
Vers une distribution des 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
prenom.nom@mines-douai.fr

RÉSUMÉ. La gestion des réseaux de voies navigables devrait subir des modifications notables en raison d'une volonté d'accroître le trafic fluvial dans un contexte de changement climatique. Ces nouvelles contraintes vont imposer une gestion adaptative et résiliente de la ressource en eau conduisant à une planification de sa répartition sur l'ensemble de ces réseaux. Une approche basée sur les MDP est proposée afin de répondre à cette problématique. Elle permet la coordination de plusieurs entités sur plusieurs pas de temps. Son utilisation sur un sous-réseau de voies navigables conduit à une minimisation de l'impact négatif lors d'événements de montées des eaux ou de sécheresse. Malgré des premiers résultats prometteurs, les mécanismes de passage à l'échelle permettant de considérer une application réelle ne sont pas totalement définis. Différentes approches proposées dans la littérature pour le passage à l'échelle sont discutées en dégageant leurs avantages et limitations. Parmi elles, une modélisation distribuée est privilégiée et un nouvel algorithme de résolution est proposé. Il est testé sur un sous-réseau de 7 biefs.

ABSTRACT. Inland waterways networks management methods should undergo significant changes due to a commitment to increase the waterway traffic in a context of climate change. These new constraints will impose an adaptive and resilient management of the water resource leading to an optimal planning of its distribution over the integrity of the inland waterway network, while taking into account the uncertainties arising from their operation. A MDP based approach is proposed to address this problem. It allows the coordination of multiple entities over multiple time steps. Its use on a subnetwork of the waterway made from 2 reaches leads to a minimization of the impacts of flood and drought periods. Despite promising results, the scaling mechanisms for considering a real application are not fully defined. Different approaches proposed in the literature for scaling are discussed by identifying their advantages and limitations. Among them, a distributed modeling is privileged and a new resolution algorithm is proposed. It is tested on a subnetwork composed of 7 reaches.

MOTS-CLÉS : processus de décision markovien, réseau de voies navigables, grand modèle.

KEYWORDS: Markov decision process, inland waterway network, large model.

DOI:10.3166/RIA.31.183-205 © 2017 Lavoisier

1. Introduction

Comme en atteste l'accord de la COP21 signé le 12 décembre 2015 à Paris, le changement climatique est une préoccupation actuelle majeure des sociétés modernes. Cet accord qui a été validé par tous les participants et récemment entériné par la Chine et les États-Unis vise à mettre en œuvre toutes les mesures possibles pour limiter le réchauffement mondial entre 1,5 °C et 2 °C d'ici 2100. Ces prédictions tenant compte de nouvelles mesures d'adaptation sont basées sur les rapports du GIEC *Groupe d'experts Intergouvernemental sur l'Évolution du Climat*. Dans le domaine du transport, le dernier rapport du GIEC préconise des mesures d'adaptation notamment par la promotion de modes de transport alternatif (IPCC, 2014). En ce sens, le transport par voies navigables est reconnu comme plus écologique et économique (Brand *et al.*, 2012a).

Aussi, il est attendu une hausse notable de la demande de transport avec une augmentation estimée à + 35 % pour le transport fluvial en Europe d'ici 2050 (Beuthe *et al.*, 2012). En parallèle, plusieurs études sur les effets du changement climatique ont été réalisées ces dernières années (EnviCom, 2008 ; IWAC, 2009 ; Arkell, Darch, 2006 ; Bates *et al.*, 2008 ; Wanders, Wada, 2015 ; Beuthe *et al.*, 2014 ; Pachauri *et al.*, 2014). Ces études montrent que l'intensité et la fréquence des périodes d'inondation et de sécheresse vont augmenter, pouvant impacter le transport par voies navigables.

La conception de stratégies de gestion des réseaux de voies navigables en tenant compte de ces nouvelles contraintes est une priorité pour les gestionnaires. Les réseaux de voies navigables sont des systèmes de grandes dimensions aménagés par l'Homme pour répondre à ses besoins. Ils ont une forte interaction avec l'environnement naturel. Sous l'hypothèse d'une connaissance exhaustive de l'ensemble de ces interactions, des approches de gestion adaptative de réseaux de voies navigables dans un contexte de changement climatique ont été proposées dans (Nouasse *et al.*, 2015 ; 2016) à partir de problèmes de satisfaction de contraintes ou de techniques d'optimisation quadratique (Duviella *et al.*, 2016). Cependant, une majorité de ces interactions n'est que partiellement connue. Il s'agit de rejets sauvages, d'échanges avec les nappes phréatiques, de l'influence locale de phénomènes météorologiques... La gestion de tels réseaux est donc soumise à de nombreuses incertitudes. Ainsi, une approche basée sur une modélisation stochastique semble tout adaptée.

L'objectif consiste à concevoir des stratégies de gestion adaptative et résiliente de réseaux de voies navigables dans un contexte de changement climatique et d'accroissement de la navigation. Dans cet article, une modélisation stochastique de réseaux de voies navigables basée sur les processus de décision markoviens (*Markov Decision Processes* - MDP) est proposée afin d'optimiser la planification de la répartition de la ressource en eau sur l'ensemble des réseaux en tenant compte des nombreuses incertitudes inhérentes à leur fonctionnement. Les MDP sont largement utilisés pour la planification de modèles stochastiques. Ils permettent de modéliser l'évolution de systèmes incertains et conduisent à un plan optimal pour toutes les configurations possibles des modèles conçus. Cependant, leur utilisation se heurte à des difficultés de résolution, notamment à cause de la dimension des espaces de recherche à par-

courir. Dans une majorité d'applications, ils induisent des modèles intraitables. Les politiques de contrôle optimal nécessitent alors des algorithmes spécifiques de résolution (Boutilier *et al.*, 1999 ; Nair *et al.*, 2005).

La complexité dans cette application vient de la dimension des réseaux de voies navigables et des nombreuses possibilités de déplacement d'eau. En effet, il est nécessaire de considérer l'ensemble des actions conjointes (déplacements d'eau) et l'ensemble des configurations possibles sur la totalité du réseau de voies navigables étudié. Cela induit une explosion combinatoire du nombre d'états. Cet article vise dans un premier temps à décrire l'approche de modélisation par MDP de réseaux de voies navigables, qui à notre connaissance n'a jamais été utilisé dans le contexte du transport fluvial. Dans un deuxième temps, les limitations d'un tel modèle lorsqu'il est conçu de manière centralisée sont discutées. Une modélisation distribuée et décentralisée est proposée dans un troisième temps afin de contourner ces limitations.

La problématique de gestion d'un réseau de voies navigables dans le contexte du changement climatique et d'accroissement du trafic fluvial est présentée dans la section 2. Une modélisation naïve du réseau en utilisant les MDP est proposée en section 3. En section 4, l'étude d'un réseau composé de deux biefs permet d'illustrer l'approche proposée et de discuter de ses limitations. La section 5 est dédiée à la description et à la comparaison de différents dérivés des MDP permettant de lever les limitations liées à l'étude de systèmes à grande échelle. Enfin, de façon à contourner ces limitations, une modélisation via un MDP distribué est proposée dans la section 6. Elle est illustrée sur un réseau composé de sept biefs.

2. Gestion 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 (Mallidis *et al.*, 2012 ; Mihic *et al.*, 2011), tout en assurant un transport discret, efficace et sûr des biens (Brand *et al.*, 2012b). Il est constitué majoritairement des rivières canalisées et des canaux artificiels, le tout interconnecté. Ces rivières et canaux sont composés de biefs reliés par des écluses. Un bief est donc une partie d'une rivière ou d'un canal séparé par au moins deux écluses.

L'objectif principal des gestionnaires des voies navigables est de maintenir un certain niveau d'eau dans tous les biefs du réseau afin de permettre la navigation. Ce niveau doit respecter les conditions définies par le 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*.

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 vers le bief aval. D'autres perturbations du niveau

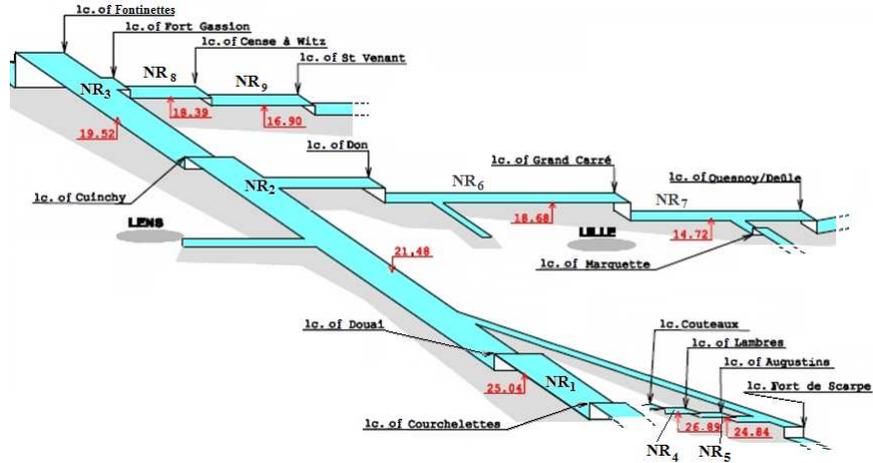


Figure 1. Extrait du réseau de voies navigables du nord de la France

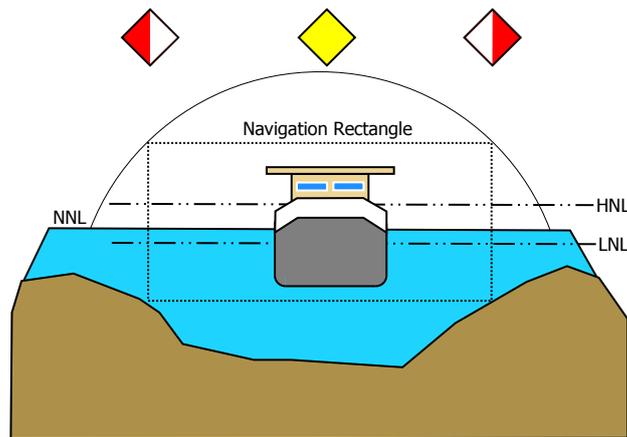


Figure 2. Rectangle de navigation

d'eau peuvent exister telles que des échanges avec les eaux souterraines, 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 ouvrages, tels que des vannes ou des barrages sont utilisés pour envoyer de l'eau en aval et des pompes, lorsqu'elles sont disponibles, peuvent renvoyer de l'eau en amont. Ce sont les principaux ouvrages qui sont utilisés pour la répartition de la ressource en eau entre les biefs du réseau.

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. Cependant, de nouvelles contraintes liées au changement

climatique et à une volonté d'augmenter le trafic fluvial devraient impacter les stratégies de gestion actuelles. L'objectif principal est donc d'anticiper l'impact de ces nouvelles contraintes en concevant des approches de gestion adaptative visant à assurer en chaque point du réseau à chaque instant les conditions de navigation requises. Cela consiste à déterminer un plan global de répartition de la ressource en eau sur la totalité du réseau en tenant compte des incertitudes liées aux événements climatiques et à la demande de navigation. Une planification de la répartition de la ressource en eau sur plusieurs pas de temps permet ainsi une meilleure anticipation des événements possibles. Les informations sur l'état courant du réseau de voies navigables sont recueillies en temps réel grâce au réseau de capteurs de niveau équipant les biefs.

3. 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 termes de volumes d'eau) et le contrôle est incertain du fait des entrées/sorties d'eau incontrôlées.

3.1. Définition

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 une assignation 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 (Bellman, 1957) définie pour chaque état par :

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

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

Divers algorithmes existent pour résoudre de manière optimale un MDP. Une version notable est *Value Iteration* (Puterman, 1994). Il construit itérativement la fonction de valeur V , en utilisant l'équation (3) pour un nombre spécifié d'itérations corres-

pondant à l'horizon fini de planification. Une version à horizon infini existe avec un facteur d'atténuation modulant l'importance des récompenses futures et immédiates.

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

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.

3.2. Application aux réseaux de voies navigables

Le but de l'approche proposée est de planifier la meilleure suite d'actions pour l'ensemble du réseau de voies navigables sur τ pas de temps, tout en sachant que certaines conditions pourront évoluer à chaque pas de temps et ainsi affecter les objectifs de navigation. Par exemple, le temps peut devenir pluvieux augmentant le niveau d'eau dans les biefs affectés; une augmentation imprévue du trafic fluvial sur certains biefs implique une plus grande utilisation des écluses et donc de transferts de volumes d'eau supplémentaires.

Des demi-journées (12 heures) sont utilisées comme pas de temps, de façon à séparer les périodes de navigation le jour et les périodes d'inactivité la nuit. L'utilisation de pas de temps larges permet de considérer le niveau d'eau d'un bief comme uniforme et de lisser 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. Comme les volumes d'eau observés (obtenus à partir des mesures de niveau) et transférés entre les biefs du réseau sont continus, ils ont dû être discrétisés sous forme d'intervalles. Ainsi, la plage de fonctionnement de chaque bief est divisée en intervalles réguliers à l'exception du premier et du dernier intervalle, considérés de taille infinie. Ces derniers qui correspondent aux valeurs en dehors du rectangle de navigation 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 pas d'intervalle de taille infinie (voir figure 3).

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 (voir figure 3). L'intervalle 0 correspond à tout volume inférieur au LNL et l'intervalle ri_{out} représente tout volume supérieur au HNL.

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

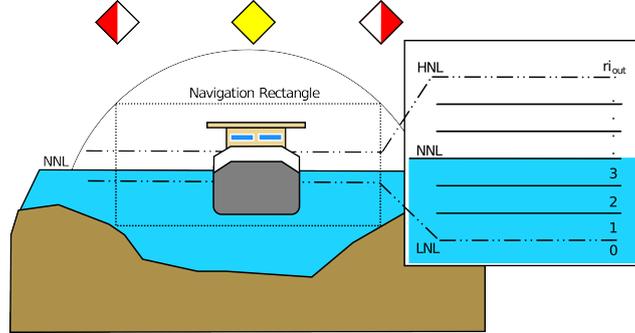


Figure 3. Discretisation du volume d'eau d'un bief en intervalles

où N représente le nombre de biefs dans le réseau. L'identification du pas de temps courant dans l'état permet d'exprimer dans le temps les probabilités sur les entrées sorties incontrôlées. Le temps étant inclus dans la représentation des états et n'étant pas cyclique, la politique optimale sera trouvée au pire en τ itérations. L'horizon de résolution du MDP sera donc fixé à τ .

Similairement, l'ensemble d'actions A est défini comme la combinaison des intervalles de volume des points de transfert comme autant de sous-actions réalisées de façons distribuées sur le réseau. Les actions étant indépendantes du temps, soit :

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

où $A_{i,j}$ représente l'ensemble des intervalles du point de transfert reliant le bief i au bief j , avec le bief 0 qui correspond aux rivières n'appartenant pas au réseau et à tout autre élément externe pouvant apporter ou retirer des volumes d'eau aux biefs considérés.

Le volume à transférer du bief i vers le bief j prévu par l'action $a \in A$ est noté $a_{i,j} \in A_{i,j}$. Cependant, il existe un nombre très limité de points de transfert effectif. Pour tout point de transfert $A_{i,j}$ entre deux biefs i et j non connecté, aucun transfert d'eau n'est autorisé ($A_{i,j} = \{0\}$). Pour simplifier la notation, a_i représente la partie de l'action qui affecte le bief i tel que :

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

avec \oplus et \ominus deux opérateurs, définis sur $(\mathbb{R} \cup \mathcal{I})^2 \rightarrow \mathbb{R}$, qui peuvent respectivement ajouter et soustraire des intervalles de nombre et/ou des nombres; le résultat étant toujours un réel, et \mathcal{I} étant l'ensemble de tous les intervalles possibles de réseau étudié. Ces opérateurs sont respectivement une simple addition et soustraction qui utilisent la valeur du membre s'il s'agit d'un réel, et 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 définis aux pas de temps t et t' , s' n'est atteignable que si $t' = t + 1$. Dans cette fonction de transition, l'état d'un bief est indépendant de celui des autres biefs. En effet, il ne dépend que des volumes d'eau entrants et sortants. Pour définir formellement cette fonction de transition, il est nécessaire de décrire au préalable les sources d'incertitudes impactant le réseau étudié.

Une première source d'incertitudes sur les transitions vient des déplacements d'eau non contrôlés, modélisés par une liste de variations temporelles notées V_t . De la pluie sur un bief, par exemple, est une variation temporelle. Ces variations temporelles sont des changements locaux à un ou plusieurs biefs pendant un ou plusieurs pas de temps avec une certaine probabilité. Comme elles ne sont pas dans l'espace d'action, elles affectent uniquement la fonction de transition. V_t contient toutes les combinaisons possibles de variations temporelles pouvant affecter le réseau modélisé à un pas de temps t . Si le réseau est affecté par la variation $v_t \in V_t$, alors le volume non contrôlé qui affecte le bief i est noté $v_t(i)$. La probabilité que cette combinaison se produise au pas de temps t est représenté par $P(v_t)$.

Une seconde source d'incertitudes est induite par la discrétisation en intervalles des volumes transférés et des volumes contenus dans les biefs. Cette discrétisation cause une approximation dans la représentation des états. En considérant cette seconde source d'incertitudes, une probabilité $P(r_{i_{s'}} | r_{i_s}, a_i + v_t(i))$ est définie. Elle correspond à la probabilité que le volume d'eau du bief i appartenant à l'intervalle r_{i_s} à l'instant t soit inclus dans l'intervalle $r_{i_{s'}}$ au pas de temps $t' = t + 1$ si l'action a_i est effectuée avec un déplacement incontrôlé de volume $v_t(i)$ en i . Ainsi :

$$P(r_{i_{s'}} | r_{i_s}, a_i + v_t(i)) = \begin{cases} p_{=} & \text{si } r_{i_s} \oplus a_i + v_t(i) \in r_{i_{s'}} \\ p_{+} & \text{si } r_{i_s} \oplus a_i + v_t(i) \in r_{i_{s'}} - 1 \\ p_{-} & \text{si } r_{i_s} \oplus a_i + v_t(i) \in r_{i_{s'}} + 1 \\ 0 & \text{sinon} \end{cases} \quad (7)$$

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

$$p_{+} + p_{-} + p_{=} = 1$$

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

$$T(s, a, s') = \sum_{v_t \in V_t} \left(P(v_t) \times \prod_{i=1}^N P(r_{i_{s'}} | r_{i_s}, a_i + v_t(i)) \right) \quad (8)$$

L'évolution du système est donc caractérisé par une somme de produit de probabilité basée sur des sous-actions $a_{i,j}$ interconnectées par des super sous-action a_i .

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

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

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 g . Lorsque $ri_{s'}$ n'est que partiellement en dehors du rectangle de navigation, il est remplacé par $\frac{g}{2}$. Ces permutations ont pour objectif de pénaliser fortement le non-respect du rectangle de navigation. f est une fonction de coût lié à l'utilisation des différents ouvrages d'art définie en fonction des caractéristiques du réseau.

4. Gestion d'un réseau composé de deux biefs

Un réseau de navigation réaliste est défini afin d'implémenter et de tester l'approche proposée. Il est composé de deux biefs et de six points de transfert (voir figure 4).

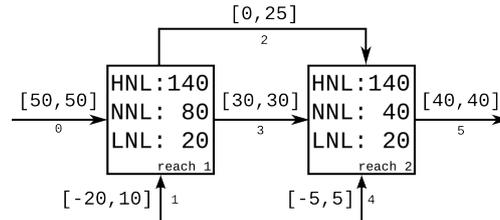


Figure 4. Réseau de voies navigables

4.1. Caractéristiques du réseau

Sur la figure 4, les biefs sont représentés par des carrés, avec le rectangle de navigation spécifié en unité de volume. 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.

Les volumes des deux biefs sont divisés en 9 intervalles dont 7 sont de taille régulière, fixée à 20 unités. Le premier et le dernier intervalle sont considérés de taille infinie. Les volumes des points de transfert 1, 2 et 4 sont divisés en intervalles réguliers de taille 5. Ce qui conduit respectivement à 6, 5 et 2 intervalles pour chacun de ces points de transfert. Les points de transfert 0, 3 et 5 représentent les écluses. Ils transfèrent une quantité d'eau constante, correspondant au trafic fluvial.

Les scénarios de fonctionnement du réseau étudié sont construits sur une période de 8 journées et 8 nuits, soit 16 pas de temps. Il est à noter que les écluses ne sont

pas utilisées la nuit puisqu'aucune navigation n'est permise. La planification étant réalisée sur 16 pas de temps, l'ensemble d'états S du modèle (voir relation 4) est tel que $S = \{0, \dots, 16\} \times \prod_{i=1}^2 [0, r_{i_{out}}] = 9 \times 9 \times (16 + 1)$, soit 1 377 états. Un pas de temps supplémentaire est ajouté pour marquer la fin de planification. Tous les états du dernier pas de temps sont considérés absorbants ($T(s, a, s) = 1$ et $R(s, a, s) = 0$). Similairement, le nombre d'actions A (voir relation 5) est $A = \prod_{i,j \in [0,2]^2} A_{i,j} = 1 \times 6 \times 5 \times 1 \times 2 \times 1$, soit 60 actions possibles.

La valeur de g est réglée arbitrairement à 10 000 afin de pénaliser le non respect du rectangle de navigation. Finalement, 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. Scénarios de fonctionnement

Plusieurs scénarios de fonctionnement ont été définis afin de tester l'approche proposée. Le premier scénario correspond à une situation idéale où les deux biefs sont initialement à leur niveau normal de navigation et durant laquelle aucune perturbation n'est introduite. Dans le second scénario, le premier bief est initialement proche de son niveau minimal de navigation et le second bief est proche de son niveau maximal, sans perturbation. Le troisième scénario est similaire au précédent à la différence près que le premier bief est initialement proche de son niveau maximal et le second de son niveau minimal. Finalement, le dernier scénario est défini à partir du premier scénario en y ajoutant une perturbation importante. Cette perturbation très probable ne dure qu'un seul pas de temps et n'affecte qu'un seul point de transfert. Cependant, son amplitude est telle qu'elle peut occasionner un débordement du premier bief. Il est important de noter que les simulations des quatre scénarios définis sont réalisées en considérant des volumes continus, de façon à être le plus proche possible du comportement réel du réseau.

Puisque les actions sont des intervalles de volumes, les scénarios ont été testés en transférant une valeur aléatoire dans ces intervalles, plutôt que de choisir la meilleure valeur ou la moyenne. Le but est 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. L'intérêt est de visualiser les conséquences de la sélection aléatoire des volumes transférés sur les politiques mises en œuvre. Les résultats obtenus sont représentés aux figures 5, 6, 7 et 8. Chacune permet de visualiser les 5 simulations réalisées par scénario.

La planification optimale du transfert des volumes entre les biefs en considérant le premier scénario permet de maintenir le niveau normal de navigation dans les deux biefs à chaque pas de temps. La politique obtenue conduit à des niveaux qui ne sont que relativement proches des objectifs définis (voir figure 5). Les écarts sont liés à l'approximation résultante de la discrétisation en intervalles. En effet, comme un in-

tervalle est représenté par sa moyenne, il est possible que le volume d'un bief augmente ou baisse tout en restant dans le même intervalle. C'est la raison de l'écart à l'optimalité. Néanmoins, avec un choix intelligent des volumes à transférer dans les intervalles, il est possible de maintenir le NNL de chaque bief en suivant la politique produite.

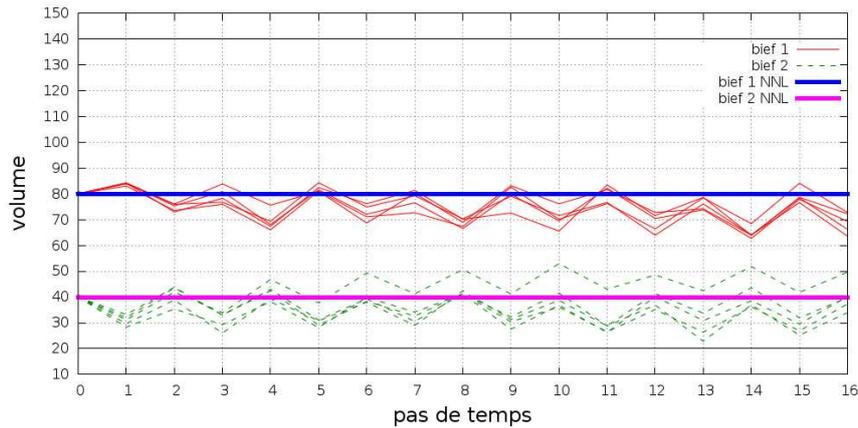


Figure 5. Scénario 1 : biefs avec NNL comme conditions initiales, sans perturbation

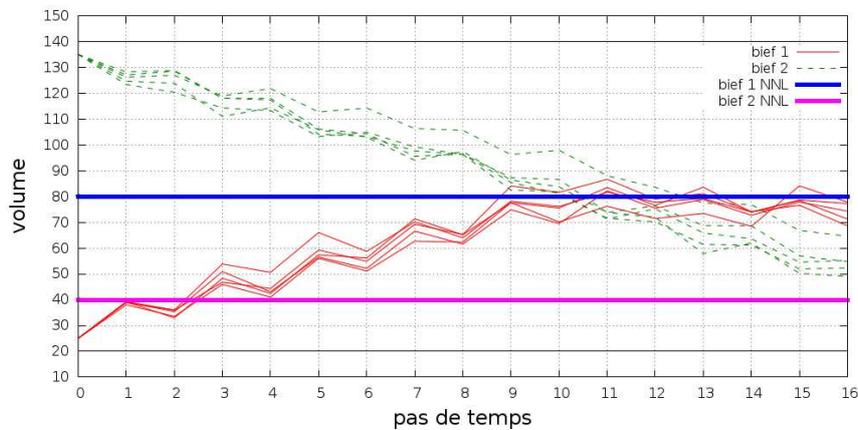


Figure 6. Scénario 2 : bief 1 proche du LNL et bief 2 du HNL comme conditions initiales, sans perturbation

Les résultats de la politique face à des événements qui avaient amené les biefs aux limites du rectangle de navigation sont visibles sur les figures 6 et 7 (scénarios 2 et 3 resp.). Dans les deux cas, la planification conduit au respect des objectifs de gestion. Il est cependant possible de remarquer que le respect des objectifs est plus rapide lors du scénario 3 ; ceci étant directement lié à la configuration du réseau. Il est en effet plus simple au bief amont de se vider dans le bief aval lorsque celui-ci doit être alimenté.

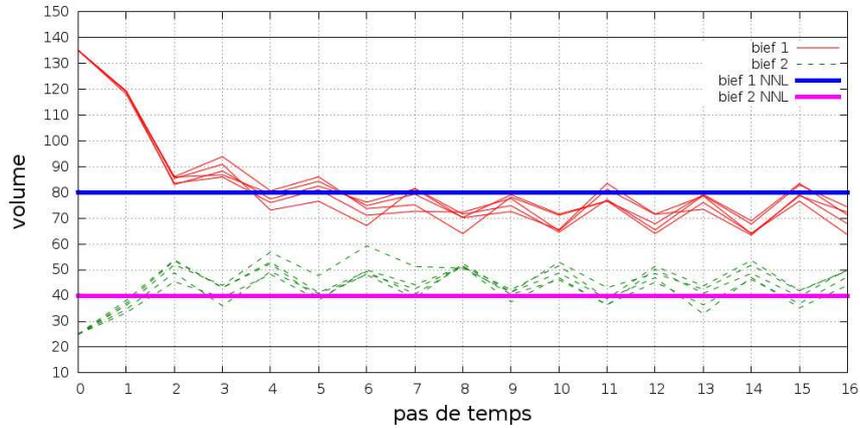


Figure 7. Scénario 3 : bief 1 proche du HNL et bief 2 du LNL comme conditions initiales, sans perturbation

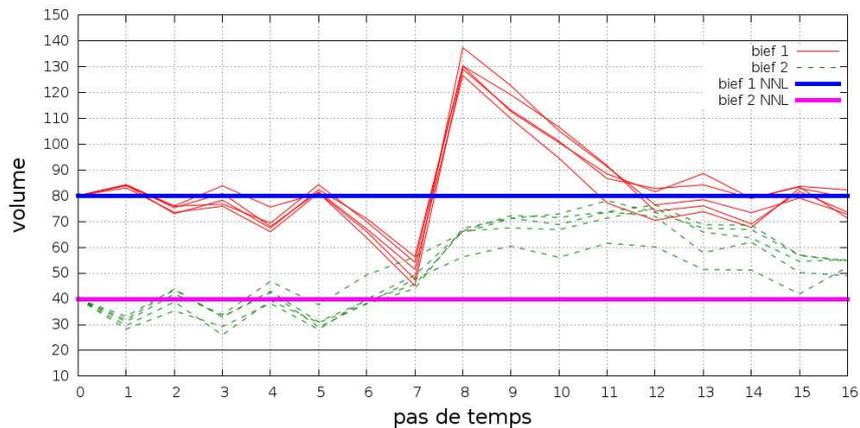


Figure 8. Scénario 4 : biefs avec NNL comme conditions initiales, avec perturbation

Dans le dernier scénario, une forte pluie localisée au niveau du bief amont est supposée se produire entre le pas de temps 7 et 8. Son amplitude est telle que si le bief amont est proche de son NNL avant son occurrence, le bief déborde. La politique mise en œuvre permet l'anticipation de cet événement. Afin de se prémunir contre ce débordement, le bief amont commence à se vider quelques pas de temps avant l'occurrence de la perturbation, l'éloignant de son NNL, tout en préservant le respect du rectangle 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 le réseau de voies navigables étudié $1377^2 \times 60 = 1.14 \times 10^8$ 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 d'une matrice creuse est bénéfique.

Lors du calcul de la politique optimale du MDP, plusieurs résultats ont pu être observés : les politiques donnent les résultats attendus; la construction de la fonction de transition est une des étapes les plus longues; enfin, la résolution est très rapide, les horizons étant généralement très petits.

Lorsque la taille du réseau étudié augmente et/ou les pas de discrétisation des intervalles diminuent, l'approche proposée se heurte 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. Par exemple, le réseau de voies navigables du nord de la France contient près de 50 biefs. Par conséquent, l'approche naïve proposée ne pourrait pas être mise en œuvre. Il serait matériellement impossible de construire l'ensemble des états nécessaire à son application.

5. Approches existantes dédiées au passage à l'échelle

Pour contourner les difficultés liées aux limitations spatiales, des extensions des MDP ont été définies dans la littérature telle que la représentation factorisée du modèle, ou la décomposition du MDP en sous-MDP locaux. Les approches discutées se limitent à celles permettant une planification sur l'ensemble de l'espace d'états. Dans cet article, les approximations comme les recherches arborescentes de Monte-Carlo (Kocsis, Szepesvári, 2006) où la politique calculée ne s'applique qu'aux états les plus probables, ne sont pas présentées. 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, introduite dans (Boutilier *et al.*, 1995), vise à représenter de manière compacte la fonction de transition et de récompense. 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 le cas d'étude proposé, 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, et seules les actions l'influencent. En effet, 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. De plus, il est supposé qu'il est toujours possible de déplacer de l'eau, puisqu'un bief n'est jamais ni plein ni vide. Les actions conduisant à de tels cas sont interdites dans le modèle. Les états des biefs sont indépendants et il serait possible d'utiliser cette propriété pour factoriser le MDP proposé.

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 (Boutilier *et al.*, 1999 ; Dean, Lin, 1995). 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 elle requiert de générer une partition de l'ensemble d'états (Parr, 1998 ; Sabbadin, 2002). S'il n'y a pas de décomposition évidente, le partitionnement est un problème très difficile (Bichot, Siarry, 2011) qui 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. Une option consisterait cependant à 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-MDP : normal, inondation, sécheresse, il est probable que la politique mise en œuvre conserve le système dans les états normaux (proche du NNL) avec peu de dépendance entre les trois sous-MDP. 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 MDP distribués semblent être une méthode *ad hoc* pour résoudre les problèmes coopératifs d'une modélisation multi-agent. Une telle approche est utilisée pour résoudre des MDP décentralisés (Chades *et al.*, 2002 ; Nair *et al.*, 2005), 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 protocoles de communication 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 (Lozenguez *et al.*, 2015).

Cette approche combine les idées de la factorisation et la décomposition. Le MDP est divisé en plusieurs sous-MDP 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 MDP distribués. Chaque itération modifiera la structure de chaque sous-MDP, *i.e.* les valeurs de la fonction de transition et/ou de récompense. Le calcul s'arrêtera lorsque les politiques seront stables pour tous les sous-MDP (agents). L'espace d'états exploré pour calculer la politique peut 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 étant donné que les points de transfert sont déjà physiquement répartis 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 MDP distribués semblent très prometteurs pour répondre à cette problématique puisqu'ils peuvent réduire significativement la complexité de calcul en permettant 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. Modélisation par MDP distribué

L'adaptation de notre cas d'étude à l'approche distribuée nécessite de définir un agent comme un sous-ensemble de points de transfert du réseau. Chaque point de transfert ne peut être assigné qu'à un seul agent. Notons $\alpha = \{A_{i,j}, \dots, A_{i',j'}\}$ la représentation d'un agent et $reach_\alpha = \{i, \dots, i', j, \dots, j'\}$ l'ensemble des biefs affectés par cet agent. Deux agents affectant un même bief sont considérés comme voisins.

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

$$S_\alpha = \{0, \dots, \tau\} \times \prod_{i \in reach_\alpha} [0, r_{i_{out}}]$$

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. Un choix cohérent consiste à ce que chaque agent forme un sous-graphe connexe, dans le graphe non orienté formé par le réseau, avec les biefs comme sommets et les points de transfert comme arcs. Ce type de décomposition permet de minimiser le nombre de biefs affectés par chaque agent.

6.1. Algorithme distribué

L'algorithme de résolution proposé (voir Algorithme 1) est inspiré de l'algorithme *LID – JESP* (Nair *et al.*, 2005). Il se base sur des agents n'ayant qu'une vision locale du système et dont les transitions peuvent dépendre d'actions d'autres agents. À chaque itération, au plus un agent par voisinage aura la possibilité d'améliorer sa politique. La résolution sera considérée comme terminée lorsqu'aucun agent ne pourra améliorer sa politique ou lorsque tous les agents auront atteint un cycle. Pour que la résolution ne s'arrête que lorsque tous les agents n'arrivent plus à s'améliorer, un critère d'arrêt, sous la forme d'un compteur, est utilisé. Chaque agent α possède un compteur compteur_α , initialisé à une valeur $d > 0$. Ce compteur diminue lorsqu'un agent n'arrive plus à s'améliorer ou est dans un cycle. Dans les autres cas, le compteur est réinitialisé à d (ligne 11). À la fin de chaque itération, les agents échangent leur compteur avec leurs voisins et conservent le plus grand compteur (ligne 12). Si le compteur atteint 0 alors l'agent a terminé sa résolution et possède sa politique finale (ligne 13).

PROPOSITION. — *L'algorithme s'arrêtera en $d = \max_{\alpha, \beta \in Ag^2} \text{dist}(\alpha, \beta)$ itérations si et seulement si les agents ne peuvent plus s'améliorer ou sont dans un cycle.*

Où dist est défini comme

$$\text{dist}(\alpha, \beta) = \begin{cases} 1 & \text{si } \beta \in N_\alpha \\ 1 + \min_{\gamma \in N_\alpha} \text{dist}(\gamma, \beta) & \text{sinon} \end{cases}$$

PREUVE. — Supposons que l'agent α n'entre pas dans l'itération c car il a terminé sa résolution ($\text{compteur}_\alpha(c-1) = 0$), mais certains agents ne sont pas dans un optimum local, cela signifie qu'à l'itération $c-d$, il existe au moins un agent β qui peut s'améliorer. Cela entraîne la réinitialisation de son compteur ($\text{compteur}_\beta(c-d) = d$). Comme tous les compteurs sont décrémentés d'au plus 1 à chaque itération et qu'ils sont partagés avec le voisinage, à l'itération $c-d + \text{dist}(\alpha, \beta)$ le compteur de l'agent α possédera une borne inférieure : $d - \text{dist}(\alpha, \beta)$. Cela permet de déduire la borne minimale du compteur de l'agent α pour sa dernière itération : $\text{compteur}_\alpha(c-1) \geq d - \text{dist}(\alpha, \beta) + 1 - d + \text{dist}(\alpha, \beta) = 1$. Puisque son compteur ne sera pas nul, l'agent

α effectuera l'itération c . Par contradiction, l'algorithme ne s'arrêtera que si tous les agents sont dans un optimum local.

À l'inverse, si tous les agents ont atteint un optimum local les compteurs ne seront jamais réinitialisés à d et diminueront de 1 à chaque itération. Donc après d itérations, $\forall \alpha \text{ compteur}_\alpha = 0$ et les agents s'arrêtent. ■

À chaque itération, chaque agent échangera sa politique et sa fonction de récompense, approximées aux états inconnus du voisin concerné, à tout son voisinage (ligne 3). Une politique approximée est une assignation de distribution de probabilité d'actions pour chaque état de l'agent qui reçoit la politique. $SV(s_\alpha)$ est l'ensemble des combinaisons possibles d'actions des différentes politiques reçues par l'agent α . Il est ainsi possible de définir T_α :

$$T_\alpha(s, a, s') = \sum_{sv \in SV(s)} \sum_{v_t \in V_t} P(sv) \times P(v_t) \times \prod_{i=1}^N P(ri_{s'} | ri_s, a + sv(i) \oplus v_t(i)) \quad (10)$$

Si les agents sont *altruistes*, alors la fonction de récompense sera redéfinie (ligne 6) comme la somme de la fonction de récompense locale avec les fonctions de récompenses partielles reçues. Dans le cas d'agents *égocentriques*, la fonction de récompense de l'agent sera sa fonction de récompense locale qui n'évoluera donc pas dans le temps.

Ces nouvelles fonctions de transition et de récompense vont être utilisées pour calculer la nouvelle politique optimale locale de l'agent (ligne 8). La politique optimale de ce modèle sera comparée à la politique actuelle de l'agent (ligne 9) afin de déterminer si la nouvelle politique améliore l'agent. Le gain de l'agent sera comparé avec celui de ses voisins et seul l'agent avec le plus grand gain dans son voisinage pourra conserver la politique calculée à cette itération (ligne 10).

Le gain de la nouvelle politique par rapport à la politique actuelle est calculé grâce à une fonction heuristique. Cette heuristique influencera fortement la qualité des résultats, la rapidité de résolution ainsi que la capacité à converger.

Pour ce cas d'étude, le gain est défini comme l'écart de probabilité d'atteindre les états les plus récompensés entre les deux politiques :

$$\text{gain}(\pi', \pi) = \frac{1}{\tau + 1} \times \sum_{t=0}^{\tau} P(NNL|\pi', t) - P(NNL|\pi, t)$$

La politique initiale consiste en la meilleure planification, à court terme, pour les biefs visibles de l'agent en ne prenant en compte que ses points de transfert.

où it représente l'itération actuelle, G_{it} (resp C_{it}) l'ensemble des gains (resp. compteur) du voisinage de l'agent lors de l'itération it

Algorithme 1 Algorithme de résolution distribuée

-
- 1: Construire une politique initiale π_0
 - 2: **répéter**
 - 3: Échange de π_{it} et R_{it} à ses voisins
 - 4: Construction de T_{it+1} à partir des politiques reçues
 - 5: **si** agent altruiste **alors**
 - 6: Construction de R_{it+1} à partir de fonctions de récompenses reçues
 - 7: **fin si**
 - 8: Calcul de π' politique optimale de $M' : \langle S, A, T_{it+1}, R_{it+1} \rangle$
 - 9: $g_{it} \leftarrow \text{gain}(\pi', \pi_{it})$ dans M'
 - 10: $\pi_{it+1} \leftarrow \pi'$ si $g_{it} = \max(G_{it})$ sinon π_{it}
 - 11: $\text{compteur} \leftarrow d$ si $g_{it} > 0$ sinon $\text{compteur} - 1$
 - 12: $\text{compteur} \leftarrow \max(C_{it})$
 - 13: **jusqu'à que** $\text{compteur} = 0$
-

6.2. Gestion d'un réseau composé de sept biefs

Un réseau (voir figure 9), composé de 7 biefs, représentés par des carrés, et 14 points de transferts, représentés par des arcs, a été créé afin de tester notre approche. Le volume de chaque bief a été décomposé en 8 intervalles, 6 de taille 8 et 2 de taille infinie. Chaque point de transfert a été divisé en intervalles de taille 5, donnant 3, 4 ou 6 intervalles selon les capacités de transfert. La planification couvrira 8 pas de temps. De ce fait pour une approche centralisée, $|S| = 8^7 \times (8 + 1)$ et $|A| = 3^6 \times 4^2 \times 6^6$, la matrice creuse de la fonction de transition devrait stocker un maximum de $|S| \times |A| \times |S| \times \frac{1}{t} = 2.15 \times 10^{22}$ valeurs. Malheureusement, il n'a pas été possible de résoudre un scénario avec un unique agent afin de comparer les résultats optimaux avec les résultats de l'approche distribuée.

Plusieurs décompositions en agents ont été proposées grâce à l'utilisation d'un algorithme heuristique stochastique visant à minimiser la fonction de transition de chaque agent. Ces différentes décomposition ont permis d'observer l'impact de la décomposition sur les résultats. Pour comparer les politiques obtenues pour chaque décomposition de chaque scénario, plusieurs critères ont été observés : le temps de calcul de la politique jointe, la taille globale de la fonction de transition (somme du nombre de valeurs stockées dans la fonction de transition de chaque agent), pourcentage de fois qu'un bief se trouve hors de son rectangle de navigation (out) et pourcentage d'écart moyen des biefs au>NNL relativement aux LNL et HNL (avg). Les deux dernières valeurs sont des moyennes sur 50 000 exécutions d'une même simulation ou le niveau initial de chaque bief correspond à son>NNL.

Similairement à la version centralisée, la simulation d'une politique s'effectue en choisissant une valeur aléatoire dans les intervalles de l'action donnée par la politique pour l'état actuel. Le calcul des politiques jointes optimales s'est déroulé sur un cluster en utilisant le framework JADE (Bellifemine *et al.*, 1999) pour le développement des agents et de leur protocole.

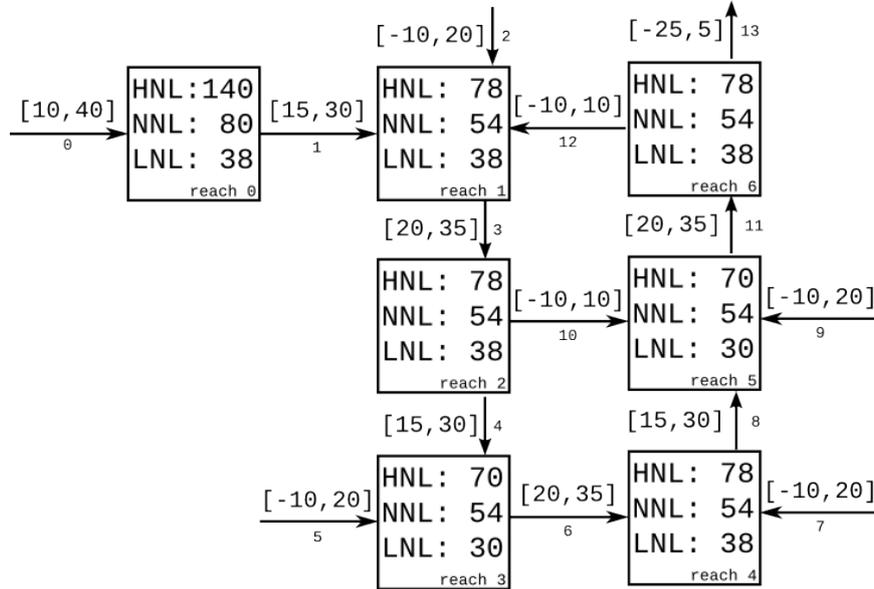


Figure 9. Scénario à 7 biefs

Notons au passage que la taille de l'intervalle de discrétisation des actions correspond approximativement à 20 % de l'écart entre le niveau idéal de navigation et ses bornes supérieure et inférieure, tandis que l'intervalle de discrétisation des états correspond à 33 %. Cela signifie que changer l'intervalle d'une action rapprochera ou éloignera de 20 % (relativement aux LNL et HNL) les biefs concernés de leur NNL.

Tableau 1. Résultats de l'approche distribuée altruiste

agents	taille (nb valeurs)	temps (s)	out (%)	avg (%)
6	$5,799\,936 \times 10^6$	605	0,000	17,39
7	$2,850\,816 \times 10^6$	446	0,000	18,98
8	$4,751\,36 \times 10^5$	84	0,000	17,86
9	$4,345\,60 \times 10^5$	50	0,000	17,73
10	$3,530\,24 \times 10^5$	33	0,011	18,30
11	$2,919\,68 \times 10^5$	38	0,043	18,91
12	$2,309\,12 \times 10^5$	38	0,021	17,10
13	$1,698\,56 \times 10^5$	35	0,006	17,28
14	$1,088\,00 \times 10^5$	23	0,001	16,28

Les résultats des tableaux 1 et 2 montrent que les politiques générées ne conduisent que rarement à la sortie du rectangle de navigation, ceci en choisissant de manière aléatoire les volumes à transférer dans chaque intervalle obtenu par la politique. Cet aspect aléatoire dans l'application des politiques implique parfois des choix incohérents par rapport à l'application, tel que choisir de transférer le maximum d'eau alors

Tableau 2. Résultats de l'approche distribuée égocentrique

agents	taille (nb valeurs)	temps (s)	out (%)	avg (%)
6	$5,799\,936 \times 10^6$	591	0,000	17,13
7	$2,850\,816 \times 10^6$	167	0,001	15,38
8	$4,751\,36 \times 10^5$	68	0,000	16,45
9	$4,345\,60 \times 10^5$	35	0,000	17,27
10	$3,530\,24 \times 10^5$	32	0,021	17,16
11	$2,919\,68 \times 10^5$	30	0,023	17,83
12	$2,309\,12 \times 10^5$	31	0,016	17,27
13	$1,698\,56 \times 10^5$	30	0,009	16,64
14	$1,088\,00 \times 10^5$	26	0,007	15,57

que le bief concerné est déjà à la limite de son rectangle de navigation. Cela semble concerner la majorité des simulations sortant du rectangle de navigation. De plus, le niveau de chaque bief reste relativement proche de son>NNL, avec un écart moyen inférieur à 20 % de la distance entre le>NNL et ses bornes, sachant que pour ces exemples 20 % de distance correspond à la taille d'un intervalle d'action. Une réduction attendue de la mémoire utilisée et du temps de calcul est observé lorsque le nombre d'agent augmente.

Contrairement, à ce qui pourrait être attendu, augmenter le nombre d'agents dans le réseau ne diminue pas forcément la qualité de la politique obtenue. Plusieurs hypothèses pourraient l'expliquer. Lorsque le nombre d'agent augmente, la vision des agents diminue, et donc d'un point de vue global la qualité des politiques calculées devrait diminuer. Cependant, l'augmentation du nombre d'agents réduit la part d'inconnue sur le réseau contrôlé de chaque voisin, ce qui réduirait la quantité d'incertitudes sur les politiques et fonctions de récompense échangés. Cette réduction des incertitudes permettrait d'améliorer la qualité des politiques calculées. Enfin, chaque scénario ne possède pas forcément de partition de qualité pour chaque taille de partition.

L'approche égocentrique (voir tableau 2) semble donner de meilleurs résultats, particulièrement au niveau des pourcentages d'écart des biefs au>NNL. Cela pourrait s'expliquer par la différence de la taille du système que chaque agent veut optimiser ainsi que la quantité d'approximations utilisées pour le calcul. Chaque agent égocentrique ne cherche qu'à optimiser les biefs qu'il affecte en se basant sur des approximations sur les actions de ses voisins. À l'inverse, un agent altruiste essaye d'optimiser l'ensemble des biefs affectés par les agents de son voisinage en utilisant les approximations des actions de ses voisins et des niveaux des biefs, via les fonctions de récompense. Les approximations sur l'état des biefs inconnus de l'agent apportées par les fonctions de récompenses partielles transmises par les agents altruistes ne semblent donc pas permettre d'optimiser les actions d'un agent pour l'ensemble de son voisinage.

En conclusion, les résultats obtenus par ces tests sur ce scénario réaliste sont très encourageants. Les biefs restent globalement proche de leur NNL et ne sortent qu'exceptionnellement de leur rectangle de navigation. Hélas, il n'a pas été possible de comparer les résultats distribués à ceux optimaux de l'approche centralisée à cause de l'incapacité de résoudre un problème de taille aussi importante. L'approche égocentrique semble donner de meilleurs résultats en un temps de calcul plus réduit grâce à la réduction du nombre de calculs et de communications due à l'absence d'utilisation de fonctions de récompense approximées des voisins.

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 risquent d'augmenter à cause du changement climatique.

En utilisant des MDP, 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. Plusieurs pistes permettraient alors de contourner cette limitation : les approches factorisées, décomposées et distribuées des MDP. Les approches factorisées et distribuées tirent avantage de la corrélation des variables dans la définition des états et des actions.

Ainsi une approche basée sur les MDP distribués est proposée, avec une modélisation orientée agents du réseau de voies navigables, où un agent est défini comme un ensemble de points de transfert. Cette approche permet de diminuer fortement les ressources nécessaires pour le calcul de la politique globale ainsi que le temps requis pour le calcul. La qualité des résultats et la convergence ne sont, a priori, pas garanties. Néanmoins, les premiers résultats obtenus sont encourageants. Les résultats seraient dépendants de la décomposition en agents. Le choix de la décomposition reste un problème difficile puisque minimiser le nombre d'agents, en conservant une capacité de résolution, ne garantit pas une amélioration de la politique.

Les travaux futurs basés sur cette approche consisteront à comparer différentes méthodes de partitionnement en agents, avec différentes heuristiques dans le but de déterminer les approches permettant de minimiser l'espace mémoire requis, le temps de calcul et l'écart à la politique jointe optimale. Cet étude sur le partitionnement sera réalisée en parallèle avec une détermination de règles de discrétisation pertinentes. L'utilisation de l'algorithme distribué sur des données réelles permettra de comparer les résultats obtenus à ceux proposés par des experts des réseaux de voies navigables. L'impact de la politique initiale sur la politique finale sera étudié et plusieurs méthodes de calcul de politique initiale seront comparées. Finalement afin de réduire la taille des intervalles d'action de la politique finale, un chaînage de résolution sera envisagé. Effectuer une nouvelle résolution en ne gardant de l'ensemble d'actions initial que les actions permises par la dernière politique calculée permettrait ainsi une discrétisation plus fine des actions.

D'autres approches pour la modélisation du réseau, tel que les réseaux de Petri ou les graphes de flot, pourraient être étudiées et comparées à l'approche actuelle dans le futur.

Bibliographie

- Arkel B., Darch G. (2006). Impact of climate change on London's transport network. *Proceedings of the ICE - Municipal Engineer*, vol. 159, p. 231-237.
- Bates B., Kundzewicz Z., Wu S., Palutikof J. (2008). Climate change and water. *Technical report, Intergovernmental Panel on Climate Change, Geneva*.
- Bellifemine F., Poggi A., Rimassa G. (1999). Jade—a fipa-compliant agent framework. In *Proceedings of paam*, vol. 99, p. 33.
- Bellman R. (1957). A Markovian Decision Process. *Journal of Mathematics and Mechanics*, vol. 6, n° 4, p. 679–684.
- Beuthe M., Jourquin B., Urbain N., Bruinsma F., Lingemann I., Ubbels B. *et al.* (2012). Estimating the impacts of water depth and new infrastructures on transport by inland navigation: A multimodal approach for the rhine corridor. *Procedia - Social and Behavioral Sciences - Proceedings of EWGT2012 - 15th Meeting of the EURO Working Group on Transportation*, vol. 54, p. 387 - 401.
- Beuthe M., Jourquin B., Urbain N., Lingemann I., Ubbels B. (2014). Climate change impacts on transport on the rhine and danube: A multimodal approach. *Transportation Research Part D: Transport and Environment*, vol. 27, p. 6 - 11.
- Bichot C.-E., Siarry P. (2011). *Graph Partitioning*. Wiley-ISTE.
- Boutillier C., Dean T., Hanks S. (1999). Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, vol. 11, p. 1–94.
- Boutillier C., Dearden R., Goldszmidt M., others. (1995). Exploiting structure in policy construction. In *IJCAI*, vol. 14, p. 1104–1113.
- Brand C., Tran M., Anable J. (2012a). The UK transport carbon model: An integrated life cycle approach to explore low carbon futures. *Energy Policy*, vol. 41, p. 107-124.
- Brand C., Tran M., Anable J. (2012b). The UK transport carbon model: An integrated life cycle approach to explore low carbon futures. *Energy Policy*, vol. 41, p. 107–124. Consulté sur <http://www.sciencedirect.com/science/article/pii/S0301421510006348>
- Chades I., Scherrer B., Charpillet F. (2002). A Heuristic Approach for Solving Decentralized-POMDP: Assessment on the Pursuit Problem. In *SAC '02: Proceedings of the 2002 ACM symposium on Applied computing*, p. 57–62. Madrid, Spain, ACM.
- Dean T., Lin S. hong. (1995). Decomposition techniques for planning in stochastic domains. In *In proceedings of the fourteenth international joint conference on artificial intelligence (ijcai-95)*, p. 1121–1127. Morgan Kaufmann.
- Duviella E., Nouasse H., Doniec A., Chuquet K. (2016). Dynamic optimization approaches for resource allocation planning in inland navigation networks. *ETFA'2016, Berlin, Allemagne, September 6-9*.

- EnviCom. (2008). Climate change and navigation - waterborne transport, ports and waterways: A review of climate change drivers, impacts, responses and mitigation. *EnviCom - Task Group 3*.
- IPCC. (2014). Climate change 2014. *The Core Writing Team, R. K. Pachauri and L. Meyer, Synthesis Report*. Consulté sur https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full.pdf
- IWAC. (2009). Climate change mitigation and adaptation. implications for inland waterways in england and wales. *Report*.
- Kocsis L., Szepesvári C. (2006). Bandit Based Monte-Carlo Planning. In J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.), *Machine Learning: ECML 2006*, vol. 4212, p. 282–293. Springer Berlin Heidelberg.
- Lozenguez G., Adouane L., Beynier A., Mouaddib A.-I., Martinet P. (2015). Punctual versus continuous auction coordination for multi-robot and multi-task topological navigation. *Autonomous Robots*, p. 1–15.
- Mallidis I., Dekker R., Vlachos D. (2012). The impact of greening on supply chain design and cost: a case for a developing region. *Journal of Transport Geography*, vol. 22, p. 118–128. Consulté sur <http://www.sciencedirect.com/science/article/pii/S096669231200004X>
- Mihic S., Golusin M., Mihajlovic M. (2011). Policy and promotion of sustainable inland waterway transport in Europe – Danube River. *Renewable and Sustainable Energy Reviews*, vol. 15, n° 4, p. 1801–1809. Consulté sur <http://www.sciencedirect.com/science/article/pii/S1364032110004028>
- Nair R., Varakantham P., Tambe M., Yokoo M. (2005). Networked Distributed POMDPs: A Synthesis of Distributed Constraint Optimization and POMDPs. In *National Conference on Artificial Intelligence*, p. 7.
- Nouasse H., Doniec A., Duviella E., Chuquet K. (2016). Efficient management of inland navigation reaches equipped with lift pumps in a climate change context. *4th IAHR Europe Congress, Liege, Belgium 27-29 July*.
- Nouasse H., Rajaoarisoa L., Doniec A., Duviella E., Chuquet K., Chiron P. *et al.* (2015). 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*, p. 1–6. IEEE.
- Pachauri R. K., Allen M., Barros V., Broome J., Cramer W., Christ R. *et al.* (2014). 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.
- Parr R. (1998). Flexible Decomposition Algorithms for Weakly Coupled Markov Decision Problems. In *14th Conference on Uncertainty in Artificial Intelligence*, p. 422–430.
- Puterman M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc.
- Sabbadin R. (2002). Graph partitioning techniques for Markov Decision Processes decomposition. In *15th European Conference on Artificial Intelligence*, p. 670–674.
- Wanders N., Wada Y. (2015). Human and climate impacts on the 21st century hydrological drought. *Journal of Hydrology*, vol. 526, p. 208–220.

