



Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions

Yuanyuan Yang¹, Zefang Shen¹, Andrew Bissett², and Raphael A. Viscarra Rossel¹

¹Soil and Landscape Science, School of Molecular and Life Sciences, Curtin University,
GPO Box U1987, Perth WA 6845, Australia

²CSIRO Oceans and Atmosphere, GPO BOX 1538, Hobart TAS 7001, Australia

Correspondence: Raphael A. Viscarra Rossel (r.viscarra-rossel@curtin.edu.au)

Received: 21 July 2021 – Discussion started: 13 September 2021

Revised: 7 February 2022 – Accepted: 19 February 2022 – Published: 25 March 2022

Abstract. Soil fungi play important roles in the functioning of ecosystems, but they are challenging to measure. Using a continental-scale dataset, we developed and evaluated a new method to estimate the relative abundance of the dominant phyla and diversity of fungi in Australian soil. The method relies on the development of spectrotransfer functions with state-of-the-art machine learning and uses publicly available data on soil and environmental proxies for edaphic, climatic, biotic and topographic factors, and visible–near infrared (vis–NIR) wavelengths, to estimate the relative abundances of Ascomycota, Basidiomycota, Glomeromycota, Mortierellomycota and Mucoromycota and community diversity measured with the abundance-based coverage estimator (ACE) index. The algorithms tested were partial least squares regression (PLSR), random forest (RF), Cubist, support vector machines (SVM), Gaussian process regression (GPR), extreme gradient boosting (XGBoost) and one-dimensional convolutional neural networks (1D-CNNs). The spectrotransfer functions were validated with a 10-fold cross-validation ($n = 577$). The 1D-CNNs outperformed the other algorithms and could explain between 45 % and 73 % of fungal relative abundance and diversity. The models were interpretable, and showed that soil nutrients, pH, bulk density, ecosystem water balance (a proxy for aridity) and net primary productivity were important predictors, as were specific vis–NIR wavelengths that correspond to organic functional groups, iron oxide and clay minerals. Estimates of the relative abundance for Mortierellomycota and Mucoromycota produced $R^2 \geq 0.60$, while estimates of the abundance of the Ascomycota and Basidiomycota produced R^2 values of 0.5 and 0.58 respectively. The spectrotransfer functions for the Glomeromycota and diversity were the poorest with R^2 values of 0.48 and 0.45 respectively. There is no doubt that the method provides estimates that are less accurate than more direct measurements with conventional molecular approaches. However, once the spectrotransfer functions are developed, they can be used with very little cost, and could serve to supplement the more expensive and laborious molecular approaches for a better understanding of soil fungal abundance and diversity under different agronomic and ecological settings.

1 Introduction

Soil fungi are important components of microbial communities, which inhabit dynamic soil environments. They play critical functional roles as decomposers, mutualists and pathogens (Li et al., 2019). They impact nutrient cycling and ecosystem services, such as soil carbon fixation, fertility and productivity (Vetrovsky et al., 2019; Delgadobaquerizo et al., 2016). Given the important functions that soil fungi perform, it is important to better characterise and understand their communities over large scales. However, data on soil fungi are few or largely unavailable because the measurement of soil fungi, which needs field sampling, followed by culture-based analysis or DNA sequencing, are laborious, time-consuming and costly. Using soil sensing technologies, such as spectroscopy together with molecular approaches, could greatly improve the utility of fungal inventory data (Hart et al., 2020).

Improvements in soil analytical methodologies provide an opportunity to increase sampling density for deriving a more detailed understanding of soil properties, their spatial variation and soil conditions, and to improve decision-making. Spectroscopic techniques, such as visible–near infrared (vis–NIR) spectroscopy, have been developed to provide rapid estimates of soil properties (Viscarra Rossel et al., 2016). Soil vis–NIR spectra are largely nonspecific because of the overlapping absorptions of soil constituents (Stenberg et al., 2010). Complex absorption patterns generated from soil constituents need to be mathematically extracted from the spectra and there are various methods that can be used to model soil properties with spectra. They include multivariate statistical methods such as partial least squares regression (PLSR), and machine learning with different algorithms, including neural networks (Viscarra Rossel and Behrens, 2010; Morellos et al., 2016; Liu et al., 2018; Tsakiridis et al., 2020; Shen and Viscarra Rossel, 2021). Thus, vis–NIR spectra can integrally characterise the soil's mineral–organic composition, and combined with multivariate modelling, soil spectroscopy provides a rapid and cost-efficient method for soil characterisation (Viscarra Rossel and Brus, 2018).

Although there are no vis–NIR absorptions that can be directly assigned to soil microbial communities or diversity, soil microbes are dependent on fundamental soil composition: its minerals, organic matter and water content. For example, they rely on organic matter for energy and on clay minerals and iron oxides for the supply of essential elements in order to grow (Müller, 2015). These organic and mineral properties are well represented and have a direct response in soil vis–NIR spectra (Stenberg et al., 2010). Therefore, vis–NIR spectra have been used to model various functional soil properties, such as soil organic carbon, cation exchange capacity, pH, clay content (Shi et al., 2015) and soil microbial communities (Davinic et al., 2012; Yang et al., 2019). For the latter, if microbial biomass is present in the soil organic mat-

ter, then the spectra might well detect their functional constituents.

There are studies that use environmental proxies (or covariates) at continental and global scales to model soil microbial properties using various methods, including linear regression and machine learning (Serna-Chavez et al., 2013; Griffiths et al., 2011; Vetrovsky et al., 2019; Yang et al., 2019; Delgadobaquerizo et al., 2018a). However, we found no published studies that used vis–NIR spectra or a combination of spectra with other soil and environmental covariates (i.e. spectrotransfer functions) to infer fungal abundance or diversity. In a previous study, Yang et al. (2019) showed that vis–NIR spectra combined with other soil and environmental data could estimate soil bacterial abundance and diversity. Here, our hypotheses are (i) spectroscopic models with machine learning can estimate soil fungal abundance and diversity at the continental scale, and (ii) spectrotransfer functions with additional predictors for capturing other soil and environmental properties that affect soil fungi will improve the accuracy of the estimates.

Thus, our objective is to develop and test the spectroscopic method for estimating soil fungal abundance and diversity over a large scale, and our aims are to

1. compare the modelling of fungal abundance and diversity with vis–NIR spectra only (spectroscopic models), with readily available soil and environmental data only (environmental models) and with the combined set of vis–NIR spectra and readily available soil and environmental data (spectrotransfer functions), and
2. test different statistical and machine learning algorithms for the modelling.

2 Methods

2.1 Soil sampling and laboratory analyses

We used 577 soil samples from the Biomes of Australian Soil Environments (BASE) project (Bissett et al., 2016). In that project, sampling was undertaken from in that supports diverse plant communities across Australia. The sampling was carried out during the growing season when hydrothermal conditions are most conducive to typical plant growth. In the higher rainfall forested regions of the continent, the soil samples were collected mostly in spring and summer from September to February. In the shrublands and grasslands of the semi-arid and arid interior, soil samples were collected in spring from September to November. In the transitional zone between the south-east coast and the more arid interior, soil samples were collected in mainly autumn from March to May. Samples came from two soil depths (0–0.1 and 0.2–0.3 m), covering five typical Australian ecosystem types comprising cropland, forest, grassland, shrubland and woodland (Fig. 1a). Woodlands in Australia represent ecosystems which contain widely spaced trees, the crowns of which do

not touch. Woodlands consist of areas with fewer and more scattered trees than forests. In temperate Australia, woodlands are mainly dominated by Eucalyptus species. Temperate woodlands occur predominantly in regions with a mean annual rainfall of between 250 and 800 mm, forming a transitional zone between the higher rainfall forested margins of the continent and the shrub and grasslands of the arid interior. Each sample was partitioned into subsamples for DNA sequencing (see below) and air-dried and crushed to a particle size of ≤ 2 mm for physicochemical analyses. The soil properties analysed were total organic carbon and soil nutrients (e.g. ammonium, nitrate, phosphorus, potassium), pH, exchangeable cations (aluminium, sodium, magnesium, calcium) and texture (sand, silt, clay). The methods are described in Bissett et al. (2016). Subsamples of the ≤ 2 mm portions were used for the spectroscopic analysis (see below).

2.2 Derivation of fungal abundance and diversity

The methods for DNA extraction and sequencing are detailed in Bissett et al. (2016). Briefly, the soil DNA was extracted in triplicate following methods used in the Earth Microbiome Project (<http://www.earthmicrobiome.org/emp-standard-protocols/dna-extraction-protocol/>, last access: 22 March 2022). Sequencing occurred with an Illumina MiSEQ, which is described in the BASE protocols (<https://ccgapps.com.au/bpa-metadata/base/information>, last access: 22 March 2022). Summarising, amplicons targeting the fungal ITS region were prepared and sequenced for each sample. The ITS amplicons were sequenced using 300 bp paired end sequencing. ITS1 regions were extracted using ITSx (Bengtsson-Palme et al., 2013). Sequences comprising full and partial ITS1 regions were passed to the operational taxonomic units (OTUs) selection and assigning workflow (Bissett et al., 2016), which followed guidelines described in the BASE protocols (<https://ccgapps.com.au/bpa-metadata/base/information>, last access: 22 March 2022). These are based on the most current version of UNITE database (version 8.2, updated 15 January 2020) for molecular identification of fungi (Nilsson et al., 2018). We used the final sample-by-OTU data matrix and annotated taxonomy file for the analyses of fungal diversity and composition.

To eliminate bias in the diversity comparison caused by unbalanced sequencing, samples were resampled at the same sequencing depth using functions of the RAM library in R software (R Core Team, 2014). The BASE dataset sought to produce as many sequences as resources allow with a minimum sequencing number of 10 000 per sample. Here, 11 000 sequences (the median number of sequences in the samples) were used as the resampling depth, because the majority of samples only had this amount of sequences, but also because the rarefaction curves started to flatten out for all 577 samples at this sequencing depth. This suggested that

the sequencing number was sufficient (Fig. S2 in the Supplement). To quantify community diversity, we then calculated the abundance-based coverage estimator (ACE) index (Lozupone and Knight, 2008) from the resampled sample-by-OTU matrix. The relative abundance of fungal phyla were then determined using the ratio of the number of sequences classified at each phylum to the total number of sequences of each sample.

2.3 Soil visible–near-infrared spectroscopy

We measured the diffuse reflectance spectra of all air-dried ≤ 2 mm soil samples with the Labspec[®] vis–NIR spectrometer (Malvern Panalytical, Boulder, Colorado, USA) following the protocols described in Viscarra Rossel et al. (2016). The spectral range of the spectrometer is 350 to 2500 nm. Due to the low signal-to-noise ratio at the start and end of each spectrum, for our analysis, we kept only spectra in the range between 380 and 2450 nm. As the spectra are highly collinear, to reduce redundancy in the data, we resampled them to a resolution of 10 nm. The measurements were performed with the instrument's high intensity contact probe (Malvern Panalytical, Boulder, Colorado, USA), and a Spectralon[®] white reference panel was used for calibration once every 10 measurements.

For the modelling and interpretation, we first transformed the reflectance (R) spectra to apparent absorbance, using $A = \log_{10}(1/R)$, and then used the Savitzky–Golay method with a window of size 7, a quadratic polynomial and first derivative method (Savitzky and Golay, 1964), in order to remove baseline effects and to improve the signal-to-noise ratio. To visualise the spectra, we further fitted each reflectance (R) spectrum with a convex hull and computed the deviations from the hull (Clark and Roush, 1984). These continuum removed (CR) spectra help to visualise the characteristic absorptions more clearly than the Savitzky–Golay first derivatives (SG1Der) absorbance spectra.

2.4 Modelling soil fungal abundance and diversity

We developed spectroscopic models, environmental models and spectrotransfer functions for estimating soil fungal abundance and diversity (see below). The spectroscopic models used only the vis–NIR spectra, the environmental models used only the publicly available soil and environmental data that represent the soil forming factors soil, climate, vegetation, terrain and parent material (Jenny, 1994) and the spectrotransfer functions used the vis–NIR spectra together with soil and environmental data.

We assembled a set of readily available soil and environmental maps that represented climate, terrain, vegetation and parent material. To relate these covariates to the fungal data, we extracted values from these maps using the geographic coordinates of the sample set. The soil property data came from Australia-wide fine spatial resolution (90 × 90 m)

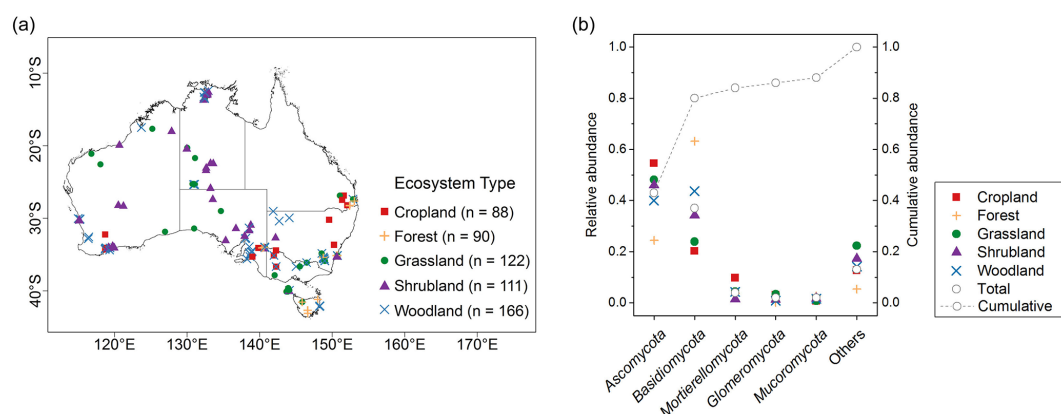


Figure 1. (a) Sampling sites and the range of ecosystem types across Australia. (b) The mean relative abundances of dominant fungal phyla and unclassified “Others” taxa in five ecosystem types. Individual abundance of each phylum and their cumulative abundance were shown in the graph.

digital soil maps of total organic carbon, total nitrogen, total phosphorus, bulk density, effective cation exchange capacity, available water capacity, pH and soil texture (sand, silt and clay; Viscarra Rossel et al., 2015), as well as maps of the clay minerals kaolinite, illite and smectite (Viscarra Rossel, 2011). To represent climate, we used data on mean annual temperature (MAT), mean annual precipitation (MAP), solar radiation and evapotranspiration (Xu and Hutchinson, 2011) and the Prescott index (PI; Prescott, 1950), which is calculated as the ratio of precipitation to evapotranspiration. To capture functional landscape characteristics, we used a digital elevation model (DEM) from the 3 arcsec Shuttle Radar Topography Mission (SRTM) and derived terrain attributes (Gallant et al., 2011). To represent vegetation, we used data on net primary productivity (NPP; Haverd et al., 2013) and on the fraction of photosynthetically active radiation intercepted by the sunlit canopy of the evergreen (Fpar-e) and woody (Fpar-r) vegetation (Donohue et al., 2009). To represent parent material, we used gamma radiometrics, which comprises data on potassium, uranium and thorium (Minty et al., 2009). Supplement Table S2 lists these data and their main characteristics.

The spectra and the covariates were centred and scaled before the modelling of fungal abundance and diversity. The algorithms that we tested were PLSR (Wold et al., 2001), Gaussian process regression (GPR; Rasmussen and Williams, 2005), support vector machines (SVM; Suykens et al., 2002), random forest (RF; Breiman, 2001), Cubist (Quinlan, 1992), extreme gradient boosting (XGBoost; Friedman, 2001) and optimised one-dimensional convolutional neural networks (1D-CNNs; Shen and Viscarra Rossel, 2021). The algorithms and their implementation are described in the Supplement linked to this article.

The predictability of the spectroscopic models and the spectrotransfer functions were assessed using 10-fold cross-validation. We evaluated the estimates using the Nash–

Sutcliffe model efficiency, otherwise known as the coefficient of determination (R^2), which represent the fraction of the explained variance based on the 1 : 1 line of estimated versus measured values (Janssen and Heuberger, 1995). The R^2 was computed as $1 - \text{RSS}/\text{TSS}$, where RSS is the residual sum of squares and TSS is the total sum of squares. The root mean squared error (RMSE) measures inaccuracy, the standard deviation of the error (SDE) measures imprecision and the mean error (ME) measures bias (Viscarra Rossel and McBratney, 1998). Inaccuracy (RMSE) embraces both the bias (ME) and the imprecision (SDE; Viscarra Rossel and McBratney, 1998). Their relationship is given by $\text{RMSE}^2 = \text{ME}^2 + \text{SDE}^2$.

To interpret the models, we calculated their variable importance as follows. For the PLSR, GPR, SVM, Cubist, RF and XGBoost models, variable importance was calculated using the varImp function in the caret library (Kuhn et al., 2008) of the software R. To calculate the variable importance of the CNN models, we used permutation variable importance. In our case, we ran 1000 permutations and measured the decrease in RMSE after a predictor was permuted (randomly rearranged). The permutation breaks the relationship between the predictor and the response variables, and a reduction in RMSE indicates how much the model depends on the particular predictor. An advantage of this approach is that it can be applied on any estimator and does not require re-training the model (Breiman, 2001; Fisher et al., 2019). In order to compare the importance between different fungal phyla and diversity, we scaled the importance values between 0 and 1. In the results, we only report the variable importance of the model that performed best.

3 Results

In total, more than 60 million quality-filtered sequences in the whole dataset were obtained, with an average of 107 310

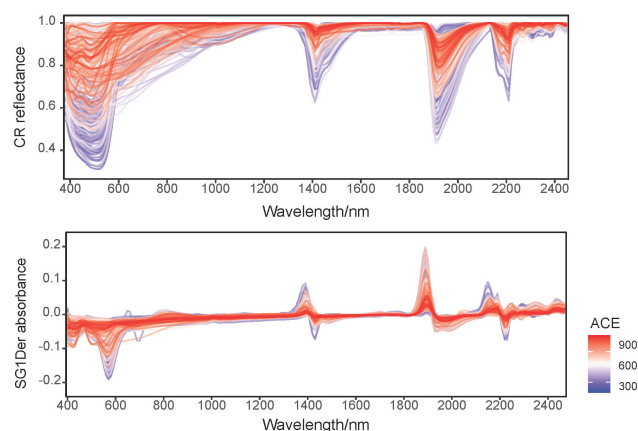


Figure 2. Continuum removed (CR) spectra and the Savitzky–Golay first derivatives (SG1Der) absorbance spectra curves coloured by fungal ACE diversity.

sequences per sample. When we clustered the sequences at 97 % similarity level, 202 200 OTUs were detected. Each sample had an average of 666 OTUs. In total, 16 phyla were identified; 5 dominant phyla, with relative abundance of approximately > 2 %, were present in most soils. This represented nearly 88 % of the sequence number. The relative abundance of fungal phyla varied across ecosystem types (Fig. 1b).

Ascomycota (mean 0.43, SD 0.21) was the most abundant phylum, followed by Basidiomycota (mean 0.37, SD 0.24; Table 1). Dominant fungal phyla showed a high degree of variability, with an averaging 83 % coefficient of variation (CV). The ACE index showed a wide range from 81 to 1823 (mean 563, SD 315). The rich soil biodiversity of the data resulted from extensive soil sampling taken from diverse vegetation, soils and climates across Australia.

Figure 2 shows the CR reflectance and SG1Der absorbance spectra with the characteristic absorption features. Soil with different fungal diversity show variations in absorptions, particularly around those that are due to iron oxides (400–800 nm), minerals (around 1400, 1900 and 2200 nm) and organic compounds (throughout the vis–NIR spectrum; Stenberg et al., 2010). Soil with lower fungal diversity showed a more pronounced absorbance around 600 nm as shown in Fig. 2. In our study, the soil with lower fungal diversity mainly came from central and western Australia. In these areas, soil is subjected to intense weathering regimes and can accumulate large quantities of iron oxides (total soil Fe_2O_3 larger than 10 %) in surficial environments, and strongly absorbed in the visible region (Viscarra Rossel et al., 2010). These highly iron-rich lateritic soils occur with acidic pH and high H_2O and Al activity, which has been shown to be not conducive to the development of fungal diversity (Viscarra Rossel et al., 2010).

3.1 Modelling

With the different algorithms, the spectroscopic models (i.e. with only the vis–NIR spectra) could explain 9 %–45 % of the variation in fungal phyla relative abundance and diversity. Spectroscopic models of the Glomeromycota were the least successful, with R^2 values ranging from 0.09 using SVM to 0.30 using 1D-CNN, while those of the Mortierellomycota produced the largest R^2 values, ranging from 0.32 using XGBoost to 0.45 using 1D-CNN (Fig. 3). The models of diversity had R^2 values ranging from 0.14 with PLSR to 0.35 using 1D-CNN.

The models derived with the readily available soil and environment data could explain 14 %–60 % of the variation in fungal phyla relative abundance and diversity with the different algorithms. These environmental models generally performed better than the spectroscopic models, with an average 10 % additional variance explained.

Combining the vis–NIR spectra and soil and environmental data further improved the models and their explanatory power. The spectrotransfer functions (i.e. with the combined set of vis–NIR spectra and other soil and environmental data) performed, on average, 20 % better than the spectroscopic models and 10 % better than environmental models. Depending on the algorithm used, they could explain between 17 % and 73 % of the variation in fungal phyla relative abundance and diversity (Fig. 3). The spectrotransfer functions of Glomeromycota produced R^2 values, ranging from 0.17 using PLSR to 0.48 using 1D-CNN. The spectrotransfer functions of the Mortierellomycota and Mucoromycota produced the largest R^2 values ranging from 0.51 to 0.73 (Fig. 3).

Generally, PLSR and GPR were the least successful methods, while SVM, RF, Cubist and XGBoost were similarly successful at estimating fungal phyla relative abundance and diversity (Fig. 3). The 1D-CNN spectrotransfer functions were 13 %–31 % more successful compared to other machine learning methods as they could explain between 45 % and 73 % of the variation in fungal relative abundance and diversity (Fig. 3).

3.2 1D-CNNs spectrotransfer functions

The final architecture and optimised hyperparameters of the 1D-CNNs are given in Supplement Table S3. As deep learning models are dataset dependent, the optimisation returned a different architecture for each response variable. Overall, the 1D-CNNs used simple architecture with less than four convolutional layers (Supplement Table S3). Scatter plots of the measured versus estimated values of relative abundance and diversity using 1D-CNNs spectrotransfer functions and their validation statistics are shown in Fig. 4. Estimates of the relative abundance of Mortierellomycota and Mucoromycota produced R^2 values ≥ 0.60 , while estimates of Ascomycota and Basidiomycota produced values of $0.5 \leq R^2 < 0.6$. Estimates of Glomeromycota and ACE produced

Table 1. Descriptive statistics of relative abundance of dominant phyla and community diversity ($n = 577$).

Variables	Mean	Median	SD	Range	Coeff. var. (%)
Abundance					
<i>Ascomycota</i>	0.43	0.42	0.21	0.04–0.98	49
<i>Basidiomycota</i>	0.37	0.32	0.24	0.01–0.92	65
<i>Mortierellomycota</i>	0.04	0.02	0.04	0.00–0.36	100
<i>Glomeromycota</i>	0.02	0.01	0.01	0.00–0.41	50
<i>Mucoromycota</i>	0.02	0.01	0.03	0.00–0.55	150
Diversity					
ACE	563	503	315	81–1823	56

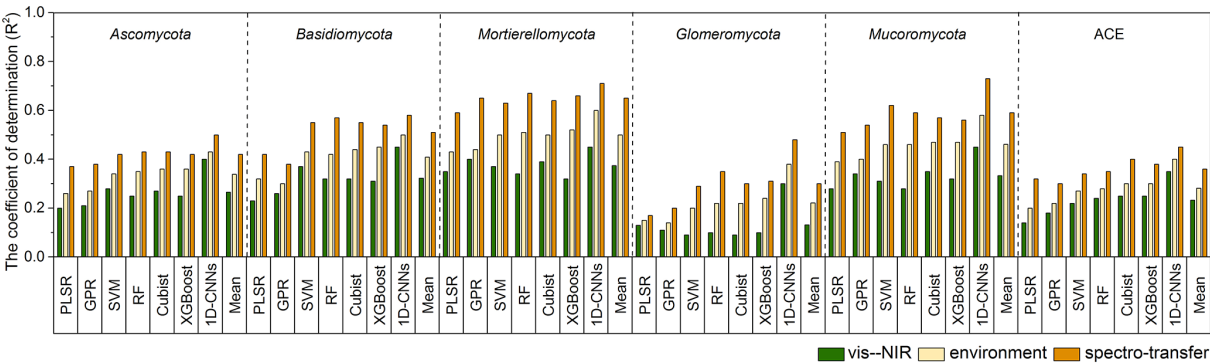


Figure 3. Coefficient of determination (R^2) for the vis–NIR spectroscopic models, soil and environmental models and the spectrotransfer functions that used a combined set of the vis–NIR and readily available soil and environmental covariates used to estimate soil fungal phyla abundance and diversity ($n = 577$). The different statistical and machine learning methods were PLSR, GPR, SVM, RF, Cubist, XGBoost and optimised 1D-CNNs.

values of $0.4 \leq R^2 < 0.5$. The estimates were relatively unbiased (small ME), although generally small values were overestimated and large values were underestimated (Fig. 4). Imprecision contributed to the majority of the RMSE. The imprecision of our estimates was a result of the absence of repeated sampling and the high adaptability of soil fungi to the wide range of environments.

The important variables in the 1D-CNNs spectrotransfer functions of phyla relative abundance and diversity were vis–NIR wavelengths representing organic matter, iron oxide and clay minerals (Fig. 5).

The identified wavelengths mostly coincided with absorptions that are related to carbon functional groups found in organic matter, including C–H, N–H and C–O, with a smaller number of wavelengths coinciding with those that are related to clay minerals and iron oxides (Table 2). The organic functional groups, C–H alkyl and methyls, N–H of amines and C–O of carbohydrates, which might indicate the presence of relatively labile forms of carbon, were important in the models of fungal phyla but not of ACE diversity. The C=O of amides and carboxylic acids, which represent stable forms of carbon were not as important in modelling (Fig. 5). Other wavelengths that represent iron oxides and clay minerals were also

important in the models, indicating the different ecological niches and physiological characteristics (Table 2).

Other soil properties, such as total organic carbon and pH were important variables in the spectrotransfer functions of *Ascomycota* and *Basidiomycota*, and fungal diversity. Total organic carbon and total nitrogen were important in the spectrotransfer functions of *Mortierellomycota* and *Mucoromycota* and bulk density was important in the spectrotransfer functions of *Glomeromycota*, *Ascomycota* and ACE diversity (Fig. 5). As well as soil properties, climatic factors such as PI and PET, and vegetation, represented by Fpar-e and NPP were also important in the modelling of fungal phyla relative abundance and community diversity. The variables that we used to represent terrain and the parent material exerted less influence in the models (Fig. 5).

4 Discussion

Soil fungi play essential and diverse functional roles in ecosystems. However, they are challenging to investigate due to laborious, time-consuming and costly field sampling, and laboratory analysis. We show that spectrotransfer functions with readily accessible vis–NIR spectra and publicly avail-

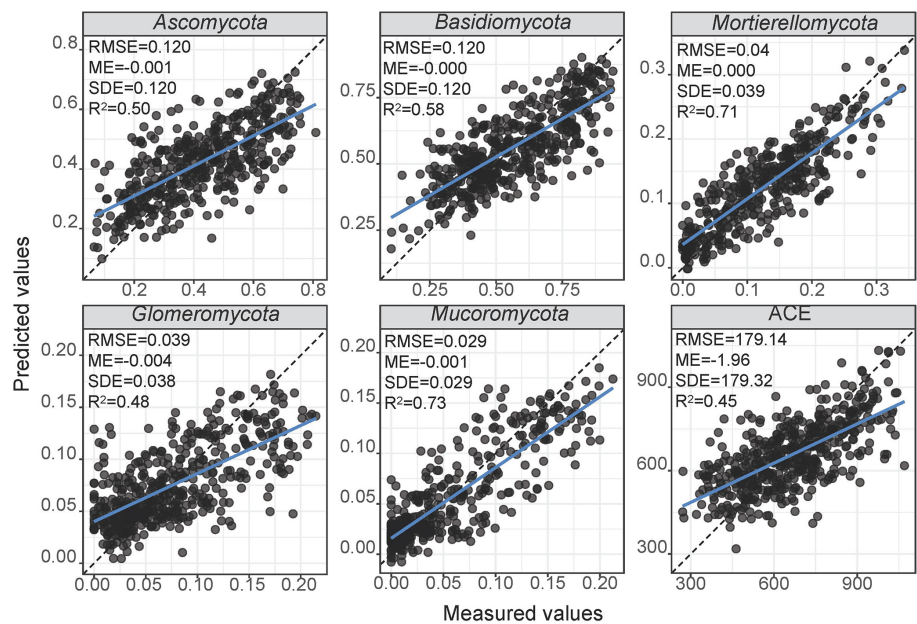


Figure 4. Performance of the CNN spectrotransfer functions for estimation of the relative abundance of dominant fungal phyla and diversity index. The spectrotransfer functions used vis–NIR spectra with other publicly available data on soil environmental variables. The plots show measured vs. estimated values using a 10-fold cross validation. The grey points represent no overlap with any other points, and the black points represent at least two points that overlap.

Table 2. Absorption wavelength assignment (in nanometres) for the most important vis–NIR wavelengths in the 1D-CNN models. The assignment of vis–NIR absorptions are based on Viscarra Rossel and Behrens (2010) and Stenberg et al. (2010).

	ACE	Ascomycota	Basidiomycota	Mortierellomycota	Glomeromycota	Mucoromycota
Iron oxides	390	390	410, 460			
Clay minerals	2190, 2240	1330, 2190, 2210	1330, 2140	1360, 2140		1330, 2150
Organics						
C-H of aromatics		1630, 1650				
N-H of amine		2070, 2090, 2110	1010	2060	780, 2030	2060
C-H of alkyl asymmetric-symmetric doublet			890, 1290	1250, 1280	1740	1270, 1280
C=O of carboxylic acids						
C=O of amides						
C-H of aliphatics						
C-H of methyls		1840, 2440	1770, 1800, 1810 1830, 2450		1880	
C-OH of phenolics						
C-O of carbohydrates	2260		2410, 2290	2300		2300

able soil and environmental data could variably estimate (R^2 ranging from 0.45 to 0.73) soil fungal abundance and diversity measured with ITS gene metabarcoding. The spectrotransfer functions explained less than 60 % of the variance in the two dominant phyla, the Ascomycota and Basid-

iomycota, representing 80 % of the total fungal relative abundance. In comparison, the spectrotransfer functions could explain more than 70 % of the variance in the Mortierellomycota and Mucoromycota, which were less abundant in soil. The reason for the different predictability might be the coarse

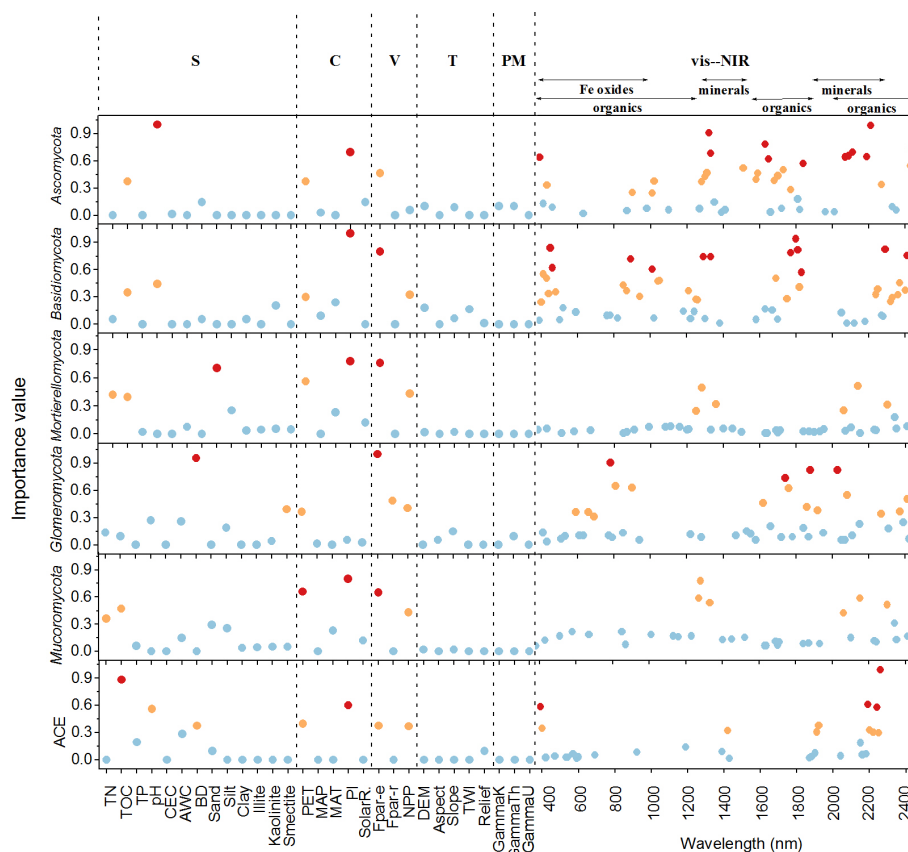


Figure 5. Important predictors of relative abundance of fungal phyla and diversity index measured by the variable importance of the 1D-CNNs spectrotransfer functions ($n = 577$) derived with publicly accessible data that represent soil (S), climate (C), vegetation (V), terrain (T), parent material (PM) and vis–NIR spectra. The dots in red, orange and blue indicated the most, medium and least important levels. The importance value for the majority of wavelengths were low and close to zero, thus these wavelengths were not shown to make the figure clearer.

phylum-level identity. Compared with the Mortierellomycota and Mucoromycota, the Ascomycota and Basidiomycota are more complex phylogenetic classifications and consist of more diverse taxa with different phenotypic traits. These taxa have distinct ecological functions and environmental preferences, which might have reduced the predictability of their relative abundance at the phylum level. Classifying taxa with similar habitat preferences or studying at a finer taxonomic resolution might provide better predictability and understanding of soil fungal communities. The spectrotransfer function for the ACE index could only explain around 50 % of the variance in diversity. The reason might be that local geography, environmental conditions and difficult-to-proxy long-term natural selection and evolution affect community diversity.

The general concept of using proxies has been used in other studies to attempt more rapid estimation of microbial properties towards the diagnosis of soil quality. For example, Horrigue et al. (2016) developed a statistical predictive model of soil microbial biomass according to environmental parameters including soil physicochemical and climatic

characteristics across France. Their model ($R^2 = 0.67$) provided a reference value of microbial biomass for a given pedoclimatic condition to enable rapid diagnosis of soil quality across France. Other similar studies exist, for example Griffiths et al. (2016), who focused on the estimation of bacterial community structure and diversity at the Europe scale. ITS gene metabarcoding analyses are expensive, laborious and require specialised laboratories and methods, while spectroscopic measurements are faster, less expensive and soil–environmental data are more readily available. When many measures are needed, for example, to assess, characterise and improve our understanding of soil fungal communities and their associated functions at different scales, the approach could complement molecular techniques (Hart et al., 2020). For instance, to characterise spatial variation (i.e. for mapping), one needs many measurements that would be too expensive with only metabarcoding. In this case, estimates with the spectrotransfer functions ($R^2 = 0.45$ – 0.73) could complement the metabarcoding analysis to represent the variability present better. As a whole, the spatial characterisation will be more accurate than when only taking a few very precise

measurements. This is the rationale for the characterisation of soil properties in space and time with sensing (Viscarra Rossel et al., 2011). The soil covariates in the model are derived from digital soil maps and not from measured soil samples. The reason is that using measured data would increase the cost of the approach significantly, making the approach less attractive. We note that the uncertainty in the spectrotransfer estimates caused by using the digital soil map predictors will propagate to the spectrotransfer functions and thereby lower the precision of the estimates.

We do not expect that the spectrotransfer method will produce estimates that are as accurate as more conventional molecular methods, even with further improvements in modelling and better covariates. This is because we understand that the modelling of living organisms is dynamic and hugely complex. Fungi vary over space and time (Duan et al., 2018), often showing that their prevalence in different habitats differs seasonally (Talley et al., 2002). The inconsistent correlations of fungi with climate and plant hosts observed in various ecosystems may be due to seasonal variation and spatial heterogeneity across single time point studies (Kivlin and Hawkes, 2016). Thus, temporal sampling is needed to capture the seasonal dynamics of microbial communities.

Our research uses soil fungal measurements at a single point in time and there are likely to be many undetermined controlling factors, including seasonal variability and complex biological interactions. Despite this drawback, our approach allows us to infer the distribution of soil fungal communities and diversity more simply and at a lesser cost, to help better understand the diversity and biogeography of soil fungi in different habitats. Thus, our approach shows promise and could complement molecular methods. We hope that our study will stimulate further research towards achieving more widespread characterisation of fungal abundance and diversity, which will help to deepen our understanding of fungal biology, biogeography and their environmental controls. Different spectra, new sensing technologies and improved methods could also improve the spectrotransfer approach.

Out of the seven statistical and machine learning models tested, the optimised 1D-CNNs were the most successful at estimating fungal phyla relative abundance and diversity, consistently producing the highest cross-validation R^2 values. The reason might be that the 1D-CNNs can automatically “learn” the non-linear and complex relations between the soil fungal variables and the covariables. The models extract large features during convolution and adjust the weights of each covariate during the model iterations, which are also back-propagated (Lecun et al., 2015). Although 1D-CNNs have been used for the spectroscopy modelling of soil physicochemical properties (Ng et al., 2019; Tsakiridis et al., 2020; Shen and Viscarra Rossel, 2021), to our best knowledge, this present study is the first to develop spectrotransfer functions for estimating soil fungal abundance and diversity.

Our results shown that the 1D-CNN spectroscopic models (with only vis–NIR spectra) could explain, on average, 40 %

of the variation in the relative abundance of fungal phyla and community diversity (R^2 values of 0.30–0.45). This is because these spectra characterise the soil’s organic and mineral composition, which serves to supply energy and the elements that fungi use to promote vital activities (Müller, 2015). Microbial activities are closely associated with the types and amounts of organic matter and our results indicate that the most important vis–NIR wavelengths in the modelling of fungal relative abundance and community diversity corresponds to functional groups in the different types of organic compounds in the soils (Viscarra Rossel and Hicks, 2015) (Fig. 5 and Table S2 in Supplement).

The 1D-CNN spectrotransfer functions (with vis–NIR spectra and other soil and environmental data) improved the modelling. This suggests that other variables that represent climate, soil nutrients, pH and vegetation are important predictors of fungal growth. Their use in the spectrotransfer functions provided additional and supplementary information for the modelling. On average, these models could explain 60 % of the variation of fungal phyla relative abundance and diversity (R^2 values of 0.45–0.73).

The soil organic and mineral composition, represented by the vis–NIR spectra, were the most important predictors in the models for fungal relative abundance and community diversity. Additionally, total organic carbon and pH were important predictors of fungal diversity and the relative abundance of Ascomycota and Basidiomycota. Although most soil fungi do not require strict pH ranges for habitation and growth (Rousk et al., 2009; Zhao and Shen, 2018), some basophilic or acidophilic fungi are sensitive to changes in pH (Gai et al., 2006), and saprophytic fungi are thought to be more sensitive to soil pH compared to other fungi (Kivlin and Hawkes, 2016). Soil bulk density was an important predictor of fungal diversity and the relative abundance of Glomeromycota. Many fungi, including those that form arbuscular mycorrhiza, such as Glomeromycota, infect plants roots achieving mutualistic symbiosis (Schubler et al., 2001). Denser soil bulk density could reduce the availability of soil nutrients and water, leading to poor development of plant roots and a smaller infection rate for the symbiosis. PI and evapotranspiration were the most important climatic predictors of fungal abundance and diversity in the models. PI represents the soil–water balance, which has been shown to affect soil microbial growth in various studies (Bachar et al., 2010; Blankinship et al., 2011; Maestre et al., 2015; Delgadobaquerizo et al., 2018b) because soil–water stress could strongly restrict microbial activity and distribution by controlling the availability of soil nutrients, pH and oxygen (Delgadobaquerizo et al., 2018b). NPP and Fpar-e were important predictors of fungal diversity and the relative abundance of the five dominant phyla. Larger values of NPP and Fpar-e occur due to greater biomass production and thus more accumulation of litter and coarse organic matter in soil. Soil fungi are some of the decomposers of litter and soil organic matter, including cellulose and lignin, which are often resis-

tant to bacterial decomposition (Treseder and Lennon, 2015; Nicolas et al., 2019).

5 Conclusions

Our study contributes to the development of methods that could complement, not replace, molecular approaches for characterising and better understanding the diversity and biogeography of soil fungi. We have shown that deep learning spectrotransfer functions are a promising new method for estimating soil fungal communities' relative abundance and diversity. The optimised 1D-CNNs outperformed the six other machine learning algorithms tested for estimating the relative abundance of fungal phyla and diversity. The spectrotransfer functions (with vis–NIR spectra and soil and environmental data) produced more accurate estimates (R^2 0.45–0.73) than the spectroscopic models (only vis–NIR spectra; R^2 0.36–0.55) and models with only the soil and environmental data (R^2 0.38–0.60). In addition to the soil organic and mineral composition, represented by vis–NIR spectra, other edaphic, climatic and biotic factors including soil nutrients, pH, bulk density, potential evapotranspiration, the soil–water balance and net primary productivity were important predictors in the modelling. We hope that our study will provide food-for-thought for further research on the measurement and estimation of fungal abundance and diversity. We believe that improvements will be possible as new technologies and methodologies develop that will also help to deepen our understanding of fungal biology and biogeography.

Code availability. The code used for the analyses presented in this work is available from the corresponding author on reasonable request.

Data availability. The Biomes of Australian Soil Environments (BASE) data are available online from the BioPlatforms Australia data portal <https://doi.org/10.4227/71/561c9bc670099> (BioPlatforms Australia, 2022). The spectra are deposited at <https://doi.org/10.5281/zenodo.6265730> (Bissett and Viscarra Rossel, 2022).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/soil-8-223-2022-supplement>.

Author contributions. RAVR conceived the research and designed the study. YY and ZS carried out the experiments. YY and RAVR drafted the manuscript. RAVR edited the manuscript with input from all authors. AB derived the data on fungal relative abundance and diversity and edited the manuscript. All authors discussed and interpreted the results and produced the final manuscript.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We acknowledge the contribution of the Biomes of Australian Soil Environments (BASE) consortium in the generation of data used in this publication. The BASE project is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS). We thank Shuo Li for his assistance with the spectroscopic measurements.

Financial support. This research has been supported by the Research Office at Curtin University, and the development of the convolutional neural networks was supported by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

Review statement. This paper was edited by Nicolas P. A. Saby and reviewed by Lauric Cécillon and one anonymous referee.

References

- Bachar, A., Al-Ashhab, A., Soares, M. I., Sklarz, M. Y., Angel, R., Ungar, E. D., and Gillor, O.: Soil microbial abundance and diversity along a low precipitation gradient, *Microbial. Ecol.*, 60, 453–461, 2010.
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., Veldre, V., and Nilsson, R. H.: Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data, *Meth. Ecol. Evol.*, 4, 914–919, <https://doi.org/10.1111/2041-210X.12073>, 2013.
- BioPlatforms Australia: Biomes of Australian Soil Environments (BASE), BioPlatforms Australia [data set], <https://doi.org/10.4227/71/561c9bc670099>, last access: 22 March 2022.
- Bissett, A. and Viscarra Rossel, R.: Soil visible–near infrared (vis–NIR) spectra for the Biomes of Australian Soil Environments (BASE) soil microbial diversity database (1.0), Zenodo [data set], <https://doi.org/10.5281/zenodo.6265730>, 2022.
- Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., Breed, M. F., Brown, B., Brown, M. V., Bruggen, J., Byrne, M., Caddy-Retalic, S., Carmody, B., Coates, D. J., Correa, C., Ferrari, B. C., Gupta, V. V. S. R., Hamonts, K., Haslem, A., Hugenholtz, P., Karan, M., Koval, J., Lowe, A. J., Macdonald, S., McGrath, L., Martin, D., Matt, M., North, K. I., Paungfoo-Lonhienne, C., Pendall, E., Phillips, L., Pirzl,

- R., Powell, J. R., Ragan, M. A., Schmidt, S., Seymour, N., Snape, I., Stephen, J. R., Stevens, M., Tinning, M., Williams, K., Yeoh, Y. K., Zammit, C. M., and Young, A.: Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database, *GigaScience*, 5, s13742-016-0126-5, <https://doi.org/10.1186/s13742-016-0126-5>, 2016.
- Blankinship, J. C., Niklaus, P. A., and Hungate, B. A.: A meta-analysis of responses of soil biota to global change, *Oecologia*, 165, 553–565, 2011.
- Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, 2001.
- Clark, R. N. and Roush, T. L.: Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications, *J. Geophys. Res.*, 89, 6329–6340, 1984.
- Davinic, M., Fultz, L. M., Acosta-Martinez, V., Calderón, F. J., Cox, S. B., Dowd, S. E., Allen, V. G., Zak, J. C., and Moore-Kucera, J.: Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition, *Soil Biol. Biochem.*, 46, 63–72, 2012.
- Delgadobaquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., Berdugo, M., Campbell, C. D., and Singh, B. K.: Microbial diversity drives multifunctionality in terrestrial ecosystems, *Nat. Commun.*, 7, 10541–10541, 2016.
- Delgadobaquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-gonzalez, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., and Fierer, N.: A global atlas of the dominant bacteria found in soil, *Science*, 359, 320–325, 2018a.
- Delgadobaquerizo, M., Reith, F., Dennis, P. G., Hamonts, K., Powell, J. R., Young, A. G., Singh, B. K., and Bissett, A.: Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere, *Ecology*, 99, 583–596, 2018b.
- Donohue, R. J., McVicar, T., and Roderick, M. L.: Climate-related trends in Australian vegetation cover as inferred from satellite observations, 1981–2006, *Glob. Change Biol.*, 15, 1025–1039, 2009.
- Duan, Y. B., Xie, N. D., Song, Z. Q., Ward, C. S., Yung, C. M., Hunt, D. E., Johnson, Z. I., and Wang, G. Y.: A high-resolution time series reveals distinct seasonal patterns of planktonic fungi at a temperate coastal ocean site (Beaufort, North Carolina, USA), *Appl. Environ. Microbiol.*, 84, e00967–18, <https://doi.org/10.1128/AEM.00967-18>, 2018.
- Fisher, A., Rudin, C., and Dominici, F.: All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously, *J. Mach. Learn. Res.*, 20, 1–81, 2019.
- Friedman, J. H.: Greedy function approximation: A gradient boosting machine, *Ann. Stat.*, 29, 1189–1232, 2001.
- Gai, J., Christie, P., Feng, G., and Li, X. L.: Twenty years of research on community composition and species distribution of arbuscular mycorrhizal fungi in China: a review, *Mycorrhiza*, 16, 229–239, 2006.
- Gallant, J., Wilson, N., Dowling, T., Read, A., and Inskeep, C.: SRTM-derived 1 second digital elevation models version 1.0, Geoscience Australia: Canberra, ACT, 2011.
- Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whiteley, A. S.: The bacterial biogeography of British soils, *Environ. Microbiol.*, 13, 1642–1654, 2011.
- Griffiths, R. I., Thomson, B. C., Plassart, P., Gweon, H. S., D., S., Creamer, R. E., Lemanceau, P., and Bailey, M. J.: Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets, *Appl. Soil Ecol.*, 97, 61–68, 2016.
- Hart, M. M., Cross, A. T., D'Agui, H. M., Dixon, K. W., Van der Heyde, M., Mickan, B., Horst, C., Grez, B. M., Valliere, J. M., Viscarra Rossel, R. A., Whiteley, A., Wong, W. S., Zhong, H., and Nevill, P.: Examining assumptions of soil microbial ecology in the monitoring of ecological restoration, *Ecol. Solut. Evid.*, 1, e12031, <https://doi.org/10.1002/2688-8319.12031>, 2020.
- Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Pickett-Heaps, C., Roxburgh, S. H., van Gorsel, E., Viscarra Rossel, R. A., and Wang, Z.: Multiple observation types reduce uncertainty in Australia's terrestrial carbon and water cycles, *Biogeosciences*, 10, 2011–2040, <https://doi.org/10.5194/bg-10-2011-2013>, 2013.
- Horrigue, W., Dequiedt, S., Chemidlin Prévost-Bouré, N., Jolivet, C., Saby, N. P., Arrouays, D., Bispo, A., Maron, P., and Ranjard, L.: Predictive model of soil molecular microbial biomass, *Ecol. Ind.*, 64, 203–211, 2016.
- Janssen, P. and Heuberger, P.: Calibration of process-oriented models, *Ecol. Modell.*, 83, 55–66, 1995.
- Jenny, H.: Factors of Soil Formation: A System of Quantitative Pedology, Courier Corporation, New York, ISBN 978-0-48668-128-3, 1994.
- Kivlin, S. N. and Hawkes, C. V.: Tree species, spatial heterogeneity, and seasonality drive soil fungal abundance, richness, and composition in Neotropical rainforests, *Environ. Microbiol.*, 18, 4662–4673, 2016.
- Kuhn, M., Leeuw, J. D., and Zeileis, A.: Building Predictive Models in R Using the caret Package, *J. Stat. Softw.*, 28, 1–26, 2008.
- Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>, 2015.
- Li, J., Delgadobaquerizo, M., Wang, J., Hu, H., Cai, Z., Zhu, Y., and Singh, B. K.: Fungal richness contributes to multifunctionality in boreal forest soil, *Soil Biol. Biochem.*, 136, 107526, <https://doi.org/10.1016/j.soilbio.2019.107526>, 2019.
- Liu, L., Ji, M., and Buchroithner, M.: Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery, *Sensors (Switzerland)*, 18, 9, <https://doi.org/10.3390/s18093169>, 2018.
- Lozupone, C. A. and Knight, R.: Species divergence and the measurement of microbial diversity, *FEMS Microbiol. Rev.*, 32, 557–578, 2008.
- Maestre, F. T., Delgadobaquerizo, M., Jeffries, T. C., Eldridge, D. J., Ochoa, V., Gozalo, B., Quero, J. L., Garciagomez, M., Gallardo, A., Ulrich, W., Bowker, M. A., Arredondo, T., Barraza-Zepeda, C., Bran, D., Florentino, A., Gaitán, J., Gutiérrez, J. R., Huber-Sannwald, E., Jankju, M., Mau, R. L., Miriti, M., Naseri, K., Ospina, A., Stavi, I., Wang, D., Woods, N. N., Yuan, X., Zaady, E., and Singh, B. K.: Increasing aridity reduces soil microbial diversity and abundance in global drylands, *P. Natl. Acad. Sci. USA*, 112, 15684–15689, 2015.
- Minty, B., Franklin, R., Milligan, P., Richardson, M., and Wilford, J.: The radiometric map of Australia, *Explor. Geophys.*, 40, 325–333, 2009.
- Morellos, A., Pantazi, X., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., and Mouazen, A. M.: Machine learning based prediction of soil total nitrogen, organic

- carbon and moisture content by using VIS-NIR spectroscopy, *Biosyst. Eng.*, 152, 104–116, 2016.
- Müller, B.: Experimental interactions between clay minerals and bacteria: a review, *Pedosphere*, 25, 799–810, 2015.
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., and McBratney, A. B.: Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra, *Geoderma*, 352, 251–267, <https://doi.org/10.1016/j.geoderma.2019.06.016>, 2019.
- Nicolas, C., Martinbertelsen, T., Floudas, D., Bentzer, J., Smits, M. M., Johansson, T., Troein, C., Persson, P., and Tunlid, A.: The soil organic matter decomposition mechanisms in ectomycorrhizal fungi are tuned for liberating soil organic nitrogen, *ISME J.*, 13, 977–988, 2019.
- Nilsson, R. H., Larsson, K.-H., Taylor, A. F., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., and Abarenkov, K.: The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications, *Nucl. Acids Res.*, 47, D259–D264, <https://doi.org/10.1093/nar/gky1022>, 2018.
- Prescott, J. A.: A climatic index for the leaching factor in soil formation, *J. Soil Sci.*, 1, 9–19, 1950.
- Quinlan, J. R.: Learning with continuous classes, in: 5th Australian joint conference on artificial intelligence, vol. 92, pp. 343–348, Singapore, 1992.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org> (last access: 22 March 2022), 2014.
- Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 8–31, ISBN 026218253X, 2005.
- Rousk, J., Brookes, P. C., and Baath, E.: Contrasting Soil pH Effects on Fungal and Bacterial Growth Suggest Functional Redundancy in Carbon Mineralization, *Appl. Environ. Microbiol.*, 75, 1589–1596, 2009.
- Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.*, 36, 1627–1639, 1964.
- Schubler, A., Schwarzott, D., and Walker, C.: A new fungal phylum, the Glomeromycota: phylogeny and evolution, *Fung. Biol.*, 105, 1413–1421, 2001.
- Serna-Chavez, H. M., Fierer, N., and Bodegom, P. M.: Global drivers and patterns of microbial abundance in soil, *Glob. Ecol. Biogeogr.*, 22, 1162–1172, 2013.
- Shen, Z. and Viscarra Rossel, R. A.: Automated spectroscopic modelling with optimised convolutional neural networks, *Sci. Rep.*, 11, 208, <https://doi.org/10.1038/s41598-020-80486-9>, 2021.
- Shi, Z., Ji, W., Viscarra Rossel, R. A., Chen, S., and Zhou, Y.: Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library, *Eur. J. Soil Sci.*, 66, 679–687, 2015.
- Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., and Wetterlind, J.: Visible and Near Infrared Spectroscopy in Soil Science, *Adv. Agron.*, 107, 163–215, 2010.
- Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Moor, B. D., and Vandewalle, J.: Least Squares Support Vector Machines, 29–70, <https://doi.org/10.1142/5089>, World Scientific Collaborate, Hackensack, USA 2002.
- Talley, S. M., Coley, P. D., and Kursar, T. A.: The effects of weather on fungal abundance and richness among 25 communities in the Intermountain West, *BMC Ecol.*, 2, 7, <https://doi.org/10.1186/1472-6785-2-7>, 2002.
- Treseder, K. K. and Lennon, J. T.: Fungal Traits That Drive Ecosystem Dynamics on Land, *Microbiol. Mol. Biol. Rev.*, 79, 243–262, 2015.
- Tsakiridis, N. L., Keramaris, K. D., Theocharis, J. B., and Zalidis, G. C.: Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network, *Geoderma*, 367, 114208, <https://doi.org/10.1016/j.geoderma.2020.114208>, 2020.
- Vetrovsky, T., Kohout, P., Kopecky, M., Machac, A., Man, M., Bahnmann, B. D., Brabcova, V., Choi, J., Meszarosova, L., Human, Z. R., et al.: A meta-analysis of global fungal distribution reveals climate-driven patterns, *Nat. Commun.*, 10, 5142, <https://doi.org/10.1038/s41467-019-13164-8>, 2019.
- Viscarra Rossel, R., Bui, E., Caritat, P., and McKenzie, N. J.: Mapping iron oxides and the color of Australian soil using visible–near-infrared reflectance spectra, *J. Geophys. Res.*, 115, F04031, <https://doi.org/10.1029/2009JF001645>, 2010.
- Viscarra Rossel, R., Adamchuk, V., Sudduth, K., McKenzie, N., and Lobsey, C.: Proximal Soil Sensing. An Effective Approach for Soil Measurements in Space and Time, *Adv. Agron.*, 113, 243–291, <https://doi.org/10.1016/B978-0-12-386473-4.00010-5>, 2011.
- Viscarra Rossel, R. A.: Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra, *J. Geophys. Res.-Ea. Surf.*, 116, F04023, <https://doi.org/10.1029/2011JF001977>, 2011.
- Viscarra Rossel, R. A. and Behrens, T.: Using data mining to model and interpret soil diffuse reflectance spectra, *Geoderma*, 158, 46–54, 2010.
- Viscarra Rossel, R. A. and Brus, D. J.: The cost-efficiency and reliability of two methods for soil organic C accounting, *Land Degrad. Develop.*, 29, 506–520, 2018.
- Viscarra Rossel, R. A. and Hicks, W. S.: Soil organic carbon and its fractions estimated by visible-near infrared transfer functions, *Eur. J. Soil Sci.*, 66, 438–450, <https://doi.org/10.1111/ejss.12237>, 2015.
- Viscarra Rossel, R. A. and McBratney, A. B.: Soil chemical analytical accuracy and costs: implications from precision agriculture, *Aust. J. Exp. Agr.*, 38, 765–775, 1998.
- Viscarra Rossel, R. A., Chen, C., Grundy, M., Searle, R., Clifford, D., and Campbell, P.: The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project, *Soil Res.*, 53, 845–864, 2015.
- Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D., Dematté, J., Shepherd, K., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthes, B., Bartholomeus, H., Bayer, A., Bernoux, M., Bottcher, K., Brodsky, L., Du, C., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C., Knadel, M., Morras, H., Nocita, M., Ramírez López, L., Roudier, P., Campos, E., Sanborn, P., Sellitto, V., Sudduth, K., Rawlins, B., Walter, C., Winowiecki, L., Hong, S., and Ji, W.: A global spectral library

- to characterize the world's soil, *Earth-Sci. Rev.*, 155, 198–230, <https://doi.org/10.1016/j.earscirev.2016.01.012>, 2016.
- Wold, S., Sjöström, M., and Eriksson, L.: PLS-Regression: A basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130, 2001.
- Xu, T. and Hutchinson, M.: ANUCLIM version 6.1 user guide, The Australian National University, Fenner School of Environment and Society, Canberra, 2011.
- Yang, Y., Viscarra Rossel, R., Li, S., Bissett, A., Lee, J., Shi, Z., Behrens, T., and Court, L. N.: Soil bacterial abundance and diversity better explained and predicted with spectro-transfer functions, *Soil Biol. Biochem.*, 129, 29–38, 2019.
- Zhao, X. Q. and Shen, R. F.: Aluminum-Nitrogen Interactions in the Soil-Plant System, *Front. Plant Sci.*, 9, 807, <https://doi.org/10.3389/fpls.2018.00807>, 2018.