Statistical analysis of oilseed rape dispersion data along a road network.

Katarzyna K. Adamczyk, Sandrine Pivard, Annie A. Bouvier, Jane Lecomte, Pierre-Henri Gouyon, Sylvie Huet

HAL Id: hal-02831483
https://hal.inrae.fr/hal-02831483v1

Submitted on 7 Jun 2020

Statistical analysis of oilseed rape dispersion data along a road network

Katarzyna Adamczyk[1][*] & Sandrine Pivard[2] & Annie Bouvier[1] & Jane Lecomte[2] &
Pierre-Henry Gouyon[2] & Sylvie Huet[1]

[1]*INRA, MIA Department, Jouy-en-Josas*
[2]*Laboratory of Ecology, Systematic and Evolution, University Paris-Sud*

[*] email address : Katarzyna.Adamczyk@jouy.inra.fr

The release of genetically modified oilseed rape may involve some undesirable effects for the environment. Herbicide resistant cultivars may transfer the resistance gene to the conventional crops or some weedy species. The risk of transgene spread is amplified by the presence of abundant *feral populations* of oilseed rape growing on the road verges. In order to study the origin and the dynamics of these populations a ground survey has been conducted in an agricultural region of winter oilseed rape production in the center of France each year from 2000 to 2003. The oilseed rape fields and the feral populations were located by a GPS system on about a 100 km long road network and their characteristics were taken.

We present the statistical methods used to analyse the results of this ecological survey. The set of the data contains the information about the factors likely to affect the feral population occurence and persistence. Because of the large scale of the survey some possibly important variables are not available, like transport intensity or weeding treatment. The purpose of the analysis is to model the probability of feral population occurence in 2003 conditionnally on the explanatory variables taking into account spatio-temporal origin of the feral population and dealing with the unknown sources of the variability. We begin with the exploratory analysis using the Random Forest algorithm for ranking the variables. Then we propose the Mixed Effect Logistic Model to describe the relationship between the probability of feral plant occurence and the explanatory variables.

Key words : oilseed rape, risk assessment, Random Forest algorithm, Mixed Effect Logistic Model

Bibliographie

[1] Pessel, F. D., Lecomte, J., Emeriau, V., Krouti, M., Messan, A., Gouyon, P-H. (2001) Persistence of oilseed rape in natural habitats : consequence for release of transgenic crops.*Theoretical and Applied Genetics* 102, 841-846.

[2] Pinheiro, J.C. , Bates, D.M. (2000) *Mixed-effects model in S and S-Plus. Statistics and computing.* Springer-Verlag.

[3] Breiman, L. (2001) Random Forests. *Machine Learning* 45, 5-32.