



HAL
open science

Échantillonnage sous contraintes en viticulture de précision

Baptiste Oger, Bruno Tisseyre, Philippe Vismara

► **To cite this version:**

Baptiste Oger, Bruno Tisseyre, Philippe Vismara. Échantillonnage sous contraintes en viticulture de précision. JFPC 2018 - 14es Journées Francophones de Programmation par Contraintes, Jun 2018, Amiens, France. pp.119-122. lirmm-01924365

HAL Id: lirmm-01924365

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01924365v1>

Submitted on 15 Nov 2018

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.

Echantillonnage sous contraintes en viticulture de précision

Baptiste OGER^{1,2} * Bruno TISSEYRE² Philippe VISMARA^{1,3}

¹ MISTEA, Montpellier SupAgro, INRA, Université Montpellier, France

² ITAP, Montpellier SupAgro, IRSTEA, Université Montpellier, France

³ LIRMM, Montpellier SupAgro, Université Montpellier, CNRS, France
{baptiste.oger, bruno.tisseyre, philippe.vismara}@supagro.fr

Résumé

Le développement de l'agriculture de précision nécessite de collecter des données par échantillonnage de manière plus rationnelle. Cet échantillonnage s'appuie sur des outils statistiques qui n'intègrent pas toujours les contraintes opérationnelles. L'objet de cette étude est d'utiliser le raisonnement par contraintes pour optimiser l'échantillonnage dans une parcelle de vigne. Il s'agit à la fois de choisir, dans la parcelle, les points qui respectent les contraintes d'échantillonnage tout en optimisant le déplacement du technicien. C'est donc un problème de tournée, pour lequel la Programmation par Contrainte a développé de nombreux algorithmes de filtrage ces dernières années, notamment la contrainte WeightedSub-Circuit [Briot et al. 2017] bien adaptée aux tournées ne passant que par un sous-ensemble de points. Nous proposons ici un modèle pour une version simplifiée du problème et les premiers résultats obtenus avec le solveur Choco.

Abstract

The development of precision agriculture pushes growers to collect sampling data more efficiently. Sampling methods rely on statistical tools that do not always consider operational constraints. The purpose of this study is to use Constraint Programming to optimize sampling in vineyard block. It is both about choosing sampling points that meet operational constraints in the vineyard and optimizing the path taken by the practitioner. It is thus a routing problem for which Constraint Programming developed many filtering algorithms in the last few years, in particular the constraint WeightedSub-Circuit [Briot et al. 2017], well suited for routes connecting a small subset of points. We expose in this paper a first model for a simplified approach of the problem as well as the first results that we get from the Choco solver.

*Papier doctorant : Baptiste OGER^{1,2} est auteur principal.

1 Introduction

Malgré l'apparition de nouvelles méthodes d'acquisition de données, de nombreuses mesures continuent d'être effectuées par échantillonnage manuel dans les parcelles agricoles. Initialement effectué de façon aléatoire, cet échantillonnage peut aussi s'appuyer sur des outils statistiques et intégrer différentes contraintes opérationnelles, en particulier associées à l'environnement structuré des rangs de vigne. Afin de mettre au point une méthodologie d'échantillonnage en viticulture de précision, cet article s'intéresse à la conception d'un outil d'optimisation sous contraintes capable de déterminer les points à échantillonner tout en optimisant le trajet du technicien. Ce type de problème se prête bien à la programmation par contraintes, d'autant plus qu'il s'agit de pouvoir faire évoluer le modèle afin d'y incorporer différents critères.

En viticulture, l'estimation du rendement est importante pour faciliter l'organisation des vendanges. Des études récentes ont mis en évidence une corrélation entre le rendement et le NDVI (Normalized Difference Vegetation Index) qui fournit un indice de couverture végétale en chaque point de la parcelle [4]. Le NDVI est ainsi une donnée auxiliaire, c'est à dire une donnée reliée à la variable que l'on cherche à estimer. Elle est accessible à moindre coût par télédétection avec un niveau de précision en chaque point qui dépend de la résolution de l'image. En travaillant sur des parcelles pour lesquelles cette donnée auxiliaire est disponible, on souhaite exploiter cette corrélation pour orienter le choix des points d'échantillonnage.

Une contrainte majeure de l'échantillonnage est le temps disponible par parcelle. Celui-ci est généralement assez faible et le temps passé pour les déplace-

ments du technicien dans les rangs de vigne se fait au détriment du nombre de mesures. Ce n'est pas sans conséquences sur la qualité de l'estimation finale. L'optimisation du parcours emprunté par l'opérateur est donc une part importante du problème. Cette optimisation s'apparente à un problème de tournée de véhicule (VRP) [11].

Depuis plusieurs années, la Programmation par Contraintes permet de résoudre efficacement des problèmes faisant intervenir des contraintes de circuit, notamment quand elles sont combinées à d'autres types de contraintes. C'est le cas pour des problèmes de voyageur de commerce (TSP) ou des problèmes de tournée de véhicule (VRP) [9, 5, 7, 1, 6]. Ces contraintes ont été appliquées à de nombreux domaines y compris en viticulture de précision [2].

La particularité du problème d'échantillonnage est qu'il mêle problème de tournée et sélection des points de passage parmi un ensemble plus large de sites candidats. La contrainte WSC (WeightedSubCircuits) [3] permet justement de gérer ce genre de situation où les points de passage sont soumis à d'autres contraintes.

La suite de l'article présente une formalisation du problème d'échantillonnage en viticulture de précision et une première modélisation sous la forme d'un problème d'optimisation sous contraintes. Nous présentons ensuite quelques résultats préliminaires obtenus avec le solveur Choco sur des données obtenues sur l'unité expérimentale de Pech Rouge (INRA).

2 Le problème de l'échantillonnage en viticulture de précision

L'estimation précoce du rendement est une donnée fondamentale pour l'organisation logistique des vendanges. Cette opération, réalisée quelques jours en amont des vendanges permet d'optimiser les opérations de récolte (gestion des équipements, de la main d'œuvre etc.) et la qualité du produit. Dans ce but un opérateur est envoyé sur la parcelle pour faire des mesures de comptages sur plusieurs sites d'échantillonnage. Ces quelques mesures de rendement ponctuelles permettent ensuite d'obtenir une estimation globale du rendement de la parcelle.

Le nombre de sites qui seront échantillonnés au cours du parcours est supposé fixé à une valeur N donnée. L'objectif sera d'effectuer les N mesures le plus rapidement possible mais on pourra faire varier N pour déterminer le nombre d'échantillonnages réalisables dans un temps donné.

Afin de pouvoir optimiser le chemin parcouru par l'opérateur lors de son travail d'échantillonnage, il est nécessaire de calculer les distances effectives entre deux points de la parcelle. Les rangs des parcelles de vignes

étant généralement palissées, il n'est pas possible de couper à travers la parcelle. La distance entre deux sites d'un même rang sera donc la distance euclidienne convertie en temps de parcours. Dans le cas où les deux sites ne sont pas dans le même rang le chemin qui les sépare passe par le bord de la parcelle. Cette matrice de distances (en temps) sera notée \mathcal{M} par la suite.

Les points d'échantillonnage sont choisis parmi un ensemble de sites candidats. Il s'agit d'un sous-ensemble de pieds de vigne qui peuvent être répartis sur une grille régulière ou obtenus par d'autres méthodes. Par exemple, la figure 1 montre une parcelle expérimentale où 53 sites candidats ont été échantillonnés afin de pouvoir tester différentes méthodes d'échantillonnage sur un sous-ensemble de ces points.



FIGURE 1 – Parcelle avec les sites d'échantillonnage candidats (rouge) et les extrémités de rangs (bleue).

L'intérêt d'utiliser une variable auxiliaire comme le NDVI pour l'échantillonnage est double : il s'agit d'une part d'aider à sélectionner les points d'échantillonnage les plus pertinents et, d'autre part, d'établir la corrélation entre les données mesurées et les données auxiliaires. Cette corrélation permettra ensuite d'obtenir une estimation de la variable d'intérêt (le rendement) pour l'ensemble des points pour lesquels on connaît la valeur de la variable auxiliaire.

Pour choisir les points suivant cette approche on peut s'inspirer de la méthode de Kennard & Stone [8]. Cette méthode commence par la sélection des deux points les plus distants dans l'espace des variables auxiliaires. Dans le cas étudié ici, avec une seule variable auxiliaire (le NDVI), cet espace n'aura donc qu'une seule dimension et les points choisis correspondront aux valeurs minimales et maximales du NDVI. Les points suivants sont ensuite choisis de manière itérative en sélectionnant à chaque fois le point dont le NDVI sera le plus éloigné des valeurs déjà sélectionnées, toujours dans l'espace des variables auxiliaires. À chaque étape, un seul point d'échantillonnage est sélectionné. Dans la section suivante, nous proposerons une

variante de cette approche afin de proposer plusieurs points candidats à chaque étape, ce qui permettra de les sélectionner d'après les autres contraintes.

3 Modélisation par un COP

Le problème de l'échantillonnage pour l'estimation du rendement peut être modélisé comme un problème d'optimisation sous contraintes (COP). Il s'agit d'un modèle préliminaire qui doit permettre de tester différentes variantes du problème d'échantillonnage.

3.1 Variables et domaines

On suppose que les K sites d'échantillonnage candidats sont numérotés de 0 à $K - 1$. La position initiale du technicien est associée au site K .

Un ensemble de $N + 1$ variables $\{P_i\}_{i \in 0..N}$ désignera les N sites qui seront échantillonnés. La variable $P_N = K$ correspond au point de départ du technicien. Chacun des autres P_i désigne un point d'échantillonnage ayant un certain niveau de NDVI (imposé par la contrainte 2) mais sans lien avec l'ordre de parcours. $D(P_i) = \{0, \dots, K - 1\}$ pour $i < N$.

Pour décrire le trajet du technicien, on définit $K + 1$ variables telles que $\forall j \in 0 \dots K, D(Next_j) = \{0, \dots, K\}$. La variable $Next_j$ désigne le point suivant j dans la tournée du technicien.

Si le technicien ne visite pas le point j alors la contrainte 4 (WSC) imposera $Next_j = j$.

La variable $Cost$ est la variable à optimiser. Elle correspond au coût (en temps) du trajet du technicien.

Pour faciliter l'écriture du modèle, on introduit une variable ensembliste $Visited \subseteq \{0, \dots, K\}$ contenant l'ensemble des points échantillonnés.

Une contrainte de channeling (Union dans le solveur Choco) permet de la lier aux variables P_i :

$$Visited = \{j \in 0 \dots K \mid \exists i \in 0..N, P_i = j\} \quad (1)$$

3.2 Prise en compte de la variable auxiliaire

L'objectif est de choisir les P_i pour que les valeurs auxiliaires (NDVI) associées à ces points soient réparties au mieux dans l'espace des variables auxiliaires.

Considérons le cas simple à une dimension (NDVI). On peut déterminer l'ensemble des points choisis par la méthode de Kennard & Stone [8]. Comme le montre la figure 2, il s'agit d'un ensemble de points qui subdivisent successivement l'espace des NDVI en deux. On définit pour cela une fonction ϕ_i telle que :

$$\begin{aligned} \phi_0 &= NDVI_{min} & \phi_1 &= NDVI_{max} \\ \forall i \in 2, \dots, N - 1, \phi_i &= \frac{1 + 2^{\lambda(i-1)}}{2^{\lambda+1}} \times (\phi_1 - \phi_0) + \phi_0 \end{aligned}$$

avec $\lambda = \lfloor \log(\frac{i-1}{\log(2)}) \rfloor$

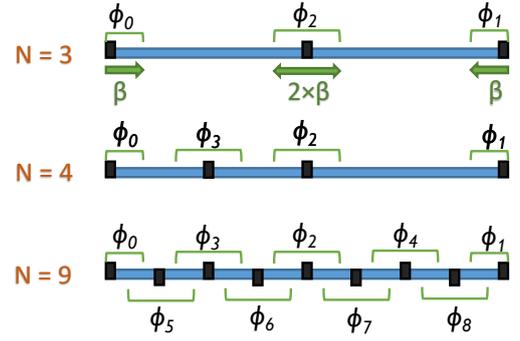


FIGURE 2 – Répartition des valeurs dans l'espace de la variable auxiliaire, suivant le nombre N d'échantillons.

Dans notre modèle, chaque site d'échantillonnage P_i sera choisi parmi l'ensemble des sites candidats ayant un NDVI proche à plus ou moins une constante β près de la valeur ϕ_i . Il s'agit donc d'une contrainte unaire sur les variables P_i :

$$|NDVI_{P_i} - \phi_i| < \beta \quad (2)$$

Par ailleurs, suivant la valeur de β , les domaines des $P_i, i \in \{0, \dots, N - 1\}$, peuvent se recouper. On pose donc une contrainte de différence :

$$AllDifferent(P_0, \dots, P_N) \quad (3)$$

3.3 Contrainte de tournée

L'optimisation du trajet du technicien est gérée par la contrainte WeightedSubCircuit (WSC) de [3] qui est utilisée ici avec un seul circuit¹ :

$$WSC[\mathcal{M}](Visited, \{Next_j\}_{j \in 0..K}, Cost) \quad (4)$$

Cette contrainte garantit que l'ensemble des $Next_i \neq i$ définissent un circuit Hamiltonien de coût $Cost$ (d'après les temps de parcours \mathcal{M}) sur l'ensemble des points contenus dans $Visited$.

4 Résultats

Le modèle décrit précédemment a été implémenté avec le solveur Choco [10] dans le langage Java. Il a été testé sur des données provenant d'une parcelle du vignoble de l'INRA Pech Rouge dans l'Aude (figure 1). Les calculs ont été effectués sur une machine Linux avec Intel(R) Xeon(R) CPU E5-2680 2.40GHz. La table 1 donne le résultat des expérimentations. La première colonne donne le nombre de site d'échantillonnage à instancier, la deuxième colonne le paramètre β qui influe sur la taille des domaines, la taille

1. la variable Z de la contrainte initiale est ici confondue avec $Cost$. De plus, l'ensemble Set_{dummy} des sommets non visités n'apparaît pas puisqu'il est égal au complémentaire de $Visited$

N	β	$ D(P_i) $	Temps CPU	Nœuds
5	5	3.80	0.2	452
5	10	7.80	1	6 531
5	15	13.00	5	31 537
6	10	7.17	2	23 444
6	15	11.83	19	153 056
7	10	7.14	8	59 763
7	15	11.29	111	640 643
8	10	9.50	100	1 119 260
8	15	13.37	1277	9 623 221
9	10	9.56	552	3 211 244
9	15	13.33	9 569	49 990 919
10	10	8.80	441	6 214 723
10	15	12.50	13 614	102 793 574

TABLE 1 – Résultats préliminaires

moyenne de ces derniers pour les N sites est donnée dans la troisième colonne, les dernières colonnes donnent respectivement le temps de calcul (sec) et le nombre de nœuds explorés.

On remarque que le solveur arrive à trouver la solution optimale dans un temps assez court pour des instances allant jusqu'à $N = 7$, ce qui correspond à un nombre d'échantillonnage courant. Au delà, le temps de calcul augmente avec la taille de l'instance mais reste exploitable pour tester des modèles, ce qui est le seul objectif de l'implémentation actuelle.

Il s'agit ici d'un modèle fonctionnel mais simplifié et incomplet du problème d'échantillonnage. Il est destiné à être complété par de nouvelles contraintes afin d'améliorer la prise en compte des contraintes opérationnelles (déplacement, choix des sites d'échantillonnages...), et d'affiner la qualité de l'estimation qui résulte de l'échantillonnage (prise en compte de l'autocorrélation spatiale des données, intégration de nouvelles approches statistiques ...). Dans un second temps, quand le problème sera bien spécifié, plusieurs pistes devront être explorées pour diminuer les temps de calcul (contraintes de symétrie, contraintes redondantes, ordre d'instanciation, ...).

5 Conclusion

Dans cet article, nous avons abordé le problème de l'échantillonnage sous contraintes. Nous avons proposé un premier modèle permettant de choisir les points d'échantillonnage en s'appuyant sur une variable auxiliaire tout en optimisant le trajet du technicien. Ce modèle utilise la contrainte `WeightedSubCircuit` [3] et les variables et contraintes ensemblistes du solveur `Choco`. L'implémentation de ce modèle, sans être très performante, va nous permettre d'explorer différentes variantes d'échantillonnages en viticulture de précision.

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'Avenir portant la référence ANR-16-CONV-0004

Références

- [1] P. Benchimol, W. Jan van Hoeve, J.-C. Régim, L.-M. Rousseau, and M. Rueher. Improved filtering for weighted circuit constraints. *Constraints*, 17(3) :205–233, 2012.
- [2] N. Briot, C. Bessiere, and P. Vismara. A constraint-based approach to the differential harvest problem. In *Proc. 21st International Conference on Principles and Practice of Constraint Programming (CP 2015)*, volume 9255 of *LNCS*, pages 541–556. Springer, 2015.
- [3] N. Briot, C. Bessiere, and P. Vismara. Une contrainte de circuit adaptée aux tournées multiples. In *Actes des Treizièmes Journées Francophones de Programmation par Contraintes (JFPC 2017)*, pages 137–144, 2017.
- [4] E. Carrillo, A. Matese, J. Rousseau, and B. Tisseyre. Use of multispectral airborne imagery to improve yield sampling in viticulture. *Precision Agriculture*, 17(1) :74–92, 2015.
- [5] Y. Caseau and F. Laburthe. Solving small tpsps with constraints. In *Logic Programming, Proceedings of the Fourteenth International Conference on Logic Programming*, pages 316–330, 1997.
- [6] S. Ducomman, H. Cambazard, and B. Penz. Alternative filtering for the weighted circuit constraint : Comparing lower bounds for the TSP and solving TSPTW. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 3390–3396, 2016.
- [7] F. Focacci, A. Lodi, and M. Milano. Embedding relaxations in global constraints for solving TSP and TSPTW. *Annals of Mathematics and Artificial Intelligence*, 34(4) :291–311, 2002.
- [8] W. Kennard and L. A. Stone. Computer Aided Design of Experiments. *Technometrics*, 11 :137–148, 1969.
- [9] G. Pesant, M. Gendreau, J.-Y. Potvin, and J.-M. Rousseau. An exact constraint logic programming algorithm for the traveling salesman problem with time windows. *Transportation Science*, 32(1) :12–29, 1998.
- [10] C. Prud'homme, J.-G. Fages, and X. Lorca. *Choco Documentation*. TASC, INRIA Rennes, LINA CNRS UMR 6241, COSLING S.A.S., 2016.
- [11] P. Toth and D. Vigo. *Vehicle routing : problems, methods, and applications*. SIAM, 2014.