MAPPING THE RESILIENCE CHARACTERISTICS OF SOIL BY USING MULTI-OBJECTIVE REGRESSION MODELS

Debeljak M.[(1)], Griffiths B.[(2)], Hallett P.[(3)], Kocev D.[(4)], Towers W.[(5)] Džeroski S. [(6)]

[(1)] Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, marko.debeljak@ijs.si
[(2)] Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK, bryan.griffiths@scri.ac.uk
[(3)] Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK, paul.hallett@scri.ac.uk
[(4)] Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, dragi.kocev@ijs.si
[(5)] Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK, w.towers@macaulay.ac.uk
[(6)] Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, saso.dzeroski@ijs.si

Key words: soil resilience, soil resistance, multi-objective regression trees, soil functions, soil map

Abstract: Soil resilience is a key component in the environmental sustainability of ecosystems. It defines the ability of soil to recover from external stresses that may occur through either climate, soil management or industrial pollution. Recent research has shown that the resilience of a single soil to physical and biological stresses varies considerably and that biological and physical resilience in soil can be interdependent properties (Griffiths et al. 2005).

An effort to determine the underlying relationships between physical and biological resilience has been made by Kuan et al. (2007), where a wide range of soils from across Scotland has been examined in an attempt to isolate the impacts of different soil properties and any interdependency between biological and physical resilience. Key soil properties that were anticipated to drive physical and biological resilience included organic matter content, clay content, pH and land use. The results have shown that there were no soils uniformly resilient to a range of physical and biological stresses: this implies that no single measure can properly describe resilience and any index of soil quality should take this into account.

As a result, a major limitation in the work of Kuan et al. (2007) was the interpretation of the vast amount of data collected into a usable index of soil quality. To address this problem, the recently developed methodology of multi-objective regression modelling (Blockeel et al. 1998, Struyf and Džeroski 2006), was applied to examine the underlying relationships between different attributes describing soil exposure to stress. We took the data and the results from the published work of Kuan et al. (2007) on the biological and physical resilience of soil as input data in our research. The dataset consisted of ten attributes describing physical properties (soil texture: sand, silt, clay), chemical properties (pH, C, N, SOM(soil organic matter)), FAO soil classification (Order and Suborder) of 26 soil samples taken throughout Scotland, and eight indicators of physical (resistance to compression: 1/Cc, recovery from compression: Ce/Cc, overburden stress: eg, recovery from overburden stress after two days cycles: eg2dc) and biological (heat, copper; Cu)) resistance and resilience soil properties. Thus, the flat table of data consisted of 26 by 18 data entries.

To examine the existence of underlying relationships between soil properties and its resilience and resistance capacity we applied multi-objective regression modelling on this dataset. We used the tool CLUS (Blockeel and Struyf 2002) for learning multi-objective regression trees. The main reason for applying these machine learning techniques was to overcome the problems indicated in the work of Kuan et al. (2006), namely that no single measure can properly describe resilience and that dependencies between the assessed resilience and resistance attributes needs to be taken into account.

A regression tree is a predictive model for a numeric target (Breiman et al. 1984). Multi-objective regression trees (MORTs) (Blockeel et al. 1989) are regression trees capable of predicting several numeric targets simultaneously (only one tree is learned, which predicts all targets), and are as such a generalization of regression trees. In this approach, the prediction is a vector of numeric values (one component of the vector for each target attribute).

Models have been made from the data given in the described flat table, where attributes describing physical, chemical and classification properties were treated as independent attributes and were used to predict the physical and biological resistance and resilience capacity of the examined soil samples.

We produced a number of models depending on different combination of dependent attributes which we tried to predict by a particular model. For their validation we applied Leave One Out (LOO)

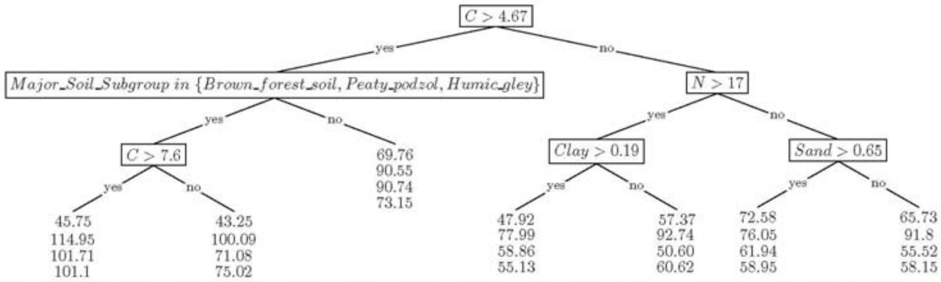
validation, which is commonly used when the number of cases is small (i.e. 26 samples). The final selection of the models has been done according to their performance, as measured by the Root Mean Squared Error (RMSE) and the Correlation Coefficient (CC).

The results revealed a pattern that describes the relationships between the attributes included in the analysis. An example model which predicts biological resilience and resistance soil properties is given in Fig. 1.

Because of the increasing importance of mapping soil functions to advice on land use and environmental management, we used selected models to make a map of soil resilience for Scotland. The models were used as filters for existing GIS datasets about physical and chemical properties of Scottish soils. The map (Fig. 2) generated indicates remarkably distinctive areas in terms of soil biological resilience and resilience.

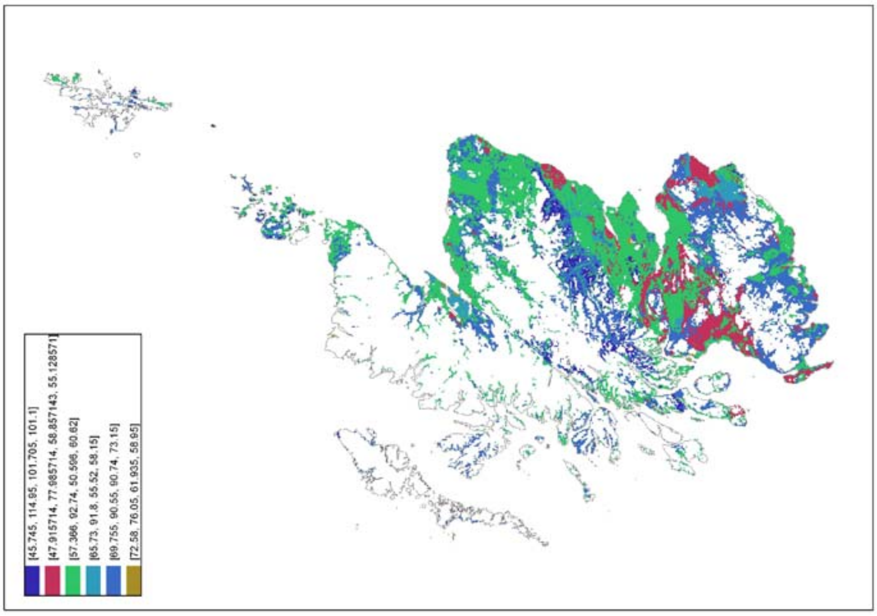
Multi objective regression modelling has been confirmed as an advanced technique to overcome problems of single attribute modelling. The research has confirmed the existence of relationships between biological and physical soil resilience which were integrated into regression models and applied on GIS datasets to produce various types of soil resilience maps. The research succeeded to find causal links between soil properties and soil function which is an important step toward sustainable land use.

Fig.1: A multi-objective regression model predicting biological soil resilience.



Numbers in leafs correspond to the values of heat resilience, heat resistance, copper resistance and copper resilience (top down)

Fig.2: Map of biological soil resilience for Scotland



References:

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth.

Blockeel, H., De Raedt, L., Ramon, J., 1998. Top-down induction of clustering trees. In: Shavlik, J. (Ed.), Proceedings of the 15th International Conference on Machine Learning, pp. 55–63.

Blockeel, H., Struyf, J., 2002. Efficient algorithms for decision tree cross-validation. J. Mach. Learn. Res. 3 (Dec), 621–650.

Griffiths,B.S., Hallett,P.D., Kuan,H.L., Pitkin,Y. & Aitken,M.N. 2005. Biological and physical resilience of soil amended with heavy metal-contaminated sewage sludge. European Journal of Soil Science, 56, 197-206.

Kuan, H.L, Hallett, P.D., Griffiths, B.S. Gregory, A.S., Watts, C.W. & Whitmore, A.P. 2007. The biological and physical stability and resilience of a selection of Scottish soils to stresses. European Journal of Soil Science, 58, 811-821

Struyf, J., Dzeroskim S., 2006. Constraint Based Induction of Multi-ObjectiveRegression Trees. Knowledge Discovery in Inductive Databases, 4th International Workshop, KDID'05, Revised, Selected and Inductive Papers (Bonchi, F. and Boulicaut, J.-F., eds.), vol 3933, Lecture Notes in Computer Science, pp. 222-233