Received: 13 April 2023 | Revised: 22 July 2023 | Accepted: 31 August 2023

DOI: 10.1111/jvs.13209

Journal of Vegetation Science IAVS

REPORT

Vegetation of Southern Patagonia in the 1970s – Digitization of a gray-literature data set as a monitoring baseline in a changing world

Francesco Maria Sabatini[1,2] | Georg Hähn[3] | Karina Speziale[4] | Ana María Cingolani[5] | Gabriella Damasceno[3,6] | Helge Bruelheide[3,6]

[1]BIOME Lab, Department of Biological, Geological and Environmental Sciences (BiGeA), Alma Mater Studiorum University of Bologna, Bologna, Italy

[2]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Suchdol, Czech Republic

[3]Martin Luther University Halle-Wittenberg, Institute of Biology/ Geobotany and Botanical Garden, Halle (Saale), Germany

[4]INIBOMA (CONICET-UNCOMA), Department of Ecology, Bariloche, Argentina

[5]Instituto Multidisciplinario de Biología Vegetal (CONICET- UNC), Córdoba, Argentina

[6]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Leipzig, Germany

Correspondence
Francesco Maria Sabatini, BIOME Lab, Department of Biological, Geological and Environmental Sciences (BiGeA), Alma Mater Studiorum University of Bologna, Via Irnerio 42, 40126, Bologna, Italy.
Email: francescomaria.sabatini@unibo.it

Funding information
Deutsche Forschungsgemeinschaft; Deutsches Zentrum für integrative Biodiversitätsforschung Halle-Jena-Leipzig; Ministero dell'Istruzione, dell'Università e della Ricerca

Co-ordinating Editor: Hans Henrik Bruun

Abstract

Monitoring vegetation trends against objective baselines is fundamental to quantify the impacts of global change on plant biodiversity. Vegetation plot time series are a gold standard in vegetation monitoring, but such data are missing for many regions. Southern Patagonia is an example of a region strongly impacted by climate change but lacking time series data. Monitoring in this region could benefit from a comparison with vegetation survey data gathered between 1975 and 1979, as part of the multidisciplinary research program "Transecta botánica de la Patagonia austral" (hereafter Transecta). Published in 1985, it contains data on 668 vegetation plots, which were so far inaccessible to most researchers: Transecta has never been reprinted, nor fully digitized, and can only be found in specialized libraries. Here, we created a reproducible workflow, documenting how vegetation plot data from historical sources can be extracted and harmonized. The resulting open-access database we created fills a major regional gap and provides a needed baseline to assess the impacts of global change on southern Patagonia vegetation. By making these data available, we hope to inspire a new generation of vegetation scientists to resurvey the area and continue the legacy of the pioneer researchers who compiled Transecta.

KEYWORDS
biodiversity data, gray literature, historical baseline, Patagonia, relevés, vegetation plots

Francesco Maria Sabatini and Georg Hähn contributed equally.

1 | INTRODUCTION

We are in the midst of a global biodiversity crisis (Barber et al., 2020). The rate of extinction is now up to 1000 times higher than the geologically recent rates before human actions inflated them (De Vos et al., 2015), and more than a million species are estimated to be at risk (IPBES, 2019). Land use change is the main driver of this crisis (IPBES, 2019). Agriculture, deforestation and urbanization deeply shape all landscapes, with more than two thirds of terrestrial ecosystems being significantly altered by humans (Haddad et al., 2015; Díaz et al., 2019). The increasing impacts of other drivers of biodiversity loss, such as climate change, overexploitation of wild populations and the spread of exotic species, make the prospects for biodiversity even gloomier (Di Marco et al., 2019).

To understand the impact of global change on ecological communities, we need baseline historical data of local biodiversity across biomes, ecoregions and habitat types but these are often incomplete or geographically biased (Sutherland et al., 2004; Meyer et al., 2016; Sabatini et al., 2022). When objective data are not available for a specific region, there is the risk that our assessments of biodiversity status and trends are biased (Bowler et al., 2022) and suffer from the so-called shifting baseline syndrome (Pauly, 1995). This term indicates the tendency of humans to consider the state of the ecosystems encountered during their childhood (or early professional life, in case of scientists) as normal and natural, even if they were, in fact, depleted or in bad conditions already (Pauly, 1995). For instance, a large-scale online questionnaire comparing public perceptions with long-term biological change data in the UK showed that the perceptions of older participants are more in line with biological data, suggesting that older people might feel a more urgent need for conservation actions compared to younger people (Jones et al., 2020). Generation after generation, the shifting baseline syndrome might lead to lowering standards and to a severe underappreciation of environmental damage and biodiversity loss.

To avoid this generational trap, we need a precise quantification of the spatial patterns of biodiversity, and a detailed knowledge of how these have changed in time (Meyer et al., 2016). Historical biodiversity data are instrumental to this goal. The gold standard are vegetation plot time series, that is, vegetation plots resurveyed with a standard methodology over multiple decades (Jandt et al., 2022a). Time series have been widely used to document the impact of climate change (Steinbauer et al., 2018), land-use change, eutrophication (Diekmann et al., 2014; Ridding et al., 2020), and biological invasions (Petrášová et al., 2013), and for revealing that few winner species are locally expanding at the expense of many receding species (Jandt et al., 2022b). Yet, for many regions of the world vegetation time series are not available, which means that reconstructing an objective baseline of biodiversity remains difficult (Dornelas et al., 2018). A useful complement to vegetation time series could be provided by historical vegetation plot data contained in the gray literature, that is, reports, working papers, and government documents which are published outside of the traditional academic channels (Haddaway & Bayliss, 2015). Accessing gray literature is difficult, being rarely available in academic libraries, seldom accessible in digital format, and because of additional barriers related to language and copyright issues. For many regions, such as the tropics, there is often as much ecological and conservation-relevant information in the gray literature compared to peer-reviewed research (Corlett, 2011). Gray literature therefore represents an untapped source for obtaining information on past biodiversity distribution to be used as a monitoring baseline in many regions (Jandt et al., 2022a).

One such region is southern Patagonia. While vegetation time series are scarcely available for this region, a baseline knowledge of the flora and vegetation of southern Patagonia was gathered during an intensive, international, multidisciplinary research program conducted between 1975 and 1979. This program took the name "Transecta botánica de la Patagonia austral" ("Transecta" hereafter) and was led by D. Moore (Royal Society, UK), O. Boelcke (CONICET, Argentina) and E. Pisano (CONICYT, Chile). Over five consecutive growing seasons, a large team of botanists, vegetation scientists, geographers and geologists studied the distribution of vascular and non-vascular flora in a 450 km long and 55 km wide transect going from the Atlantic to the Pacific Ocean between 51° and 52° S latitude. Altogether, they collected more than 9000 plant vouchers, and surveyed several hundred vegetation plots across diverse habitats and environmental conditions according to the Braun-Blanquet method, and built vegetation maps at a scale of 1:250,000. All results were published in 1985 (Boelcke et al., 1985).

Since its publication, Transecta has had a major impact on local biodiversity research. It provided a much needed framework for understanding the main aspects of the vegetation landscape of the area (León et al., 1998), as a diverse mosaic of steppe (Romo et al., 2012), forest (Damascos, 1997; Promis et al., 2008) and wetland (Blanco & de la Balze, 2004; Kleinebecker & Vogel, 2008; San Martín et al., 2013). The knowledge gathered in Transecta was also used for considerations on the land use potential for forestry and pasture (Seibert, 1987), fodder production (Olivares Espinoza & Schmidt, 2008) or livestock farming (Oliva et al., 2012). It also served as a baseline for assessing ecosystem services, such as carbon storage potential (Peri, 2011), and long-term recovery of the vegetation after a large fire (Quintanilla, 2008). Yet, the reach of the book remained limited to regional research, with very little resonance in the international literature.

Five decades later, the data collected by Transecta have become invaluable for benchmarking the impact of global change on the vegetation, especially considering the major climatic changes that southern Patagonia is facing (Castillo Marín, 2016). The Andean part of Patagonia, for instance, has already experienced more than 1°C of temperature increase over the last decades, yearly precipitation is decreasing and almost all of the glaciers are retiring, with serious repercussions on water availability (Castillo Marín, 2016). Yet, as far as we know, the vegetation data contained in Transecta have not been used to test these or other impacts of global change, probably due to their limited accessibility: the final report of Transecta has never been translated to English and it is available only in paper format in few specialized libraries. Most importantly, the data contained therein have never been digitized, with few exceptions (Pliscoff et al., 2008; Alvarez & Luebert, 2022), which limits their use by a wider community of vegetation scientists.

A questionnaire we sent around to Argentinian botanists ($n = 31$) and environmental scientists ($n = 25$) confirmed that the impact of the book has been hindered by its limited accessibility. Only 14 out of 56 respondents know the book, and only seven have access to it. The average age of those owning a hardcopy was 60 years old (range 43–69), which highlights that the accessibility of Transecta has a generational bias. Most botanist respondents answered that they would (12/31) or probably would (16/31) use the data if available, and these proportions were equally high among non-botanist respondents (9/25 and 11/25, respectively). Moreover, 15 respondents commented on the importance of the survey carried out by the Transecta project as a key reference about the Patagonian environment, stressing the need of making the data available.

In short, Transecta clearly qualifies as a cornerstone of vegetation research in Patagonia. Here, we (1) created a workflow for extracting, digitizing and harmonizing the vegetation plot data from the final report; (2) made the data available in an open-access online database; and (3) conducted some exploratory analysis to sketch its content and to inspire future use. By making the data available, we contribute to overcoming the main obstacle which limited the scientific and social impact of Transecta in the last four decades, that is, the arduous accessibility of the final report in the pre-digital era.

2 | DATABASE ORGANIZATION

The Transecta database contains information on 668 vegetation plots, and is formed by two main matrices, which are relationally linked through the key column "PlotID." Plot IDs are coded as a string of table number, group number and plot number, separated by an underscore (e.g., "txx_gxx_pxxx"), and correspond to the numbering in Transecta (Roig & Faggi, 1985). The spatial distribution of the plots is indicated in Figure 1.

All plot-level information is contained in the "header" matrix, which includes metadata and sampling design information (Appendix S1). When available, it also contains information on the environmental conditions of the plot, and a general description of the vegetation structure. Based on the species composition and the ancillary information, we attributed each vegetation plot to its formation type, according to a global hierarchical classification system that combines vegetation attributes (physiognomy, structure, and floristics) and their response to ecological and biogeographic factors (Faber-Langendoen et al., 2016). This hierarchical classification was complemented with five non-mutually-exclusive Boolean categories: "Forest," "Shrubland," "Grassland," "Wetland," "Sparse vegetation," where intermediate vegetation types are indicated as combinations of categories (e.g., Savanna: Forest = True & Grassland = True). A summary of all the 26 variables in the header matrix is provided in Appendix S1.

The "**DT**" matrix contains information on the species composing each plot and their respective abundances. It is structured in long format and totals 9509 records for 589 species. For each record we reported both the species name as originally reported in Transecta, as well as the taxon name resolved after taxonomic standardization, a process described in the "Technical validation" section, below. Each record also contains an abundance value, following the Braun-Blanquet cover–abundance ordinal scale (Braun-Blanquet, 1928). The **DT** matrix also contains four additional columns, reporting on the taxonomic group of the species, the height of some selected species as estimated in the field, the taxonomic level at which the original species names were matched during standardization, and

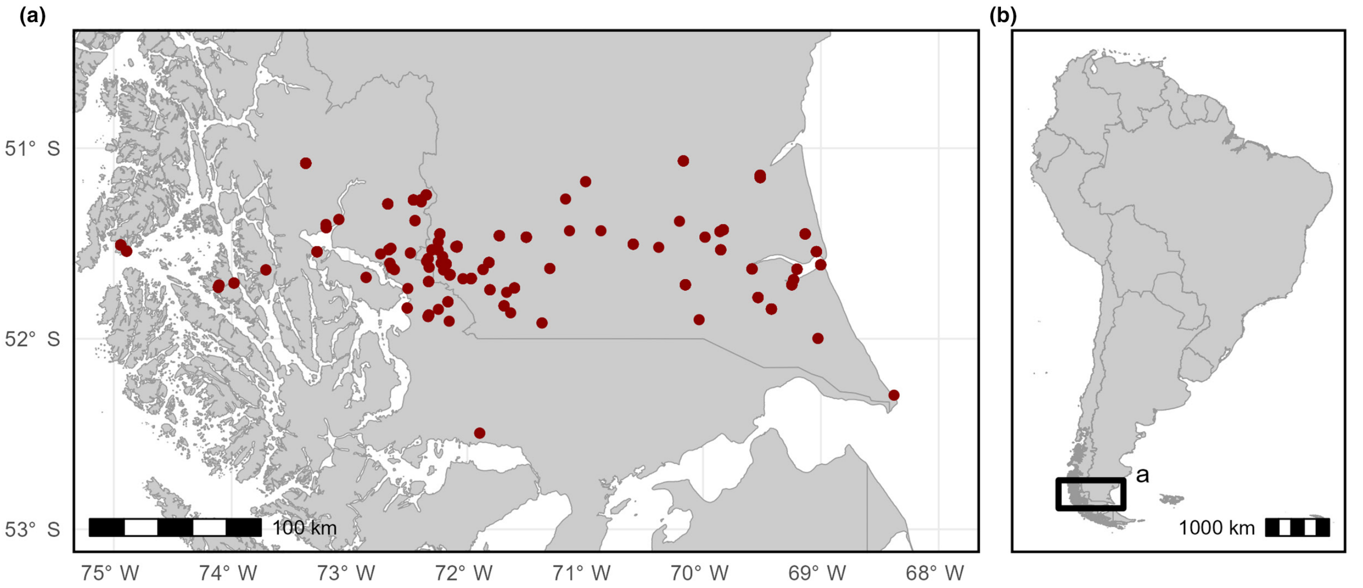

FIGURE 1 Spatial distribution of vegetation plots digitized from the book Transecta botánica de la Patagonia austral (Boelcke et al., 1985). Due to the uncertain spatial location of many plots, each dot in the map might be associated with more than one vegetation plot

some occasional species level notes, for instance related to the growth stage, or to distinguish congeneric species not identified at species level across plots, or to indicate species with particular indicator value from a phytosociological point of view. A description of all variables contained in the DT matrix is provided in Appendix S1.

3 | TECHNICAL VALIDATION

The vegetation plot data were extracted from the data table in Chapter 13 of Transecta (Roig & Faggi, 1985). After scanning the tables, we ran an optical character recognition (OCR) algorithm in ADOBE Acrobat for extracting all characters, followed by a thorough manual revision ensuring the table structure (plots by species, with abundances at the row–column intersection) was accurately reproduced. Additional information from the text was added on a plot-level basis, if available.

The locations of all relevés were manually extracted from the toponyms reported in the printed tables in Transecta or annex paragraphs describing the data. Each location was then georeferenced using either Google Maps (maps.google.com) or the Directorio cartográfico de España y América (www.dices.net). For each location, we also estimated the location uncertainty, that is, the estimated precision of the geographical coordinates (in km), depending on our level of confidence in the identification of the correct location. The location uncertainty was quite coarse, being on average (median) equal to 5 km. The best and worst location uncertainties were equal to 0.5 and 250 km, respectively. The highest uncertainty was assigned to a set of plots ($n=57$) without toponyms. For some of these plots ($n=39$), we could ascertain from the text whether they were on the Argentinian or Chilean side. In this case, we manually assigned the coordinates of the centroid of the Argentinian or Chilean side of Transecta, respectively, with a location uncertainty equal to the radius of the smallest circle including the whole Transecta region in that country (200 km, in both cases). For the remaining $n=18$ plots, we assigned the centroid of the whole Transecta, with a location uncertainty equal to 250 km. All spatial coordinates were visually checked, to ensure that they were plausible, and within the spatial range of Transecta (Figure 1).

The information required to classify each plot into its respective formation was also extracted manually, and the classification of plots into their formation type (Faber-Langendoen et al., 2016) was expert-based, based on (1) the global distribution of formation types; (2) the physiognomy of plots, as described in the main text of Transecta; and (3) the dominant species of each plot. A cross-table linking the vegetation physiognomy reported in Transecta to the world formation types is provided in Appendix S2.

We then submitted these initial data to an R Markdown script to (1) check for and correct typos in the species names; (2) assign each bit of information from the original table to the correct column in the output table, identifying which entries related to, for instance, species abundance, species height or plot-level information; (3) extract missing plot-level data (e.g., slope or aspect) from the table notes, using string recognition functions; (4) check all plant species names against the Argentinian checklist (Zuloaga et al., 2019) and resolve species synonyms; (5) check unresolved species names ($n=201$) with The Plant List through the Taxonstand Package (version 2.4, accessed March 24, 2023, Cayuela et al., 2021). Unresolved species names ($n=83$) were left as they appear in Transecta; and (6) create clean outputs of all data (header + DT tables).

4 | EXPLORATORY ANALYSIS

Most of the plots belonged to herbaceous and shrub vegetation, especially temperate grasslands and shrubland ($n=235$, Figure 2). Scrub and herb coastal vegetation (which includes matorral vegetation) was sampled in 56 plots. About one fourth of the plots represent forest ecosystems, mostly forest and woodland ($n=145$) and secondarily flooded and swamp forest ($n=11$). Wet freshwater marsh, wet meadows and shrublands were sampled in 80 plots, bogs and salt marshes were sampled in 48 and four plots, respectively. The remaining plots were in alpine tundra ($n=72$) and cliff, scree and other rock vegetation ($n=22$). Plots in different formations differed in average species richness (Figure 3). Grassland and shrubland plots were the most species-rich communities (median = 19; interquartile range [IQR] = 11), followed by alpine tundra (median = 16; IQR = 10.8) and forest and woodland (median = 13; IQR = 7). Communities having the lowest species richness were those of salt marshes (median = 5; IQR = 0.75) and cliff, scree and other rock vegetation (median = 6; IQR = 4.5). The overlap across formations was high. It must be noted, however, that the actual size of the vegetation plot sampled in the field is only rarely reported in Transecta. Different formations are likely to have been sampled in vegetation plots of different sizes, as is a tradition in phytosociological studies. Also given the uneven

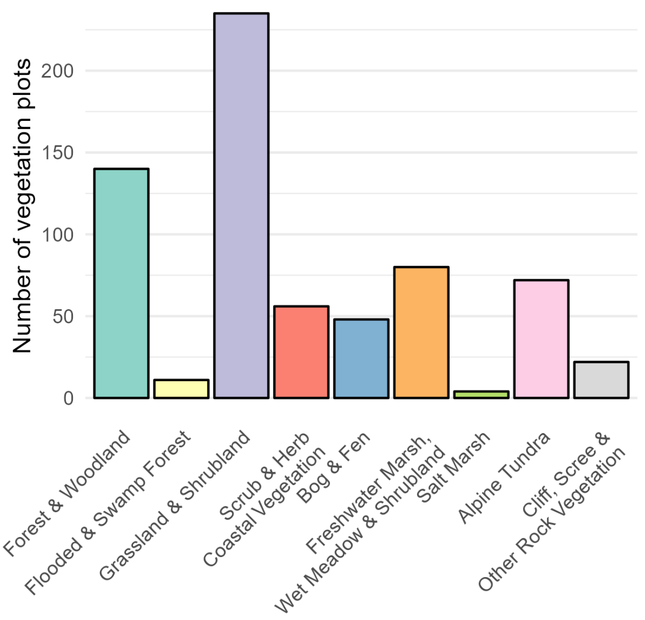

FIGURE 2 Number of plots per vegetation type.

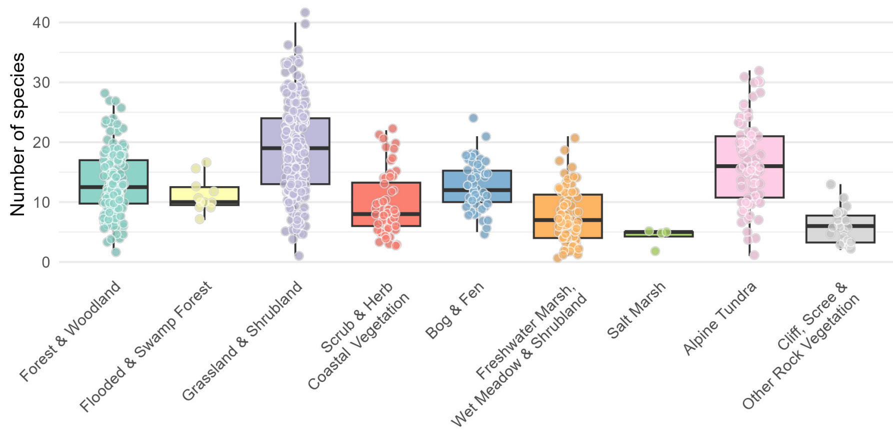

FIGURE 3 Species richness per vegetation type. Each dot represents a vegetation plot.

sampling size across formations, these comparisons should be taken as merely indicative.

To show the compositional similarities among plots in different formations, we calculated a transformation-based principal component analysis (tb-PCA), where species cover data were pre-transformed using the Hellinger transformation (Legendre & Gallagher, 2001). To interpret the principal components (PCs), we passively projected explanatory variables using the "envfit" function in the *vegan* package in R (Oksanen et al., 2022). We used a set of five explanatory variables: latitude, longitude, elevation, yearly mean temperature and yearly mean precipitation. The approximate elevation of each vegetation plot was extracted based on the plot spatial coordinates using the *elevatr* package in R (Hollister, 2021). We did so because elevational data, as measured in the field, are only rarely reported in the data set. Elevation was extracted with a zoom scale = 4, corresponding to an approximately 7-km resolution at 45° latitude, which we deemed adequate given the median location uncertainty of our data. We extracted climatic data from *WorldClim v2.1* using the "worldclim_global" function from the *geodata* package (Hijmans et al., 2023), setting a resolution of 10 arcmin (~10 km). While we excluded all plots having a location uncertainty above 20 km from the following analyses, the remaining spatial mismatch between the geographical uncertainty of the most imprecisely located plot and the resolution of the elevation and climatic data requires the results to be interpreted with caution.

The plots in Transecta encompassed a large compositional variability, with the first three axes of the tb-PCA only accounting for 17.5% of explained variance. PC1 clearly discriminated between grassland and shrubland plots vs the other vegetation types and aligned well with a gradient in elevation and temperature. PC2 discriminated flooded and swamp forests vs forest and woodland, and mostly represented a gradient in longitude and precipitation. Alpine tundra plots clearly stood out at the low end of PC3, which represents a latitudinal gradient. All predictors were significantly associated to one of the PCs, but only longitude and precipitation had an R^2 above 0.1 (Figure 4).

The spatial pattern of formations across Transecta reflects the distribution of the main vegetation types in southern Patagonia (Figure 5b). While the Atlantic side of Transecta was mostly represented by grassland and shrubland plots, with local occurrences of cliff, scree and rock vegetation (e.g., in the Cerro Norte area), the central and western parts were more often covered by forest ecosystems. Most forest plots occurred in proximity to the Argentinean–Chilean border, as well as in high-elevation sites on the islands of the Queen Adelaide Archipelago. Matorral coastal vegetation occurred on both sides of Transecta.

We also observed a pattern in the geographical distribution of species richness along the east–west gradient (Figure 5a). Species richness was relatively low in the plots on the Atlantic coast. Proceeding toward the inland, it increased to a maximum toward the central part of Transecta (that is, between 71 and 72° W). On the Chilean coast, species richness decreased slowly from the continental coast toward the outer islands of the Queen Adelaide Archipelago.

Terrain elevation varied across the longitudinal gradient, reaching a maximum in the middle of the transect where the species richness was highest. To further investigate whether elevation is a determinant of species richness, we used linear regression. The relationship between species richness and elevation was weakly negative (−0.1 species per 100 m elevation) and significant ($p < 0.001$), but the explained variation was negligible (Adj-$R^2 = 0.02$). Inspired by the findings of Kambach et al. (2023) we also tested an alternative model including the interaction between elevation and formation. This model returned a much higher explained variation (Adj-$R^2 = 0.36$, Appendix S3). Elevation had a marginally significant overall negative effect on species richness (−0.6 species per 100 m elevation, $p < 0.001$), but the interaction with formation was highly significant, as revealed by comparing the two models with an ANOVA test ($F = 22.11$, $p < 0.001$), showing that the relationship between elevation and species richness is habitat-dependent (Figure 6). The richness of forests decreased with elevation while the richness of grasslands increased, at least within the elevational range of our data.

We reported pairwise comparisons of regression coefficients in Appendix S5, as calculated using the "emtrends" function in the *emmeans* package (Lenth, 2023). The only significant contrasts, based on Tukey-adjusted *p*-values for multiple testing, were those comparing alpine tundra to forest & woodland, scrub & herb coastal vegetation, and freshwater marsh, wet meadow & shrublands.

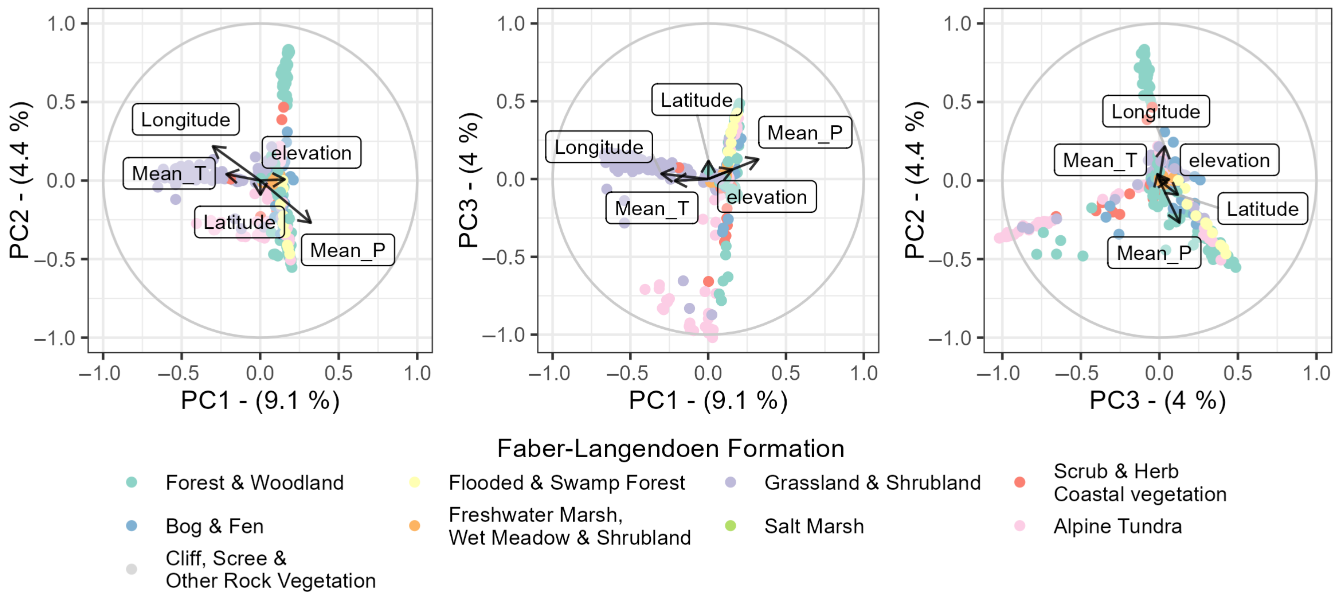

FIGURE 4 Transformation-based principal component analysis (tb-PCA) where species cover data were pre-transformed using the Hellinger transformation. Environmental predictors were passively projected using the "envfit" function in *vegan*. Numbers in parentheses next to the axis labels represent explained variation (in percentage).

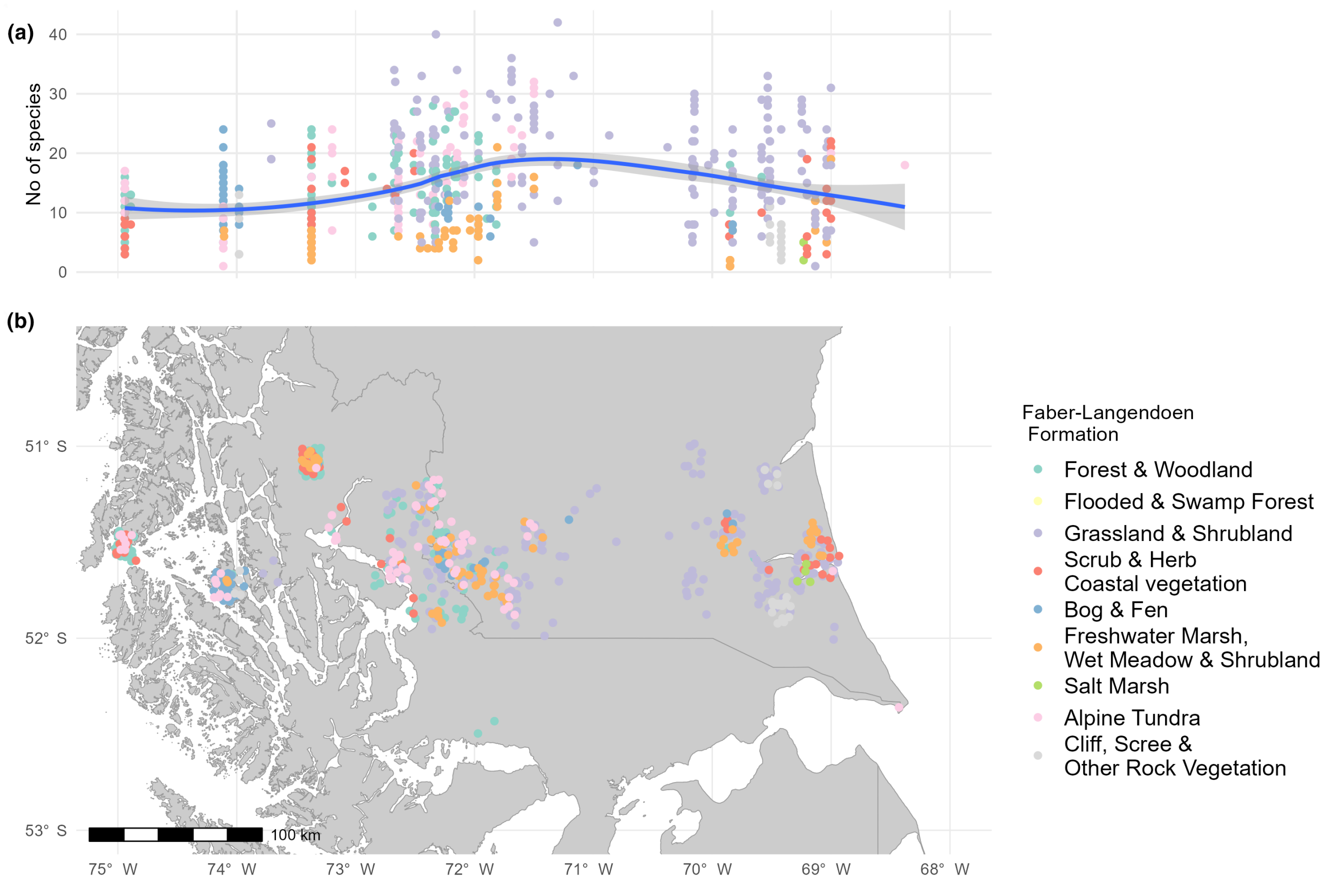

FIGURE 5 (a) Species richness per vegetation type across the longitudinal gradient. A LOWESS (Locally Weighted Scatterplot Smoothing) smoother with 95% confidence intervals was fitted to the scatterplot. (b) Spatial distribution of plots in the study area, classified based on their formation. Plots having the same coordinates (for lack of better information) were jittered in order to show the diversity of formations. Plots with unknown location (therefore having a location uncertainty of 200 km or higher) were removed.

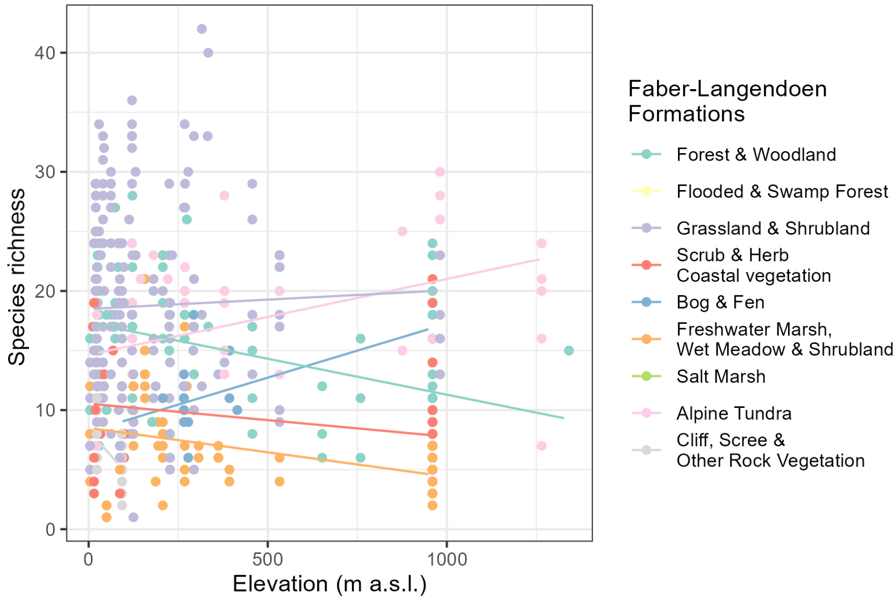

FIGURE 6 Modeled relation between elevation and species richness across different formations. A colorblind-friendly version of this graph is available in Appendix S4.

5 | DATA USAGE

The digital version of the Transecta database can be downloaded from https://doi.org/10.25829/idiv.3554-ks7d98 and used without limitation. The database has also been included in the Global Index of Vegetation Databases (GIVD): https://www.givd.info/ID/S-A-00-007. Data are distributed according to a CC-BY license. Users are asked to cite both the original data source (Roig & Faggi, 1985. *Transecta botánica de la Patagonia austral. Análisis geobotánico de la vegetación. Consejo Nacional Investigaciones Científicas y Técnicas, Instituto de La Patagonia y The Royal Society, Buenos Aires*) and this work. Please note that we do not claim any ownership of the original data here. Rather, our intellectual contribution is limited to the procedure for the semi-automatic extraction of publicly available data, and to the procedures of data validation and harmonization.

The database contains two main files:

1. 3554_DT-Transecta-Patagonia-v04-2023.csv
2. 3554_Header-Transecta-Patagonia-v04-2023.csv

The header and DT tables are also provided in.Rdata format (R-Image-Transecta-Patagonia-v04-2023.Rdata), to facilitate R users.

All scripts used to extract the information from the OCR-recognized text, as well as to clean, harmonize and explore the data are available on GitHub (https://github.com/georghaehn/Transecta-Patagonia-Digitalization). All code is distributed under a GNU General Public License v3.0. We provide two R scripts and one html report (R Core Team, 2022):

3. 00.Transecta-Patagonia_processing.Rmd
4. 00.Transecta-Patagonia_processing.html
5. 01.GraphsTables_Analysis.R

The first is an R Markdown script documenting the whole data creation and cleaning workflow. It comes with the knitted .html computing notebook, commenting and showing each step. The third file is for reproducing analyses, graphs and tables presented in this manuscript.

6 | EXPECTED IMPACTS AND LIMITATIONS

The latest years have seen a new wave of macroecological studies, upscaling ecological questions normally investigated at fine spatial scale to continental or even global extents (Cai et al., 2021, 2023; Sporbert et al., 2021; Testolin et al., 2021a, 2021b; Graco-Roza et al., 2022; Kambach et al., 2023). The reason for this renaissance is the aggregation of regional and global databases of once-scattered biodiversity data (Pärtel et al., 2022). Yet, most of this recent work is based on a biased distribution of available biodiversity data, with most of the data being located in developed countries, while many regions of the Global South remain underrepresented (Sabatini et al., 2021, 2022).

Our work fills one of these regional gaps, southern Patagonia. Together with other data existing for the region (Collantes et al., 1999; Anchorena & Cingolani, 2002; Pliscoff et al., 2008; Alvarez & Luebert, 2022), we expect our work to be helpful in both regional and global studies dealing with plant diversity patterns, vegetation change and the impacts of land use and climate change on vegetation. We also hope these data may inspire a new generation of vegetation scientists to resample some of the plots digitized here, therefore continuing the legacy of the pioneer researchers who collected the data in the first place. By distributing the data open access, we hope they can also be included in international vegetation plot databases, such as sPlot (Bruelheide et al., 2019), to strengthen future macroecological and biodiversity research.

Even if due care was paid when creating this database, it does not come without limitations. The first limitation relates to the uncertain geographical position of many vegetation plots. Relocating vegetation plots can be tricky, especially when data were collected in the pre-digital era and do not come with GPS coordinates. For each plot, we provide an estimate of its location uncertainty. When trying to extract ancillary climatic, soil or topographic data based on the plot locations, users should be aware of this problem and avoid neglecting this potential source of bias.

The second limitation relates to the many missing data, especially with regard to plot-level information. For many plots, there are very limited data on the structure of the vegetation, the actual soil conditions and even on the area surveyed in the vegetation plot. While this might limit the usefulness of these data for ecological research, we believe that even just data on the presence, absence and co-occurrence of plant species can deliver significant ecological information (Siefert et al., 2023) and provide a baseline for assessing global change impacts on vegetation since the 70s.

7 | CONCLUSIONS

To quantify the negative consequences of global change for biodiversity we need data to compare the current situation to a baseline. Vegetation scientists collected a huge amount of such data in the 20th century. Not all these data have already been digitized, harmonized and made accessible, however. Mobilizing data from the gray literature, especially from those regions which are currently underrepresented in global biodiversity databases, is crucial to fill critical data gaps and get a more complete and less biased understanding of biodiversity patterns and trends (Corlett, 2011). Here, we have mobilized 668 vegetation plot data, which were previously mostly stored only in paper form, and have made these data easily accessible to the scientific community. While these data are extremely valuable per se, to facilitate their digitization we also produced semi-automatic procedures for text mining, data parsing, harmonization and taxonomic standardization. While not immediately transferable to other data sources, these procedures can serve as a blueprint and facilitate other researchers trying to mobilize other historical biodiversity data stored in the gray literature of other underinvestigated regions.

AUTHOR CONTRIBUTIONS

Francesco Maria Sabatini and Helge Bruelheide designed the study. Georg Hähn digitized the data and created the workflow for cleaning and harmonizing the data set, with significant input by Francesco Maria Sabatini. Karina Speziale and Ana María Cingolani provided the species list and checked for consistency of taxonomic attributions. Francesco Maria Sabatini analyzed the data and created figures and tables, with significant contributions by Georg Hähn, Gabriella Damasceno and Helge Bruelheide. Francesco Maria Sabatini wrote the first draft with the support of Georg Hähn. All authors contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

Francesco Maria Sabatini gratefully acknowledges the financial support the Italian Ministry of University and Research, within the Rita-Levi Montalcini 2019 programme. We appreciate the support of the German Research Foundation for funding sPlot as one of the iDiv strategic projects (DFG FZT 118, 202548816). Francesco Maria Sabatini, Georg Hähn, Gabriella Damasceno and Helge Bruelheide were funded under this scheme. Karina Speziale acknowledges PICT 2018-1623. Finally, we would like to thank all the people who contributed to the design, data collection, analysis and reporting of Transecta project for gathering such an inestimable legacy of plant diversity data.

FUNDING INFORMATION

Francesco Maria Sabatini gratefully acknowledges the financial support the Italian Ministry of University and Research, within the Rita-Levi Montalcini 2019 program. We appreciate the support of the German Research Foundation for funding sPlot as one of the iDiv strategic projects (DFG FZT 118, 202548816). Francesco Maria Sabatini, Georg Hähn, Gabriella Damasceno and Helge Bruelheide were funded under this scheme.

DATA AVAILABILITY STATEMENT

The digital version of Transecta database can be downloaded from https://doi.org/10.25829/idiv.3554-ks7d98 and used without limitation. Data are distributed according to a CC-BY license.

All scripts used to extract the information from the OCR-recognized text, as well as to clean, harmonize and explore the data are available on GitHub (https://github.com/georghaehn/Transecta-Patagonia-Digitalization). All code is distributed under a GNU General Public License v3.0.

ORCID

Francesco Maria Sabatini https://orcid.org/0000-0002-7202-7697
Georg Hähn https://orcid.org/0000-0003-3733-1498
Karina Speziale https://orcid.org/0000-0003-2224-2097
Ana María Cingolani https://orcid.org/0000-0002-6173-1814
Gabriella Damasceno https://orcid.org/0000-0001-5103-484X
Helge Bruelheide https://orcid.org/0000-0003-3135-0356

REFERENCES

Alvarez, M. & Luebert, F. (2022) Chilean vegetation in the context of the Braun-Blanquet approach and a comparison with EcoVeg formations. *Vegetation Classification and Survey*, 3, 45–52.

Anchorena, J. & Cingolani, A. (2002) Identifying habitat types in a disturbed area of the forest-steppe ecotone of Patagonia. *Plant Ecology*, 158, 97–112.

Barber, C., Petersen, R., Young, V., Mackey, B. & Kormos, C. (2020) The nexus report: nature based solutions to the biodiversity and climate crisis. *F20 Foundations, Campaign for Nature and SEE Foundation*. Available from: http://foundation.see.org.cn/uploads/pdf/2020.pdf

Blanco, D.E. & de la Balze, V.M. (2004) *Los Turbales de la Patagonia. Bases para su inventario y la conservación de su biodiversidad*. Buenos Aires, Argentina: Wetlands International-América del Sur.

Boelcke, O., Moore, D. & Roig, F. (1985) *Transecta botánica de la patagonia austral*. Buenos Aires, Argentina: Consejo Nacional de Investigaciones Científicas y Técnicas (AR), Instituto de la Patagonia (CL), Royal Society (UK).

Bowler, D.E., Callaghan, C.T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C. et al. (2022) Temporal trends in the spatial bias of species occurrence records. *Ecography*, 2022, e06219.

Braun-Blanquet, J. (1928) *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Wien: Springer Verlag.

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S.M., Chytrý, M. et al. (2019) sPlot – a new tool for global vegetation analyses. *Journal of Vegetation Science*, 30, 161–186.

Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F. et al. (2023) Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist*, 237, 1432–1445.

Cai, Q., Welk, E., Ji, C., Fang, W., Sabatini, F.M., Zhu, J. et al. (2021) The relationship between niche breadth and range size of beech (Fagus) species worldwide. *Journal of Biogeography*, 48, 1240–1253.

Castillo Marín, N. (2016) *El cambio Climatico en Argentina*. Buenos Aires, Argentina: Agencia de Cooperación Internacional del Japón (JICA) & Dirección de Cambio Climático Argentina.

Cayuela, L., Macarro, I., Stein, A. & Oksanen, J. (2021) *Taxonstand: taxonomic standardization of plant species names*. R package version 2.4. Available from: https://CRAN.R-project.org/package=Taxonstand

Collantes, M.B., Anchorena, J. & Cingolani, A.M. (1999) The steppes of Tierra del Fuego: floristic and growthform patterns controlled by soil fertility and moisture. *Plant Ecology*, 140, 61–75.

Corlett, R.T. (2011) Trouble with the gray literature. *Biotropica*, 43, 3–5.

Damascos, M. (1997) Formas de crecimiento en los bosques patagónicos del sur de Argentina y Chile. *Revista Chilena de Historia Natural*, 70, 465–479.

De Vos, J.M., Joppa, L.N., Gittleman, J.L., Stephens, P.R. & Pimm, S.L. (2015) Estimating the normal background rate of species extinction. *Conservation Biology*, 29, 452–462.

Di Marco, M., Harwood, T.D., Hoskins, A.J., Ware, C., Hill, S.L. & Ferrier, S. (2019) Projecting impacts of global climate and land-use scenarios on plant biodiversity using compositional-turnover modelling. *Global Change Biology*, 25, 2763–2778.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A. et al. (2019) Pervasive human-driven decline of life on earth points to the need for transformative change. *Science*, 366, eaax3100.

Diekmann, M., Jandt, U., Alard, D., Bleeker, A., Corcket, E., Gowing, D.J.G. et al. (2014) Long-term changes in calcareous grassland vegetation in North-Western Germany – no decline in species richness, but a shift in species composition. *Biological Conservation*, 172, 170–179.

Dornelas, M., Antão, L.H., Moyes, F., Bates, A.E., Magurran, A.E., Adam, D. et al. (2018) BioTIME: a database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, 27, 760–786.

Faber-Langendoen, D., Keeler-Wolf, T., Meidinger, D., Josse, C., Weakley, A., Tart, D. et al. (2016) *Classification and description of world formation types*. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A.T.R., Alahuhta, J., Altman, J. et al. (2022) Distance decay 2.0 – a global synthesis of taxonomic and functional turnover in ecological communities. *Global Ecology and Biogeography*, 31, 1399–1421.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D. et al. (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1, e1500052.

Haddaway, N.R. & Bayliss, H.R. (2015) Shades of grey: two forms of grey literature important for reviews in conservation. *Biological Conservation*, 191, 827–829.

Hijmans, R.J., Barbosa, M., Ghosh, A. & Mandel, A. (2023) *geodata: Download geographic data*. R package version 0.5-8. Available from: https://CRAN.R-project.org/package=geodata

Hollister, J.W. (2021) *elevatr: access elevation data from various APIs*. R package version 0.4.1. Available from: https://CRAN.R-project.org/package=elevatr/

IPBES (Ed.). (2019) (Ed.)*Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services*. Bonn, Germany: IPBES secretariat.

Jandt, U., Bruelheide, H., Berg, C., Bernhardt-Römermann, M., Blüml, V., Bode, F. et al. (2022a) ReSurveyGermany: vegetation-plot time-series over the past hundred years in Germany. *Scientific Data*, 9, 631.

Jandt, U., Bruelheide, H., Jansen, F., Bonn, A., Grescho, V., Klenke, R.A. et al. (2022b) More losses than gains during one century of plant biodiversity change in Germany. *Nature*, 611, 512–518. Available from: https://doi.org/10.1038/s41586-022-05320-w

Jones, L.P., Turvey, S.T., Massimino, D. & Papworth, S.K. (2020) Investigating the implications of shifting baseline syndrome on conservation. *People and Nature*, 2, 1131–1144.

Kambach, S., Sabatini, F.M., Attorre, F., Biurrun, I., Boenisch, G., Bonari, G. et al. (2023) Climate-trait relationships exhibit strong habitat specificity in plant communities across Europe. *Nature Communications*, 14, 712.

Kleinebecker, T. & Vogel, A. (2008) Die Vegetation ombrotropher Moore Südpatagoniens. *Abhandlungen aus dem Westfälischen Museum für Naturkunde*, 70, 439–452.

Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280.

Lenth, R. (2023) *emmeans: estimated marginal means, aka least-squares means*. R package version 1.8.7. Available from: https://CRAN.R-project.org/package=emmeans

León, R.J.C., Bran, D., Collantes, M., Paruelo, J.M. & Soriano, A. (1998) Grandes unidades de vegetación de la Patagonia extra andina. *Ecología Austral*, 8, 125–144.

Meyer, C., Weigelt, P. & Kreft, H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006.

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P. et al. (2022) *vegan: community ecology package*. R package version 2.6-4. Available from: https://CRAN.R-project.org/package=vegan

Oliva, G., Ferrante, D., Puig, S. & Williams, M. (2012) Sustainable sheep management using continuous grazing and variable stocking rates in Patagonia: a case study. *The Rangeland Journal*, 34, 285–295.

Olivares Espinoza, A. & Schmidt, H. (2008) Einfluss des deckungsgrades eines ñirre (Nothofagus Antarctica)-bestandes auf wachstum und qualität der darunterliegenden grasweide im chilenischen Patagonien = Effect of forest canopy cover of a Nothofagus Antarctica stand on growth and quality of grass pasture in the Chilean Patagonia. *Forstarchiv*, 79, 83–88.

Pärtel, M., Sabatini, F.M., Morueta-Holme, N., Kreft, H. & Dengler, J. (2022) Macroecology of vegetation — lessons learnt from the virtual special issue. *Journal of Vegetation Science*, 33, e13121.

Pauly, D. (1995) Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology & Evolution*, 10, 430.

Peri, P.L. (2011) Carbon storage in cold temperate ecosystems in southern Patagonia, Argentina. In: Atazedeh, I. (Ed.) *Biomass and remote sensing of biomass*. Rijeka, Croatia: InTech Publisher, pp. 213–226.

Petrášová, M., Jarolímek, I. & Medvecká, J. (2013) Neophytes in Pannonian hardwood floodplain forests – history, present situation and trends. *Forest Ecology and Management*, 308, 31–39.

Pliscoff, P., Luebert, F. & de, D. (2008) *Digitalización de datos e información, depuración y estandarización de Pisos de Vegetación de Chile*. Santiago, Chile: Informe Final IABIN Ecosystem Grant, Corporación taller la Era.

Promis, Á., Cruz, G., Reif, A. & Gärtner, S. (2008) Nothofagus betuloides (Mirb.) Oerst 1871 (fagales: Nothofagaceae) forests in Southern Patagonia and Tierra del Fuego (Bosques de Nothofagus betuloides

(Mirb.) Oerst 1871 (Fagales: Nothofagaceae) en la Patagonia Austral y Tierra del Fuego). *Anales Del Instituto De La Patagonia*, 36, 53–68.
Quintanilla, V. (2008) Estado de recuperación del bosque nativo en una cuenca nordpatagónica de Chile, perturbada por grandes fuegos acaecidos 50 años atrás (44°-45°S). *Revista de Geografía Norte Grande*, 28, 73–92.
R Core Team. (2022) *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
Ridding, L.E., Bullock, J.M., Pescott, O.L., Hawes, P., Walls, R., Pereira, M.G. et al. (2020) Long-term change in calcareous grassland vegetation and drivers over three time periods between 1970 and 2016. *Plant Ecology*, 221, 377–394.
Roig, F.A. & Faggi, A.M. (1985) *Transecta botánica de la Patagonia austral. Análisis geobotánico de la vegetación*. Buenos Aires: Consejo Nacional Investigaciones Científicas y Técnicas, Instituto de La Patagonia y The Royal Society.
Romo, À., Ferrari, S., Romo, N., Albrieu, C. & Kofalt, R. (2012) Fitodiversidad en la estepa magallánica húmeda de la Patagonia Argentina. In: Cunill, R., Pèlachs, A., Pérez-Obiol, R. & Soriano, J.M. (Eds.) *Las zonas de montaña: gestión y biodiversidad. VII CONGRESO ESPAÑOL DE BIOGEOGRAFÍA*, *Pirineo 2012*. Sant Pere de Ribes: Universidad Autonoma de Barcelona.
Sabatini, F.M., Jiménez-Alfaro, B., Jandt, U., Chytrý, M., Field, R., Kessler, M. et al. (2022) Global patterns of vascular plant alpha diversity. *Nature Communications*, 13, 4683.
Sabatini, F.M., Lenoir, J., Hattab, T., Arnst, E.A., Chytrý, M., Dengler, J. et al. (2021) sPlotOpen – an environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography*, 30, 1740–1764.
San Martín, C., Pérez, Y., Alvarez, M., Salazar, C. & Ramírez, C. (2013) Diversidad vegetal de lagunas temporales en la estepa patagónica chilena noroccidental. *Anales del Instituto de la Patagonia*, 41, 111–116.
Seibert, P. (1987) Ökologische Bewertung und Bewertung des Landnutzungspotentials nach naturräumlichen Einheiten in der Transecta Botánica de la Patagonia austral (Ecological evaluation and evaluation of land-use potential within landscape units of the Transecta botanica de la Patagonia austral). *Erdkunde*, 41, 226–240.
Siefert, A., Laughlin, D.C. & Sabatini, F.M. (2023) *You shall know a species by the company it keeps: leveraging co-occurrence data to improve ecological prediction*. https://doi.org/10.1101/2023.02.15.528518
Sporbert, M., Welk, E., Seidler, G., Jandt, U., Aćić, S., Biurrun, I. et al. (2021) Different sets of traits explain abundance and distribution patterns of European plants at different spatial scales. *Journal of Vegetation Science*, 32, e13016.
Steinbauer, M.J., Grytnes, J.-A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H. et al. (2018) Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature*, 556, 231–234.
Sutherland, W.J., Pullin, A.S., Dolman, P.M. & Knight, T.M. (2004) The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19, 305–308.
Testolin, R., Attorre, F., Borchardt, P., Brand, R.F., Bruelheide, H., Chytrý, M. et al. (2021a) Global patterns and drivers of alpine plant species richness. *Global Ecology and Biogeography*, 30, 1218–1231.
Testolin, R., Carmona, C.P., Attorre, F., Borchardt, P., Bruelheide, H., Dolezal, J. et al. (2021b) Global functional variation in alpine vegetation. *Journal of Vegetation Science*, 32, e13000.
Zuloaga, F.O., Belgrano, M.J. & Zanotti, C.A. (2019) Actualización del Catálogo de las Plantas Vasculares del Cono Sur. *Darwiniana, Nueva Serie*, 7, 208–278.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Description of the variables contained in the "header" and "DT" matrices

Appendix S2. Cross-table linking the vegetation description in Transecta to the World Formation Types

Appendix S3. Modeled response of species richness to elevation for each world formation type

Appendix S4. Regression slopes of the linear model predicting species richness as a function of elevation, world formation type and their interaction

Appendix S5. Pairwise comparisons of regression coefficients of the interaction term between world formation type and elevation

How to cite this article: Sabatini, F.M., Hähn, G., Speziale, K., Cingolani, A.M., Damasceno, G. & Bruelheide, H. (2023) Vegetation of Southern Patagonia in the 1970s – Digitization of a gray-literature data set as a monitoring baseline in a changing world. *Journal of Vegetation Science*, 34, e13209. Available from: https://doi.org/10.1111/jvs.13209