**Response to reviewers: Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions by Yang et al.**

# Reviewer 1

## General comments

Trying to estimate soil fungi with vis-NIR spectroscopy is venturing down a very slippery slope. Soil fungal abundance and diversity have no spectrally active components. This is recognized by the authors in the Introduction. At best, estimation of soil fungal abundance and diversity is based on indirect correlation with properties that are spectrally active, such as clay minerals functional groups. However, it is not even clear from the literature whether soil fungal abundance and diversity have much relationship with these spectrally active soil components. With the use of complex machine learning models, there is a serious risk of finding accurate results simply because the model finds fortuitous relationships between spectra/variables and the property of interest. Without any surprise, in this study the model performing best is the most complex one: the deep learning model. The simplest model, PLSR, performed poorly in nearly all cases. PLSR is yet the benchmark method for spectroscopic modelling, and is not per se a simple model, it uses the whole spectrum and makes statistical decomposition of the spectral data (principal component analysis). This clearly suggests that the results of this study cannot be trusted because the good results obtained by the 1D-CNN models are not based on direct nor known indirect relationship with spectrally active soil components. At best, there is some kind of indirect relationship with spectrally active soil components which serve as proxy for the soil fungal abundance and diversity, but it is difficult to prove that this relationship indeed exists based on the literature. The reliability of this relationship would also then depend strongly on the relationship found among the data by the deep learning model and on the specific study case and

calibration data. For this reason alone, I do not endorse publication of this manuscript. There is no reason from the existing literature to justify the use of vis-NIR spectroscopy to estimate soil fungal abundance and diversity. It is misleading to claim that spectroscopy can "serve to supplement the more expensive and laborious molecular approaches".

**Authors:** Where to begin...let us see if we can help the reviewer get a grip as we would not like him/her to slip down the slope that he/she is on!

1. On the point that fungal abundance and diversity do not have 'spectrally active components'. Correct. Nowhere in our manuscript do we write that they do. Most functional soil properties, e.g. cation exchange capacity, pH, clay content, etc., do not have 'spectrally active components'. Yet, spectroscopy is used to estimate them because spectra hold information on clay minerals, iron oxides, carbonates, colour (and chromophores), organic functional groups present in organic matter, water, particle size. It is now well understood that estimates of such functinal properties are due to (first or higher-order) correlations to those 'spectrally active components'. It is in this manner that fungi can correlate to the spectra, albeit with rather poor predictability, as we show in Figure 3 (submitted ms). Hence, the idea for spetro-transfer functions, but more on this later.

2. Interestingly, the reviewer does not mention that the spectra measure organic functional groups in organic matter. Spectroscopy provides a chemical fingerprint of these functional groups (e.g., carbonyls, carboxyls, hydroxyls, phenols, and phenyls). Although there are no absorptions that can be directly assigned to soil fungi or microbial communities more generally, soil microbes are dependent on organic compounds for energy, but also on clay minerals and iron oxides for the supply of essential elements (Müller, 2015). For example, the decomposers of saprotrophic fungi are involved in decomposing complex organic material. They are dominant in soil containing more lignocellulose, recalcitrant

aromatic polymers that consist of phenylpropane units joined by C–C and ether linkages (Tunlid et al., 2013). Mutualists of mycorrhizal fungi are dominant in soil containing more decomposed lignin polymers and humic-rich organic matter that can form supramolecular aggregates. These aggregates are stabilised by hydrophobic interactions and hydrogen bonding (Tunlid et al., 2013). Thus, soil fungi are linked with 'spectrally active components' in soil, at the very least from the perspective of carbon supply.

3. The literature has shown that spectra can estimate biological soil properties (as described above). For example, various authors have shown that vis–NIR spectra can estimate microbial biomass, enzyme activity in soil, and respiration (Coûteaux et al., 2003; Cohen et al., 2005; Zornoza et al., 2008; Chodak, 2011). There have been several reports on estimating soil microbial community composition with vis–NIR reflectance spectroscopy too. For instance, Zornoza et al. (2008) and Davinic et al. (2012) performed vis–NIR modelling of groups of microorganisms measured with traditional PLFA profiling. They could well estimate ($R^2$ >0.7) total PLFA biomass, total fungi, mycorrhizal fungi, total bacteria, Gram-positive bacteria and Actinomycetes. Using mid-infrared (mid-IR) spectroscopy, Nath et al. (2021) showed that they could estimate various soil biological properties. We found no research that uses vis–NIR spectra to infer soil microbial community composition based on 16S rRNA and ITS gene metabarcoding. Hence the added novelty of our study.

4. On the comments around PLSR and the machine learning methods. Glad to know the reviewer understands how PLSR and machine learning methods work. We do too, so we will keep this brief and refer to Figure 3 in the manuscript to make a few points that seem to have evaded the reviewer: (i) Our results clearly show that the spectra alone can not well estimate fungal phyla abundance and diversity, (ii) The soil and environmental data could better estimate fungi because they control the abundance and distribution of fungi—of course, this

has been well-reported (Delgadobaquerizo et al., 2018; Vetrovsky et al., 2019; Tedersoo et al., 2014 ), (iii) By combining the spectra with the soil-environment data (i.e. developing spectro-transfer functions), we could improve the modelling and achieve the best accuracies, and (iv) The PLSR estimates of the different phyla and ACE diversity were not too dissimilar to other machine learning methods, but yes, our optimised CNNs did perform somewhat better.

5. We have shown that spectro-transfer functions could fairly accurately (although as R2 notes, not superbly well) estimate fungal community composition and diversity measured with ITS gene metabarcoding using a large dataset from diverse ecosystems under different conditions. As stated in the manuscript, we are under no illusion that our method will replace metabarcoding. However, we propose that it could complement metabarcoding when many measurements are needed. ITS gene metabarcoding analyses are expensive, laborious and require specialised laboratories and methods. Once a spectro-transfer function is developed, vis–NIR spectroscopy is much easier, faster and less costly to measure, and soil-environmental data are more readily available. Since soil fungal properties are sensitive and reliable indicators of soil health, there has been much research to find variables related to soil fungal community to assess degradation processes, restoration strategies or management practices. Consequently, we think it worthwhile to research the development of spectro-transfer functions to model soil fungal composition and diversity.

6. Finally, we hope to have addressed the reviewer's concerns. We do not ask the reviewer to trust our models or us for that matter; however, we do ask that he/she reads our manuscript, tries to understand our results, interprets them accurately and provides us with a fair, unbiased assessment.

In addiction to this major conceptual problem, I have also several important comments that need to be addressed, please see below.

## Specific comments

This paper is very similar to the Soil Biology and Biochemistry paper: same methodology, same data (fungi instead of bacteria, but still on the Australian BASE dataset), same covariates, same concept.

**Authors:** The general concept of the spectro-transfer functions is similar in both papers, but because we found modelling of fungi more challenging, we needed to research different methods and we thought useful to report them. This paper looks at fungal composition and diversity measured by ITS metabarcoding. To develop the spectro-transfer functions we tested different strategies (using spectra, environmental data, and a combination) and learning methods (7 different algorithms).

Combining spectral data and terrain into the modelling is not recommended and not common is spectroscopic research. The authors provide as input to the model a vector comprising both spectral data and environmental data found at site. The rational for doing this is unclear, and providing too much data to complex models will only aggravate the problem of finding spurious relationship among the data. This way of modelling brings additional problem with the covariates used. For example, the covariates called "mineralogy", i.e. kaolinite, illite and smectite (Appendix), are an interpolation of vis-NIR spectra band depth. So these covariates add redundant information. Combining spectral data and environmental variables into a single vector used for prediction is going too far into "unconscious" soil modelling.

**Authors:** We wonder who does not recommend using different data in modelling? and simply because it isn't commonly done, does the reviewer suggest that we shouldn't either? A couple of rather strange and subjective statements. Our rationale is clearly stated: we wished to test spectro-transfer functions. Of course, there are other examples in the soil and environmental sciences literature where spectra and other environmental data are combined in modelling. For example (amongst many other), Pullanagari et al. (2018) integrated airborne hyperspectral, topographic, and soil data for estimating pasture quality with Random Forest. Guo et al. (2019)

combined environmental factors and vis–NIR spectra to estimate soil organic matter with geospatial techniques. On the covariates, we see no issue with using data on soil clay minerals in the modelling, even if they were derived from specific absorptions in the near-infrared. The data were measured independently. We are unsure of the meaning of 'unconcious soil modelling', however, we hope that the reviewer is conscious that we perform the modelling using best-practice, and that we interpret the modelling (see Figure 5), finding that the interpretation was as reasonable as one might expect with this type of modelling.

To my understanding (because the writing needs to be improved and made more precise) the modelling is made on the absorbance spectra whereas the interpretation is made on a different model based on the continuum removed reflectance spectra. This is not correct. Interpretation should be made on the same model that the authors choose to be the best, not on a model fitted on different data. The authors are interpreting a model that was not used during modelling.

**Authors:** On the matter of the writing, it would serve the reviewer (and us) well if he/she could point out where and how our writing is lacking so that we might better address the comment. The manuscript was read and edited by the native English speaking authors as well as colleagues in our institutions before submission. We note that Reviewer 2 wrote that the paper is '...overall well written and is very well structured'. About the form of the spectra used for interpretation, there may be some misunderstanding here. Figure 2 is a stand-alone representation of the continuum removed spectra, which is simply to show that fungi actually show response in the vis–NIR. We used the continuum removed spectra because it helps to visualise the absorptions, more clearly than the absorbance first derivative spectra. Yes, the modelling was performed on the absorbance spectra, that is correct. Because the models were derived using absorbance spectra, their interpretation was made using these spectra and that interpretation is reported in Figure 5. To prevent confusion, we will include both types of spectra in Figure 2, the continuum removed and the

absorbance, first derivative spectra.

The BASE dataset contains more than 577 samples. Also, in the SBB paper 681 samples are used for bacteria. Why are not all data used? What is the reason to discard some observations? This should be clearly described in the Methods section of the manuscript.

**Authors:** We wrote this in the submitted manuscript, lines 86–90. When we calculated community diversity, to remove the bias induced by unbalanced sequencing, each sample was resampled at the same sequencing depth. The BASE dataset sought to produce as many sequences as resources allow with a minimum sequencing number of 10,000 per sample. Here, each sample was re-sampled at depth of 11,000 sequences to eliminate the unbalanced sequencing as shown in the Supplementary Figure 1. We chose 11,000 sequences as the resampling depth mainly because many samples only had this sequences number but also the rate of increase in the rarefaction curves is small at this depth. Doing this produced 577 data, which is therefore different that of the bacterial study.

The environmental variables used as covariates are outdated (Appendix). For example, the prescott index map for Australia is made at 5 km, but indeed downscaled at 90 m. The mineral maps (kaolinite, illite and smectite) are an interpolated product based on the band depths of vis-NIR spectra -should not be used here to avoid redundancy. The soil texture maps of 2015 are outdated, there are new ones since several months (see the paper https://www.publish.csiro.au/sr/pdf/SR20284). The vegetation maps are also very outdated. The authors used the old ones based on 250 m and 1 km resolution products, but there are new ones since several months based on Landsat 30 m data, see for example: http://data.auscover.org.au/xwiki/bin/view/Product+pages/Landsat+Seasonal+Fractional+Cover. Also, all the topographic covariates are now available at 30 m resolution and easy to download through the TERN repository: https://shiny.esoil.io/Covariates/.

**Authors:** The spatial data that we used is from around 2011–2017. The BASE soils were collected within that period so there is no reason to use newer data in this research. We see no reason why we should re-run all our analysis simply because there are newer maps of soil texture (Note that based on our assessments, when tested with independent data, the new texture maps provide insignificant improvements in overall accuracy compared to Figure 8 in Viscarra Rossel et al. (2015)). Similarly for the comments around resolution. Please note that the Prescott index layer that we used was produced at an approximate 90 m pixel resolution (by CSIRO terrain data analysts) using down-scaled climatic data. Also, why should we re-run all our analysis simply because there are newer, higher resolution datasets? We think it more important to use datasets of similar age. We have responded to the comments on the mineralogy maps.

The use of various machine learning methods is not of interest and there is the risk that this study simply becomes a comparison of validation statistics. Not clear what the added value is of applying seven methods. In this way, the manuscript runs the risk of being about the models and their comparison, rather than about understanding whether vis-NIR spectral data can effectively predict soil fungi. I think SOIL' readership is rather interested in the latter. Any model that here predicts better than another is case-study specific, and is unlikely to interest soil scientists who are the readers of this journal. Further, a model is usually chosen carefully with the problem in hand. Several of the models used by the authors are in fact very similar, they are all non-linear. The problem is here that none of the models described in the Appendix are thoroughly explained and it is not clear how the authors actually implemented all this. I could make many questions on each of the models, but my best advice here would be to reduce the model count to the minimum (PLSR and DL), otherwise most readers will probably see this manuscript as a programming exercise.

**Authors:** We disagree that the testing of the different methods is of no interest and

disagree with the suggestion to only report results for PLSR and DL. The purpose of our paper is to develop predictive spectro-transfer functions for fungal abundance and diversity and this requires us to test different algorithms for the modelling. Contrary to the comment that this might confuse readers about the purpose of the manuscript, our result show that all of the algorithms perform similarly (see Figure 3), and this should provide readers with confidence that the response captured by the modelling is 'real' and the modelling with the spectro-transfer functions robust. On the comment that 'a model is usually chosen carefully with the problem in hand'—this is not necessarily our experience with empirical modelling and machine learning. Usually, depending on the purpose, the datasets, etc., one needs to test different algorithms to find the most appropriate one. On the comment around the similarity of the algorithms, we tested one linear model (PLSR), three tree-based models (RF, XGBoost, Cubist) that work quite differently, SVM, Gaussian Processes and optimised CNNs. We chose those different methods because they provide a good and varied range of methods. We would be very surprised if after reading our manuscript readers thought that it was a 'programming exercise', particularly as we devote large portions of the results and discussion to interpreting our results and also provide Table 2 and Figure 5 for this purpose. We provide clear descriptions of the algorithms and their implementation in the Methods and Supplementary Information. Our descriptions are not exhaustive because these methods are not new and have been widely used and reported in other literature, which we cite.

It is unclear why the authors describe the accuracy statistics this way and what was done. With the information provided, the interpretation of these statistics is also unprecise. For example, the coefficient of determination of the linear model can be interpreted as amount of variance explained only in the case of a linear model with intercept. So the $R^2$ that the authors report in the manuscript is similar to a squared correlation coefficient, and should not be interpreted as percent of variance explained. See also the Wikipedia page, last paragraph of the intro:

https://en.wikipedia.org/wiki/Coefficient$_$of$_$determination. Further, decomposition of the RMSE into a bias and variance component is very well known in the literature and adds nothing to the paper. How the authors call it with "inaccuracy" and "imprecision" is not well accepted in the statistical literature. The decomposition is so well known that it is described in Wikipedia: https://en.wikipedia.org/wiki/Mean$_$squared$_$error or in any introductory statistic book, see, for example, page 298 in the book "Elementary Statistics for Geographer", by Burt et al. (2009), Third Edition. https://www.routledge.com/Elementary-Statistics-for-Geographers/Burt-Barber-Rigby-Robeson-Horner/p/book/9781572304840. Note that in both cases, they do not use terms such as "inaccuracy" and "imprecision" and I would agree not to use them.

**Authors:** Our reporting of the evaluation statistics are useful and correct. Reporting of the RMSE, ME and SDE describes the relevant errors because they represent the bias (ME), the imprecision (SDE) and thus the inaccuracy (RMSE) of the estimates, noting that $RMSE^2 = ME^2 + SDE^2$. That is, if the model predictions are unbiased, then imprecision will be the only contributor to their inaccuracy, and this seems to be the case for most of our results. We have cited Viscarra Rossel and Mcbratney (1998) who described the reporting of erros in this way, following the paper by Kempthorne and Allmaras (1965). The reviewer can find more detail in those texts. We additionally report the $R^2$, which is a more 'contextualized' measure of performance that can be used for more general assessments and comparisons to other work. $R^2$ we report definitely represent the proportion of the variation in the dependent variable that is predictable from the independent variable(s).

The methods section of the manuscript does not clearly describe what was done. For example, there are many methods available to compute the variable importance. At lines 137-140, the only description is "VarImp for caret" and permutation importance for the 1D-CNN models. This is not a description that allows one to

reproduce these results. What is done in caret VarImp? For the permutation importance of the 1D-CNN model, how is correlation among wavelengths accounted for (permutation is very sensitive to correlation)? How many permutations? What is the metric used (RMSE difference, ratio?)? What is the unit?

**Authors:** We disagree that our methods are unclear. The manuscript was also read by other colleagues before submission and Reviewer 2 could understand the methods, writing that the paper is '...overall well written and is very well structured'. However, we agree that the description of variable importance is a little brief. We will revise and improve the description of our implementation of the variable importance. Permutation importance measures the changes in a metric (e.g. RMSE) after a feature is permuted, where the relationship between the feature and the targets are broken. It can be applied on any estimator and does not require retraining the model. We used permutation importance and the RMSE) with 1000 permutations to calculate the variable importance of the CNN models. The importance values measures the RMSE increase after permutating the particular feature. In order to compare the importance between different fungal phyla and diversity, we scaled the importance values between 0 and 1. Note that we are aware of the issues with using permutation, however, we felt that for this study and with vis–NIR spectra, it would suffice if we used a large number of permutations to calculate the variable importance. We understand permutation importance has its limitations and that there are other methods to decipher CNNs, however they too have strengths and limitations (Fisher, et al., 2019). Although we think that it would be out of scope here, if the editor suggests, we could use another method that can better handle multicollinearity, e.g. SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017). However, doing so might not necessarily improve our results or change our conclusions, but it would significantly complicate and lengthen the manuscript.

In Fig. 5, the importance should have a unit. What is the unit? How comes that covariates and spectra, which have very different units, can all have importance

values between 0 and 1? Usually a standardization can be made to the covariates and spectral data prior to use permutation on them. If this was done, please report it in the methods section and also use the same standardized input data for modelling. There is no reason to use different data and model for modelling and interpretation.

**Authors:** The spectra and the covariates were centred and scaled before the modelling. This is clearly stated in the submitted manuscript line 126. For Permutation importance, please see our previous response.

# 1 References


Coûteaux, M.M., Berg, B., Rovira, P., 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. Soil Biology and Biochemistry, 35 (12), 1587–1600.

Cohen, M.J., Prenger, J.P., DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. Journal of Environmental Quality, 34(4), 1422–1434.

Chodak, M., 2011. Near-infrared spectroscopy for rapid estimation of microbial properties in reclaimed mine soils. Journal of Plant Nutrition and Soil Science, 174 (5), 702–709.

Davinic, M., Fultz, L.M., Acosta-Martinez, V., et al., 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. Soil Biology and Biochemistry, 46 (1), 63–72.

Delgadobaquerizo, M., Reith, F., Dennis, P. G., et al., 2018. Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere, Ecology, 99, 583–596.

Fisher, A., Rudin, C. and Dominici, F., 2019. 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(177), pp.1-81

Guo, L., Zhang, H., Chen, Y. et al.,2019. Combining Environmental Factors and Lab VNIR Spectral Data to Predict SOM by Geospatial Techniques. Chin. Geogr. Sci., 29, 258–269.

Horwitz B., Mukherjee P., Mukherjee M., Kubicek C. (eds) Genomics of Soil- and Plant-Associated Fungi. Soil Biology, vol 36. Springer, Berlin, Heidelberg.

Kempthorne, O., Allmaras, R. R., 1965. Errors of observation. In 'Methods of Soil Analysis. Part 1. Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling'. Agronomy Monograph No. 9.(Ed. C. A. Black.) (ASA/SSSA: Madison, WI, USA.)

Lundberg, Scott M., and Su-In Lee. 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems.

Müller, B., 2015. Experimental interactions between clay minerals and bacteria: a review, Pedosphere, 25, 799–810.

Nath, D., Laik,R., Meena,V. S., et al., 2021. Can mid-infrared (mid-IR) spectroscopy evaluate soil conditions by predicting soil biological properties? Soil Security, 4, 100008.

Pullanagari, R., Kereszturi,G., Yule, I. 2018. Integrating Airborne Hyperspectral, Topographic, and Soil Data for Estimating Pasture Quality Using Recursive Feature Elimination with Random Forest Regression" Remote Sensing, 10(7), 1117.

Tedersoo, L., Bahram, M., Polme, S., et al., 2014. Fungal biogeography. Global diversity and geography of soil fungi. Science, 346 (6213), 1256688.

Tunlid A. et al., 2013. Genomics and Spectroscopy Provide Novel Insights into the Mechanisms of Litter Decomposition and Nitrogen Assimilation by Ectomycorrhizal Fungi. In:

Vetrovsky, T., Kohout, P., Kopecky, M., et al., 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns., Nature Communications, 10, 5142.

Viscarra Rossel, R. A. and McBratney, A. B., 1998.Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture, 38, 765–775.

Zornoza, R., Guerrero, C., Mataix-Solera, J., et al., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in mediterranean soils. Soil Biology and Biochemistry. 40 (7), 1923–1930.