*Research Article*

# Identifying the Underlying Factors and Variables Governing Macronutrients in Cultivated Tropical Peatland Using Regression Tree Approach

## *Identifikasi Faktor dan Peubah Dasar yang Mengontrol Unsur Hara Makro pada Lahan Gambut Tropis Budidaya Menggunakan Pendekatan Pohon Regresi*

**Heru Bagus Pulunggono[1]*, Yusuf Azmi Madani[2], Moh Zulfajrin[2], Yusrizal[3,4]**

[1] Department of Soil Science, Faculty of Agriculture, IPB University, Bogor, 16680, West Java, Indonesia

[2] Bachelor of Agriculture, Department of Soil Science, Faculty of Agriculture, IPB University, Bogor, 16680, West Java, Indonesia

[3] Graduated School of Soil Science, IPB University, Bogor, 16680, West Java, Indonesia

[4] Department of Agrotechnology, Faculty of Agriculture, Teuku Umar University, Meulaboh, 23615, West Aceh, Indonesia

*email: heruipb@yahoo.co.id

Recived: July 2022
Accepted: August 2022
Published: August 2022

p-ISSN: 2723-7974
e-ISSN: 2723-7966
doi: 10.52045/jca.v3i1.353

Website: https://ojs.untika.ac.id/index.php/faperta

***Abstract:*** The capability of machine learning/ML algorithms to analyze the effect of human and environmental factors and variables in controlling soil nutrients has been profoundly studied over the last decades. Unfortunately, ML utilization to estimate macronutrients and their governing factors in cultivated tropical peat soil are extremely scarce. In this study, we trained regression tree/RT, ML-based pedotransfer models to predict total N, P, and K in peat soils based on oil palm/OP and OP+bush datasets. Our results indicated that the dataset might contain outliers, non-linear relationships, and heteroscedasticity, allowing RT-based models to perform better compared to multiple linear regression/MLR models (as a benchmark) in estimating total N and P in both datasets, contrastingly, not in K. The difference of important variables in each RT-based model partially showed the vital role of land use in nutrient modeling in peat. The depth of sample collection, organic C, and ash content became the prominent factor and variables in regulating the entire predicted nutrients. Meanwhile, the distance from the oil palm tree and pH were the salient features of total P prediction models in OP and OP+bush sites, respectively. This study proposed employing ML-based pedotransfer models in analyzing and interpreting complex tropical peat data as an alternative to linear-based regression. Our initial study also shed more light on the development possibility of the pedotransfer models that agricultural practician, researchers, companies, and farmers can use to predict macronutrients, both in tabular and spatial terms, in cultivated tropical peatlands.

*Keywords:* artificial intelligence, GINI relative importance, machine learning, ICE-PDP

***Abstrak:*** Kemampuan algoritma pembelajaran mesin/ML untuk menganalisis pengaruh faktor dan variabel manusia dan lingkungan dalam mengontrol unsur hara dalam tanah telah dipelajari secara mendalam selama beberapa dekade terakhir. Sayangnya, studi mengenai penggunaan ML untuk memprediksi kandungan unsur hara makro dan berbagai faktor pengontrolnya dalam tanah gambut yang dibudidayakan sangatlah jarang. Dalam kajian ini, kami melatih model pedotransfer berbasis pohon regresi/RT ML untuk memprediksi total N, P, dan K menggunakan dataset kelapa sawit/OP dan OP+semak pada lahan gambut. Hasil analisis kami mengindikasikan bahwa dataset yang digunakan mengandung outlier, hubungan non-linier, dan heteroskedastisitas, yang memungkinkan model berbasis RT berkinerja lebih baik dibandingkan model regresi linier berganda/MLR (sebagai referensi) dalam mengestimasi total N dan P di kedua dataset. Tren yang berkebalikan terjadi pada prediksi K total. Perbedaan peubah penting pada setiap model berbasis RT secara parsial menunjukkan peran vital penggunaan lahan dalam pemodelan hara di gambut. Kedalaman pengambilan sampel, C organik dan kadar abu menjadi faktor dan peubah utama yang mengatur seluruh unsur hara yang diprediksi. Sementara itu, jarak dari pohon kelapa sawit dan pH merupakan peubah yang menonjol dari model prediksi P total secara berurutan pada lokasi OP dan OP+bush. Studi ini mengusulkan penggunaan model pedotransfer berbasis ML sebagai alternatif dari regresi berbasis linier dalam menganalisis dan menafsirkan data gambut tropika yang kompleks. Studi awal

***Citation:***
Pulunggono HB, Madani YA, Zulfajrin M & Yusrizal. 2022. Identifying the Underlying Factors and Variables Governing Macronutrients in Oil Palm Cultivated Tropical Peatland Using Regression Tree Approach. *Celebes Agricultural*. 3(1): 43-61. doi: 10.52045/jca.v3i1.353

kami juga menjelaskan kemungkinan pengembangan model pedotransfer yang dapat digunakan oleh praktisi, peneliti, perusahaan pertanian, dan petani untuk memprediksi unsur hara makro, baik secara tabular maupun spasial, padahal lahan gambut tropis yang dibudidayakan.

*Kata kunci:* kecerdasan buatan, tingkat kepentingan relatif berbasis GINI, pembelajaran mesin, ICE-PDP

## INTRODUCTION

Tropical peatlands hold great importance in the agricultural sector of Southeast Asia/SEA. Due to the rising scarcity of mineral soil areas, farmers and plantation companies in the SEA lowland areas are currently utilizing peatlands as their second option for farming. In Indonesia, peatlands became the essential natural resource, particularly for the people who lived in the coastal lowland of Eastern Sumatera, as well as West and Central Kalimantan. In general, they are benefitted from OP smallholding or plantation companies (Jelsma et al., 2017; Pacheco et al., 2017; Bou Dib et al., 2018). Ditjen Perkebunan (2011) reported that about 20 percent of Indonesian OP plantations were located in peatland areas. Compared to other crops, a well-managed OP plantation could also support a complex ecosystem that harbors relatively high biodiversity (Koh, 2008; Hood et al., 2020; Pashkevich et al., 2022).

Determining the amounts of N, P, and K in peat are essential to OP plantation in obtaining and sustaining optimum yields. They are the major nutrients that OP plants essentially absorb at high levels (Wahid et al., 2005; Woittiez et al., 2017). The NPK-contained fertilizer demands are so high for the OP plantations since the natural concentration of the nutrients in peat soil does not meet the requirement of OP to produce an economically feasible yield (Ng et al., 1990; Pardon et al., 2016). The parent material of peats contained less P and K contents than mineral soils. Owing to a slower release of peat mineralization process, available N in peat is under the OP sufficiency levels for producing high yields, which requires an additional N fertilization (Mutert et al., 1999; Ariffin et al., 2019). Consequently, under-fertilized OP often undergoes a hidden hunger and yield below the average. At the same time, the soil nutrients were continuously mined, thus may leading to land degradation (Majumdar et al., 2016; Woittiez et al., 2018; Woittiez et al., 2019).

Conducting optimal fertilization in OP plantations requires soil and plant samplings, laboratory determination, data analysis, and reporting. These first and second procedures are considered arduous and tedious, especially for large-scale companies. Furthermore, laboratory determination resulted in a negative environmental impact since the destructive analyses require hazardous chemical extractants. Modeling techniques based on the pedotransfer function provide an alternative pathway to estimate soil nutrient contents using soil and environmental-related factors and variables (or predictors/covariates; Van Looy et al., 2017; Grunwald et al., 2021). Linear-based regression model employed with an ordinary least square/OLS algorithm was a common example. However, problems emerge while the models only rely on a linear approach in determining nutrients that possess complex and non-linear relationships with their predictors. The interactions among soil, microorganism, and plant roots were very complex with non-linear interrelationships and highly collinear (Vereecken et al., 2016; York et al., 2016; Gregory et al., 2022).

The complexity of the problems of the linear approach was also exacerbated by the variability of nutrient concentrations and characteristics in the spatial perspective, both laterally and vertically, as affected by management zonation at the block level (Nelson et al., 2015; Pulunggono et al., 2022a). Specific zone or microsite occurs on OP plantation following best management practices/BMP (RSPO, 2012). Moreover, the entire microsites had their specific ecosystem functioning, particularly in providing nutrients to the peat. For instance, amelioration and fertilization are carried out in the fertilization/weeded cycle, whereas pruned fronds are stacked in the row between the OP tree. Moreover, the interrows and stacked fronds are often left overgrown by the understory's fast-growing weeds or planted with leguminous land cover crops (Ariffin et al., 2019; Luke et al., 2019; Agus et al., 2019). Based on that, total N, P, and K at the surface peat near oil palm tree were likely higher than the farther distance or deeper layer due to fertilizer addition and peat mineralization in the fertilization circle. Total N might follow these trends to some length. However, the decomposition of understory detrital materials at the farthest distance from the oil palm tree might prohibit the decreasing pattern. The decreasing trends of K were obstructed by lateral and vertical leaching since K is considered highly mobile (Sparks, 2003). Meanwhile, P was an immobile nutrient in the soil (Sparks, 2003), which may follow these distance and depth trends. Hence, the complex site-specific micro-ecosystems and interrelationships explained above can not be simplified by performing linear-based regression.

During the last decades, machine learning/ML, a part of artificial intelligence/AI, was rapidly employed as the main algorithm for soil (Padarian et al., 2020; Grunwald et al., 2021), agriculture (Meshram et al., 2021), and environmental modeling (Zhong et al., 2021). ML algorithms have an exceptional ability to automatically learn the complex pattern from data, store the learned information, and use it to improve their performance when learning a new task. This capability is considered more powerful and flexible than commonly widespread modeling algorithms, such as OLS employed by linear-based regression, particularly for the model that consists of non-linear interrelationships between its variables (Jordan & Mithel, 2015; Grunwald et al., 2021). Several researchers reported promising performance of the ML-based regression (or pedotransfer) model in predicting or mapping soil properties in tropical mineral soil, such as cation exchange capacity (Teixeira et al., 2020), soil organic carbon (Gomes et al., 2019; Poppiel et al., 2019), soil macro- and micronutrients (Pulunggono et al., 2022b), and soil acidity (Teixeira et al., 2020; Pulunggono et al., 2022b). The regression tree (RT) algorithm is one of the firstly developed ML (Breiman et al., 1984; see review by Loh, 2014). RT was considered successful in modeling several chemical and physical characteristics of mineral soils compared to linear-based models and their derivatives (Vega et al., 2009; Pinheiro et al., 2018; Qiu et al., 2016; Rawal et al., 2019). Despite their widespread usage in other fields of study, RT-based pedotransfer models' capability to estimate soil nutrients using tropical peat data was understudied.

Based on the previous information, we hypothesized that the nutrient variability was less controlled by land use since there are differences in management both in OP and its surrounding bush environment, yet considering they might undergo similar hydrological regimes. Concerning the ability of RT to handle complex and non-linear interrelationships among soil variables, we believed that RT-based N, P, and K models are more convenient compared to MLR, regardless of the data types. Moreover, the depth of sample collection could be used as a general factor for the entire data and for each different land use due to the nutrient enrichment or concentration at the surface either from fertilizer addition (OP data) or peat mineralization (OP+bush data). Lastly, the effect of distance from the oil palm tree and the depth of the sample collection was constrained heavily in OP-based models. Therefore, this study intends to determine the underlying factors governing total N, P, and K in OP cultivated tropical peatlands and their surrounding environment using a common machine learning approach, namely the regression tree/RT method.

## MATERIALS AND METHODS

### Sampling Site and Laboratory Analyses

This study was conducted in oil palm cultivated peatland, Buatan Village, Koto Gasip District, Siak Regency, Riau Province, Indonesia, as shown in Figure 1. Field sampling and laboratory analysis were carried out from January 2020 to June 2021. This study used three factors with an unbalanced design. The first factor was land use, consisting of oil palm/OP and bush. The bush was located adjacent to the OP area, allowing the possible impact of OP management (*e.g.*, drainage, groundwater level/GWL) to influence the bush's soil system. The second factor was the depth of the sample collected, which consisted of 0-20, 20-40, and 40-60 cm depths. OP land use was furthermore divided based on three distances from the oil palm tree, namely 1.5, 3, and 4.5 meters. Approximately 2 kg of peat materials were compositely sampled using an eijelkamp hand auger of 50 cm in length. All samples were 120 in total, where 90 and 30 samples were taken from OP and bush sites, respectively.

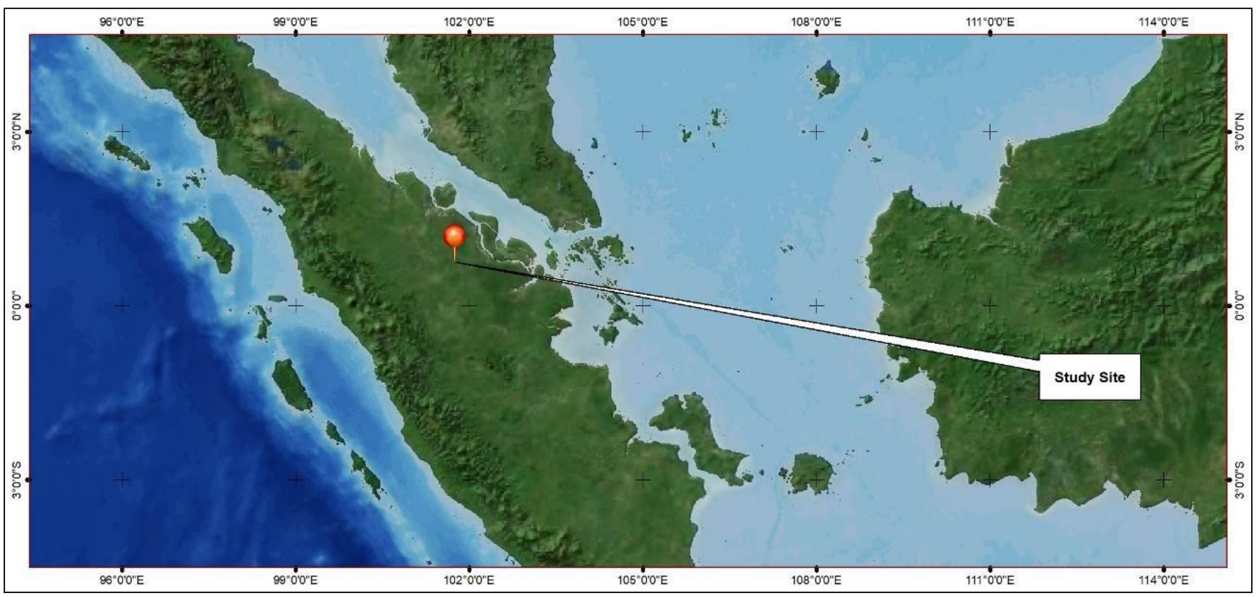



**Figure 1.** The study location

The laboratory analysis was carried out in Soil Chemistry and Fertility Laboratorium, Department of Soil Science and Land Resource, Faculty of Agriculture, IPB University. The analyzed peat chemical properties were pH, ash content, and organic C, as well as total N, P, and K. Peat pH was measured using $H_2O$ with a ratio of 1:5. Ash content and organic C were analyzed using the lost on ignition/LoI method. Total N was determined using the Kjeldahl method, whereas total P and K were determined using HCl25%.

## Statistical analyses

All graphical contents and analyses were conducted in the RStudio environment (R Core Team, 2022), except for correlation heatmaps. The packages employed in this study were tidyverse (Wickam, 2022), rpart (Therneau and Atkinson, 2022), caret (Kuhn et al., 2022), vip (Grenwell et al., 2020), rpart.plot (Milborrow, 2022), Metrics (Hamner et al., 2018), and ModelMetrics (Hunt, 2020) packages. Firstly, the tidyverse package was used to build boxplots. Then, the RT algorithm was executed utilizing the rpart and caret packages. Next, the vip package was used to extract and visualize the variable importance, partial dependence/PD, and individual conditional expectation/ICE plots. Lastly, to visualize the splitting procedure of RT after pruned, we used a rpart.plot package. Metrics and ModelMetrics packages were used to calculate model performance metrics. Furthermore, correlation heatmaps were visualized in Minitab software using the Pearson product-moment correlation method.

Before the analysis was run, the entire values (factors and variables) were qualitatively inspected, resulting in no extreme value. Therefore, we avoid the quantitative outlier detection since its consideration might be biased with respect to the high variability of tropical peat data. Moreover, we visually inspected each variable's maximum and minimum values to ensure they were inside the common ranges. Then, we run the analysis of variance and Tukey Honestly Significant Difference/HSD, as well as Pearson correlation tests to provide an initial indication of predictors' relationships with estimated nutrients and potential collinearity among the predictors. Furthermore, the datasets were randomly split into training (70%) and validation (30%) data (`seed 42`). The models were then assembled using the entire data (OP+bush) and OP data to predict the total N, P, and K in peat. This study did not consider bush data for the third model due to insufficient observation. OP+bush models involved the depth sample collection, pH, organic C, and ash content as their predictors. Moreover, OP models consisted of similar predictors to the previous models with the addition of distance from the oil palm tree. The entire predictors/covariates were chosen due to their widely known relationships with N, P, and K in soils.

This study performed the non-parametric statistical machine learning approach, namely regression tree/RT. RT was firstly formulated by Breiman, Friedman, Olshen, and Stone (Breiman et al., 1984). Thanks to its high interpretability, flexibility, and convenient usage, RT was widely used in practice and research (Ahn, 2014; Loh, 2014). Using the "divide and conquer" approach, RT partition the data recursively, diminishing the sum of squared error of model prediction/outcomes at each splits. The RT algorithm continuously undergoes the splitting process until a certain condition is achieved. The particular criterion was developed considering the trade-off between local accuracy and model complexity with the outcome's variances and model interpretability (Breiman et al., 1984). This study used the combination of cost complexity (`cp`), minimum observation in each node (`minsplit`), and maximum depth of the growing tree (`maxdepth`) parameters to penalize the tree from growing too deep (or 'pruning'), thereby preventing model overfitting. To select the best parameterization of each model, we used the grid search of a certain range of minsplit and maxdepth combined with the ten-fold cross-validation (`xval= 10`), as summarized in Table 1. This study also performed standard multiple linear regression (MLR) as a benchmark compared to RT's model performances, which selected RMSE, MAE, and Bias as the metrics. The relative variable importances were calculated using the Gini importance/GI or mean decrease in impurity/MDI approach for RT models and the absolute value of the t-statistic for MLR models (Grenwell et al., 2020). To gain more clearance on RT-based model interpretability, the marginal effects of each variable/predictor on the predicted nutrients were computed and visually plotted using partial dependence/PD and individual conditional expectation/ICE curves.

**Table 1.** Model parameterization of RT algorithm specified for each model employed in this study

| Model | Minsplit | Maxdepth | cp |
|---|---|---|---|
| *All (OP+Palm) Dataset* | | | |
| N | 5 | 6 | 0.02 |
| P | 21 | 17 | 0.01 |
| K | 24 | 21 | 0.01 |
| *OP Dataset* | | | |
| N | 7 | 7 | 0.03 |
| P | 11 | 4 | 0.01 |
| K | 18 | 16 | 0.01 |

## RESULTS

### Chemical Properties of the Studied Peat

The results of peat chemical analyses are summarized in Figure 2. The figures showed that the entire chemical variables were in the range of common values found on peat, as reported by other researchers (*e.g.*, Dhandapani et al., 2020). It could be observed that both land uses seemingly had similar ranges of soil chemical variables, overlapping one another. Other factors also exhibited similar trends, except for total P and K. Both variables were likely higher at the peat surface compared to the lowest sampling depth. However, the peat sampled at 20-40 cm had values in the range of other sampling depths.

The observed means of organic C were higher in the bush site compared to the OP site. However, contrasting patterns were observed in pH, ash content, and total P. Similar means at both sites were exhibited by total N and K. Based on the distance from the oil palm tree, ash content and total N showed relatively similar trends. Their means were increased along with the increasing distance up to three meters. Then, they dropped again at their farthest distances.

In contrast, the organic C content of the studied peat showed the pattern oppositely to the previously described pattern of other variables. They denoted the lowest mean at three meters from the oil palm tree, whereas the first and the last were relatively high. The resemblances also likely occurred for the ranges and means of ash content, total N and P sampled at the bush and the farthest distance from the oil palm tree. Obstructed effects of the distances from oil palm trees were exhibited by pH and total K, which showed relatively flat curves along the increase of the distance from the oil palm tree. Based on sampling depth, pH, ash content, and total P and K showed decreasing trends along with the increase in sampling depth. Opposite trends were exhibited by organic C and total N.

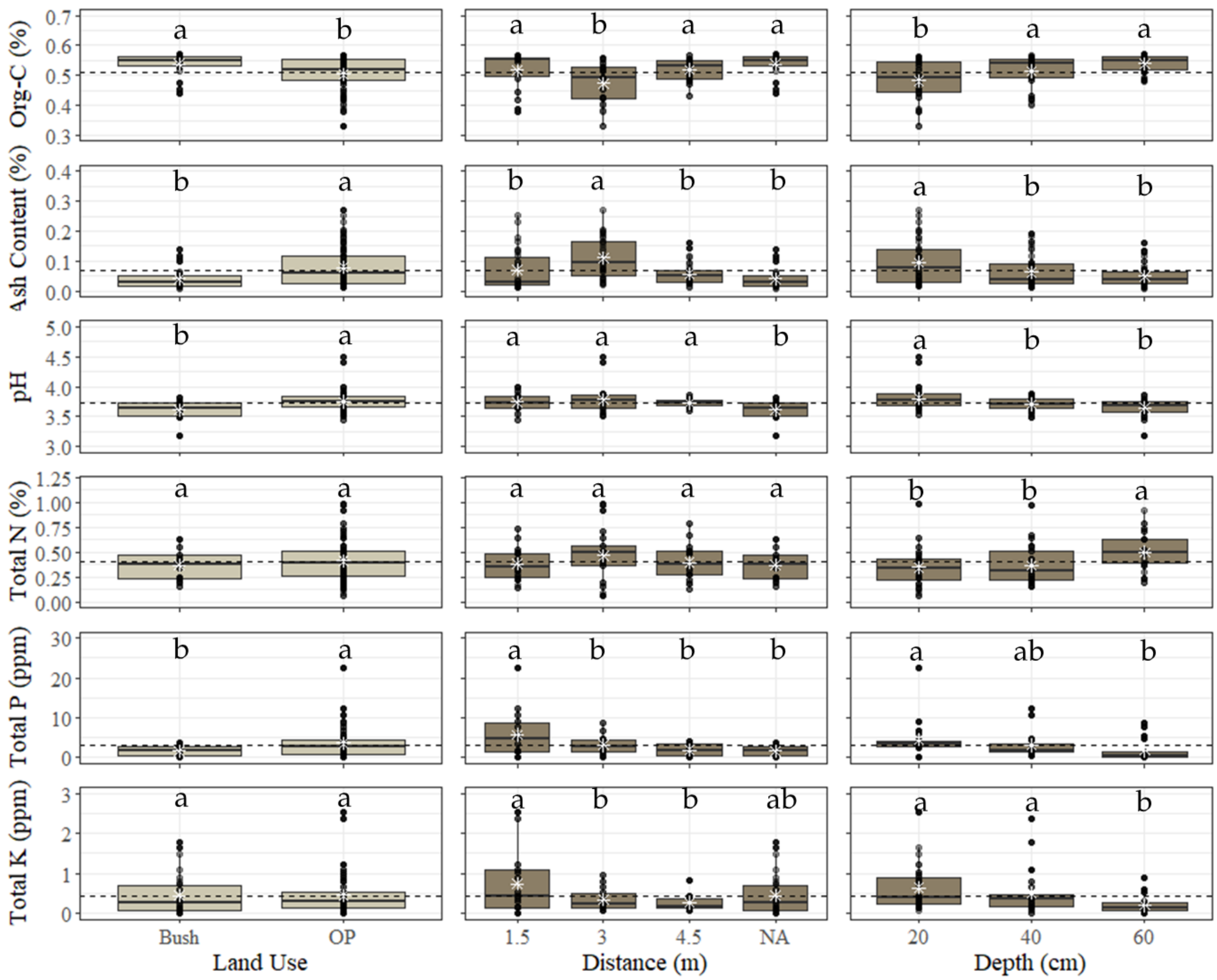



**Figure 2.** Chemical characteristics of studied peat based on land use, distance from the oil palm tree, and depth of sample collected. Note that NA's in distance graphs indicated the samples collected outside the OP area. The white asterisks and black horizontal lines inside the boxplot indicate mean and median values, respectively, whereas the black dotted lines indicate means from the entire data. Depth, Distance, and Org-C represent the depth of sample collection, distance from the oil palm tree, and organic C, respectively. The letter above the boxplots indicated differences concerning Tukey HSD test.

The correlation heatmaps are shown in Figure 3. All correlations were low to moderate on the entire variables, leaving only organic C and ash content with high correlations. OP data also showed similar relationship strengths. Total N generally exhibited a low correlation with other factors and variables, except for depth which had a moderately positive correlation in both OP+bush and OP data. Using OP+bush data, total P and K had moderate correlations with all factors and variables. Meanwhile, similar relationship strengths were denoted by the same variables using OP data, except for their insignificant strength with organic C and ash content (only for total P).

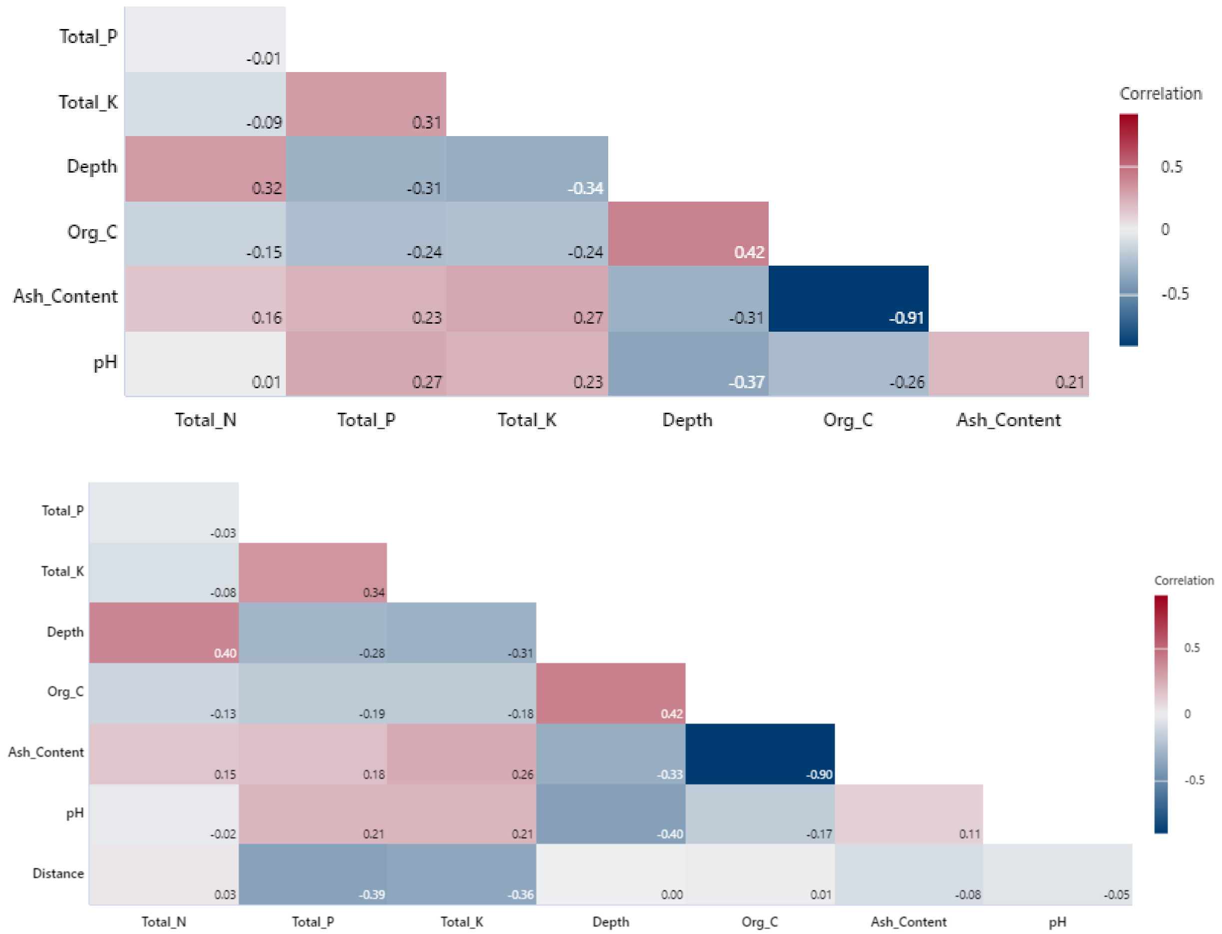



**Figure 3.** Correlation matrix supplied to heatmaps of all variables used in this study. The upper graph represents all data (OP+bush; N=120). The lower graph represents only OP data (N=90). Depth, Distance, and Org_C represent the depth of sample collection, distance from the oil palm tree, and organic C, respectively.

## RT-based Pedotransfer Model Evaluation

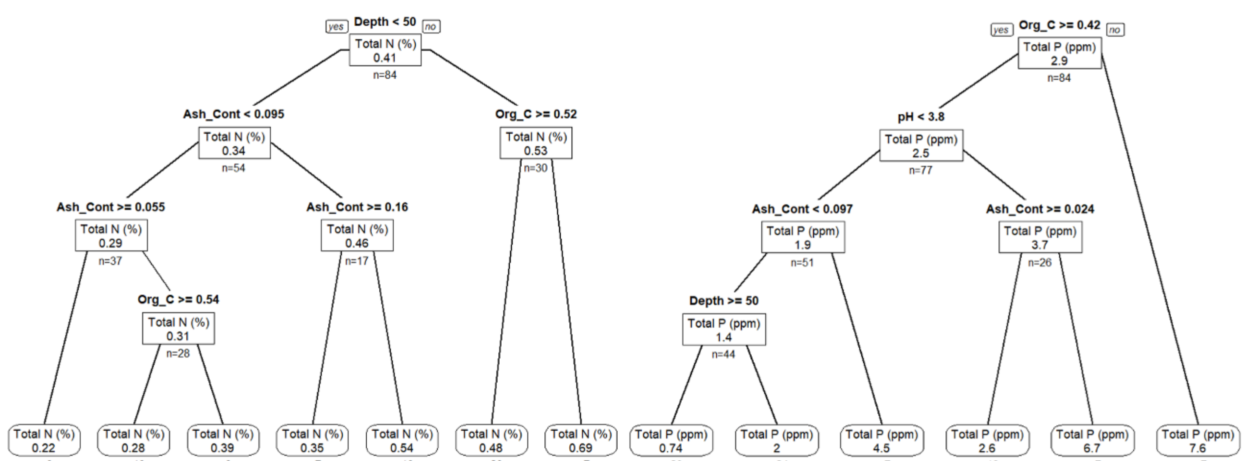
The model evaluation for the entire models employed in this study was presented in Table 2. In general, RT prediction of total N and P showed a slight improvement compared to MLR. However, the entire metrics of RT's models of total K exhibited poor performance than MLR. Generally, OP+bush models showed the predicted value of RT models overestimated the observed values. Oppositely, the contrasting patterns were denoted by OP-based models. Besides their huge algorithm differences, MLR and RT showed the same over-and underestimation patterns in bias regardless of the dataset types.

**Table 2.** Model evaluation from the entire RT models performed in this study in comparison with MLR models

| Dataset | Algorithm Type | Predicted Output | RMSE | MAE | Bias |
|---|---|---|---|---|---|
| OP+Bush | MLR | Total N | 0.18358 | 0.15554 | -0.01506 |
| | | Total P | 0.19548 | 0.12016 | 0.02820 |
| | | Total K | 0.64168 | 0.41581 | 0.11363 |
| | RT | Total N | 0.19149 | 0.14406 | -0.01017 |
| | | Total P | 0.19017 | 0.11210 | 0.02291 |
| | | Total K | 0.66864 | 0.44671 | 0.07962 |
| OP | MLR | Total N | 0.17566 | 0.14446 | -0.01823 |
| | | Total P | 4.37354 | 2.78739 | -0.11845 |
| | | Total K | 0.53507 | 0.36313 | 0.03928 |
| | RT | Total N | 0.17350 | 0.13196 | -0.03047 |
| | | Total P | 4.16727 | 2.75208 | -0.02595 |
| | | Total K | 0.61454 | 0.41430 | 0.03388 |

## RT-based Pedotransfer Model Interpretation

The RT approach in predicting total N, P, and K was shown in figures 4 (OP+bush data) and 5 (OP data), whereas the MLR model summary was presented in Table 3. The results showed that the entire OP+bush models contained the organic C, ash content, and depth of sample collection as their primary splitting rules. Meanwhile, ash content was a common variable in the entire OP-based models. Therefore, organic C and depth were considered variables that primarily influenced total N and P predictions in the OP area. Furthermore, the RT algorithm considered the distance from the oil palm tree as the factor that specifically influenced total P prediction. However, we did not observe similar effects on total N and K at the same site. When utilizing OP+bush data, both RT and MLR models agreed on depth as one of their building block variables, regardless of the predicted nutrients. In addition, pH and distance from the oil palm tree were involved in both RT and MLR model development.



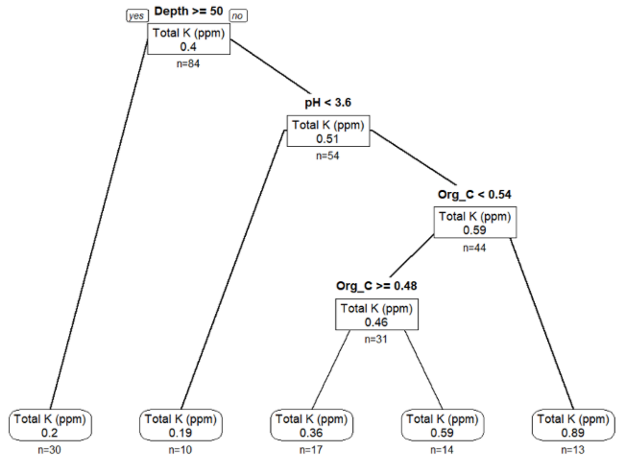



**Figure 4.** RT diagrams of the total N (upper-left), P (upper-right), and K (lower) using OP+bush data. Note that the values inside the boxes are means. The values outside the boxes represent the number of observations concerning the total amount of training data.

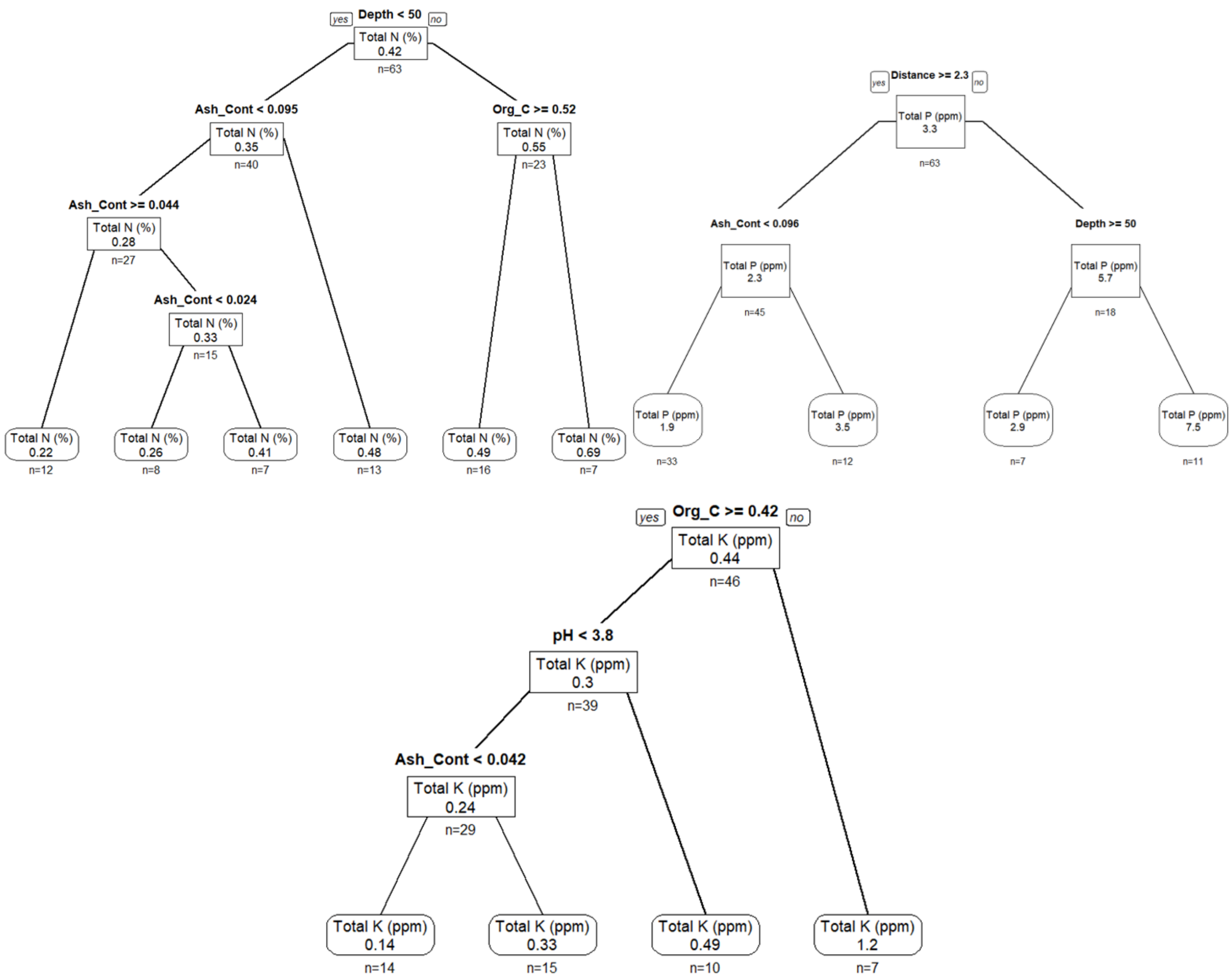



**Figure 5.** RT diagrams of the total N (upper-left), P (upper-right), and K (lower) using OP data. Note that the values inside the boxes are means. The values outside the boxes represent the number of observations concerning the total training data.

**Table 3.** Summary of the entire MLR models used in this study

| Factors | OP+Bush Dataset | | | OP Dataset | | |
|---|---|---|---|---|---|---|
| | Total N | Total P | Total K | Total N | Total P | Total K |
| Intercept | -0.043002 | -0.11077 | -1.703783 | 0.134981 | 5.91966 | -0.497520 |
| Distance | - | - | - | -0.005788 | **-1.32472** | **-0.128856** |
| Depth | **0.006764** | **-0.07056** | **-0.011289** | **0.320700** | -0.05087 | -0.004825 |
| Org_C | -0.429896 | 0.16669 | 2.292648 | -0.027073 | 2.55642 | -0.601756 |
| Ash Content | 0.716152 | 0.20206 | 1.138609 | 0.265964 | 14.75097 | 2.670156 |
| pH | 0.094643 | **0.23339** | 0.353830 | 0.146463 | 0.34123 | 0.426085 |

**Note:** Depth, Distance, and Org_C represent the depth of sample collection, distance from the oil palm tree, and organic C, respectively. Bold values indicated statistical significance using 90% confidence intervals.

Corroborating the previous results presented in Figures 4 and 5, the Gini/MDI-based relative variable importance of RT models in Figure 6 revealed that organic C was generally the most critical predictor in OP+bush or OP models. Moreover, ash content, depth, and pH were respectively the second, third, and fourth variables of RT models that were important to predict total N, P, and K using OP+bush and OP data. This study also found that the distance from the oil palm tree was the least important variable in predicting total N and K in the OP site. Particularly for total K prediction using OP data, we observed that both the distance from the oil palm tree and the depth of sample collection were considered unimportant, shown by their lowest relative importance compared to other variables. Unlike RT models, the organic C content showed less contribution to the MLR models developed on OP+bush or OP data.

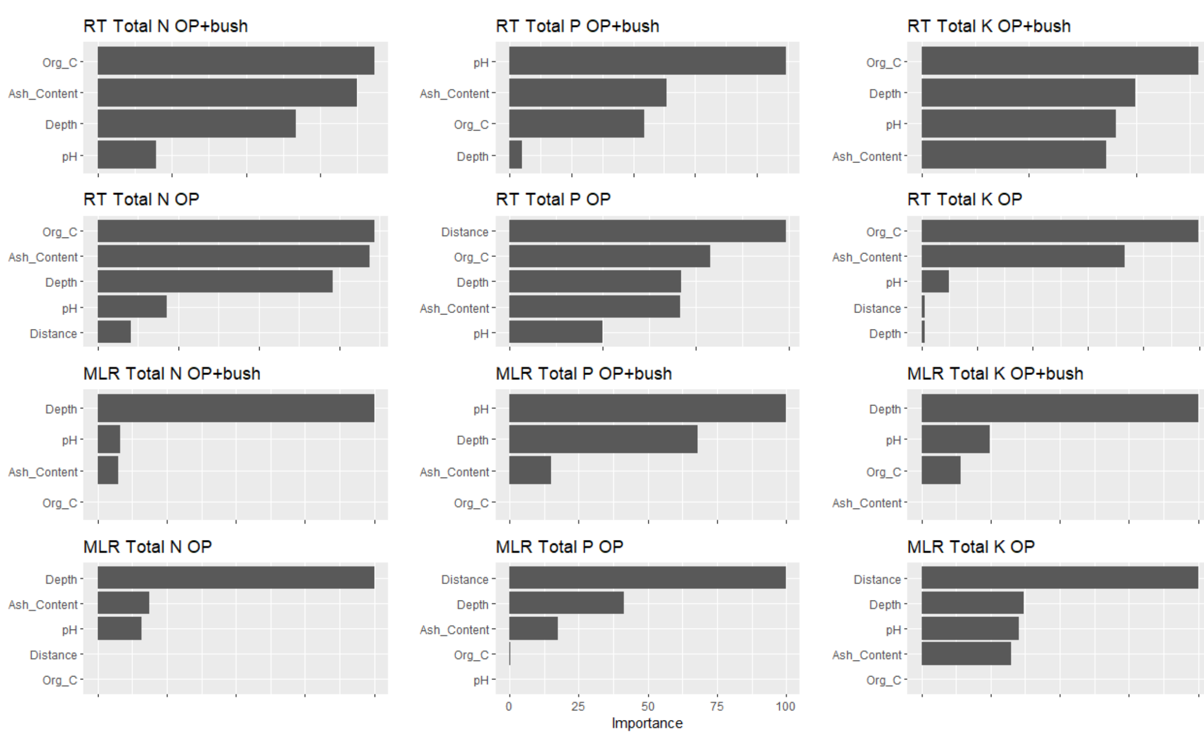



**Figure 6.** Relative variable importance from the entire models performed in this study.

The PD and ICE plots for the entire RT models employed in this study were presented in Figure 7. The figure showed that most variables exhibit non-linear relationships with the predicted nutrients (total N, P, and K; expressed as yhat). However, the marginal effect of PD and ICE curves denoted low or no relationships with the predicted nutrients, for example, pH for N prediction using OP+bush data and distance from the oil palm tree for N and K prediction using OP data.

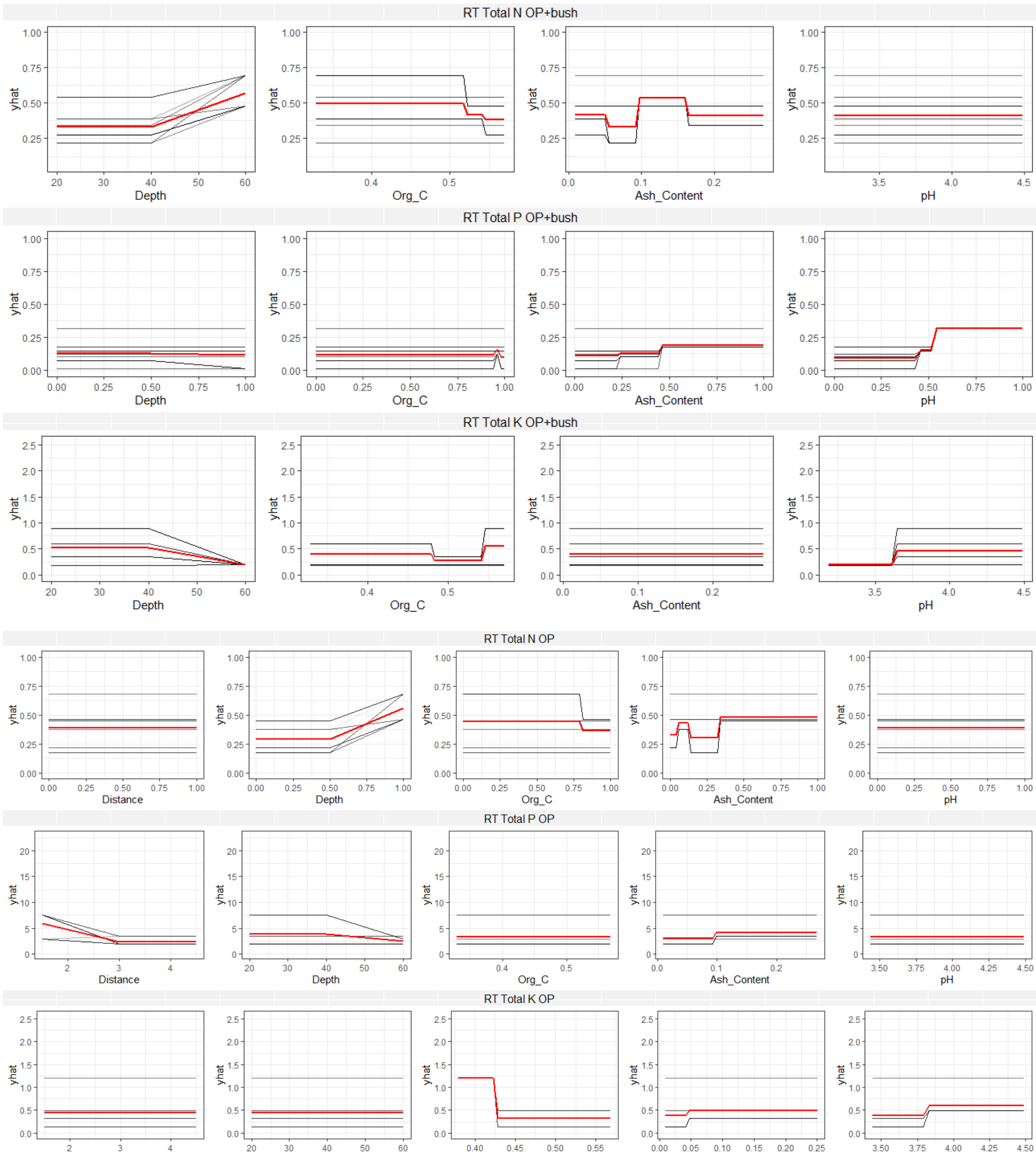



**Figure 7.** PD (red lines) and ICE (black lines) curves of the RT models in predicting total N, P, and K (expressed as yhat in the Y axis) using OP+bush and OP data.

## DISCUSSION

This study emphasized the importance of using a non-parametric ML approach, namely regression tree/RT, to analyze and interpret highly variable tropical peat data. This study also corroborates previous findings and results, reporting the promising performances of several RT-based models over their MLR counterparts.

Our RT approach showed that land use might be a primary factor in estimating total N, P, and K in cultivated tropical peatlands and their surrounding area. This was shown partially by the difference between the factors and variables controlling total N, P, and K in OP and OP+bush sites. Based on the RT approach (Figures 4, 5, and 6), organic C, ash content, and depth were the common variables and factors regulating total N, P, and K in OP+bush and OP sites. Notably, pH possessed high relative importance in developing an RT-based model for total P prediction at the OP+bush site (Figures 4 and 6). Furthermore, among other variables, distance from the oil palm tree became the salient feature that notably governed total P in the OP site (Figures 5 and 6).

The differences in variables controlling nutrients in the soil as affected by different land use can be explained by the difference in management and nutrients' chemical characteristics. The surface peat in the OP site was continuously fertilized and ameliorated along the fertilization circle following best management practices/BMP of the oil palm plantation in peatland (RSPO, 2012; Pauli et al., 2014; Corley & Tinker, 2015). This condition caused the distance from the oil palm tree, and the depth of sample collection becomes important for immobile nutrients such as P. Peat located closer to the oil palm tree might contain a higher P concentration compared to farther distance. Furthermore, a similar condition also undergoes between the surface and subsurface layers, respectively (Fairhurst, 1996; Woittiez et al., 2019). Meanwhile, the lessening effect of both variables on the total K model (Figure 6) was possibly caused by high K mobility. K had less tightened bonding with organic exchange sites than other ions (*i.e.,* Ca, Mg; Sparks, 2003; Tipping, 2004), provoking cation competition and vertical-lateral leachings.

Furthermore, our study's results also proved the "ecological fallacy" that could be made when comparing the difference between individual land use (OP) and the aggregate group (OP+bush) data. The relative strength of the variables in regulating more mobile nutrients, *i.e.,* total N and K prediction, exhibited slight differences between OP+bush and OP data, both in RT and MLR models (Figures 4, 5, and 6). This indicates the behavioral similarities of nutrients and their predictor between individual and aggregate data. These results were possibly due to relatively similar environmental conditions, particularly in hydrology or water management, since the sampling at the bush site was conducted adjacent to the OP site. Moreover, to avoid the fallacy, it can not be assumed that this similarity would also be undergone at the bush site with respect to our observation insufficiency. The difference was evident in immobile nutrients such as P (Figures 4, 5, and 6), suggesting the carefulness in interpreting individual and aggregate data and statistics. These contrasting conditions related to particular nutrients are an interesting subject. However, further study must be conducted using more observations that account for spatial and temporal variability to provide reliable models and gather sufficient evidence.

Similar marginal effects were shown by the distance from the oil palm tree on N prediction in the OP site (Figures 6 and 7). The BMP management allowed the fast-growing weeds (Ashton-Butt et al., 2018; Luke et al., 2019; Rahman et al., 2021) or leguminous plants (Arifin et al., 2015; Agus et al., 2019) to occupy the interrow and frond stack as understory cover crops/UCC. N mineralization derived from UCC's litter decomposition could enhance N in peat outside of the fertilization circle, hence, minimizing the effect of distance. However, the exact impact seemingly did not occur in K due to its low immobilization in the UCC plant (Joo et al., 1994) and poor K source of peat materials (Mohidin et al., 2019; Dhandapani et al., 2020). The obstructed effects of the distance from the oil palm to total N and K in the OP site could also arise while the OP management applies empty fruit bunches/EFB or EFB-based compost outside of the fertilization circle (*e.g.,* Rahman et al., 2021). Composted EFB contained relatively high N and K compared to other nutrients (Mardhiana et al., 2021), whereas EFB ash had K in a substantial amount (Anyaoha et al., 2018). Therefore, the EFB application could boost K in peat at the farthest distance from the oil palm tree, resulting in no distance relationships. Thus, incorporating bush data leads to more obfuscated effects on the sample collection depth since no additional nutrients were added to the peat (Figure 6).

As shown in Figure 2, the entire ranges of total N, P, and K overlap while divided by land use, distance from the oil palm tree, and depth of sampling collection. The obstructed effect would be detected by performing the classical parametric test (*e.g.,* F-test, T-test, MLR; Figure 2; Table 3). MLR capability was seemingly constrained (Table 3) since most of our predictors exhibited non-linear relationships with total N, P, and K, regardless of the dataset differences (Figure 7). Concerning common modeling considerations, all of our models might suffer low observation numbers, noisy variables (Figure 2), inadequate related predictors (Figures 3 and 7), and heteroscedasticity (Figure 3). Nevertheless, most of our RT models showed a relatively robust prediction against the confounding factors described above. In both datasets, RT-based total N and P models gained slight performance improvements compared to MLR (Table 2). The RT models reportedly outperformed MLR and its derivatives (*e.g.,* stepwise backward or forward eliminations, general linear model/GLM) in predicting sorption and retention of heavy metals (Vega et al., 2009), topsoil texture (Pinheiro et al., 2018), soil heavy metal content (Qiu et al., 2016), and base saturation percentages (Rawal et al., 2019). The higher capability of RT compared to MLR was also reported in other fields, such as airborne bacterial hazard assessment (Yoo et al., 2018) and landslide susceptibility (Pourgashemi & Rahmati, 2018).

Improved performances of our RT models' prediction would be achieved when the models are developed using outlier-removed and large quantities datasets. To build proper RT models, redesigning our sampling methodology could also be considered, concerning some gradual changes in total N, P, and K observed at several continuous factors (Figure 2). For example, the peat variables collected at 20-40 cm depth had intermediate values and ranges over their upper and lower depths. Similar patterns were also observed in the variables of peat collected at a 3 m distance from the oil palm tree, which exhibited the intermediate values and ranges of peat variables at near and farther distances. Since the RT algorithm works in partitioning terms, this approach is likely to prefer sharp or abrupt changes of their input variables and suffer from smooth and gradual variable changes (Trontelj ml. & Chambers, 2021). Without removing the intermediate-source classes in attributed factors, more flexible ML-based models (*i.e.,* random forest, gradient boosting regression, support vector regression) must be employed to obtain considerable prediction agreements.

## CONCLUSIONS

This study demonstrated the importance of the ML approach, namely regression tree/RT, in analyzing factors and variables controlling total N, P, and K in tropical cultivated peatlands and their surrounding environment. Our RT approach partially showed that land use might be vital in regulating N, P, and K in our study area. The depth of sample collection was found as an essential factor in predicting total N, P, and K in cultivated tropical peatlands and their surrounding area; meanwhile, a similar influence was exhibited by organic C and ash content. Notably, pH possessed high relative importance in developing an RT-based model for total P prediction at the OP+bush site; the same nutrient was importantly governed by the distance from the oil palm tree in the OP site. Despite the limitation of RT algorithm that prefers sharp or abrupt changes of the input variables values, this study's results indicated that integrating ML-based pedotransfer functions in the soil scientist statistical inventory is essential in analyzing and interpreting complex tropical peat data. Furthermore, our initial study also pinpointed the possibility of the tabular and spatial-based pedotransfer model development for agricultural practician, researchers, companies, and farmers to predict macronutrients in cultivated tropical peatlands.

## ACKNOWLEDGMENT

This research is a part of the $CO_2$ Emission Factor of Peatland Drained for Oil Palm Plantations Research granted by Indonesian Oil Palm Plantation Fund Management. The authors thanked the management and field staff of Astra Agro Lestari Tbk for their assistance during the field campaign.

## REFERENCES

Agus C, Ilfana ZR, Azmi FF, Rachmanadi D, Widiyatno, Wulandari D, Santosa PB, Harun MK, Yuwati TW & Lestari T. 2019. The effect of tropical peat land-use changes on plant diversity and soil properties. *International Journal of Environmental Science and Technology*, 17(3):1703–1712. https://doi.org/10.1007/s13762-019-02579-x

Ahn H. 2014. Fifty Years of Classification and Regression Trees. *International Statistical Review*, 82(3):357–359. https://doi.org/10.1111/insr.12061

Anyaoha KE, Sakrabani R, Patchigolla K & Mouazen AM. 2018. Critical evaluation of oil palm fresh fruit bunch solid wastes as soil amendments: Prospects and challenges. *Resources, Conservation and Recycling*, 136:399–409. https://doi.org/10.1016/j.resconrec.2018.04.022

Ariffin MR, Ahmed OH, Md Isa I & Khairuddin MN. 2019. Effects of moisture level on nitrogen availability in tropical peat soil cultivated with oil palm (*Elaeis guineensis* Jacq.). *AGRIVITA Journal of Agricultural Science*, 41(1):166-174. https://doi.org/10.17503/agrivita.v41i1.20

Arifin, Atmojo SW, Setyono P & Dewi WS. 2015. Temperature effect investigation toward peat surface $CO_2$ emissions by planting leguminous cover crops in oil palm plantations in West Kalimantan. *Journal of Agricultural Science and Technology B5*, 170-183. https://doi.org/10.17265/2161-6264/2015.03.002

Ashton-Butt A, Aryawan AAK, Hood ASC, Naim M, Purnomo D, Suhardi, Wahyuningsih R, Willcock S, Poppy GM, Caliman J-P, Turner EC, Foster WA, Peh KS-H & Snaddon JL. 2018. Understory vegetation in oil palm plantations benefits soil biodiversity and decomposition rates. *Frontiers in Forests and Global Change*, 1:10. https://doi.org/10.3389/ffgc.2018.00010

Bou Dib J, Krishna, VV, Alamsyah Z & Qaim M. 2018. Land-use change and livelihoods of non-farm households: The role of income from employment in oil palm and rubber in rural Indonesia. *Land Use Policy*. 76:828–838. https://doi.org/10.1016/j.landusepol.2018.03.020

Breiman L, Friedman JH, Olshen RA & Stone CJ. 1984. *Classification And Regression Trees*. Boca Raton (US):Routledge. https://doi.org/10.1201/9781315139470

Corley RHV & Tinker PB. 2016. *The Oil Palm, Fifth Edition*. West Sussex (UK): Blackwell Science Ltd. https://doi.org/10.1002/9781118953297

Dadap NC, Hoyt AM, Cobb AR, Oner D, Kozinski M, Fua PV, Rao K, Harvey CF & Connings AG. 2021. Drainage canals in Southeast Asian peatlands increase carbon emissions. *AGU Advances 2*, e2020AV000321. https://doi.org/10.1029/2020AV000321

Dhandapani S, Evers S, Ritz K & Sjögersten S. 2020. Nutrient and trace element concentrations influence greenhouse gas emissions from Malaysian tropical peatlands. *Soil Use and Management*, 37:138–150. https://doi.org/10.1111/sum.12669

Ditjen Perkebunan. 2011. Sustainable Palm Oil Development Policy. *In Indonesia:* Kebijakan Pengembangan Kelapa Sawit Berkelanjutan. Seminar on RSPO Implementation in Indonesia. Jakarta, 10 Februari 2011.

Fairhurst T. 1996. *Management of Nutrients for Efficient Use in Smallholder Oil Palm Plantations*. PhD thesis. London(UK): Department of Biological Sciences, Wye College, University of London

Gomes LC, Faria RM, de Souza E, Veloso GV, Schaefer CEGR & Filho EIF. 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340:337-350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gregory PJ, George TS & Paterson E. 2022. New methods for new questions about rhizosphere/plant root interactions, *Plant and Soil.* https://doi.org/10.1007/s11104-022-05437-x

Grenwell B, Boehmke B & Gray B. 2020. Package 'vip'. Variable Importance Plots. Retrieved from https://cran.r-project.org/web/packages/vip/index.html

Grunwald S. 2021. Grand challenges in pedometrics-AI research. *Frontier in Soil Science*, 1:714323. https://doi.org/10.3389/fsoil.2021.714323

Hamner B, Frasco M & LeDell E. 2018. Package 'Metrics'. Evaluation Metrics for Machine Learning. Retrieved from https://cran.r-project.org/web/packages/Metrics/

Hood AS, Advento AD, Stone J, Fayle TM, Fairnie ALM, Waters HS, Foster WA, Snaddon JL, Sudharto Ps, Caliman J-P, Naim M & Turner EC. 2020. Removing understory vegetation in oil palm agroforestry reduces ground-foraging ant abundance but not species richness. *Basic and Applied Ecology*, https://doi.org/10.1016/j.baae.2020.07.002

Hunt T. 2020. Package 'ModelMetrics'. Rapid Calculation of Model Metrics. Retrieved from https://cran.r-project.org/web/packages/ModelMetrics/

Jelsma I, Schoneveld GC, Zoomers A & van Westen ACM. 2017. Unpacking Indonesia's independent oil palm smallholders: An actor-disaggregated approach to identifying environmental and social performance challenges. *Land Use Policy*, 69:281–297. https://doi.org/10.1016/j.landusepol.2017.08.012

Joo GK, Soon CP & Kiang KK. 1994. *K Nutrition for Mature Oil Palm in Malaysia*. IPI Research Topics No. 17. Basel (CH): International Potash Institute

Jordan MI & Mitchell TM. 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260. https://doi.org/10.1126/science.aaa8415

Koh LP. 2008. Can oil palm plantations be made more hospitable for forest butterflies and birds? *Journal of Applied Ecology*, 45(4):1002–1009. https://doi.org/10.1111/j.1365-2664.2008.01491.x

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Core Team, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C & Hunt T. 2022. Package 'caret'. Classification and Regression Training. Retrieved from https://cran.r-project.org/web/packages/caret/index.html

Loh W-Y. 2014. Fifty years of classification and regression trees. *International Statistical Review.* 82(3):329–348. https://doi.org/10.1111/insr.12016

Luke SH, Purnomo D, Advento AD, Aryawan AAK, Naim M, Pikstein RN, Ps S, Rambe TDS, Soeprapto, Caliman J-P, Snaddon JL, Foster WA & Turner EC. 2019. Effects of understory vegetation management on plant communities in oil palm plantations in Sumatra, Indonesia. *Frontiers in Forests and Global Change,* 2:33. https://doi.org/10.3389/ffgc.2019.00033

Padarian J, Minasny B & McBratney AB. 2020. Machine learning and soil sciences: a review aided by machine learning tools. *Soil*, 6(1):35–52. https://doi.org/10.5194/soil-6-35-2020

Majumdar K, Sanyal SK, Dutta SK, Satyanarayana T & Singh VK. 2016. Nutrient mining: addressing the challenges to soil resources and food security. In Singh U, Praharaj CS, Singh SS, Singh NP. (Eds.). *Biofortification of Food Crops,* Springer. p.177–198. https://doi.org/10.1007/978-81-322-2716-8_14

Mardhiana, Apriyani D, Adiwena M & Pradana AP. 2021. Growth and yield performance of maize at red-yellow podzolic acid soil after oil palm empty fruit bunches compost and rice husk charcoal application. *Journal of Degraded and Mining Lands Management*, 8(2): 2653-2660. https://doi.org/10.15243/jdmlm. 2021.082.2653

Meshram V, Patil K, Meshram V, Hanchate D & Ramkteke SD. 2021. Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences,* 1:100010. https://doi.org/10.1016/j.ailsci.2021.100010

Milborrow S. 2022. Package 'rpart.plot'. Retrieved from https://cran.r-project.org/web/packages/rpart.plot/index.html

Mohidin H, Man S, Hanafi MM, Rufai S, Idris J, Fonguimgo T, Idris AS & Rafii YM. 2019. Optimum levels of N, P, and K nutrition for oil palm seedlings grown in tropical peat soil. *Journal of Plant Nutrition,* 42:1461-1471. https://doi.org/10.1080/01904167.2019.1628975

Mutert E, Fairhurst TH & von Uexküll HR. 1999. Agronomic management of oil palms on deep peat. *Better Crops International,* 13(1):22-27.

Nelson PN, Banabas M, Goodrick I, Webb MJ, Huth NI & O'Grady D. 2015. Soil sampling in oil palm plantations: a practical design that accounts for lateral variability at the tree scale. *Plant and Soil,* 394(1-2):421–429. https://doi.org/10.1007/s11104-015-2490-9

Ng SK, von Uexküll HR, Thong KC & Ooi SH. 1990. Maximum exploitation of genetic yield potentials of some major tropical tree crops in Malaysia. In: Proceedings of Symposium on Maximum Yield Research, Kyoto, Japan, 17 August, 1990. A Satellite Symposium of the 14th International Congress of Soil Science. International Society of Soil Science, pp. 120-130.

Pacheco P, Gnych S, Dermawan A, Komarudin H & Okarda B. 2017. The palm oil global value chain: Implications for economic growth and social and environmental sustainability. Working Paper 220. Bogor (ID): CIFOR.

Pashkevich MD, Luke SH, Ketut Aryawan AA, Waters HS, Caliman J-P, Dupérré N, Naim M, Potapov AM & Turner EC. 2022. Riparian buffers made of mature oil palms have inconsistent impacts on oil palm ecosystems. *Ecological Applications*, 32(4): e2552. https://doi.org/10.1002/eap.2552

Pauli N, Donough C, Oberthür T, Cock J, Verdooren R, Rahmadsyah, Abdurrohim G, Indrasuara K, Lubis A, Dolong T & Pasuquin JM. 2014. Changes in soil quality indicators under oil palm plantations following application of "best management practices" in a four-year field trial. *Agriculture, Ecosystems and Environment*, 195:98–111. https://doi.org/10.1016/j.agee.2014.05.005

Pardon L, Bessou C, Nelson PN, Dubos B, Ollivier J, Marichal R, Caliman J-P & Gabrielle B. 2016. Key unknowns in nitrogen budget for oil palm plantations. A review. *Agronomy for Sustainable Development*, 36(1). https://doi.org/10.1007/s13593-016-0353-2

Pinheiro HSK, Carvalho Júnior W, Chagas CS, Anjos LHC & Owens PR. 2018. Prediction of topsoil texture through regression trees and multiple linear regressions. *Revista Brasileira de Ciência do Solo,* 42:e0170167. https://doi.org/10.1590/18069657rbcs20170167

Poppiel RR, Lacerda MPC, Safanelli JL, Rizzo R, Oliveira MP, Novais JJ & Demattê JAM. 2019. Mapping at 30 m resolution of soil attributes at multiple depths in Midwest Brazil. *Remote Sensing,* 11(24):2905. https://doi.org/10.3390/rs11242905

Pourghasemi HR & Rahmati O. 2018. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena*, 162: 177–192. https://doi.org/10.1016/j.catena.2017.11.022

Pulunggono HB, Nurazizah LL, Zulfajrin M, Anwar S & Sabiham S. 2022a. Assessing the distribution of total Fe, Cu, and Zn in tropical peat at an oil palm plantation and their relationship with several environmental factors. *Journal of Degraded and Mining Lands Management,* 9(2): 3349-3358. https://doi.org/10.15243/jdmlm.2022.092.3349

Pulunggono HB, Kartika VW, Nadalia D, Nurazizah LL & Zulfajrin M. 2022b. Evaluating the changes of Ultisol chemical properties and fertility characteristics due to animal manure amelioration. *Journal of Degraded and Mining Lands Management,* 9(3):3545-3560. https://doi.org/10.15243/jdmlm.2022.093.3545

Qiu L, Wang K, Long W, Wang K, Hu W & Amable GS. 2016. A comparative assessment of the influences of human impacts on soil Cd concentrations based on stepwise linear regression, classification and regression tree, and random forest models. *PLOS ONE*, 11(3):e0151131. https://doi.org/10.1371/journal.pone.0151131

R Core Team. 2022. R*: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rahman N, Giller KE, de Neergaard A, Magid J, van de Ven G & Bruun TB. 2021. The effects of management practices on soil organic carbon stocks of oil palm plantations in Sumatra, Indonesia. *Journal of Environmental Management*, 278, 111446. https://doi.org/10.1016/j.jenvman.2020.111446

Rawal A, Chakraborty S, Li B, Lewis K, Godoy M, Paulette L & Weindorf DC. 2019. Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. *Geoderma,* 338:375-382. https://doi.org/10.1016/j.geoderma.2018.12.032

RSPO [Roundtable on Sustainable Palm Oil]. 2012. RSPO Manual on Best Management Practices (BMPs) for Existing Oil Palm Cultivation on Peat. In lim KH, lim SS, Parish F, Suharto R. (Eds.). Kuala Lumpur (MY):Roundtable on Sustainable Palm Oil.

Sparks DL. 2003. *Environmental Soil Chemistry*. San Diego (US). Academic Press.

Teixeira A, Pelegrino MHP, Faria EM, Silva SHG, Gonçalves MGM, Júnior FWA, Gomide L, de Pádua Junior AL, de Souza IA, Chakraborty S, Weindorf DC, Guilherme LRG & Curi N. 2020. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. *Geoderma,* 361:114132, https://doi.org/10.1016/j.geoderma.2019.114132

Therneau T, Atkinson B, Ripley B. 2022. Package 'rpart'. Retrieved from https://cran.r-project.org/web/packages/rpart/index.html

Tipping E. 2004. *Cation Binding by Humic Substances*. Cambridge University Press, Cambridge. p140

Trontelj ml. J & Chambers O. 2021. Machine learning strategy for soil nutrients prediction using spectroscopic method. *Sensors*, 2:4208. https://doi.org/10.3390/s21124208

Van Looy K, Bouma J, Herbst M, Koestel J, Minasny B, Mishra U, Montzka C, Nemes A, Pachepsky YA, Padarian J, Schaap MG, Tóth B, Verhoef A, Vanderborght J, van der Ploeg MJ, Weihermüller L, Zacharias S, Zhang Y & Vereecken H. 2017. Pedotransfer functions in

Earth system science: Challenges and perspectives. *Reviews of Geophysics,* 55:1199–1256. https://doi.org/10.1002/2017RG000581

Vega FA, Matías JM, Andrade ML, Reigosa MJ & Covelo EF. 2009. Classification and regression trees (CARTs) for modelling the sorption and retention of heavy metals by soil. *Journal of Hazardous Materials*, 167(1-3), 615–624. https://doi.org/10.1016/j.jhazmat.2009.01.016

Vereecken H, Pachepsky Y, Simmer C, Rihani J, Kunoth A, Korres W, Graf A, -Hendricks Franssen HJ, Thiele-Eiche I & Shao Y. 2016. On the role of patterns in understanding the functioning of soil-vegetation-atmosphere systems. *Journal of Hydrology,* 542:63–86. https://doi.org/10.1016/j.jhydrol.2016.08.053

Wahid MB, Abdullah SNA & Henson IE. 2005. Oil palm — Achievements and potential. *Plant Production Science*, 8(3): 288-297. https://doi.org/10.1626/pps.8.288

Wickham H. 2022. Package 'tidyverse'. Easily Install and Load the 'Tidyverse'. Retrieved from https://cran.r-project.org/web/packages/tidyverse/index.html

Woittiez LS, Slingerland M, Rafik R & Giller KE. 2018. Nutritional imbalance in smallholder oil palm plantations in Indonesia. *Nutrient Cycling in Agroecosystems*, 111(1):73–86. https://doi.org/10.1007/s10705-018-9919-5

Woittiez lS, Turhina S, Deccy D, Slingerland M, van Noordwijk M & Giller KE. 2019. Fertilizer application practices and nutrient deficiencies in smallholder oil palm plantations in Indonesia. *Experimental Agriculture,* 1–17. https://doi.org/10.1017/s0014479718000182

Woittiez LS, van Wijk MT, Slingerland M, van Noordwijk M & Giller KE. 2017. Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy*, 83:57–77. https://doi.org/10.1016/j.eja.2016.11.002

Yoo K, Yoo H, Lee JM, Shukla SK & Park J. 2018. Classification and regression tree approach for prediction of potential hazards of urban airborne bacteria during Asian dust events. *Scientific Reports,* 8:11823. https://doi.org/10.1038/s41598-018-29796-7

York LM, Carminati A, Mooney SJ, Ritz K & Bennett MJ. 2016. The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. *Journal of Experimental Botany*, 67(12):3629–3643. https://doi.org/10.1093/jxb/erw108

Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, Ma X, Marrone BL, Ren ZJ, Schrier J, Shi W, Tan H, Wang T, Wang X, Wong BM, Xiao X, Yu X, Zhu J-J & Zhang H. 2021. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environmental Science & Technology*, https://doi.org/10.1021/acs.est.1c01339