Formalisation et résolution d'un problème en Santé Animale avec le cadre Leader-Follower MDP

A.-F. Viet[1] C. Belloc[1] R. Sabbadin[2]

[1] BioEpAR, INRA, Oniris, Nantes
[2] MIAT, INRA, Toulouse

anne-france.viet@oniris-nantes.fr ; catherine.belloc@oniris-nantes.fr ; regis.sabbadin@inra.fr

Résumé

Le contrôle des maladies transmissibles est une préoccupation majeure du secteur des productions animales. Pour les maladies non réglementées, chaque éleveur décide la mise en place éventuelle d'actions de maîtrise. Pour aider à la coordination entre éleveurs, des organisations proposent des approches collectives de maîtrise, s'appuyant sur des incitations. Nous souhaitons concevoir un outil d'aide à la coordination, basé sur le cadre des Leader-Follower Markov Decision Processes *(LF-MDP). La résolution exacte étant difficile, nous avons proposé des méthodes de résolution exacte et approchée (basée sur l'agrégation d'états). Nous avons appliqué ces méthodes sur le cas d'étude du virus du Syndrome Dysgénésique Respiratoire Porcin. La résolution exacte a pu être effectuée jusqu'à* $n = 20$ *suiveurs. La résolution approchée a pu être effectuée jusqu'à* $n = 100$ *suiveurs et les solutions exacte et approchée sont proches pour* $n \leq 20$.

Mots Clef

Epidémiologie animale, Processus Décisionnel de Markov, Théorie des jeux

Abstract

Transmissible disease control is a major concern for the livestock sector. For unregulated diseases, each farmer chooses to implement control actions. To help coordination between farmers, organizations propose collective control approaches using incentives. Our objective is to design a tool, based on the Leader-Follower Markov Decision Processes *framework (LF-MDP), to help coordination. Exact solution is hard, however we have proposed an exact solution method and an approximate (using state aggregation) one. We have applied these methods to a case study concerning the Porcine Reproductive and Respiratory Syndrome (PRRS) virus. Exact solution is possible for up to* $n = 20$ *followers. Approximate solution could be performed for up to* $n = 100$ *followers and we observed similar exact and approximate solutions for* $n \leq 20$.

Keywords

Animal epidemiology, Markov Decision Process, Game theory

1 Introduction

Le contrôle des maladies transmissibles est une préoccupation majeure pour le secteur des productions animales. On sépare classiquement les maladies épidémiques des maladies endémiques. Les maladies épidémiques ont une propagation souvent très rapide et sont majoritairement réglementées (mise en place obligatoire de mesures de maîtrise). En revanche, les maladies endémiques sont souvent non réglementées et sont présentes à divers niveaux de prévalence. Leur présence induit une diminution de compétitivité des élevages liée à des pertes économiques [7, 2]. Pour ces maladies, chaque éleveur choisit de mettre en place ou non des actions de maîtrise en se basant sur différents critères dont sa perception du risque. Les critères et la perception diffèrent d'un éleveur à un autre. Or, pour les maladies transmissibles, le choix d'une stratégie de maîtrise dans une exploitation va influencer la propagation de l'agent pathogène dans cette exploitation, et donc le risque de transmission aux autres exploitations (par voisinage ou achat d'animaux), créant une externalité positive [12]. D'autres élevages peuvent choisir de ne rien faire du fait des actions de leurs voisins, adoptant alors un comportant de "free rider" [11]. Pour limiter ces comportements, il est nécessaire de coordonner les décisions des élevages. Pour aider à la coordination entre élevages, des organisations professionnelles regroupant un ensemble d'élevages proposent parfois des approches collectives de maîtrise d'une maladie pouvant s'appuyer sur des incitations financières. Comme le collectif n'impose pas d'actions de maîtrise à ses membres, il doit tenir compte du comportement possible des éleveurs. Un outil permettant de définir ces stratégies collectives serait utile. L'enjeu est donc de proposer des approches d'aide à la décision collective tenant compte des prises de décisions des éleveurs. Les décisions sont prises régulièrement (séquentielles) et réajustées en tenant compte de la situation épidémiologique courante (adapta-

tion). On s'intéresse donc aux approches d'optimisation pour la prise de décisions séquentielles en environnement incertain. Les méthodes de l'Intelligence Artificielle sont d'intérêt pour répondre à nos questions. En santé animale, peu de travaux utilisant de telles approches ont été proposés et ils sont souvent limités à un seul décideur.

Du point de vue de l'Intelligence Artificielle, notre question en santé animale peut être vue comme une forme de problème décisionnel multi-agent et séquentiel. Tous les agents (éleveurs) maximisent leurs objectifs propres lorsqu'ils décident de mettre en place ou non une mesure de maîtrise. En parallèle, un pseudo-agent du type leader (l'organisation) a pour objectif de minimiser l'impact de la maladie dans l'organisation en donnant des incitations aux éleveurs pour que leurs comportements se rapprochent de l'optimum collectif. Du fait des interactions très spécifiques entre l'organisation et les éleveurs, nous sommes face à un problème de type *Leader-Follower Markov Decision Process (LF-MDP)* [23]. La résolution des LF-MDP est la plupart du temps réalisée par simulation, par des approches de type *Apprentissage par Renforcement* [23, 10]. Pour envisager une résolution exacte, nous nous sommes intéressés à un sous-problème plus simple, dont les hypothèses étaient acceptables pour répondre à nos questions en santé animale. Nous avons supposé que le nombre de statuts des suiveurs vis à vis de la maladie était petit et que les suiveurs ne sont pas individuellement identifiés par l'organisation, excepté par leur statut vis à vis de la maladie. Cette dernière hypothèse signifie que 2 éleveurs ayant le même statut vont utiliser la même politique (qui peut être stochastique). Nous avons pu résoudre de manière exacte ce problème pour des petits nombres de suiveurs ($n \leq 20$). Pour de plus grandes valeurs, nous avons proposé une méthode de résolution approchée, par agrégation d'états.

L'objectif de ce papier est de présenter la formulation par un LF-MDP du problème de coordination de décision individuelle par une organisation ainsi que les approches développées pour sa résolution exacte et approchée. Après un état de l'art, nous présentons notre contribution à la résolution de LF-MDP et ensuite un cas d'étude permettant d'illustrer notre approche sur une question en santé animale.

2 Etat de l'art

En santé animale, il existe des travaux s'intéressant aux problématiques de coordination de décisions. Le plus souvent, ceux-ci s'intéressent à la modélisation et à la simulation de la propagation d'une épidémie dans une région. Ils se limitent alors à évaluer, par simulation, plusieurs scénarios représentant différentes actions imposées par un collectif et supposées mises en place dans tous les élevages [15, 24]. Il existe toutefois quelques approches pour la conception par optimisation de stratégies de contrôle collectives [9, 13], ou individuelles [25, 19, 20].

Le non suivi des actions proposées par une organisation par une partie des éleveurs est un facteur impactant la propagation [27]. De rares travaux prennent en compte le non-suivi des actions proposées, dans le cadre de l'optimisation de stratégies collectives [26].

Quelques travaux, en économie [23] et en Intelligence Artificielle / Recherche Opérationnelle [28], se sont intéressés à la coordination d'un leader et de plusieurs suiveurs, dans des problèmes de décision. Néanmoins, ces travaux se limitent à des comparaisons par simulation de politiques de coordination prédéfinies ou à l'étude de systèmes ne comprenant que quelques individus ou un seul pas de temps. En négligeant le rôle particulier de l'organisation, des approches de type *Competitive Markov Decision Process* aussi appelées *Stochastic Game* [8] ou de type *Mean Field Game* [14] seraient envisageables pour traiter le problème qui nous intéresse. Du fait des interactions très spécifiques entre l'organisation et les éleveurs, nous sommes plutôt dans les problèmes de type *Leader-Follower Markov Decision Process (LF-MDP)* [23] et *Dynamic Principal-Agent Problem* [18]. Pour résoudre ces problèmes, il est nécessaire de résoudre un grand nombre de jeux à n joueurs, à chaque pas de temps pour construire la politique des n éleveurs. La complexité (non-polynomiale) de cette étape se répercute sur la résolution des LF-MDP. C'est pourquoi la plupart des approches existantes opèrent par simulation, en utilisant des méthodes *d'apprentissage par renforcement* [23, 10]. Mais ces méthodes ne permettent pas de traiter facilement le contrôle des maladies non réglementées en santé animale, où il faut considérer : (1) des horizons de plusieurs pas de temps et (2) plusieurs centaines d'éleveurs en interaction. Nous avons, dans [21, 22], proposé une approche générique de résolution pour ces problèmes, entrelaçant programmation dynamique et calculs d'équilibres de Nash. Cette approche est basée sur une hypothèse simplificatrice de *substituabilité* des suiveurs dans les LF-MDP. Pour le cas d'étude qui nous occupe ici, elle permet de résoudre, de manière exacte, des problèmes avec 20 suiveurs. En utilisant une approximation basée sur l'aggrégation des suiveurs, nous pouvons résoudre, de manière approchée, des problèmes avec au moins 100 suiveurs.

3 Formalisation et résolution

Nous considérons 2 types d'agents : un leader (l'organisation) et des suiveurs (les éleveurs). Le leader cherche à maximiser sa propre fonction d'utilité qui est dépendante de la quantité d'incitations versées (fonction des décisions des suiveurs) et de l'état du système (niveau d'infection dans l'organisation). Les suiveurs maximisent leur propre fonction d'utilité dépendant de leur statut individuel (pertes dues à la maladie) et des actions de maîtrise si elles sont mises en place (coût des actions diminué des incitations du leader si il en distribue). Dans notre situation, le leader influence les fonctions de récompense des suiveurs via les incitations. Les décisions individuelles, de maîtrise ou non, des suiveurs influencent la dynamique du système (propagation de la maladie) et donc indirectement la fonction de récompense du leader.

Nous présentons d'abord le cadre LF-MDP (Section 3.1) proposé par [23]. Puis, nous décrivons brièvement un algorithme de type *Backward Induction* "naïf", de complexité exponentielle en le nombre de suiveurs, permettant de résoudre un LF-MDP (Section 3.2). Enfin, nous rappelons les principes et la complexité des algorithmes de résolution, exacte et approchée, que nous avons proposés dans [22] (Section 3.3).

3.1 Formalisation LF-MDP

Etats, actions, transitions et récompenses. Un LF-MDP [23] modélise un processus de décision séquentielle dans l'incertain impliquant un *leader* et $n \geq 1$ *suiveurs* (*followers*). Il se définit, en horizon fini [1], par : $\mathcal{M} =< n, \Sigma, A^L, \{A_i^F\}_{i=1..n}, T, r^L, \{r_i^F\}_{i=1..n}, H >$, où :

- Σ est l'espace d'états joint du leader et des suiveurs. Dans le cas général, il peut être factorisé : $\Sigma = S^L \times S_1^F \times \ldots \times S_n^F$.
- $A^L = \{1, \ldots, m\}$ est l'ensemble fini des actions du leader.
- $A_i^F = \{1, \ldots, p_i\}$ est l'ensemble fini des actions du suiveur i. Pour simplifier l'exposé, nous supposerons que ces ensembles sont identiques pour tous les suiveurs : $A^F = \{1, \ldots, p\}$.
- $T : \Sigma \times (A^F)^n \times \Sigma \rightarrow [0, 1]$ est la fonction de transition jointe. $T\left(\sigma' | \sigma, \{a_i^F\}_{i=1..n}\right)$ est la probabilité de passer de l'état σ à l'état σ', lorsque les actions des suiveurs sont $a^F = \{a_i^F\}_{i=1..n}$. Notons que l'action du leader n'influence pas directement ces probabilités.
- $r^L : \Sigma \times A^L \times (A^F)^n \rightarrow \mathbb{R}$ est la fonction de récompense instantanée du leader.
- $r_i^F : \Sigma \times A^L \times A^F \rightarrow \mathbb{R}$ est la fonction de récompense instantanée du suiveur i.
- H est le nombre d'étapes de décision.

Stratégies du leader et des suiveurs. Nous supposons, comme classiquement dans les Processus Décisionnel de Markov (MDP, pour *Markov Decision Processes*) à horizon fini, que les agents choisissent leurs actions à l'étape t en suivant des stratégies *non-stationnaires*, $\delta_t^L, \{\delta_{t,i}^F\}_{i=1\ldots n}$. Nous nous restreindrons à des stratégies *Markoviennes*, *stochastiques*, car les stratégies d'équilibre d'un LF-MDP vérifient ces hypothèses [23].

$\delta_t^L(a^L|\sigma)$ est la probabilité que $a^L \in A^L$ soit choisie par le leader à l'étape t, connaissant l'état courant $\sigma \in \Sigma$. $\delta_{t,i}^F(a_i^F|\sigma, a^L)$ est la probabilité que $a_i^F \in A^F$ soit choisie par le suiveur i à l'étape t, connaissant l'état courant σ et l'action a_L choisie par le leader [2].

Dans un LF-MDP, la stratégie optimale du leader est *déterministe* [23, 22] : $\delta_t^L(a^L) \in \{0, 1\}$. Aussi, nous écrivons : $a^L = \delta_t^L(\sigma)$. Dans le cas (non général) où les stratégies des suiveurs sont déterministes, nous écrirons également : $a_i^F = \delta_{t,i}^F(\sigma, a^L)$.

Valeurs d'une stratégie jointe, stratégies d'équilibre. Soit $\Delta = \left\{\delta_t^L, \{\delta_{t,i}^F\}_{i=1\ldots n}\right\}_{t=1\ldots H}$ une stratégie jointe fixée du leader et des suiveurs. Les *valeurs* Q_Δ^L et $Q_\Delta^{F,i}$ de cette stratégie jointe pour le leader et pour les suiveurs sont définies, à chaque pas de temps et dans chaque état joint, par :

$$Q_\Delta^L(\sigma, t) = E\Big[\sum_{t'=t}^{H} r_{t'}^L \mid \Delta, \sigma\Big], \quad (1)$$

$$Q_\Delta^{F,i}(\sigma, t) = E\Big[\sum_{t'=t}^{H} r_{t',i}^F \mid \Delta, \sigma\Big]. \quad (2)$$

Résoudre un LF-MDP consiste à trouver une stratégie d'équilibre, $\Delta^* = \left\{\delta_t^{L*}, \{\delta_{t,i}^{F*}\}_{i=1\ldots n}\right\}_{t=1\ldots H}$, pour le leader et les suiveurs.

Définition 1 (stratégie d'équilibre d'un LF-MDP) *$\Delta^* = \left\{\delta_t^{L*}, \{\delta_{t,i}^{F*}\}_{i=1\ldots n}\right\}_{t=1\ldots H}$ est une stratégie d'équilibre si et seulement si elle vérifie, $\forall t, \delta_t^L, \{\delta_{t,i}^F\}, \sigma \in \Sigma$:*

$$Q_{\Delta^*}^L(\sigma, t) \geq Q_{\Delta^{*\downarrow\delta^L}}^L(\sigma, t), \forall \delta^L, \quad (3)$$

$$Q_{\Delta^*}^{F,i}(\sigma, t) \geq Q_{\Delta^{*\downarrow\delta_i^F}}^{F,i}(\sigma, t), \forall i, \delta_i^F. \quad (4)$$

$\Delta^{*\downarrow\delta^L}$ (resp. $\Delta^{*\downarrow\delta_i^F}$) est une stratégie jointe où les δ_t^{L*} (resp. $\delta_{t,i}^{F*}$) ont été remplacées par des stratégies arbitraires δ_t^L (resp. $\delta_{t,i}^F$), $\forall t$ ($\forall i$).

Dans [22], nous avons exploité les résultats de [23, 8] montrant l'existence d'une stratégie d'équilibre dans laquelle les stratégies du leader sont déterministes, pour définir un algorithme exact, par programmation dynamique arrière, de calcul de stratégies jointes d'équilibre d'un LF-MDP.

3.2 Résolution

Soit $\mathcal{M}$, un LF-MDP. Une stratégie d'équilibre Δ^* peut être calculée par l'algorithme suivant [21] :

Pas de temps final. Au pas de temps final, H, tout suiveur i appliquant l'action $a_i^F \in A^F$ alors que l'état joint est σ et que l'action du leader est a^L, reçoit une récompense immédiate $r_i^F(\sigma, a^L, a_i^F)$, indépendante des actions des autres suiveurs. On peut toujours trouver une stratégie $\delta_{H,i}^{F*}$ déterministe [3] :

$$\delta_{H,i}^{F*}(\sigma, a^L) \in \arg\max_{a_i^F \in A^F} r_i^F(\sigma, a^L, a_i^F) \text{ et}$$

$$Q_{\Delta^*, a^L}^{F,i}(\sigma, H) = \max_{a_i^F \in A^F} r_i^F(\sigma, a^L, a_i^F), \forall (\sigma, a^L). \quad (5)$$

Nous définissons la récompense espérée du leader à tout pas de temps $t \in \{1, \ldots, H\}$, pour une stratégie jointe

1. Nous considérerons, pour raisons de simplicité des notations, des fonctions de transition et de récompense stationnaires, mais cette limitation est triviale à lever.

2. La caractéristique principale d'un LF-MDP est que le leader communique son action aux suiveurs, afin d'influencer leurs choix d'action.

3. Au cas où plusieurs actions généreraient la même récompense, maximale, il suffit de choisir l'une d'elles, arbitrairement.

stochastique des suiveurs, $\delta_t^F = \{\delta_{t,i}^F\}_{i=1..n}$:

$$r^L_{\delta_t^F}(\sigma, a^L) = \sum_{a^F} \Big(\prod_{i=1}^{n} \delta^F_{t,i}(a_i^F|\sigma, a^L) \Big) r^L(\sigma, a^L, a^F). \tag{6}$$

Donc, pour le leader au pas de temps H :

$$\begin{aligned} \delta_H^{L*}(\sigma) &\in \arg\max_{a^L \in A^L} r^L_{\delta_H^{F*}}(\sigma, a^L), \qquad (7) \\ Q^L_{\Delta^*}(\sigma, H) &= \max_{a^L \in \mathcal{A}^L} r^L_{\delta_H^{F*}}(\sigma, a^L), \forall \sigma. \end{aligned}$$

Etapes d'induction. La stratégie jointe (stochastique) d'équilibre des suiveurs à l'étape $t < H$ est calculée inductivement, en fonction des stratégies d'équilibre des pas de temps suivants [4], par la résolution de jeux sous forme normale à n joueurs, pour chaque couple (σ, a^L).
La *valeur* pour le suiveur i de l'action jointe des suiveurs, a^F, dans l'état σ au pas de temps t, après une action du leader, a_L, et en supposant qu'une stratégie jointe d'équilibre est appliquée par la suite, est :

$$\begin{aligned} G^t_{\sigma,a^L,\Delta^*}(i, a^F) &= r_i^F(\sigma, a^L, a_i^F) \qquad (8) \\ &+ \sum_{\sigma'} T\left(\sigma'|\sigma, a^F\right) Q^{F,i}_{\Delta^*}(\sigma', t+1). \end{aligned}$$

Soit $\{\alpha_1^*, \ldots, \alpha_n^*\}$, une solution du jeu $G^t_{\sigma,a^L,\Delta^*}$. $\alpha_i^*(a_F)$ est la probabilité que i "joue" $a_F \in A^F$. Une politique d'équilibre des suiveurs est obtenue par : $\delta^{F*}_{t,i}(a_i^F|\sigma, a^L) = \alpha_i^*(a_i^F)$ et

$$Q^{F,i}_{\Delta^*,a^L}(\sigma, t) = \sum_{a^F} \Big(\prod_{j=1}^{n} \alpha_j^*(a_j^F) \Big) G^t_{\sigma,a^L,\Delta^*}(i, a^F). \tag{9}$$

Puisqu'une action a_L du leader détermine un équilibre de Nash pour les suiveurs, à travers l'équation (9), la stratégie optimale du leader peut être calculée en résolvant un simple MDP non-stationnaire $< \Sigma, A^L, \{T_{\delta_t^{F*}}\}, \{r^L_{\delta_t^{F*}}\}_{t=1..H}, H >$, dans lequel les $\{r^L_{\delta_t^{F*}}\}_{t=1..H}$ ont été définis plus haut et

$$T_{\delta_t^{F*}}(\sigma'|\sigma, a^L) = \sum_{a^F} \prod_{j=1}^{n} \delta^{F*}_{t,j}(a_j^F|\sigma, a^L) T(\sigma'|\sigma, a^F). \tag{10}$$

Les fonctions $T_{\delta_t^{F*}}$ et $r^L_{\delta_t^{F*}}$ sont calculées au fur et à mesure qu'elles sont utilisées dans l'algorithme d'induction arrière. Cet algorithme calcule des stratégies δ_t^{L*} et des fonctions de valeur $Q^{L*}_{\Delta^*}$ optimales, par induction arrière :

$$\begin{aligned} \delta_t^{L*}(\sigma) &\in \arg\max_{a^L \in \mathcal{A}^L} \Big\{ r^L_{\delta_t^{F*}}(\sigma, a^L) \\ &+ \sum_{\sigma' \in \Sigma} T_{\delta_t^{F*}}(\sigma'|\sigma, a^L) Q^{L*}_{\Delta^*}(\sigma', t+1) \Big\}, \\ Q^L_{\Delta^*}(\sigma, t) &= \max_{a^L \in \mathcal{A}^L} \Big\{ r^L_{\delta_t^{F*}}(\sigma, a^L) \qquad (11) \\ &+ \sum_{\sigma' \in \Sigma} T_{\delta_t^{F*}}(\sigma'|\sigma, a^L) Q^{L*}_{\Delta^*}(\sigma', t+1) \Big\}. \end{aligned}$$

4. Cette stratégie jointe, découlant de la recherche d'équilibres de Nash dans des jeux sous forme normale, n'est pas unique.

La complexité de cet algorithme est calculée dans [22]. Dans cet article, on montre également comment la complexité algorithmique peut être réduite grâce à certaines hypothèses sur la structure du problème. Dans la Section suivante nous décrivons brièvement ces résultats. Pour une description plus complète, le lecteur se référera à [22].

3.3 Considérations de complexité algorithmique

Les différentes étapes de l'algorithme LF-MDP générique ont des complexités variées.

Etape 1 : Génération de jeux en forme normale. Afin de calculer les stratégies d'équilibre des suiveurs, il est nécessaire de construire des jeux en forme normale, $G^t_{\sigma,a^L,\Delta^*}$ (Equation 8). Chaque jeu comprend $O\left(n \times |A^F|^n\right)$ éléments, et il y a $|\Sigma| \times |A^L|$ tels jeux. La complexité temporelle de la construction de ces jeux est donc en $O\left(n \times |A^F|^n \times |\Sigma| \times |A^L|\right)$. Néanmoins, il n'est nécessaire de stocker qu'un jeu à la fois dans l'algorithme.

Etape 2 : Calcul et stockage des stratégies jointes d'équilibre des suiveurs. Les stratégies jointes d'équilibre des suiveurs sont obtenues à partir de la résolution des jeux précédemment définis. Leur stockage nécessite un espace en $O(n \times |\Sigma| \times |A^F| \times |A^L|)$. De plus, leur calcul nécessite de résoudre de nombreux jeux, chaque résolution étant elle même difficile [5].

Etape 3 : Calcul des fonctions de transition et de récompense du leader. Les fonctions de transition $T_{\delta_t^{F*}}$ sont calculées grâce à l'équation 10. Ce calcul nécessite un espace en $O\left(|\Sigma|^2 \times |A^L|\right)$ et un temps en $O\left(n \times |A^F|^n \times |\Sigma|^2 \times |A^L|\right)$. Le calcul des fonctions de récompense $r_{\delta_t^{F*}}$ nécessite un espace en $O\left(|\Sigma| \times |A^L|\right)$ et un temps en $O\left(n \times |A^F|^n \times |\Sigma| \times |A^L|\right)$, en utilisant l'équation 6.

Etape 4 : Calcul de la stratégie optimale du leader. Le calcul de δ_t^{L*} nécessite un espace en $O(|\Sigma|)$ et un temps en $O\left(|\Sigma|^2 \times |A^L|\right)$, en utilisant l'équation 11.

Réduction de la complexité. Dans [22], nous avons montré que la propriété de *substituabilité* permettait de simplifier la résolution d'un LF-MDP. Les suiveurs sont substituables dans un LF-MDP, si, pour chacun d'entre eux, seul son état propre et les *nombres* de suiveurs dans chaque état déterminent sa fonction de transition et de récompense. Sous cette hypothèse, un algorithme de résolution de LF-MDP de complexité réduite peut être défini, dont l'espace d'états Σ^L est formé de l'ensemble des vecteurs d'entiers positifs ou nuls, $c = (c_1, \ldots, c_k)$, dont la somme est n. Nous obtenons alors $|\Sigma^L| = O\left(n^k\right)$ et, plus généralement, la taille des divers objets du LF-MDP réduit devient polynomiale en n. En conséquence, la résolution devient polynomiale [6].

5. Chaque résolution est un problème PPAD-complet, où PPAD est une classe de complexité supposée inclure *strictement* la classe P.

6. Sauf la construction de $\bar{T}$, qui reste exponentielle, mais peut être "approchée" en temps polynomial.

TABLE 1 – Complexité de différents éléments utiles dans la résolution d'un LF-MDP dans le cadre d'une résolution naïve, ou exploitant la substituabilité seule, ou incluant également une agrégation des nombres de suiveurs dans chaque état.

	Résolution naïve	Substitutabilité	+ Agrégation
Taille de l'espace d'états	$\lvert\Sigma\rvert = O\left(k^n\right)$	$\lvert\Sigma^L\rvert = O\left(n^k\right)$	$\lvert\overline{\Sigma^L}\rvert = O\left(K^k\right)$
Taille de l'espace d'actions des suiveurs	$\lvert A^F\rvert^n$	$\lvert A^F\rvert^k$	$\lvert A^F\rvert^k$
Taille des matrices de transition	$\lvert T\rvert = O\left(k^{2n}\lvert A^F\rvert^n\right)$	$\lvert\bar{T}\rvert = O\left(n^{2k}\lvert A^F\rvert^k\right)$	$\lvert\hat{T}\rvert = O\left(K^{2k}\lvert A^F\rvert^k\right)$
Taille d'un jeu	$\lvert G^t_{\sigma,a^L,\Delta^*}\rvert = O\left(n\lvert A^F\rvert^n\right)$	$\lvert G^t_{c,a^L,\Delta^*}\rvert = O\left(k\lvert A^F\rvert^k\right)$	$\lvert G^t_{\kappa,a^L,\Delta^*}\rvert = O\left(k\lvert A^F\rvert^k\right)$
Nb jeux par étape	$nb = \lvert A^F\rvert^n\lvert A^L\rvert$	$nb = \lvert A^F\rvert^k\lvert A^L\rvert$	$nb = K^k\lvert A^L\rvert$

Néanmoins, une complexité en $O(n^k)$ peut être problématique, si $n \geq 100$... Aussi, nous avons proposé une méthode de résolution approchée des LF-MDP substituables, consistant à *agréger* les vecteurs c. Pour se faire, nous avons proposé de considérer une partition de l'ensemble $\{0,\ldots,n\}$ en $K+2$ intervalles non-homogènes, où K est un entier divisant n : $I_0 = \{0\}$, $I_{K+1} = \{n\}$ et $I_i = \{\frac{(i-1)n}{K}+1,\ldots,\frac{i.n}{K}\}, \forall i = 1,\ldots,K$. Nous avons définis les *états agrégés* comme des k-uplets d'entiers $(\kappa_1,\ldots,\kappa_k) \in \{0,\ldots,K\}^k$. L'ensemble des états agrégés correspond à l'ensemble des combinaisons d'ensembles $I_{\kappa_1} \times \ldots \times I_{\kappa_k}$, telles qu'il existe un état $c \in \Sigma^L, c_h \in I_{\kappa_h}, \forall h = 1,\ldots,k$. L'ensemble $\overline{\Sigma^L}$ de ces états agrégés vérifie $|\overline{\Sigma^L}| = O(K^k)$. Sa taille est donc indépendante de n. Plus généralement, tous les objets d'un LF-MDP agrégé ont une taille indépendante de n, ce qui permet une résolution en temps quasiment indépendant de n. Seule, encore une fois, la construction de la matrice de transition approchée reste exponentielle en n, mais peut être approchée par simulation. La table 1 regroupe des éléments de complexité des différents sous-modèles.

4 Cas d'étude

4.1 Modèle

Notre cas d'étude concerne la coordination d'un ensemble d'éleveurs afin de limiter la propagation du virus du Syndrome Dysgénésique Respiratoire Porcin (SDRP) [17]. SDRP est une maladie endémique non réglementée. Elle impacte la santé et le bien-être des animaux ainsi que l'économie de l'exploitation. Dans certaines zones (par exemple : [16, 5]), des actions sont proposées par des organisations, dont des incitations financières permettant de limiter le coût des actions de maîtrise pour les éleveurs. Pour les autres zones souhaitant mettre en place une coordination, un outil d'aide à la décision serait d'intérêt.

Pour définir notre modèle LF-MDP, nous présentons d'abord les statuts et les transitions entre statuts au niveau individuel en fonction des décisions individuelles. Pour les transitions, par souci de lisibilité, on donne un diagramme par action, mais des éleveurs peuvent retenir des actions différentes au même instant s'ils ont des statuts différents ou même si leurs statuts sont identiques, si la stratégie obtenue lors du calcul de l'équilibre de Nash est mixte.

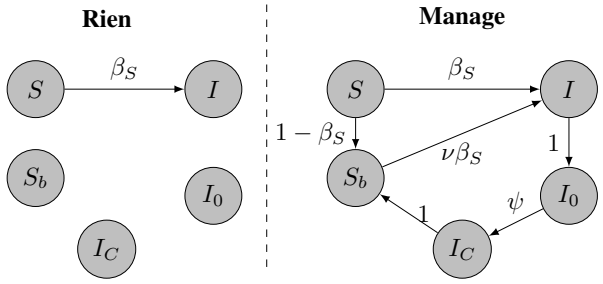

FIGURE 1 – Transitions individuelles pour les suiveurs, en fonction des actions retenues. Les auto-transitions ne sont pas représentées pour simplifier le graphique.

Pour le SDRP, nous avons défini, pour les suiveurs, 5 statuts vis à vis de la maladie et 2 actions possibles "Rien" et "Manage" avec les transitions suivantes (Fig.1) :

- S (non infecté) : un suiveur S devient I (infecté) avec une probabilité β_S. Sinon, soit il reste S (si $a^F_S = 0 \iff$ Rien") soit il devient S_b (si $a^F_S = 1 \iff$ Manage, correspondant à prendre des mesures de biosécurité).
- S_b (non infecté avec management) : seul $a^F_{S_b} = 1 \iff$ Manage (correspondant à de la biosécurité) est possible (pas d'arrêt de la maîtrise). En S_b le risque d'infection et donc de passage en I est réduit comparé à S. La probabilité est $\nu\beta_S$, avec $\nu < 1$ qui dépend de la qualité de la biosécurité.
- I (infecté) : soit le suiveur reste I (si $a^F_I = 0 \iff$ Rien) soit il devient I_0 (si $a^F_I = 1 \iff$ Manage, correspondant à de la vaccination).
- I_0 (infecté en cours de contrôle) : Seul $a^F_{I_0} = 1 \iff$ Manage (correspondant à la vaccination) est possible (pas d'arrêt de la maîtrise). La transition vers I_C est définie avec une probabilité ψ, modélisant un temps de séjour en I_0 stochastique, représentant un délai avant un contrôle effectif.
- I_C (infecté contrôlé) : soit le suiveur reste I_C si $a^F_{I_C} = 0 \iff$ Rien (maintien de la vaccination) soit il devient S_b si $a^F_{I_C} = 1 \iff$ Manage pour revenir non infecté (dépeuplement).

Pour la transmission, on suppose que les suiveurs sont tous en contact les uns avec les autres de par leur situation géographique et les achats/ventes d'animaux. En conséquence, seul la proportion totale d'infectés est à prendre en compte pour le taux de transmission. Cela nous permet de faire une hypothèse de *substituabilité* des suiveurs. Le paramètre β_S a une forme fréquence-dépendant [1] :

$$\beta_S(c) = \frac{1}{n}\sum_{h=1}^{k}\beta(h)c_h + \beta_{out} \tag{12}$$

où $\beta(h)$ est le taux de transmission de la maladie par les suiveurs du statut h et β_{out} le taux de transmission par l'extérieur. Ces paramètres ont été fixés par des experts à $\beta(I) = 0.08, \beta(I_0) = 0.06, \beta(I_C) = 0.01$ et $\beta_{out} = 0.005$.

Les états du leader correspondent à la répartition des suiveurs dans les différents statuts : $\Sigma^L = \{c = (n_S; n_{Sb}; n_I; n_{I0}; n_{Ic})$ avec $n_{Sb} + n_I + n_{I0} + n_{Ic} = n\}$ où n_h le nombre de suiveurs dans le statut h. La taille de l'espace d'états est $|\Sigma^L| = C^4_{n+4}$. Deux actions sont possibles : $a^L = 0 \Leftrightarrow Rien$ et $a^L = 1 \Leftrightarrow Incitation$. Dans le cas $a^L = 1$ une proportion ($perc \leq 1$) des coûts de l'action Manage des suiveurs est prise en charge par le leader. Les transitions entre états sont calculées à partir des transitions entre statuts , de manière similaire à [26].

Les fonctions de récompense des suiveurs (r_F) et du leader (r_L) sont les suivantes, pour $\sigma = (s_1, \ldots, s_n)$:

$$r^L(\sigma, a^L, a^F) = -c^L(a^L) - \sum_{i=1}^{n} c^F(s_i)q^L(a^L, a_i^F) - L^L(s_i),$$

$$r^F(\sigma, a^L, a_i^F) = -E_{s'_i}\left[L^F(s'_i)\right] - \sum_{i=1}^{n} c^F(s_i)q^F(a^L, a_i^F),$$

avec

- c^L, le coût pour le leader de la mise en place des incitations,
- $L^F(s_i)$ les pertes dans le statut s_i pour le suiveur,
- $L^L(s_i) = red \times L^F(s_i)$ les pertes dans le statut s_i pour le leader correspondant à une proportion ($red \leq 1$) de celles des suiveurs,
- $c^F(s_i)$ le coût de l'action Manage pour les suiveurs,
- $q^L(a^L, a_i^F)$ la proportion prise en charge par le leader pour les suiveurs retenant l'action Manage : $q^L(a^L, a_i^F) = perc \leq 1$ si $a^L = 1$ et $a_i^F = 1$ et 0 sinon,
- $q^F(a^L, a_i^F)$ la proportion restant à la charge des suiveurs retenant l'action Manage ($a_i^F = 1$) : $q^F(a^L, a_i^F) = 1 - perc$ si $a^L = 1$ et 1 si $a^L = 0$.

Pour la résolution, seuls 3 statuts nécessitent des décisions individuelles (S, I, I_C). Par conséquent, la taille de l'espace d'actions des suiveurs est 2^3 et la taille d'un jeu 3×2^3.

4.2 Evaluation de l'approximation

Pour valider l'approximation présentée dans la section 3.3, nous avons réalisé, pour $n \in \{12, 15, 20\}$, une résolution exacte et des résolutions approchées avec différentes valeurs de K compatibles avec n. Pour chaque combinaison de K et n, nous avons calculé les politiques exacte (δ^*) et approchée (δ^K). Puis, nous avons comparé le comportement du modèle avec ces différentes politiques.

Pour le paramétrage du modèle, nous avons considéré plusieurs jeux de paramètres générés en faisant varier les valeurs des paramètres dans des intervalles de possibles donnés par les experts pour évaluer les comportements de notre modèle. Nous avons considéré deux distributions initiales (à $t = 0$) (i) Γ_U uniforme sur tous les états et (ii) Γ_E uniforme seulement sur les états représentant une situation endémique (état avec au moins 40% dans les statuts $S + S_b$ et au moins 40% dans le statut I_C).

Nous avons comparé les politiques et le comportement du modèle selon 6 indicateurs

- $\#Diff$ Le nombre de pas de temps où la politique approchée diffère de la politique exacte,
- Max_Gap la valeur maximale de la proportion d'états par pas de temps où il y a une différence d'actions entre les politiques,
- DB_x la distance de Battacharyya [3] entre les distributions sur les états du leader au pas de temps final, en partant d'une distribution initiale Γ_x avec $x = U$ ou E,
- $RMSE_x$ l'écart entre les valeurs espérées calculées sur tous les états si $x = U$ et uniquement les états endémique si $x = E$ en utilisant la formule :

$$RMSE_x = \sqrt{\sum_{c \in \Sigma^L}((V^K(c) - V^*(c))^2 \times \Gamma_x)},$$

avec $x = U$ ou E et $V^K(.)$ la fonction de valeur de la politique δ^K.

4.3 Résultats

Plusieurs jeux de paramètres testés conduisent à très peu d'incitations du leader. Dans certains jeux de paramètres, la politique consiste à ne rien faire sauf à un seul pas de temps où l'incitation est retenue pour quelques états. Pour ces jeux de paramètres, les différences liées à l'approximation sont faibles. Pour les illustrations de ce papier, nous avons sélectionné 2 jeux de paramètres avec des incitations réparties sur plusieurs pas de temps (Table 2). Pour ces deux jeux, les profils d'incitations sont conservés : les instants où il y a au moins un état avec une incitation sont les mêmes quelle que soit l'approximation. Néanmoins, les états avec incitations changent selon K.

Il y a quelques légères différences en termes de politique et de RMSE, mais elles restent faibles (Fig. 2). Il n'y a pas de relation claire entre la valeur de K et les écarts à la résolution exacte.

Avec notre approximation, le problème peut être résolu pour $n = 100$ avec $K = 5$. Mais du fait de la taille du modèle ($|\Sigma^L| = 4\,598\,251$), nous n'avons pas pu calculer la fonction de valeur sur l'espace complet. Si on considère uniquement les états endémiques pour l'application au

TABLE 2 – Valeurs des paramètres de transition des suiveurs et des récompenses (leader et suiveurs) dans 2 jeux

Jeu	ν	ψ	$L^F(s_i)$ (pertes)	$c^F(s_i)$ (coûts des suiveurs)	$c^L(a_L)$ (coût leader)	$perc$	red
A	0.5	0.5	(0,0,6,5,4)	(4,1,4,2,101)	(0,3)	0.5	0.7
B	0.7	0.5	(0,0,4.8,5.6,2.8)	(7.84,1.4,11.76,2.8,101.4)	(0,4.2)	0.7	0.7

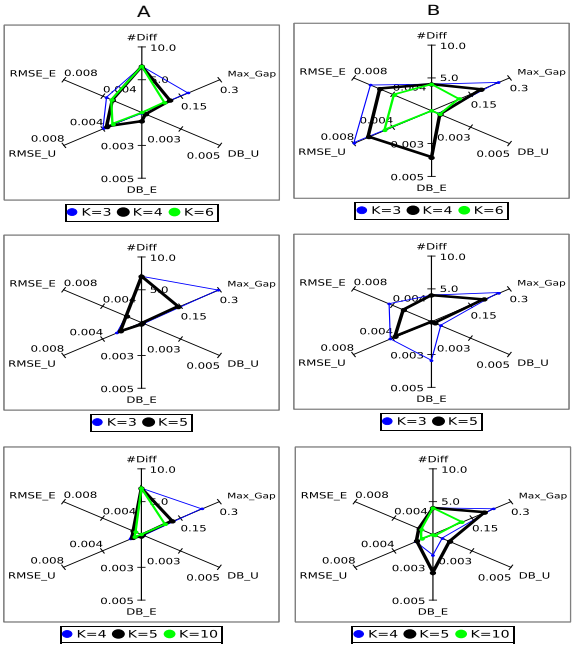

FIGURE 2 – Représentation des différences entre la résolution exacte et approchée, sur différents indicateurs et pour différentes approximations (K) pour $N = 12$ (ligne du haut), $N = 15$ (ligne du milieu) et $N = 20$ (ligne du bas)

SDRP, il est néanmoins possible de calculer par simulation les fonctions de valeurs pour ces états.

5 Conclusion

Nous avons proposé une approche pour la gestion collective de maladies animales transmissibles non réglementées via la coordination de décisions individuelles, dans le cadre LF-MDP. Nous avons développé un algorithme générique permettant la résolution exacte ($n \leq 20$) ou approchée ($n \leq 100$) d'un problème de contrôle collectif du SDRP. Dans notre cas d'étude, nous avons montré par comparaison, la proximité des solutions exacte et approchée pour $n \leq 20$. Les écarts sont d'autant plus acceptables que notre approche est principalement un outil de réflexion pour les décideurs. Pour illustrer l'intérêt de la politique obtenue, qui est adaptative, l'impact de celle ci devra être comparé à celui des stratégies classiquement considérées qui sont non-adaptatives, comme dans [19, 20].

Ces travaux peuvent être poursuivis selon différents axes. Tout d'abord, bien que la résolution approchée soit possible avec $n = 100$, le calcul exact des matrices de transitions est très long pour $n > 100$ (plusieurs heures/jours). Ces matrices peuvent être approchées par des simulations non exhaustives des transitions, ce qui permettrait un important gain de temps. Ensuite, nous avons supposé que l'horizon de la prise de décision individuelle était le même que celui de l'organisation. Or les éleveurs raisonnent souvent à un horizon plus court, parfois d'un seul pas de temps. La prise en compte d'horizons différents pour le leader et les suiveurs est possible dans le cadre LF-MDP et il serait intéressant d'étudier l'impact sur la politique optimale du leader de la considération d'un horizon court pour les suiveurs. Enfin, les éleveurs peuvent présenter une *aversion au risque* plus ou moins importante. L'effet de l'attitude vis à vis du risque de gestionnaires forestiers sur leurs politiques de gestion a déjà été étudié dans le cadre MDP [4, 6]. L'extension de ces travaux au cadre LF-MDP est un problème ouvert, important en épidémiologie animale.

Enfin, le cadre LF-MDP est générique et pourrait être adopté dans d'autres domaines que la santé animale. Notre approche peut être utilisée dans de nombreux problèmes où des décideurs maximisant leur profit individuel sont sous l'influence d'un coordinateur (via des actions impactant les jeux entre suiveurs) visant à atteindre son propre objectif. De tels problèmes peuvent se rencontrer en épidémiologie humaine ou végétale (dans ce dernier cas, on se place au niveau de parcelles agricoles gérées par des décideurs indépendant). Toujours dans le cadre de la gestion Environnementale, la taxation des émissions de carbone est un autre domaine d'application potentiel.

Remerciements

Ce travail a été soutenu par l'ANR, projets ANR-10-BINF-07 (MIHMES) et ANR-13-AGRO-0001-04 (AgroBioSE) et par l'Europe (fond FEDER, Pays-de-la-Loire).

Références

[1] M. Begon, M. Bennett, R.G. Bower, N.P. French , S.M. Hazel , J. Turner, A clarification of transmission terms in host-microparasite models : numbers, densities and areas. Epidemiology and Infection, 129, pp. 147-153, 2002.

[2] R. Bennett, The 'direct costs' of livestock disease : the development of a system of models for the analysis of 30 endemic livestock diseases in Great Britain. Journal of Agricultural Econonic 54 pp. 55-71, 2003

[3] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their pro-

bability distributions. Bulletin of the Calcutta Mathematical Society, 35, pp. 99-109, 1943.

[4] M. Brunette, S. Couture, J. Laye, Optimising forest management under storm risk with a Markov decision process model. Journal of Environmental Economics and Policy, 4, pp. 141-163, 2015.

[5] C.A. Corzo, E. Mondaca, S. Wayne, M. Torremorell, S. Dee, P. Davies, R.B. Morrison, Control and elimination of porcine reproductive and respiratory syndrome virus. Virus research, 154, pp. 185-192, 2010.

[6] S. Couture, M.J. Cros, R. Sabbadin, Risk aversion and optimal management of an uneven-aged forest under risk of windthrow : A Markov decision process approach. Journal of Forest Economics, 25, pp. 94-114, 2016.

[7] K. Ekboir, The role of the public sector in the development and implementation of animal health policies. Preventive veterinary Medicine 40 pp. 101-115, 1999.

[8] J. Filar, K. Vrieze, Competitive Markov Decision Processes, Springer, 1996.

[9] L. Ge, M. Mourits, A.R. Kristensen, R. Huirne, A modelling approach to support dynamic decision-making in the control of FMD epidemics, Preventive veterinary medicine, 95, pp. 167-174, 2010.

[10] J. Hu, M. P. Wellman, Nash Q-learning for general-sum stochastic games, Journal of Machine-Learning Research, 4, pp. 1039-1069, 2003.

[11] Y. Ibuka, M. Li, J. Vietri, G.B. Chapman, A.P. Galvani, Free-riding behavior in vaccination decisions : An experimental study. PloS one 9 pp. e87164, 2014.

[12] E. Klein, R. Laxminaryan, D.L. Smith, C.A. Gilligan, Economic incentives and mathematical models of disease. Environment and Development Economics 12 pp. 707-732,2007.

[13] M. Kobayashi, T.E. Carpenter, B.F. Dickey, R.E. Howitt, A dynamic, optimal disease control model for foot-and-mouth disease : I. Model description. Preventive Veterinary Medicine 79 pp. 257-273, 2007.

[14] L. Laguzet, G. Turinici, Individual Vaccination as Nash Equilibrium in a SIR Model with Application to the 2009-2010 Influenza A (H1N1), Bulletin of Mathematical Biology, 77, pp. 1955-1984, 2015.

[15] A. Le Menach, E. Vergu, R.F. Grais, D.L. Smith, A. Flahault, Key strategies for reducing spread of avian influenza among commercial poultry holdings : lessons for transmission to humans. Proceedings of the Royal Society B, Biological Science 273, pp. 2467-2475, 2006.

[16] M-F. Le Potier, P. Blanquefort, E. Morvan, E. Albina, Results of a control programme for the porcine reproductive and respiratory syndrome in the French 'Pays de la Loire' region. Veterinary microbiology, 55 pp. 355-360, 1997.

[17] G. Nodelijk, Porcine Reproductive and Respiratory Syndrome (PRRS) with special reference to clinical aspects and diagnosis : A review. Veterinary Quarterly, 24, pp. 95-100, 2002.

[18] E.L. Plambeck, S.A. Zenios, Performance-based incentives in a dynamic principal-agent model. Manufacturing & service operations management 2, pp. 240-263, 2000.

[19] O. Rat-Aspert, C. Fourichon, Modelling collective effectiveness of voluntary vaccination with and without incentives, Preventive Veterinary Medicine, 93, pp. 265-275, 2010.

[20] O. Rat-Aspert, S.Krebs, Individual and collective management of endemic animal diseases : an economic approach, In : 2012 Conference of the International Association of Agricultural Economists, August 18-24, 2012, Foz do Iguacu, Brazil

[21] R. Sabbadin, A.F. Viet, A Tractable Leader-Follower MDP Model for Animal Disease Management, In : *27th AAAI Conference on Artificial Intelligence*, pp. 1320-1326, 2013.

[22] R. Sabbadin, A.F. Viet, Leader-Follower MDP model with factored state space and many followers - followers abstraction, structured dynamics and state aggregation, In : *22nd European Conference on Artificial Intelligence (ECAI 2016)*, pp. 116-124, 2016.

[23] K. Tharakunnel, S. Bhattacharyya, Single-leader-multiple-follower games with boundedly rational agents. Journal of Economic Dynamics and Control, 33, pp. 1593-1603,2009.

[24] M.J. Tildesley, N.J. Savill, D.J. Shaw, R. Deardon, S.P. Brooks, M.E.J. Woolhouse, B.T. Grenfell, M.J. Keeling, Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 440, pp. 83-86, 2006.

[25] N. Toft, A.R. Kristensen, E. Jørgensen, A framework for decision support related to infectious diseases in slaughter pig fattening units. Agricultural Systems 85, pp. 120-137, 2005.

[26] A.-F. Viet, L. Jeanpierre, M. Bouzid, A.-I. Mouaddib, Using Markov Decision Process to define an adaptive strategy to control the spread of an animal disease. Computer and Electronics in Agriculture, 80, pp. 71-79, 2012.

[27] A. Vonk Noordegraaf, J.A.A.M. Buijtels, A.A. Dijkhuizen, P. Franken, J.A. Stegeman, J. Verhoeff, An epidemiological and economic simulation model to evaluate the spread and control of infectious bovine rhinotracheitis in the Netherlands. Preventive Veterinary Medicine 36 pp. 219-238, 1998.

[28] C. Wernz, A. Deshmukh, Unifying temporal and organizational scales in multiscale decision-making. European Journal of Operational Research 223, pp 739-751, 2012.