Estimating Soil Organic Carbon from Multispectral Images using Physics-Informed Neural Networks

James Sargeant[1], Shyh Wei Teng[1], Manzur Murshed[2], Manoranjan Paul[3], and David Brennan[4]

[1] Institute of Innovation, Science and Sustainability, Federation University, Australia. {j.sargeant,shyh.wei.teng}@federation.edu.au
[2] School of Information Technology, Deakin University, Australia. m.murshed@deakin.edu.au
[3] School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia. mpaul@csu.edu.au
[4] Wimmera Catchment Management Authority, Horsham, Australia. david.brennan@wcma.vic.gov.au

Abstract. Understanding the amount of Soil Organic Carbon (SOC) at farm and field scale is a necessary precursor to effective management, important for both agricultural productivity and to reduce CO_2 emissions. To avoid the prohibitive cost of measurement, SOC can be estimated by using multispectral images. In this study, we propose a novel Physics-Informed Convolutional Neural Network (CNN) to model well-known but noisy relationship between a soil index and SOC using the network's loss function. This study is also conducted by resampling the European Land Use/Classification Area Survey (LUCAS) dataset to Sentinel-2 bands. Our experimental results show that our proposed network converges more quickly, has a lower root mean squared error (RMSE) and is more robust (as measured by the standard deviation of RMSE over multiple trials) than a compatible standard CNN. The operation of the novel Physics-Informed CNN is explained in terms of the components of the loss function.

Keywords: Physics-Informed Neural Network · Multispectral Images · Soil Organic Carbon

1 Introduction

The large number of samples required to make informed business decisions on farm management practices of Soil Organic Carbon (SOC) stocks makes traditional laboratory-based measurement of SOC too expensive. SOC can also be estimated from multispectral images of soil, typically through the use of soil and vegetation indices (derived from raw bands), through the raw band themselves or though a combination of indices and bands. None of these three methods have been shown to be superior to one another, and all have strengths and weaknesses.

Where indices and raw bands are both used, they are used in the same manner. This is despite significant differences in the underlying physics of the way they convey information about SOC (and other soil properties).

Physics-Informed Neural Networks (PINNs) have not previously been used to estimate SOC. In this paper we propose the novel use of a PINN to estimate SOC from multispectral images by separately treating the raw bands and a single soil and vegetation index. Raw bands are used as inputs into a Convolutional Neural Network (CNN). The soil and vegetation index is used in the loss function to ensure that the method takes advantage of imprecise soil properties. The loss function is further enhanced so that the method initially uses the soil and vegetation index to converge more quickly to a solution and then refines that solution based on the raw bands.

Our proposed PINN architecture converges more quickly, has better accuracy and is more robust to noise than an equivalent CNN that does not use a physics-based loss function. Moreover, the operation of the method is more explainable than a standard CNN.

The remainder of this paper is as follows. Sec. 2 discusses the various properties of SOC, multispectral images, and soil and vegetation indices that inform the use of a PINN to estimate SOC. Sec. 3 discusses the architecture of our proposed PINN. Sec. 5 discusses the relative performance of the PINN method compared to a standard method (*i.e.* using a standard loss function) of the same architecture, and also discusses the performance of an alternate loss function. Sec. 5 also explains the operation of the PINN method in terms of the various components of the loss function. Sec. 6 concludes the paper.

2 Related Works

2.1 Climate Change and Soil Organic Carbon

SOC plays a vital role in soil quality, adding structure, water holding capacity, nutrient holding capacity and increasing bacterial activity, all of which are essential to productive growth.

Approximately 12% of the earth's land area is used for cropping [1]. With agriculture contributing significantly to CO_2 emissions, there is increasing interest in making sure that farm practices are as carbon neutral as possible [22].

At field (or paddock) scale, farmers wishing to participate in the soil carbon sequestration market or to demonstrate that their practices are carbon neutral (for market access reasons) will need to understand:

- Where the soil carbon is located?
- How much soil carbon is present?
- How is the soil carbon to be measured or estimated?
- What farm practice changes are needed to conserve or increase soil carbon?

2.2 Non-machine Learning Based Measurement of SOC

Measurement of SOC is an expensive and time-consuming task, as it requires field collection and subsequent laboratory analysis. The cost of SOC measurement is approximately AUD$500 per sample and thus prohibitive given the number of samples required for the detailed understanding required to manage this resource at farm and field scales.

An alternative to laboratory measurement is estimation of SOC through analysis of the reflected visible to near infrared light. Organic compounds have reflectance "overtones" at specific wavelengths in the VIS, NIR and SWIR ranges [3, 29, 30]. Viscarra Rossel et al. [30] found moderate positive peaks for SOC at 410, 570 and 660 nm. Chabrillat et al. [6] identified that organic carbon responded to wavelengths in the approximate ranges of 750 – 1150 nm and 2200 – 2300 nm. These studies give theoretical weight to the idea that SOC can be estimated from reflectance spectra of bare soils.

Many studies (*e.g.* [11, 16–18, 21, 25–27]) have confirmed that SOC can be estimated from reflectance spectra of bare soil samples.

Multispectral satellites, such as Sentinel-2 and Landsat 4-8, have 6 to 13 narrow bands across the visible to near infra-red bands. Some of the wavelengths captured by these satellites are particularly sensitive to the amount of SOC present in the top few centimetres of the soil. Some multispectral satellites also have bands in the long-wave infra-red part of the spectrum.

The advantages of multispectral images are:

- Low cost of images (zero cost for Sentinel-2 and Landsat images).
- Short revisit times – Sentinel 2 images are available every five days.
- Reasonable resolution (*e.g.* 10 metres for some Sentinel-2 bands).

Various soil and vegetation indices can be created from the different wavelengths in a multispectral image. These soil and vegetation indices have been shown to have good predictive power of SOC.

Soil indices are based on the empirical observation that the relationship between the near-infrared and the red reflectance of any particular soil property is generally linear. Moreover, this relationship can be generalized beyond near-infrared and red. However, exploiting this linear relationship to estimate SOC from one index is not particularly accurate due to the complexity of soil and the interplay of soil properties.

Spectral indices have advantages over raw bands insofar as they reduce spectral reflection errors, the effects of moisture, the effects of soil roughness and the effects of the atmosphere [4, 16].

However, there are some disadvantages of indices when compared with using raw bands:

- Using spectral indices alone implicitly causes a lot of spectral information to be ignored.
- Many of the indices used for SOC research are "Vegetation" indices. The rationale for using such indices appears to be that vegetation is affected by SOC content – thus vegetation indices become "proxies" for SOC.

– Sophisticated neural network methods can identify subtle relationships between the bands to more precisely determine SOC (accuracy being a different matter). Neural networks, in particular convolutional neural networks, have also been shown to have good predictive power of SOC. However, neural networks have some drawbacks, such as the need for large (and expensive) training datasets [18,21] and a lack of explainability .

2.3 Machine and Deep Learning of SOC

Studies that estimate SOC using remote sensing at farm and field scales have largely used traditional machine learning techniques, with Random Forest, Cubist, Partial Least Squares Regression and Support Vector Regression being the most popular. However, reasonable accuracy was dependent on using non-spectral inputs in addition to indices and/or raw bands [10,13,14].

The few studies that used neural network techniques to estimate SOC at farm and field scales were limited to simple one, two or three hidden layers. Accuracy was typically superior compared to machine learning methods [12,16], though not always [13].

CNNs were originally developed for image classification purposes and were shown to be superior to other neural network architectures. Subsequently, they have been adapted to many other tasks with good effect.

1D-CNNs have been shown to be effective at estimating SOC in laboratory settings [18,24–26] Shen & Viscarra Rossel [18] note that CNNs "can potentially outperform conventional statistical and machine learning models", and also note that 1D-CNNs produce more accurate predictions than 2D-CNNs. However, only a few CNN studies have involved satellite images (*e.g.* [33]) and no farm or field studies using CNNs were identified in the literature search.

A major drawback with CNN (and other deep learning methods) is that they require a large amount of training data [18,21]. Vaudour et al. [28] and Odebiri et al. [20] suggest that the lack of studies that use deep learning to estimate SOC is due to the lack of large datasets, computational resources and the problems associated with being able to explain the resulting models. Reeves et al (2012), cited in England & Viscarra Rossel [8] caution that SOC spectroscopic models that are developed and validated with insufficient samples may be unstable and provide incorrect results.

2.4 Physics-Informed Neural Networks

More recently, PINNs [23] have been developed to overcome some of the problems associated with neural networks. These methods use known physical relationships between dependent and independent variables to ensure that the model output is consistent with the laws of physics. The advantages of PINNs over regular neural networks include:

– Lower number of "ground truth" samples are required [2].
– The method can converge to a solution more quickly.

– The model can be used for estimation when the inputs are outside the range used in training [2].

PINNs have been used to good effect with problems that have a well-known, though highly complex, relationship between model inputs and outputs, typically involving partial differential equations [2,5,15,23].

A literature review we conducted did not identify any studies that used PINNs to estimate SOC.

3 Dataset

The European Land Use/Classification Area Survey (aka LUCAS) [7] is a well-known, publicly available dataset that is frequently used to test soil property estimation techniques.

The LUCAS dataset contains a file containing analytical records for a range of soil, topographic, land class and land use properties. The LUCAS dataset also contains one or more (typically two) hyperspectral records for each analytical data record, organized by country, allowing combined analytical and spectral records to be created for one or more countries.

Combined records were created for each country in the LUCAS dataset, and the spectra were resampled to Sentinel-2 bands using the specific spectral response of the Sentinel-2 multispectral imager [9].

A linear regression analysis was conducted for each country to determine the correlations between soil organic carbon and 40 different indices, calculated on the basis of the resampled Sentinel-2 bands.

The index with the highest coefficient of determination (R^2 was selected for further analysis. Countries with high numbers of spectral records (>1000) were then combined into a single dataset along with the SOC values from the analytical file. This was used as input data to the neural network.

4 Proposed Method

4.1 Convolutional Neural Network Architecture

We have chosen 1D-CNN as the backbone architecture for our proposed PINN for the following reasons:

– 1D-CNNs are able to identify subtle relationships between the different bands.
– Padarian et al. [21] found good performance (R^2, RMSE) when using the entire LUCAS dataset with a 1D-CNN with little spectral pre-treatment.
– Shen & Viscarra Rossel [25] used a CNN with a 1:20 down-sampled version of a significant subset of the LUCAS dataset, achieving a coefficient of determination (R^2)of 0.89.

The architecture of the 1D-CNN is shown in Fig. 1. Our improvements over a standard CNN are:

– Addition of a Softplus activation function ($log(1+e^x)$) just before the output to force SOC to be greater than zero.
– Unlike other models that use one or more soil and vegetation indices, the index is not used as an input. Instead, the index is introduced into the loss function.

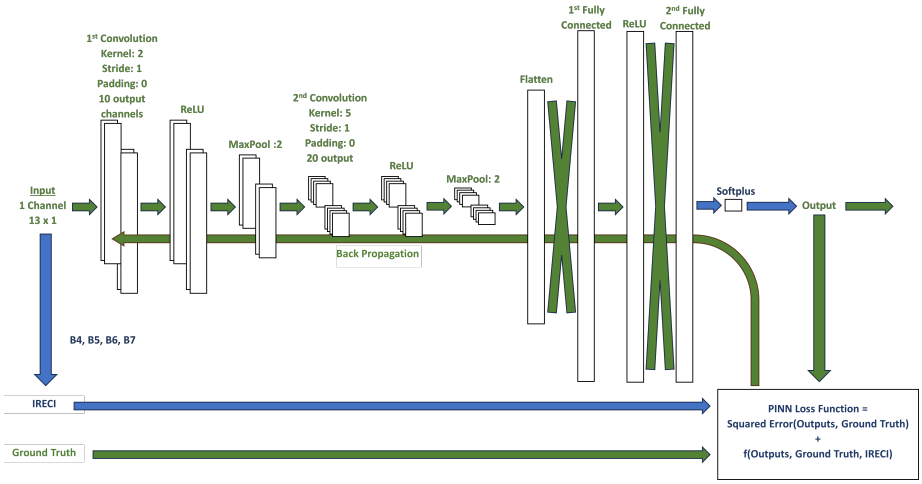

Fig. 1: 1D-CNN PINN Architecture. Improvements to a conventional CNN shown in blue.

4.2 PINN Loss Function Development

As applied to estimating SOC from Sentinel 2 reflectance bands, a "pure" neural network attempts to learn the relationship between SOC and the spectral reflectance bands of Sentinel 2:

$$SOC = f(B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B10, B11, B12) \quad (1)$$

During training, the neural network adjusts the weights associated with each band and feature to minimize the Euclidean distance between the predicted value from the model and the "ground truth". The resulting loss function is:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(SOC_{Actual_i} - SOC_{Model_i})^2 \quad (2)$$

or

$$\mathcal{L} = \mathcal{L}_{MSE} \quad (3)$$

The optimizer function inside numerically calculates the partial derivatives (and the first and second moments in the case of the Adam optimizer algorithm) of

each of the inputs (Sentinel 2 bands B1 to B12) to determine the direction in which to alter every weight in order to minimize the entire loss function.

The issues with this technique are:

- The optimizer may adjust the weights in such a way that the loss function converges on a local rather than a global minimum.
- The resulting model may not be reliable for inputs outside the training range.
- The resulting model may have been over-fitted.
- The resulting model is not explainable.
- A large number of measurements are required to train the model.

In our method, the loss function is augmented by an additional term that captures the physical relationship between the index and SOC.

$$\mathcal{L} = \mathcal{L}_{MSE} + \mathcal{L}_{Physics} \tag{4}$$

Two candidates for Loss $_{\text{Physics}}$ were tested, a candidate loss function Eq. (5) and an alternate loss function Eq. (6):

$$\mathcal{L}_{Physics} = \frac{1}{N}\sum_{i=1}^{N}(SOC_{Actual_i} - SOC_{Computed_i})^2 \tag{5}$$

and

$$\mathcal{L}_{Physics} = \frac{1}{N}\sum_{i=1}^{N}(SOC_{Model_i} - SOC_{Computed_i})^2 \tag{6}$$

The loss functions thus become (without loss of information):

$$\mathcal{L} = \sum_{i=1}^{N}(SOC_{Actual_i} - SOC_{Model_i})^2 + \sum_{i=1}^{N}(SOC_{Actual_i} - SOC_{Computed_i})^2 \tag{7}$$

$$\mathcal{L} = \sum_{i=1}^{N}(SOC_{Actual_i} - SOC_{Model_i})^2 + \sum_{i=1}^{N}(SOC_{Model_i} - SOC_{Computed_i})^2 \tag{8}$$

where

$$SOC_{Computed} = m \cdot Index + c \tag{9}$$

With m and c determined by conducting a linear regression of the training data.

The loss function can only be minimized by ensuring that both halves are minimized, *i.e.* the model must fit the data and ensure that Eq. (9) is satisfied to the best extent possible.

Because the optimizer is provided with the mathematical relationship between the Loss function and the index, and with the mathematical relationship between the index and the specific bands that form that index, it can (numerically) calculate the partial derivatives of the entire loss function with respect to

those specific bands. The optimizer then uses these gradients to determine the next values of the model parameters.

Eqs. (7) and (8) were further modified to include a parameter Ω, whose value was selected during training but constrained to the range [0,1]. Ω is initially set to zero.

$$\mathcal{L} = \Omega \cdot \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i})^2 + (1-\Omega) \cdot \sum_{i=1}^{N} (SOC_{Acutal_i} - SOC_{Computed_i})^2 \tag{10}$$

$$\mathcal{L} = \Omega \cdot \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i})^2 + (1-\Omega) \cdot \sum_{i=1}^{N} (SOC_{Model_i} - SOC_{Computed_i})^2 \tag{11}$$

The rationale for including Ω is to initially guide the network to a solution based on the physics-based index and then to use all of the information contained in the individual bands to refine that solution.

5 Results and Discussion

A one-dimensional convolutional neural network (1D-CNN) comprising two convolutional layers and two fully connected layers (Fig. 1) is used to estimate topsoil SOC from 13 spectral bands that correspond to Sentinel-2 bands.

The accuracy of the network (as measured by RMSE) when using a conventional mean-squared error loss function is compared to the accuracy of the network using loss functions using IRECI inspired by PINN research [2,5,15,23]. An additional term is added to the loss function so that the network initially focuses on the index and then transfers its focus to the conventional mean squared error loss function to refine the index-generated model.

Experiments were conducted using a Lenovo Thinkpad X1 Carbon Generation 5: i7-12700H Processor, NVIDIA® GeForce RTX™3050 Ti Laptop GPU 4GB, 32 GB DDR5 memory, 1 TB SSD storage. Programs were written using the Pytorch Lightning library.

5.1 Index selection

County-by country linear regressions were carried out to determine the coefficient of determination between SOC and 40 soil and vegetation indices. Tab. 1 shows countries and indices where the coefficient of determination (R^2) were signigicant. IRECI was selected for on the basis that it was the index with the highest R^2 for four countries, and the highest R2 for countries with more than 1000 samples.

In terms of Sentinel-2 satellite bands, IRECI is calculated as:

$$IRECI = \frac{B7 - B4}{(B5/B6)} \tag{12}$$

The strong performance of IRECI is consistent with [19], which identified an imperfect linear relationship between SOC and IRECI in two Western Australian wheatbelt farms.

Table 1: Linear regression coefficients of determination (R^2) with maximum R^2 for each country shown in bold. Columns are individual countries (with number of sample in dataset for each country shown in the heading). Rows are R^2 for each index in left-most column

Index	**EE 388 Samples**	**FI 2296 Samples**	**IE 392 Samples**	**LT 704 Samples**	**LV 614 Samples**	**SE 3780 Samples**	**SI 294 Samples**
BI2	0.7052	0.6940	0.7907	0.6903	**0.7177**	0.6259	0.6539
IRECI	**0.6784**	**0.7066**	**0.8007**	0.6853	0.6974	**0.6276**	0.7490
MCARI	0.6048	0.5837	0.6585	**0.6992**	0.6650	0.4432	0.7570
MTVI	0.4290	0.4304	0.5750	0.5598	0.4729	0.3090	**0.7600**

A dataset comprising samples from Finland (FI) and Sweden (SE) was created and downsampled to Sentinel-2 bands.

Now, Eq. (9) becomes:

$$SOC_{Computed} = m \cdot IRECI + c \tag{13}$$

The loss functions for the PINNs thus become:

$$\mathcal{L} = \Omega \cdot \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i})^2 + (1-\Omega) \cdot \sum_{i=1}^{N} (SOC_{Acutal_i} - m \cdot IRECI_i - c)^2 \tag{14}$$

$$\mathcal{L} = \Omega \cdot \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i})^2 + (1-\Omega) \cdot \sum_{i=1}^{N} (SOC_{Model_i} - m \cdot IRECI_i - c)^2 \tag{15}$$

The values of m and c were provided to the algorith by extracting them from a linear regression of the training set prior to model training. Hyper-parameters and model design were not optimised but produced adequate results when compared with other LUCAS 1D-CNN studies [21,26,31].

5.2 Candidate Loss Function

10 runs of 500 epochs were executed. For each run, training sets of 4253 (70%) samples were selected at random from the 6076 samples in the dataset. The remainder of the samples were used for validation. For each run, two models

were trained with the same training set, a data only model (Eq. (16)) and a PINN model using the candidate loss function (Eq. (14)).

$$\mathcal{L} = \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i}) \tag{16}$$

Average convergence behaviour for accuracy as indicated by the median and mean of the root mean square error (RMSE) over the 10 runs is shown in Fig. 2. The PINN that used the candidate loss function showed faster convergence and a lower final RMSE (median of 54.76 g/kg, mean of 55.03 g/kg) than the DataOnly model (median of 55.13 g/kg, mean of 55.44 g/kg). While the difference is small, the Wimmera Catchment Management Authority [32] notes that even a one percent improvement in accuracy represents significant economic value (in the order of millions of dollars) when applied to its large area of 30,000 square kilometres.

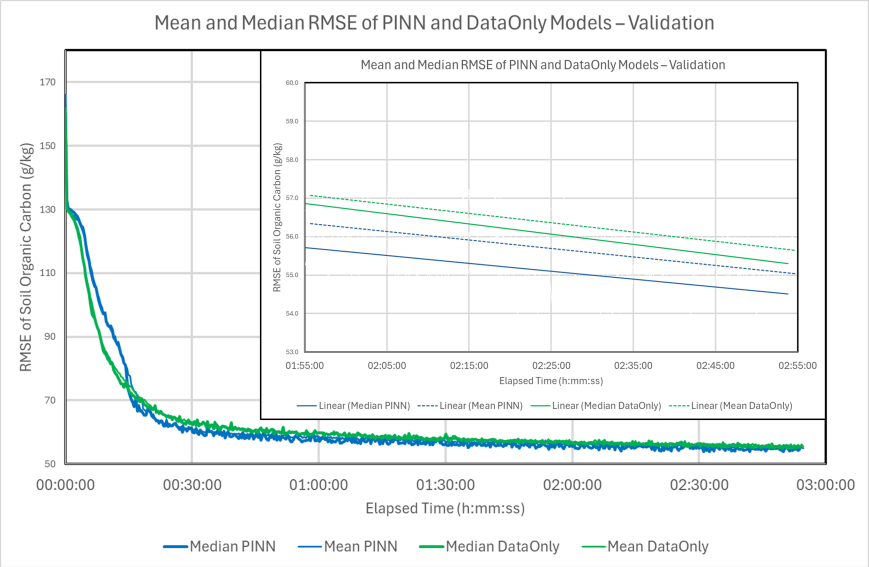

Fig. 2: Median and Mean RMSE of PINN and DataOnly Inset shows the smoothed median and mean RMSE over the last hour of training.

We consider the source dataset to be noisy for two reasons:

1. Individual soil and vegetation indices are imperfect predictors of soil organic carbon due to soil complexity [4, 16].
2. Unlike other studies (*e.g.* [25]), the LUCAS dataset was intentionally not cleaned in order to preserve any gross errors that might be present.

Our novel PINN method is more robust than the standard DataOnly method. The standard deviation of RMSE at the end of each validation batch was calculated across the 10 trials for both the PINN method and the DataOnly method. Fig. 3 shows that the PINN method had a significantly lower standard deviation than the DataOnly method even though both methods used the same training and validation data split for each trial. Our novel PINN method thus appears to be more robust and less susceptible to noise in comparison to the DataOnly method.

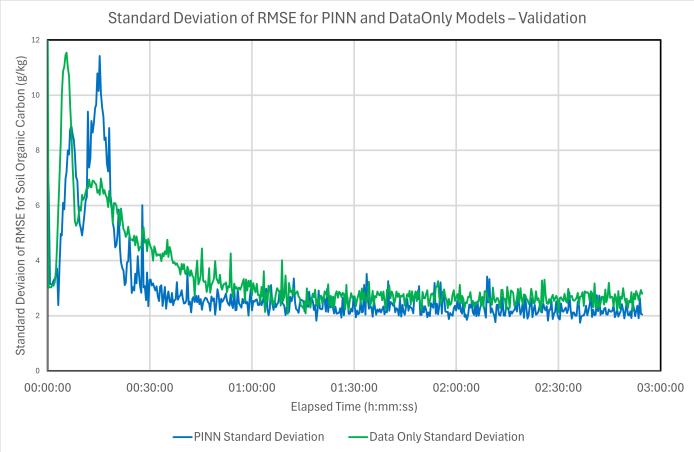

Fig. 3: Standard deviation of RMSE for PINN and DataOnly Methods.

Tab. 2 shows that showed a substantially lower standard deviation (calculated at the end of each validation batch) for the PINN method when compared to the DataOnly method and so the dataset can be assumed to be noisy. Moreover, the LUCAS dataset was intentionally not cleaned in order to preserve any gross errors that might be present. The PINN method thus appears to be more robust to noise in comparison to the DataOnly method.

The computational time of the PINN model is comparable to the computational time of the baseline model, as shown in Fig. 2. This result is expected due to the low additional complexity of calculating the physics component of the PINN. As the PINN model leads to quicker convergence (Fig. 2) it is potentially more efficient than the baseline model.

5.3 Operation of the PINN Method

The PINN loss function has two components, the data only component and the physics component, as described by Eq. (14). The convergence of the Loss function towards its minimum value is shown for one trial in Fig. 4. Initially, the

right-hand component (the physics component) is much smaller than the left-hand component (the data component), so Ω (RHS scale) tends to zero in order to minimize the overall loss function. The data component of the loss decreases due to the effect of model training until such time as it reaches the same value as the physics component. At this time, Ω begins to increase (slowly at first) towards one, causing the influence of the physics component of loss to decrease and the influence of the data component to increase. The solid line in Fig. 4 shows the total loss initially following the physics component (dashed line) and then transitioning to the data component (dotted line), indicating that the PINN performed as intended, using the index to converge to an approximate solution based on the physics associated with indices and reflectance, and then using the detailed characteristics of all the reflectance data to refine that approximate solution.

Changing the initial value of Ω from zero to one had little effect on the operation of the method, as the low initial value of the physics component (compared to the data component) quickly caused Ω to drop to zero, whereupon the method behaved as described above.

When Ω first reaches one, indicating that the PINN method has switched over to only using the data component and is thus equivalent to the Data Only method, the RMSE is significantly lower (Tab. 2). This shows that the physics component of the method – which only uses the index – has guided the method towards a better minimum than the Data Only method. The standard deviation of RMSE is also significantly lower for the PINN method than it for the DataOnly method when Ω first reaches one, showing the improved robustness of the PINN method at the point in time in training where the two methods have an equivalent loss function.

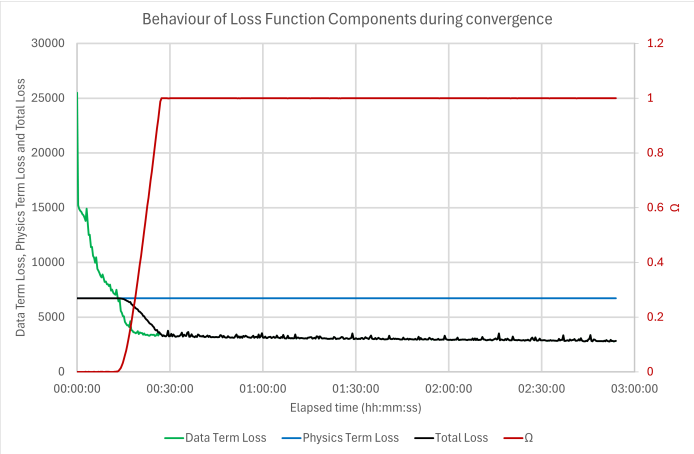

Fig. 4: Behaviour of PINN Loss Function Components during convergence.

Table 2: RMSE of PINN and DataOnly Methods when Ω first reaches one

Method	Trial										Mean	Median	Std Dev
	1	2	3	4	5	6	7	8	9	10			
DataOnly	61.1	70.3	65.0	71.7	75.7	64.2	60.9	73.6	66.6	61.8	67.1	65.8	5.1
PINN	58.6	60.0	62.3	59.7	61.8	61.7	61.6	65.5	59.3	59.0	60.9	60.8	2.0

5.4 Alternate PINN Method

The alternate PINN loss function Eq. (15) was tested using six runs of the method. The alternate PINN method did not provide an accurate solution (Fig. 5).

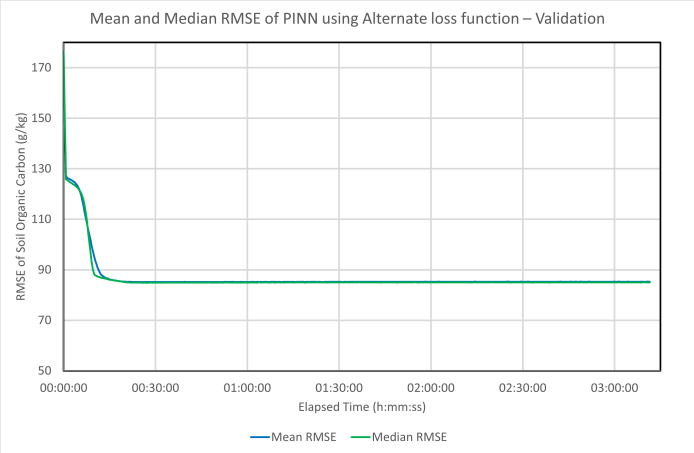

Fig. 5: Mean and Median RMSE of PINN using Alternate loss function.

Fig. 6 shows the value of the physics loss component quickly reduces to about 220 as the method converges towards the solution suggested by the IRECI linear equation. As the data component of the loss function is always greater than the physics component, Ω stays at zero (even if it is initially set at a higher value of 0.5). Thus, the method did not use the data component and RMSE was limited to the value suggested by IRECI. This behaviour is usual in PINNs, and a common way to deal with the matter is to introduce a further factor, λ, to cause an overlap between the data and physics component of the loss function:

$$\mathcal{L} = \Omega \cdot \sum_{i=1}^{N} (SOC_{Actual_i} - SOC_{Model_i})^2 + \lambda \cdot (1-\Omega) \cdot \sum_{i=1}^{N} (SOC_{Model_i} - m \cdot IRECI_i - c)^2 \tag{17}$$

The value of λ is arbitrary and has to be determined on a case-by-case basis for each network, so no exploration of the loss function in Eq. (17) was performed.

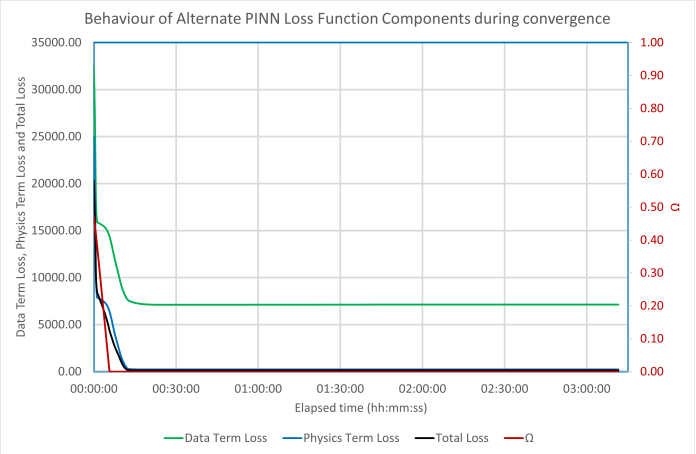

Fig. 6: Behaviour of Alternate PINN Loss Function Components during convergence.

5.5 Summary

The PINN method converges more quickly and achieves a lower RMSE than the standard (DataOnly) method. Moreover, the PINN method is more robust and less susceptible to noise.

Even though the physics term in the loss function of our PINN method does not change during training, the PINN model has a lower RMSE when compared to the baseline model at the 15 and 30 minute marks (just before and after the PINN model loss function transitions from the physics term to the data term).

6 Conclusion

In this paper we developed a novel PINN method that was capable of estimating SOC from multispectral images by employing the relationship between a soil and vegetation index and SOC in the loss function. Our proposed method converged more quickly, is more accurate and is more robust than a standard CNN method having an identical architecture. Our proposed method switches during training from an index-based approach to a data-based approach to overcome the noisy and imprecise relationship between the soil and vegetation index and SOC. Enhancements to our method could include the use of multiple indices rather than a single index. Further research is required to determine the performance of this method with a reduction in the size of the training dataset.

Acknowledgements The work has been supported and funded by the *Cooperative Research Centre for High Performance Soils* whose activities are funded by the Australian Government's Cooperative Research Centre Program, with contribution from Federation University Australia.

References

1. Acharya, U., Lal, R., Chandra, R.: Data driven approach on in-situ soil carbon measurement. CARBON MANAGEMENT **13**(1), 401–419 (2022). https://doi.org/10.1080/17583004.2022.2106310
2. Badora, M., Bartosik, P., Graziano, A., Szolc, T.: Using physics-informed neural networks with small datasets to predict the length of gas turbine nozzle cracks. Advanced Engineering Informatics **58**, 102232 (2023). https://doi.org/https://doi.org/10.1016/j.aei.2023.102232, https://www.sciencedirect.com/science/article/pii/S1474034623003609
3. Ben-Dor, E., Inbar, Y., Chen, Y.: The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote sensing of environment **61**(1), 1–15 (1997). https://doi.org/10.1016/S0034-4257(96)00120-4, p33 Q60 F04 9732871 ObjectType-Article-2 SourceType-Scholarly Journals-1 ObjectType-Feature-1 content type line 23
4. Budak, M., Günal, E., Kılıç, M., Çelik, I., Sırrı, M., Acir, N.: Improvement of spatial estimation for soil organic carbon stocks in yuksekova plain using sentinel 2 imagery and gradient descent–boosted regression tree. Environmental Science and Pollution Research **30**(18), 53253–53274 (2023). https://doi.org/10.1007/s11356-023-26064-8, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149019766&doi=10.1007%2fs11356-023-26064-8&partnerID=40&md5=935bbbf2cc962937244585318a271fb5, export Date: 04 August 2023; Cited By: 0
5. Cao, F., Gao, F., Guo, X., Yuan, D.: Physics-informed neural networks with parameter asymptotic strategy for learning singularly perturbed convection-dominated problem. Computers & Mathematics with Applications **150**, 229–242 (2023). https://doi.org/https://doi.org/10.1016/j.camwa.2023.09.030, https://www.sciencedirect.com/science/article/pii/S0898122123004157
6. Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B.: Imaging spectroscopy for soil mapping and monitoring. SURVEYS IN GEOPHYSICS **40**(3), 361–399 (2019). https://doi.org/10.1007/s10712-019-09524-0
7. Commission, E., Centre, J.R., Jones, A., Fernández-Ugalde, O., Scarpa, S.: LUCAS 2015 topsoil survey – Presentation of dataset and results. Publications Office (2020). https://doi.org/doi/10.2760/616084
8. England, J.R., Viscarra Rossel, R.A.: Proximal sensing for soil carbon accounting. SOIL **4**(2), 101–122 (2018). https://doi.org/10.5194/soil-4-101-2018, https://dx.doi.org/10.5194/soil-4-101-2018
9. European Space Agency: Sentinel-2 spectral response functions (s2-srf) (2023), https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses
10. Garosi, Y., Ayoubi, S., Nussbaum, M., Sheklabadi, M.: Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of iran. Geoderma Regional **29**, e00513 (2022). https://doi.org/https://doi.org/10.1016/j.geodrs.2022.e00513, https://www.sciencedirect.com/science/article/pii/S2352009422000335
11. Gholizadeh, A., Žižžala, D., Saberioon, M., Borůvka, L.: Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral

imaging. Remote Sensing of Environment **218**, 89–103 (2018). https://doi.org/10.1016/j.rse.2018.09.015, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053836935&doi=10.1016%2fj.rse.2018.09.015&partnerID=40&md5=26c21702d4fb56fb7d2541b5f23594ef, export Date: 03 August 2023; Cited By: 209

12. Guo, H.L., Zhang, R.R., Dai, W.H., Zhou, X.W., Zhang, D.J., Yang, Y.H., Cui, J.: Mapping soil organic matter content based on feature band selection with zy1-02d hyperspectral satellite data in the agricultural region. AGRONOMY-BASEL **12**(9) (2022). https://doi.org/10.3390/agronomy12092111
13. Hateffard, F., Szatmári, G., Novák, T.J.: Applicability of machine learning models for predicting soil organic carbon content and bulk density under different soil conditions. Soil Science Annual **74**(1) (2023). https://doi.org/10.37501/soilsa/165879, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85165428835&doi=10.37501%2fsoilsa%2f165879&partnerID=40&md5=f2e4eca72f08465999d8c9c338818d39, export Date: 04 August 2023; Cited By: 0
14. Kaya, F., Keshavarzi, A., Francaviglia, R., Kaplan, G., Başayiğit, L., Dedeoğlu, M.: Assessing machine learning-based prediction under different agricultural practices for digital mapping of soil organic carbon and available phosphorus (2022). https://doi.org/10.3390/agriculture12071062
15. Liu, B., Wang, Y., Rabczuk, T., Olofsson, T., Lu, W.: Multi-scale modeling in thermal conductivity of polyurethane incorporated with phase change materials using physics-informed neural networks. Renewable Energy **220**, 119565 (2024). https://doi.org/https://doi.org/10.1016/j.renene.2023.119565, https://www.sciencedirect.com/science/article/pii/S0960148123014805
16. Liu, Q., He, L., Guo, L., Wang, M., Deng, D., Lv, P., Wang, R., Jia, Z., Hu, Z., Wu, G., Shi, T.: Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. CATENA **219**, 106603 (2022). https://doi.org/https://doi.org/10.1016/j.catena.2022.106603, https://www.sciencedirect.com/science/article/pii/S0341816222005896
17. Madugundu, R., Al-Gaadi, K.A., Tola, E., Edrris, M., Edrees, H., Alameen, A., Fulleros, R.B.: Estimation of soil organic carbon in agricultural fields: A remote sensing approach. Journal of Environmental Biology **43**(1), 73–84 (2022). https://doi.org/10.22438/jeb/43/1/MRN-1873, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122220559&doi=10.22438%2fjeb%2f43%2f1%2fMRN-1873&partnerID=40&md5=552edc18dab4d38c845aea5708e250eb, export Date: 02 August 2023; Cited By: 1
18. Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., 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 (2019). https://doi.org/https://doi.org/10.1016/j.geoderma.2019.06.016, https://www.sciencedirect.com/science/article/pii/S0016706119300588
19. Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H.: A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and sar data fusion. Science of The Total Environment **804**, 150187 (2022). https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.150187, https://www.sciencedirect.com/science/article/pii/S0048969721052645
20. Odebiri, O., Odindi, J., Mutanga, O.: Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review. International

Journal of Applied Earth Observation and Geoinformation **102**, 102389 (2021). https://doi.org/https://doi.org/10.1016/j.jag.2021.102389, https://www.sciencedirect.com/science/article/pii/S0303243421000969
21. Padarian, J., Minasny, B., McBratney, A.B.: Using deep learning to predict soil properties from regional spectral data. Geoderma Regional **16**, e00198 (2019). https://doi.org/https://doi.org/10.1016/j.geodrs.2018.e00198, https://www.sciencedirect.com/science/article/pii/S2352009418302785
22. Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., Helming, K.: Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? Journal of Environmental Management **330**, 117142 (2023). https://doi.org/https://doi.org/10.1016/j.jenvman.2022.117142, https://www.sciencedirect.com/science/article/pii/S0301479722027153
23. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics **378**, 686–707 (2019). https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045, https://www.sciencedirect.com/science/article/pii/S0021999118307125
24. Shen, Z., Ramirez-Lopez, L., Behrens, T., Cui, L., Zhang, M., Walden, L., Wetterlind, J., Shi, Z., Sudduth, K.A., Baumann, P., Song, Y., Catambay, K., Viscarra Rossel, R.A.: Deep transfer learning of global spectra for local soil carbon monitoring. ISPRS Journal of Photogrammetry and Remote Sensing **188**, 190–200 (2022). https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.04.009, https://www.sciencedirect.com/science/article/pii/S092427162200106X
25. Shen, Z., Viscarra Rossel, R.A.: Automated spectroscopic modelling with optimised convolutional neural networks. Scientific Reports **11**(1) (2021). https://doi.org/10.1038/s41598-020-80486-9, https://dx.doi.org/10.1038/s41598-020-80486-9
26. Tsakiridis, N.L., Keramaris, K.D., Theocharis, J.B., 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 (2020). https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114208, https://www.sciencedirect.com/science/article/pii/S0016706119308870
27. Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., Zalidis, G.: An integrated methodology using open soil spectral libraries and earth observation data for soil organic carbon estimations in support of soil-related sdgs. Remote Sensing of Environment **244**, 111793 (2020). https://doi.org/https://doi.org/10.1016/j.rse.2020.111793, https://www.sciencedirect.com/science/article/pii/S0034425720301632
28. Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D.: Temporal mosaicking approaches of sentinel-2 images for extending topsoil organic carbon content mapping in croplands. International Journal of Applied Earth Observation and Geoinformation **96**, 102277 (2021). https://doi.org/https://doi.org/10.1016/j.jag.2020.102277, https://www.sciencedirect.com/science/article/pii/S030324342030920X
29. Viscarra Rossel, R.A., Behrens, T.: Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma **158**(1), 46–54 (2010). https://doi.org/https://doi.org/10.1016/j.geoderma.2009.12.025, https://www.sciencedirect.com/science/article/pii/S0016706109004315

30. Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O.: Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma **131**(1), 59–75 (2006). https://doi.org/https://doi.org/10.1016/j.geoderma.2005.03.007, https://www.sciencedirect.com/science/article/pii/S0016706105000728
31. Wadoux, A.M.J.C.: Using deep learning for multivariate mapping of soil with quantified uncertainty. Geoderma **351**, 59–70 (2019). https://doi.org/10.1016/j.geoderma.2019.05.012
32. Wimmera CMA: Wimmera cma (2024), https://wcma.vic.gov.au/
33. Xu, X., Du, C., Ma, F., Qiu, Z., Zhou, J.: A framework for high-resolution mapping of soil organic matter (som) by the integration of fourier mid-infrared attenuation total reflectance spectroscopy (ftir-atr), sentinel-2 images, and dem derivatives. Remote Sensing **15**(4) (2023). https://doi.org/10.3390/rs15041072, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149255932&doi=10.3390%2frs15041072&partnerID=40&md5=429e173fa9b005ae1285d5847b67434f, export Date: 01 August 2023; Cited By: 2