

HotSpotSDM – ein WPS-Prozess für die Modellierung ökologischer Artenfunde

Anna Charlotte Kirchner¹, Christian Aden¹

¹Carl von Ossietzky Universität Oldenburg · charlotte.kirchner@uni-oldenburg.de

Zusammenfassung: Der OGC Web Processing Service Standard ermöglicht eine Verarbeitung von räumlichen Daten über das Internet. Dies lässt sich nutzen, um komplizierte Analysemethoden zu automatisieren und somit einem größeren Anwenderkreis zugänglich zu machen. Wir nutzen eine Serverkonfiguration mit PyWPS und R Statistics, zusammen mit Daten aus den freien Open Access Datenbanken WorldClim und GBIF, um am Beispiel von biologischen Artenfunden eine automatisierte Habitatmodellierung mit einer zentralen Speicherung von Ergebnissen durchzuführen. Mittels dieses Beispiels soll das Potenzial für die Nutzung von WPS anhand von interdisziplinären Fragestellungen in Ökologie und den Umweltwissenschaften verdeutlicht werden.

Schlüsselwörter: WPS, Habitatmodellierung, Prozessautomatisierung, INSPIRE, Citizen Science, GDI

Abstract: *The OGC Web Processing Service standard enables analysis of spatial data over the internet. This can be used to automate complicated analyzing methods and make them available to a wider audience. Using a combined setup of PyWPS and R Statistics and the free and open access databases WorldClim and GBIF, we aim to demonstrate the processing of species occurrence data using a process for automated species distribution modelling with centralized data storage. Based on this example we are pointing out the potential advantages of using WPS as a tool for solving interdisciplinary issues within the disciplines of ecology and environmental sciences.*

Keywords: *WPS, species distribution modelling, process automation, INSPIRE, citizen science, SDI*

1 Motivation und Stand der Technik

Mit dem *Standard for Web Processing Services* wurde im Jahr 2007 vom OGC der Weg für die webbasierte Verarbeitung von Geodaten geebnet (OGC 05-007r7). Eine Zentralisierung von Prozessen auf einen Geoserver bietet bei häufig genutzten und komplexen Methoden diverse Vorteile gegenüber einer traditionellen Bearbeitung und kann interdisziplinär genutzt werden, um solche teils sehr rechenintensive Prozesse für eine breitere Gruppe von Anwendern verfügbar zu machen. Vor allem, bei der Bearbeitung großer Datenmengen kann eine solche Zentralisierung viele Vorteile mit sich bringen (VITOLO et al. 2015) und hat somit das Potenzial, sich in Zukunft zu einer Standardmethode zu entwickeln.

Um das Potenzial von Web Processing im Bereich der Ökologie und des Artenschutzes zu verdeutlichen, haben wir einen automatisierten Prozess für die ökologische Nischenmodellierung implementiert, mit dem bestehende Datenlücken gefüllt werden können.

Ökologische Nischenmodellierung (auch Species Distribution Modelling (SDM) oder Habitatmodellierung) ist eine statistische Methode zum Vorhersagen von Artenvorkommen an nicht untersuchten Orten auf Basis von bekannten Fundorten und einer Reihe von Umweltparametern (GUISAN & ZIMMERMANN 2000). Es hat sich zu einer beliebten Methode in diversen Bereichen von Umweltwissenschaften und Biodiversitätsschutz entwickelt (FRANK-

LIN 2013), wo es unter anderem dafür genutzt wird, Datenlücken ähnlich wie bei einem geostatistischen Interpolationsverfahren zu schließen und erfreut sich in der Community einer immer größer werdenden Nutzergemeinde. Auch in der Geodateninfrastruktur der Europäischen Kommission (INSPIRE-Richtlinie 2007/2/EG) ist SDM im Geodaten thema „Verteilung der Arten“ als Datenquelle für die Artenverteilung aufgeführt.

Im Gegensatz zu den bekannten geostatistischen Interpolationsverfahren, wird bei SDM der spezifische Toleranzbereich einer Art, ihre ökologische Nische, in Bezug auf einen oder mehrere Umweltparameter gesucht um eine Approximation ihrer Verbreitung in einem Zielgebiet zu erhalten. Bei einem SDM handelt es sich um eine mathematische Darstellung dieser ökologischen Nische, ein räumlicher Bezug wird jedoch erst durch eine Prognose auf Basis von georeferenzierten Prädiktorvariablen, wie z. B. Temperatur, hergestellt.

Eine ökologische Nischenmodellierung, die „fit for purpose“, also geeignet für die entsprechenden Anforderungen ist, erfordert sowohl Hintergrundwissen über den zu modellierenden Organismus als auch Software- und Programmierkenntnisse. Nicht zuletzt sind für eine Darstellung der Ergebnisse GIS-Kenntnisse unabdingbar.

Diese Kombination an benötigten Skills, (sowie ihre Abwesenheit bei vielen potenziellen Anwendern,) macht SDM zu einem idealen Beispiel für die Umsetzung einer umweltwissenschaftlichen Methode als automatisierter Web Processing Service.

Es existieren eine Reihe von Desktop-Anwendungen (z. B. MaxEnt, openModeller), sowie diverse Pakete für R Statistics (z. B. dismo, biomod2) mit denen zwar teils eine intuitive Ausführung von SDM ermöglicht wird, die jedoch eine Anzahl von Schwächen aufweisen:

- (1) Die Laufzeiten der rechenintensiven Modellierungen sind abhängig von der Rechenleistung der eigenen Hardware.
- (2) Die Resultate sind häufig nicht, oder nur schwer, vergleichbar mit anderen Modellen.
- (3) Speicherung der Resultate erfolgt meist lokal und nicht zentralisiert in Geodatenbanken.
- (4) Eine Automatisierung der Modellierung ist häufig nicht, oder nur mit Programmierkenntnissen, möglich.
- (5) Eine Verkettung der Modellierung mit anderen Verarbeitungsschritten (Process Chaining) ist nicht, oder nur schwer möglich.

Als Lösungsansatz wurde bereits von mehreren Entwicklern ein webbasierter Ansatz gewählt. Vielversprechend stellt sich hierbei der von DUBOIS et al. (2013) beschriebene *eHabitat* dar, der ein auf Open-Source-Programmen aufgebauter, (laut Autoren) erweiterbarer, Web Processing Service ist, der viele Ansprüche abdeckt, die auch wir als Ziel gesetzt haben. Eine Nutzung dieses Prozesses über Process Chaining ist jedoch durch die Inflexibilität der eigenen Weiterentwicklung und einen fehlenden stabilen Zugriff des WPS nicht möglich. Auch der *OpenModeller Web Service* (OMWS) (DE GIOVANNI et al. 2015) ist eine relativ neue, webbasierte Anwendung zur ökologischen Nischenmodellierung, bei der jedoch keine OGC-Standards zum Tragen kommen.

Unser Ansatz stellt durch die kombinierte Nutzung von webbasierter Verarbeitung über den OGC-WPS-Standard mittels PyWPS und der einfach erweiterbaren und vergleichbaren Modellierung mit dem R Paket *sdm* (NAIMI & ARAÚJO 2016) eine einfache, aber effektive Methode dar, um die zuvor benannten Probleme bewältigen zu können.

2 Methode

Ziel ist es, einen WPS bereitzustellen, mit dem theoretisch selbst ein ungeschulter Laie für eine beliebige selbst gewählte Art eine relativ verlässliche Habitatmodellierung durchführen kann. Dazu muss dieser SDM-Prozess einer Reihe von Anforderungen gerecht werden:

- (1) Es soll möglich sein, auch ohne Hintergrundwissen über die Implementierung des Verfahrens eine Habitatmodellierung durchzuführen.
- (2) Erfahrenere Nutzer (z. B. Wissenschaftler) sollen die Möglichkeit haben, eigene Daten über den Prozess zu verarbeiten und Modellparameter anzupassen. Bei minimaler Datenlage soll der Prozess jedoch auch funktionieren, wenn nur ein Artname übergeben wird.
- (3) Durch Nutzung einer Reihe von Best Practice Validierungsmethoden soll garantiert werden, dass Modelle mit zu hoher Unbestimmtheit ausgeschlossen werden.
- (4) Das ausgegebene Modell soll über eine konservative Grenzwertbildung in eine binäre Präsenz/Absenz-Verbreitungskarte verwandelt werden, die in Folge weiterverwendet werden kann.
- (5) Die Ergebnisse sollen zentral in einer Geodatenbank gespeichert werden und zusätzlich als Download zur Verfügung gestellt werden.

Die zugrunde liegende Modellierung der Daten findet in R Statistics (v. 3.2.2) statt. Das R Skript wird in einen PyWPS-Prozess eingebunden (vgl. ČEPICKÝ 2008), der von außen die Inputs und Outputs regelt, sowie eine Verbindung zum GeoCMS HotSpot (ADEN 2012) beinhaltet und Ergebnisse zentralisiert in der Geodatenbank des Servers abspeichert.

Die Ergebnisse bieten eine Grundlage für die Plausibilitätskontrolle der von Citizen Scientists erfassten Artendaten. Diese Methode wird im Rahmen der Entwicklung des GeoCMS-Moduls *GoInspired* genutzt, das die Transformation von Citizen Science Daten in das Geodaten thema „Verteilung der Arten“ der EU-weit umzusetzenden INSPIRE-Richtlinie (2007/2/EG) ermöglicht. Weiterhin soll das Modellierungsergebnis Informationen für die Bereiche im Raum aufzeigen, an denen keine Artenerfassungen stattgefunden haben (ADEN & KIRCHNER 2016). Daher ist es ein Ziel, den Ergebnisdatensatz so aufzubereiten, dass eine Abbildung auf den im Modul *GoInspired* verwendeten Gebietseinheiten (Species Distribution Units, SDU) möglich ist (z. B. 10 × 10 km Grid).

2.1 Funktionsweise SDM

Bei der Habitatmodellierung werden für eine Art zunächst die Werte von verschiedenen Umwelt-Prädiktorvariablen extrahiert, die an den georeferenzierten Fundorten vorliegen. Im nächsten Schritt werden auf Grundlage dieser Werte mithilfe verschiedener Algorithmen Modelle erstellt, auf deren Basis im letzten Schritt dann die Verbreitung der Art in einem nicht untersuchten Gebiet oder zu einer anderen Zeit, z. B. in der Zukunft, anhand der gewählten Prädiktoren prognostiziert werden kann.

2.2 Daten

Artenfunde: Es gibt zwei Möglichkeiten, georeferenzierte Artenfunde in den Prozess einzulesen. Eigene Fundorte können als CSV Datei oder Shapefile eingelesen werden. Sollten keine eigenen Funddaten vorhanden sein, werden automatisch Daten aus der freien und Open

Access Datenbank der Global Biodiversity Information Facility (s. GBIF 2016a) heruntergeladen und zur Modellierung verwendet. Die Nutzung der GBIF-Datenbank in unserem Projekt hat verschiedene praktische Gründe:

- (1) GBIF enthält mit über einer halben Milliarde an Artenfunden und Funden zu über 1,5 Millionen Spezies eine enorm große Vielfalt an Informationen.
- (2) Als unabhängige Organisation kooperiert GBIF mit einer Reihe von Staaten und Organisationen, u. a. auch mit der Europäischen Kommission.
- (3) Ein automatisierter Zugriff auf die Daten ist über das `dismo` Paket in R relativ einfach.

Zwar wurde in der Vergangenheit Probleme bemängelt, die bei der Verwendung von Daten aus der GBIF-Datenbank zur Habitatmodellierung auftreten können (u. a. YESSON et al. (2007) und Beck et al. (2014)), GBIF arbeitet jedoch beständig an der Verbesserung der Dateninfrastruktur, sowie der Verwendbarkeit der Daten für SDM (GBIF 2016b).

Umweltparameter: Als *default*-Einstellung werden für das Gebiet, das die Artenfunde umfasst, von WorldClim (WORLDCLIM 2016) ein Datensatz über 19 bioklimatische Parameter (Bioclim) in der Auflösung 30 Bogensekunden heruntergeladen und zugeschnitten. Alternativ kann auch ein Set von eigenen Umweltparametern verwendet werden, die als Raster bereitliegen müssen.

2.3 Programmierung in R

R Statistics (R CORE TEAM 2015) eignet sich besonders gut zur Umsetzung des Vorhabens, da Pakete sowohl für die Verarbeitung von räumlichen Daten, als auch für Species Distribution Modelling existieren. Zudem existiert über das `rpy2`-Modul die Möglichkeit zur Interoperabilität zwischen R und PyWPS, und somit die Voraussetzung zur Umsetzung von R Skripten als WPS-Prozess.

2.3.1 Daten: Download und Vorbereitung

Die GBIF-Artenfunde können über die `gbif()` Funktion aus dem Paket `dismo` automatisch in R gedownloadet werden. Der Datensatz wird infolge von nicht georeferenzierten Funden, sowie Funden die vor dem Jahr 2000 erfasst wurden, bereinigt und als *SpatialPoints* Objekt abgespeichert.

Der Datensatz für die Bioclim-Parameter kann über die `getData()`-Funktion im `raster` Paket heruntergeladen werden. Er wird auf die Ausdehnung der Artenfunde zurechtgeschnitten und in einen *RasterStack* konvertiert.

Die Daten werden anschließend für die Modellierung in einem `sdmData{sdm}` Objekt zusammengeführt.

2.3.2 Modellierung

Wie jede Habitatmodellierung besteht die Modellierung nach der Bereitstellung von Präsenzdaten und Umweltparametern aus zwei weiteren Schritten:

- (1) Für die Fundorte werden die Werte der Umweltparameter extrahiert, um ein Modell anzupassen.

- (2) Das Modell wird genutzt, um Vorhersagen über die Verbreitungsmuster in einer Zielregion zu treffen.

Hierzu wurde das `sdm` Paket v. 0.0-11 (NAIMI & ARAÚJO 2015) verwendet, das auf andere zuvor implementierte Pakete für Habitatmodellierung in R zugreift und für eine Vergleichbarkeit der einzelnen Algorithmen sorgt. Um Modellungenauigkeiten zu reduzieren, haben sich als Ansatz zusammengesetzte Modelle (ensemble-models) bewährt. Für den Prozess wurde deshalb ein solches Modell aus 5 verschiedenen Algorithmen (s. Tabelle 1) gewählt, die einer dreifachen Kreuzvalidierung unterzogen werden und deren `mean()` dann kalkuliert und als Raster ausgegeben wird.

Die offiziell veröffentlichte Version des `sdm` Pakets (NAIMI & ARAÚJO 2016) beinhaltet neben den genannten und im `HotSpotSDM` angewendeten Algorithmen, noch 12 weitere Algorithmen. Zum Zeitpunkt der Implementierung der aktuellen Version von `HotSpotSDM` (`sdm` v. 0.0-11) war nur eine beschränkte Anzahl von Algorithmen implementiert. Eine zukünftige Version des `HotSpotSDM` kann jedoch problemlos um weitere Algorithmen, wie z. B. `Bioclim` oder `Mahalanobis`, erweitert werden.

Tabelle 1: Verwendete Algorithmen für das Ensemble Model

Algorithmus		Typ
GLM	Generalisierte Lineare Modelle	Regressionsanalyse
GAM	Generalisierte additive Modelle	Regressionsanalyse
BRT	Boosted Regression Trees	Maschinelles Lernen
RF	Random Forest	Maschinelles Lernen
SVM	Support Vector Machine	Maschinelles Lernen

2.3.3 Ergebnisaufbereitung

Um boolesche (`true/false`) Aussagen über die Verbreitung der Art in den INSPIRE-Species Distribution Units zu treffen, wird zunächst der `mean()` aller Fundorte innerhalb des Rasters als Schwellenwert für eine Präsenz/Absenz-Karte kalkuliert. Dieses binäre Raster wird zusammen mit den realen Artenfunden mit dem 10×10 km Grid der INSPIRE SDUs verschnitten und dort nach dem in den Kategorien Artvorkommen positiv (`Fund+`, `SDM+` oder `Fund+/SDM+`) oder negativ pro Gridzelle dargestellt.

Außerdem wird eine Übersichtsgrafik mit den Ergebnissen (`SDM`, `SDUs` und Fundorte im Zielgebiet) sowie einiger Statistik (`ROC-Kurve`, `AUC`, ...) geplottet und als PDF ausgegeben.

2.4 Umsetzung des `HotSpotSDM` `PyWPS`-Prozess

Inputs

Im Minimalprozess besteht der Input vom Nutzer alleine durch die Texteingabe des Artnamens der gewünschten Spezies, der als *LiteralInput* an `PyWPS` übergeben wird. Die weiteren Modellparameter werden in diesem Fall über eine Reihe von *default*-Werten gesetzt.

Nutzerspezifische Prozesse können neben eigenen georeferenzierten Artfunden (CSV oder Shape/GML) auch zusätzlich eigene Umweltparameter (GeoTIFF) sowie ein eigenes Zielgebiet (Shape/GML) enthalten, die entweder über ihren Speicherort auf dem Server als *LiteralInput*, oder bei lokalen Daten als *ComplexInput* eingelesen werden können.

Outputs

Der Prozess speichert die gewählten Ergebnisse auf dem Server ab und stellt sie zum Download bereit.

Neben dem Species Distribution Model (GeoTIFF) und dem auf die Species Distribution Units berechneten Vorkommen (Shape/GML) können optional auch die für das Modell verwendeten GBIF Fundorte (Shape/GML), sowie eine Zusammenfassung (PDF) mit Grafiken und deskriptiver Statistik ausgegeben werden

3 Ergebnisse und Diskussion

3.1 Potenzial

HotSpotSDM wurde implementiert, um das Problem geringer Datenverfügbarkeit zu minimieren. In diesem konkreten Fall wird der Prozess als Validierungsverfahren für ehrenamtlich erhobene Artenfunde genutzt. Autonom oder auch im Hintergrund von Umweltinformationssystemen genutzt, könnte mithilfe des Prozesses dem interessierten Nutzer für jede beliebige Art relativ schnell eine Karte mit ihrer potenziellen Verbreitung erstellt werden.

Eine Annäherung an die Datenverarbeitung über das Internet bietet viele Vorteile. Die hier genutzten Konfigurationsmöglichkeiten von PyWPS nicht nur mit R Statistics, sondern auch mit anderer Software wie beispielsweise GRASS GIS, bieten eine sehr gute Grundlage für die automatisierte Verarbeitung georeferenzierter ökologischer Daten. Auf spezielle Fragestellungen angepasst, könnten beispielsweise ähnliche Prozesse über ein Process Chaining in der Wissenschaft dabei helfen, komplexe Fragestellungen zu bearbeiten und dabei nicht nur durch Umsetzung von Standards eine höhere Vergleichbarkeit von Daten zu erzielen, sondern auch über eine zentralisierte Speicherung von Ergebnissen in Geodatenbanken oder Bereitstellung in GDIs das Datenmanagement erleichtern.

Nutzung als Validierungsverfahren für Citizen Science Daten

Ursprünglich wurde HotSpotSDM nicht als Stand-alone-Prozess konzipiert. Die grundlegende Idee war es, eine Variante des Prozesses als eine von verschiedenen Validierungsmethoden, für Funde aus Citizen Science Projekten in das *GoInspired* Modul des GeoCMS HotSpot zu integrieren. Neben Expertenwissen (*estimateExpert*), ist auch SDM eine Methode die genutzt werden kann um die Verbreitung von Arten auf nationaler Ebene darzustellen. Diese ist unter dem Synonym *predictionModelling* in der INSPIRE Data Specification on Species Distribution verankert (INSPIRE-Richtlinie 2007/2/EG).

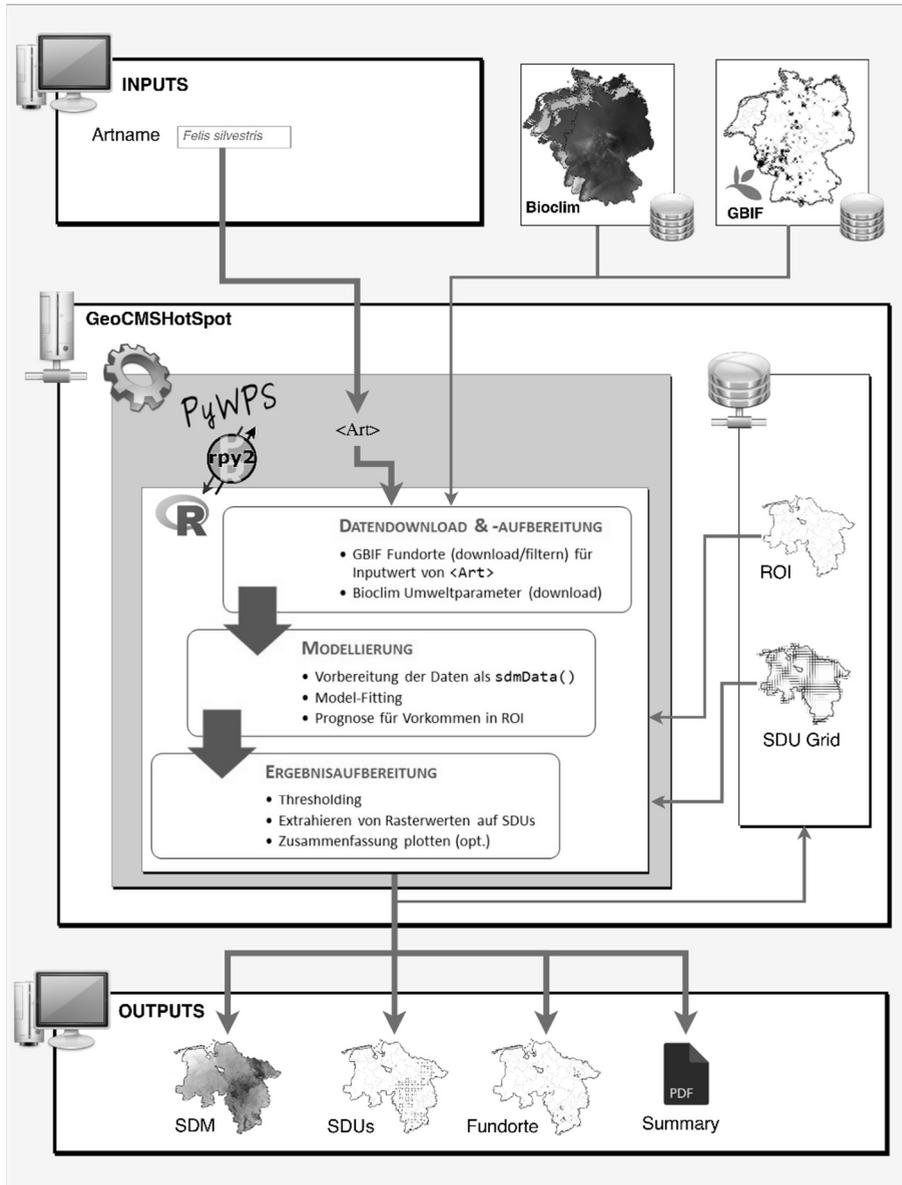


Abb. 1: WPS-Setup und Prozessablauf für das Minimalbeispiel des SDM Prozesses. Der Nutzer übergibt dem Prozess als Input einen wissenschaftlichen Artnamen. Fundorte und Umweltparameter für das Gebiet Deutschland und Niederlande (*default*) werden downgeloadet und modelliert. Die Verbreitung der Art wird für Niedersachsen vorhergesagt und weiterverarbeitet. Die Ergebnisse werden in der Datenbank abgespeichert und das SDM, die Verbreitung je SDU, GBIF Fundorte und eine Übersicht dem Nutzer als Download bereitgestellt.

Zur Überprüfung von Funddaten wird getestet, ob die INSPIRE Species Distribution Unit, in der der Fund sich befindet, vom SDM positiv auf potenzielles Artvorkommen getestet wurde. Da die von Bürgern erhobenen Daten als SDU (10×10 km) dargestellt werden sollen, wird hierbei kein hoch aufgelöstes Modell benötigt.

3.2 Grafische Nutzeroberfläche (GUI)

Der native Aufruf von Prozessen über PyWPS ist wenig intuitiv und kann ein Problem für die Nutzung darstellen. Über eine grafische Nutzeroberfläche, ähnlich dem Schema in **Abb. 2**, könnte die Nutzung des Prozesses vereinfacht werden. Zwar ist eine spezielle GUI zum aktuellen Stand noch nicht implementiert, der Prozess kann jedoch ohne Probleme über den QGIS WPS-Client angestoßen werden. Aufgrund der langen Laufzeit kam es in Tests zwar zu einem Timeout Error des QGIS WPS-Clients, der angestoßene Prozess wird jedoch ausgeführt und die verarbeiteten Daten werden auf dem Server gespeichert.



Abb. 2: Schema für eine Prozess-GUI (nicht implementiert). Über eine einfache Eingabemaske kann der Prozess vom Nutzer gestartet werden. Die Outputs werden direkt online dargestellt und können zur Weiterverwendung gedownloadet werden.

3.3 Durchführung und Laufzeit

Durch die relativ hohe Auflösung der verwendeten Rasterdaten (30 Bogensekunden) und die implementierten Download-Funktionen hat der Prozess zum aktuellen Zeitpunkt noch relativ hohe Laufzeiten. Um die Laufzeiten zu verbessern, werden folgende Lösungen angestrebt:

- (1) Verbesserung der verwendeten Implementierung durch Verwendung von *rpcc* und Objektorientierung.
- (2) Für die am häufigsten verwendeten Regionen sollten die entsprechenden Umweltparameter in der Geodatenbank des Servers gespeichert werden, sodass nur in Notfällen ein Download von Bioclim-Daten nötig wird.
- (3) Bei Nutzung von GBIF-Funden und *default*-Einstellungen soll über einen Timestamp überprüft werden, ob bereits ein aktuelles Modell vorliegt. In diesem Fall wird dieses Modell dann vom Server abgerufen und ausgegeben.

3.4 Qualitätssicherung: Kontrollmechanismen zur Ergebnisüberprüfung

Modelle sind Abstraktionen der Realität und können keine hundertprozentige Kopie der Wirklichkeit sein. Die Wahl, ob ein Modell gut oder schlecht ist, hängt deshalb häufig nicht nur von den verwendeten Daten, sondern auch vom Verwendungszweck ab und benötigt zur Evaluation im besten Fall Expertenwissen und Artenkenntnisse.

Damit der HotSpotSDM-Prozess trotzdem konstant verlässliche Modelle kalkuliert, sollten also weitere Kontrollmechanismen basierend auf Best Practice Methoden der Habitatmodellierung implementiert werden. Diese sollten z. B. bei unzureichender Datenlage die Prozessausführung stoppen oder eine Warnung an den Nutzer ausgeben. Folgende Methoden sind bisher umgesetzt und/oder geplant:

- (1) Reduktion der Prädiktorvariablen durch automatisierte Entfernung von Autokorrelation ($VIF > 10$).
- (2) Prozessabbruch bei zu geringer Anzahl von Artenfunden (geplant).
- (3) Prozessabbruch bei zu niedrigen AUC-Werten (geplant).
- (4) Kreuzvalidierung der Ergebnisse und Verwendung von zusammengesetzten Modellen.
- (5) Konservatives Thresholding des Modells für die Weiterverarbeitung und Darstellung in den INSPIRE SDUs.

4 Ausblick

Neben der Implementierung von Qualitätssicherungsmethoden, sollen in der nächsten Phase neben Bioclim für den terrestrischen Bereich, mit dem Datensatz von Bio-Oracle auch Umweltprädiktoren für den marinen Bereich zur Verfügung gestellt werden.

Mit der offiziell veröffentlichten Version des `sdm` Pakets bietet sich ein großer Fundus an neuen Möglichkeiten. Diese gilt es in den aktuellen Prozess zu integrieren um in Folge den HotSpotSDM Prozess unter Verwendung von aktuellen und von der GBIF-Datenbank unabhängig erhobenen Daten zu testen und zu kalibrieren.

Literatur und Software

- ADEN, C. (2012), Current approaches for networks between NGOs, citizen scientists and public authorities using mobile devices for environmental data collection and online portals for data processing. *Service Oriented Mapping*, 335-352.
- ADEN, C. & KIRCHNER, A. C. (2016), GoInspired – Kontrolle und Transformation von Arten-daten aus Citizen-Science-Projekten für INSPIRE. In: *AGIT – Journal für Angewandte Geoinformatik*, 2-2016, 602-611.
- BECK, J., BÖLLER, M., ERHARDT, A. & SCHWANGHART, W. (2014), Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10-15.
- ČEPICKÝ, J. (2008), OGC Web Processing Service and its usage. *Papers of GIS Ostrava 2008*. gis.vsb.cz/GIS_Ostrava/GIS_Ova_2008/sbornik/Lists/Papers/025.pdf (09.12.2012).

- DE GIOVANNI, R., TORRES, E., AMARAL, R. B., BLANQUER, I., REBELLO, V. & CANHOS, V. P. (2015), OMWS: A Web Service Interface for Ecological Niche Modelling. *Biodiversity Informatics*, 10, 35-44.
- DE SOUZA MUÑOZ, M. E., DE GIOVANNI, R., DE SIQUEIRA, M. F., SUTTON, T., BREWER, P., PEREIRA, R. S. & CANHOS, V. P. (2011), openModeller: a generic approach to species' potential distribution modelling. *Geoinformatica*, 111-135.
- DUBOIS, G., SCHULZ, M., SKØIEN, J., BASTIN, L. & PEEDELL, S. (2013), eHabitat, a multi-purpose Web Processing Service for ecological modeling. *Environmental Modelling & Software*, 41, 123-133.
- EC (2013), Data Specification on Species distribution – Draft Technical Guidelines. http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SD_v3.0rc3.pdf (02/2016).
- ELITH, J. & LEATHWICK, J. R. (2009), Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40 (1), 677.
- FRANKLIN, J. (2013), Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions*, 19 (10), 1217-1223.
- GBIF (2016a), Final Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling: Are Species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF).
- GBIF (2016b), Global Biodiversity Information Facility. <http://www.gbif.org> (02/2016).
- GUISAN, A. & ZIMMERMANN, N. E. (2000), Predictive habitat distribution models in ecology. *Ecological modelling*, 135 (2), 147-186.
- NAIMI, B., & ARAÚJO, M. B. (2015), sdm: Species Distribution Modelling (R package version 0.0-11/r18.).
- NAIMI, B. & ARAÚJO, M. B. (2016), sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*.
- OGC (2015), WPS 2.0 Interface Standard. Open Geospatial Consortium. <http://docs.opengeospatial.org/is/14-065/14-065.html> (01/2016).
- R CORE TEAM (2015), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- VITOLO, C., ELKHATIB, Y., REUSSER, D., MACLEOD, C. J. & BUYTAERT, W. (2015). Web technologies for environmental Big Data. *Environmental Modelling & Software*, 63, 185-198.
- WORLDCLIM (2016). *WorldClim – Global Climate Data*. Retrieved Februar 1, 2016, from Free climate data for ecological modeling and GIS: <http://www.worldclim.org/>
- YESSON, C., BREWER, P. W., SUTTON, T. et al. (2007), How Global Is the Global Biodiversity Information Facility? *PLoS ONE*, 2 (11). doi: 10.1371/journal.pone.0001124.