Analysis of time series data on the economic and ecological components of vegetation by using predictive clustering trees

Debeljak Marko[1], Geoff R. Squire[2], Dragi Kocev[1], Cathy Hawes[2], Mark W. Young[2], Sašo Džeroski[1]

[1] Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia. E-mail: firstname.lastname@ijs.si

[2] Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK. E-mail: firstname.lastname@scri.ac.uk

In this work, we present an approach to modelling interdependent time series variables that support different functions in a managed ecosystem. The data contains information about the crop structure (spring or winter rape, conventional or genetically modified) and the management regimes (crop history, crop characteristics, data on crop cover and height), representing the capture of resource to build economic offtake (e.g. by grass, wheat, oilseeds), as well as on weeds, representing the capture of resource to support the food web (vegetative, flowering and seeding 'weeds'). For optimal agricultural management, crops and weeds should coexist, in a way that the 'weeds' should not strongly reduce the economic output of the crops, while the crops and their management should not reduce the 'weeds' to a point where the food web is impaired. The general limits within which the two might fluctuate without serious loss of function have been established experimentally. There are two open questions: whether crops and weeds can coexist within their limits in the same field, and if so, which factors encourage this coexistence. We analyse paired time series data, obtained by measuring the percentage of cover of crop and weed vegetation every, 7 to 14 days for 5 months at 130 sites in the UK. The sites cover a wide range of localities and management regimes. We first cluster the crop cover and crop height time series (separately) by using the k-medoids algorithm. Roughly five clusters for each are used as representative profiles: Each of the two clusterings gives rise to an additional attribute used in the further analysis. We then learn predictive clustering trees (generalization of decision trees) to predict the weed profile: these identify the factors (vegetation, crop history, herbicide management) that most influence the weed profiles. This helped us to identify the conditions under which coexistence of weeds and crops can be achieved. The practical use of the produced models will be to guide further field experiments and to monitor ecological changes (due to, e.g., the cultivation of biotech crops).