

## Information Management

## Native and non-native aquatic plants of South America: comparing and integrating GBIF records with literature data

Vanessa Lozano<sup>1,\*</sup>, Daniel S. Chapman<sup>2</sup> and Giuseppe Brundu<sup>1</sup><sup>1</sup>Department of Agriculture, University of Sassari, Viale Italia 39, 07100 Sassari, Italy<sup>2</sup>NERC Centre for Ecology & Hydrology, Bush Estate, Edinburgh EH26 0QB, UKAuthor e-mails: [vlozano@uniss.it](mailto:vlozano@uniss.it) (VL); [dcha@ceh.ac.uk](mailto:dcha@ceh.ac.uk) (DSC); [gbrundu@uniss.it](mailto:gbrundu@uniss.it) (GB)

\*Corresponding author

Received: 27 December 2016 / Accepted: 28 June 2017 / Published online: 14 July 2017

Handling editor: Elena Tricarico

**Editor's note:**

This study was first presented at the special session on aquatic invasive species at the 33rd Congress of the International Society of Limnology (SIL) (31 July – 5 August 2016, Torino, Italy) (<http://limnology.org/meetings/past-sil-congress/>). This special session has provided a venue for the exchange of information on ecological impacts of non-native species in inland waters.

**Abstract**

The Global Biodiversity Information Facility (GBIF) is at the moment one of the largest and most widely used biodiversity databases. Nevertheless, there are still some limitations, e.g. in terms of plant species status (native vs. non-native) and geographic resolution of records. At the same time, it is well known that alien plant invasions in inland freshwaters can alter community structure, ecosystem functions and services with significant negative impacts on biodiversity and human activities. We assessed if the GBIF database has a geospatial homogeneous information for native and non-native aquatic plant species for South America and whether or not literature resources not yet digitalized (floras, checklists and other papers) could provide additional information. We selected a set of 40 native and 40 non-native aquatic species. These 80 species included a sub-set of 40 alien species previously evaluated with the USAqWRA scheme (US Aquatic Weed Risk Assessment). Species with non-reliable identification, duplicates of the same collection, records poorly georeferenced were removed from the dataset. New records were manually compiled through classical literature research. All the georeferenced records (GBIF + literature) were used for the mapping and the comparative analysis. As a result, we can conclude that the two datasets provide quite significantly different information and the combination of the two offers new information that would not exist in a single data source. Nevertheless, a careful quality evaluation of the primary information, both in the case of literature and GBIF should be conducted, before the data is used for further analyses.

**Key words:** alien aquatic plants, biodiversity occurrence data, Global Biodiversity Information Facility, ModestR software, risk assessment

**Introduction**

The Global Biodiversity Information Facility (GBIF, <http://www.gbif.org/>) is one of the largest and most widely used biodiversity database (Jetz et al. 2012; Beck et al. 2014), and it offers freely and universally primary biodiversity data (Roberts and Moritz 2011). This kind of information, together with tools to analyze it (e.g. Geographic Information Systems software,

and statistical analysis packages), has facilitated large-scale analyses and interpretation of biodiversity and distribution data (García-Roselló et al. 2015; Maldonado et al. 2015) for both native and non-native plant species.

Most data on plant species distribution are stored in different sources, including checklists, herbaria, floras and field observations, and are based on point occurrence records, representing what is generally

referred as the primary distribution data, i.e., the occurrence of a particular plant species at a particular location at a particular point in time (Soberón and Peterson 2004). Millions of these records from herbaria and other sources have been mobilised via international data-sharing networks and databases (Edwards et al. 2000) although there might be constraints to a generalised use of this occurrence data due to its coverage (or thematic resolution) and level of accuracy.

Coverage has several components and sub-components, but three main aspects are most commonly considered. The first one is the “taxonomic coverage”, i.e., how many of the existing species and valid lower taxa are well documented (Funk et al. 1999; Hortal et al. 2007; Brummitt et al. 2015) and how frequently the taxonomic and nomenclature resolution and precision of the database is updated and cross-checked. The second is the “geographical coverage”, i.e., how precisely and completely species locations and resulting ranges are documented within records (Feeley and Silman 2011). Finally, the “temporal coverage”, i.e., the time resolution of the database, based, e.g. on more or less continuous recording of species through time (Brummitt et al. 2015) and on the verification of the persistency along time of a given species in one historical locality (Troia et al. 2016). However, gaps and biases usually exist in the available biodiversity information (Boakes et al. 2010; Feeley and Silman 2011; Sousa-Baena et al. 2014; García-Roselló et al. 2015) and data limitations may occur as a result of an inadequate financial and institutional support (Vollmar et al. 2010; Amano and Sutherland 2013) or of different sampling efforts focusing more on regions with certain appeal like endemism, species richness or protected areas (Petřík et al. 2010; Yang et al. 2014). Petřík et al. (2010), showed that the bias in grid mapping of flora seems to be dependent on spatial scale. In addition, the number of botanists involved and duration of the study are associated with some level of bias in estimates of species richness.

GBIF records are widely used in ecology, evolution and conservation (Meyer et al. 2015) and have been used for many different purposes, e.g. to identify native ranges of invasive alien species or species climatic and environmental requirements (Peterson 2003; Suarez and Tsutsui 2004; Chapman 2005; García-Roselló et al. 2015).

In addition to these well-known problems, specific limitations in the GBIF database become evident when one wants to use it as a tool for the analysis of plant invasions, e.g. concerning the plant species status (alien vs. native; casual vs. naturalized vs.

invasive; archaeophyte vs. neophyte), because the invasive status and residence time cannot be inferred for most of the records. GBIF created a dedicated working group to address the enhancement of the system to be used in the field of biological invasions (McGeoch et al. 2016). In fact, it has been remarked that the information available in GBIF has been in some cases used to rapidly assess patterns of diversity and allodiversity, without much attention being paid to the quality and reliability of the data (García-Roselló et al. 2015; Maldonado et al. 2015).

However, GBIF distribution data for alien plants, after expert review, may help in identifying not only highly invaded areas, but also the overall distribution, in predicting locations susceptible to further establishment (Duursma et al. 2013) and in identifying areas that are at greatest risk from future invasions. The knowledge of the spatial distribution of invasive species and invaded habitats is one of the pillars supporting an effective strategy for their management and control (Thuiller et al. 2005). In this concern, the identification of invasive species risk hotspots is a useful tool to prioritize management of plant invasions at large scale (Liang et al. 2014; Adhikari et al. 2015).

Among invaded ecosystems, freshwater ecosystems and habitats, especially lakes and streams are particularly vulnerable (Strayer 2010; Simberloff 2013; Boltovskoy and Correa 2015; Brundu 2015) and prone to dramatic biodiversity loss (Ricciardi and Rasmussen 1999) because of their high concentration of species per surface area (Thomaz et al. 2015). Nutrients in suspension and in sediments are also important determinants of aquatic plants invasion (Engelhardt 2011). While in some continents, intra-continental propagule pressure can be assumed to have been larger, because of the shorter distances, South America have species with restricted ranges, and as the consequence they are less likely to have been dispersed outside their native ranges (van Kleunen et al. 2015). In addition, freshwater ecosystems are often difficult to survey and monitor, so there might be a general scarcity of information about the distribution of invasive alien aquatic plants in many part of the worlds, as is the case of South America (Lozano and Brundu 2016). In particular, Chile, Brazil (Brazil’s Atlantic Forest), Ecuador and Tropical Andes, where biodiversity hotspots are mainly represented, offers a unique opportunity to study biological invasions because they hold a unique native flora with high levels of endemism, extraordinary richness and diverse climatic gradients (Myers et al. 2000; Pauchard et al. 2004). In Brazil, for example, the Guiana Shield constitutes a geological, hydrographical and biogeographic region in the Amazonian

Basin that is considered a very important biodiversity hotspot (Delnatte and Meyer 2012). However, the degree of susceptibility of ecosystems to invasion in these regions is poorly understood and investigated (Thomaz et al. 2015).

The main goal of this study was to assess the state of data availability for aquatic plant species in South America. Therefore, the present research aimed to: (1) evaluate the increase in reliability offered by the merging of information “manually” extracted from literature, for a set of native and non-native aquatic species of South America, with the information for the same species held in GBIF, and (2) evaluate the relationship between the density of non-native aquatic species and large-scale expected predictors such as the Human Influence Index (HII) and the distribution of protected areas in 16 regions of South America.

## Methods

### *Species selection and study area*

For the purposes of the present research, we selected a set of 80 species composed by 40 native and 40 non-native aquatic plant species thriving in South America (Supplementary material Table S1). The set included both helophytes (growing in anaerobic saturated soils) and hydrophytes, free-floating, floating (rooted) and submerged freshwater vascular plants. The selection of these 80 species was based on the availability of reliable information found in the literature concerning taxonomical, geographical and biological traits data, which are a fundamental prerequisite for any reliable modelling and risk assessment.

The 16 South American regions (there after called “regions”) considered in the study were defined as follows: (1) Argentina, (2) Bolivia, (3) Brazil, (4) Chile, (5) Colombia, (6) Ecuador, (7) Falklands Islands, (8) French Guiana, (9) Galapagos, (10) Guyana, (11) Paraguay, (12) Peru, (13) South Georgia and South Sandwich Islands, (14) Suriname, (15) Uruguay and (16) Venezuela.

### *Collection of distribution records*

We first created two distribution datasets for the selected 80 species, using two different methodologies. The first dataset (hereafter called literature dataset) included the geographical coordinates (Lat/Long WGS84) of species records that were collected through classical literature research (national or regional floras: <http://www.floraargentina.edu.ar>, <http://floradobrasil.jbrj.gov.br>, <http://www.lib.udec.cl>, <http://www2.darwin.edu.ar>, flora checklists and papers) for the 16 regions of South America. The

second dataset (hereafter called GBIF dataset) was created using the ModestR software (freely available at the website <http://www.ipez.es/ModestR>, accessed on 2016) (Pelayo-Villamil et al. 2012; García-Roselló et al. 2013). We retrieved all available distributional data for the 80 selected aquatic species in South America from the GBIF portal (<http://www.gbif.org>, accessed on 2016). Acknowledgments for all the sources of the downloaded records from GBIF are shown in the supplementary materials (Table S4). Finally, we merged the two datasets in a new “integrated dataset”, including new information (e.g. status of invasion and the risk level for non-native species using the USAQWRA scheme).

### *Taxonomical and geographical validation of distribution records*

The synonyms used in the literature were handled in accordance to The Plant List portal (<http://www.theplantlist.org/>) and crosschecked using IPNI (International Plant Name Index, <http://www.ipni.org/>), to ensure that all records were assigned to an accepted valid name in agreement with GBIF taxonomic treatment. For species reported in the literature without georeferenced localities, but with an accurate description of the collection locality, geographical coordinates were assigned using Google Earth. On the other hand, the GBIF data were checked and cleaned using the menu facility of ModestR (García-Roselló et al. 2014). Species with non-reliable identification, duplicates of the same collection (discriminating between real duplicates and records of the same specimen sent to different collections), records without (georeferenced) locations or with latitude/longitude equal to 0° and records on the sea (i.e., coordinates that did not project onto land) were removed from the dataset. In addition, the software automatically classified valid and invalid samples depending on whether records were within or outside the inland freshwater. In addition, the software allowed retrieval for all species at the same time by including a file with the species names following a simple taxonomic classification, correcting wrong or invalid synonyms. Once the taxonomic data was introduced, distribution maps for each species was stored in the ModestR database.

### *Invasive status and distribution maps*

Four a priori status categories were defined according to expert opinion, classifying each species in one of the following status categories for each of the 16 regions defined in the present study, or for part of the regions. The four status categories were as

follows: alien non-invasive (NNV), alien invasive (INV), native (IND) and absent (ABS). We used the scores of the 35 non-native species previously evaluated by Lozano and Brundu (2016) with the US Aquatic Weed Risk Assessment (USAqWRA). We assessed the five additional alien aquatic species with the USAqWRA scheme (originally shaped by Gordon et al. 2012), i.e., *Agrostis stolonifera* L., *Aponogeton distachyos* L.f., *Eleocharis acicularis* (L.) Roem. & Schult., *Nasturtium officinale* R.Br., and *Nymphaea micrantha* Guill. & Perr. (Supplementary material Table S1, S5 and S6). The USAqWRA scores and invasion status, per species/region, improved the information downloaded from GBIF, that at the moment does provide only limited features related to biological invasions. This information allowed to correlate the invasive species risk in South America (i.e., the scores derived from the USAqWRA scheme) with large-scale expected predictors such as the Human Influence Index (HII) and the location of the protected areas (PAs).

The whole set of cleaned records, for alien invasive and alien non-invasive species obtained from the “integrated dataset”, was used to map species distribution in each of the 16 regions or part of regions in South America. Distribution maps were created with: 1) GBIF presence records with the addition of records retrieved from literature and 2) GBIF presence records with the invasion status according to expert opinion (see Table S5 and S7). We also produced choropleth maps, based on the number of records, to highlight species density at regional level. Finally, we mapped the allover diversity of the 16 investigated regions.

### GIS & statistical data analysis

We tested the difference between the raw and cleaned records within the two datasets for the 80 species, respectively for GBIF and literature. We also tested the difference between the native and non-native records downloaded from the GBIF and literature dataset applying a t-test (see Table 1). Additionally, we evaluated whether the information collected by literature did improve the information obtained through GBIF.

We downloaded the data set at continental-level (grid format, 1×1 km cell size) for the Human Influence Index (HII) available at the Socioeconomic Data and Applications Center [Wildlife Conservation Society - WCS; Center for International Earth Science Information Network - CIESIN - Columbia University 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4M61H5F>], accessed on December 2016].

Afterward, we extracted the HII values for each pairs of coordinates, corresponding to each species records, in the 16 regions of South America. After projecting the set of cleaned records from GBIF and literature dataset into the HII layer, we used the Pearson's correlation to evaluate the relationship between the non-native species HII scores and their USAqWRA scores, at each geographical location. A Wilcoxon test was performed to check differences between the HII scores at individual points where native species were recorded vs. the points where non-native species were recorded. We addressed the possible bias due to spatial autocorrelation treating HII data with Generalized Least Square Models. GLS was fitted using the function `gls`, with “nlme” R package (Dormann et al. 2007; see Table S10). On the other side, polygon layers of nature reserves or protected areas (PAs) data sets for South America, were obtained from the World Database on Protected Areas (WDPA, <http://www.wdpa.org>, accessed on December 2016). We assessed the relationship between the non-native species proportion of records inside and outside the protected areas and their USAqWRA scores, using a generalized linear model (logistic regression). The risk of invasion was evaluated when the records of the non-native species were within or outside the PAs and in accordance with USAqWRA scores, using the software R (R Core Team 2015). Chi-square test on the contingency table between native/non-native records and inside/outside PAs was also performed. We addressed the spatial autocorrelation treating PAs data with autocovariate regression. The regression was conceived for binary data (as autologistic regression). We used the function `autocov_dist`, with “spdep” R package (Dormann et al. 2007; see Table S10).

## Results

### Number of records and distribution in the 16 South American regions

The GBIF database held valid georeferenced distribution records for 79 of the 80 aquatic plant species investigated in the present study. Importantly, there were no GBIF records for the alien species *Aponogeton distachyos* L.f. (cape pondweed), in South America. On the other hand, the distribution records downloaded from GBIF covered only 15 of the 16 regions, i.e., excluding South Georgia and South Sandwich Islands. The data downloaded from GBIF contained 10,735 raw records (Table 1). Overall, cleaning and validation led to an exclusion of 1,825 records.

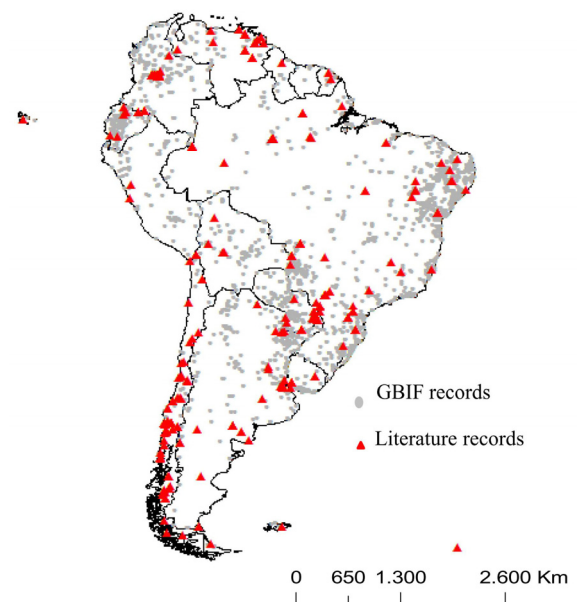
**Table 1.** Difference of the main descriptors of primary biodiversity information. Studies were performed comparing between GBIF and literature datasets: the total number and total records of native, non-invasive and invasive species, and the total records found in the 16 regions of South America, after the cleaning process with the software ModestR.

Feature	GBIF	Literature	p value
No. native species	40	40	–
No. non-invasive alien species	17	18	0.839
No. invasive alien species	22	22	–
Total no. of species	79	80	0.977
Total no. of raw records	10,735	452	< 0.001
Total no. of cleaned records	8,910 (82.99%)	427 (94.46%)	< 0.001
No. of records native species	4,536	126	0.001
No. of records non-invasive alien species	1,565	78	< 0.001
No. of records invasive alien species	2,809	223	< 0.001
No. of regions with records	15	16	0.895
No. sources (GBIF) and manuscripts (Literature)	559	210	0.356

The GBIF data were supplemented with records collected from literature sources. This resulted in the addition of 427 records, obtaining a final total number of 9,337 records (integrated dataset). The difference between the total number of records provided by GBIF and those obtained from the literature dataset, for the 80 species, were highly significant ( $p$  value < 0.001) (Table 1, Figure 1). The average records for species found in the GBIF dataset was 111.3 while for the literature dataset was 5.33. The literature dataset provided information lacking in the GBIF dataset, as it was possible to add records in regions not documented by GBIF such as South Georgia and South Sandwich Islands (Table S2). Although the additional number of records (4.79%) provided by the literature dataset was relatively low, we had an increase in terms of new/different species coverage of 1.26% and of 6.66% coverage for regions: the overall bias rate was considerably lower (5.53%). In addition, literature search provided information such as life form, plant traits and invasion status according to expert opinion.

#### *Native and non-native status*

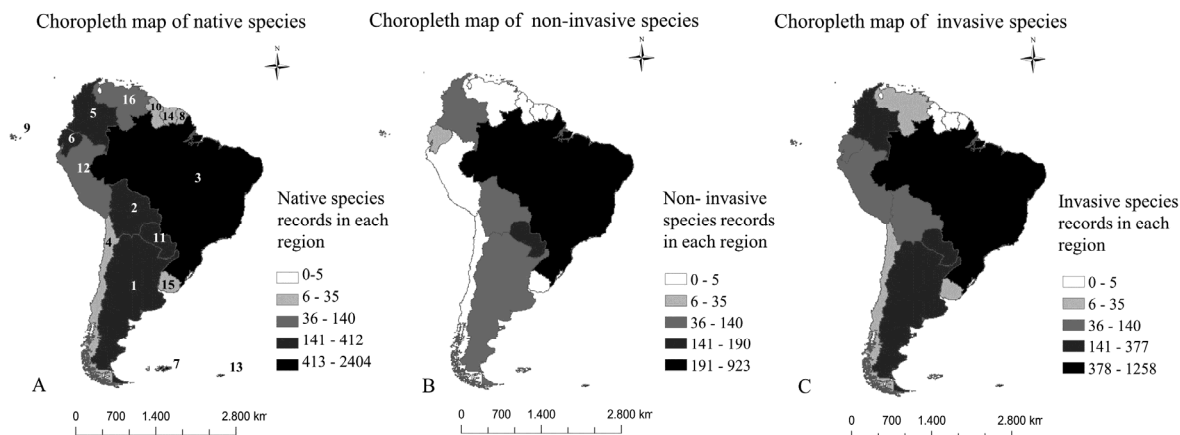
The records available in the GBIF dataset for the selected 40 native and 40 non-native (alien non-invasive + alien invasive) aquatic plant species were in almost equal proportions: 4,536 for native and 4,374 for non-native species (Table 1). The literature dataset provided 126 records for native species, and 301 for non-native species (Table 1). The choropleth maps highlight areas differing in the number of records for native, non-invasive and invasive alien species (Figure 2). Noteworthy, the regions with the higher occurrence of non-native aquatic species records were Brazil (2,182), Colombia (454) and Argentina (444) (Supplementary material Figure S1). In both



**Figure 1.** Map showing the distribution of records downloaded from GBIF (grey circle) and collected from literature sources (red triangle), for South America.

datasets, the total records for alien invasive species were higher than the records for alien non-invasive species (Table 1, Figure 2B, C). The country with the higher density of native and non-native species was Brazil. Regions like French Guiana, Guyana, Suriname and Venezuela, had the tendency to hold more native species than the non-native (Figure 2), and those with the higher density of native and non-native species per square kilometers were Ecuador and Paraguay (Figure S2).

The data downloaded from GBIF showed a massive tendency towards denser species concentration of native and non-native species in Brazil (53), Argentina (52)



**Figure 2.** Choropleth maps showing the density of records of aquatic species in each of the 16 regions of South America, downloaded from GBIF dataset, cleaned and increased through the addition of records from literature. (A) Records of native species; (B) Records of non-invasive alien species per region; (C) Records of invasive alien species per region. The 16 South American regions considered in the study were defined as follows: (1) Argentina, (2) Bolivia, (3) Brazil, (4) Chile, (5) Colombia, (6) Ecuador, (7) Falklands Islands, (8) French Guiana, (9) Galapagos, (10) Guyana, (11) Paraguay, (12) Peru, (13) South Georgia and South Sandwich Islands, (14) Suriname, (15) Uruguay and (16) Venezuela.

and Colombia (50). In the data collected manually from the literature the denser concentrations were found in Brazil (38), Chile (30) and Argentina (28) (Table S3).

According to GBIF *Eichhornia azurea* (Sw.) Kunth, *E. crassipes* (Mart.) Solms, *Ludwigia octovalvis* (Jacq.) P.H.Raven, *Nymphoides indica* (L.) Kuntze, *Pistia stratiotes* L., *Salvinia auriculata* Aubl. and *Utricularia foliosa* L. were the species with the highest numbers of records across the 16 regions of South America. Importantly, among these species, *E. crassipes* (76) and *P. stratiotes* (67) were classified as invasive alien as they reached the highest scores with the USAqWRA, meaning that they represent a major risk (Table S1 and S2). In the literature dataset, the most commonly cited species were: *E. crassipes*, *Hydrocotyle ranunculoides* L.f., *P. stratiotes*, and *U. foliosa*. The literature dataset improved the results adding new records and new species not present in the GBIF dataset, such as *Alisma lanceolatum* With., *Alisma plantago-aquatica* L., *A. distachyos*, *H. ranunculoides*, *Lemna minor* L., *Potamogeton nodosus* Poir. and *Ranunculus aquatilis* L. (Table S2).

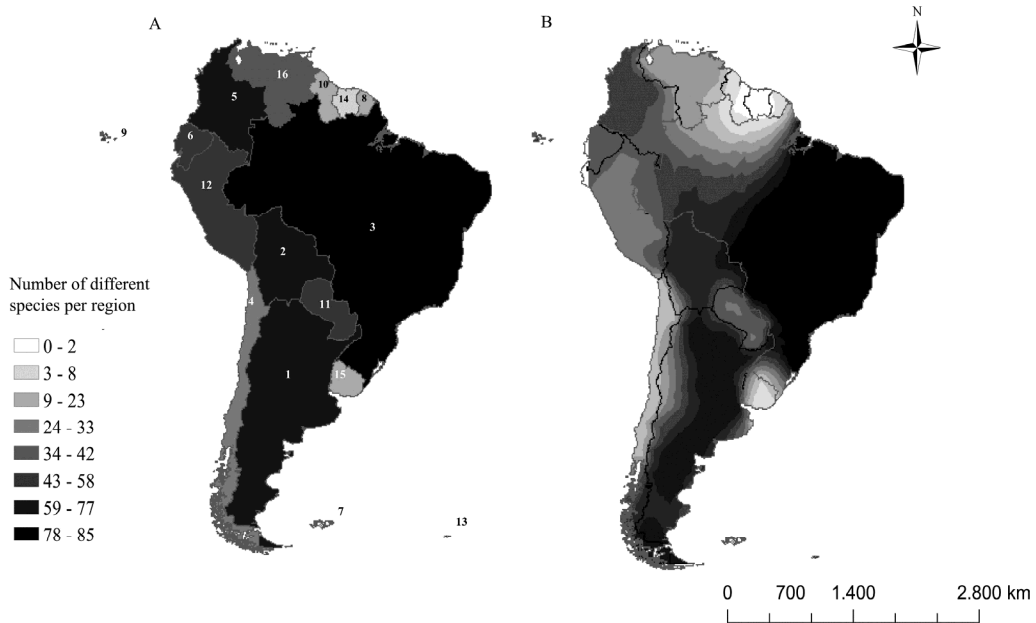
The reliability of GBIF and literature datasets, after cleaning process using ModestR, was 47% and 59% respectively (Table S2). This means that the literature dataset considered in the present study contained a higher proportion of species with reliable information in comparison to GBIF.

The allodiversity of the 16 investigated regions (i.e., the number of alien species present in a specific area, *sensu* Barthlott et al. 1999) is shown in Figure

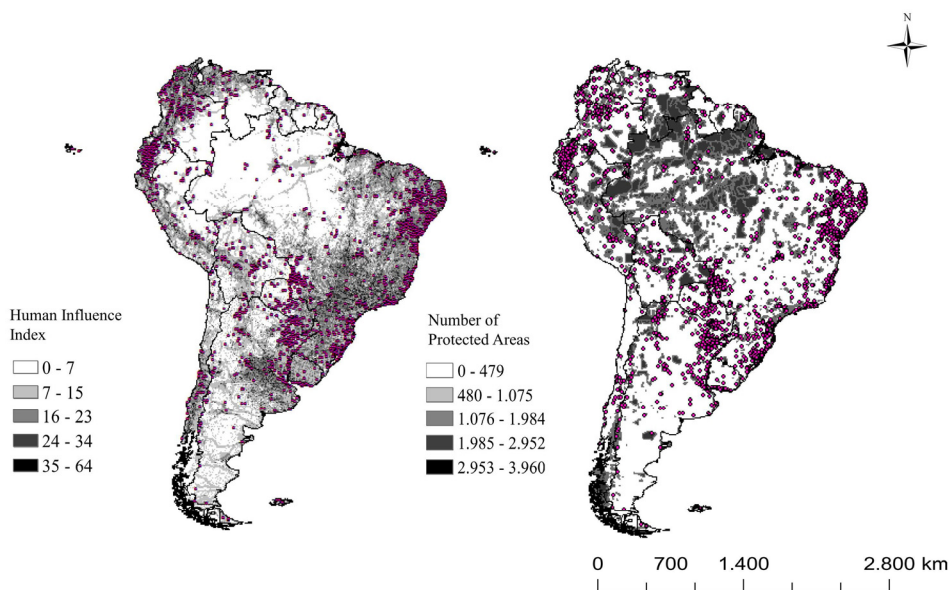
3A. According to our results, the regions holding the highest number of different alien aquatic plant species were Argentina and Brazil. The ordinary Kriging map of Figure 3B shows with a better spatial resolution those parts of the regions holding the highest numbers and densities of alien invasive and alien non-invasive species.

When the Human Influence Index was considered, the alien invasive and alien non-invasive species with a high level of risk according to the USAqWRA scheme were positively correlated ( $t = 3.5851$ ,  $df = 4421$ ,  $p$  value  $< 0.001$ ) with those locations with the higher level of anthropisation (Figure 4 and S3). In addition, we observed that in Brazil, Colombia, Ecuador, Uruguay and Venezuela the records were mostly found along the coast, probably close to the main ports, and this could be related to the pathway of introduction (intentional, e.g. ornamental, Table S1) or secondary release. We predicted a significantly higher HII for non-native occurrences ( $p$  value  $< 0.001$ ). We did not find spatial autocorrelation with the HII scores and the USAqWRA scores ( $p$  value = 0.0014).

The correlation between the distribution of the most invasive species according to the USAqWRA scheme and the PAs in South America was found significant ( $p$  value = 0.034) and with a negative correlation coefficient, meaning that there is a higher probability of founding the most risky alien species outside the PAs (Figure 4). The chi-square test was significant ( $p$  value  $< 0.001$ ), meaning that there is a higher number of records of non-native species outside the PAs in comparison to native species records.



**Figure 3.** Number of different invasive alien and non-invasive alien species present in each of the 16 regions of South America, downloaded from GBIF dataset, cleaned and increased through the addition of records from literature. A) Choropleth map of the number of non-native species. B) Kriging of estimated non-native species density according to the number of different species in the 16 regions of South America. The 16 South American regions considered in the study were defined as follows: (1) Argentina, (2) Bolivia, (3) Brazil, (4) Chile, (5) Colombia, (6) Ecuador, (7) Falklands Islands, (8) French Guiana, (9) Galapagos, (10) Guyana, (11) Paraguay, (12) Peru, (13) South Georgia and South Sandwich Islands, (14) Suriname, (15) Uruguay and (16) Venezuela.



**Figure 4.** Human Influence Index (HII) map on the left side and the Protected Areas (PAs) map on the right side. The records of the invasive alien and non-invasive alien species downloaded from GBIF dataset and integrated through the addition of records compiled manually are shown in both maps (dots). The HII grid layer was downloaded from the Wildlife Conservation Society [WCS; Center for International Earth Science Information Network – CIESIN – Columbia University 2005] and the PAs map was downloaded from the World Database of Protected Areas (WDPA).

## Discussion

Globalization facilitates the spread of aquatic invasive plants as international commerce develops (Perrings et al. 2005; Donaldson et al. 2014; Seebens et al. 2015, 2017), and most of the aquatic alien species have been deliberately introduced as ornamental or for other commercial uses. After being introduced they might escape into the environment and also South America is unfortunately negatively affected by this process (Table S1). For example, *Arundo donax* L., was introduced in the Galapagos as an ornamental (Guézou et al. 2014). *Hippuris vulgaris* L., was introduced in Chile from Europe as ornamental (Ramírez and San Martín 2006). In Chile, 21% of the aquatic and riparian flora has been introduced. It is likely, that this percentage will increase as the country develops and together with it the water bodies are subjected to greater disturbances and prone to accidental escapes or even to intentional releases (Urrutia et al. 2016). To this concern, ports can be one of the main entrance points for aquatic alien species from other countries arriving as stowaways (e.g. ship hull fouling or transport with ballast water, Hulme 2009). Although GBIF and literature data may be biased and have limits in the coverage, especially in poorly investigated regions, we can expect that part of the difference in allo-diversity detected in South America using a sample of 80 aquatic species, might be due to the distribution of ports acting as points of entry, and to the intense trade of ornamental plants. The native species considered in the present research could be considered as a “control group” whose distribution pattern reflects the sampling biases in the data when the effects of introduction dynamics and pathways are not important. Therefore, we can assume that differences between native and non-native records (e.g. non-natives in higher HII) could be a result of introduction and secondary release pathways.

In our study, similarly to other studies where the primary biodiversity information is used (Sousa-Baena et al. 2014; García-Roselló et al. 2015; Maldonado et al. 2015; Meyer et al. 2015), a critical point that decreased the data reliability was the inaccurate georeferencing (17.0% of wrong/missing locations). Hijmans et al. (1999) suggested that a relatively large proportion of all available records are not correctly georeferenced. Feeley and Silman (2011), reported the extreme lack of collections data in GBIF (and a similar database for Brazil named SpeciesLink; <http://splink.cria.org.br/>) for tropical plant species. They estimated that about 65% of tropical plants lack available geo-referenced collections. This lack of reliable spatial information over

vast extents demonstrates that for many regions with large conservation opportunities there are not sufficient occurrence data to support even the most sophisticated modeling approaches (Meyer et al. 2015). Feeley and Silman (2011) termed this lack of knowledge as the “data void”. They pointed out the importance of investigating species responses to climate change through species distribution modeling to predict rates of habitat loss and the associated extinction risks. Nevertheless, Collen et al. (2008), found for the tropical South America that species distributions and their responses to climate change is potentially crippled by a lack of basic data. Using “presence-only” data, a minimum of 20–50 collections per taxa are generally required to produce accurate species distribution models. Due to the paucity of digitized collections, very few tropical species meet this criterion (Feeley and Silman 2011).

Sousa-Baena et al. (2014), pointed out that incompleteness not only is due to the lack of collection effort, but may also correspond to existing knowledge that is not digital or not accessible. In accordance with Maldonado et al. 2015, we would like to emphasize that the sources of information should be always accompanied by good metadata, including specific details on how the coordinates were obtained, and on whether the coordinate assignment was done manually (e.g. literature sources) or automatically (GBIF data). Periodically, researchers will need to re-evaluate coverage and completeness, and this information will need to incorporate additional coverage information. An advantage of enhancing GBIF dataset with occurrence records collected manually is that they might increase information about local patterns of occurrence, species abundances or community composition.

Feeley (2015) quantified the amount of occurrence data available through GBIF for plant species in tropical South America and examined how data availability had changed through time. He found that most of this increase was due to the inclusion of additional pre-existing records rather than new collections. This increase was driven in large part by the incorporation of SpeciesLink data into GBIF. The greatest density of collections comes from the Northern Andean Paramo and Andean ecoregion, consistently with part of our data occurrence. In tropical South America, more than 10% is still represented by no collection and the reason is that the vast majority of species are sterile, therefore many collections are not identified to species or are identified incorrectly.

The importance and advantage of increasing the digitized records (e.g. in South America) is due to the fact that many ecoregions are very poorly



represented in the GBIF collections database. Our results suggested that literature records can improve the coverage of the GBIF dataset, e.g. in Argentina, Brazil and Chile. For example, the Cerrado is one of the South America's largest, most diverse, and most threatened ecoregions but it is not well-represented (Feeley and Silman 2009). In contrast, the Andean ecoregion are well-represented, maybe their collection intensities were higher and there is not a lack of access (due to physical or bureaucratic impediments) (Feeley 2015). Major rivers, such as the Tocantins and Tapajós in Pará State, and the Rio Negro and Rio Madeira in Amazonas State, are sometimes associated with higher information content. Nevertheless, for our dataset the information in this well-known place is still lacking.

Importantly, in contrast with the general trend, the biodiversity hotspot regions in Brazil (Myers et al. 2000) have the highest concentration of the invasive alien plants and include a large number of protected areas (World Conservation Union and UNEP-World Conservation Monitoring Centre 2007). Protected areas are usually characterised by high levels of biodiversity, unique habitats, pristine ecosystems or protected or endangered species (Yang et al. 2014; Kumschick et al. 2015). The Andes biodiversity hotspot is one of the most diverse regions and supports many endemic species of high conservation priority (Myers et al. 2000), yet the lack of usable data interferes in conservation efforts. Immediate efforts are needed to increase the quality and number of data available from this and other underrepresented systems (Feeley and Silman 2010).

In accordance with our results, Feeley and Silman (2011) reported that in Ecuador a relatively large number of collections are available online, thanks to the efforts of local herbaria, including the Museo de Historia Natural (QCNE), the Pontificia Universidad Católica del Ecuador (QCA), and the Universidad Central del Ecuador (QAP).

There is a clear need for more frequent and intensive collection campaigns, not just for our set of aquatic species but in general, and research efforts in the structure and dynamic across the Amazon and Tropical South America (Feeley 2015).

The fact that Protected areas are holding a low quantity of invasive species (Foxcroft et al. 2017), could be related to a lower sampling effort (Figure S4). Nevertheless, in the past years, there have been attempts to standardize inventory data (i.e., plot data) in the Amazon forest (e.g. the Amazon Tree Diversity Network (ATDN, <http://web.science.uu.nl/Amazon/atdn/>) and the RAINFOR Amazon Forest Inventory Network (<http://www.rainfor.org/>) (Feeley 2015) and this information would be useful to reduce

the artefact of sampling bias. Schulman et al. (2007), showed that much of the Amazonian Basin shows little or no evidence of botanical exploration. Therefore, geographical gaps and the small number of herbarium collections available impede accurate mapping of plant distributions and mapping biodiversity (Hopkins 2007).

Among the 59 alien plant species that are reported as invaders from 135 protected areas from around the world (in Foxcroft et al. 2013) there are many aquatic plants such as *Arundo donax*, *Hydrilla verticillata*, *Pistia stratiotes* and *Salvinia molesta*. According to our results and in combination with the USAqWRA scores, those species were the most invasive species in South America, prone to causes biodiversity loss in PAs.

The Human Influence Index could be considered a useful proxy to detect areas where the alien species could arrive and establish. Our results reflected that alien species were favoured in locations with a high effect of anthropisation. According to Gallardo et al. (2015) transport networks are at the moment one of the most important driver for the entry and the distribution of invasive plants (e.g. port proximity determined the presence of freshwater invaders) and directly linked to the vectors and pathways of introduction and secondary release for invasive species. These findings confirm that the relationship between invasive species and the human influence are quite important to explain highest risk values in areas where propagule pressure can be presumed high (i.e., close to transport networks and densely populated areas).

In accordance with García-Roselló et al. (2015), the inclusion of species in localities from which they had not been recorded by the use of predicted maps generally involves an increase in species richness. Extrapolations of individual species ranges, alternatively, do not appear to affect the geographical position of hotspots or patterns of global species richness.

## Conclusion

The GIBF and literature datasets provided significantly different information and the combination of the two offered new information and a better coverage that would not exist in a single data source. Nevertheless, a careful quality evaluation of the primary biodiversity information, both in the case of literature and GBIF should be conducted, before the data is used for further analyses in macroecological studies.

The identification of invasive species risk hotspots for aquatic invasive plant species could promote the development of prevention and control strategies. Particularly, the biodiversity hotspots and the protected areas should be efficiently prevented and

monitored. The human influence amplifying the potential for invasion could be translated into highest cumulative risk scores in close relation to the location of commercial ports, dense populated areas and intensely used landscapes. The methodology used in the present research, if applied on a larger dataset including all non-native species, could facilitate prevention and monitoring, at least for some regions of South America.

Finally, we would like to stress that GBIF data and tools are very valuable and important. However constant efforts at increasing sample sizes through the generation of new data and the publishing of existing datasets are particularly required of native and alien aquatic plants in South America.

## Acknowledgements

We gratefully thank A. Pauchard, J. Urrutia, R. Bustamante, S. Magela-Thomaz and L.J. Cumana Campos for providing useful information and literature. The University of Sassari Visiting Professor programme supported DC.

## References

- Adhikari D, Tiwary R, Barik SK (2015) Modelling Hotspots for Invasive Alien Plants in India. *PLoS ONE* 10: e0134665, <https://doi.org/10.1371/journal.pone.0134665>
- Amano T, Sutherland WJ (2013) Four barriers to the global understanding of biodiversity conservation: wealth, language, geographical location and security. *Proceedings of the Royal Society B* 280: 20122649, <https://doi.org/10.1098/rspb.2012.2649>
- Barthlott W, Biedinger N, Braun G, Feig F, Kier G, Mutke J (1999) Terminological and methodological aspects of the mapping and analysis of the global biodiversity. *Acta Botanica Fennica* 162: 103–110
- 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, <https://doi.org/10.1016/j.ecoinf.2013.11.002>
- Boakes EH, McGowan PJ, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol* 8: e1000385, <https://doi.org/10.1371/journal.pbio.1000385>
- Boltovskoy D, Correa N (2015) Ecosystem impacts of the invasive bivalve *Limnoperna fortunei* (golden mussel) in South America. *Hydrobiologia* 746: 81–95, <https://doi.org/10.1007/s10750-014-1882-9>
- Brummitt N, Bachman SP, Aletrari E, Chadburn H, Griffiths-Lee J, Lutz M, Moat J, Rivers MC, Syfert MM, Lughadha EMN (2015) The Sampled Red List Index for Plants, phase II: ground-truthing specimen-based conservation assessments. *Philosophical Transactions of the Royal Society B* 370: 20140015, <https://doi.org/10.1098/rstb.2014.0015>
- Brundu G (2015) Plant invaders in European and Mediterranean inland waters: profiles, distribution, and threats. *Hydrobiologia* 746: 61–79, <https://doi.org/10.1007/s10750-014-1910-9>
- Chapman AD (2005) Uses of primary species-occurrence data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen
- Collen B, Ram M, Zamin T, McRae L (2008) The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science* 1: 75–88, <https://doi.org/10.1177/194008290800100202>
- Delnatte C, Meyer JY (2012) Plant introduction, naturalization, and invasion in French Guiana (South America). *Biological Invasions* 14: 915–927, <https://doi.org/10.1007/s10530-011-0129-1>
- Donaldson JE, Hui C, Richardson DM, Robertson MP, Webber BL, Wilson JR (2014) Invasion trajectory of alien trees: the role of introduction pathway and planting history. *Global Change Biology* 20: 1527–1537, <https://doi.org/10.1111/gcb.12486>
- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kuhn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schroder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 606–628, <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
- Duursma DE, Gallagher RV, Roger E, Hughes L, Downey PO, Leishman MR (2013) Next-generation invaders? Hotspots for naturalised sleeper weeds in Australia under future climates. *PLoS ONE* 8: e84222, <https://doi.org/10.1371/journal.pone.0084222>
- Edwards JL, Lane MA, Nielsen ES (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289: 2312–2314, <https://doi.org/10.1126/science.289.5488.2312>
- Engelhardt KAM (2011) Eutrophication, aquatic. In: Simberloff D, Rejmánek M (eds), *Encyclopedia of Biological Invasions*. University of California Press, Berkeley, pp 209–213
- Feeley KJ, Silman MR (2009) Extinction risks of Amazonian plant species. *Proceedings of the National Academy of Sciences* 106: 12382–12387, <https://doi.org/10.1073/pnas.0900698106>
- Feeley KJ, Silman MR (2010) Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. *Journal of Biogeography* 37: 733–740, <https://doi.org/10.1111/j.1365-2699.2009.02240.x>
- Feeley KJ, Silman MR (2011) The data void in modeling current and future distributions of tropical species. *Global Change Biology* 17: 626–630, <https://doi.org/10.1111/j.1365-2486.2010.02239.x>
- Feeley KJ (2015) Are We Filling the Data Void? An Assessment of the Amount and Extent of Plant Collection Records and Census Data Available for Tropical South America. *PLoS ONE* 10: e0125629, <https://doi.org/10.1371/journal.pone.0125629>
- Foxcroft LC, Pyšek P, Richardson DM, Genovesi P (eds) (2013) Plant invasions in protected areas: patterns, problems and challenges. Springer, Dordrecht, pp 656, <https://doi.org/10.1007/978-94-007-7750-7>
- Foxcroft LC, Pyšek P, Richardson DM, Genovesi P, MacFadyen S (2017) Plant invasion science in protected areas: Progress and priorities. *Biological Invasions* 19: 1353–1378, <https://doi.org/10.1007/s10530-016-1367-z>
- Funk V, Zermoglio MF, Nasir N (1999) Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. *Biodiversity & Conservation* 8: 727–751, <https://doi.org/10.1023/A:1008877222842>
- Gallardo B, Zieritz A, Aldridge DC (2015) The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. *PLoS ONE* 10, e0125801, <https://doi.org/10.1371/journal.pone.0125801>
- García-Roselló E, Guisande C, González-Dacosta J, Heine J, Pelayo-Villamil P, Manjarrés-Hernández A, Vaamonde A, Granado-Lorencio C (2013) ModestR: a software tool for managing and analyzing species distribution map databases. *Ecography* 36: 1202–1207, <https://doi.org/10.1111/j.1600-0587.2013.00374.x>
- García-Roselló E, Guisande C, Heine J, Pelayo-Villamil P, Manjarrés-Hernández A, González Vilas L, González-Dacosta J, Vaamonde A, Granado-Lorencio C (2014) Using ModestR to download, import and clean species distribution records. *Methods in Ecology and Evolution* 5: 708–713, <https://doi.org/10.1111/2041-210X.12209>
- García-Roselló E, Guisande C, Manjarrés-Hernández A, González-Dacosta J, Heine J, Pelayo-Villamil P, González-Vilas L, Vari

- RP, Vaamonde A, Granado-Lorencio C (2015) Can we derive macroecological patterns from primary Global Biodiversity Information Facility data? *Global Ecology and Biogeography* 24: 335–347, <https://doi.org/10.1111/geb.12260>
- Gordon DR, Gantz CA, Jerde CL, Chadderton WL, Keller RP, Champion PD (2012) Weed risk assessment for aquatic plants: modification of a New Zealand system for the United States. *PLoS ONE* 7: e40031, <https://doi.org/10.1371/journal.pone.0040031>
- Guézou A, Chamorro S, Pozo P, Guerrero AM, Atkinson R, Buddenhagen C, Jaramillo Díaz P, Gardener M (2014) CDF Checklist of Galapagos Introduced Plants - FCD Lista de especies de Plantas introducidas Galápagos. In: Bungartz F, Herrera H, Jaramillo P, Tirado, N, Jiménez-Uzcátegui G, Ruiz D, Guézou A, Ziemmeck F (eds), Charles Darwin Foundation Galapagos Species Checklist - Lista de Especies de Galápagos de la Fundación Charles Darwin. Charles Darwin Foundation / Fundación Charles Darwin, Puerto Ayora, Galapagos: <http://www.darwinfoundation.org/datazone/checklists/introduced-pecies/introduced-plants/> (last updated: 21 January 2014)
- Hijmans RJ, Schreuder M, De la Cruz J, Guarino L (1999) Using GIS to check co-ordinates of genebank accessions. *Genetic Resources and Crop Evolution* 46: 291–296, <https://doi.org/10.1023/A:1008628005016>
- Hopkins MJ (2007) Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography* 34: 1400–1411, <https://doi.org/10.1111/j.1365-2699.2007.01737.x>
- Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of Biodiversity Databases: Case Study on Seed-Plant Diversity in Tenerife, Canary Islands. *Conservation Biology* 21: 853–863, <https://doi.org/10.1111/j.1523-1739.2007.00686.x>
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* 46: 10–18, <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Jetz W, McPherson JM, Guralnick RP (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution* 27: 151–159, <https://doi.org/10.1016/j.tree.2011.09.007>
- Kumschick S, Gaertner M, Vilà M, Essl F, Jeschke JM, Pyšek P, Ricciardi A, Bacher S, Blackburn TM, Dick JTA, Evans T, Hulme PE, Kühn I, Mrugała A, Pergl J, Rabitsch W, Richardson DM, Sendek A, Winter M (2015) Ecological impacts of alien species: quantification, scope, caveats and recommendations. *Bioscience* 65: 55–63, <https://doi.org/10.1093/biosci/biu193>
- Liang L, Clark JT, Kong N, Rieske LK, Fei S (2014) Spatial analysis facilitates invasive species risk assessment. *Forest Ecology and Management* 315: 22–29, <https://doi.org/10.1016/j.foreco.2013.12.019>
- Lozano V, Brundu G (2016) Prioritisation of aquatic invasive alien plants in South America with the US Aquatic Weed Risk Assessment. *Hydrobiologia* 1–16, <https://doi.org/10.1007/s10750-016-2858-8>
- Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A (2015) Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography* 24: 973–984, <https://doi.org/10.1111/geb.12326>
- Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. *Nature Communications* 6, <https://doi.org/10.1038/ncomms9221>
- McGeoch MA, Groom QJ, Pagad S, Petrosyan V, Ruiz G, Wilson J (2016) Data fitness for use in research on alien and invasive species. Copenhagen: GBIF Secretariat. <http://www.gbif.org/resourcere/82958> (accessed on 5 January 2017)
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858, <https://doi.org/10.1038/35002501>
- Pauchard A, Cavieres L, Bustamante R, Becerra P, Rapoport E (2004) Increasing the understanding of plant invasions in southern South America: first symposium on Alien Plant Invasions in Chile. *Biological Invasions* 6: 255–257, <https://doi.org/10.1023/B:BINV.0000022137.61633.09>
- Pelayo-Villamil P, Guisande C, González-Vilas L, Carvajal-Quintero JD, Jiménez-Segura LF, García-Roselló E, Heine J, González-Dacosta J, Manjarrés-Hernández A, Vaamonde A (2012) ModestR: Una herramienta informática para el estudio de los ecosistemas acuáticos de Colombia. *Actualidades Biológicas* 34(97): 225–239
- Perrings C, Dehnen-Schmutz K, Touza J, Williamson M (2005) How to manage biological invasions under globalization. *Trends in Ecology & Evolution* 20: 212–215, <https://doi.org/10.1016/j.tree.2005.02.011>
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* 78: 419–433, <https://doi.org/10.1086/378926>
- Petřík P, Pergl J, Wild J (2010) Recording effort biases the species richness cited in plant distribution atlases. *Perspectives in Plant Ecology, Evolution and Systematics* 12: 57–65, <https://doi.org/10.1016/j.ppees.2009.06.004>
- Ramírez C, San Martín C (2006) Diversidad de macrófitos chilenos. In: Vila I, Veloso A, Schlatter R, Ramírez C (eds), Macrófitos y vertebrados de los sistemas límnicos de Chile. Editorial Universitaria, Santiago, Chile, pp 358–363
- Ricciardi A, Rasmussen JB (1999) Extinction rates of North American freshwater fauna. *Conservation Biology* 13: 1220–1222, <https://doi.org/10.1046/j.1523-1739.1999.98380.x>
- Roberts T, Moritz A (2011) A framework for publishing primary biodiversity data. *BMC Bioinformatics* 12: 11, <http://www.biomedcentral.com/1471-2105/12/S15/11>
- Schulman L, Toivonen T, Ruokolainen K (2007) Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography* 34: 1388–1399, <https://doi.org/10.1111/j.1365-2699.2007.01716.x>
- Seebens H, Essl F, Dawson W, Fuentes N, Moser D, Pergl J, Pyšek P, Kleunen M van, Weber E, Winter M, Blasius B (2015) Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology* 21: 4128–4140, <https://doi.org/10.1111/gcb.13021>
- Seebens H, Blackburn T, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapo L, Dawson W, Dullinger S, Fuentes N, Jäger H, Kartesz J, Kenis M, Kreft H, Kühn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* 8: 14435, <https://doi.org/10.1038/ncomms14435>
- Simberloff D (2013) Invasive species: What everyone needs to know. Oxford University Press, New York, 352 pp
- Soberón J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 359: 689–698, <https://doi.org/10.1098/rstb.2003.1439>
- Sousa-Baena MS, Garcia LC, Peterson AT (2014) Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* 20: 369–381, <https://doi.org/10.1111/ddi.12136>
- Strayer DL (2010) Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology* 55: 152–174, <https://doi.org/10.1111/j.1365-2427.2009.02380.x>
- Suarez AV, Tsutsui ND (2004) The value of museum collections for research and society. *BioScience* 54: 66–74, [https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
- R Core Team (2015) R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org/>

- Thomaz SM, Kovalenko KE, Havel JE, Kats LB (2015) Aquatic invasive species: general trends in the literature and introduction to the special issue. *Hydrobiologia* 746: 1–12, <https://doi.org/10.1007/s10750-014-2150-8>
- Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* 11: 2234–2250, <http://dx.doi.org/10.1111/j.1365-2486.2005.001018.x>
- Troia MJ, McManamay RA (2016) Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution* 6(14): 4654–4669, <http://dx.doi.org/10.1002/eec3.2225>
- Urrutia EJ, Sánchez González P, Pauchard Cortés A (2016) Flora acuática y palustre introducida en Chile. Laboratorio de Invasiones Biológicas LIB, 87 pp
- van Kleunen M, Dawson W, Essl F, Pergl J, Winter M, Weber E, Kreft H, Weigelt P, Kartesz J, Nishino M, Antonova LA, Barcelona JF, Cabezas FJ, Cardenas D, Cardenas-Toro J, Castaño N, Chacón E, Chatelain C, Ebel AL, Figueiredo E, Fuentes N, Groom QJ, Henderson L, Inderjit, Kupriyanov A, Masciadri S, Meerman J, Morozova O, Moser D, Nickrent DL, Patzelt A, Peller PB, Baptiste MP, Poopath M, Schulze M, Seebens H, Shu WS, Thomas J, Velayos M, Wieringa JJ, Pyšek P (2015) Global exchange and accumulation of non-native plants. *Nature* 525: 100–103, <https://doi.org/10.1038/nature14910>
- Vollmar A, Macklin JA, Ford L (2010) Natural history specimen digitization: challenges and concerns. *Biodiversity Informatics* 7, <https://doi.org/10.17161/bi.v7i2.3992>
- Wildlife Conservation Society - WCS; Center for International Earth Science Information Network - CIESIN - Columbia University (2005) Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), <http://dx.doi.org/10.7927/H4M61H5F>
- World Conservation Union and UNEP-World Conservation Monitoring Centre (2007) World Database on Protected Areas, Version 2007, WCMC, Cambridge, UK, 2007, <https://www.unep-wcmc.org>
- Yang W, Ma K, Kreft H (2014) Environmental and socio-economic factors shaping the geography of floristic collections in China. *Global Ecology and Biogeography* 23: 1284–1292, <https://doi.org/10.1111/geb.12225>

### Supplementary material

The following supplementary material is available for this article:

**Table S1.** List of the investigated 80 aquatic plant species varying for native origin.

**Table S2.** List of the investigated 40 native and 40 non-native aquatic plant species according to GBIF and literature records.

**Table S3.** The 16 regions of South America considered in the present study and the taxonomic diversity of the 80 native and non-native aquatic plant species according to GBIF database and literature sources.

**Table S4.** Species data downloaded from GBIF dataset with the software ModestR.

**Table S5.** Qualitative a priori categorization for the 40 South American alien aquatic species evaluated with the US Aquatic Weed Risk Assessment scheme.

**Table S6.** Total scores for the 40 South American alien aquatic species evaluated with the US Aquatic Weed Risk Assessment scheme.

**Table S7.** Qualitative a priori categorization for the 40 South American native aquatic species for each region.

**Table S8.** GBIF and Literature records for South American regions (data clean).

**Table S9.** Contingency table between native/non-native records and inside/outside protected area.

**Table S10.** Script and output of Generalized Least Squares to evaluate the spatial autocorrelation in Human Influence Index data.

**Table S11.** Kriging parameters of estimated non-native species density according the number of different species in the 16 regions of South America.

**Figure S1.** Distribution of species records, downloaded from GBIF.

**Figure S2.** Choropleth maps on the density of aquatic species in each of the 16 regions of South America.

**Figure S3.** Box plots of the Human Influence Index (HII) for invasive alien and non-invasive alien species.

**Figure S4.** Box plots of the number of records in the protected areas for invasive alien and non-invasive alien species.

This material is available as part of online article from:

[http://www.reabic.net/journals/mbi/2017/Supplements/MBI\\_2017\\_Lozano\\_et\\_al\\_Supplementary\\_Tables.xlsx](http://www.reabic.net/journals/mbi/2017/Supplements/MBI_2017_Lozano_et_al_Supplementary_Tables.xlsx)

[http://www.reabic.net/journals/mbi/2017/Supplements/MBI\\_2017\\_Lozano\\_et\\_al\\_Supplementary\\_Figures.pdf](http://www.reabic.net/journals/mbi/2017/Supplements/MBI_2017_Lozano_et_al_Supplementary_Figures.pdf)