Weeds sampling for map reconstruction: a Markov random field approach

Mathieu Bonneau Bonneau, Sabrina S. Gaba, Nathalie Dubois Peyrard Peyrard, Régis Sabbadin

Weeds Sampling for Map Reconstruction: a Markov Random Field Approach

M. Bonneau[1,2], S. Gaba[2], N. Peyrard[1], R. Sabbadin[1]
[1]Unité de Biométrie et Intelligence Artificielle - UR 875
INRA Toulouse - France
[2]UMR Biologie et Gestion des Adventices
INRA Dijon - Université de Dijon - France

1 Introduction

In the past 15 years, there has been a growing interest for the study of the spatial repartition of weeds in crops, mainly because this is a prerequisite to herbicides use reduction. There has been a large variety of statistical methods developped for this problem ([5], [7], [10]). However, one common point of all of these methods is that they are based on in situ collection of data about weeds spatial repartition. A crucial problem is then to choose where, in the field, data should be collected. Since exhaustive sampling of a field is too costly, a lot of attention has been paid to the development of spatial sampling methods ([12], [4], [6] [9]). Classical spatial stochastic model of weeds counts are based on Cox processes [3] or kriging [7]. In this work we propose to deal with abundance classes and to adopt a Markov Random Field (MRF) framework.

In a companion paper [2], we present an approach for spatial sampling which is based on MRF. This approach relies on an a priori model of the repartition of weeds in crops. It also relies on a model of sampling costs (time spent to sample), in order to mimic field constraints. The goal of this talk is to present the modelling choices that we have made in order to apply the approach [2] to the sampling and reconstruction problem for a real case study with a large data set of partial samples, in various conditions (weeds, crop, date...). In section 2 we present the model selection study that we have performed in order to build the a priori MRF model of weeds repartition. Then, in section 2, we present the sampling (time) cost model that we have built. Finally, in section 4 we discuss the use of the sampling approach [2] for weeds sampling in crop fields.

2 A MRF model of the abundance repartition of a weed species

2.1 Candidates pairwise MRF models

Let us recall briefly the definition of a pairwise MRF distribution. Let $X = (X_1, \dots, X_n)$ be discrete random variables taking values in $\Omega^n = \{0, \dots, K\}^n$. $V = \{1, \dots, n\}$ is the set of indices of the vector X and an element $i \in V$ will be called a *site*. If $G = (V, E)$ is the graph associated with the MRF, then $\forall (x_1, \dots, x_n) \in \Omega^n$,

$$\mathbb{P}\big(x_1, \dots, x_n\big) \propto \prod_{i \in V} e^{\psi_i(x_i)} \prod_{(i,j) \in E} e^{\psi_{ij}(x_i, x_j)}.$$

For modeling the map distribution of a particular weed species, we define G as a regular rectangular grid representing a decomposition of the field into quadrats (which are also the sampling units). We considered a first order neighbourhood (2 closest neighbours in each field direction). The variable X_i is the abundance class on quadrat i. For example using Barralis classes [1] : $\Omega = \{0, \dots, 6\}$ with 0 corresponding to the absence of the species. The choice of an appropriate MRF model for mapping weed abundance classes distribution amounts to the choice of adapted potential functions ψ_i and ψ_{ij}. We considered several options : Potts model with or without external field and with or without anisotropy. The more complex is the Potts model with external field and with aniisotropy :

$$\forall (i,j) \in E, \forall k, l \in \Omega \begin{cases} \psi_i(k) & = \alpha_k, \\ \psi_{ij}(k,l) & = \beta_s \mathbb{1}_{\{k=l\}} \text{ if edge } (i,j) \text{ is along tillage direction} \\ \psi_{ij}(k,l) & = \beta_o \mathbb{1}_{\{k=l\}} \text{ if edge } (i,j) \text{ is orthogonal to tillage direction} \end{cases}$$

where α_k, β_s and β_o are real valued parameters. The three other models are derived by setting all the αs equal and/or $\beta_s = \beta_o$. We also considered an alternative to the Potts model, where we impose a smooth spatial variation of the abundance classes : the order-2 potentials are modified as follows

$$\forall (i,j) \in E, \forall k, l \in \Omega \begin{cases} \psi_{ij}(k,l) & = \beta_s (1 - \frac{|k-l|}{K}) \text{ if edge } (i,j) \text{ is along tillage direction} \\ \psi_{ij}(k,l) & = \beta_o (1 - \frac{|k-l|}{K}) \text{ if edge } (i,j) \text{ is orthogonal to tillage direction} \end{cases}$$

2.2 Model selection

The analysis was performed on 6 species, sampled in different cropping systems, at different periods of the year. For each situation, the data available consist of samples of abundance classes of the weed species within a crop field. We used variational versions of the EM algorithm and the BIC criterion [8] to estimate the parameters of each of the eight candidate models and estimate their BIC score. We obtained the following conclusions : *i*) for a large majority of situations, the isotropic Potts model without external field is the best candidate to represent abudance maps distribution, and *ii*) the MRF model with smooth variation is clearly not adapted. The latter conclusion is in coherence with results from the litterature which claim that variations of weeds abundances are often abrupt within a crop field.

3 Cost of sampling

Sampling is adaptive and divided into H steps. One quadrat is sampled at each step. The cost incured to sampling plan A defines the effort necessary for executing this sampling plan. From discussions with experts, we defined this cost based on the time spent to execute A. If $A = \{a_1, \dots, a_H\}$ are the indices of observed quadrats, we suppose that the overall time cost, denoted $c(A)$, is the sum of times spent for observing each quadrat. That is $c(A) = \sum_{i=1}^{H} c(a_i)$. We propose a linear model which expresses the time spent for observing a quadrat as a function of variables $(Z_1, \dots, Z_5)$, representing respectively : the period of observation, the number of weed individuals, the number of species, crop and farming practices. Period of observation is a binary variable with value $\{$*favorable, unfavorable*$\}$ depending of the recovery stage of the crop. We consider five different farming practices, depending on the quantity of pesticide used. For fitting the parameters of this model we use a 18300 length dataset which is a result of a nine-years experiment in Dijon-Epoisses. Eight different crops have been tested. Coefficients of the linear model were fitted using a linear regression with the software R.

4 Applying LSDP to weeds sampling

Once a model of the abundance repartition of a weed species is established, it can be used for finding new sampling policies which realise a trade-off beetween quality of the reconstructed map and cost of sampling. One way to compute this policy is based on the LSDP algorithm described in [2]. The main constraint for applying LSDP to weeds sampling problems is the large number of quadrats within a field. For now the LSDP algorithm gives interesting results for problems with 100 quadrats which is much less than the possible number of quadrats within a field. For exemple [11] report that the size of experimental fields usually varies between 0.019 and $173ha$. The same authors report that a quadrat size varies usually between 0.025 and $1.46m^2$. For solving this problem one solution is to divide the overall field into subfields and solve the sampling problem into each subfield. Another possibility is to combine heuristic(s) strategy(ies) and the LSDP algorithm for solving the sampling problem. This two approaches are currently investigated.

5 Discussion

In this work, we propose an alternative to classical kriging approaches or point processes models for representing the spatial distribution of weeds abundance. This seems to be more adapted to the observed non smooth spatial variation of weeds abundance. We are currently testing this hypothesis by extending our model selection work to a new candidate : the log normal Cox process [3]. It could also be worth investigating the adaptation of our Reinforcement-Learning approach [2] to propose sampling strategies relying on this latter model.

Then, the combination of simple heuristic strategies and more complex ones (like the LSDP one) should lead to a promising avenue for designing spatial sampling strategies for weeds with a satisfying trade-off between evaluation complexity and map reconstruction quality.

Références

[1] G. Barralis. Méthode d'étude des groupements adventices des cultures annuelles. In *Colloque International sur l'écologie et la Biologie des Mauvaise herbes*, pages 59–68, Dijon-France, 1976.

[2] M. Bonneau, N. Peyrard, and R. Sabbadin. A reinforcement-learning algorithm for sampling design in markov random fields. In *Proc. of SSIAB 2012*.

[3] A. Bourgeois, S. Gaba, N. Munier-Jolain, B. Borgy, P. Monestiez, and S. Soubeyrand. Inferring weed spatial distribution from multi-type data. *Ecological Modelling*, 226 :92–98, 2012.

[4] R.D. Cousens, R.W. Brown, A.B. McBratney, B. Whelan, and M. Moerkerk. Sampling strategy is important for producing weed maps : a case study using kriging. *Weed science*, 50(4) :542–546, July 2002.

[5] M.R.T. Dale, P. Dixon, M.J. Fortin, P. Legendre, D.E. Myers, and M.S. Rosenberg. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25(5) :558–577, October 2002.

[6] J. de Gruijter, D. Brus, M. Bierkens, and K. Knotters. *Sampling for Natural Resource Monitoring*. Springer, 2006.

[7] J.A. Dille, M. Milner, J.J. Groeteke, D.A. Mortensen, and M.M. Williams. How good is your weed map ? a comparaison of spatial interpolators. *Weed science*, 51(1) :44–55, January 2003.

[8] F. Forbes and N. Peyrard. Hidden markov random field model selection criteria based on mean field-like approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 :1089–1101, 2003.

[9] WG Müller. *Collecting spatial Data*. Springer Verlag : Heidelberg, 2007. 3rd ed.

[10] J.N. Perry, M. Liebhold, S. Rosenberg, J. Dungan, M. Miriti, A. Jakomulska, and S. Citron-Pousty. Illustrations and guidelines for selecting statistical methods for quatifying spatial pattern in ecological data. *Ecography*, 25 :578–600, 2002.

[11] L.J. Rew and R.D. Cousens. Spatial distribution of weeds in arable crops : are current sampling and analytical methods appropriate ? *Blackwell Science Ltd Weed Research*, 41 :1–8, 2001.

[12] L.J. Wiles. Sampling to make maps for site specific weed management. *Weed science*, 53(2) :228–235, March 2005.