Solid Earth Discuss.,
https://doi.org/10.5194/se-2019-58-AC1, 2019

Interactive comment on "Prediction of seismic p-wave velocity using machine learning" *by* Ines Dumke and Christian Berndt

Ines Dumke and Christian Berndt

cberndt@geomar.de

Received and published: 9 August 2019

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 re-

sults, 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.

AC: As we explain in the Methods section, due to our leave-location-out approach all predictions are made for locations that were withheld from the training data and therefore act as unknown locations. Validation of the prediction involved a comparison against the true vp data, but these data were in no case part of the prediction model. We explain below (last "specific comment") why we refrain from making predictions for completely new locations as it is beyond the scope of this paper, i.e. the purpose of this paper is to demonstrate the method and to discuss its advantages and limitations. When more training data become available the method can be used to make predictions elsewhere - probably first for limited areas and then globally.

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?

AC: It is not true that all the data are from sites above oceanic crust. In fact, 142 of the 333 boreholes - 42% - were drilled on continental crust, e.g. in continental shelf regions. As the thermal regime of continental crust is different to that of oceanic crust - with old continental crust being of lower temperatures than young oceanic crust -, which affects density and hence p-wave velocity, we thought it reasonable to differentiate between the two types of crust and their ages. CM: no changes made in the manuscript

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.

AC: We do not claim that categorical predictors "generally" do not have "any importance" in prediction performance. In the referenced line (now p. 8 line 10-11), we use the term "negligible importance", i.e. almost zero, and we explain that this only refers to the results of our own study, not to studies involving categorical predictors in general. The number of boreholes per predictor (for which the predictor is 1) varies between 2 (0.6%) and 191 (57%), on average, it is 42 (12.7%). We therefore agree that predictors with a very low representation will also be of low importance, and that this should be added as an explanation in the Discussion.

CM: We added the following sentence to the end of section 2.1 in the Methods: "Across the categorical predictors, the number of boreholes for which a predictor was set to 1 varied between 2 (0.6 %) and 191 (57.4 %); on average, the geological setting represented by a categorical predictor applied to 42 boreholes (12.7 %)."

We also included the sentence "The poor representation of some predictors, such as "cold_vent", "mud_volcano" and "hydroth_vent" in the dataset, causing these predictors to be 0 for all boreholes in some test folds, may likely explain the low importance of these predictors in the predictor ranking." in the last paragraph of section 4.2 in the Discussion (p. 10 lines 2-5).

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

AC: As far as we know, predictions with depth have not been done before, and we agree that this should be stated in the text.

CM: We added a sentence to this paragraph (lines 21-23): "These studies were in general restricted to the prediction of one value per geographic location; the prediction of multiple values, such as depth profiles, has, to our knowledge, not been attempted before."

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.

AC: We agree that this should already be stated in the Abstract.

CM: We changed the sentence in line 9-10 (now lines 10-11) to read: "Here, we present a machine learning approach to predict the overall trend of seismic p-wave velocity (vp) as a function of depth (z) for any marine location."

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?)

AC: We did not mean to imply that one of the two runs provides better results than the other, and we also do not claim this anywhere. We merely stated the differences. However, we agree that this could have been easily misunderstood due to our ill use of the term "performance" - we meant performance to refer to both the highest % correct and the lowest error (i.e., in the same model), but we seem to have used it in other ways too, which must have been confusing. We now explain in more detail what we mean by performance and how our predictions were evaluated.

CM: Paragraph 3 in the Methods section 2.3 was changed to read "Performance of the RF model was evaluated in two ways: (1) by standard error metrics and (2) by the proportion of boreholes with predicted vp(z) superior to that of empirical functions. The standard error metrics root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R2) were calculated based on the comparison of the predicted and true vp(z) curves for each borehole in the test fold. RMSE, MAE and R2 of all test folds were then averaged to give final performance values." Throughout the manuscript, we also replaced the term "performance" where necessary, to make its use consistent. In the sentence reference above (now lines 19-22), we replaced "performance" by "prediction scores".

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

AC: We do not want to go so far as to give a final global spatial resolution for the prediction of vp. Our main aim was to investigate if it is at all possible to achieve realistic predictions of vp(z). We have shown that this is generally the case, however, our results also clearly indicate that more input data are required to overcome low prediction performance due to lack of suitable data. For this reason, we think that the prediction model needs to be improved further before a "final resolution" should be given. - In any case, one final resolution value likely would not be sufficient. Due to the heterogeneous depth distribution of the boreholes used (in addition to the heterogeneous spatial distribution), the resolution would vary with depth. Thus, separate resolution values would need to be determined for different depths (here: range 0-2500 m), which would likely be confusing and not very helpful for the reader.

CM: no changes made in the manuscript

Page 9 Section 4.2 (Most important predictors for the prediction of vp(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?

AC: We did not mean that we can only use predictors with depth measurements, we obviously also used depth-independent predictors. The point here (which was not well explained in the text) was that we could only use predictors that were available (or could be determined) for every borehole location. This did not apply to many of the e.g. porosity measurements, which had been measured in boreholes (with a depth component) but often not at the borehole locations at which vp had been measured. Even in the relatively few boreholes where both porosity (or density, pressure etc) and vp had been measured, the depth ranges did not always match - so there would have been depths with vp data but no porosity data. It was impossible to also account for such cases, which is why we decided to leave these parameters out. As reviewer 2 also asked for an explanation regarding choice of predictors, we clarified this in the Methods section 2.1.2. We also agree that a parameter like seafloor porosity, which is available as a global grid (we are assuming that the reviewer is referring to the grid by Martin et al., 2015), could easily have been added as a predictor. We did not do this at the time, and we hope the reviewer will understand that it is now too late to add new predictors to our study - as we state in section 4.2, there are several other predictors that could potentially be added, but this would have to be done in a future study. CM: We have clarified our choice of predictors by adding the following passage to the Methods section 2.1.2 (p. 4 lines 5-13): "... These predictors were parameters that were assumed to influence p-wave velocity. However, only predictors that could be obtained for each of the 333 borehole locations were used. Predictors such as latitude (lat), longitude (long), and water depth (wdepth) were taken from the borehole's metadata, whereas other predictors were extracted from freely available global datasets and grids (Table 1). In addition, predictors describing the borehole's geological setting were determined from the site descriptions given in the proceedings of each drilling campaign. Some parameters known to influence seismic velocity - e.g. porosity, density, or pressure - had to be left out as suitable datasets were not available. Although some of these parameters had been measured in DSDP, ODP and IODP boreholes, they had not necessarily been logged at the same locations and depths at which vp data had been measured, and therefore could not be obtained at all of the 333 boreholes used."

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.

AC: No, we do not agree. As with the final spatial resolution, providing a final global prediction of vp at this stage (i.e. when the prediction model still requires optimization and is therefore not final yet) is neither feasible nor helpful. In fact, it would maybe give this method a bad reputation to deploy it prematurely. Furthermore, we show that one "final global prediction" would not be sufficient. We assume the reviewer expects a global map of final prediction values, similar to Fig. 4 in Taylor et al. (2019) or Fig. 1c in Martin et al. (2015). While such a map may be useful in cases with only one prediction value per location, in our case - taking into account the depth component of the predicted vp - a whole range of prediction maps would seem necessary, one for each depth. However, none of these maps would be of much use on its own. It would only show the variation of velocity at a certain depth, but we are interested in the variation (or trend) of velocity with depth (i.e., a profile), which is much better illustrated by the predicted vp(z) profiles (of which we show sufficient examples). Thus, we do not think a final global prediction is useful.

CM: no changes made in the manuscript

Technical corrections

Page 8 delete "the" on line 21: "by the at least 60% of test locations" CM: deleted "the" (now p. 9 line 8)

Page 8 line 3 consider changing "our results show that vp(z) profiles" to "our results

show that the general trend of vp(z) profiles" CM: We changed this sentence accordingly. (now line 24)

Page 16 Figure 2 caption (e) change "less good" to different word (substandard?) CM: We changed this to "lower-quality prediction".

Page 23 table 3, change words so they have consistent capitalization between table columns (e.g. Long and long) CM: We changed the capitalized letters accordingly (also in Table 2).

Page 12 Lee et al., 2019 citation is missing the publication year. AC: This paper was fully published just before we submitted our manuscript and we forgot to update the reference correctly. CM: Added publication year.

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-58, 2019.