MDP à grande échelle : étude de cas des voies navigables

Guillaume Desquesnes Guillaume Lozenguez Arnaud Doniec Éric Duviella

Mines Douai, IA, F-59508 Douai, FRANCE,
Univ. Lille, F-59000 Lille, FRANCE

prenom.nom@mines-douai.fr

Résumé

Les réseaux de voies navigables devraient subir des changements importants en raison d'une volonté d'augmenter le trafic naval et des effets du changement climatique. Ces changements nécessitent une gestion adaptative et résiliente de la ressource en eau et requièrent donc une planification plus intelligente. Un modèle représentatif du réseau, utilisant des MDPs, est proposé et testé afin d'optimiser la gestion de l'eau. Il fournit des résultats prometteurs, le modèle proposé permet de coordonner plusieurs entités sur plusieurs pas de temps de façon à éviter les inondations et sécheresses dans le réseau. Cependant, la solution proposée ne permet pas de passer à l'échelle et n'est pas utilisable dans une application réelle. Les avantages et limitations de plusieurs approches de la littérature qui pourrait permettre de passer à l'échelle sont présentés et discutés sous le prisme de notre étude de cas.

Mots Clef

Processus de décision markovien, Réseau de voies navigables, Grand modèle

Abstract

Inland waterway networks are likely to go through heavy changes due to a will in increasing the boat traffic and to the effects of climate change. Those changes would lead to a greater need of an automatic and intelligent planning for an adaptive and resilient water management. A representative model of the network is proposed and tested using MDPs with promising results on the water management optimization. The proposed model permits to coordinate multiple entities over multiple time steps in order to avoid a flood and drought, in the waterway network. However, the proposed model suffers a lack of scalability and is unable to represent a real case application. The advantages and limitations of several approaches of the literature are discussed according to our case study.

Keywords

Markov Decision Process, Inland waterway network, Large model

1 Introduction

Le changement climatique est une problématique majeure de notre société moderne. Ces dernières années, les effets du changement climatique sur le réseau des voies navigables ont été étudiés. Le consensus général est que l'intensité et la fréquence des périodes d'inondation et de sécheresse vont augmenter [14]. En parallèle, l'utilisation des voies navigables pour décongestionner le trafic routier et ferroviaire est en vogue. Une augmentation du trafic naval est donc attendue dans les années à venir.

Un réseau de voies navigables est un réseau hydrographique aménagé par les hommes qui interagit avec un environnement naturel. Une majorité de ces interactions n'est que partiellement connue : rejets illégaux, échanges avec les nappes phréatiques, influence locale de la météo, ... Le contrôle d'un tel réseau est donc soumis aux incertitudes et une modélisation stochastique semble donc la plus adaptée.

Actuellement, la supervision et la conduite du système repose principalement sur l'expertise d'opérateurs humains. Toutes ces évolutions ont tendance à complexifier la gestion de l'eau dans les réseaux de voies navigables et rendent a priori la planification par l'opérateur humain de moins en moins pertinente pour optimiser la gestion de l'eau du réseau.

Les processus de décision markoviens (ou Markov Decision Processes - MDP) sont largement utilisés pour la planification de modèles stochastiques et permettent d'obtenir un plan pour toutes les configurations possibles du modèle. À notre connaissance, les MDPs n'ont pas encore été utilisés pour modéliser les réseaux de voies navigables de façon à les rendre résilients et plus stables. Néanmoins, des travaux ont déjà été réalisés sur ce sujet, en utilisant des réseaux de flots avec l'hypothèse forte d'un modèle déterministe [13].

Les MDPs permettent de modéliser l'évolution d'un système incertain, mais induisent un modèle intraitable dans une majorité d'applications. Ces MDPs trop grands rendent les politiques de contrôle optimal dures à calculer et nécessitent des algorithmes spécifiques [3, 12]. Une modélisation stochastique du réseau est proposée, permettant de planifier une coordination de toutes les entités du réseau. La complexité dans cette application vient d'abord

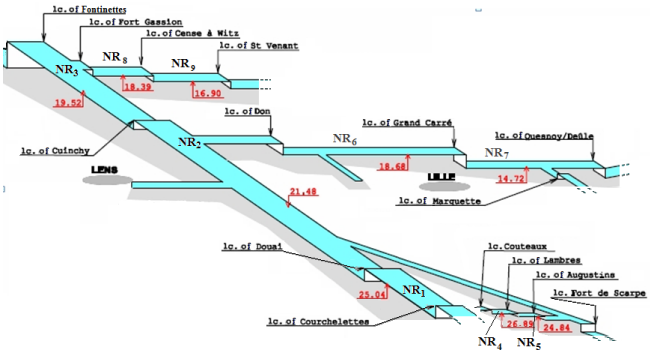

FIGURE 1 – Extrait du réseau de voies navigables du nord de la France

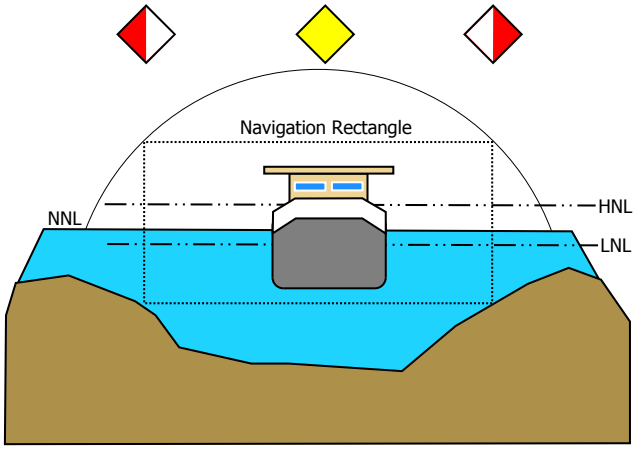

FIGURE 2 – Rectangle de navigation

des possibilités de déplacement d'eau réparties sur un territoire. Un contrôle optimal représente l'ensemble des actions conjointes (déplacements d'eau) en considérant l'ensemble des configurations possibles de l'intégralité du réseau ce qui induit une explosion combinatoire du nombre d'états. Cet article vise à montrer les limitations d'un tel modèle optimal et centralisé appliqué à un réseau de voies navigables et à discuter des approches qui permettraient de distribuer le modèle.

Dans cet article, le problème de la gestion d'un réseau de voies navigables dans le cadre du changement climatique est présenté dans la section 2. Une modélisation naïve du réseau en utilisant les MDPs est présentée en section 3 et de premiers résultats nous permettent de discuter de ses limitations en section 4. Une description et une comparaison de différents dérivés des MDPs pour des modèles à grande échelle et distribués sont proposées en section 5, toujours dans le cadre de la supervision d'un réseau de voies navigables. Enfin une ébauche de modélisation du problème via un MDP distribué est discutée dans la section 6.

2 Supervision d'un réseau de voies navigables

Un réseau de voies navigables (voir figure 1) est un système à grande échelle utilisé majoritairement pour la navigation. Il fournit à la fois des avantages économiques et environnementaux [10, 11], tout en assurant un transport discret, efficace et sûr des biens [5]. Il est constitué majoritairement des rivières canalisées et des canaux artificiels, le tout séparé par des écluses. Tout morceau de réseau entre deux écluses est appelé un bief.

Le niveau d'eau d'un bief doit respecter les conditions du rectangle de navigation (voir figure 2) tout en étant le plus proche possible du niveau normal de navigation ou normal navigation level (NNL). Les bornes inférieures et supérieures du rectangle de navigation sont appelées respectivement niveau de navigation inférieur ou lower navigation level (LNL) et niveau de navigation supérieur ou higher navigation level (HNL). L'objectif principal des opérateurs est de maintenir un niveau d'eau acceptable dans tous les biefs du réseau afin de permettre la navigation.

Dans une situation normale, le passage de bateaux par les écluses est la principale perturbation du niveau d'eau, puisque l'utilisation d'une écluse draine une quantité importante d'eau du bief amont pour la déverser dans le bief aval. D'autres perturbations du niveau d'eau peuvent exister telles que des échanges avec le sol, les rivières naturelles, les précipitations ainsi que d'autres échanges inconnus et non contrôlés tels que des rejets illégaux. Les écluses ne sont pas dédiées au contrôle du niveau de l'eau, des portes sont utilisées pour envoyer de l'eau en aval et des pompes peuvent renvoyer de l'eau en amont.

Pour le moment, la navigation n'est autorisée que pendant la journée, avec quelques exceptions, notamment le dimanche. La gestion des biefs est basée sur une expertise humaine obtenue au fil du temps. Mais, dans un contexte de changement climatique, qui augmentera les effets des inondations et des sécheresses, combinés à une volonté d'augmenter le trafic notamment en permettant la navigation sur 24h, l'expertise humaine devrait montrer ses limites.

L'objectif principal est d'assurer en chaque point du réseau les conditions de navigation. Cela consiste à déterminer un planning global, pour l'intégralité du réseau, en prenant en compte les incertitudes du problème, telles que le climat et le trafic, le tout dans un contexte de changement climatique et d'augmentation du trafic. Planifier sur plusieurs pas de temps permet une meilleure anticipation des événements possibles, en utilisant la capacité d'obtenir des informations sur l'état du réseau en temps réel grâce à des capteurs de niveau répartis dans les biefs.

3 Utilisation des processus décisionnels markoviens

Un processus décisionnel markovien (MDP) permet de modéliser de façon générique les possibilités de contrôle d'un système dynamique et stochastique sous forme d'un automate probabiliste. Cette modélisation est bien adaptée au réseau de voies navigables puisque l'état du réseau est

totalement observable (en terme de volumes d'eau) et le contrôle est incertain du fait des entrées / sorties d'eau incontrôlées.

3.1 MDP

Un MDP est défini par un tuple $\langle S, A, T, R \rangle$, où S et A représentent respectivement les ensembles d'états et d'actions qui définissent le système et ses possibilités de contrôle. T est la fonction de transition définie par $T : S \times A \times S \rightarrow [0, 1]$. $T(s, a, s')$ est la probabilité d'atteindre l'état s' en effectuant l'action a depuis l'état s, avec $s, s' \in S$ et $a \in A$. La fonction de récompense R définie par $R : S \times A \times S \rightarrow \mathbb{R}$, $R(s, a, s')$ donne la récompense obtenue lorsque l'agent arrive en s' après avoir effectué l'action a en s.
Une politique $\pi : S \rightarrow A$ est un assignement d'une action à chaque état du système. La résolution optimale d'un MDP consiste à trouver la politique optimale π^* qui maximise la récompense espérée. π^* maximise la fonction de valeur de l'équation de Bellman [1] définie pour chaque état par :

$$V^{\pi}(s) = \sum_{s' \in S} T(s, a, s') \times (R(s, a, s') + \gamma V^{\pi}(s')) \quad (1)$$

avec $a = \pi(s)$

$$\pi^*(s) = \arg\max_{a \in A} (\sum_{s' \in S} T(s, a, s') \times (R(s, a, s') + \gamma V^{\pi^*}(s'))) \quad (2)$$

Le paramètre $\gamma \in [0, 1]$ permet de varier l'importance entre les récompenses futures ou immédiates. Si γ a une valeur proche de 0 les récompenses immédiates seront préférées, tandis que pour γ proche de 1, des pénalités à court terme pourront être acceptées si elles mènent a des récompenses importantes à long terme. Divers algorithmes existent pour résoudre de manière optimale un MDP. Une version notable est *Value Iteration* [16]. Il construit itérativement la fonction de valeur V, en utilisant l'équation 3 après un nombre spécifié d'itérations ou jusqu'à convergence. Le dernier V_i obtenu est utilisé pour générer la politique optimale grâce à l'équation 2. La première fonction de valeur V_0 étant initialisé à 0.

$$V_{i+1}(s) = \max_{a \in A} (\sum_{s' \in S} T(s, a, s') \times (R(s, a, s') + \gamma V_i(s'))) \quad (3)$$

3.2 Approche naïve du contrôle du réseau des voies navigables

L'objectif est de planifier la meilleure suite d'actions pour l'ensemble du réseau sur t pas de temps, tout en sachant que certaines conditions pourront être différentes à chaque pas de temps et peuvent affecter la navigation. Par exemple, le temps peut devenir pluvieux, augmentant le niveau de l'eau dans les biefs affectés, ou encore une augmentation du trafic naval sur certains biefs impliquant une plus grande utilisation des écluses.

Des demi-journées sont utilisées comme pas de temps, de façon à séparer les périodes de navigation le jour et les périodes inactives la nuit. L'utilisation de pas de temps larges permet de considérer le niveau d'eau d'un bief comme uniforme et de réduire les incertitudes sur le trafic et autres variations temporelles.
Un état du modèle est défini comme une assignation de volume à chaque bief du réseau pour chaque pas de temps. De même, une action est une assignation de volume d'eau transférée par chaque point de transfert contrôlé (les ouvrages). Le formalisme MDP nécessite des ensembles d'états et d'actions discrétisés, mais, comme les volumes observés et transférés du système sont continus, ils ont dû être discrétisés sous forme d'intervalles.
Chaque bief est divisé en intervalle, tous de même taille, à l'exception du premier et dernier intervalle qui comportent les valeurs en dehors du rectangle de navigation. Ces deux intervalles sont considérés de taille infinie.
Les points de transfert utilisent une partition en intervalles similaires à celle des biefs, cependant les volumes transférés étant considérés comme étant parfaitement contrôlables, ils n'ont donc pas d'intervalle de taille infinie.
Plus formellement, l'ensemble d'états S du modèle est défini comme la combinaison de tous les intervalles possibles de chaque bief pour chaque pas de temps. Pour un bief i, les intervalles sont obtenus par une discrétisation régulière des volumes partant de moins d'eau que le minimum autorisé 0 jusqu'à plus que le maximum autorisé ri_{out}.

$$S = \{0, \ldots, t\} \times \prod_{i=1}^{N} [0, ri_{out}] \quad (4)$$

où N représente le nombre de biefs dans le réseau.

Similairement, l'ensemble d'actions A est défini comme la combinaison des intervalles de volume des points de transfert. Les actions étant indépendantes du temps, nous avons simplement :

$$A = \prod_{i,j \in [0,N]^2} L_{i,j} \quad (5)$$

où $L_{i,j}$ représente l'ensemble des intervalles du point de transfert reliant le bief i au bief j et le bief 0 correspond aux rivières externes, ou autres éléments externes, reliés au bief. Il existe, en fait, un nombre très limité de points de transfert, la plupart des $L_{i,j}$ ne permettent donc pas de transfert et sont notés $L_{i,j} = \emptyset$.
Nous notons $a_{i,j} \in L_{i,j}$ le volume à transférer du bief i à j prévu par l'action $a \in A$. Pour simplifier la notation, a_i représente la partie de l'action qui affecte le bief i tel que :

$$a_i = \sum_{j=0}^{N} (a_{i,j} \ominus a_{j,i}) \quad (6)$$

Avec $\oplus$ et $\ominus$ deux opérateurs, définis sur $(\mathbb{R} \cup \mathcal{I})^2 \rightarrow \mathbb{R}$, qui peuvent respectivement ajouter et soustraire des inter-

valles de nombre et/ou des nombres, le résultat étant toujours un réel. $\mathcal{I}$ étant l'ensemble de tous les intervalles possibles de notre réseau. Nos opérateurs sont respectivement une simple addition et soustraction, en utilisant la valeur du membre s'il s'agit d'un réel, ou la moyenne du membre s'il s'agit d'un intervalle.

La fonction de transition $T(s, a, s')$ représente la probabilité d'atteindre l'état s' après avoir effectué l'action a depuis l'état s en prenant en compte les possibles variations temporelles. Trivialement, pour s et s' respectivement défini aux pas de temps t et t' alors, s' n'est atteignable que si $t' = t + 1$.

L'état d'un bief ne dépend que des volumes d'eau entrants et sortants, et est donc indépendant de celui des autres biefs dans la fonction de transition. Une première source d'incertitude sur les transitions vient des déplacements d'eau non contrôlés, modélisés par une liste de variations temporelles notées Var.

Les variations temporelles sont des changements locaux à un ou plusieurs biefs ou points de transfert pendant un ou plusieurs pas de temps avec une certaine probabilité. De la pluie sur un bief, par exemple, est une variation temporelle. Comme les variations temporelles ne sont pas dans l'espace d'action, elles affectent uniquement la fonction de transition. Le volume non contrôlé qui affecte le bief i est noté $v_i \in Var$ et $P(v_i|t)$ représente la probabilité qu'il arrive au pas de temps t.

La seconde source d'incertitude vient de la discrétisation en intervalle des volumes transférés et des biefs, qui est la cause d'une approximation dans la représentation des états. Nous définissons $P(ri_{s'}|ri_s,\ a_i + v_i)$ la probabilité que le volume d'eau du bief i au pas de temps $t_s + 1$ soit inclus dans l'intervalle $ri_{s'}$ si l'action a_i est effectuée avec un déplacement incontrôlé de volume v_i en i en partant de l'intervalle ri_s.

$$P(ri_{s'}|ri_s,\ a_i{+}v_i) = \begin{cases} p_= & \text{si } ri_s \oplus a_i + v_i \in ri_{s'} \\ p_+ & \text{si } ri_s \oplus a_i + v_i \in ri_{s'} + 1 \\ p_- & \text{si } ri_s \oplus a_i + v_i \in ri_{s'} - 1 \\ 0 & \text{sinon} \end{cases} \tag{7}$$

où $p_=$ est la probabilité d'atteindre l'intervalle attendu en prenant en compte l'approximation des intervalles, p_+ (resp. p_-) est la probabilité d'atteindre l'intervalle correspondant à un niveau d'eau supérieur (resp. inférieur), en respectant

$$p_+ + p_- + p_= = 1 \quad \text{et} \quad p_+ = p_-$$

La fonction de transition est construite à partir du produit des deux sources d'incertitudes :

$$T(s, a, s') = \prod_{i=1}^{N} \left(\sum_{\forall v_i \in Var} P(v_i|t) \times P(ri_{s'}|ri_s,\ a_i + v_i) \right) \tag{8}$$

Finalement, la fonction de récompense est définie, avec les experts, de façon à pénaliser fortement l'écart au niveau normal de navigation pour chaque bief, tout en ajoutant un faible coût aux déplacements d'eau. Plus formellement :

$$R(s, a, s') = -1 \times \left(\sum_{i=1}^{N} (NNL_i \ominus ri_{s'})^2 + a_i \right) \tag{9}$$

où NNL_i est le volume objectif correspondant au niveau normal de navigation du bief i. Lorsque $ri_{s'}$ est en dehors du rectangle de navigation, il est remplacé par une grande valeur c, et par $\frac{c}{2}$ si $ri_{s'}$ n'est que partiellement en dehors du rectangle de navigation.

4 Essai sur un réseau

Afin de tester cette approche, un réseau de navigation réaliste a été imaginé (voir figure 3), composé de deux biefs et de six points de transfert. Notre approche sera testée sur plusieurs scénarios de ce modèle, ce qui impliquera la création de plusieurs MDPs.

4.1 Caractéristiques du réseau

Sur la figure 3, les biefs sont représentés par des carrés, avec le rectangle de navigation spécifié en unité de volume, et les arcs correspondent aux points de transfert avec une capacité de transfert minimale et maximale. Une valeur négative signifie que le point de transfert peut être utilisé pour importer et exporter de l'eau. Ce réseau a été simulé sur une durée de 8 journées et 8 nuits, nous donnant 16 pas de temps. Les simulations se déroulent dans un contexte d'absence de navigation de nuit, où les écluses ne sont pas utilisées de nuit.

Les volumes des deux biefs sont divisés en 9 intervalles de tailles $\{\infty, 20, 20, 20, 20, 20, 20, \infty\}$. Les volumes des points de transfert 1, 2 et 4 sont divisés en intervalles de taille 5. Ce qui correspond respectivement à 6, 5 et 2 intervalles. Les points de transfert 0, 3 et 5 représentent les écluses et transfèrent une quantité d'eau constante, correspondant au trafic fluvial. Comme nous planifions sur 16 pas de temps, nous avons $\prod_{\forall i} |[0, ri_{out}]| \times |0, \ldots, t| = 9 \times 9 \times (16+1) = 1377$ états. Un pas de temps supplémentaire est ajouté pour marquer la fin de planification, et tous les états durant ce pas de temps sont considérés absorbants, ce qui signifie $T(s, a, s) = 1$ et $R(s, a, s) = 0$, et n'ont donc aucune influence sur la planification. Similairement, le nombre d'actions est $\prod_{\forall (i,j)} |L_{i,j}| = 1 \times 6 \times 5 \times 1 \times 2 \times 1 = 60$.

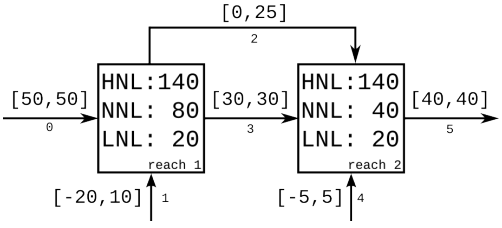

FIGURE 3 – Réseau de voies navigables

Nous avons utilisé $c = 10000$, une valeur arbitrairement grande, comme pénalité pour être en dehors du rectangle de navigation. Dans une optique de validation des performances de l'approche, les probabilités sur les intervalles ont été assignées arbitrairement de la façon suivante $p_= = 0.9$ et $p_+ = p_- = 0.05$.

4.2 Quelques résultats

Pour analyser empiriquement, la qualité des politiques obtenues par notre modélisation, nous avons effectué plusieurs simulations correspondant à 4 scénarios. Le premier scénario correspond à un scénario idéal. Les deux biefs sont initialement à leur niveau normal de navigation et il n'y a aucune perturbation. Dans le second scénario, le premier bief est près de son niveau minimal de navigation et le second près de son niveau maximal toujours sans perturbation. Le troisième scénario est similaire au précédent, avec les niveaux de départ des biefs inversés. Finalement, le dernier scénario se base sur le premier scénario, en y ajoutant une perturbation importante. Cette perturbation très probable ne dure qu'un pas de temps et n'affecte qu'un point de transfert. Cependant, elle peut possiblement faire déborder un bief si elle n'est pas anticipée par la planification.

Puisque les actions sont des intervalles de volumes, les scénarios ont été testés en transférant une valeur aléatoire dans les intervalles, plutôt que de choisir la meilleure valeur ou la moyenne. Cela dans le but d'avoir une meilleure perception de la qualité des intervalles choisis par la politique. Comme les volumes à transférer sont choisis de façon aléatoire, 5 simulations ont été effectuées pour chaque scénario. Cela permet d'avoir une meilleure visualisation des résultats possibles sans trop surcharger les figures. Il est important de noter que les simulations pour tester les politiques se déroulent dans un système continu, de façon à être le plus proche d'un système réel.

Ce réseau a été créé de façon à ce qu'une planification optimale du premier scénario permette de maintenir le niveau normal de navigation dans les deux biefs sur tous les pas de temps. Il est possible d'observer, sur la figure 4 que la politique obtenue par cette approche n'en est que relativement proche. Cet écart est lié à la discrétisation en intervalles et à l'approximation qui en résulte. Comme un intervalle est représenté par sa moyenne, il est possible que le volume d'un bief augmente ou baisse tout en restant dans le même intervalle et cela peut induire un écart à l'optimalité.

La réaction du réseau face à un événement qui avait amené les biefs aux limites du rectangle de navigation est visible sur les figures 5 et 6. Il est possible de remarquer que la récupération est plus rapide dans le second cas, ceci étant lié à la configuration du réseau. Il est en effet plus simple pour un bief d'envoyer de l'eau dans un bief aval qui doit se remplir, que lorsqu'il faut utiliser les sources extérieures afin d'éviter de nuire au reste du réseau.

Dans le dernier scénario, une forte pluie est supposée se produire entre le pas de temps 7 et 8, ce qui ferait déborder le premier bief, si celui-ci se trouvait proche de son NNL.

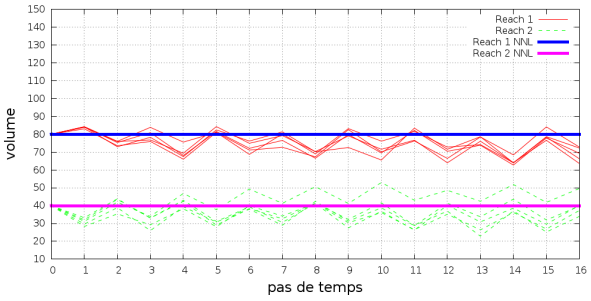

FIGURE 4 – Scénario 1 : partant des NNL sans perturbation

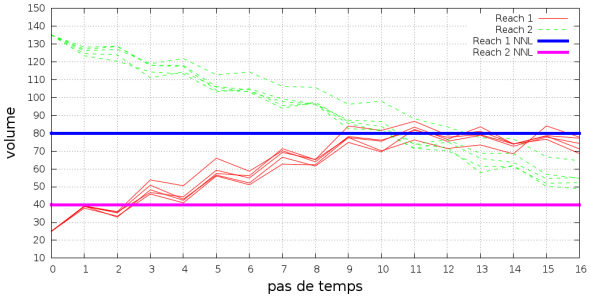

FIGURE 5 – Scénario 2 : partant du LNL et du HNL sans perturbation

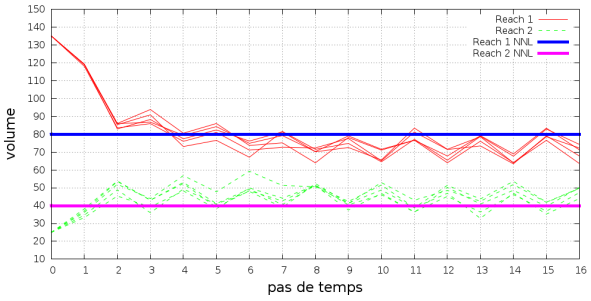

FIGURE 6 – Scénario 3 : partant du HNL et du LNL sans perturbation

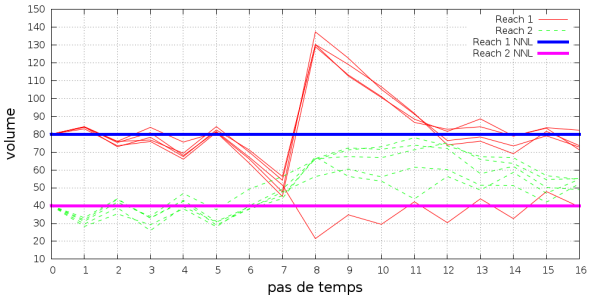

FIGURE 7 – Scénario 4 : partant des NNL et avec perturbation

Et il est observable que la politique, afin de se prémunir contre ce débordement, commence à vider le premier bief quelques pas de temps avant la perturbation, l'éloignant du NNL, tout en préservant les rectangles de navigation.

4.3 Limitation du passage à l'échelle

Une implémentation naïve de la fonction de transition consiste à créer une matrice de taille $|S|^2 \times |A|$, qui contiendrait pour notre exemple $1377^2 \times 60 = 1.137\,677\,40 \times 10^8$ valeurs. En supposant que le stockage en mémoire de chaque valeur nécessite 8 octets, cela impliquerait un minimum 0.91 Go d'espace mémoire pour stocker la fonction de transition.

Cependant, la plupart des transitions entre états sont impossibles donc nulles et cela, principalement du fait de la continuité temporelle. Pour un état s au pas temps t, seuls les états au pas de temps $t + 1$ sont atteignables. Pour ce type de configuration, une matrice creuse permet de réduire drastiquement le nombre de valeurs à stocker. Seuls les indices des valeurs et leurs valeurs ont à être stockés pour les valeurs non nulles. Tant que la matrice est plus qu'à moitié vide, l'utilisation de matrice creuse est bénéfique.

Lors du calcul de la politique optimale du MDP, plusieurs résultats ont pu être observés : les politiques donnent des résultats attendus ; la construction de la fonction de transition est une des étapes les plus longues ; enfin, la convergence des algorithmes est très rapide. La définition du temps n'étant pas cyclique, le choix des actions au dernier pas de temps ne dépend que de la récompense immédiate (une seule itération) et récursivement, les actions des états à un pas de temps t ne dépendent que des états pour les pas de temps supérieurs ($\forall t' > t$). Cela donne donc une borne maximale sur le nombre d'itérations nécessaires pour trouver la politique optimale. Elle est égale au nombre total de pas de temps modélisés, ici 17. Le nombre d'itérations étant fixé indépendamment de γ, il est possible d'utiliser $\gamma = 1$ afin d'obtenir les meilleurs résultats.

5 Approches pour contourner les limitations de l'approche naïve

Lorsque l'on augmente la taille du réseau étudié et/ou la précision des intervalles, nous nous heurtons rapidement à des problèmes de limitations de mémoire. En effet, l'ensemble d'états croît, par construction, exponentiellement par rapport au nombre de biefs. Or les applications réelles contiennent un nombre important de biefs. Le réseau de voies navigables du nord de la France contient près de 50 biefs. Notre approche naïve serait par exemple incapable de construire l'ensemble des états nécessaire à notre application. Pour contourner ce type de limitations spatiales, des extensions des MDP ont été définies dans la littérature telle que la représentation factorisée du modèle. Une autre technique consiste à décomposer, à diviser le MDP en sous-MDPs locaux. Nos recherches se limitent sur les approches permettant une planification sur l'ensemble de l'espace d'états. Pour le moment, nous ne sommes pas intéressés par les approximations comme les recherches arborescentes de Monte-Carlo [8], où la politique calculée ne s'applique qu'aux états les plus probables. Cette approche requiert la connaissance des états initiaux et un mécanisme continu si le système dérive vers des états inconnus.

5.1 MDP factorisé

L'approche des MDP factorisés vise à représenter de manière compacte la fonction de transition et de récompense, introduite dans *Boutilier, Dearden, Goldszmidt et al.* [4]. Pour cette raison, les états sont représentés par une assignation des variables. Chaque variable peut avoir une influence sur la valeur d'une variable spécifique au pas de temps suivant. L'idée derrière les MDP factorisés est d'exploser l'espace d'état pour regrouper les parties similaires des fonctions de transition et de récompense.

Dans notre cas, l'espace d'états (resp. d'actions) est obtenu par le produit cartésien de l'espace d'état de chaque bief (resp. espace d'action de chaque point de transfert). De plus l'espace d'état d'un bief ne dépend, en règle générale, pas directement de l'état de ses voisins, seules les actions l'influencent. Lorsqu'un bief reçoit de l'eau, les volumes des biefs amont et aval ne sont pas utilisés pour déterminer le nouveau volume, seuls les volumes échangés sont utilisés. Nous supposons qu'il est toujours possible de déplacer de l'eau, un bief n'étant jamais ni plein ni vide. Les actions conduisant à de tels cas sont interdites dans le modèle. Les états des biefs sont indépendants et nous pourrions l'utiliser pour factoriser notre MDP.

5.2 MDP décomposé

La décomposition d'un MDP permet de réduire la complexité du calcul de la politique en construisant une hiérarchie entre des problèmes locaux et une solution globale [3, 7]. Elle est particulièrement efficace dans des problèmes spatiaux puisqu'elle est basée sur les aspects topologiques des transitions.

Dans la majorité des problèmes réels, la décomposition de MDP pourrait simplifier le calcul de la politique (avec ou sans garantie d'optimalité), mais requiert de générer une partition de l'ensemble d'états [15, 17]. S'il n'y a pas de décomposition évidente, le partitionnement est un problème très difficile [2] et pourrait pénaliser l'approche par décomposition.

Un MDP de voies navigables n'est pas facilement décomposable, bien que chaque état représente un aperçu du réseau entier. Cependant, une option pourrait être de considérer plusieurs niveaux de détérioration des conditions de navigation. Chaque sous-MDP correspondant à un niveau de détérioration produira une politique visant à rétablir les conditions normales de navigation. Par exemple, si le MDP est divisé en trois sous-MDPs : normal, inondation, sécheresse, nous pouvons nous attendre que la politique gardera le système dans les états normaux (proche du NNL) avec peu de dépendance entre les trois sous-MDPs. De cette manière, résoudre en premier le sous-MDP normal puis les deux autres permettrait d'accélérer le calcul de la politique. Cependant, la décomposition n'aura pas d'impact sur la taille de la fonction transition, seul son calcul pourra s'effectuer en plusieurs temps.

5.3 MDP distribué

Les MDPs distribués semblent être une méthode ad hoc pour résoudre les problèmes coopératifs d'une modélisation multi-agents. Une telle approche est utilisée pour résoudre des MDPs décentralisés [6, 12], un framework où la politique doit être distribuée sur les agents et utilisée de façon décentralisée. Chaque agent est responsable du calcul de sa propre politique en prenant en compte ses objectifs. Des mécanismes orientés protocole permettent aux agents d'adapter leur politique afin d'atteindre un intérêt commun. Les MDP distribués sont utilisés, par exemple, dans une mission robotique, pour traiter la coordination de voyageurs de commerce [9].
Cette approche combine les idées de la factorisation et la décomposition. Le MDP est divisé en plusieurs sous-MDPs en partitionnant l'ensemble d'états et d'actions. Chaque sous-MDP est ensuite responsable d'un sous-ensemble des variables du problème et ignore les autres. Dans une modélisation orientée agents, chaque sous MDP correspondra aux capacités d'un agent dans le groupe (perceptions individuelles et actions).
Une approche itérative est utilisée pour résoudre les MDPs distribués, chaque itération modifiera la structure de chaque sous-MDP (valeurs de la fonction de transition et/ou récompense). Le calcul s'arrêtera lorsque les politiques seront stables pour tous les sous-MDPs (agents). L'espace d'états exploré pour calculer la politique pourrait être significativement réduit. Cela permet d'accélérer le calcul, sans qu'il y ait cependant de garantie sur l'optimalité de la solution.
Le réseau de voies navigables devrait être facilement distribuable puisque les points de transfert sont déjà distribués sur un territoire. Un agent serait responsable du contrôle d'un ou plusieurs points de transfert et le mécanisme de coordination serait basé sur les biefs communs à un ou plusieurs agents. Les MDPs distribués semblent très prometteurs pour répondre à notre problème, car ils peuvent réduire significativement la complexité de calcul et permettent une définition flexible du réseau. Cependant, la résolution de MDP distribué est une approche récente, sans framework générique établi. Les résultats restent donc incertains.

6 Vers une modélisation distribuée

Nous avons commencé à explorer l'adaptation de notre cas d'étude à l'approche distribuée. Nous définissons un agent comme un sous-ensemble de points de transfert du réseau. Un choix cohérent, consiste à ce que l'agent forme un sous-graphe connexe, dans le graphe formé par le réseau, avec les biefs comme sommets et les points de transfert comme arcs. De plus, chaque point de transfert ne peut être assigné qu'à un agent. Notons $\alpha = \{L_{i,j}, \ldots, L_{i',j'}\}$ la représentation d'un agent et $reach_\alpha = \{i, \ldots, i', j, \ldots j'\}$ l'ensemble des biefs affectés par cet agent. Nous considérons que si deux agents affectent un même bief alors ces agents sont voisins.

Un état d'un agent représente l'état des biefs affectés par les points de transfert de l'agent. Nous avons donc l'ensemble d'états de l'agent α :

$$S_\alpha = \{1, \ldots, t\} \times \prod_{i \in reach_\alpha} [0, ri_{out}]$$

Trivialement, l'ensemble d'actions d'un agent est la combinaison des intervalles de volumes de ses points de transfert.

$$A_\alpha = \prod_{l \in \alpha} l$$

La modélisation des états et des agents est la même que pour l'approche naïve, le découpage du réseau en agent reste néanmoins à définir.
L'algorithme de résolution envisagé est une variante de l'algorithme $LID - JESP$ [12]. Il se déroulerait de la façon suivante.

1. Calculer une politique gloutonne : π_α
2. Partager cette politique avec ses voisins
3. Recevoir les politiques des voisins
4. Mettre à jour T_α avec les politiques reçues
5. Déterminer l'amélioration possible g_α de π_α sur le nouveau MDP
6. Partager g_α avec ses voisins
7. Si un voisin à une plus grande amélioration possible alors retourner en 3.
8. Construire π_α à partir de $\langle S_\alpha, A_\alpha, T_\alpha, R_\alpha \rangle$
9. Si l'algorithme converge la politique optimale locale est π_α
10. Sinon retourner en 2.

L'objectif en utilisant cette approche distribuée consiste à valider une réduction de la complexité des calculs et permettre une définition flexible du réseau tout en limitant les pertes de qualité sur les solutions produites de supervision du réseau. Si la réduction au moins en taille de la modélisation est assez évidente, les points critiques d'une telle approche sont liés au processus distribué (convergence et qualité globale des politiques jointes).

7 Conclusion

Dans cet article, une approche orientée MDP est présentée pour optimiser la gestion de l'eau dans un réseau de voies navigables avec une vue globale et une planification sur un horizon de gestion. Cette approche vise à réduire l'impact des sécheresses et des inondations qui risque d'augmenter à cause du changement climatique, dans les années à venir. En utilisant des MDPs, il est possible de modéliser la dynamique et les incertitudes d'un tel système afin d'optimiser les conditions de navigation. Cependant, ce modèle est rapidement limité par la taille de l'espace d'état. Mais plusieurs pistes permettraient de contourner cette limitation : les approches factorisées, décomposées et distribuées des

MDPs. Les approches factorisées et distribuées tirent avantage de la corrélation des variables dans la définition des états et des actions.
Une approche basée sur les MDPs distribués est envisagée plus concrètement avec une modélisation orientée agents du réseau de voies navigables. Cette approche devrait permettre un bon contrôle des ressources de calcul nécessaire pour produire une solution intéressante, mais sans garanties a priori sur son optimalité.
Dans des travaux futurs basés sur cette approche, nous chercherons à valider une bonne gestion des ressources pour gérer des réseaux de toutes tailles. L'étude du calcul de la probabilité d'atteindre l'intervalle attendu est souhaité, en parallèle d'une détermination de règles de discrétisation pertinantes, en se basant sur de donnéess réelles d'exploitation.

Références

[1] R. Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(4) :679–684, 1957.

[2] C.-E. Bichot et P. Siarry. *Graph Partitioning*. Wiley-ISTE, 2011.

[3] C. Boutilier, T. Dean, et S. Hanks. Decision-theoretic planning : Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11 :1–94, 1999.

[4] C. Boutilier, R. Dearden, M. Goldszmidt, et others. Exploiting structure in policy construction. In *IJCAI*, volume 14, pages 1104–1113, 1995.

[5] C. Brand, M. Tran, et J. Anable. The UK transport carbon model : An integrated life cycle approach to explore low carbon futures. *Energy Policy*, 41 :107–124, 2012.

[6] I. Chades, B. Scherrer, et F. Charpillet. A Heuristic Approach for Solving Decentralized-POMDP : Assessment on the Pursuit Problem. In *SAC '02 : Proceedings of the 2002 ACM symposium on Applied computing*, pages 57–62, Madrid, Spain, 2002. ACM.

[7] T. Dean et S. hong Lin. Decomposition techniques for planning in stochastic domains. In *In Proceedings Of The Fourteenth International Joint Conference On Artificial Intelligence (IJCAI-95)*, pages 1121–1127. Morgan Kaufmann, 1995.

[8] L. Kocsis et C. Szepesvári. Bandit Based Monte-Carlo Planning. In J. Fürnkranz, T. Scheffer, et M. Spiliopoulou, editors, *Machine Learning : ECML 2006*, volume 4212 of *Lecture Notes in Computer Science*, pages 282–293. Springer Berlin Heidelberg, 2006.

[9] G. Lozenguez, L. Adouane, A. Beynier, A.-I. Mouaddib, et P. Martinet. Punctual versus continuous auction coordination for multi-robot and multi-task topological navigation. *Autonomous Robots*, pages 1–15, 2015.

[10] I. Mallidis, R. Dekker, et D. Vlachos. The impact of greening on supply chain design and cost : a case for a developing region. *Journal of Transport Geography*, 22 :118–128, 2012.

[11] S. Mihic, M. Golusin, et M. Mihajlovic. Policy and promotion of sustainable inland waterway transport in Europe – Danube River. *Renewable and Sustainable Energy Reviews*, 15(4) :1801–1809, 2011.

[12] R. Nair, P. Varakantham, M. Tambe, et M. Yokoo. Networked Distributed POMDPs : A Synthesis of Distributed Constraint Optimization and POMDPs. In *National Conference on Artificial Intelligence*, page 7, 2005.

[13] H. Nouasse, L. Rajaoarisoa, A. Doniec, E. Duviella, K. Chuquet, P. Chiron, et B. Archimede. Study of drought impact on inland navigation systems based on a flow network model. In *Information, Communication and Automation Technologies (ICAT), 2015 XXV International Conference on*, pages 1–6. IEEE, 2015.

[14] R. K. Pachauri, M. Allen, V. Barros, J. Broome, W. Cramer, R. Christ, J. Church, L. Clarke, Q. Dahe, P. Dasgupta, et others. Climate Change 2014 : Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 2014.

[15] R. Parr. Flexible Decomposition Algorithms for Weakly Coupled Markov Decision Problems. In *14th Conference on Uncertainty in Artificial Intelligence*, pages 422–430, 1998.

[16] M. L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[17] R. Sabbadin. Graph partitioning techniques for Markov Decision Processes decomposition. In *15th Eureopean Conference on Artificial Intelligence*, pages 670–674, 2002.