



HAL
open science

Comment la fouille de données peut renouveler la démarche de construction des indicateurs agro-environnementaux ? Exemple de l'indicateur Icarab

Abdelhak Rouabah, Françoise Lasserre-Joulin, Jean Villerd, Béatrice Serradj-Noirtin, Helmut Meiss, Régis Wartelle, Julie Maillet-Mezeray

► To cite this version:

Abdelhak Rouabah, Françoise Lasserre-Joulin, Jean Villerd, Béatrice Serradj-Noirtin, Helmut Meiss, et al.. Comment la fouille de données peut renouveler la démarche de construction des indicateurs agro-environnementaux ? Exemple de l'indicateur Icarab. Groupe Français des Pesticides, May 2017, NANCY, France. hal-02918921

HAL Id: hal-02918921

<https://hal.science/hal-02918921>

Submitted on 21 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comment la fouille de données peut renouveler la démarche de construction des indicateurs agro-environnementaux ? Exemple de l'indicateur Icarab

Abdelhak ROUABAH (2), Françoise LASSERRE-JOULIN (1), Jean VILLERD (1),
Béatrice SERRADJ-NOIRTIN (1), Helmut MEISS (1), Régis WARTELLE (3), Julie
MAILLET-MEZERAY (4)

(1) UMR 1121 Université de Lorraine / INRA Nancy-Colmar, 2 avenue de la Forêt de Haye TSA
40602 54518 Vandoeuvre-lès-Nancy ; francoise.lasserre@univ-lorraine.fr ; jean.villerd@univ-lorraine.fr ; beatrice.noirtin@univ-lorraine.fr ; helmut.meiss@univ-lorraine.fr

(2) INRA, UMR 1137 Ecologie et Ecophysiologie Forestière, Campus des Aiguillettes, Faculté des
Sciences et Techniques, 54506 Vandoeuvre-lès-Nancy cedex abdelhak.rouabah@univ-lorraine.fr

(3) Chambre Régionale d'Agriculture des Hauts de France 19 bis rue A. Dumas 80096 Amiens
cedex 03 r.wartelle@hautsdefrance.chambagri.fr

(4) Arvalis Institut du végétal pendant la période 2009-2013 jmezeray@yahoo.fr

Introduction :

La biodiversité est un concept complexe à décrire, avec différents niveaux et échelles emboîtés (Noss, 1990). De ce fait, il existe encore très peu de modèles mathématiques permettant de prédire la présence et l'abondance d'espèces, notamment en fonction des pratiques agricoles et des facteurs environnementaux.

C'est pourquoi la recherche sur les indicateurs agro-écologiques est en plein essor. Elle peut venir en soutien à l'évaluation des systèmes agricoles et à la prise de décision nécessaire quand les modèles complexes (et plus précis) ne sont pas encore disponibles. Une particularité de ce type d'outil est d'utiliser des données facilement accessibles et d'être utilisable par des non spécialistes.

Les premières générations d'indicateurs ont été élaborées sur la base des connaissances des experts intégrées dans des arbres de décision, ce qui constitue le noyau dur de l'analyse multicritère. Cette dernière demeure un outil précieux car elle permet d'intégrer l'ensemble des connaissances disponibles dans un domaine donné et de contourner les manques d'information en sélectionnant des variables facilement accessibles et corrélées avec les variations du phénomène décrit. Cependant,

l'agrégation des données demeure une étape complexe, sujette à un degré d'imprécision difficilement mesurable (Girardin et al, 1999).

Dans cette étude, nous avons eu recours aux méthodes de fouille de données liées à l'intelligence artificielle pour construire un nouveau type d'indicateurs, non plus basé sur la connaissance d'experts (méthode de type *top-down*), mais sur des mesures recueillies sur le terrain (méthode de type *bottom-up*). Le prototype d'indicateur agro-écologique construit, intitulé Icarab, a pour objectif de prédire la présence de carabes, selon leurs caractéristiques fonctionnelles liées à la prédation (taille, régime alimentaire) en fonction des pratiques agricoles et de caractéristiques du paysage. Nous en présentons ici la démarche expérimentale qui en a sous-tendu la conception, et les principales caractéristiques de l'indicateur.

Matériel et Méthodes

1/ Sites d'études et méthodes d'échantillonnage:

L'étude a été conduite dans deux régions de France, Picardie et Région Centre, dans le cadre des programmes Casdar « Entomophages en grandes cultures » (2009-11) et Auximore (2012-14). Entre 13 et 15 parcelles par région ont été sélectionnées selon le type de système (biologique/conventionnel), le type de paysage (plus ou moins ouvert), la présence ou non d'une haie ou bande enherbée en bordure, et le type de culture. Dans chacune des parcelles, 4 à 5 lignes de 3 pots Barber ont été installées en fonction de la distance à la bordure de référence : le milieu de la bordure, 5m, 30m et 70m à partir de la bordure dans la parcelle, et au milieu de la parcelle si celle-ci faisait plus de 140m de large. Les pots ont été remplis au deux tiers d'eau, de sel et de quelques gouttes de détergent, et relevés toutes les semaines entre avril et juillet 2010, 2011 et 2012.

Les carabes ont été déterminés à l'espèce, et classés en groupes en fonction de leur taille et de leur régime alimentaire (à dominance carnivore ou phytophage). Une étude récente a effectivement montré que la taille était un proxy intéressant du niveau de prédation (Rouabah et al., 2014). L'activité-densité totale, par classe de taille, de régime alimentaire et, des espèces les plus abondantes (*Pterostichus melanarius*, *Poecilus cupreus*) ainsi que la diversité totale de carabes correspondent aux variables expliquées de l'indicateur Icarab.

2/ Description des pratiques agricoles et caractéristiques paysagères (variables explicatives):

Les pratiques agricoles de l'année 2010 ont été extraites de la Base de données Systemre pilotée par Arvalis Institut du végétal. Il s'agissait notamment du type de travail du sol (labour, travail du sol superficiel, semis direct), du type d'amendement

(minéral, organique...), de l'application ou non de produits phytosanitaires (selon la famille), de la nature de la culture en cours et de son précédent.

Nous avons également introduit les descripteurs paysagers enregistrés en 2010, dont la composition du paysage sur un rayon de 1500 m autour du centroïde de chaque parcelle (pourcentage de surface de chaque culture). Ont été également calculés des indices de configuration du paysage tels que l'hétérogénéité paysagère ainsi que la densité des zones semi-naturelles (représentée par la somme des surfaces correspondantes).

4/ Fouille de données

Le jeu de données a été soumis à des méthodes de construction d'arbres de régression qui permet de proposer des classifications de variables quelle que soit leur nature et leurs relations (De'ath and Fabricius, 2000, De'ath, 2002). Ces méthodes de segmentation peuvent produire des arbres de décision à partir de bases de données avec différents attributs (variables descriptives paysage et pratiques) associés à différents objets (ici, les sites des pots Barber) ; les arbres obtenus sont très proches du prototype d'indicateur recherché. Le package *mypart* du logiciel R (*R DEVELOPMENT CORE TEAM*, 2013) a été utilisé et a permis d'obtenir un degré de précision pour chaque arbre obtenu.

Résultats-Discussion:

1/ structure d'un arbre de décision prédisant la diversité totale des carabes) :

L'arbre de décision de la Figure 1 montre que la proportion de céréales dans le paysage constitue une variable clé, à influence plutôt négative sur la diversité de la communauté complète des carabes. Dans un deuxième niveau, interviennent des variables telles que la proportion de surface en pommes de terre et de surface urbanisées qui elles aussi ont une influence négative sur la diversité. Les valeurs de diversité de carabes correspondant aux feuilles à l'extrémité des branches de l'arbre ne doivent être interprétées qu'en valeurs relatives et non absolues.

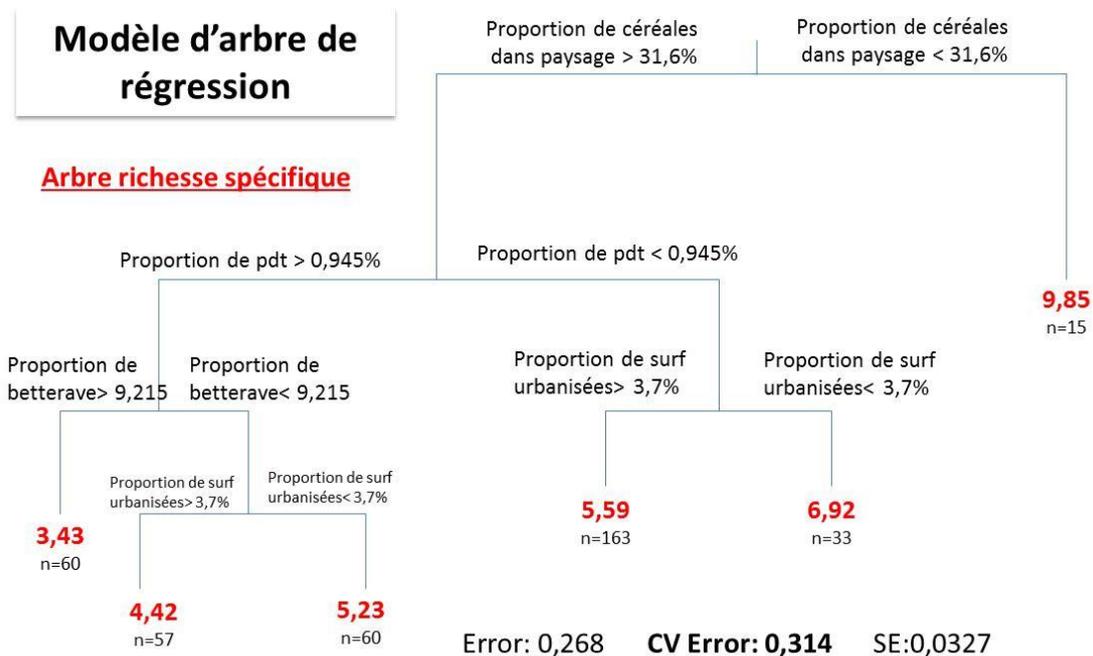


Figure 1 : Exemple d'arbre de régression obtenu pour prédire la diversité des carabes. Les chiffres en rouge correspondent aux moyennes prédites de diversité de carabes, la valeur de n au nombre de situations correspondantes dans la base de données. Les valeurs indiquées pour chaque variable au niveau des branches correspondent à la valeur seuil qui départagent les données en deux groupes (exemple : 31,6% pour le pourcentage de céréales dans le paysage). Les valeurs de CV error indique le degré de confiance (0,686, soit 68,6%) à accorder à l'arbre. Pdt=pomme de terre.

2/ Exemple de sorties de l'indicateur :

L'indicateur Icarab se présente actuellement sous la forme d'un tableur excel facilement mobilisable par un utilisateur non spécialiste. Il est libre de droit. Chaque arbre correspond à un onglet et l'utilisateur peut prendre en compte les résultats par variable ; l'outil délivre également une synthèse sur l'ensemble de ces dernières (Tableau 1)

Tableau 1 : extrait de la feuille « recap_1^{er}_arbres » regroupant l'ensemble des valeurs prédites par les arbres. En rouge, figurent les valeurs plus faibles, considérées à priori défavorables pour l'agriculteur. La colonne « Parcelle Picardie » correspond aux résultats obtenus pour une parcelle de référence en Picardie. Les simulations 1, 2 et 3 correspondent respectivement à une modification de la variable % de céréales dans le paysage, % de zones semi-naturelles et hétérogénéité du paysage pour cette même parcelle.

	N° arbre	Parcelle Picardie	simul 1	simul 2	simul 3
			>41% céréales	>37% Ncrop	>1,7 Het pays
Diversité carabes	arbre 1	9,85	6,92	9,85	9,85
Ab totale sans <i>Pterostichus melanarius</i>	arbre 5	417	155	417	417
Abondance gros carabes (L>13,4 mm)	arbre 8	203	203	48,2	48,2
Abondance gros carabes sans <i>Pterostichus melanarius</i>	arbre 11	23,1	23,1	64,9	64,9
Abondance moyens carabes (9mm<L<13,4mm)	arbre 15	143	3320	143	143
Abondancemoyens carabes sans <i>Poecilus cupreus</i>	arbre 16	11,9	11,9	11,9	11,9
Abondance petits carabes (L<9mm)	arbre 20	332	33,2	332	332
Abondance <i>Pterostichus melanarius</i>	arbre 22	43,1	43,1	43,1	43,1

3/ Discussion : Mode d'utilisation de l'indicateur :

L'intérêt majeur de cet outil réside dans la comparaison relative des situations prédites. Le nombre de carabes prédits (basée sur un cumul de 12 semaines de relevé) ne présente pas de valeur *per se*, et peut présenter inévitablement certains écarts avec la réalité. Par contre, l'outil a montré la validité des écarts entre parcelles présentant des profils contrastés (validation en cours).

L'indicateur Icarab se prête donc bien à l'animation de groupes d'agriculteurs voulant réfléchir collectivement à leurs pratiques agricoles, ainsi qu'à la gestion commune du paysage environnant leurs parcelles agricoles (haies, chemins enherbés,.....).

L'indicateur se prête moins à l'utilisation individuelle par parcelle, surtout s'il n'y a pas de discussion avec un conseiller agricole qui peut aider à mettre en relief et/ou nuancer la portée des résultats en fonction de ceux obtenus par d'autres utilisateurs.

Conclusion

Les méthodes de fouille de données procurent un appui intéressant pour construire des indicateurs prédictifs. Toutefois, celles-ci sont totalement dépendantes de la qualité des données expérimentales d'entrée : il est nécessaire que ces dernières

soient le plus représentatives possibles des situations étudiées, et les extrapolations doivent être réalisées avec beaucoup de prudence.

De telles méthodes basées sur des larges jeux de données de terrain sont complémentaires des indicateurs basés sur les dires d'experts et n'ont pas la prétention de remplacer ces derniers. Elles peuvent permettre de conforter certaines hypothèses qui sous-tendent la structure d'arbres de décision multicritères. Inversement, l'analyse multicritère, basée sur la connaissance des experts, permet de réaliser des regroupements judicieux de certaines classes de variables, voire de choix de variables explicatives pour la fouille de données.

Il apparaît donc pertinent de mener de front ces deux types de démarches pour construire des jeux d'indicateurs agro-écologiques qui puissent se compléter mutuellement et améliorer la qualité des prédictions dans un domaine donné.

Remerciements :

Les auteurs remercient l'ensemble des stagiaires et techniciens qui ont participé au tri et à l'identification des 800 000 carabes récoltés sur les trois années du programme Casdar Entomophage. Ils remercient également Jean-David Chapelin-Viscardi pour les déterminations des espèces de carabes les plus complexes.

Bibliographie

- DE'ATH, G., 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83, 1105–1117.
- DE'ATH, G., FABRICIUS, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81, 3178–3192.
- GIRARDIN, P., BOCKSTALLER C., VAN DER WERF, H.M.G., 1999. Indicators, tools to evaluate the environmental impacts of farming systems. *J. Sust. Agric.* 13, 5-21.
- NOSS, R.F., 1990. Indicators for monitoring biodiversity, a hierarchical approach. *Conserv. Biol.* 4, 355-364.
- R DEVELOPMENT CORE TEAM, 2013. A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN: 3–900051-07–0 <http://www.r-project.org/> (accessed on 26.05.13).
- ROUABAH A., LASSERRE-JOULIN F., AMIAUD B., and PLANTUREUX S. 2014. Emergent effects of ground beetles size diversity on the strength of prey suppression. *Ecological Entomology*, 39, 47-57