

## SDM-Workflow – Nutzerhandbuch

### 1. Benötigte Software

#### R

<https://www.r-project.org/>

#### RStudio

<https://www.rstudio.com/products/rstudio/>

### 2. R-Skripte für den Workflow

<https://github.com/LarissaNow/SDM-workflow-CASPIAN>

### 3. Betriebssystem

Dieser Workflow wurde mit einem Windows-System mit Englischem Zahlenformat geschrieben und bisher ausschließlich für selbiges getestet.

### 4. Installation der benötigten Software

Sie benötigen zum Ausführen des Workflows die Programmiersprache R, sowie die Software RStudio. Falls Sie diese noch nicht auf Ihrem Rechner installiert haben, laden Sie bitte die neusten Versionen unter den obenstehenden Links herunter und installieren Sie sie.

### 5. Download der benötigten R-Skripte und Daten

Rufen Sie den oben bereit gestellten Github-Link auf und laden Sie alle dort bereit gestellten R-Skripte, das R-Project „SDMs“, sowie die Landbedeckungsdaten und (falls gewünscht) das Shapefile mit dem Umriss von Deutschland herunter. Stellen Sie sicher, dass alle Dateien auf Ihrem Rechner in demselben Ordner hinterlegt sind.

### 6. Öffnen der notwendigen Skripte und Dateneingabe

Öffnen Sie zunächst das R-Project „SDMs“. Innerhalb dieses öffnen Sie nun das R-Skript „run\_SDM\_workflow“. Geben Sie nun in Zeile 31 – 56 des R-Skriptes alle benötigten Informationen ein.

- ✓ **identifier:** Hier sollten Sie für jeden Lauf des Skriptes einen eindeutigen Namen eingeben. Dies kann beispielsweise eine Kombination aus dem Artnamen und dem Datum sein. Die Eingabe sollte als character erfolgen, was durch Anführungszeichen angegeben wird.
- ✓ **mypespecies:** Geben Sie hier bitte den wissenschaftlichen Artnamen der Art von Interesse ein. Dies sollte in folgendem Format erfolgen: „Gattung art“, z.B. „Lonicera japonica“.
- ✓ **occ:** Hier können Sie optional eine Tabelle mit Vorkommenspunkten der Art von Interesse einlesen. Diese sollte im .csv-Format (Komma-getrennt), und folgendermaßen aufgebaut sein:

	decimalLongitude	decimalLatitude	Presence
1	-96.818884	33.01209	1
2	-73.746405	40.752605	1
3	-77.522455	39.273881	1
4	-86.235125	32.360087	1
5	-73.607063	40.627328	1
6	-79.979987	39.587404	1

Anstelle eigener Daten kann hier auch die im ersten Schritt des SDM-Workflows erstellte Tabelle mit Vorkommenspunkten eingelesen werden, beispielsweise nach Unterbrechung der Berechnungen oder wenn Sie diese Tabelle händisch editiert haben.

- ✓ **envir**: Geben Sie hier bitte den Namen der Klimavariablen von Interesse als character ein. Die Auswahl der verschiedenen Klimavariablen finden Sie im Anhang dieses Nutzerhandbuchs.
- ✓ **landcov**: Geben Sie hier bitte den Namen der Landbedeckungsklassen von Interesse als character ein. Die Auswahl der verschiedenen Landbedeckungsklassen finden Sie im Anhang dieses Nutzerhandbuchs.
- ✓ **filecropbase**: Hier können Sie optional eine geographische Ausdehnung für die Projektion der Umwelteignung angeben. Falls Sie hier NULL angeben, wird global/Europaweit projiziert.
- ✓ **GermanShapefile**: Hier können Sie optional das Shapefile mit dem Umriss von Deutschland laden, wenn Sie die Abbildungen der Umwelteignung, die im Laufe des Workflows erstellt werden, auf Deutschland beschränken möchten
- ✓ In Zeile 52 können Sie optional die Liste mit den Abwesenheitspunkten laden, falls Sie diesen Schritt bereits laufen gelassen haben. Ansonsten führen Sie diese Zeile bitte nicht aus.
- ✓ In Zeile 54 können Sie optional die Liste mit der Umwelteignung laden, falls Sie diesen Schritt bereits laufen gelassen haben. Ansonsten führen Sie diese Zeile bitte nicht aus.
- ✓ **netshp**: Hier können Sie das Shapefile mit dem Verkehrsnetz laden, für das im letzten Schritt des Workflows die mittlere Umwelteignung je Segment berechnet wird.

## 7. Ausführen des Skriptes

Es ist zu empfehlen das R-Skript „run\_SDM\_workflow“ Zeile für Zeile auszuführen und auf mögliche Fehler und Benachrichtigungen in der Konsole zu achten. Beginnen Sie mit Zeile 16 – 22 und 28, führen Sie dann alle relevanten Zeilen der Dateneingabe aus (s.h. oben) und fahren Sie dann Zeile für Zeile mit den unterschiedlichen Arbeitsschritten fort. Achten Sie hierbei auch auf die Annotationen im Skript, da diese wichtige Informationen und weitere Erklärungen enthalten können.

## 8. Dateien die im Laufe des Workflows auf Ihrem Computer gespeichert werden

- ✓ **output**: Ein Ordner in dem alle Dateien abgelegt werden, die im Laufe des Workflows entstehen
- ✓ **occtemp.csv**: Von GBIF herunter geladene Vorkommensdaten in Tabellenformat (als csv-Datei)
- ✓ **PAlist**, „**identifizier**“, **.RData**: Optional. Liste mit den fünf Abwesenheitsdatensets
- ✓ **suitability**, „**identifizier**“, **.RData**: Optional. Liste mit der Umwelteignung je Modelllauf
- ✓ **averageSuitability\_**, „**identifizier**“, **.csv**: Mittlere Umwelteignung in Tabellenformat (als csv-Datei)
- ✓ **Suitability\_**, „**identifizier**“, **.pdf**: Abbildung der modellierten mittleren Umwelteignung
- ✓ **SuitabilityOccurrences\_**, „**identifizier**“, **.pdf**: Abbildung der modellierten mittleren Umwelteignung mit den Vorkommensdaten
- ✓ **Traffic\_net\_env\_**, „**identifizier**“, **.shp**, **.dbf**, **.shx**, **.prj**: Optional. Shapefile mit dem Verkehrsnetz und der jeweiligen mittleren Umwelteignung für jedes Verkehrssegment, sowie der Standardabweichung der Umwelteignung je Verkehrssegment

## 9. Hintergrund

SDM steht für Species Distribution Model. Dies sind Modelle, die basierend auf Vorkommensdaten von Arten und assoziierten Umweltvariablen berechnen können, wie hoch die Wahrscheinlichkeit ist, dass eine Art in einer bestimmten Gegend vorkommt, was als Annäherung an die Umwelteignung dieser Gegend für die Art verstanden werden kann. SDMs können auch genutzt werden um Zukunftsprojektionen des Vorkommens der Art zu erstellen. SDMs können mit einer Vielzahl unterschiedliche Umweltvariablen informiert werden. Hier nutzen wir Klimavariablen, sowie Informationen zur Landbedeckung. Die Ausbreitungsfähigkeit der Arten ist explizit nicht Teil des hier präsentierten SDM Workflows, da dieser mit einem Modell zur Ausbreitung der Arten entlang von Verkehrsnetzen verknüpft werden kann. Es ist wichtig im Hinterkopf zu behalten, dass die Ergebnisse von SDMs immer mit Vorsicht interpretiert werden müssen! Zum einen kann mit Vorkommensdaten ausschließlich die realisierte Nische einer Art, nicht aber ihre fundamentale Nische abgeschätzt werden. Zum anderen können die Ergebnisse von SDMs ungenau oder fehlerhaft sein, wenn die zu Grunde liegenden Vorkommensdaten ungenau, fehlerhaft oder unvollständig sind.

## 10. Anhang

### **I. Auswahl an Klimavariablen**

- ✓ **bio1** = Annual Mean Temperature, mittlere Jahrestemperatur
- ✓ **bio2** = Mean Diurnal Range, mittlerer Tagesbereich (Temperatur)
- ✓ **bio3** = Isothermality, Isothermalität
- ✓ **bio4** = Temperature Seasonality, Temperatursaisonalität
- ✓ **bio5** = Max Temperature of Warmest Month, maximale Temperatur des wärmsten Monats
- ✓ **bio6** = Min Temperature of Coldest Month, minimale Temperatur des kältesten Monats
- ✓ **bio7** = Temperature Annual Range, jährlicher Temperaturbereich
- ✓ **bio8** = Mean Temperature of Wettest Quarter, mittlere Temperatur des feuchtesten Monats
- ✓ **bio9** = Mean Temperature of Driest Quarter, mittlere Temperatur des trockensten Monats
- ✓ **bio10** = Mean Temperature of Warmest Quarter, mittlere Temperatur des wärmsten Quartals
- ✓ **bio11** = Mean Temperature of Coldest Quarter, mittlere Temperatur des kältesten Quartals
- ✓ **bio12** = Annual Precipitation, Jahresniederschlag
- ✓ **bio13** = Precipitation of Wettest Month, Niederschlag des feuchtesten Monats
- ✓ **bio14** = Precipitation of Driest Month, Niederschlag des trockensten Monats
- ✓ **bio15** = Precipitation Seasonality, Niederschlagssaisonalität
- ✓ **bio16** = Precipitation of Wettest Quarter, Niederschlag des feuchtesten Quartals
- ✓ **bio17** = Precipitation of Driest Quarter, Niederschlag des trockensten Quartals
- ✓ **bio18** = Precipitation of Warmest Quarter, Niederschlag des wärmsten Quartals
- ✓ **bio19** = Precipitation of Coldest Quarter, Niederschlag des kältesten Quartals

### **II. Quelle der Klimadaten**

In dem SDM-Workflow werden Klimadaten von WorldClim in einer Auflösung von 2.5 Minuten genutzt.

<https://www.worldclim.org/data/bioclim.html>

<https://www.worldclim.org/data/worldclim21.html>

### III. Auswahl an Landbedeckungsklassen

- ✓ **LC1** = 1.1.1. Continuous urban fabric
- ✓ **LC2** = 1.1.2. Discontinuous urban fabric
- ✓ **LC3** = 1.2.1. Industrial or commercial units
- ✓ **LC4** = 1.2.2. Road and rail networks and associated land
- ✓ **LC5** = 1.2.3. Port areas
- ✓ **LC6** = 1.2.4. Airports
- ✓ **LC7** = 1.3.1. Mineral extraction sites
- ✓ **LC8** = 1.3.2. Dump sites
- ✓ **LC9** = 1.3.3. Construction sites
- ✓ **LC10** = 1.4.1. Green urban areas
- ✓ **LC11** = 1.4.2. Sport and leisure facilities
- ✓ **LC12** = 2.1.1. Non-irrigated arable land
- ✓ **LC13** = 2.1.2. Permanently irrigated land
- ✓ **LC14** = 2.1.3. Rice fields
- ✓ **LC15** = 2.2.1. Vineyards
- ✓ **LC16** = 2.2.2. Fruit trees and berry plantations
- ✓ **LC17** = 2.2.3. Olive groves
- ✓ **LC18** = 2.3.1. Pastures
- ✓ **LC19** = 2.4.1. Annual crops associated with permanent crops
- ✓ **LC20** = 2.4.2. Complex cultivation patterns
- ✓ **LC21** = 2.4.3. Land principally occupied by agriculture
- ✓ **LC22** = 2.4.4. Agro-forestry areas
- ✓ **LC23** = 3.1.1. Broad-leaved forest
- ✓ **LC24** = 3.1.2. Coniferous forest
- ✓ **LC25** = 3.1.3. Mixed forest
- ✓ **LC26** = 3.2.1. Natural grassland
- ✓ **LC27** = 3.2.2. Moors and heathland
- ✓ **LC28** = 3.2.3. Sclerophyllous vegetation
- ✓ **LC29** = 3.2.4. Transitional woodland-scrub
- ✓ **LC30** = 3.3.1. Beaches, dunes, sands
- ✓ **LC31** = 3.3.2. Bare rocks
- ✓ **LC32** = 3.3.3. Sparsely vegetated areas
- ✓ **LC33** = 3.3.4. Burnt areas
- ✓ **LC34** = 3.3.5. Glaciers and perpetual snow
- ✓ **LC35** = 4.1.1. Inland marshes
- ✓ **LC36** = 4.1.2. Peat bogs
- ✓ **LC37** = 4.2.1. Salt marshes
- ✓ **LC38** = 4.2.2. Salines
- ✓ **LC39** = 4.2.3. Intertidal flats
- ✓ **LC40** = 5.1.1. Water courses
- ✓ **LC41** = 5.1.2. Water bodies
- ✓ **LC42** = 5.2.1. Coastal lagoons
- ✓ **LC43** = 5.2.2. Estuaries
- ✓ **LC44** = 5.2.3. Sea and ocean

#### **IV. Quelle der Landbedeckungsklassen und weitere Informationen**

Die von uns bereitgestellten Landbedeckungsdaten basieren auf den CORINE Landbedeckungsdaten des Copernicus Land Monitoring Services. Wir nutzen hier Daten für das Jahr 2012. Die Daten wurden von uns von einer 100 m-Auflösung auf eine 2.5 Minuten-Auflösung aggregiert. Dabei wurde für jede Landbedeckungsklasse der prozentuale Anteil dieser Klasse je Gitterzelle berechnet. Das hierfür verwendete R-Skript („Prepare\_LC\_data“) ist ebenfalls auf GitHub bereit gestellt, so dass es Ihnen möglich ist, die Berechnung nachzuvollziehen. Ausführen müssen Sie dieses Skript aber nicht.

<https://land.copernicus.eu/pan-european/corine-land-cover>

#### **V. GBIF Vorkommensdaten**

Wenn Sie keine eignen Vorkommensdaten einlesen, werden im ersten Schritt des Workflows automatisch Vorkommensdaten von GBIF (Global Biodiversity Information Facility) herunter geladen und bereinigt. Dies sind Punktdaten, die aus verschiedenen Datensets zentral zusammen geführt werden.

<https://www.gbif.org/>

<https://www.gbif.org/occurrence/search>