ResNeTS: A ResNet for Time Series Analysis of Sentinel-2 Data Applied to Grassland Plant-Biodiversity Prediction

Álvaro G. Dieste, Francisco Argüello, Dora B. Heras, *Member, IEEE*, Paul Magdon, Anja Linstädter, Olena Dubovyk, and Javier Muro

***Abstract*—Analyzing time series from remote sensing data can aid in understanding spectral-temporal phenomena in ecosystems, such as the seasonal variation of plant components. Lately, deep learning has emerged as a strong method for mapping environmental variables from this data due to its exceptional predictive capabilities. This work studies the adaptation of the ResNet computer vision architecture for time series analysis of Sentinel-2 data. The resulting deep learning architecture, ResNeTS, stacks sequential convolutions to build a deep and narrow network, aligning with the design principles of leading convolutional architectures in computer vision. Experiments were carried out for predicting different plant-biodiversity indices, namely, species richness, and Shannon and Simpson indices, for temperate grassland ecosystems. The results show that ResNeTS can achieve moderate improvements in terms of accuracy compared to other state-of-the-art architectures, such as InceptionTime (up to +0.021 r^2), with reduced computational costs owing to its streamlined architecture.**

***Index Terms*—Biodiversity prediction, deep learning, multispectral imaging, remote sensing, residual network (ResNet), sentinel-2, time series analysis.**

Received 15 February 2024; revised 9 May 2024, 29 June 2024, and 5 August 2024; accepted 27 August 2024. Date of publication 3 September 2024; date of current version 2 October 2024. The work of Álvaro Goldar Dieste was supported in part by Xunta de Galicia - Consellería de Cultura, Educación, Formación Profesional e Universidades, under Grant ED481A-2022/257. This work was supported in part by the German Research Foundation (DFG) under the Priority Program 1374, through the Project Sensing Biodiversity Across Scales (SeBAS), under Grant DU 1596/1-1 and Grant LI 1842/4-1, in part by the Agencia Estatal de Investigación, Government of Spain, through MCIN/AEI/10.13039/501100011033 and "European Union NextGenerationEU/PRTR", under Grant PID2022-141623NB-I00 and Grant TED2021-130367B-I00, in part by the Xunta de Galicia - Consellería de Cultura, Educación, Formación Profesional e Universidades, through Galician Research Center accreditation under Grant ED431G-2019/04, and through Reference Competitive Group accreditation under Grant ED431C-2022/16, and in part by the European Regional Development Fund (ERDF/EU). *(Corresponding author: Álvaro G. Dieste.)*

Álvaro G. Dieste and Dora B. Heras are with the Singular Research Center on Intelligent Technologies (CiTIUS), University of Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: alvaro.goldar.dieste@usc.es; dora.blanco@usc.es).

Francisco Argüello is with the Department of Electronics and Computing, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: francisco.arguello@usc.es).

Paul Magdon is with the University of Applied Sciences and Arts (HAWK), 37077 Göttingen, Germany (e-mail: paul.magdon@hawk.de).

Anja Linstädter is with the Biodiversity Research/Systematic Botany, University of Potsdam, 14469 Potsdam, Germany (e-mail: linstaedter@uni-potsdam.de).

Olena Dubovyk is with the Geography Institute, Department of Earth System Sciences, University of Hamburg, 20146 Hamburg, Germany (e-mail: olena.dubovyk@uni-hamburg.de).

Javier Muro is with the Institute of Farm Economics, Thünen Institute, 38116 Braunschweig, Germany (e-mail: javier.muro@thuenen.de).

The code developed for this work will be available at https://github.com/alvrogd/ResNeTS. All data necessary to replicate the study is available at the repository of the Biodiversity Exploratories BExIS (https://www.bexis.uni-jena.de/); accession numbers 30969 (Sentinel-2 time series) and 27386 (Species inventories).

Digital Object Identifier 10.1109/JSTARS.2024.3454271

Nomenclature

CNN	Convolutional neural network.
FCN	Fully convolutional network.
Grad-CAM	Gradient-weighted class activation mapping.
GRU	Gated recurrent unit.
LAI	Leaf area index.
LSTM	Long short-term memory.
MLP	Multilayer perceptron.
MODIS	Moderate resolution imaging spectroradiometer.
NDVI	Normalized difference vegetation index.
r^2	Coefficient of determination.
ResNet	Residual network.
ResNeTS	Residual network for time series.
RMSE	Root mean squared error.
rRMSE	Relative root mean squared error.
sRMSE	Systematic root mean squared error.
uRMSE	Unsystematic root mean squared error.
RNN	Recurrent neural network.
SAR	Synthetic aperture radar.
SGD	Stochastic gradient descent.
SHAP	Shapley additive explanations.

I. Introduction

THE current biodiversity crisis has tremendous impacts on the proper functioning of ecosystems and its derived benefits to humans [1]. For instance, biodiversity is vital for food production as it directly influences the maintenance of fundamental ecosystem functions, such as soil fertilization, pest and disease regulation, erosion control, biomass production, and pollination of crops and trees [2], [3]. In this context, biodiversity conservation, i.e., protecting and preserving the richness and variety of ecosystems, habitats, and species, is of imperative importance [4], [5]. This is particularly true for European grasslands, which are an integral part of Europe's agricultural landscapes [6]. They have a key role in ecosystem health and

functionality, as they act as habitats for many plant and animal species. Human intervention, including the transformation of natural grasslands, intensive cultivation, pollution, and climate change, can affect the diversity and stability of grassland ecosystems in many ways and at different scales. In recent decades, the decline in grassland diversity across Europe has threatened biodiversity and is a major environmental concern [7].

To effectively address these challenges, the application of remote sensing data from satellite images has become indispensable [8], [9], [10]. In particular, time series analysis of sequences of satellite images allows for the extraction of temporal characteristics critical for understanding ecosystem dynamics. By capturing seasonal variations with this data, researchers can enrich the understanding of ecosystems, helping the development of precise solutions for environmental challenges such as land cover classification [11] and phenological metrics extraction [12].

Building on this technological foundation, researchers have initiated multiple efforts to monitor grassland biodiversity, using temporal data extracted from satellite imagery at different scales. The biodiversity mapping includes alpha diversity, which quantifies the biodiversity within a community or ecosystem [13], [14]; beta diversity, which measures the diversity between two communities [15]; and gamma diversity, a combination of the other two metrics that assesses the biodiversity at landscape scale between different ecosystems [16]. Another way to map biodiversity is using radiative transfer models to map functional diversity: the different strategies for growth, reproduction, and survival of organisms that determine ecosystem services, resilience, and stability [17]. Much of this research is confined to hyperspectral sensors (e.g., plane or UAV mounted), with a much more limited spatial and temporal coverage than multispectral sensors (e.g., Landsat, Sentinel-2). This considerably reduces the practical applicability of these approaches for systematic monitoring [18]. In addition, the problem of spatial autocorrelation in empirical models is too often not addressed, and thus, model results cannot be trusted outside the training area [19].

Deep learning approaches have recently shown high predictive capabilities and a unique ability to reveal higher hierarchical patterns from spectral-temporal data such as time series [20]. The application of these powerful models can offer substantial advantages for biodiversity monitoring using remote sensing data, facilitating the move beyond traditional land cover type classifications and spurious correlations between biodiversity and spectral indices [21]. This approach also circumvents the reliance on controversial assumptions, such as the spectral diversity hypothesis [22]. While the applications of deep learning methods for biodiversity monitoring are relatively scarce, their potential for time series analysis of remote sensing data can be appreciated in a broader context of applications:

1) *Multilayer Perceptron (MLP) [23]:* This architecture marked a first milestone by outperforming traditional machine learning approaches in analyzing time series. Numerous studies have since explored and continued using it for a wide range of applications. For instance, J. Muro et al. [24] used a MLP to estimate biomass and species richness of grasslands in three regions of Germany using Sentinel-2 surface reflectance, achieving better estimations than a Random Forest regressor [25]. Other interesting applications of MLPs are described in [26] and [27].
   Despite their capabilities, a notable drawback of MLPs is their limited ability to model temporal and dimensional patterns in multivariate time series effectively. This limitation arises from their simultaneous analysis of all data points, without considering the sequential nature of the data. To overcome these limitations, more advanced deep learning architectures, such as the RNN [28] and the CNN [29] were eventually introduced.
2) *RNNs:* These include the LSTM [30] and GRU [31]. They process time series data step by step, updating an internal state that models relations with previous steps. For instance, H. Ma and S. Liang [32] used both a LSTM and a GRU to generate a global LAI from MODIS data with better accuracy compared to a MLP approach.
   More advanced applications use their bidirectional variants, Bi-LSTM and Bi-GRU, which process time series in both forward and backward directions to understand temporal patterns better. For example, H. C. d. C. Filho et al. [33] leveraged a Bi-LSTM architecture to detect rice crops in southern Brazil from Sentinel-1 SAR time series, Rußwurm and Körner [34] used another Bi-LSTM for crop classification from Sentinel-2 data, and Garioud et al. [35] employed a Bi-GRU architecture to estimate NDVI in two areas in France using SAR and optical data from Sentinel-2. More innovative approaches incorporate additional components, such as attention mechanisms [36], into these bidirectional architectures to further enhance their capabilities [37], [38].
3) *CNNs:* While initially developed for visual information processing, they also exhibit potential for time series analysis. Originally, CNNs slide 2-D convolutional kernels across an image, extracting dimensional and spatial patterns between pixels and their neighbors to capture visual features effectively. In the early attempts of using CNNs for time series analysis, a common approach involved converting the time series data into images through feature engineering [39], [40]. This transformation aimed to map the time series temporal and dimensional aspects onto the image representation's height and width, but it proved to be suboptimal compared to RNNs [39], [41].
   Eventually, CNNs started to be specifically designed for time series analysis using 1-D convolutions, thus outperforming RNNs in numerous remote sensing studies. For instance, Zhao et al. [42] used a 1-D CNN for early crop classification in the Zhanjiang region of China using Sentinel-1 SAR time series, outperforming both LSTM and GRU architectures. Another study by Zhong et al. [43] showcased the advantages of a 1-D CNN over a LSTM for crop classification in California, USA, using Landsat surface reflectance.
   Following the footsteps of computer vision, time series CNNs evolved from shallow architectures with few layers to deeper, more powerful architectures, commonly known as residual CNNs [44]. These networks can be recognized by the presence of skip-connections between layers, which allow for constructing architectures with tens or hundreds

of layers while circumventing training issues such as gradient vanishing. In particular, Wang et al. [45] introduced the first residual CNN with widespread adoption in time series analysis, including remote sensing applications. For instance, Paolini et al. [46] demonstrated the capabilities of this architecture for mapping different irrigation systems across Catalonia, Spain, using data from multiple satellites.

4) *Transformers [36]:* This relatively new architecture has impacted a wide range of disciplines due to its excellent capabilities when compared to other deep learning architectures [47]. Transformers analyze time series by leveraging their novel self-attention mechanism, which simultaneously weighs the importance of all time steps while capturing complex dependencies across the entire sequence, without the sequential processing limitations of RNNs. Although the application of Transformers in analyzing time series from remote sensing data is still emerging, they have already been successfully employed in various tasks such as image land cover classification [48], [49].

Despite the recent attention on Transformers, CNNs have continued to demonstrate competitive performance in various areas, particularly in computer vision, as exemplified by architectures such as ConvNeXt [50]. Moreover, CNNs benefit from the inherent inductive bias of convolutions, making them less data-intensive during training compared to Transformers [51]. This trait can be particularly helpful for remote sensing applications, where datasets often contain limited data [52].

Among the current CNN architectures for time series analysis, InceptionTime [53] stands out for its high accuracy [54] and widespread use across various domains, including remote sensing [55]. Inspired by the Inception-v4 network from image processing [56], InceptionTime consists of modules that perform parallel convolutions of different sizes, coupled with residual connections, to handle time series of diverse lengths.

However, most innovations in CNN development are driven by image processing, with other fields such as time series analysis often experiencing a delayed adoption of these new ideas. Since the introduction of Inception-v4 in 2017, image processing CNNs have increasingly favored deep, narrow networks based on the ResNet[1] family [44]. These models prioritize a higher number of sequential modules interconnected with residual connections [50], [57], [58]. Developing a ResNet-based architecture for time series analysis could accelerate progress in the area by facilitating the integration of cutting-edge innovations from the most advanced image processing CNNs.

All these observations have inspired this work to develop ResNeTS, a deep learning architecture based on the ResNet family, now adapted for time series analysis of remote sensing data. ResNeTS is evaluated on its ability to predict three biodiversity indices over grasslands with different species compositions, using Sentinel-2 time series. In particular, ResNeTS is benchmarked against InceptionTime and other state-of-the-art architectures for time series analysis, including Transformers. Unlike InceptionTime's branching topology with parallel convolutions, ResNeTS features a simpler architecture of sequential convolutions. This design approach, when considering the characteristics of the time series to analyze, improves the accuracy of InceptionTime while reducing computational costs. These results illustrate the potential of continuing CNN development for precise and efficient time series analysis in remote sensing tasks.

The rest of this article is organized as follows. The study area and the remote sensing dataset are presented in Section II. Then, the ResNeTS architecture for time series analysis of remote sensing data is detailed in Section III. Thereafter, the experiments for evaluating the efficacy of ResNeTS for the prediction of plant-biodiversity indices of interest are presented in Section IV; an explainability analysis is conducted as well to assess the importance of each spectral and temporal features used. Next, the implications of the results for biodiversity research and management, and for further time series analysis in a broader scenario of remote sensing applications, are discussed in Section V. Finally, Section VI concludes this article.

II. Materials

A. Study Area

The Biodiversity Exploratories encompass three distinct regions distributed across Germany, Central Europe (see Fig. 1). The three regions are the UNESCO Biosphere Reserve "Schorfheide-Chorin" (SCH), the National Park "Hainich" and its surroundings (HAI), and the Biosphere Reserve "Schwäbische Alb" (ALB). Across and within these sites, species richness and compositions are highly variable, as they are influenced by abiotic environmental conditions, such as topographic position, soil characteristics, and land use intensities [59], [60].

Each exploratory region comprises 50 experimental grassland plots, representing a spectrum of management practices ranging from intensive to extensive land use. Some plots are subjected to varying degrees of cattle grazing, while others are mown at specific seasonal intervals. In addition, different levels of fertilization are implemented. To ensure standardized quantification of management levels, a land use intensity (LUI) index was developed [59]. This index is calculated annually based on factors such as mowing frequency, grazing intensity, and fertilization inputs.

Fig. 2 shows examples of different degrees of species diversity in seminatural grasslands across Germany. The grassland plots in the Biodiversity Exploratories are usually dominated by perennial grass species, such as meadow soft grass (*Holcus lanatus*), tall oat grass (*Arrhenatherum elatius*), or smooth meadow grass (*Poa pratensis*). Among the legume species, white clover (*Trifolium repens*), red clover (*Trifolium pratense*), and bird's-foot trefoil (*Lotus corniculatus*) are frequently found. As forbs, commonly occurring species are dandelion (*Taraxacum officinale*), wild carrot (*Daucus carota*), or plantain (*Plantago lanceolata*). Some of these and other species found at the Biodiversity Exploratories are shown in Fig. 3.

Using biodiversity indices can help assess the health and stability of grassland ecosystems as well as their provided functions and services. This is also vital for grassland management and conservation. Among these indices, the following

[1]The term "ResNet" will be used to refer to the family of residual networks introduced in [44], whereas "residual CNN" will be used to refer to derived networks (i.e., CNNs that integrate skip-connections in their structure).

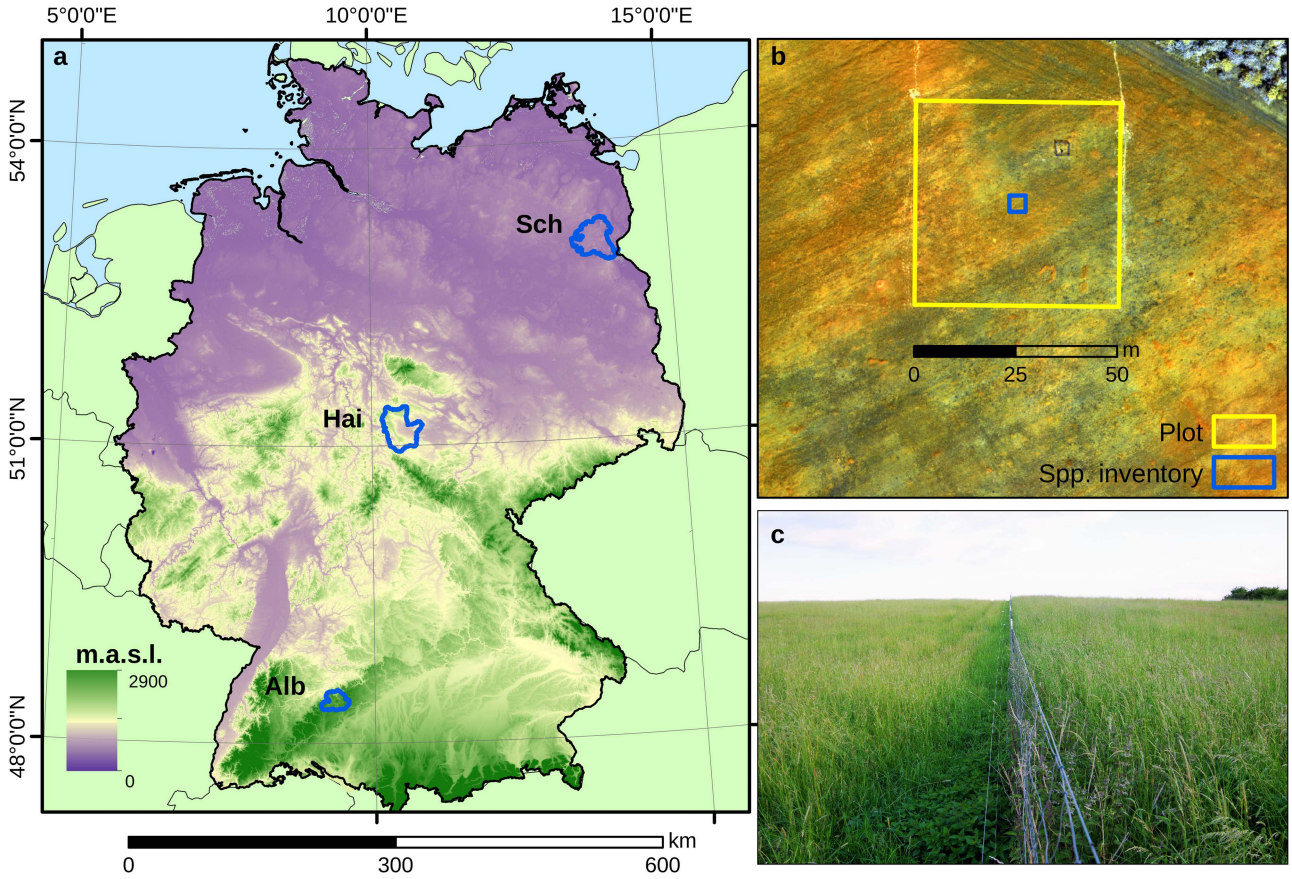

Fig. 1. (a) Location of the three study sites (Sch: Schorfheire-Chorin Biosphere Reserve, Hai: Hainich National Park, Alb: Schwäbische Alb) in Germany, Central Europe. Colors depict altitude (m.a.s.l), based on the Shuttle Radar Topography Mission digital terrain model. (b) Illustration of a monitoring plot (50 m × 50 m) and the reduced sampling area (4 m × 4 m) where plant species inventories were collected, based on a false color UAV image. (c) Example of two adjacent grasslands with different land-use intensities, resulting in different compositions and diversity of plant communities.

Fig. 2. Examples of different degrees of grassland biodiversity in the temperate regions of Germany. Biodiversity increases from left to right.

are particularly relevant and will be the object of study in this work:

1) *Species richness:* The species number per unit of area (S).
2) *Shannon index:*

$$H = -\sum_{i=1}^{S} p_i \cdot \ln(p_i) \tag{1}$$

where H = Shannon's diversity index, S = total number of species, and p_i = Relative abundance of species i.

3) *Simpson index:*

$$D = 1 - \frac{\sum_{i=1}^{S}(n_i \cdot (n_i - 1))}{N \cdot (N - 1)} \tag{2}$$

where D = Simpson's diversity index, n_i = number of individuals of the species i, N = total number of individuals in the community, and S = total number of species.

Simpson's index reflects both species richness (S) and evenness (the distribution of species relative abundances within a community). Here, 1 indicates infinite diversity (all species

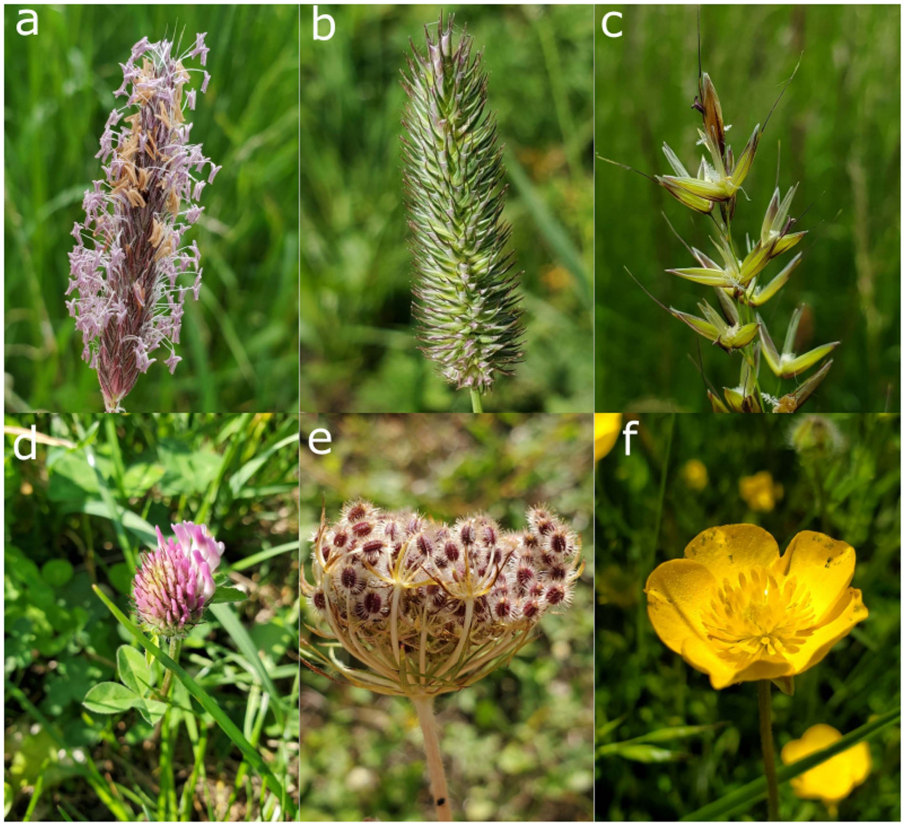

Fig. 3. Examples of plant species present in the grasslands of the Biodiversity Exploratories, with three grass species shown in the top row. (a) *Alopecurus pratensis*. (b) *Phleum pratensis*. (c) *Arrhenatherum elatius*. (d) *Trifolium pratense* (a legume species). (e) *Daucus carota* (a forb species). (f) *Ranunculus bulbosus* (another forb species).

are equally present), and 0 indicates no diversity (only one species present) [61]. Shannon's index also considers both species richness and evenness, but is more sensitive to rare or less abundant species [61]. It ranges between 0 (low richness and evenness) and 5, but most values range between 1.5 and 3.5.

B. Remote Sensing and Species Datasets

This study uses the floristic and remote sensing datasets compiled in [24]. Plant species inventories were obtained during the peak standing crop period (second half of May), from 2017 to 2020 (4 years), following a standardized sampling protocol across the 150 plots [62] (total of 600 measures). Such inventories were recorded in a permanently marked 4 m $\times$ 4 m area within each of the plots (see Fig. 1(b)), which is considered as representative of the 50 m $\times$ 50 m plot [60]. Vascular plant species' abundances were recorded annually via visual estimations of their ground cover. From this data, diversity indices (S, H, and D) were calculated for each plot. Plots with trees within or near their boundaries were excluded from the analysis, resulting in 502 valid samples. Spatial dependencies within site and across sites for this dataset were analyzed and proved nonsignificant in [24].

The Sentinel-2 images utilized in this study were processed using the FORCE software [63]. Images from 2017–2020 with a cloud cover lower than 50% were downloaded, atmospherically and topographically corrected, and resampled to 10 m. Subsequently, a synthetic time series was generated by creating an image every 14 days through radial basis function interpolation implemented in the time series analysis module of the FORCE processor. Images acquired during winter (November–March) were excluded from further analysis due to their limited significance for vegetation studies and more frequent cloud cover. The result is a synthetic gapless dataset of 10 bands and 16 time steps (see Fig. 4) representative of the spectral variation of the vegetation across the phenological cycles of several years. The median pixel value for each band was calculated for each plot and paired with each field observation, as in [24].

The resulting dataset is composed of 502 observations of each biodiversity metric (either species richness S, Shannon's diversity index H, or Simpson's diversity index D), and their corresponding temporal profiles (March–October with 16 time steps) of reflectance values for 10 bands of Sentinel-2 [64]: blue, green, red, red edge 1, red edge 2, red edge 3, near-infrared, near-infrared b, short-wave infrared 1, and short-wave infrared 2. The coastal blue, water vapor, and cirrus bands were excluded.

Deep learning studies tend to train architectures with several thousands of observations acquired from different sources [65]. Hence, the reduced size of the dataset used in this study may seem strange, but it has various reasons. First, biodiversity

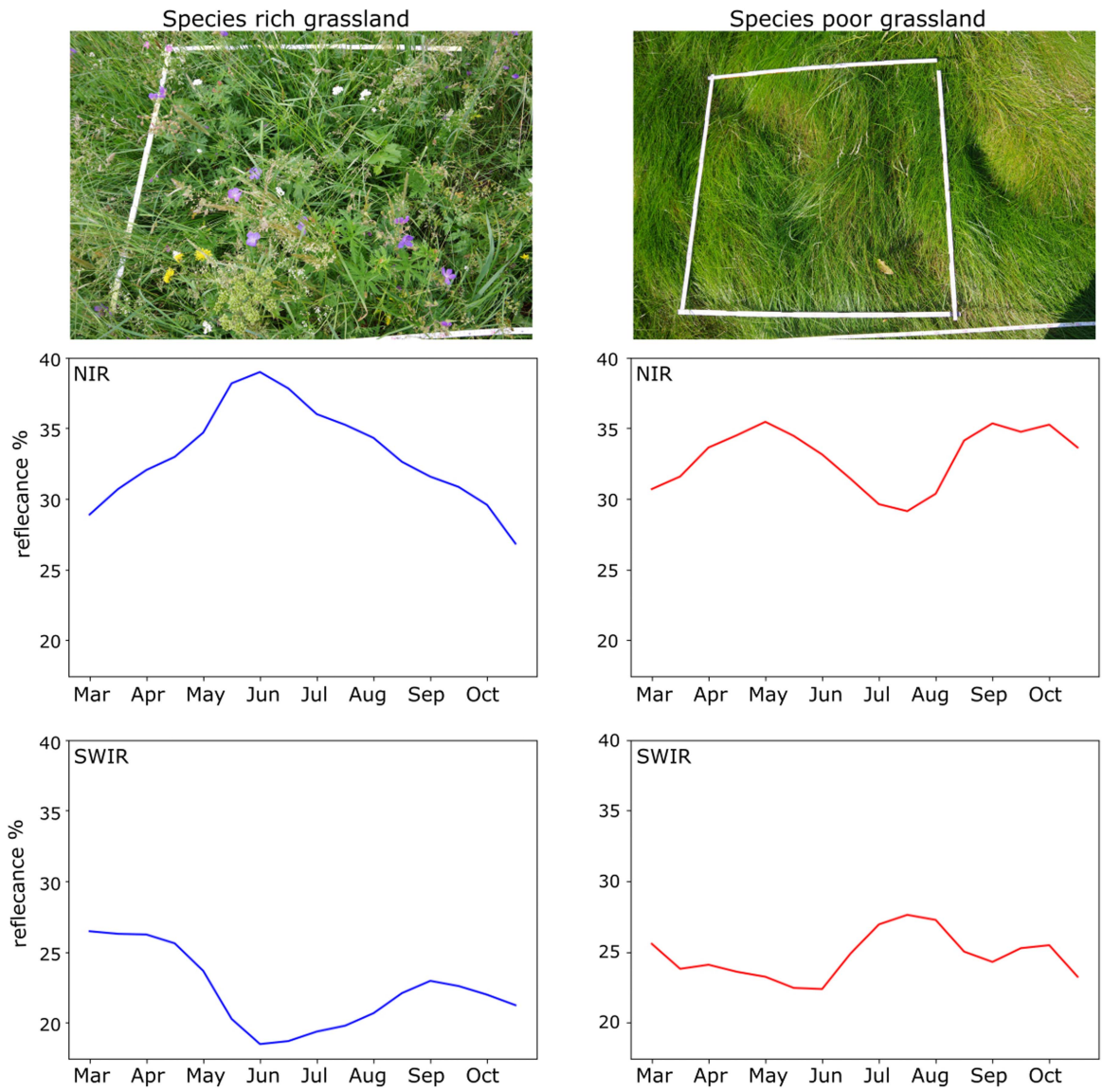

Fig. 4. Species-rich (left column) and species-poor (right column) plots along with their temporal profile in the near-infrared and short-wave infrared bands of Sentinel-2 for 2020.

observations for a single dataset must be manually collected in situ, using standardized sampling protocols such as those undertaken at the Biodiversity Exploratories, resulting in laborious annual field campaigns [66]. This inevitably limits the size of a single dataset. Second, it is often not feasible to compile local biodiversity (alpha-diversity) data from different sources for a joint analysis. For instance, species richness is directly influenced by the sampling area: the larger the sampling area, the higher the number of recorded species [67]. Inventories acquired on different spatial scales are thus difficult to compare. Unfortunately, a standardization of species richness on sampling area is also not straightforward and usually requires elaborate methods to estimate the shape of the species-area relationship [67], [68], [69]. Finally, even though the Biodiversity Exploratories started in 2008, the Sentinel-2 B was not launched until 2017, further limiting the number of observations in the dataset.

III. ResNeTS

A. ResNeTS Overview

ResNeTS is a ResNet-based network for time series analysis of Sentinel-2 data. It favors a simple design, stacking sequential convolutions to construct a narrow network. By considering in its design the characteristics of time series to analyze (i.e., length and dimensionality), ResNeTS achieves state-of-the-art accuracy with reduced computational costs.

As any ResNet-based architecture, ResNeTS' structure is divided into three sections, as shown in the left side of Fig. 5: stem, residual blocks, and head. The stem prepares the input time series for the residual blocks, which then extract meaningful features through the repeated use of convolutions. Finally, the head generates the predictions using these features.

Each convolution analyzes the time series sliding its filters step by step, aiming to extract temporal and dimensional patterns

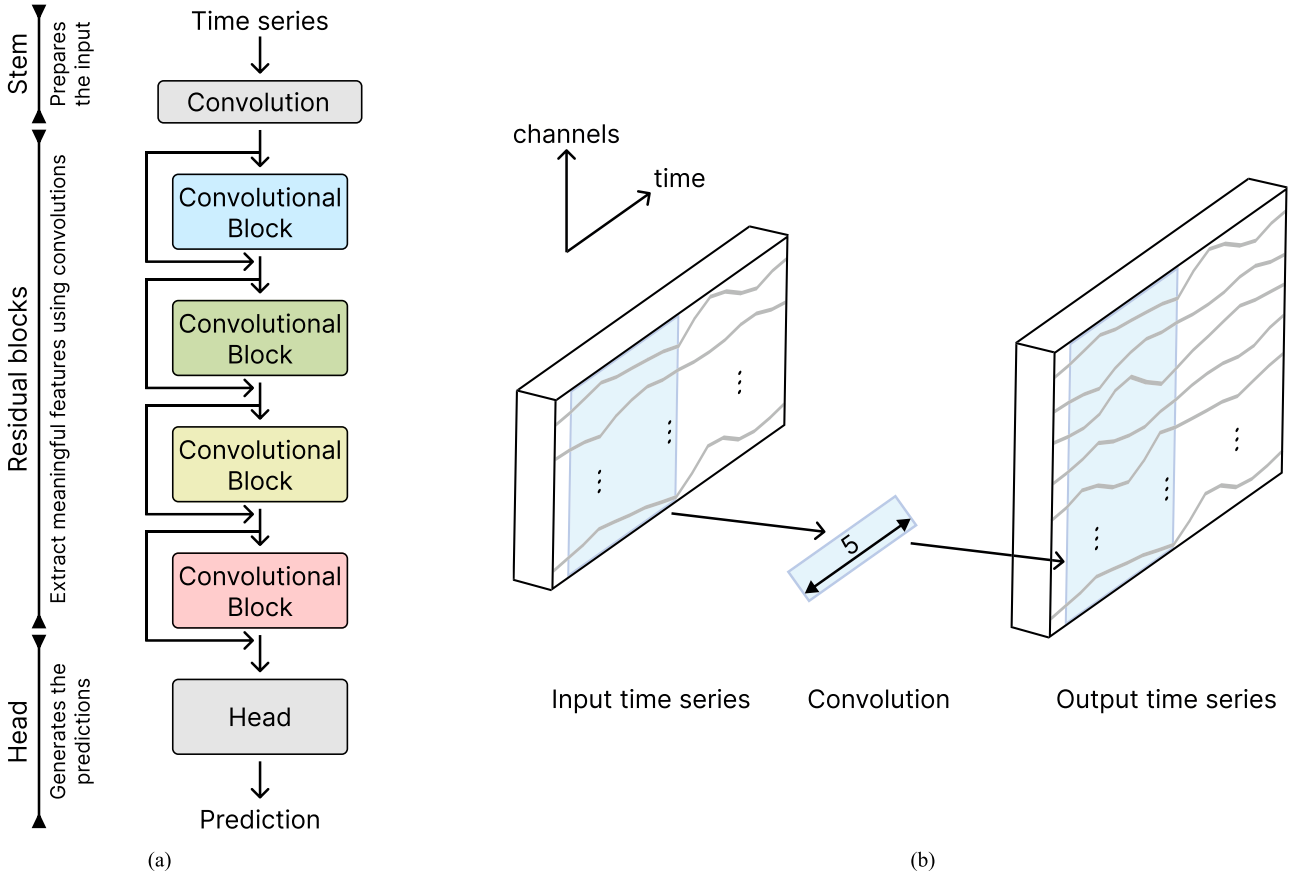

Fig. 5. Overview of ResNeTS. (a) Three sections of the ResNeTS architecture. (b) Application of a 1-D convolution to analyze a time series. The convolution uses a filter of five elements, producing an output series with more bands (taller) than the input.

through the simultaneous analysis of the current step and its neighboring steps across all dimensions. This temporal window incorporating information from all dimensions is depicted on the right side of Fig. 5.

Compared to shallow CNNs, the integration of shortcuts in residual networks allows the construction of significantly deeper architectures without encountering common training issues such as gradient vanishing. Thus, by layering a high number of convolutional layers and progressively increasing the number of filters, ResNeTS extracts increasingly complex patterns from the input data for enhanced prediction accuracy.

As previously discussed, InceptionTime, the most notable CNN for time series analysis, consists of a series of blocks that apply parallel convolutions to the data. This concept is illustrated on the left part of Fig. 6, where the four parallel convolutions are applied and combined as input for the subsequent block, along with the skip connection (left convolution). In contrast, the building blocks of ResNeTS consist of just two sequential convolutions and the corresponding skip-connection (left path).

Comparing both architectures, fewer and simpler blocks are required in ResNeTS to build an architecture that achieves an accuracy improved over InceptionTime. This results in a considerably more lightweight topology, as depicted in the right part of Fig. 6. This is because each InceptionTime block uses four parallel convolutional filters to adapt to short, medium, and long time series, whereas ResNeTS' convolutions are tuned for the target time series.

B. Initial ResNeTS Architecture

The ResNet family from computer vision [44] consists of architectures with varying sizes, denoted by the total number of layers: ResNet-18, 34, 50, 101, 152. Architecture selection for a specific problem depends on both the available computational resources and the dataset size, as larger networks can assimilate more knowledge. Given the small size of the dataset used in this study, particularly when compared to large-scale computer vision datasets, such as ImageNet [70], the ResNet-18 architecture with 2-D convolutions replaced by 1-D convolutions set the initial baseline for ResNeTS, as it provided results comparable in accuracy to larger variants.

After that, simplifications were introduced to further reduce the size of the architecture without incurring any accuracy loss, as ResNet-18 was still too large for the dataset of interest. In particular:

1) The number of repetitions per residual block was reduced from the default of two to one.
2) The number of convolutional channels was reduced to 64 in all blocks.

The optimal channel configuration was obtained from an extensive grid search on the training points of the dataset. Let B_x

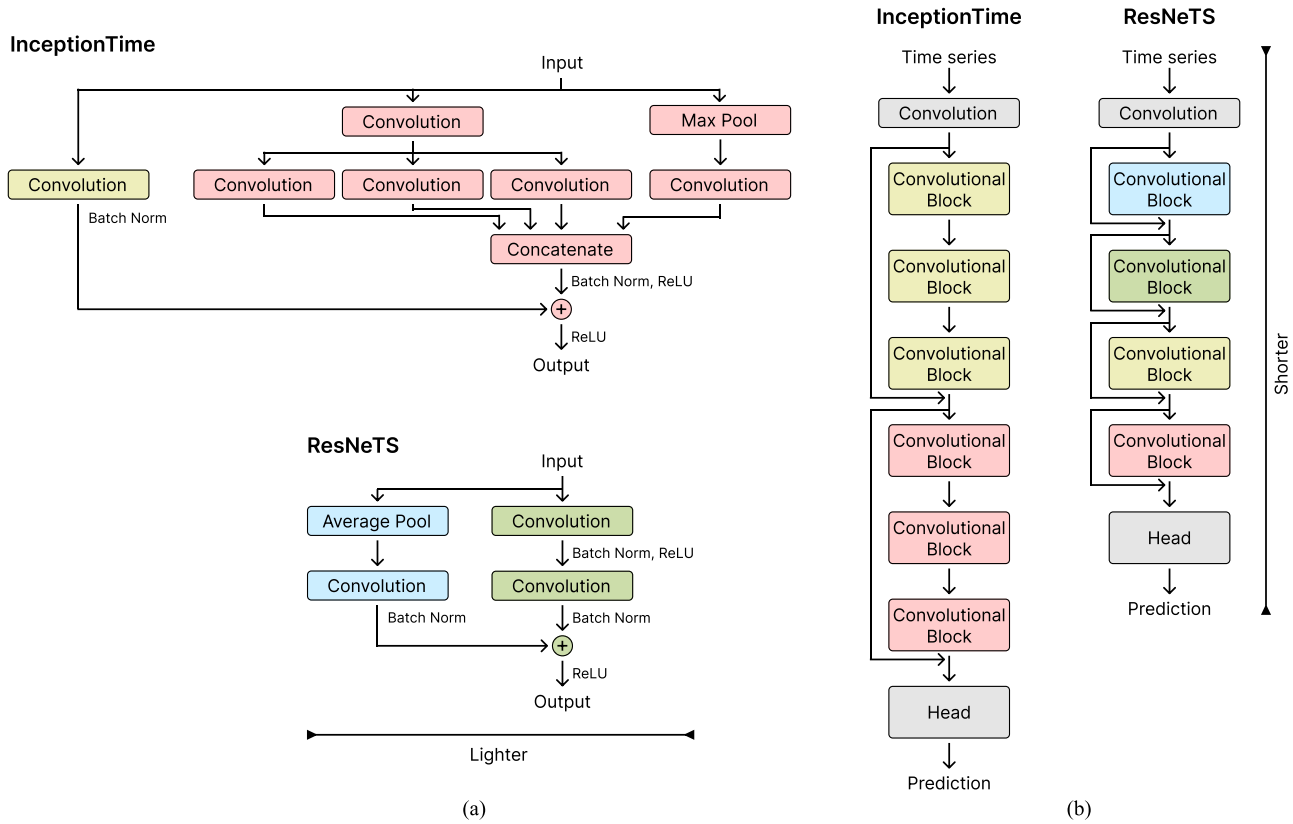

Fig. 6. Comparison between InceptionTime and ResNeTS architectures. (a) Depiction of the convolutional blocks of both architectures. (b) Depiction of the depth of both architectures.

represent the channel count of block x. In ResNet-18, the channel count is: $B_1 = 64$, $B_2 = 128$, $B_3 = 256$, and $B_4 = 512$. For ResNeTS, each block was allowed to go from 32 convolutional channels to its maximum (e.g., 128 for the second block), with a step of 32, and ensuring that: $B_1 \leq B_2 \leq B_3 \leq B4$.

Lastly, to make the ResNeTS baseline suitable for the target time series, most of the strides inherited from ResNet-18 for the progressive dimensionality reduction had to be removed. This was necessary as the series consists of only 16 time steps, whereas ResNet was initially designed for 112×112 RGB images. The four original strides in ResNet result in a total dimensionality reduction of $\times 1/16$, transforming the 112×112 image sample into a 7×7 one that enters the head layers. By retaining only one stride in ResNeTS (see Table I), the input for the head has a final size of 8 time steps, closely mirroring the original dimensions in ResNet.

Fig. 7 shows the network topology of the resulting baseline ResNeTS architecture, containing the four residual blocks detailed in Table I.

C. Larger Convolutional Kernels

The use of 1-D convolutions in CNNs for time series instead of 2-D convolutions leads to a significant reduction in computational costs. Thus, much larger convolutional filters than the standard 3×3 filters in ResNet can be used, enabling a more comprehensive understanding of temporal and dimensional patterns in time series while keeping the architecture computationally feasible. In particular, all of the convolutional layers in ResNeTS used filters of size 5, as Table I details, with no further accuracy improvements observed beyond this point for the dataset used in this study.

D. Architectural Improvements

Three minor modifications were introduced to ResNeTS to improve its accuracy further in predicting the biodiversity indices of interest. The first change was an upgrade in the skip-connections, while the other modifications were two layerwise modifications.

ResNeTS inherited its skip-connections from the ResNet family. These present a design limitation that was addressed in ResNeTS as discussed in [71]. Specifically, when traversing feature maps of different spatial sizes (i.e., downsampling the processed time series), the original skip-connections also applied a stride to its 1-size convolutions intended for channel mapping, thus disregarding part of the input feature map. This approach is shown on the left side of Fig. 8. By incorporating an extra average pooling layer within these connections when needed, as depicted on the right side of Fig. 8, all feature points are utilized during downsampling. This enables the subsequent convolutional layer to concentrate solely on mapping the channel count to produce the skip-connection. Note how the final ResNeTS architecture,

TABLE I
DETAILS OF THE LAYERS IN RESNETS ARCHITECTURE

Component	Layer	Output size	Filter size	Stride	Normalization	Activation
Stem	1-D Conv	16 × 96	1	1	Batch Norm	ReLU
Residual Block	1-D Conv	16 × 64	5	1	Batch Norm	ReLU
	1-D Conv	16 × 64	5	1	Batch Norm	–
	1-D Skip Conv	16 × 64	1	1	Batch Norm	–
	Addition	16 × 64	–	–	–	ReLU
Residual Block	1-D Conv	16 × 64	5	1	Batch Norm	ReLU
	1-D Conv	16 × 64	5	1	Batch Norm	–
	1-D Skip Conv	16 × 64	1	1	Batch Norm	–
	Addition	16 × 64	–	–	–	ReLU
Residual Block	1-D Conv	8 × 64	5	2	Batch Norm	ReLU
	1-D Conv	8 × 64	5	1	Batch Norm	–
	1-D Skip Average Pool	8 × 64	2	2	–	–
	1-D Skip Conv	8 × 64	1	1	Batch Norm	–
	Addition	8 × 64	–	–	–	ReLU
Residual Block	1-D Conv	8 × 64	5	1	Batch Norm	ReLU
	1-D Conv	8 × 64	5	1	Batch Norm	–
	1-D Skip Conv	8 × 64	1	1	Batch Norm	–
	Addition	8 × 64	–	–	–	ReLU
Head	1-D Average Pool	1 × 64	8	8	–	–
	Flatten	64	–	–	–	–
	Fully Connected	1	–	–	–	Softmax

The output size corresponds to an input time series with dimensions *length* × *bands* = 16 × 10, matching the dataset used in this study.

depicted in Fig. 7, included this pooling operation in the third block, which performed the spatial downsampling from 16 time steps to 8.

Second, since not all residual blocks in ResNeTS have a stride for dimensionality reduction, the position of the only stride was shifted among the different blocks to determine its optimal setting. In particular, the best accuracies were obtained using the stride in the third residual block, as specified in Table I.

Lastly, increasing the original channel count of the stem from 64 to 96 led to enhanced accuracies, with no significant improvements beyond this size, nor using less than 64 channels.

TABLE II
COMPARISON BETWEEN THE ORIGINAL TRAINING PROCEDURE (INHERITED FROM RESNET) AND THE IMPROVED ONE IN RESNETS

Training configuration	Original	Improved
Checkpointing	Yes	Yes
Early stopping	No	Yes
Learning rate scheduler	Reduce on plateau	Cosine
Max. training epochs	120	1500
Optimizer	SGD	AdamW
Warm-up epochs	None	150
Warm-up schedule	None	Linear

E. Modern Training Procedure

An effective training procedure is as necessary as a robust architecture for achieving optimal performance. As highlighted in [50], significant advancements have been made in training methodologies for state-of-the-art neural architectures since the introduction of the ResNet family in 2016. Therefore, updating the ResNeTS training procedure to extract its full potential was crucial. Table II compares the original training procedure from ResNet against the improved procedure used for ResNeTS, which is detailed below.

The maximum number of training epochs was increased to 1500 to match InceptionTime, and an additional 10% of linear warm-up epochs were introduced at the beginning. Checkpointing was used to retain the optimal architecture upon training completion, with validation error measured every 10 epochs to guide the process. To preserve computational resources, early stopping was also triggered if no improvements in validation error were observed for a period exceeding 10% of the maximum epochs.

Inspired by the training recipe presented in [50], the SGD optimizer was replaced with AdamW [72], and the plateau-based learning rate scheduler was substituted with a cosine learning rate scheduler [73]. To optimize the training procedure of ResNeTS for the target dataset, a comprehensive hyperparameter optimization of AdamW via grid search was conducted. This resulted in a systematic exploration of all combinations of adjustable hyperparameters across extensive ranges[2] to ensure a thorough examination of the optimization space:

1) learning rate $= \{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$.
2) $\beta_1/\beta_2 = \{0.5, 0.6, 0.7, 0.8, 0.9, 0.99, 0.999\}$.
3) ϵ/weight decay $= \{1 \times 10^{-8}, 1 \times 10^{-7}, 1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}\}$.

Each combination of hyperparameters was evaluated through ten training runs of the model using a fixed set of seeds to control

[2]Note that the values for β_1, β_2, ϵ, and weight decay were configured independently of each other.

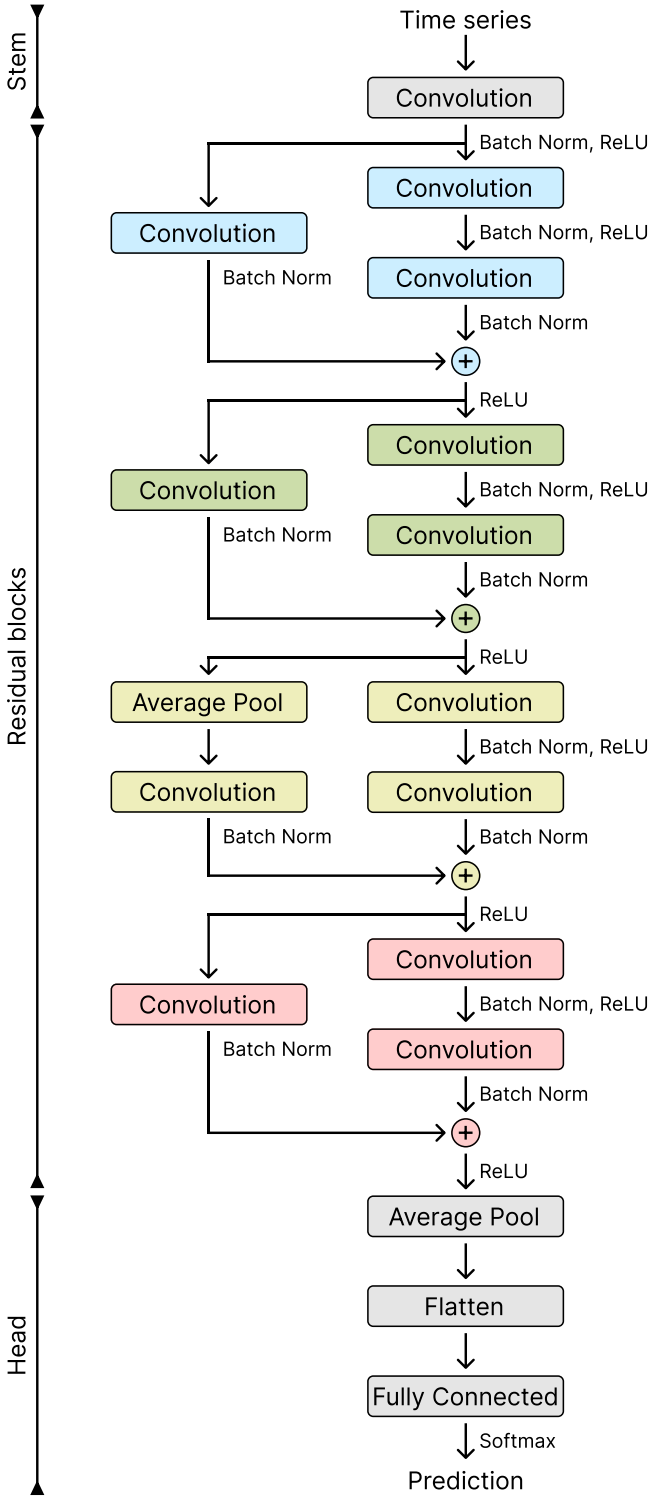

Fig. 7. Diagram of the ResNeTS architecture. A different color is assigned to each residual block.

for randomness. This enabled the identification of the configuration that produced the best average prediction metrics across all trials, resulting in the values learning rate $= 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-6}$, weight decay $= 1 \times 10^{-3}$.

F. Ensemble of Individual Models

Finally, both InceptionTime and ResNeTS were run in this study as single architectures and as ensembles of five identical architectures, InceptionTime-5 and ResNeTS-5, respectively, for a comprehensive comparison. For ensembles, each individual architecture was initialized with different random values. Once the ensemble was trained, its prediction consisted of the mean of the individual outputs, effectively reducing the variability in the quality of predictions caused by the stochastic nature of network training.

IV. Experiments

A. Experimental Setup

1) Competing Deep Learning Architectures: The capabilities of ResNeTS were compared to already established network architectures for time series analysis of remote sensing data. The first model is the MLP presented in [24] for studying the dataset used in this work. Next, the Bi-LSTM described in [34] was included to represent the capabilities of advanced RNNs. Then, the FCN described in [45] was selected to represent the early shallow CNNs used for time series analysis. The residual CNN from [45] was also included to provide a baseline of the first residual networks used for this field. Naturally, InceptionTime [53] was included to represent the state-of-the-art in residual CNNs for time series analysis. With the recent growing interest in Transformers, the architecture from [34] was included as well in the experiments. Lastly, the Rocket model [74] was employed to demonstrate the capabilities of non-deep learning models, given its high accuracy in extensive comparisons such as [54].

The MLP, Bi-LSTM, and Transformer architectures were implemented as described in the corresponding literature, since they were specifically designed for Sentinel-2 time series. Similarly, the FCN and residual CNN were used without modification due to their demonstrated adaptability across time series of various characteristics. For Rocket, although the initial implementation adhered to the default parameters for its reported adaptability, a suboptimal performance prompted an increase in the kernel count to 15 000, based on its authors' advice that a higher kernel count can improve accuracy. Finally, adjustments were made to InceptionTime to adapt its configuration to datasets with small time series, following the guidelines of its authors [53]:

1) *Bottleneck size:* The original reduction factor of $\times 0.25$ was kept to maintain the channel count throughout the network, thereby reducing the number of adjustable parameters and substantially improving runtime performance without affecting accuracy.
2) *Depth:* The number of Inception modules was also kept at the default of 6, as InceptionTime authors did not observe accuracy gains when increasing the depth for short time series.
3) *Filter length:* Filter sizes of $\{2, 4, 8\}$ were used, following the authors' recommendation for short time series.
4) *Number of filters:* Contrary to InceptionTime authors' architectural study, an increase in the filter count to 64 in all modules resulted in precision gains in these experiments, with no further improvements beyond this point.

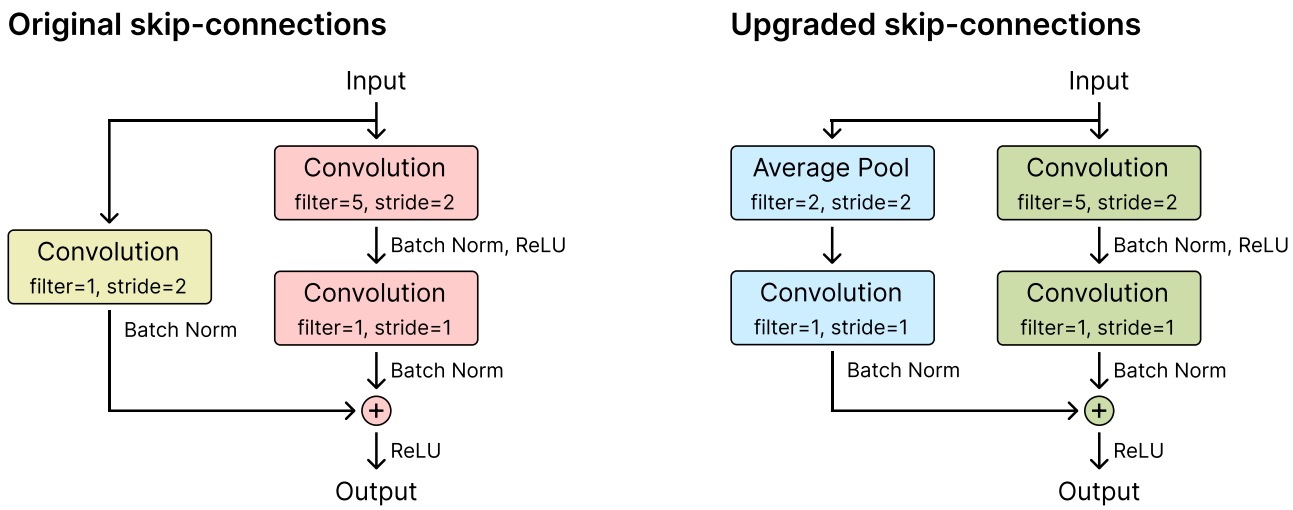

Fig. 8. Comparison between the original skip-connections and the upgraded skip-connections in ResNeTS. The numbers inside each layer represent the filter size and the stride size.

5) *Training procedure:* To ensure a fair and direct comparison between the design principles of each architecture, the modern training procedure of ResNeTS was adopted, as it incorporates more recent techniques than InceptionTime's original configuration. A comprehensive hyperparameter optimization for AdamW was also performed on the training points, testing the same ranges of values as those used for ResNeTS, resulting in selecting learning rate $= 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-2}$, and weight decay $= 1 \times 10^{-3}$.

All architectures were given a maximum of 1500 training epochs, a threshold set by InceptionTime, which had the highest epoch count among the models. In any case, checkpointing and early stopping were enabled, thereby preventing issues such as overfitting to ensure optimal results for all architectures. This is particularly necessary when running numerous training epochs on smaller architectures. Within each epoch, batches of 32 elements were supplied. While the original learning rates of most architectures were proportionally adjusted to this batch size, InceptionTime and ResNeTS had their optimizers specifically fine-tuned for these experiments, as previously described.

2) Experimental Procedure: Each model tested, including ResNeTS, was trained on the dataset described in Section II, with a single variable (Shannon, Simpson, or species richness) targeted for prediction at each instance. In particular, each combination of architecture and target variable underwent a group k-fold validation.

This k-fold process comprised five folds, each using 80% of the samples for training and 20% for testing. In addition, 20% of the training samples were set aside in a validation set to guide periodic actions such as checkpointing, early stopping, or learning rate scheduling. Thus, the sample distribution for training, validation, and testing sets stood at 64%, 16%, and 20%, respectively. All samples were normalized so the train set followed a zero mean and a standard deviation of one.

Grouping, applied during all stages of sample division, guaranteed that observations from the same plot made in consecutive years were consistently placed in the same set of training, validation, or testing. This approach effectively mitigated temporal dependencies, in pursuit of a robust analysis. Spatial autocorrelation between observations, within and across sites, was already discarded in [24]. Species compositions and numbers strongly depend on management intensity [59], which ultimately depends on the decision of each land owner rather than on location.

For each individual fold, accuracy metrics were computed alongside architecture training time. Upon completing the k-fold process, the mean and standard deviation of these metrics were reported. Given the inherently stochastic nature of training neural architectures, the k-fold process was repeated ten times for each combination of architecture and target variable, thus recording the averaged mean and standard deviation of the metrics. To ensure a fair comparison, the same ten seeds were used in all k-folds processes, guaranteeing an identical sample distribution across architectures.

In particular, the predictive capability of each architecture was determined using the following measures: the coefficient of determination (r^2), the RMSE, and the rRMSE, all calculated against the mean values of the testing set. In addition, the model-driven sRMSE and the data-driven uRMSE of the RMSE were computed to understand better the nature of the observed errors in predictions [75].

Furthermore, an explainability analysis was conducted to assess the importance of different spectral bands and time steps for the biodiversity predictions. This analysis was run on the predictions of ResNeTS-5, as it resulted in the most accurate architecture in this study.

The importance of spectral bands was evaluated using SHAP values, derived from the DeepLift algorithm [76]. For time steps, Grad-CAM attributions [77] were derived from the last convolutional block before the average pooling, and scaled from 8 to 16 steps. This block was selected following Grad-CAM's preference of complex and abstracted features of the input data. Grad-CAM was used for the time steps due to the significant role it places on the order of predictors, a critical factor for

TABLE III
RECORDED ACCURACIES FOR ALL THE PREDICTED VARIABLES USING ALL THE TESTED ARCHITECTURES

Target variable	Architecture	r^2	rRMSE	sRMSE	uRMSE
Shannon index	MLP	0.159 ± 0.042	0.234 ± 0.014	0.320 ± 0.038	0.377 ± 0.037
	Bi-LSTM	0.169 ± 0.050	0.225 ± 0.015	0.334 ± 0.057	0.335 ± 0.049
	Transformer	0.173 ± 0.063	0.215 ± 0.017	0.358 ± 0.050	0.263 ± 0.069
	FCN	0.218 ± 0.061	0.212 ± 0.013	**0.316** ± 0.033	0.318 ± 0.028
	Residual CNN	0.252 ± 0.053	0.200 ± 0.013	0.318 ± 0.033	0.278 ± 0.025
	InceptionTime	0.282 ± 0.060	0.194 ± 0.013	0.321 ± 0.030	0.257 ± 0.024
	InceptionTime-5	0.300 ± 0.066	0.191 ± 0.011	0.323 ± 0.028	0.241 ± 0.023
	ResNeTS	0.283 ± 0.054	0.193 ± 0.013	0.320 ± 0.037	0.253 ± 0.019
	ResNeTS-5	**0.308** ± 0.054	**0.188** ± 0.012	0.321 ± 0.035	**0.235** ± 0.017
	Rocket	0.232 ± 0.086	0.201 ± 0.010	0.337 ± 0.027	0.261 ± 0.026
Simpson index	MLP	0.094 ± 0.043	0.134 ± 0.013	0.086 ± 0.013	0.067 ± 0.007
	Bi-LSTM	0.088 ± 0.041	0.143 ± 0.013	0.085 ± 0.017	0.077 ± 0.012
	Transformer	0.108 ± 0.051	0.129 ± 0.017	0.094 ± 0.015	**0.043** ± 0.013
	FCN	0.111 ± 0.053	0.133 ± 0.013	**0.084** ± 0.012	0.068 ± 0.006
	Residual CNN	0.129 ± 0.062	0.128 ± 0.013	**0.084** ± 0.012	0.061 ± 0.007
	InceptionTime	0.121 ± 0.066	0.130 ± 0.013	0.085 ± 0.012	0.062 ± 0.007
	InceptionTime-5	0.146 ± 0.066	0.126 ± 0.013	**0.084** ± 0.011	0.057 ± 0.006
	ResNeTS	0.141 ± 0.058	0.125 ± 0.014	0.086 ± 0.012	0.054 ± 0.004
	ResNeTS-5	**0.166** ± 0.059	**0.121** ± 0.013	0.085 ± 0.012	0.049 ± 0.004
	Rocket	0.133 ± 0.065	0.126 ± 0.010	**0.084** ± 0.010	0.057 ± 0.004
Species richness	MLP	0.414 ± 0.052	0.284 ± 0.024	4.469 ± 1.247	7.418 ± 0.614
	Bi-LSTM	0.512 ± 0.045	0.246 ± 0.025	4.903 ± 1.065	5.694 ± 0.545
	Transformer	0.499 ± 0.069	0.250 ± 0.025	5.504 ± 1.058	5.249 ± 0.563
	FCN	0.539 ± 0.065	0.244 ± 0.027	4.408 ± 1.103	6.028 ± 0.558
	Residual CNN	0.568 ± 0.044	0.233 ± 0.021	**4.201** ± 1.034	5.753 ± 0.486
	InceptionTime	0.575 ± 0.044	0.230 ± 0.020	4.585 ± 0.920	5.352 ± 0.505
	InceptionTime-5	0.609 ± 0.035	0.222 ± 0.019	4.657 ± 0.758	**4.956** ± 0.394
	ResNeTS	0.596 ± 0.039	0.223 ± 0.018	4.318 ± 0.790	5.304 ± 0.481
	ResNeTS-5	**0.628** ± 0.036	**0.214** ± 0.015	4.283 ± 0.790	4.968 ± 0.447
	Rocket	0.312 ± 0.047	0.289 ± 0.031	6.671 ± 1.017	5.861 ± 0.361

InceptionTime-5 and ResNeTS-5 are the ensembles of five individual architectures. Higher r^2 and lower RMSE indicate better performances. The best values of each combination of variable and metric are in **bold**.

temporal analysis but less significant for the bands. By combining two explainability methods, the analysis aimed to leverage their strengths while also mitigating their individual limitations, thereby providing a more robust explanation of model predictions.

Specifically, during each step of the k-fold process of ResNeTS-5 for species richness, predictions for test data were compared to the corresponding in situ values to extract the explainability measures. For spectral bands, SHAP values were averaged over the temporal dimension; similarly, Grad-CAM attributions were averaged over the spectral dimension for time steps. These results were then averaged across the five folds of the k-fold process, and subsequently across all ten repetitions of the k-fold process, ensuring the robustness and reliability of the findings.

In addition, to explore the internal mechanisms that allow ResNeTS to capture key features in the time series data, the activations of the first convolutional block were examined across all species richness experiments. To this end, within each step of the k-fold repetitions, activation maps were extracted from each convolutional filter by averaging the outputs of the convolutional block over test data. These maps highlighted which temporal features were deemed most relevant in each learned convolution across different ResNeTS trainings, providing insights that were validated against the Grad-CAM attributions.

Regarding the activation visualization, it is straightforward to do so in the first convolutional block due to its small receptive field, which allows for a more precise correlation between the activation maps and the actual time steps in the input series. In contrast, deeper convolutional blocks entangle the impact of individual time steps due to the progressive increase of the receptive field from successive convolutions and pooling operations. Grad-CAM internally traces gradients in the last convolutional block back to their origin, taking care of this issue.

3) Execution Environment: All of the experiments were conducted on a computer equipped with an Intel Core i7-11700K CPU, 128 GB of RAM, and an NVIDIA RTX 3080 Ti GPU with 12 GB of VRAM. The system ran Ubuntu 20.04, and the source code was developed on Python 3.8.0 using PyTorch 1.13.0 [78] for neural architectures, and sktime 0.15.1 [79] for Rocket. Captum 0.6.0 provided the utilities for the explainability analysis. CUDA 11.8.0 and cuDNN 8.7.0.80 [80] were leveraged to speed up execution by using single precision arithmetic. The tool GuildAI was also used for automating and tracking all experiments.

B. Results

1) Comparison of All Architectures: Table III presents the accuracy comparison of ResNeTS and all the other architectures using r^2, rRMSE, sRMSE, and uRMSE metrics for predicting the Shannon and Simpson indices, and species richness. Moreover, visualizations of species richness, Shannon, and Simpson predictions versus in situ values are shown in the scatter plots of Figs. 9 and 10 along with the distribution of their values.

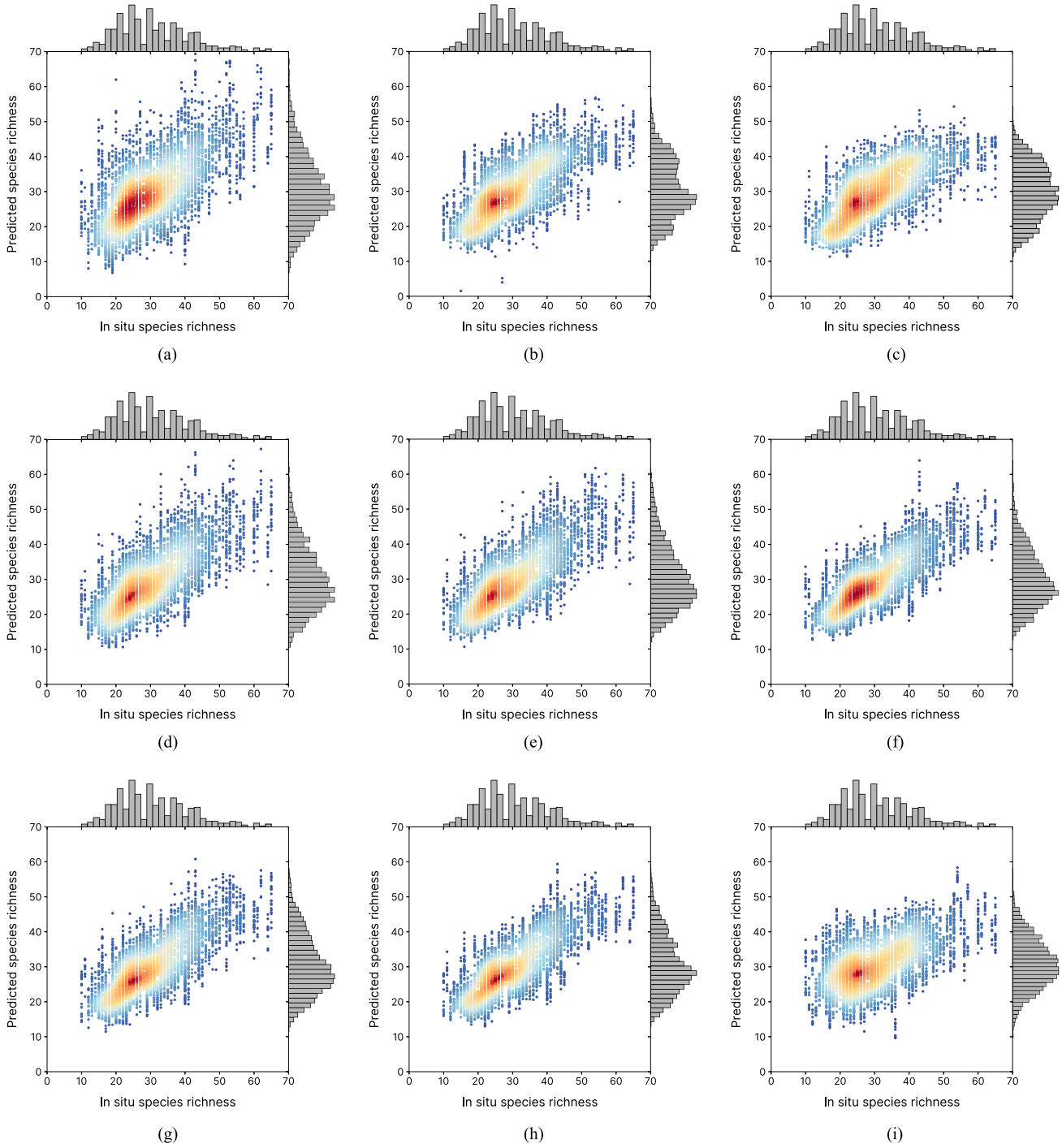

Fig. 9. Scatter plots comparing the species richness predictions of the tested architectures versus the in situ values. The predictions of each architecture are accumulated across all its experiments. (a) MLP. (b) Bi-LSTM. (c) Transformer. (d) FCN. (e) Residual CNN. (f) InceptionTime-5. (g) ResNeTS. (h) ResNeTS-5. (i) Rocket.

MLP significantly underperforms all other architectures due to its inability to capture temporal and dimensional patterns between data points in time series. Note how the RNN, CNNs, and the Transformer can model these temporal and dimensional properties, thereby achieving consistently higher r^2 coefficients and lower rRMSE values than the MLP, as Table III shows.

Even though the Bi-LSTM demonstrates an ability to understand temporal patterns, it still falls short of the MLP for the Simpson index. Moreover, when compared to the CNNs, the Bi-LSTM underperforms all architectures in every experiment. As the lower accuracy is present even when compared to the shallow FCN, the results suggest the improved capabilities of 1-D convolutions to extract more meaningful features from the time series data used.

On the one hand, the Transformer architecture improves the accuracy of the Bi-LSTM for both Shannon and Simpson indices

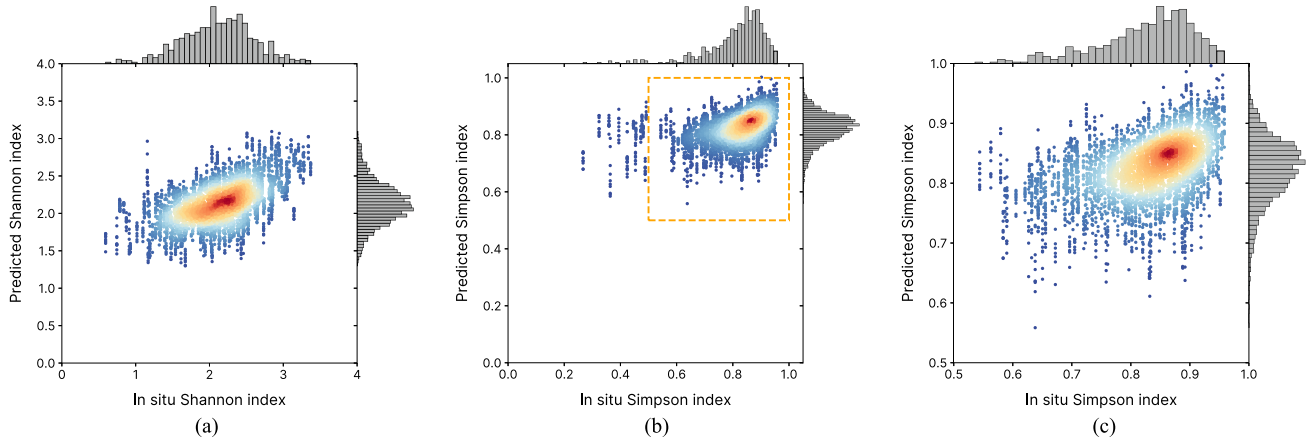

Fig. 10. Scatter plots comparing the Shannon and Simpson indices predictions of ResNeTS-5 versus the in situ values. The predictions of each variable are accumulated across all its experiments. The orange dashed box in the Simpson scatter plot marks a region with Simpson values above 0.5, where most of the recorded in situ values are found, and where an adequate relationship between the predicted and the actual values can be derived. (a) Shannon index. (b) Simpson index. (c) Simpson index (zoomed-in).

and achieves similar accuracy in predicting species richness. However, CNNs still surpass the Transformer in accuracy, indicating that convolutions are also better suited than self-attention mechanisms for analyzing the data in this study.

Focusing on the results of the CNNs, it can be seen that the progressive prediction improvements by descending on the table result from a reduction in uRMSE, while sRMSE remains at a closer level to the MLP. The range of uncertainty of the predictions across all values is also significantly lower with the CNNs with respect to the MLP, as Fig. 9 clearly shows.

Among the CNNs, the shallow FCN produces the least accurate predictions. As expected, the increased number of convolutional layers in the residual CNNs are more effective in extracting valuable information from the time series, thus leading to better predictions.

A comparison between all residual CNNs reveals ResNeTS's top-accuracy, improving the results of both the residual CNN and InceptionTime. In particular, ResNeTS's r^2 consistently achieves moderate improvements over InceptionTime's r^2 for all target variables, reaching up to +0.021 r^2.

When constructing ensembles comprising five copies of both ResNeTS and InceptionTime, an enhancement in the overall predictive performance is observed in both cases. Nonetheless, the ResNeTS ensemble continues to demonstrate competitive results compared to the InceptionTime ensemble, providing more accurate predictions for all variables.

Lastly, the Rocket regressor achieves accuracies that are lower than all the residual CNNs, and even worse than the MLP for species richness. These results contrast with studies such as [54], where Rocket can sometimes outperform advanced networks such as InceptionTime. A closer look at the scatterplot in Fig. 9(i) reveals how Rocket's predictions form a relatively wide-spread cloud of points, unlike the residual CNNs.

Up to this point, ResNeTS demonstrates a slight lead over InceptionTime in terms of accuracy. The computational performance of both architectures can become another deciding factor to further differentiate them. Table IV presents the average training time and throughput recorded for these architectures, including ensembles, aggregated across all experiments.

TABLE IV
COMPUTATIONAL COST OF TRAINING THE RESIDUAL CNNS

Metric	Architecture	Mean value	Speedup
Throughput (epochs/s)	InceptionTime	9.18	×1.35
	ResNeTS	12.39	
	InceptionTime-5	2.66	×1.61
	ResNeTS-5	4.29	
Train time (s)	InceptionTime	58.58	×1.54
	ResNeTS	37.94	
	InceptionTime-5	178.51	×1.69
	ResNeTS-5	105.78	

Higher speedups are better.

As indicated by the throughput, the lightweight architecture of ResNeTS achieves significantly better computational efficiency compared to InceptionTime. In particular, the individual ResNeTS can process 35% more samples than the individual InceptionTime within the same time frame, and the ResNeTS-5 ensemble further increases this speedup to ×1.61 over the InceptionTime-5 ensemble. A faster throughput contributes to both faster training and classification of new samples after training the architecture.

Interestingly, the training speedup of ResNeTS further increases to a maximum of ×1.69. This implies that, in addition to the faster throughput, ResNeTS architectures converge in earlier epochs while training, thus triggering the early stopping sooner.

As shown, the experiments conducted demonstrate that ResNeTS provides state-of-the-art accuracies for a residual CNN in time series analysis of remote sensing data, while being significantly faster than InceptionTime in both the training and classification stages, owing to its simpler architecture.

2) Architectural Analysis of ResNeTS: Table V displays the results of an ablation study conducted to provide empirical

TABLE V
RECORDED ACCURACIES FOR THE SPECIES RICHNESS VARIABLE WHEN APPLYING DIFFERENT MODIFICATIONS TO RESNETS

Target	Modification	r^2	rRMSE
–	Final configuration	**0.596** ± 0.039	**0.223** ± 0.018
Block Channels	2nd best: 64-64-128-256	0.591 ± 0.046	0.225 ± 0.016
	3rd best: 64-128-192-192	0.590 ± 0.040	0.225 ± 0.018
	4th best: 64-64-128-128	0.590 ± 0.044	0.226 ± 0.019
	5th best: 64-64-256-256	0.589 ± 0.042	0.226 ± 0.019
Block Repetitions	Two repetitions	0.568 ± 0.049	0.230 ± 0.018
Kernel Size	3	0.564 ± 0.040	0.232 ± 0.016
	7	0.591 ± 0.046	0.224 ± 0.015
	9	0.570 ± 0.048	0.230 ± 0.017
	11	0.575 ± 0.053	0.227 ± 0.016
	13	0.565 ± 0.059	0.230 ± 0.014
	15	0.583 ± 0.045	0.225 ± 0.016
Shortcut Pooling	Not included	0.582 ± 0.040	0.227 ± 0.015
Stem Channels	32	0.584 ± 0.042	0.226 ± 0.016
	64	0.585 ± 0.046	0.226 ± 0.016
	128	0.591 ± 0.038	0.224 ± 0.018
	160	0.592 ± 0.034	**0.223** ± 0.014
Stride Position	Block 1	0.593 ± 0.041	0.225 ± 0.017
	Block 2	0.586 ± 0.043	0.226 ± 0.015
	Block 4	0.577 ± 0.040	0.228 ± 0.015
	No stride	0.582 ± 0.047	0.228 ± 0.017
Training Procedure	Original	0.536 ± 0.056	0.240 ± 0.022

Higher r^2 values are better, while lower shap values are better. The best value of each metric is in **bold**.

support for the design process outlined in Section III for developing ResNeTS. Specifically, starting from the final architecture, various modifications are applied to evaluate their impact on prediction quality. The training hyperparameters, such as learning rate, are optimized with each modification for a fair comparison. For brevity, the results are only reported for species richness predictions, although similar trends were observed for Shannon and Simpson indices.

The architectural analysis demonstrates that, regardless of the modification made, the performance of ResNeTS consistently declines. It is noteworthy that while some modifications lead to slight decreases in accuracy, such as the removal of pooling from the shortcuts, others result in substantial downgrades, such as retaining the original training procedure.

Overall, this study demonstrates the effectiveness of the procedure outlined in Section III in successfully accommodating ResNet from image analysis to time series analysis, yielding ResNeTS, a highly competitive architecture for the dataset of interest in this study.

3) Importance of Bands and Time Steps: Figs. 11–13 summarize the results of an explainability analysis conducted on ResNeTS-5 to determine the importance of bands and time steps in predicting species richness. The focus on this variable offers a more robust understanding of predictor importance, given the higher accuracies of all architectures compared to Shannon and Simpson indices.

In particular, Fig. 11 presents the SHAP values assigned to each band, while Fig. 12 reflects the importance of each time step as approximated by Grad-CAM. The predictions of ResNeTS draw significantly from all features, although the architecture appears to leverage some more than others. Notably, the bands *red edge 2*, *near-infrared*, *near-infrared b*, and *short-wave infrared 2* were found to be of greatest importance. Moreover,

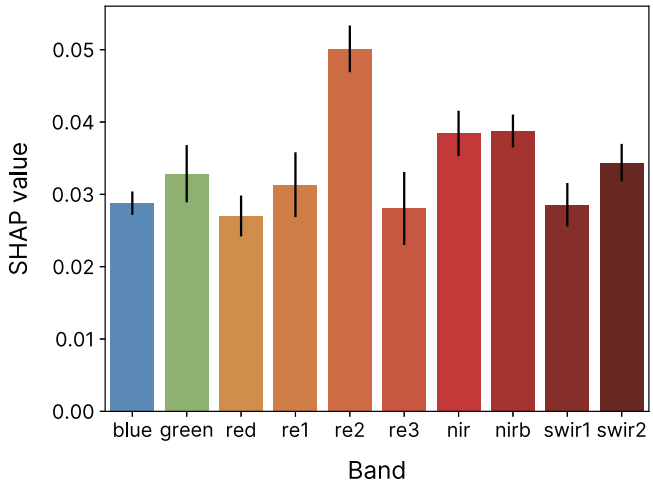

Fig. 11. Importance of each band in ResNeTS-5 predictions for species richness, given by SHAP values. The importance is measured and averaged across all experiments. Higher values are better.

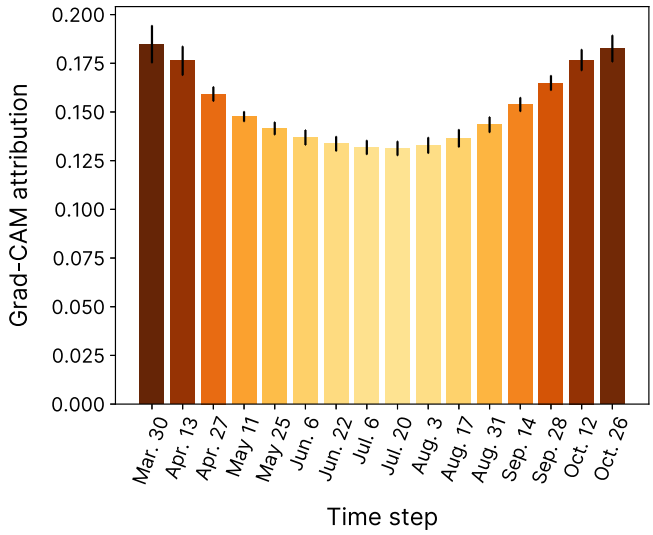

Fig. 12. Importance of each time step in ResNeTS-5 predictions for species richness, given by Grad-CAM. The importance is measured and averaged across all experiments. Higher values are better.

observations from the start and the end of the growing season (April and October, respectively) proved to be highly relevant. The consistency observed in the importance across all data points underscores the reliability of the results.

Although not displayed here due to space limitations, the activation maps visualized across different ResNeTS trainings on species richness align with the insights of the Grad-CAM analysis. Generally speaking, the maps show that the learned convolutional filters fall into three distinct yet nonexclusive categories: those emphasizing the beginning of the series, those focusing on the end, and those highlighting the middle portions, with the latter being significantly less common. Representative examples of different types of activation maps are displayed in Fig. 13.

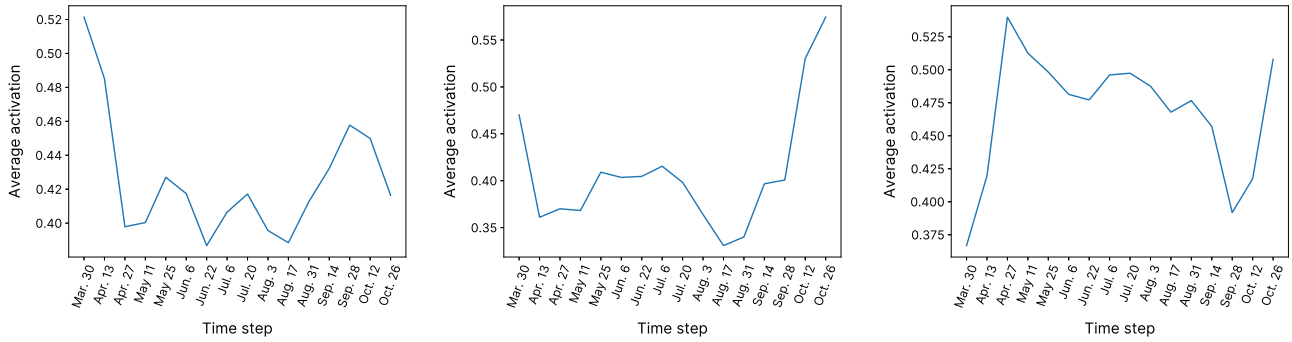

Fig. 13. Examples of different types of activation maps extracted from the first convolutional block of ResNeTS in predicting species richness. (a) Map of a convolutional filter emphasizing the beginning of the series. (b) Map of a convolutional filter focusing on the end of the series. (c) Map of a convolutional filter highlighting the middle portions and the end of the series.

V. Discussion

A. Motivation for ResNeTS

Although there is a wide array of literature predicting biodiversity with remote sensing data and machine learning, recent researchers have raised awareness of their limitations due to spatial autocorrelation [19], poor domain adaptation [81], or low performance of metrics based on the spectral diversity hypothesis [22]. An approach to partially overcome these problems is to use radiative transfer models to predict the functional diversity rather than taxonomic diversity [17]. However, functional and species diversity do not always go hand in hand. Grassland communities with a high number of species may have a low functional diversity due to high functional redundancy. Thus, species richness and distribution measures such as the ones presented in this article represent a valuable set of information for management and conservation measures.

Only a few studies make simultaneous use of the temporal and spectral dimensions to quantify taxonomic diversity; e.g., [14], [24], [82]. Other studies use only a few points in time or aggregate the temporal dimension with statistical metrics [83], [84], or aggregate the range of predicted values into a few classes [85], losing important information. Recent works still report high prediction accuracies using traditional machine learning methods without discarding spatial autocorrelation [86], which renders the use of remote sensing data potentially unnecessary [19]. [87] also pointed out the challenges encountered when modeling phenological patterns due to species diversity and image availability. Despite the success of deep learning in the field of image processing, these architectures are rarely used with full spectral-temporal data for the task of biodiversity prediction. Some architectures have shown good results when specifically fine-tuned to specific uses, such as MLPs in [24]. Although MLPs may be sufficient when predicting relatively simple biophysical variables such as above-ground biomass or LAI, more complex architectures are needed to interpret high dimensional data.

As discussed in Section I, remote sensing researchers working on tasks beyond biodiversity prediction, such as land cover classification, have already adopted advanced deep learning architectures to extend the capabilities of time series analysis [33], [34], [39], [40]. Inspired by these advancements, this work develops a powerful residual CNN to significantly improve biodiversity monitoring beyond the common MLP-based approaches.

Among the current CNNs for time series analysis, InceptionTime [53] is known for its robust performance [54]. It consists of modules that apply parallel convolutions to extract relevant features from time series data, drawing inspiration from the Inception-v4 [56] network introduced in 2017. However, most improvements in CNN development are driven by image processing, and its most advanced CNNs currently favor sequential designs over parallel ones [50], [57], [58], drawing inspiration from the success of the ResNet family [44]. This trend suggests that an effective ResNet-based CNN for time series analysis can open an interesting opportunity to advance the field by facilitating the incorporation of cutting-edge techniques from the most advanced image processing CNNs.

This observation led to the development of ResNeTS, a residual CNN resulting from adapting the ResNet architecture for time series analysis. Unlike InceptionTime, which relies on parallel convolutions with filters of varying sizes to accommodate different time series lengths, ResNeTS embraces a simpler design, consisting of sequential convolutions better suited for the characteristics of the time series studied (i.e., length and dimensionality). Consequently, ResNeTS marks a significant step toward aligning the design principles of CNNs across both time series and image processing.

B. Experimental Results

Empirical results demonstrate how ResNeTS allows the analysis of Sentinel-2 time series with improved accuracy over InceptionTime. Specifically, Shannon, Simpson, and species richness biodiversity indices are predicted, reaching gains of up to +0.02 r^2. Furthermore, ResNeTS accomplishes this with fewer convolutional blocks than InceptionTime, as reflected in Fig. 6(b). This reduction in blocks, coupled with their light design (see Fig. 6(a)), enable ResNeTS to achieve speedups of up to $\times 1.69$ in terms of execution time. This impressive speed

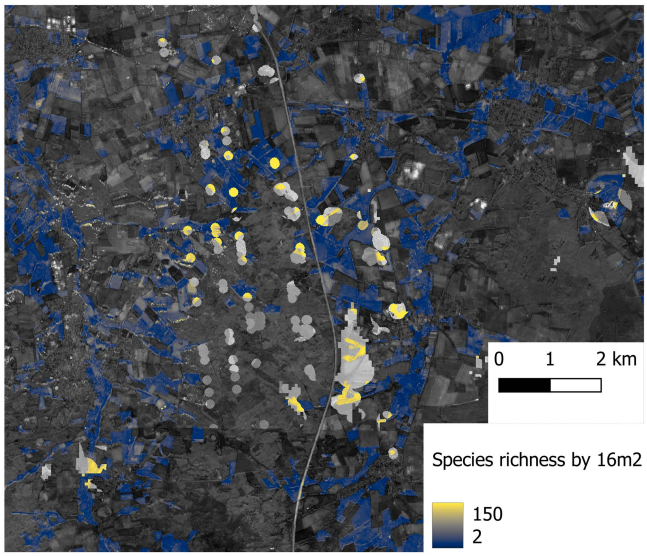

Fig. 14. Predictions over a poorly FORCE [63] interpolated time series outside the Biodiversity Exploratories due to cloud contamination. In these areas, the architecture returns values above 150 species per 16 m^2, which is not plausible, as the maximum species richness value registered is 61 per 16 m^2.

enhancement is especially beneficial for large-scale monitoring, real-time applications, and addressing more complex time series data efficiently.

The overall better performance of ResNeTS over InceptionTime and the other architectures can be better understood by looking at the uRMSE. The sRMSE represents the error that can be attributed to the architecture's bias, and ResNeTS showed only a moderate improvement (see Table III). On the other hand, the uRMSE represents the random error component, and its magnitude and variability across folds are consistently much lower in ResNeTS, compared to the other architectures. This indicates that ResNeTS is better at dealing with outliers and might perform consistently under scenarios with a slightly different data distribution. This is important for remote sensing studies, where unmasked haze or clouds can introduce some noise in the interpolated signal. Under circumstances with extremely bad quality pixels, the architecture returns unfeasible high species numbers that, in any case, are easy to detect and exclude during map production; i.e., above 150 species per 16 m^2, when the maximum number recorded at the Biodiversity Exploratories is 61 (see Fig. 14). Note that the Sentinel-2 time series used for calibrating the architecture was screened for poor-quality pixels in the Biodiversity Exploratory plots, and these extreme sources of noise do not affect the training.

It is also worth noting that, in recent years, the Transformer architecture has gained significant attention across various areas due to its competitive performance against other deep learning architectures, including CNNs. The Transformer evaluated in this study was also developed for handling Sentinel-2 time series [34]. Surprisingly, even basic CNN architectures in the experimental trials outperformed the Transformer. One potential explanation, as discussed in [51], could be the small size of the dataset used, limited by challenges such as the labor-intensive nature of in-site sample collection, and the incompatibility of aggregating data from different sources (see Section II). Other authors have already observed that Transformers are data-intensive architectures, possibly struggling in low-data regimes [88].

When compared to results from previous works, ResNeTS improved the predictions of species richness from the MLP in [24] ($r^2 = 0.42$ versus $r^2 = 0.63$), and improved the predictions of the Shannon index to acceptable levels compared to those achieved with MLPs in [14] ($r^2 = 0.23$ versus $r^2 = 0.31$). The improvement in the coefficient of determination of Simpson index predictions by ResNeTS with respect to the MLP was marginal ($r^2 = 0.16$ versus $r^2 = 0.17$). However, these errors are concentrated at low values, which correspond to plots not dominated by any particular species. This poor performance at predicting low Simpson values is probably due to the scarcity of training points for values below 0.5, as shown in Fig. 10. At mid and high Simpson values (few or no dominant species) the predictions are comparable to those of other similar research [82].

C. Model Interpretability

Regarding the feature importance of the Sentinel-2 bands, the results in this study are in agreement with [24]. This is, the *red edge 2*, *near-infrared*, *near-infrared b*, and *short-wave infrared 2* have the highest prediction importance for the model (see Fig. 11). These bands are highly sensitive to vegetation, as they can penetrate plant structures and reflect key attributes such as water and chlorophyll content. Since different plant species present distinct signatures for these characteristics, the bands enable a more accurate prediction of plant biodiversity.

Concerning the feature importance across the phenological cycle, ResNeTS assigns high importance values to the start and end of the growing season (April and October, respectively), and lower importance toward the summer (see Fig. 12). This is also to some extent in agreement with [24], where May is featured as very important, rather than April. This consistency across models suggests that the relative importance of each predictor is not model-dependent, and strengthens the interpretability of the model and its results. The flowering season starts in April/May, as well as the first mowing of the most intensive grasslands (and thus species-poor plots). During October, the browning season starts, and only the extensive grasslands (species rich) have green and senescent vegetation, whereas more intensive grasslands have either fully vegetated grass or have been recently mown (see Fig. 4).

It is worth noting that averaging the explainability measures over temporal and spectral dimensions poses a potential limitation of the analysis conducted. This approach, while facilitating the interpretation of the results, might mask finer dynamics, such as certain bands exhibiting more influence during particular time intervals; these subtleties are not captured in the aggregated values shown in Fig. 11. In addition, the analysis was run solely on species richness, motivated by the high accuracy of all models compared to the Shannon and Simpson indexes. Should future models achieve comparable accuracy levels across

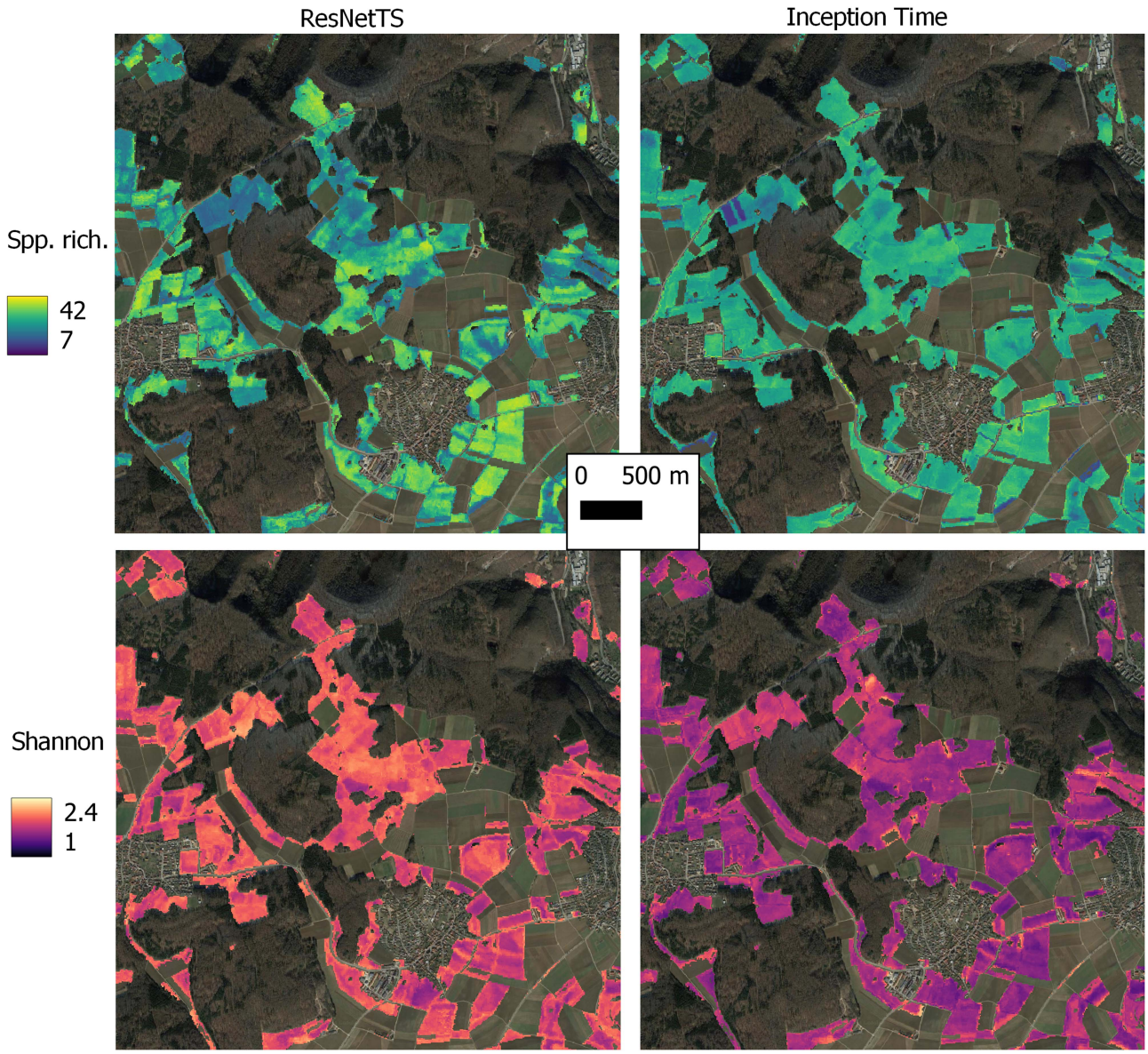

Fig. 15. Visual differences in species richness and Shannon index predictions between ResNeTS and InceptionTime across various grassland patches.

these metrics, expanding the explainability analysis could help determine whether the predictors identified as most important for species richness hold the same influence for the other biodiversity metrics.

D. Applications

LUI is one of the main drivers of species richness and composition, but its effects are in some cases nonlinear: while some meadows under a low LUI exhibit a low species richness, some pastures under intermediate LUI are home to a high number of species [3], [66]. Spatially explicit biodiversity data can contribute to a better understanding of the effects of land use on biodiversity and its relationship with ecosystem functions and services. Such data can be used to estimate other ecosystem services such as carbon storage [89] and pollination abundance [90]. Biodiversity predictions can also serve as valuable input in other models such as structure equation models, mathematic models such as Marxsan [91], or additional deep learning models [92], that attempt to disentangle the complex interactions between biodiversity, climatic and anthropogenic factors, and orient the development of conservation plans and policies.

Assessing more than one biodiversity metric across a landscape can provide a comprehensive view of biodiversity status. Fig. 15 illustrates the differences in predictions between ResNeTS and InceptionTime for the Shannon and species richness variables across various grassland patches. ResNeTS predicts a broader range of values compared to InceptionTime, as expected by its higher accuracies. Interestingly, although species richness and the Shannon diversity index often correlate, the figure also demonstrates that this is not always the case,

as they represent two independent components of taxonomic diversity [67].

The proposed ResNeTS offers a flexible architecture that can be applied to study not only plant biodiversity, but also any variable that can be inferred from the phenological profile of vegetation, such as crop yield prediction [37], phenometrics [12], or LUI [93], among many others. ResNeTS can be applied directly to other remote sensing time series without modifications, while also allowing for adjustments to achieve the most optimal performance for specific scenarios. For instance, the updated training procedure and the inclusion of pooling layers in the shortcuts are likely to provide benefits in all contexts. In contrast, in cases involving longer or higher dimensional time series, modifications such as larger kernels, increased convolution channels, or changes to the stride policy might be advantageous. For instance, a denser time series might be necessary to completely model the phenological profiles of other vegetation species. Naturally, ResNeTS can be tailored for classification tasks in addition to regression, by simply adapting the output layer to fit the problem at hand.

It is also worth noting how ResNeTS demonstrated its ability to effectively work with a small-sized dataset, presenting only a few hundred data points. Naturally, as with any machine learning model, it is still necessary to screen all data to minimize extremely bad quality inputs, such as pixels heavily obstructed by clouds (see Fig. 14). For larger datasets, it could still be worthwhile to explore larger architectures, such as increasing the number of repetitions per block, the number of convolutional channels, or even experimenting with the larger ResNet variants than ResNet-18, which served as baseline for ResNeTS.

VI. Conclusion

This work introduced ResNeTS, a residual CNN designed for the analysis of Sentinel-2 data, and demonstrated its capabilities to enhance the prediction of biodiversity metrics, specifically species richness and Shannon and Simpson indices, with regard to other state-of-the-art architectures.

In particular, ResNeTS successfully accommodates to time series analysis the well-known ResNet computer vision architecture, and incorporates enhancements such as extended convolutional kernels for improved temporal pattern recognition, state-of-the-art training methodologies, and model ensembling to reduce the variability in predictions. These innovations allow ResNeTS to maximize the use of the spectral and temporal resolution of the time series, outperforming leading networks such as InceptionTime and Transformers, in both terms of accuracy and execution time.

The advancements presented by ResNeTS extend beyond technical achievements, into biodiversity research and management. By accurately predicting biodiversity metrics across extensive regions, ResNeTS facilitates the creation of combined maps that reflect plant species richness and evenness across agricultural landscapes, and at resolutions compatible with management practices. Furthermore, integrating these predictions with climate and land-use data can help disentangle the effects of climatic and human-induced factors on biodiversity patterns, offering invaluable insights for conservation strategies and ecological studies. Finally, ResNeTS' flexibility also makes it a compelling architecture to explore in applications beyond biodiversity monitoring, such as crop yield prediction, or invasive species detection.

Acknowledgment

The authors would like to thank the managers of the three exploratories, Kirsten Reichel-Jung, Iris Steitz, Sandra Weithmann, Juliane Vogt, Miriam Teuscher, and all former managers, for their work in maintaining the plot and project infrastructure; Victoria Grießmeier for giving support through the central office; Andreas Ostrowski for managing the central database; and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser, and the late Elisabeth Kalko, for their role in setting up the Biodiversity Exploratories project. The authors would like to thank the administration of the Hainich National Park, the UNESCO Biosphere Reserve of Schwäbische Alb, and the UNESCO Biosphere Reserve Schorfheide-Chorin, as well as all landowners, for the excellent collaboration. The authors would also like to thank Ralph Bolliger for providing and maintaining the species inventory datasets, Stephan Wollawer for maintaining the remote sensing database, Florian Männer for processing the biodiversity indices, and Lisa Schwarz for the photos of plant species provided.

References

[1] *IPBES*, "Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services," 2019. Accessed: Oct. 2023. [Online]. Available: https://zenodo.org/record/3831673

[2] R. T. Guuroh, J. C. Ruppert, J. Ferner, K. Čanak, S. Schmidtlein, and A. Linstädter, "Drivers of forage provision and erosion control in west African savannas–A macroecological perspective," *Agriculture Ecosyst. Environ.*, vol. 251, pp. 257–267, 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167880917304176

[3] F. V. D. Plas, "Biodiversity and ecosystem functioning in naturally assembled communities," *Biol. Rev.*, vol. 94, pp. 1220–1245, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/brv.12499

[4] S. Díaz et al., "Pervasive human-driven decline of life on Earth points to the need for transformative change," *Science*, vol. 366, no. 6471, 2019, Art. no. eaax3100. [Online]. Available: https://www.science.org/doi/10.1126/science.aax3100

[5] J. L. McGuire, A. M. Lawing, S. Díaz, and N. C. Stenseth, "The past as a lens for biodiversity conservation on a dynamically changing planet," *Proc. Nat. Acad. Sci. USA*, vol. 120, no. 7, 2023, Art. no. e2201950120, doi: 10.1073/pnas.2201950120.

[6] J. Isselstein, B. Jeangros, and V. Pavlu, "Agronomic aspects of biodiversity targeted management of temperate grasslands in Europe–A review," *Agronomy Res.*, vol. 3, no. 2, pp. 139–151, 2005.

[7] R. L. Schils et al., "Permanent grasslands in Europe: Land use change and intensification decrease their multifunctionality," *Agriculture Ecosyst. Environ.*, vol. 330, 2022, Art. no. 107891. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167880922000408

[8] Y. Xie, Z. Sha, and M. Yu, "Remote sensing imagery in vegetation mapping: A review," *J. Plant Ecol.*, vol. 1, no. 1, pp. 9–23, Mar. 2008, doi: 10.1093/jpe/rtm005.

[9] C. Li, J. Xue, and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, 2017, Art. no. 1353691, doi: 10.1155/2017/1353691.

[10] J. Muro et al., "Land surface temperature trends as indicator of land use changes in wetlands," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 70, pp. 62–71, 2018. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0303243418301144

[11] C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 55–72, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271616000769

[12] L. Zeng, B. D. Wardlow, D. Xiang, S. Hu, and D. Li, "A review of vegetation phenological metrics extraction using time-series, multispectral satellite data," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111511. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719305309

[13] S. Sesnie, C. Espinosa, A. J.-Guerrero, and M. T.-Armijos, "Ensemble machine learning for mapping tree species alpha-diversity using multisource satellite data in an Ecuadorian seasonally dry forest," *Remote Sens.*, vol. 15, no. 3, 2023, Art. no. 583. [Online]. Available: https://www.mdpi.com/2072-4292/15/3/583

[14] J. Muro, A. Linstädter, F. A. Männer, L.-M. Schwarz, J. Hoffmann, and O. Dubovyk, "Predicting vegetation attributes with neural networks and sentinel-1 & 2," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B3-2022, pp. 945–950, 2022. [Online]. Available: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B3-2022/945/2022/

[15] A. K. Schweiger and E. Laliberté, "Plant beta-diversity across biomes captured by imaging spectroscopy," *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 2767. [Online]. Available: https://www.nature.com/articles/s41467-022-30369-6

[16] H. Polley, C. Yang, B. Wilsey, and P. Fay, "Spectral heterogeneity predicts local-scale gamma and beta diversity of mesic grasslands," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 458. [Online]. Available: http://www.mdpi.com/2072-4292/11/4/458

[17] X. Ma et al., "Inferring plant functional diversity from space: The potential of sentinel-2," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111368. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719303876

[18] J. Ferner, A. Linstädter, C. Rogass, K.-H. Südekum, and S. Schmidtlein, "Towards forage resource monitoring in subtropical savanna grasslands: Going multispectral or hyperspectral?," *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 364–384, 2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/22797254.2021.1934556

[19] P. Ploton et al., "Spatial validation reveals poor predictive performance of large-scale ecological mapping models," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 4540. [Online]. Available: http://www.nature.com/articles/s41467-020-18321-y

[20] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 24–49, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620303488

[21] H. Zhu et al., "Predicting plant diversity in beach wetland downstream of Xiaolangdi reservoir with UAV and satellite multispectral images," *Sci. Total Environ.*, vol. 819, 2022, Art. no. 153059. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0048969722001498

[22] F. E. Fassnacht, J. Müllerová, L. Conti, M. Malavasi, and S. Schmidtlein, "About the link between biodiversity and spectral variation," *Appl. Vegetation Sci.*, vol. 25, no. 1, 2022, Art. no. e12643, doi: 10.1111/avsc.12643.

[23] D. E. Rumelhart and J. L. McClelland, *Learn. Intern. Representations by Error Propag.*, Cambridge, MA, USA: MIT Press, 1987, pp. 318–362. [Online]. Available: https://ieeexplore.ieee.org/document/6302929

[24] J. Muro et al., "Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning," *Remote Sens. Environ.*, vol. 282, 2022, Art. no. 113262. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722003686

[25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[26] Y. Cai et al., "A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, pp. 35–47, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425718300610

[27] W. Xu, W. Yang, P. Chen, Y. Zhan, L. Zhang, and Y. Lan, "Cotton fiber quality estimation based on machine learning using time series UAV remote sensing data," *Remote Sens.*, vol. 15, no. 3, 2023, Art. no. 586. [Online]. Available: https://www.mdpi.com/2072-4292/15/3/586

[28] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, Jun. 1989, doi: 10.1162/neco.1989.1.2.270.

[29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[31] K. Cho, B. v. Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[32] H. Ma and S. Liang, "Development of the glass 250-m leaf area index product (version 6) from modis data using the bidirectional LSTM deep learning model," *Remote Sens. Environ.*, vol. 273, 2022, Art. no. 112985. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722000992

[33] H. C. d. C. Filho et al., "Rice crop detection using LSTM, bi-LSTM, and machine learning models from Sentinel-1 time series," *Remote Sens.*, vol. 12, no. 16, 2020, Art. no. 2655. [Online]. Available: https://www.mdpi.com/2072-4292/12/16/2655

[34] M. Rußwurm and M. Körner, "Self-attention for raw optical satellite time series classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 421–435, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301647

[35] A. Garioud, S. Valero, S. Giordano, and C. Mallet, "Recurrent-based regression of sentinel time series for continuous vegetation monitoring," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112419. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425721001371

[36] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[37] J. Xu et al., "DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping," *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111946. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720303163

[38] Z. Lin et al., "Large-scale rice mapping using multi-task spatiotemporal deep learning and sentinel-1 SAR time series," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 699. [Online]. Available: https://www.mdpi.com/2072-4292/14/3/699

[39] H. Wang, X. Zhao, X. Zhang, D. Wu, and X. Du, "Long time series land cover classification in China from 1982 to 2015 based on bi-LSTM deep learning," *Remote Sens.*, vol. 11, no. 14, 2019, Art. no. 1639. [Online]. Available: https://www.mdpi.com/2072-4292/11/14/1639

[40] Y. Xi, C. Ren, Q. Tian, Y. Ren, X. Dong, and Z. Zhang, "Exploitation of time series Sentinel-2 data and different machine learning algorithms for detailed tree species classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7589–7603, 2021.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[42] H. Zhao, Z. Chen, H. Jiang, W. Jing, L. Sun, and M. Feng, "Evaluation of three deep learning models for early crop classification using Sentinel-1a imagery time series–A case study in Zhanjiang, China," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2673. [Online]. Available: https://www.mdpi.com/2072-4292/11/22/2673

[43] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425718305418

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[45] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 1578–1585.

[46] G. Paolini, M. J. Escorihuela, O. Merlin, M. P. Sans, and J. Bellvert, "Classification of different irrigation systems at field scale using time-series of remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 10 055–10 072, 2022.

[47] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666651022000146

[48] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding unet for remote sensing image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408715.

[49] S. K. Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, "Multimodal fusion transformer for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5515620.

[50] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 11976–11986.

[51] Y. Liu, E. Sangineto, W. Bi, N. Sebe, B. Lepri, and M. Nadai, "Efficient training of visual transformers with small datasets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 23 818–23 830.

[52] T. A. Warner, A. E. Maxwell, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.

[53] H. Ismail Fawaz et al., "InceptionTime: Finding AlexNet for time series classification," *Data Mining Knowl. Discov.*, vol. 34, no. 6, pp. 1936–1962, 2020, doi: 10.1007/s10618-020-00710-y.

[54] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, "The great multivariate time series classification bake off: A review and experimental evaluation of recent algorithmic advances," *Data Mining Knowl. Discov.*, vol. 35, no. 2, pp. 401–449, 2021, doi: 10.1007/s10618-020-00727-3.

[55] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "BREIZHCROPS: A time series dataset for crop type mapping," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 2020, pp. 1545–1551, 2020.

[56] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11231

[57] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324.

[58] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. 38th Int. Conf. Mach. Learn.*, PMLR, 2021, pp. 10 096–10 106. [Online]. Available: https://proceedings.mlr.press/v139/tan21a.html

[59] N. Blüthgen et al., "A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization," *Basic Appl. Ecol.*, vol. 13, no. 3, pp. 207–220, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1439179112000424

[60] M. Fischer et al., "Implementing large-scale and long-term functional biodiversity research: The biodiversity exploratories," *Basic Appl. Ecol.*, vol. 11, no. 6, pp. 473–485, 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S143917911000099X

[61] C. J. Keylock, "Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy," *Oikos*, vol. 109, no. 1, pp. 203–207, 2005, doi: 10.1111/j.0030-1299.2005.13735.x.

[62] R. Bolliger, D. Prati, and M. Fisher, "Vegetation records for grassland EPs, 2008-2020. biodiversity exploratories information system (bexis). dataset ID = 27386," 2021. Accessed: Oct. 2023. [Online]. Available: https://www.bexis.uni-jena.de/

[63] D. Frantz, "Force–landsat sentinel-2 analysis ready data and beyond," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1124. [Online]. Available: https://www.mdpi.com/2072-4292/11/9/1124

[64] E. S. Agency, "Spectral resolution–Sentinel online," 2023. Accessed: Sep. 25, 2023. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

[65] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0924271619301108

[66] V. Busch et al., "Will i stay or will i go? Plant species-specific response and tolerance to high land-use intensity in temperate grassland ecosystems," *J. Vegetation Sci.*, vol. 30, no. 4, pp. 674–686, 2019, doi: 10.1111/jvs.12749.

[67] A. E. Magurran, "Measuring biological diversity," *Curr. Biol.*, vol. 31, no. 19, pp. R1174–R1177, 2021.

[68] W. E. Kunin et al., "Upscaling biodiversity: Estimating the species–area relationship from small samples," *Ecological Monographs*, vol. 88, no. 2, pp. 170–187, 2018. [Online]. Available: https://www.jstor.org/stable/26598644

[69] W. E. Kunin et al., "Upscaling biodiversity: Estimating the species–area relationship from small samples," *Ecological Monographs*, vol. 88, no. 2, pp. 170–187, 2018, doi: 10.1002/ecm.1284.

[70] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[71] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 558–567.

[72] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," 2017, *arXiv:1711.05101*. [Online]. Available: http://arxiv.org/abs/1711.05101

[73] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with restarts," 2016, *arXiv:1608.03983*. [Online]. Available: http://arxiv.org/abs/1608.03983

[74] A. Dempster, F. Petitjean, and G. I. Webb, "ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining Knowl. Discov.*, vol. 34, no. 5, pp. 1454–1495, Sep. 2020, doi: 10.1007/s10618-020-00701-z.

[75] C. J. Willmott, "On the validation of models," *Phys. Geogr.*, vol. 2, no. 2, pp. 184–194, 1981, doi: 10.1080/02723646.1981.10642213.

[76] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[77] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.

[78] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, Eds., 2019, vol. 32. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[79] M. Löning, A. J. Bagnall, S. Ganesh, V. Kazakov, J. Lines, and F. J. Király, "sktime: A unified interface for machine learning with time series," 2019, *arXiv:1909.07872*.

[80] S. Chetlur et al., "cuDNN: Efficient primitives for deep learning," 2014, *arXiv:1410.0759*.

[81] H. Meyer and E. Pebesma, "Machine learning-based global maps of ecological variables and the challenge of assessing them," *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 2208. [Online]. Available: https://www.nature.com/articles/s41467-022-29838-9

[82] M. Fauvel et al., "Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111536. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0034425719305553

[83] H. Gholizadeh, J. A. Gamon, C. J. Helzer, and J. C.-Bares, "Multi-temporal assessment of grassland α- and β-diversity using hyperspectral imaging," *Ecological Appl.*, vol. 30, no. 7, 2020, Art. no. e02145, doi: 10.1002/eap.2145.

[84] S. Bae et al., "Radar vision in the mapping of forest biodiversity from space," *Nature Commun.*, vol. 10, no. 1, 2019, Art. no. 4757. [Online]. Available: http://www.nature.com/articles/s41467-019-12737-x

[85] R. A. Crabbe, D. Lamb, and C. Edwards, "Discrimination of species composition types of a grazed pasture landscape using sentinel-1 and sentinel-2 data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 84, 2020, Art. no. 101978. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243419305768

[86] A. Safonova, G. Ghazaryan, S. Stiller, M. M.-Knorn, C. Nendel, and M. Ryo, "Ten deep learning techniques to address small data problems with remote sensing," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 125, Dec. 2023, Art. no. 103569. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S156984322300393X

[87] N. Younes, K. E. Joyce, and S. W. Maier, "All models of satellite-derived phenology are wrong, but some are useful: A case study from northern Australia," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 97, 2021, Art. no. 102285. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0303243420309284

[88] Z. Lu, H. Xie, C. Liu, and Y. Zhang, "Bridging the gap between vision transformers and convolutional neural networks on small datasets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 14 663–14 677. [Online]. Available: https://openreview.net/forum?id=bfz-jhJ8wn

[89] Y. Yang, D. Tilman, G. Furey, and C. Lehman, "Soil carbon sequestration accelerated by restoration of grassland biodiversity," *Nature Commun.*, vol. 10, no. 1, Feb. 2019, Art. no. 718, doi: 10.1038/s41467-019-08636-w.

[90] H. Feilhauer, D. Doktor, S. Schmidtlein, and A. K. Skidmore, "Mapping pollination types with remote sensing," *J. Vegetation Sci.*, vol. 27, no. 5, pp. 999–1011, 2016. [Online]. Available: http://www.jstor.org/stable/44132892

[91] I. Ball, H. Possingham, and M. Watts, *Spatial Conservation Prioritisation: Quantitative Methods and Computational Tools. Chapter 14: Marxan and Relatives: Software for Spatial Conservation Prioritisation*. Oxford, U.K.: Oxford Univ. Press, 2009.

[92] D. Silvestro, S. Goria, T. Sterner, and A. Antonelli, "Improving biodiversity protection through artificial intelligence," *Nature Sustainability*, vol. 5, no. 5, pp. 415–424, 2022. [Online]. Available: https://www.nature.com/articles/s41893-022-00851-6

[93] M. Lange, H. Feilhauer, I. Kühn, and D. Doktor, "Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series," *Remote Sens. Environ.*, vol. 277, 2022, Art. no. 112888. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722000025

Álvaro G. Dieste received the B.Sc. degree in computer engineering and the M.Sc. degree in high performance computing in 2021 and 2022, respectively, from the University of Santiago de Compostela, Santiago, Spain, where he is currently working toward the Ph.D. degree in computer science.

He is currently a Predoctoral Researcher with the Singular Research Center on Intelligent Technologies (CiTIUS), University of Santiago de Compostela. He integrates knowledge from deep learning, high performance computing, and big data to advance his field. His research interests focus on developing computer vision techniques for vegetation monitoring using remote sensing data.

Francisco Argüello received the B.Sc. and Ph.D. degrees in physics from the University of Santiago de Compostela, Santiago, Spain, in 1988 and 1992, respectively.

He is currently a Full Professor with the Department of Electronic and Computer Engineering, University of Santiago de Compostela. His research interests include signal and image processing, computer graphics, parallel and distributed computing, and quantum computing.

Dora B. Heras (Member, IEEE) received the M.Sc. and Ph.D. degrees in physics from the University of Santiago de Compostela, Santiago, Spain, in 1994 and 2000, respectively.

She is currently a Full Professor with the Department of Electronics and Computer Engineering, University of Santiago de Compostela. She has authored or coauthored papers on high performance computing, registration, classification, domain adaptation, and change detection applied to remotely sensed images. Her research interests include a range of topics in the combined fields of image processing, remote sensing, machine learning, and high performance computing.

Dr. Heras has been CoChair of the International Euro-Par conference since 2023. Since 2020, she has held the position of Chair for the High-Performance and Disruptive Computing in Remote Sensing (HDCRS) Working Group under the IEEE GRSS Earth Science Informatics Technical Committee (ESI TC).

Paul Magdon received the B.Sc., M.Sc., and Ph.D. degrees in forest science from the University of Göttingen, Göttingen, Germany, in 2005, 2008, and 2013, respectively.

Since 2022, he has been a Professor for geoinformation and forest planning with the University of Applied Sciences and Arts (HAWK), Göttingen. He is working on developing data- and sensor-driven environmental monitoring systems at the plot, regional, and national levels. His research interests include the development and optimization of remote sensing and GIS applications for monitoring natural resources, with the combination of field observations, geospatial analysis, and data from various sensors into coherent monitoring networks as one of his specific interests.

Anja Linstädter received the M.Sc. degree in biology from the University of Hamburg, Hamburg, Germany, in 1994, and the Ph.D. degree in ecology from the University of Cologne, Cologne, Germany, in 2001.

She is currently a Full Professor in biodiversity research and systematic botany with the Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. She has led several German and European-funded projects across African and European landscapes. As a Range Ecologist, her research interests include investigating the impacts and adaptations of grasslands under global changes.

Olena Dubovyk received the B.Sc. and M.Sc. degrees in geography from the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2007 and 2008, respectively, the joint M.Sc. degree in geo-information science and Earth observation for environmental modeling and management from the University of Southampton, Southampton, U.K., the Lund University, Lund, Sweden, and the University of Twente, Enschede, the Netherlands, in 2010, and the Ph.D. degree in remote sensing from the University of Bonn, Bonn, Germany, in 2014.

She is currently a Professor with the University of Hamburg, Hamburg, Germany. She is involved in the field of remote sensing, particularly in environmental monitoring and management. Her research interests include using Earth observation and geo-information systems for various applications including drought monitoring and assessment, as well as supporting policy and international cooperation (e.g., SDGs, Sendai indicators) and disaster risk reduction.

Javier Muro received the B.Sc. degree in environmental sciences from the University of Alcalá, Madrid, Spain, in 2009, the M.Sc. degree in ecology from the University of Turku, Turku, Finland, in 2014, and the Ph.D. degree in remote sensing from the University of Bonn, Bonn, Germany, in 2019.

Since 2022, he has been a Research Associate with the Institute of Farm Economics, Thünen Institute, Braunschweig, Germany. He has led multiple publications on modeling different land use processes in Latin America, Europe, and East Africa. His research interests include the use of Earth observation, machine learning, and data science to study the interactions between biodiversity, ecosystem services, and land management.