International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Empirical Agent-Based Modelling of Everyday Pro-environmental Behaviours at Work

Gary Polhill[a], Tony Craig[a], Amparo Alonso-Betanzos[b],
Noelia Sanchez-Maroño[b], Óscar Fontenla-Romero[b], Adina Dumitru[b],
Mirilia Bonnes[c], Marino Bonaiuto[c], Giuseppe Carrus[c], Ferdinando Fornara[c],
Fridanna Maricchiolo[c], Linda Steg[d], Angela Ruepert[d], Kees Kaizer[d],
Ricardo García-Mira[b]
[a] The James Hutton Institute, Craigiebuckler, Aberdeen, UK
{gary.polhill, tony.craig}@hutton.ac.uk
[b] Universidade da Coruña, A Coruña, Spain
{ciamparo, nsanchez, ofontenla, ricardo.garcia.mira, adina.dumitru}@udc.es
[c] Centro Interuniversitario di Ricerca in Psicologia Ambientale (CIRPA), Rome, Italy
{mirilia.bonnes, marino.bonaiuto}@uniroma1.it
{giuseppe.carrus, fridanna.maricchiolo}@uniroma3.it, ffornara@unica.it
[d] Rijksuniversiteit Groningen, NL {e.m.steg, a.m.ruepert, k.e.keizer}@rug.nl

Abstract: We report on agent-based modelling work in the LOCAW project (Low Carbon at Work: Modelling Agents and Organisations to Achieve Transition to a Low Carbon Europe). The project explored the effectiveness of various backcasting scenarios conducted with case study organisations in bringing about pro-environmental change in the workforce in the domains of transport, energy use and waste. The model used qualitative representations of workspaces in formalising each scenario, and decision trees learned from questionnaire responses to represent decision-making. We describe the process by which the decision trees were constructed, noting that the use of decision trees in agent-based models requires particular considerations owing to the potential use of explanatory variables in model dynamics. The results of the modelling in various scenarios emphasise the importance of structural environmental changes in facilitating everyday pro-environmental behaviour, but also show there is a role for psychological variables such as norms, values and efficacy. As such, the topology of social interactions is a potentially important driver, raising the interesting prospect that both workplace geography and organisational hierarchy have a role to play in influencing workplace pro-environmental behaviours.

***Keywords**:* Backcasting, Agent-Based Modelling, Decision Trees, Social Networks.

1 INTRODUCTION

The LOCAW (Low Carbon at Work: Modelling Agents and Organisations to Achieve Transition to a Low-Carbon Europe) European Framework Programme 7 project included an agent-based modelling workpackage to simulate formalisations of back-casting scenarios aimed at increasing the frequency with which various everyday pro-environmental behaviours are performed.

Everyday behaviour has been a focus of social science research since the early nineteen twenties (e.g. Lukàcs 1977), and interest in it has been growing rapidly since the nineteen eighties following the work of researchers such as Lefebvre (1971) and de Certeau (1984). In particular, norms have been shown to have a role in predicting some pro-environmental behaviours, such as recycling (Nigbur et al. 2010) and household energy use (Nolan et al. 2008), suggesting that agent-based modelling may have a useful contribution to make to the field through its ability to explicitly represent interactions among heterogeneous individuals.

In the LOCAW project, we have focused on everyday behaviours in three domains: waste, energy use and transport. The project featured six case studies: two in heavy industry, studied using life history interviews, and four remaining case studies in the public and utility sectors, studied using mixed methods. It is with the last four that this paper is concerned, as the questionnaire survey forming the quantitative part of the study was used as the basis for providing an empirical basis for decision-making in the agents through the application of decision trees. Although authors such as Gilbert (2006) have been critical of the use of questionnaires as an empirical foundation for agent-based models, Smajgl et al. (2011) argue for their application as a means of populating the agents with individual characteristics.

The remainder of this paper builds on the method used to construct decision trees published elsewhere (Sánchez Maroño et al. 2013), and describes the approach used to provide an empirical foundation for normative influence on everyday behaviour. There are several algorithms for learning decision trees, and the structure of the tree can be sensitive to the algorithm used and any parameters it takes. We demonstrate the use of the model with scenarios developed in the case study of the Italian utility company, Enel Green Power, and show some of the sensitivity associated with decision tree construction method.

2 METHOD

2.1 Questionnaire design

As is customary practice in psychological research, the questionnaire consisted of a number of questions grouped into measures of psychological constructs theorised to be predictors of everyday pro-environmental behaviour, together with questions on the frequency with which respondents perform various behaviours. Both kinds of questions are typically measured using Likert scales; in the case of measures of constructs, averages of the questions in the associated group are used.

The psychological constructs explored were egoistic, hedonistic, altruistic and biospheric values (-1: opposed to my values; 0: unimportant; 1-5: important; 6 very important; 7 of supreme importance); with other constructs measured using a scale from 1 (totally disagree) to 7 (totally agree): world-views, outcome and self-efficacy, local and general descriptive and injunctive norms, personal norms, and identity.

The behaviours measured were grouped into the three domains of transport, energy use and waste, and included questions on behaviours at home and at work. These were measured using Likert scales from 1 (never perform the behaviour) to 7 (always). To create an empirical foundation for injunctive norms, questions were also asked on the frequency with which respondents transmit norms at work encouraging their colleagues to behave pro-environmentally.

2.2 Decision tree construction

As detailed by Sánchez-Maroño et al. (2013) and is common practice in data-mining, preprocessing of the questionnaire data has been done before applying Quinlan's (1993) C4.5 algorithm to construct decision trees. This was conducted in three stages: k-means clustering to simplify the values, feature selection (Hall 1999) to preselect the explanatory variables used to predict each behaviour, and discretisation using PKID (Proportional K-Interval Discretisation; Yang and Webb 2001) to simplify the predicted output.

Here, however, we explore a refinement of this procedure aimed at addressing a particular concern with feature selection when the constructed decision trees are to be used in an agent-based model: ignoring certain explanatory variables during decision tree construction risks leaving out feedback dynamics in the model.

- Rather than use clustering to simplify the values, we adopt the practice of psychologists of taking averages of the responses to questions on each of the four values.
- In feature selection, we compare decision trees constructed using all explanatory variables with those using only variables that had at least a loose correlation with the response variable ($p < 0.1$ as measured by Kendall's tau test).

- Instead of using PKID discretisation, we simply take terciles of the response variable, again, comparing those trees predicting discretised responses with those predicting Likert scales.
- We deploy Galimberti et al.'s (2012) rpartScore() method for constructing decision trees with ordinal response variables.

2.3 Model design

We describe here the model using just the overview part of Grimm et al.'s (2010) ODD (Overview, Design concepts and Details) protocol together with the initialisation section and the submodel describing how `Contexts` are processed each time step.

Purpose

The purpose of this model is to explore the impact of various measures and integrative themes (proposed in backcasting workshops and by the LOCAW consortium) on everyday pro-environmental behaviour in the case studies of Groningen Municipality, The Netherlands and Enel Green Power, Italy.

Entities, state variables and scales

A UML (Unified Modelling Language) diagram is shown in figure 1. The internal dynamics of the model are affected by a formalisation of a `Scenario`. `Contexts`, which can be `Region` and `Post` specific determine the decision made using `data` from applicable `Persons`. (The justification for this somewhat counter-intuitive locus of agency can be related to Shove et al.'s (2012) practice theory.)

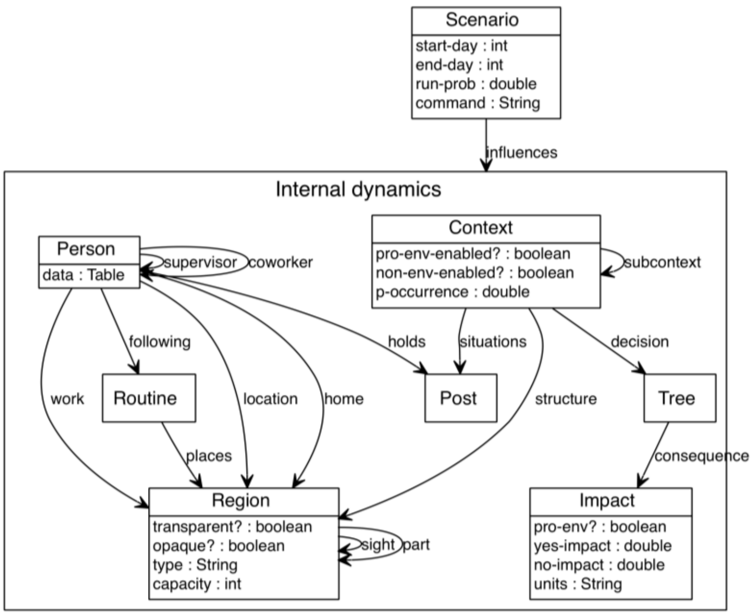


Figure 1. UML (style) class diagram of the WERC-M model

Process overview and scheduling

- `Persons` use their `Routine` to decide which `Region` they should be in
- `Contexts` find `Persons` to process them
    - `Persons` process the `Contexts` they have been recruited to
- Various optional dynamics cause adjustments to `Persons`' data attributes
- Scenarios are implemented
- Workforce is adjusted for retirement and people leaving

Initialisation

The model is initialised from a number of data files describing the components of the model. These files enable considerable flexibility in the specification of the model, and properly an instance of the model should be seen as the program source code plus the data files used to configure it. Initialisation consists of the following steps:

1. The tree file is read. The tree file uses NetLogo reporter-block code executed by the NetLogo `runresult` command to implement the decision tree as a series of nested `ifelse-value` clauses.
2. The data file is read. The values read in by this file are used to populate the `data` attribute of each `Person`.
3. The regions file is read and `Region` visibility is determined. The regions file contains specifications for all the `Regions` that appear in the model, including their `type`.
4. The organogram file is read, which specifies groups of workers by `Post`. (A post is a position or role within the company – e.g. a secretary.) It also contains a matrix used to build the social network, specifying and quantifying the workplace relationships (`coworker` or `supervisor` / `subordinate`) holders of each post have with holders of each other post.
5. The impacts file is read, which describes whether the behaviour is pro- or non-environmental, and, if available, allows quantitative values of impact with specified units to be provided for each choice.
6. The contexts file is read. This specifies the `Contexts` that will be encountered by `Persons` (or groups thereof by `Post` or `Region`) during the course of the model run.
7. The routines file is read. The routines file specifies, for each `Post`, where holders thereof are at different times of the day (here: ‘`early-morning`’, ‘`morning-break`’, ‘`late-morning`’, ‘`lunch`’, ‘`early-afternoon`’, ‘`afternoon-break`’, ‘`late-afternoon`’, ‘`evening`’.)
8. `Persons` are assigned `Posts`.
9. `Persons` are assigned `Routines` based on their `Post`.
10. `Persons` are assigned home and work `Regions` based on the `type`.
11. The social network is constructed.

Submodels: Process context

Processing a Context consists of the following steps:

1. The decision tree associated with the `Context` is run using the `Person`’s attributes to determine the outcome. The decision tree is expected to return a number between 1 and 7 as per the Likert scale on the questionnaire. This is converted to a probability of performing the behaviour using a logistic function [1] as per Kaiser et al.’s (2010) formalisation of Campbell’s (1963) paradigm for addressing the attitude-behaviour gap, the constant –4 being used to ensure the middle response corresponds to *P* = 0.5, and the factor 2 to ensure that there is a reasonable difference in probability from one response to the next without causing the curve to be too shallow:

$$P(behaviour) = \frac{1}{1+e^{-2(response-4)}} \qquad [1]$$

2. The associated `Impact` and `Context` are then used to determine whether the non- or pro-environmental options are disabled.
3. The action performed is then either the chosen action from step 1, or the structurally imposed action from step 2 if the latter has determined that options have been disabled.
4. The `Impact` of the action performed is accumulated. `Contexts` keep a record of the number of times each choice is made, and `Persons` keep records of the number of times they have a choice and what choice is made.
5. If the location of the `Person` is not an opaque `Region`, the model determines which other `Persons` could see this `Person` performing the chosen action. If the `Person` had a choice, the location is a work `Region`, and the `Person` chose the non-environmental option, then the visible `coworkers`, `supervisors` and subordinates of the `Person` transmit norms to the `Person` to behave more pro-environmentally. As a ‘meta’ pro-environmental behaviour, norm

transmission is also done using decision trees learned from questions in the questionnaire asking how often respondents encouraged their colleagues to behave pro-environmentally.

6. If the Context has a `subcontext` corresponding to the action performed, these `subcontexts` are processed immediately.

Descriptive and injunctive local norms from subordinates, supervisors and coworkers are automatically selected to be determined from the model rather than from the agent's questionnaire `data`. To compute these, the model determines whether 'most of my *colleagues* behave/think I should behave pro-environmentally at work' (where *colleagues* corresponds to the relationship in question). The following describes how these are determined for `coworkers`, with `supervisors` and subordinates being determined in a similar way:

- *Injunctive norm*: the proportion of the `Person`'s `coworkers` who have transmitted at least one norm to the agent is scaled to the range [1, 7].
- *Descriptive norm*: the proportion of the `Person`'s `coworkers` who mostly behave pro-environmentally at work is scaled to the range [1, 7]. A Person mostly behaves pro-environmentally at work if more than half of their work behaviours entail the choice of the pro-environmental option.

2.4 Formalisation of back-casting scenario

The back-casting scenario with the Enel Green Power case study identified three scenarios for the future. One simply entailed technological improvements with no significant adjustment in everyday practice in the organisation. Two other scenarios involved more significant changes. In one 'green office' scenario, the company's central offices in Rome were replaced with a number of campuses in suburbs of Rome where employees would live and work in buildings designed to minimise energy consumption. The other 'virtual office' scenario kept the offices in central Rome, but increased the use of home working. For modelling purposes, these scenarios have to be quantified. For now, the quantifications in Table 1 have been used. A baseline scenario in which there are no interventions can act as a control. Differences in timings are to allow for combinations of scenarios (e.g. technical + virtual) to be explored in the same run. Although the back-casting workshop was not specific about when these would occur, it did order technical improvements before changing work location arrangements because of the organisational and planning implications of these changes.

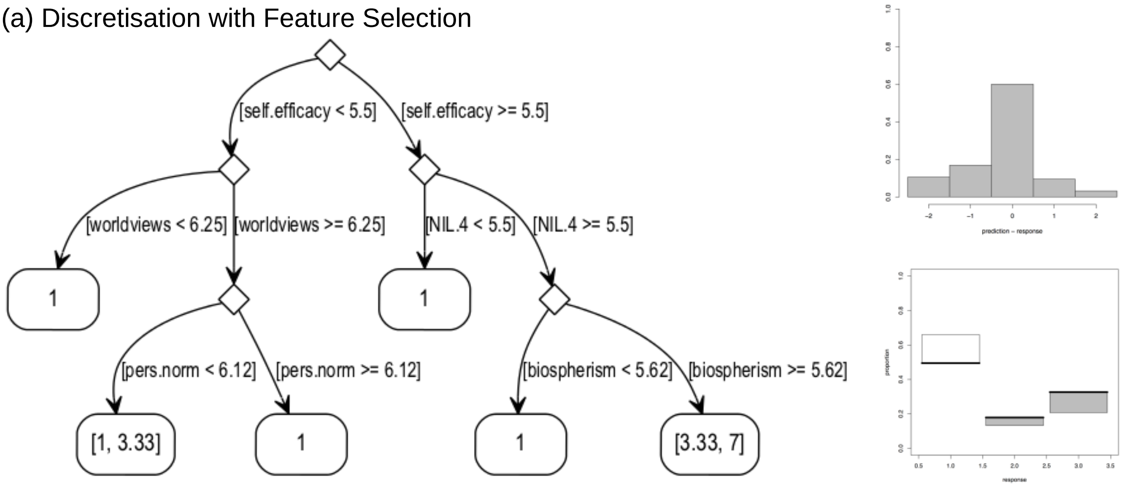
Table 1. Quantification of back-casting scenarios.

Technical Improvement: 2% reduction in impact of each behaviour from year 10 through to year 20.
Green Office: Adopt a policy of recruiting people with a high mean biospheric response (>= 5) from year 10 onwards. Move the normal work and normal home regions of each agent to one of nine campus regions in year 20. On campus, the probabilities of contexts occurring changes, also reducing energy consumption. In particular, there is no commuting, the improved architecture is assumed to reduce the frequency with which heating or air conditioning is required, and there are fewer business trips, but the use of video conferencing increases.
Virtual Office: Each year from year 20 to year 30, move 10% of the workforce with their normal place of work in the company's central offices to having their normal place of work at their home. Here too, the probabilities of contexts occurring changes for workers at the home office, with no commuting, fewer business trips and more use of video conferencing. However, without a change in building infrastructure, there is no effect on heating or A/C demand.

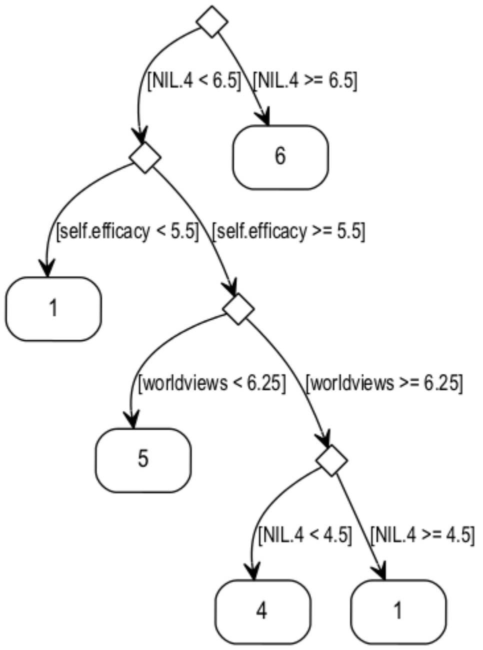
3 RESULTS AND DISCUSSION

The decision trees in figure 3 show the potential differences in structure that arise from using the possible combinations of tercile-based discretisation (or not) and tau test p-value based 'feature selection' (or not), together with graphs showing the fit of the decision tree prediction to the distribution

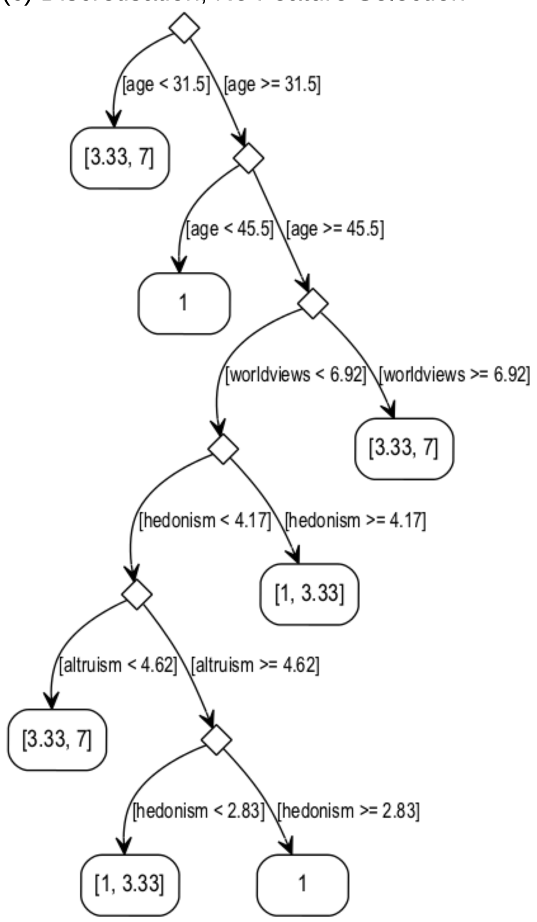
(a) Discretisation with Feature Selection

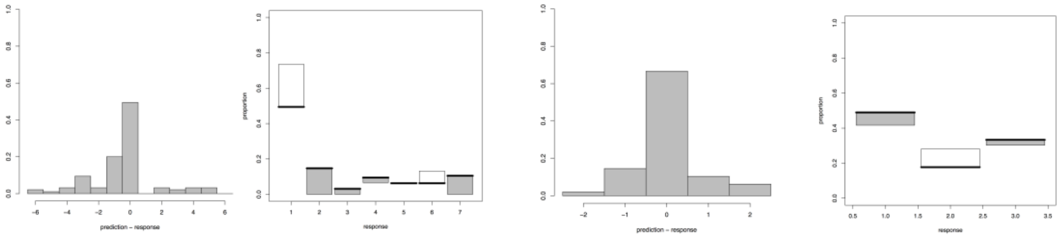


(b) Feature Selection; No Discretisation

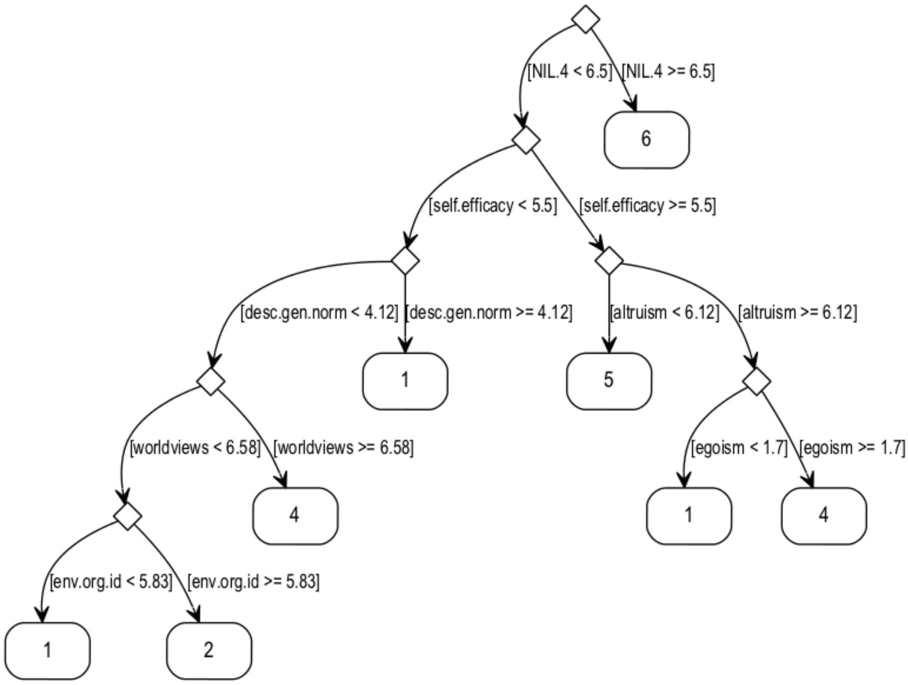


(c) Discretisation; No Feature Selection





(d) No Feature Selection or Discretisation



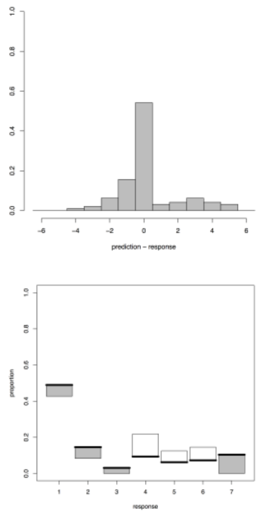


Figure 3. Decision trees arising from different construction approaches for one of the everyday behaviours (car sharing on business trips). In the graphs, the thick line is actual response distribution – white boxes show an overestimate; grey boxes an underestimate.

of respondents. Two graphs are shown. The first shows the distribution of errors in predicted response per individual. The second graph compares the distribution of predicted and actual responses across the population. In the second graph a thick line is drawn showing the actual response for each response value (1-3 for discretised terciles, 1-7 for Likert scales). A box is then drawn to show the deviation from the population response predicted by the decision tree (white: overestimate; grey: underestimate).

Discretisation can change the set of variables that feature selection identifies as being relevant, hence it should be expected that the two feature selection decision trees (figure 3(a) and 3(b)) will be different in terms of the variables used. In this case, there is considerable overlap in the variables, with eight of nine variables selected by this feature selection method being shared between the explanatory variables input to these two trees, though with different p-values to the correlation. The only difference between the sets of variables input to these two decision trees was that the discretised data included ‘biospherism’, and the non-discretised used ‘age’. Correlations were typically weakly significant, with five of the variables having a p-value of between 0.05 and 0.1 in the discretised case, and three in the non-discretised. The decision trees constructed in figure 3(a) and 3(b) end up using mostly the same variables, with the exception that the discretised tree (3(a)) uses biospherism (which was not one of the variables input to the non-discretised tree construction algorithm) and personal-norm (which was). However, despite this the trees differ in structure; for example, the non-discretised response variable tree (3(b)) gives a prominence to injunctive local norms from managers (NIL.4) that is not reflected in the discretised response variable tree (3(a)).

In this case the tree using a discretised response variable without feature selection arguably produces the best fit, with nearly 70% of predictions being in the correct tercile and a reasonably balanced distribution of misclassifications at the population level (with a tendency to underestimate the extremes). Looking at the trees in figures 3(c) and 3(d), although there is some overlap in the choice of variables with the feature selection trees, the discretised response variable tree (3(c)) in particular introduces variables that have a low correlation significance at population level.

For modelling the scenarios, we chose the decision tree for each behaviour with the best fit among the construction options. The method for choosing the tree was qualitative, but across all the behaviours, there was no consistency in whether discretisation, feature selection, or a combination of both produced the tree with the best fit. However, only three of the forty behaviours for which decision trees were generated involved choosing a tree that had being constructed without feature selection or discretisation, and in a number of cases these processes were able to achieve a marked qualitative improvement in misclassification error.

4 CONCLUSION AND FUTURE WORK

Decision trees show great promise for the empirical modelling of agent behaviour on questionnaire data, but as the results show, methods of construction affect their structure. Future work will explore in more detail the differences in such methods, and the effect they have on simulation outputs.

ACKNOWLEDGEMENTS

This work was funded by the European Commission through Framework Programme 7 grant agreement number 265155, and by the Scottish Government Rural Affairs and the Environment Portfolio Strategic Research Theme 4 (Economic Adaptation).

REFERENCES

Campbell, D.T., 1963. Social attitudes and other acquired behavioral dispositions. In Koch, S. (ed.) Psychology: A Study of a Science Volume 6. New York, NY: McGraw-Hill. pp. 94-172.

de Certeau, M., 1984. The Practice of Everyday Life. University of California Press, Berkeley, CA.

Galimberti, G., Soffritti, G., Di Maso, M., 2012. Classification trees for ordinal responses in R: The rpartScore package. Journal of Statistical Software 47 (10) 1-25. http://www.jstatsoft.org/v47/i10

Gilbert, N., 2006. Agent-based social simulation: dealing with complexity. ESRC Research Methods Festival, St. Catherine's College, Oxford, 17-20 July 2006. http://www.ccsr.ac.uk/methods/festival/programme/witha/gilbert.pdf

Hall, M.A., 1999. Correlation-based Feature Selection for Machine Learning. PhD thesis, University of Waikato, Hamilton, New Zealand.

Kaiser, F.G., Byrka, K., Hartig, T., 2010. Reviving Campbell's paradigm for attitude research. Personality and Social Psychology Review 14, 351-367.

Lefebvre, H., 1971. Everyday Life in the Modern World. Penguin, London, UK.

Lukàcs, G., 1977. History and Class Consciousness. Merlin Press, London, UK. (First pub. 1923)

Nigbur, D., Lyons, E., Uzzell, D., 2010. Attitudes, norms, identity and environmental behaviour: Using an expanded theory of planned behaviour to predict participation in a kerbside recycling programme. British Journal of Social Psychology 49, 259-284.

Nolan, J.M., Schultz, P.W., Cialdini, R.B., Goldstein, N.J., Griskevicius, V., 2008. Normative social influence is underdetected. Journal of Personality and Social Psychology 34, 913-923.

Sánchez-Maroño, N., Alonso-Betanzos, A., Fontenla-Romero, Ó., Rodríguez-García, M., Polhill, G., Craig, T., 2013. A decision-making model for environmental behavior in agent-based modeling. Lecture Notes in Computer Science 7902, 152-160.

Shove, E., Pantzar, M., Watson, M., 2012. *The Dynamics of Social Practice*. London: Sage

Smajgl, A., Brown, D.G., Valbuena, D., Huigen, M.G.A., 2011. Empirical characterisation of agent behaviours in socio-ecological systems. Environmental Modelling and Software 26, 837-844.

Yang, Y., Webb, G., 2001. Proportional k-interval discretization for naive-bayes classifiers. Proc. 12th European Conference on Machine Learning 564-575.