Prediction of seismic p-wave velocity using machine learning
Ines Dumke and Christian Berndt
Solid Earth
Referee Comment - Taylor Lee

General comments

Machine learning has been previously well established in other fields, but has not grasped attention in a similar way within the geosciences. This paper uses sparse p-wave velocity data from DSDP/ODP/IODP as training data in a machine learning algorithm (Random Forest) to predict p-wave velocity with depth. A thorough analysis was done to determine how effective machine learning is at predicting vertical velocity profiles. This analysis included comparison of p-wave velocity machine learning predictions with empirical estimates. A variety of appropriate methods were tested to improve the machine learning prediction (e.g. smoothing input data and prediction results, varying max_features and number of predictors used, 10-fold cross validation, predictor value scaling). As a result, this work provides valuable information on types of useful predictors and variables highly correlated to p-wave velocity. Additionally, this method shows in some case superior to using strictly empirical methods to estimate p-wave velocity with depth.

Results show this work is novel and useful. However, there is a major component of the analysis missing. This work contains many examples of validation of previously existing p-wave velocity but lacks demonstration on prediction of p-wave velocity in areas where no velocity data is available.

Specific comments

Page 3 Section 2.1.2 (Predictors) Line 28 mentions that the continental crust was set at 1 billion years to represent significant older crust than that of the oceanic crust. If all the observed data (DSDP/ODP/IODP) are on oceanic crust, what is the importance/meaning of defining continental crust age?

Page 7 Section 3.3 (Predictor importance) Line 20 states that categorical predictors generally do not have any importance in prediction performance. Additionally, it is again discussed in the discussion section (Section 4.2- lines 8-14 page 9). What is the variance of your sampled data set in categorical predictors? For example, for a given test data set (i.e. fold) are all of your categorical predictors for that run a 1 or 0? If all of your test data set has only one categorical value then that predictor would be of no importance.

Consider, if true, explicitly stating that predictions of this kind have not done with depth before. (page 2 ~ lines 16-20)

Minor suggestion to add in the abstract that this method is not designed to capture high variance in a p-wave velocity profile, but is instead intended to capture the overall trend of p-wave velocity profile.

It is stated and supported (Line 1 page 7; Figure 3) that the RFE CV 16 predictors prediction (green) is better than CV, max_features =22, 38 predictors however the error in the prediction is significantly higher for the green prediction with roughly the same % boreholes labelled as "good". Why do you consider green prediction to be so much better than yellow prediction? It might be useful if you explicitly state what your ultimate metric of correctness is (e.g. highest % correct or lowest error?)

What is the final global spatial resolution? E.g. prediction of p-wave velocity profile every 1-degree, 5-min, etc.?

Page 9 Section 4.2 (Most important predictors for the prediction of $v_p(z)$) Lines 2-8 discuss how certain predictors are not used (porosity, density, pressure) as not all boreholes have depth associated measurements. However, some of the predictors used in the prediction do not have a depth component (e.g. crustage). Applying this logic, why do you not use seafloor porosity (i.e. depositional porosity) or likewise predictors?

No supplemental material was provided for the global prediction of p-wave velocity with depth. This paper should include the final global prediction of p-wave velocity with depth.

Technical corrections

Page 8 delete "the" on line 21: "by the at least 60% of test locations"

Page 8 line 3 consider changing "our results show that $v_p(z)$ profiles" to "our results show that the general trend of vp(z) profiles"

Page 16 Figure 2 caption (e) change "less good" to different word (substandard?)

Page 23 table 3, change words so they have consistent capitalization between table columns (e.g. Long and long)

Page 12 Lee et al., 2019 citation is missing the publication year.