Extended Data

Heat Map Regression Trees

The analytics in the main section identify which nodes are influential in determining SQH, but they do not identify the manner in which those nodes are important. Regression trees were fitted to the belief of SQH (e.g. probability SQH=Good) using results from the 5-way study and considering all non-SQH nodes as possible explanatory variables. A regression tree sequentially identifies the value of the explanatory variables (i.e. a node state) that results in the best partition of the response variable (the SQH belief). This partition is found through an assessment of the deviance. Explicitly, the deviance summarises the goodness of fit by assessing the difference between the observed SQH belief and the model prediction. For the regression tree, this simplifies to a summation of the within group deviance, i.e.

$$D = \sum_{k=1}^{K} D_k = \sum_{k=1}^{K} \sum_{j=1}^{n_k} \left(y_j - \mu_k\right)^2 ,$$

where K is the number of groups defined by the regression tree, y_j is the j^{th} observation in group k, , $n_k$is the number of observations in group k, and μ_k is the mean SQH belief in each group, $\mu_k = \frac{1}{n_k} \sum_j y_j$. At the first iteration of the tree algorithm, the observations are split into two groups such that $|D_1 - D_2|$ is maximised. For example, the belief of SQH in the semi-natural land-use is first partitioned into 2 categories based on whether C storage is decreasing or not. This partition of SQH belief is associated with the largest difference in deviance between any two possible partitions. The tree algorithm continues sequentially partitioning the response variable until the minimum difference in deviance between any two groups (set here at 0.01) is reached.

The regression trees thus enable us to determine the combination of states that result in different SQH beliefs as depicted in the heatmaps of Fig. E1.

Simulation studies and associated analyses were done in the R statistical software environment. The RNetica package (Almond) was used to calculate the node probabilities and associated beliefs. To ensure compatibility, this was implemented in R version 3.4.2. Regression trees were fitted using the tree package (Ripley, 2019).

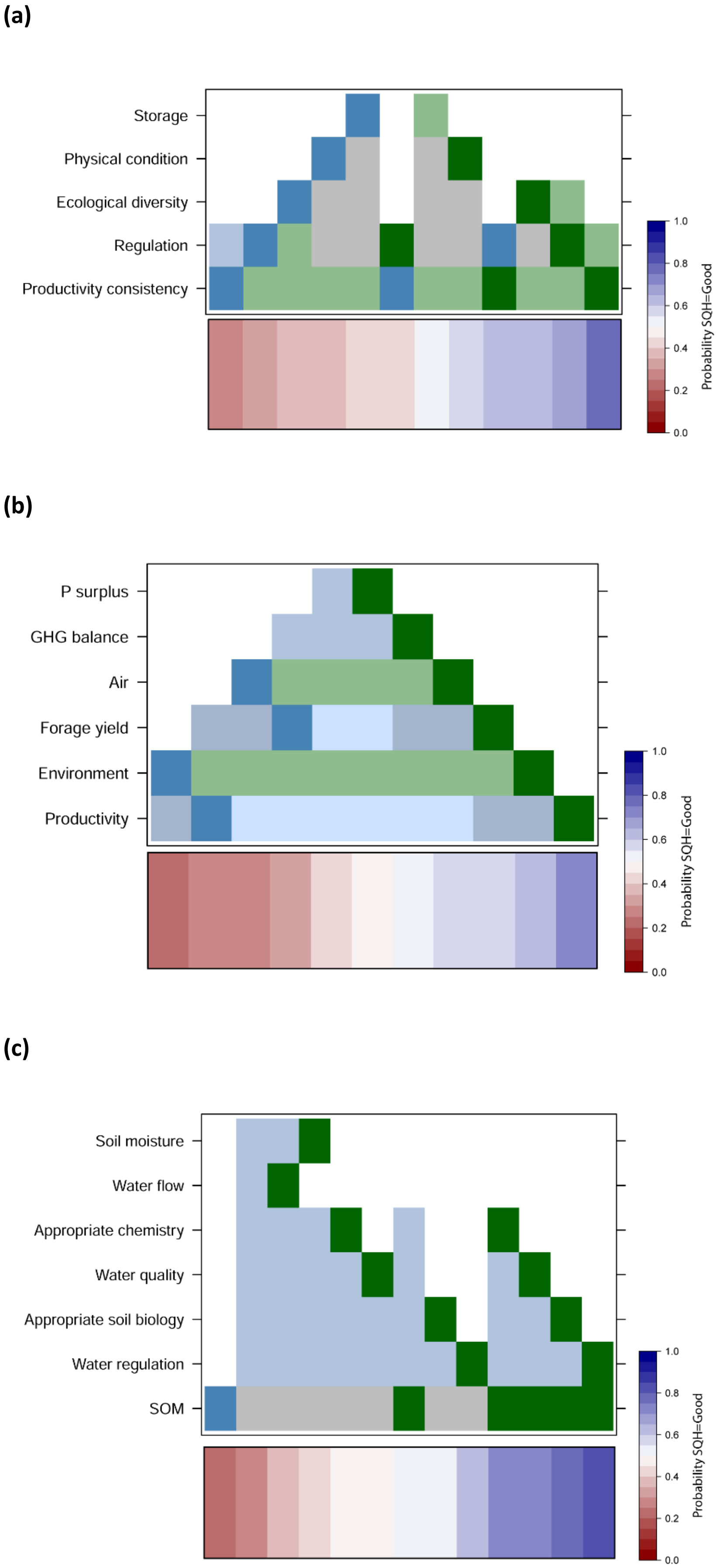

Fig E1: Heatmaps of regression trees summarizing the probabilities of finding Good SQH. *Combinations of nodes describing the greatest variation (deviance) in SQH obtained from a regression tree contingent on the probability of finding good SQH in (a) arable, (b) livestock, (b) semi-natural land-uses.*

Lower panel of each pair (a-c) indicates the probability that SQH is in a Good state according to the scale bar on the right-hand side. The corresponding set of node states that result in this SQH probability is indicated by the upper panel in each case (a-c). For each node, coloured blocks (Green or Blue) indicate that the state of that node is known. Green – node is in a favourable or optimistic state likely to result in Good SQH. *Blue – node is known to be in a negative or pessimistic state likely to lead to* Bad SQH *(e.g. Regulation=Bad). Grey tones indicate uncertainty in the node state (e.g. grey-blue = Regulation* either *unknown or Bad). White - no information on the node state.*

For example, in c) if all that is known is that SOM (C storage) is decreasing (blue), the probability of good SQH is 0.2. However, if SOM is stable (green) and water regulation is good (green), the probability of good SQH is 0.82. When SOM is stable (green), water regulation is uncertain (grey-blue), and chemistry is appropriate, the probability of good SQH is 0.76.

References

Almond, R. RNetica: R interface to Netica(R) Bayesian Network Engine. R package version 0.5-1. http://pluto.coe.fsu.edu/RNetica

Ripley, B (2019) tree: Classification and Regression Trees. R package version 1.0-40. https://CRAN.R-project.org/package=tree