data. Larger values (i.e. 2) increase scale (making E-space transitions smoother and typically larger) and smaller values (i.e. 0.5) decrease scale (making occupied E-space clusters more dense and irregular). The default value is 1.



Raw Data | **kern.smooth**=0.5 | **kern.smooth**=1 | **kern.smooth**=2 | **kern.smooth**=3

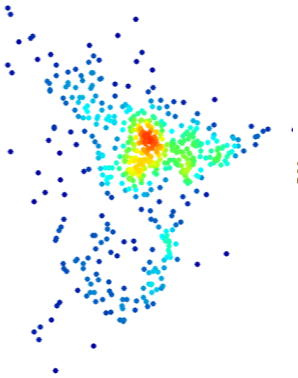
Input Parameters:

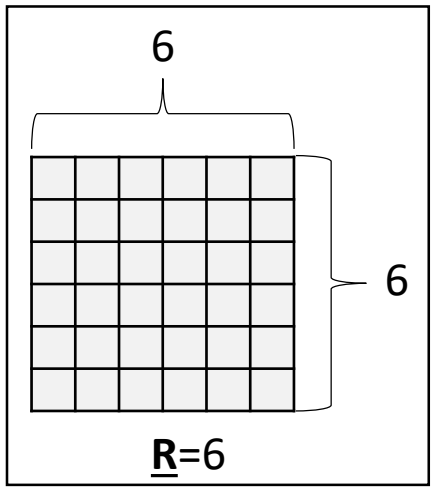
R

Quantifying Niches

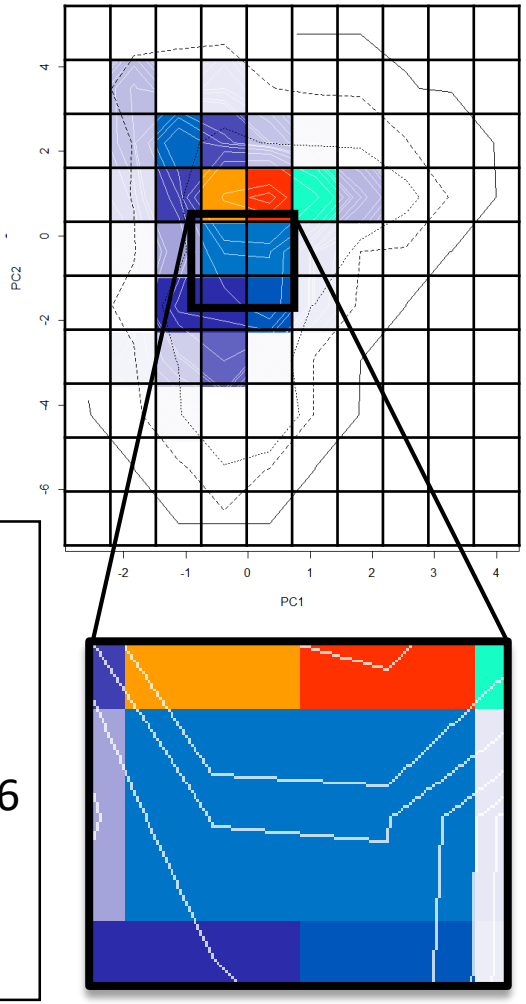
R is the resolution of grid in environmental space (RxR). The default value is 100. The larger this value is, the slower the analyses will be. Thus, you want to hit the sweet spot of having high enough, but not too high.

Raw Data

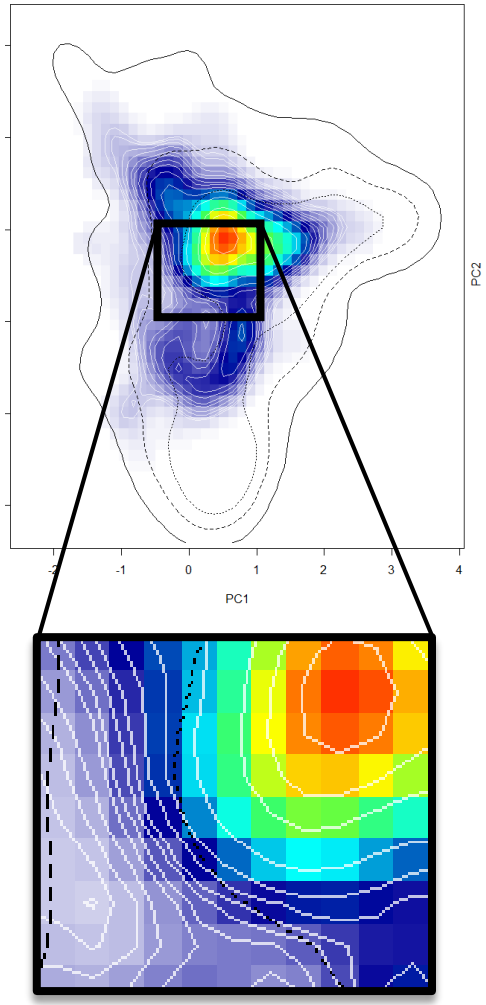




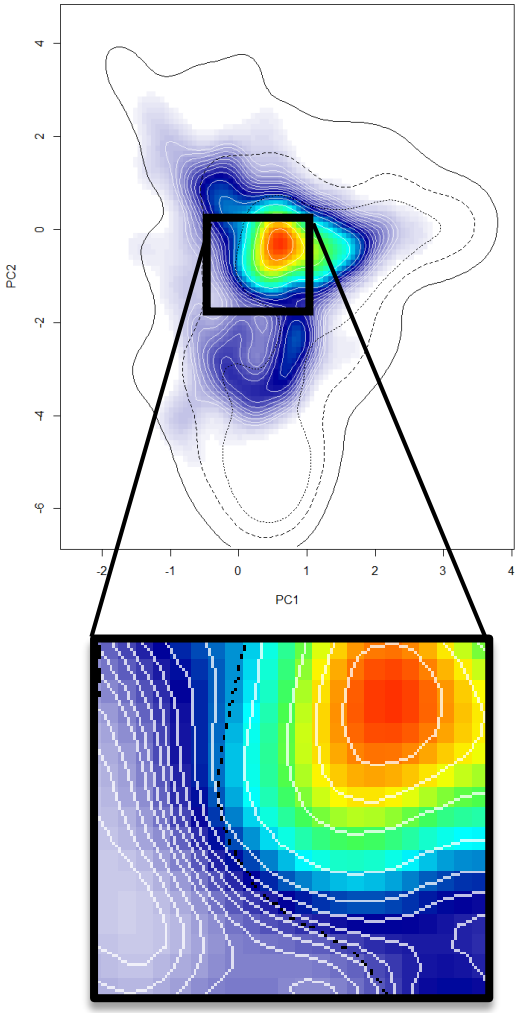
R=10



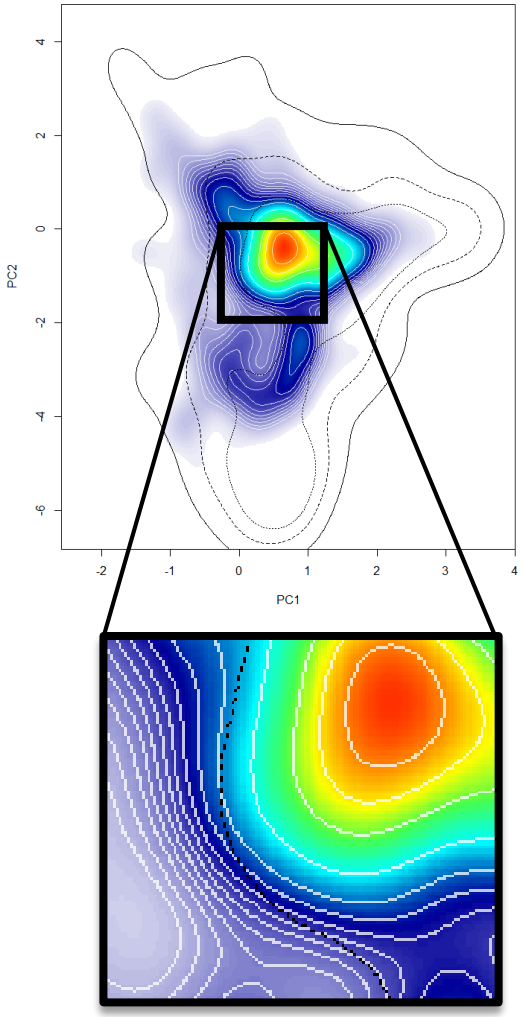
R=50



R=100



R=500

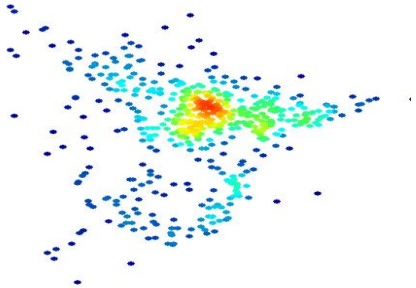


Input Parameters:

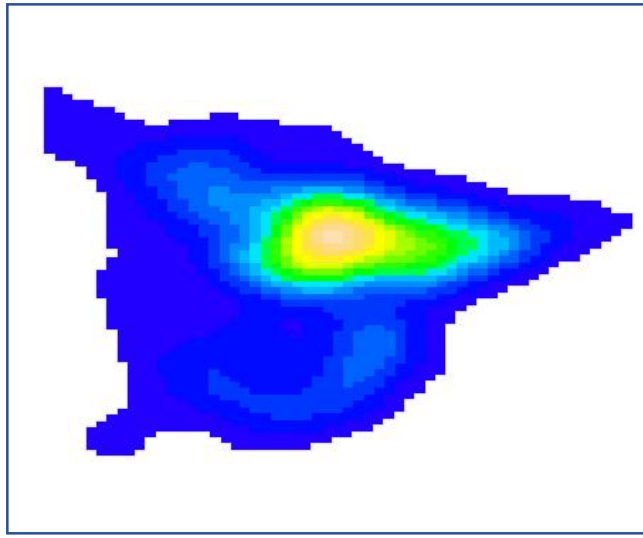
thresh.espace.z

Quantifying Niches

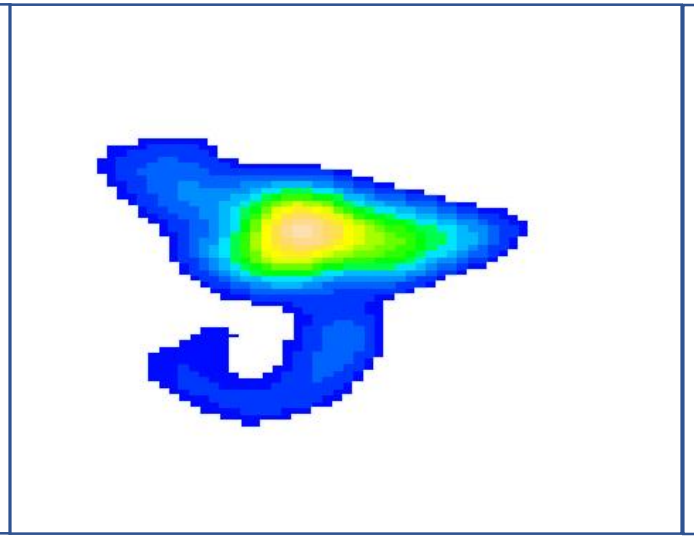
The **thresh.espace.z** parameter is an experimental parameter and controls the level at which values below the kernel density z values are removed for comparison and quantification of niches. Higher values will increase the value from which the low-density areas are removed from the environmental space. Basically values above this are retained and values below are removed. Input values should range from 0.0001-0. Default=0.001



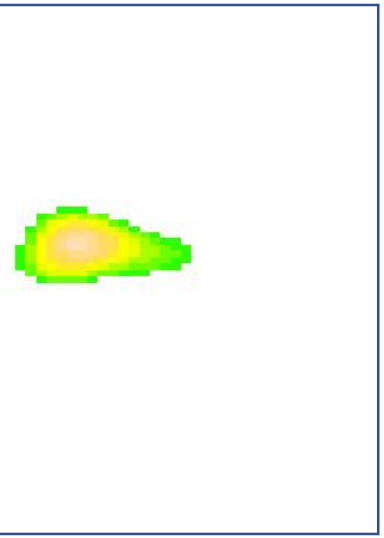
Raw Data



thresh.espace.z=0.001



thresh.espace.z=0.01



thresh.espace.z=0.1

Input Parameters:

p.overlap

Plotting Both Niches in a Single Plot

When **p.overlap**=T, both species niches are plotted in a single plot.

Histogram Density Plots

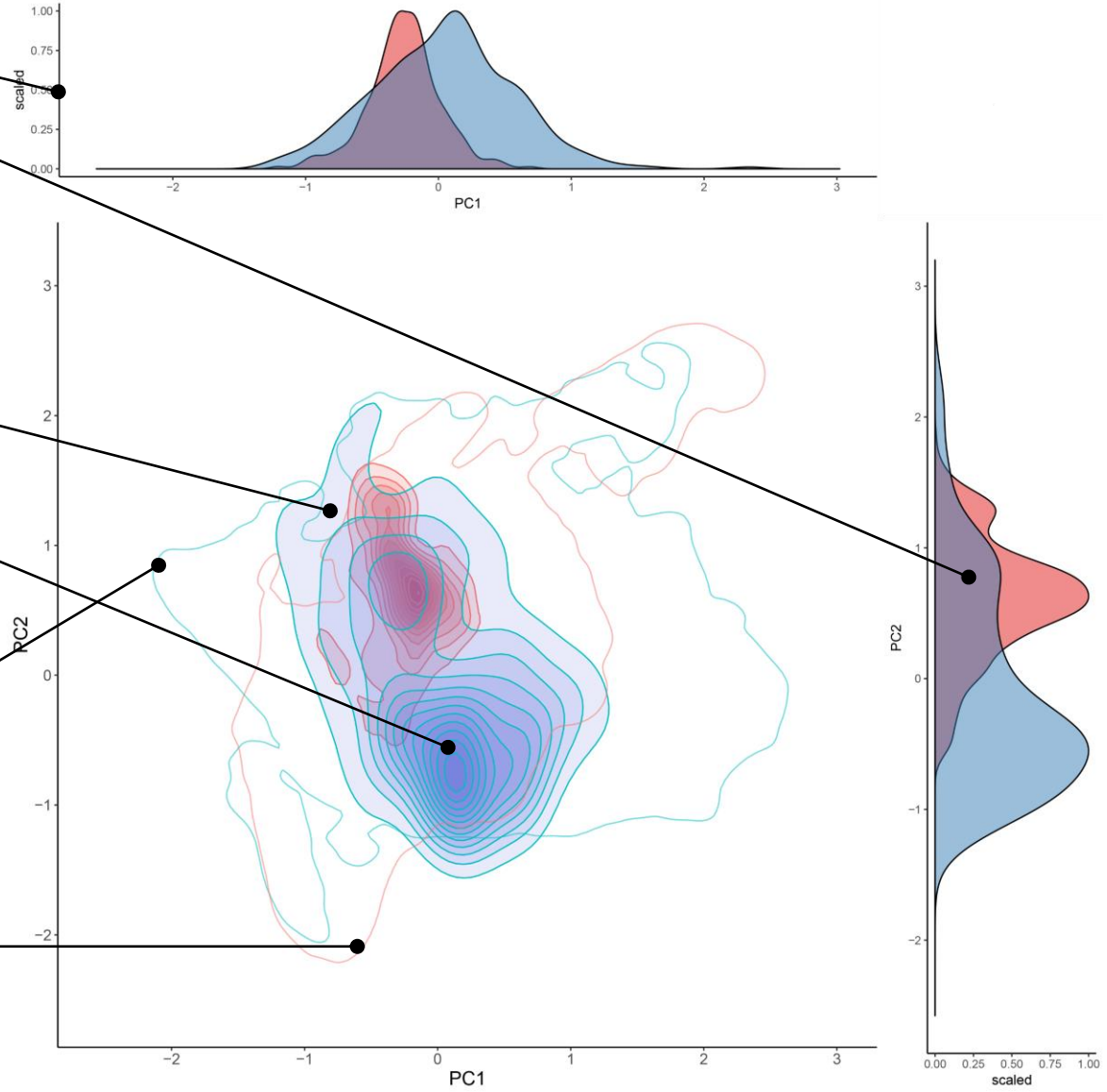
For each PC, the density of each species E-space is displayed

Filled Kernel Density Isopleths
(red= sp1 and blue= sp2*)

Lines representing the kernel density isopleths from 1-100% kernel densities. The number of bins is set at default and adjusts so that the number contour bins best display the data. Here 11 were selected, thus, as follows are the plotted kernel isopleths: 0.01,0.1,0.2,0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ,1

Empty Kernel Density Isopleths
(red= env1 and blue= env2*)

Lines representing the 1% kernel density isopleths of the environments (not species) depicting outside boundaries of E-space. Sometimes species values erroneously exceed these due to the smooth parameter of kernel densities estimates.



if swap=T, colors will be switched in plots

Input Parameters:

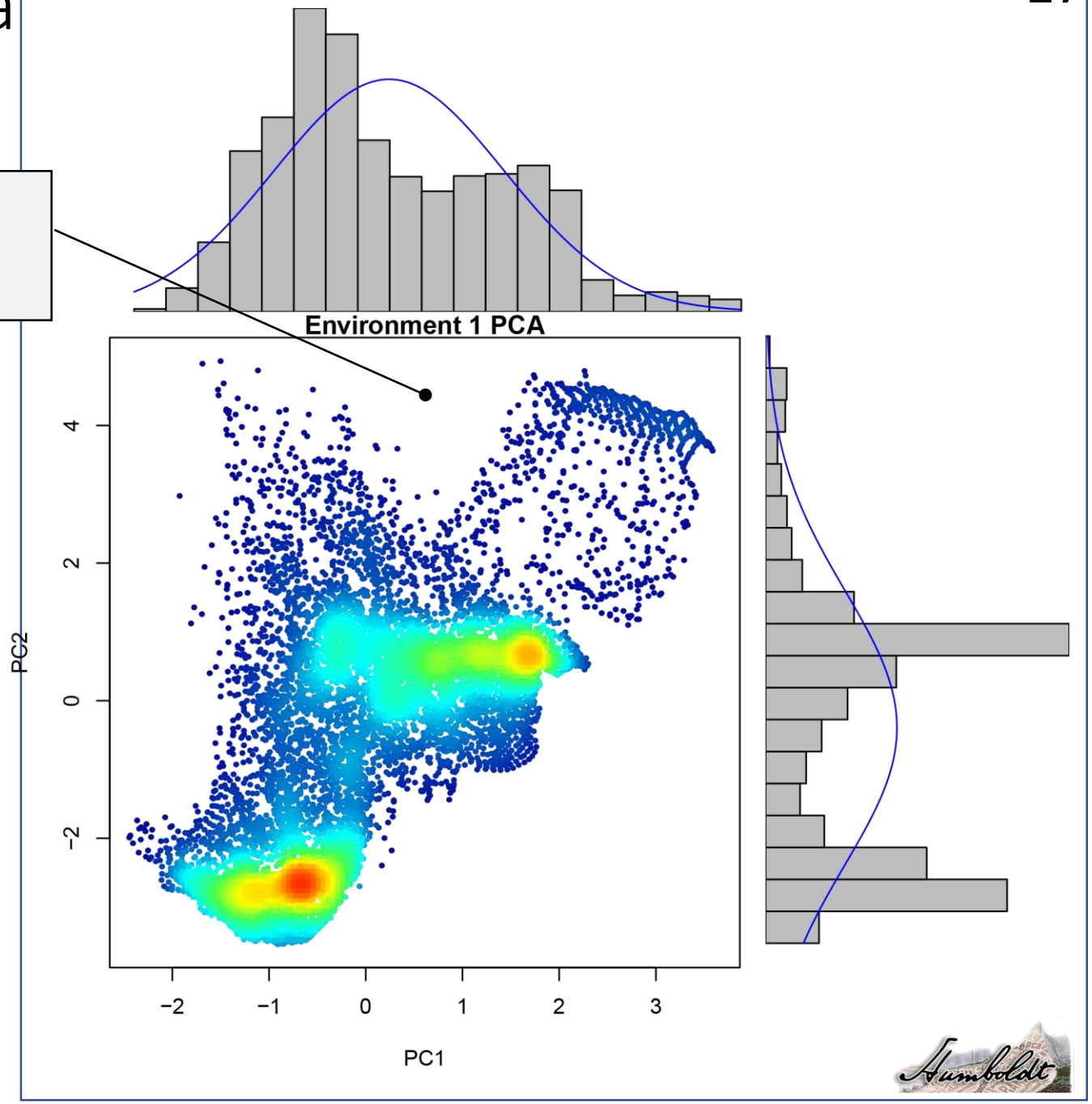
p.scatter

Plotting Raw Data

When **p.scatter**=T, raw analysis data is plotted in a series of plots

Density Plots of Raw E-space Data

A plot of raw data of species' and environment E-space. Cooler colors represent lower and warmer higher densities of E-space.



Input Parameters:

p.boxplot

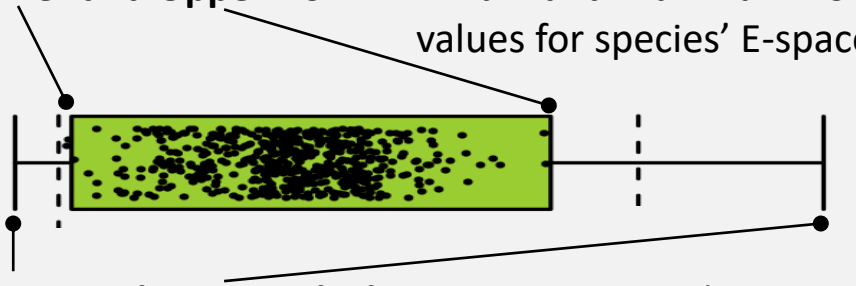
Plotting E-space Boxplots

When **p.boxplot**=T, both species E-space are plotted

E-space Box Plots

For each PC, the one dimensional E-space is Displayed.

Lower and Upper Box: Minimum and Maximum PC values for species' E-space



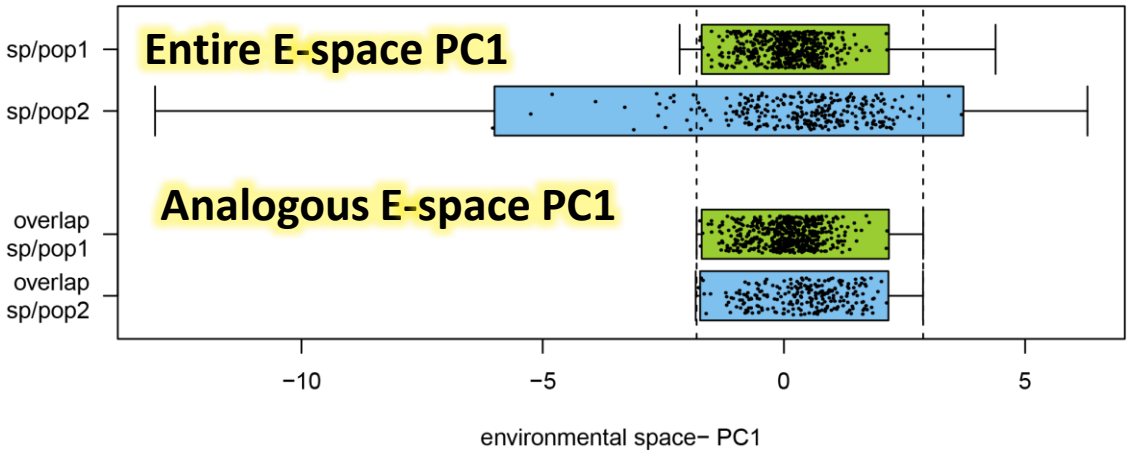
Lower and Upper Whiskers: Minimum and Maximum PC values for each environment' E-space

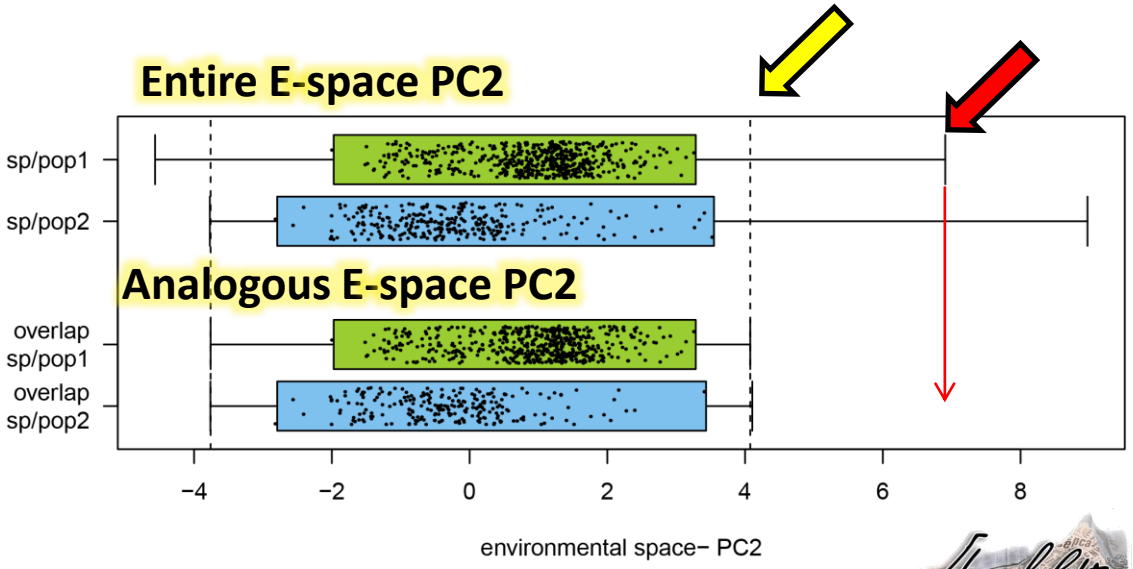
Dotted lines: Maximum and Minimum value of shared (analogous) E-space of both environments

Dots: PC values for each occurrence record

E-space truncation in 1 dimension can affect other

Sometimes the reduction of E-space in PC1, results in truncation of E-space in PC2. In absence of trimming E-space of PC1, the upper shared limit of E-space for PC2 should be ~7 (red arrow). Those values were lost in trimming PC1 and within the remaining analogous E-space, ~4 (yellow arrow) becomes the shared upper limit for PC2.

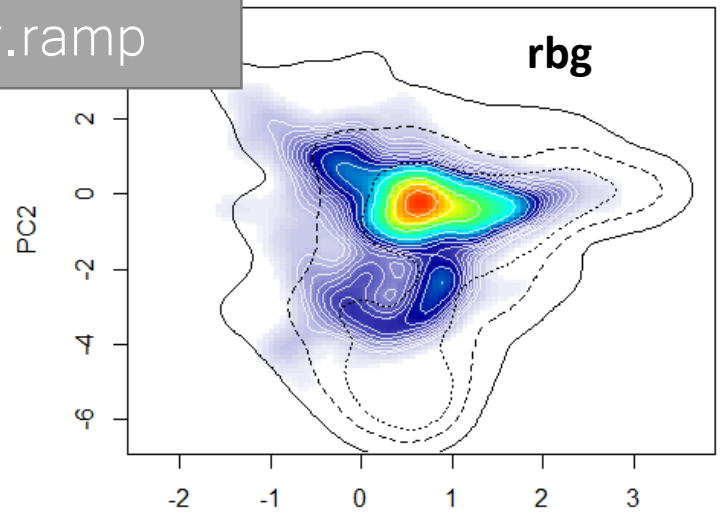




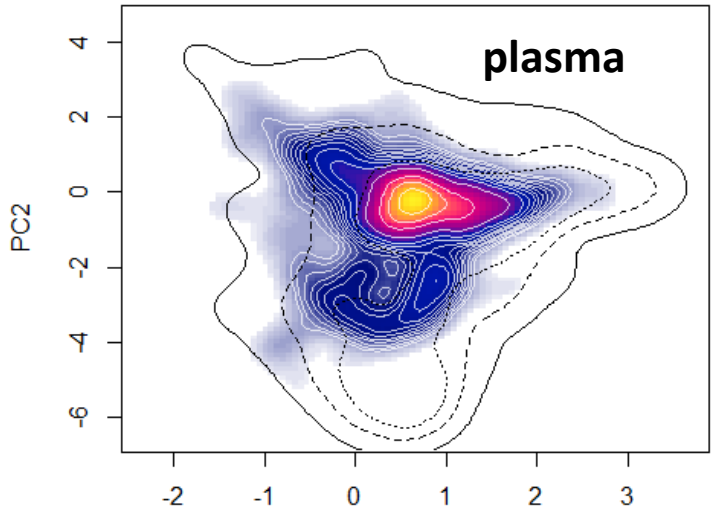
Input Parameters:

color.ramp

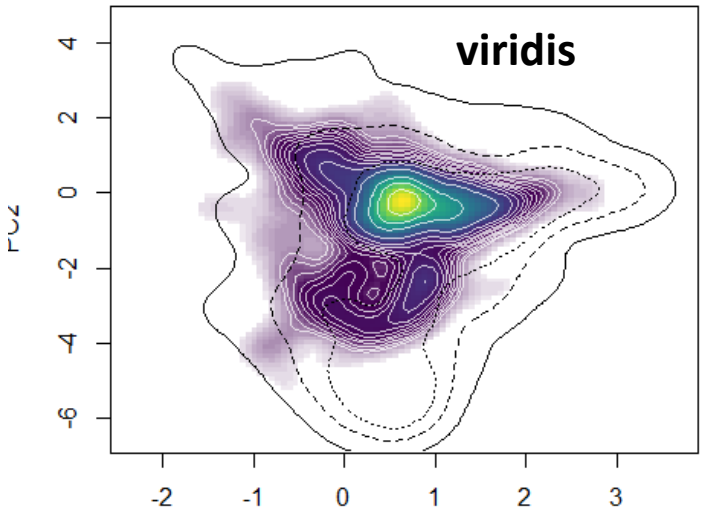
Color Schemes



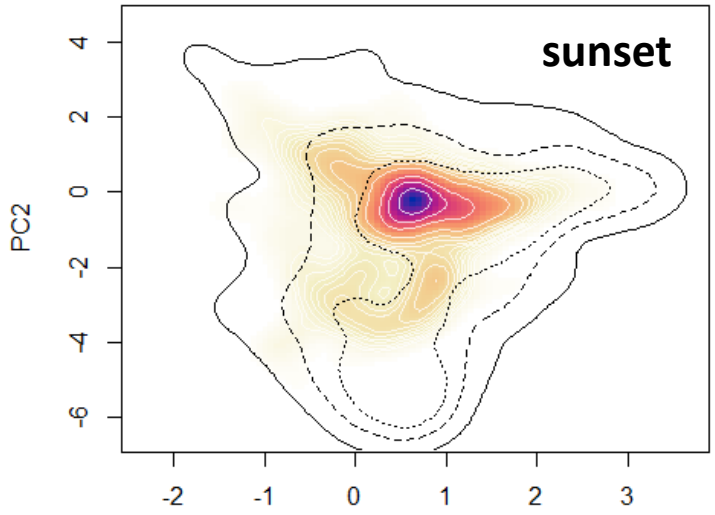
color.ramp= 1 (default)



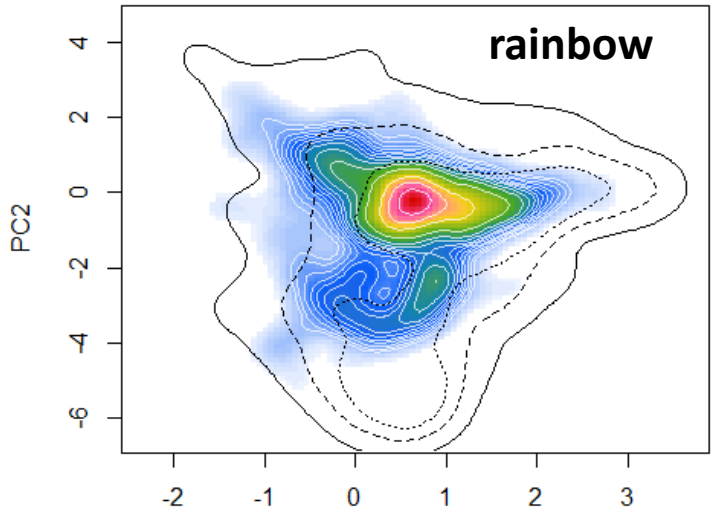
color.ramp= 2



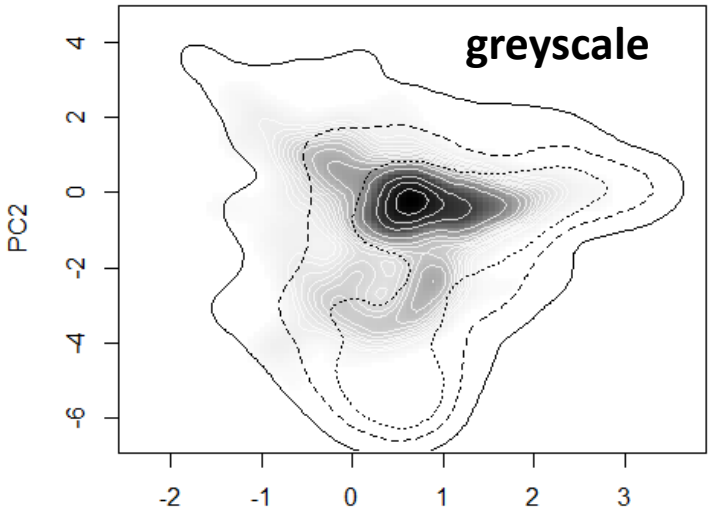
color.ramp= 3



color.ramp= 4



color.ramp= 5



color.ramp= 6