Extended information

Machine learning-based identification of key biotic and abiotic drivers of mineral weathering rate in a complex enhanced weathering experiment

Iris Janssens[1], Thomas Servotte[2], Tullia Calogiuri[3], Steven Mortier[1], Harun Niron[4], Thomas Corbett[5], Reinaldy P. Poetra[6], Lukas Rieder[6], Michiel Van Tendeloo[4], Abhijeet Singh[5], Steven Latré[1], Siegfried E. Vlaminck[4], Jens Hartmann[6], Jan Willem van Groenigen[3], Anna Neubeck[5], Alix Vidal[3], Ivan A. Janssens[7], Mathilde Hagens[8], Sara Vicca[4], and Tim Verdonck[2]

[1] *University of Antwerp - imec - IDLab, Department of Computer Science, Sint-Pietersvliet 7, 2000 Antwerp, Belgium*

[2] *University of Antwerp - imec, Department of Mathematics, Middelheimlaan 1, 2020 Antwerp, Belgium*

[3] *Wageningen University & Research, Soil Biology Group, Droevendaalsesteeg 3, 6708 PB, Wageningen, the Netherlands*

[4] *University of Antwerp, Department of Bioscience Engineering, Biobased Sustainability Engineering (SUSTAIN), Groenenborgerlaan 171, 2000 Antwerp, Belgium*

[5] *Uppsala University, Department of Earth Sciences, Villaväagen 16, 752 36 Uppsala, Sweden*

[6] *University of Hamburg, Institute for Geology, Center for Earth System Research and Sustainability, Bundesstraße 55, 20146 Hamburg, Germany*

[5] *Uppsala University, Department of Earth Sciences, Villaväagen 16, 752 36 Uppsala, Sweden*

[7] *University of Antwerp, department of Biology, Universiteitsplein 1, 2160 Wilrijk, Belgium*

[8] *Wageningen University & Research, Department of Soil chemistry, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands*

B Machine learning analysis

B.1 Tuned hyperparameters

Lasso	RF	XGB
alpha	max_depth	max_depth
	n_estimators	n_estimators
	min_samples_leaf	min_child_weight
		colsample_bytree
		subsample

Table B1: Tuned hyperparameters in the Least absolute shrinkage and selection operator regularization (Lasso), Random Forest (RF) and eXtreme Gradient Boosting (XGB) models; as defined in the *scikit-learn* [2] and *XGBoost* [1] libraries.

B.2 Least Absolute Shrinkage and Selection Operator regularization results for change in dissolved inorganic carbon

feature	coefficient	feature	coefficient
dunite SA	17.56 ± 0.9	# C. metallidurans	0.65 ± 0.64
alfalfa mixture mass	9.01 ± 0.34	# K. petricola	0.61 ± 0.22
† chlorotica ratio	6.76 ± 1.54	lava SA	0.09 ± 0.18
straw mass	4.96 ± 0.2	# carbonic anhydrase	0.06 ± 0.5
urea	4.79 ± 1.03	# urease	0.0 ± 0.0
basalt SA	4.37 ± 0.26	mineral + biochar mass	0.0 ± 0.0
# S. variegatus	2.36 ± 0.55	chlorotica ratio	-0.03 ± 0.13
digestate mass	2.23 ± 0.28	halftime inoculation	-0.09 ± 0.28
# A. pullulans	1.93 ± 0.3	sterilized	-0.33 ± 0.27
# B. subtilis	1.44 ± 0.65	P	-0.72 ± 0.54
# earthworms	1.35 ± 0.22	biochar mass	-0.75 ± 0.22
irrigation frequency	1.08 ± 0.09	# laccase	-1.41 ± 0.73
premixed	1.0 ± 0.19	# NH_4	-1.48 ± 0.14
mixed grain size	0.86 ± 0.26	# † earthworms	-4.39 ± 1.93
irrigation flow rate	0.81 ± 0.09	steel slag SA	-4.7 ± 0.35

Table B2: Regression coefficients of the Lasso prediction of the change in dissolved inorganic carbon. Depicted values are the averages and standard deviations of the coefficients of the ten models in the outer loop of the nested cross validation.

B.3 SHapley Additive eXplanation results

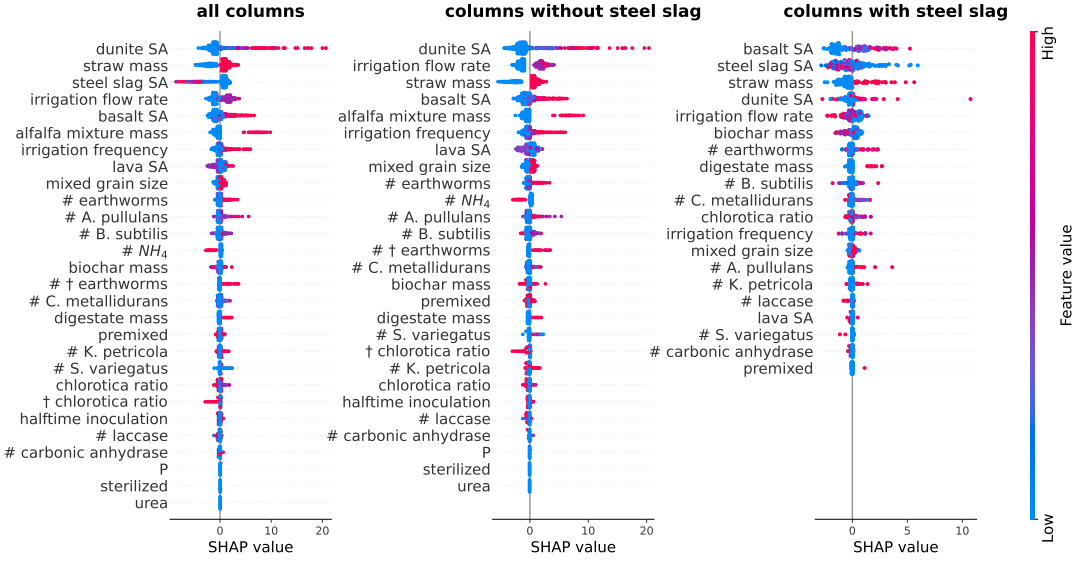

Figure B1: SHapley Additive eXplanation (SHAP) values of the varied features in the eXtreme Gradient Boosting prediction of the change in amount of dissolved inorganic carbon (ΔDIC). SA stands for surface area, # stands for added amount of biota.

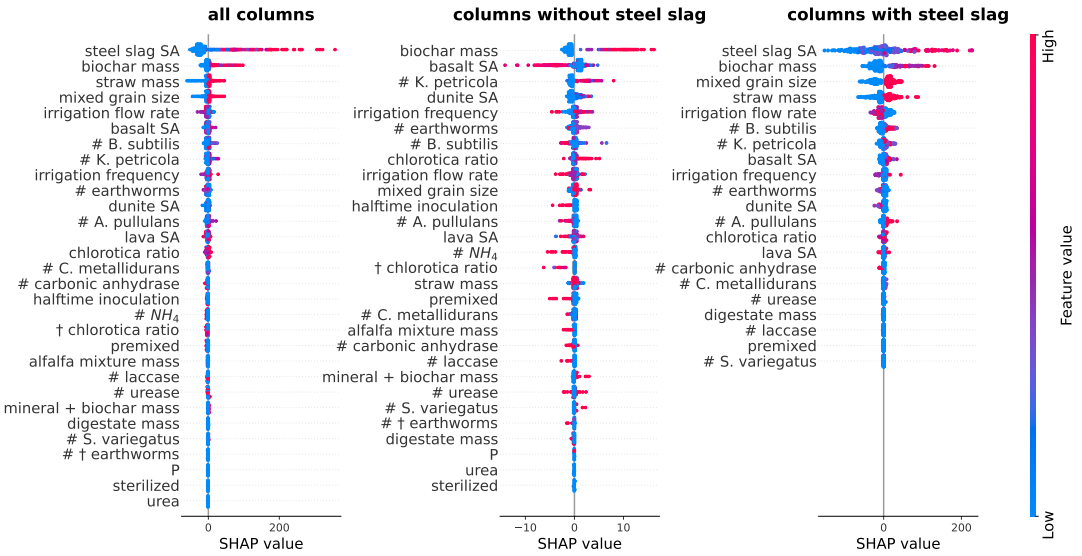

Figure B2: SHapley Additive eXplanation (SHAP) values of the varied features in the eXtreme Gradient Boosting prediction of the change in amount of solid inorganic carbon (ΔSIC). SA stands for surface area, # stands for added amount of biota.

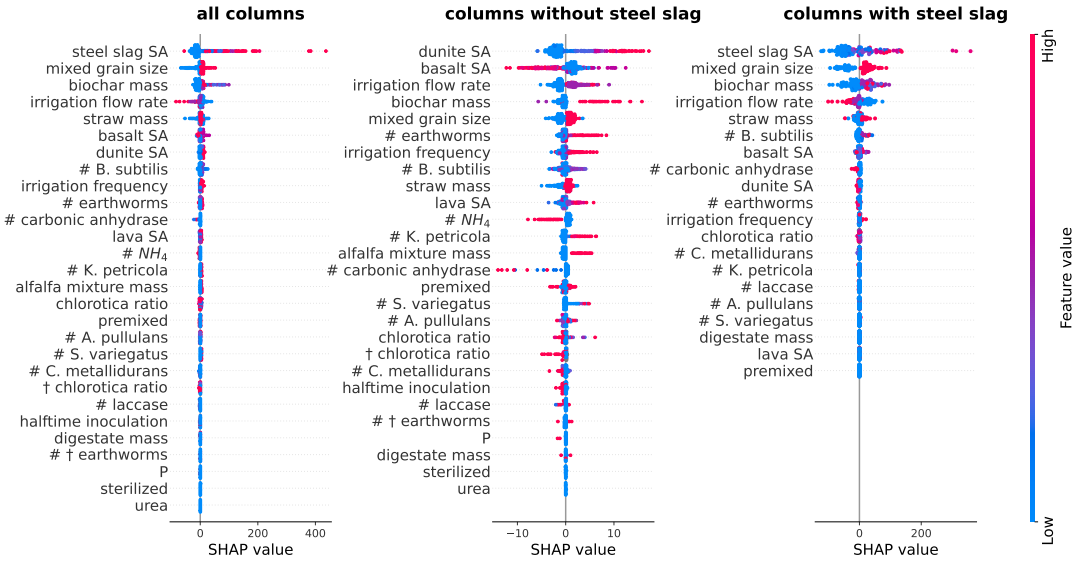

Figure B3: SHapley Additive eXplanation (SHAP) values of the varied features in the eXtreme Gradient Boosting prediction of the change in amount of total inorganic carbon (ΔTIC). SA stands for surface area, # stands for added amount of biota.

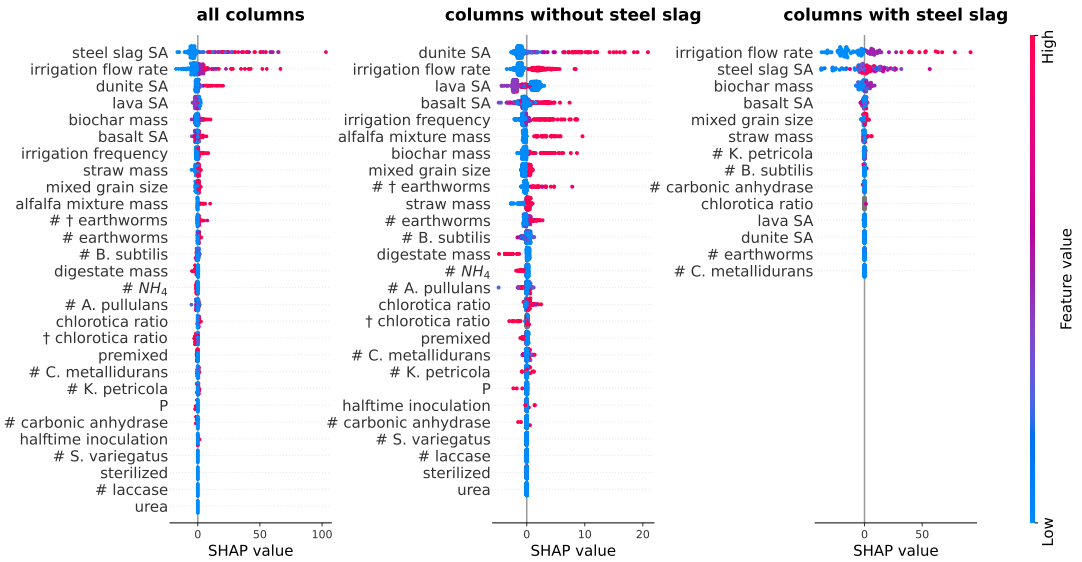

Figure B4: SHapley Additive eXplanation (SHAP) values of the varied features in the Random Forest prediction of the change in amount of total alkalinity (ΔTA). SA stands for surface area, # stands for added amount of biota.

References

[1] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.