



# Planning Large Systems with MDPs: Case Study of Inland Waterways Supervision

Guillaume Desquesnes<sup>a,b</sup>, Guillaume Lozenguez<sup>a,b</sup>, Arnaud Doniec<sup>a,b</sup>, and Éric Duviella<sup>a,b</sup>

<sup>a</sup>Mines Douai, IA, F-59508 Douai, France

<sup>b</sup>Univ. Lille, F-59500 Lille, France

{guillaume.desquesnes,guillaume.lozenguez,arnaud.doniec,eric.duviella}@mines-douai.fr

## KEYWORD

Markov decision processes; Inland waterway network; Large-scale systems

## ABSTRACT

*Inland waterway management is likely to go through heavy changes due to an expected traffic increase in a context of climate change. Those changes will require an adaptive and resilient management of the water resource. The aim is to have an optimal plan for the distribution of the water resource on the whole inland waterway network, while taking into account the uncertainties arising from the operations of such a network. A representative model using Markov decision processes is proposed to model the dynamic and the uncertainties of the waterways. The proposed model is able to coordinate multiple entities over multiple time steps in order to prevent an overflow of a test network. However, this model suffers from a lack of scalability and is unable to represent real case applications. Advantages and limitations of several approaches of the literature to circumvent this limitation are discussed according to our case study.*

## 1. Introduction

Climate change is a major concern in the modern society. It has been emphasized by the agreement of the COP21 signed in Paris the 12th of December 2015. The impact of climate change on the inland waterway network and traffic has been widely studied (EnviCom, 2008; IWAC, 2009; Arkell and Darch, 2006; Bates et al., 2008; Wanders and Wada, 2015; Beuthe et al., 2014; Pachauri et al., 2014). The general agreement is that the intensity and occurrence of flood and drought periods will increase. Furthermore, the fluvial traffic is expected to grow by 35% by 2050 (Beuthe et al., 2012).

Designing strategies to manage the inland waterway network that take into account those new constraints is a priority for the inland water managers. Inland waterway networks are large scale systems built by humans to responds to their needs. They have strong interactions with their natural surroundings. Under the hypothesis of a complete knowledge of these interactions, approaches of adaptive management of inland waterway network in a context of climate changes have been proposed in (Nouasse et al., 2015; Nouasse et al., 2016) using constraints satisfaction problems or quadratic optimization techniques (Duviella et al., 2016). However most of those interactions are only partially known. They are illegal rejects, exchange with groundwater tables, influence of local weather phenomena, ... The management of such networks is subject to numerous uncertainties. For these reasons, an approach based on a stochastic modeling seems more suitable.

The objective is to design an adaptive and resilient management strategies for the inland waterway network, in a context of climate change and traffic increase. In this article, a stochastic modeling of the inland waterway network based on Markov Decision Process (MDP) is proposed to optimize the planning of the water resource distribution across all the networks while taking into account all uncertainties inherent to their functioning. MDPs are widely used in the planning of stochastic models. They can model the evolution of uncertain systems and lead



to optimal planning for all possible configurations of the model. However, their use is hampered by resolution difficulties, especially due to the size of the search space. In a majority of applications, they induce intractable models. Obtaining the optimal control policy requires specific resolution algorithms (Boutillier et al., 1999; Nair et al., 2005).

In this application, the complexity comes from the size of the inland waterway network and the numerous possibility of water displacement. Indeed, it is required to consider the set of joint actions (water displacement) and the set of possible configurations of the inland waterway network. This induces a combinatorial explosion of the number of states. This paper aims firstly to describe the modeling of inland waterway network with MDP, which to the best of our knowledge has never been used in the context of fluvial transport. Then, the limitations of such a modeling are shown and different approaches to circumvent those limitations are discussed.

The problematic of managing an inland waterway network in a context of climate change and traffic increased is presented in section 2. A naive modeling of the network using MDP is proposed in section 3. In section 4, the study of two reaches will help illustrate the proposed approaches and its limitations. Finally, section 5 is dedicated to the description and comparison of various MDPs derivatives to eliminate those limitations.

## 2. Inland waterway network management

An inland waterway network (*see* Figure 1a) is a large-scale system, mostly used for navigation. It provides both economic and environmental benefits (Mallidis et al., 2012; Mihic et al., 2011) while providing quiet, efficient and safe transports of goods (Brand et al., 2012). It is mostly composed of interconnected canalized rivers and artificial channels, divided by locks. Any part of a river or channel separated by at least two locks is a navigational reach. For the sake of simplicity, a *navigational reach* will be called *reach* on the rest of this paper.

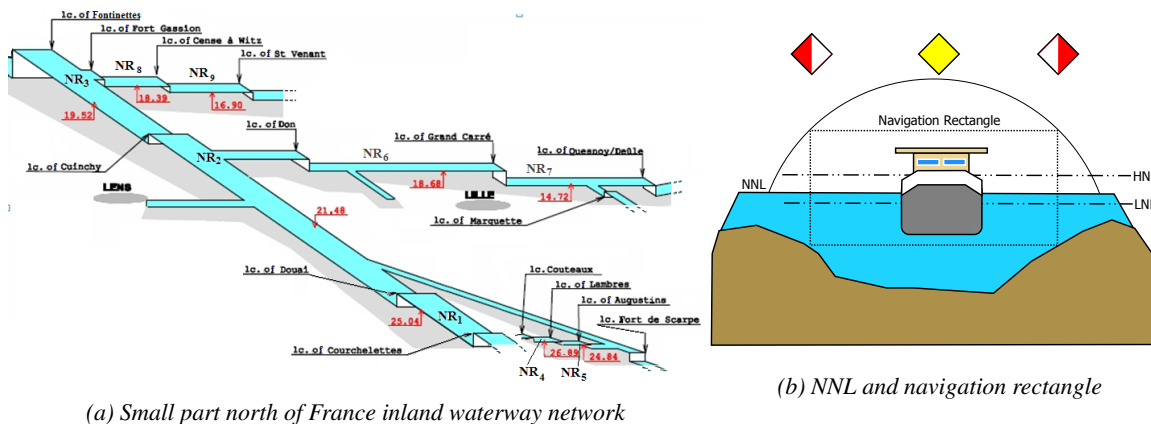


Figure 1: Figure representing waterway and navigation rectangles.

The primary objective of the waterway manager is to maintain a respectable water level in all reaches to allow navigation. This level must respect the conditions defined by the navigation rectangle (*see* Figure 1b) and be as close as possible from the Normal Navigation Level (NNL). The lower and upper bounds of the navigation rectangle are called respectively Lower Navigation Level (LNL) and Higher Navigation Level (HNL).

In normal situations, having a boat crossing a lock is the main disturbance of the water level, since using a lock drains water from the upstream reach towards the downstream reach. Furthermore, the water level is also affected by ground exchanges, natural rivers, weather and other unknown factors, like illegal discharges. Locks

are not dedicated to control the water level. Structures, such as gates or dams are used to send water downstream and when available pumps can be used to send some upstream. Those are the main structures used to displace the water resource between the reaches of the network.

Navigation is only allowed during the daytime, with few exceptions, notably on Sunday. Reaches management is based on human expertise gathered over time. However, the current management strategies should be heavily impacted by new constraints due to climate change and traffic increase. The goal is to anticipate the impact of those constraints by designing adaptive and resilient management approaches to ensure in each point of the network and at each instant the navigation requirements. It requires to determine a global planning for the water distribution on the whole network by taking into account the uncertainties of climate events and of the navigation demand. Planning the allocation of the water resource with a horizon of several time steps allows better anticipation of possible events. Information on the current state of the waterway network can be collected in real time through a network of level sensors equipping the reaches.

### 3. Markov Decision Process

#### 3.1 Definition

Markov Decision Process (MDP) is a generic framework modeling control possibility of stochastic dynamic system as probabilistic automaton. The framework is well adapted to the waterway network supervision since the state of the network is fully observable (in terms of water volumes) and the control is uncertain due to uncontrolled water transit.

A MDP is defined as a tuple  $\langle S, A, T, R \rangle$  with  $S$  and  $A$  respectively, the state and the action sets that define the system and its control possibilities.  $T$  is the transition function defined as  $T : S \times A \times S \rightarrow [0, 1]$ .  $T(s, a, s')$  is the probability to reach the state  $s'$  from  $s$  by doing action  $a \in A$ . The reward function  $R$  is defined as  $R : S \times A \times S \rightarrow \mathbb{R}$ ,  $R(s, a, s')$  gives the reward obtained by attaining  $s'$  after executing  $a$  from  $s$ .

A policy function  $\pi : S \rightarrow A$  assigns an action to each system state. Optimally solving a MDP consists in searching an optimal policy  $\pi^*$  that maximizes the expected reward.  $\pi^*$  maximizes the value function of Bellman equation (Bellman, 1957) defined on each state:

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

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

The discount factor  $\gamma \in [0, 1]$  balances the importance between future and immediate rewards. With  $\gamma$  set close to 0 the immediate action with the maximal reward will be taken while a  $\gamma$  close to 1 will permit the control to accept penalties if future gains could be important. Multiple algorithms exist to solve optimally a MDP, a notable version is *Value Iteration* (Puterman, 1994). It constructs iteratively the value function  $V$ , using equation 3, until convergence or until a specified number of iterations has been done.

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

The last  $V_i$  obtained is then used to generate the optimal policy with equation 2. The first value function  $V_0$  is initialized at 0.

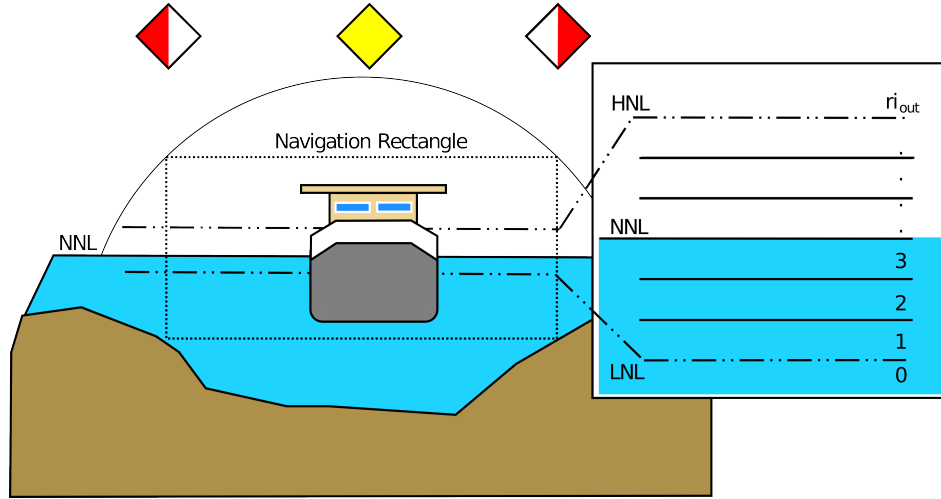


Figure 2: Discretization of a reach water volume in intervals.

### 3.2 Application to the inland waterway management

The objective is to plan the best course of actions for the whole network over  $\tau$  time steps, knowing that some conditions may differ on each time step and can affect the inland navigation. For instance, the weather might become rainy increasing the water volume of affected reaches; an unexpected fluvial traffic increase on some reaches would imply a greater locks usage and an increase in water transfers.

Periods of 12 hours are used as time steps, they model the active navigation periods during the day and the inactive at night. Using large time steps allows to consider the water level to be uniform on a reach and to smooth the uncertainties on both the traffic and other temporal variations.

#### 3.2.1 Dividing in states and actions

A state of the model is defined as an assignation of volume for all the reaches in the network at a given time steps. Similarly an action is defined as an assignation of volume for each controlled transfer point (lock, pump or gate). The MDP formalism requires discrete set of states and actions. Since the volumes observed (obtained from level measures) and transferred between reaches are continuous, they had to be discretized in intervals. Thus the operating range of each reach is divided in regular intervals with the exception of the first and last interval. Those two intervals represent all values outside of the navigation rectangle and are considered to be of infinite size. Transfer points use an interval partitioning similar to the one used for reaches. However, the volumes transferred are considered to be fully controllable, they do not have intervals of infinite size.

More formally, the set of states  $S$  of the model is defined as the combination of all possible intervals of each reach at all time steps. For a reach  $i$ , the intervals result from a regular discretization of volumes (*see* Figure 2). Interval 0 represents all volumes under the LNL and interval  $ri_{out}$  corresponds to all volume over the HNL.

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

where  $N$  represents the number of reaches in the network. Having the time step in the state allows to express temporal probabilities on uncontrolled and unknown inputs and outputs. Finally, a time step is added to represent

the end of the planning. All states of this last time step are absorbing:  $\forall a, T(s, a, s) = 1$ , with a null reward:  $\forall a, R(s, a, s) = 0$ .

Similarly, the set of actions  $A$  is defined as the combination of intervals of volume of all transfer points. An action is like a set of subactions performed in a distributed way on the network. Actions are time independent, so:

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

where  $A_{i,j}$  represent the set of all possible volume intervals of the transfer point linking reach  $i$  to reach  $j$ . Reach 0 represents external rivers and any other external element able to bring or drain water from any managed reach.

The volume transferred from reach  $i$  to reach  $j$  under action  $a \in A$  is noted  $a_{i,j} \in A_{i,j}$ . In fact, there is only a limited number of transfer points. In all transfer points  $A_{i,j}$  between two unconnected reaches  $i$  and  $j$ , no transfer is possible ( $A_{i,j} = \{0\}$ ). To simplify the notation,  $a_i$  defines the part of the action affecting the reach  $i$  as

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

Two operators  $\oplus$  and  $\ominus$  on  $(\mathbb{R} \cup \mathcal{I})^2 \rightarrow \mathbb{R}$  are defined. They can respectively add and subtract intervals of numbers and/or numbers, the result being always a real.  $\mathcal{I}$  being the set of all possible intervals of the network. Those operators are respectively a simple addition and subtraction using the value of the member if it is a real, or its average value if it is an interval.

### 3.2.2 Transition function

The transition function  $T(s, a, s')$  represent the probability to reach the state  $s'$  by doing action  $a$  from the state  $s$  while taking into account the possible temporal variation. It results from discussions with experts of the north of France waterway network. Trivially, for  $s$  at time step  $t$  and  $s'$  at time step  $t'$ ,  $s'$  is only reachable if  $t' = t + 1$ . In this transition function, the state of a reach is independent from the other reaches. Indeed, it only depends on the incoming and outgoing water volumes. To define formally the transition function, it is required to introduce the different uncertainties affecting the network.

A first source of uncertainties on the transition function comes from uncontrolled, like the navigation, or unknown water displacements. Those are modeled by temporal variations. Temporal variations are local changes on one or more reaches during one or more time steps with a certain probability and intensity. For example, rain on a reach during two time steps is a temporal variation. Since temporal variations are not actions, they are only represented in the transition function. The set of all possible combinations of temporal variations affecting the network at a time step  $t$  is denoted  $V_t$ . If the network is affected by the variation  $v_t \in V_t$  then reach  $i$  is affected by  $v_t(i)$ . The probability of such a variation happening is  $P(v_t)$ .

The second source of uncertainties results from the discretization in intervals of volume transferred and contained in reaches. This discretization causes an approximation in the state representation. Those uncertainties are reflected in  $P(r_{i,s'} | r_{i,s}, a_i, v_t(i))$ . This probability function represent the probability of the reach  $i$  to be in interval  $r_{i,s'}$  after doing the action  $a_i$  in the interval  $r_{i,s}$  under the variation  $v_t(i)$ .

$$P(r_{i,s'} | r_{i,s}, a_i, v_t(i)) = \begin{cases} p^-(r_{i,s}, a_i, v_t(i)) & \text{if } r_{i,s} \oplus a_i + v_t(i) \in r_{i,s'} \\ p^+(r_{i,s}, a_i, v_t(i)) & \text{if } r_{i,s} \oplus a_i + v_t(i) \in r_{i,s'} - 1 \\ p^-(r_{i,s}, a_i, v_t(i)) & \text{if } r_{i,s} \oplus a_i + v_t(i) \in r_{i,s'} + 1 \\ 0 & \text{else} \end{cases} \quad (6)$$

where  $p^-$  is the probability to get the expected volumes by taking into account intervals approximations and,  $p^+$  (resp.  $p^-$ ) is the probability to get the interval corresponding to a superior (resp. inferior) water level, with respect to  $p^+ + p^- + p^- = 1$ . Obtaining  $p^+$ ,  $p^-$  and  $p^-$  will be discussed in the next section.

The transition function states the product of the two sources of uncertainty:

$$T(s, a, s') = \sum_{v_t \in V_t} P(v_t) \times \prod_{i=1}^N P(ri_{s'} | ri_s, a_i + v_t(i))$$

### 3.2.3 Dealing with the discretization uncertainties

To deal with the discretization three probability functions  $p^-$ ,  $p^+$  and  $p^-$  (see relation 6) are defined.  $p^-$  represent the probability of reaching the expected interval after doing a chosen action  $a$  in current interval of a reach  $ri_s$  under a specific temporal variation  $v_t$ .  $p^-$  (resp.  $p^-$ ) represents respectively the probability to be in the interval just inferior (resp. superior) to the expected interval. The expected interval is defined as

$$ri_{ex}, \text{ with respect to } ri_s \oplus a_i + v_t(i) \in ri_{ex}$$

It is possible to define two functions  $w_i^+$  and  $w_i^-$  representing respectively the maximum and minimum volume balance (incoming water minus outgoing water), for a given action and temporal variation, that can affect reach  $i$ .

$$w_i^+(a, v_t) = v_t(i) + \sum_{j=1}^N (a_{ij}^+ \ominus a_{ji}^-)$$

$$w_i^-(a, v_t) = v_t(i) + \sum_{j=1}^N (a_{ij}^- \ominus a_{ji}^+)$$

where  $a_{ij}^+$  (resp.  $a_{ij}^-$ ) is the highest (resp. lowest) value of the interval of the subaction  $a_{ij}$ .

To simplify the notations, the following simplifications will be used:

$$w_i^+ = w_i^+(a, v_t) \quad w_i^- = w_i^-(a, v_t)$$

Two functions modeling the maximum and minimum water volume that can be achieved at the next time step are defined as:

$$\begin{aligned} \max : ri_s &\rightarrow \mathbb{R}^+ & \min : ri_s &\rightarrow \mathbb{R}^+ \\ \max(x) &= x + w_i^+ & \min(x) &= x + w_i^- \end{aligned}$$

The maximum (resp. minimum) volume achievable at the next time step for the state  $s$  when doing action  $a$  under the variation  $v_t$  is represented by the line segment  $[k, l]$  (resp.  $[n, m]$ ) with:

$$\begin{aligned} k &= (ri_s^-, \max(ri_s^-)) & l &= (ri_s^+, \max(ri_s^+)) \\ n &= (ri_s^-, \min(ri_s^-)) & m &= (ri_s^+, \min(ri_s^+)) \end{aligned}$$

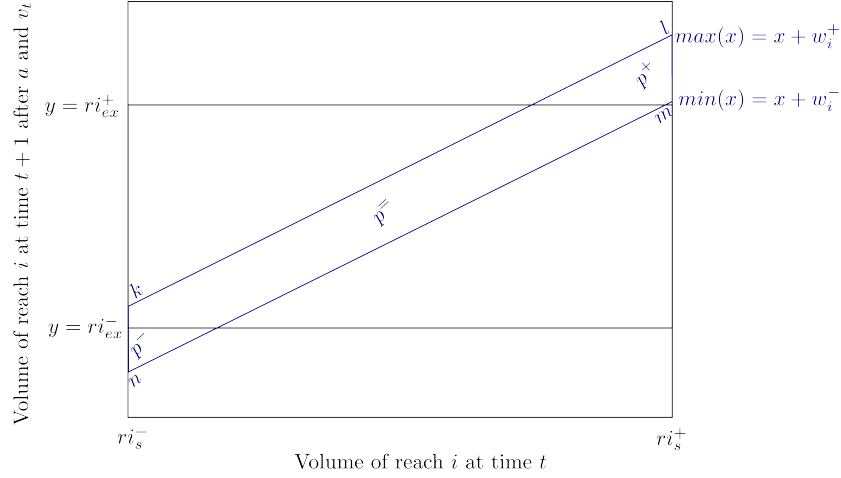


Figure 3: Construction of  $p^-$ ,  $p^=$  and  $p^+$ .

The parallelogram  $klmn$  (see Figure 3) represents all the possible water volume of reach  $i$  at the next time step depending on its current volume. Under the assumption that all possible volumes transferable by the action are equiprobable, the probability  $p^=$  of being in the interval  $ri_{ex}$  after doing action  $a$  under variation  $v_t$  from interval  $ri_s$  is the percentage of the area of  $klmn$  between the lines  $y = ri_{ex}^+$  and  $y = ri_{ex}^-$ . Similarly the probability  $p^+$  is equivalent to the percentage of area over  $y = ri_{ex}^+$  while  $p^-$  correspond to the percentage under  $y = ri_{ex}^-$ .

The area  $\mathcal{A}$  of  $klmn$  can be divided in three part  $\mathcal{A}^-$ ,  $\mathcal{A}^=$  and  $\mathcal{A}^+$  with respect to  $\mathcal{A} = \mathcal{A}^- + \mathcal{A}^= + \mathcal{A}^+$  where:

- $\mathcal{A}^-$  is the area of  $klmn$  under  $y = ri_{ex}^-$
- $\mathcal{A}^=$  is the area between  $y = ri_{ex}^-$  and  $y = ri_{ex}^+$
- $\mathcal{A}^+$  is the area of  $klmn$  over  $y = ri_{ex}^+$

It is possible to define  $p^-$ ,  $p^=$  and  $p^+$  from those areas as :

$$p^-(ri_s, a, v_t) = \frac{\mathcal{A}^-}{\mathcal{A}} \quad p^=(ri_s, a, v_t) = \frac{\mathcal{A}^=}{\mathcal{A}} \quad p^+(ri_s, a, v_t) = \frac{\mathcal{A}^+}{\mathcal{A}}$$

Using basic geometry:

$$\mathcal{A} = (ri_s^+ - ri_s^-) \times (w_i^+ - w_i^-)$$

As partially shown by Figure 3, the parts of  $klmn$  not in  $ri_{ex}$  is either shaped as a triangle or as a trapeze depending on the points location. Thus:

$$\mathcal{A}^+ = \begin{cases} 0.5 \times ((w_i^+ - ri_{ex}^+ + ri_s^+)^2 - (w_i^- - ri_{ex}^- + ri_s^-)^2) & \text{if } \min(ri_s^+) > ri_{ex}^+ \\ 0.5 \times (w_i^+ - ri_{ex}^+ + ri_s^+)^2 & \text{if } \max(ri_s^+) > ri_{ex}^+ \wedge \min(ri_s^+) \leq ri_{ex}^+ \\ 0 & \text{else} \end{cases}$$

$$\mathcal{A}^- = \begin{cases} 0.5 \times ((w_i^- - ri_{ex}^- + ri_s^-)^2 - (w_i^+ - ri_{ex}^+ + ri_s^+)^2) & \text{if } \max(ri_s^-) < ri_{ex}^- \\ 0.5 \times (w_i^- - ri_{ex}^- + ri_s^-)^2 & \text{if } \max(ri_s^-) \geq ri_{ex}^- \wedge \min(ri_s^-) < ri_{ex}^- \\ 0 & \text{else} \end{cases}$$

Since, the probability to reach the second interval over or under the expected interval is 0, this has to be reflected by the discretization. Using the previous notation:

$$\forall s, a, v_t \quad ri_s^+ + w_i^+(a, v_t) \leq ri_{ex+1}^+ \quad ri_s^- + w_i^-(a, v_t) \geq ri_{ex-1}^-$$

where  $ri_{ex+1}$  is the interval over  $ri_{ex}$  and  $ri_{ex-1}$  is the interval under  $ri_{ex}$ . This constraint aims to reduce the variations coming from the discretization by limiting the number of states reachable by a couple  $(s, a, v_t)$ .

### 3.2.4 Reward function

The aims of the planning is to maintain all reaches within their navigation rectangle while minimizing the distance to their NNL. This has been defined in cooperation with expert in the management of waterway networks and is reflected in the reward function defined as:

$$R(s, a, s') = f(a) - \sum_{i=1}^N \begin{cases} (NNL_i \ominus ri_{s'})^2 & \text{if } ri_{s'} \in Rec_i \\ g^2 & \text{if } ri_{s'} \cap Rec_i = \emptyset \\ (0.5 \times g)^2 & \text{if } ri_{s'} \cap Rec_i \neq \emptyset \text{ and } ri_{s'} \notin Rec_i \end{cases}$$

where  $Rec_i$  represents an interval of volume corresponding to the navigation rectangle of reach  $i$ .  $NNL_i$  is the volume corresponding to the NNL of reach  $i$ .  $f$  is a cost function modeling the cost of transferring water from each transfer point. This cost function will be specific for each transfer points and each network.  $g$  represents a penalty for halting the navigation when a reach leaves its navigation rectangle. Only half of the penalty is applied if the interval is not fully outside of the navigation rectangle.

## 4. Management of a network of two reaches

A realistic inland waterway network has been defined to test the proposed approach. It is composed of two reaches and six transfer points (see Figure 4).

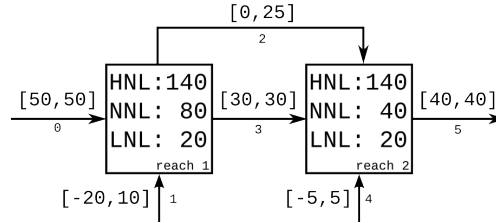


Figure 4: Waterway network.

### 4.1 Network characteristics

On Figure 4, reaches are represented by squares, with their navigation rectangle and NNL specified in units of volume. Arcs correspond to transfer point with a minimum and maximum transfer capacity. A negative value meaning this transfer point can be used to import and export water.

The possible volumes of both reaches are divided in 9 intervals, 7 of them with a fixed size of 20 units of volume. The first and last interval are considered of infinite size. The volumes of transfer points 1, 2 and 4 are divided in intervals of size 5 units of volume. This leads respectively to intervals of size 6, 5 and 2 units of



volume for those three transfer points. The remaining transfer points: 0, 3 and 5 represent locks. In this case, they transfer a constant amount of water corresponding to the navigation.

Operating scenarios of the studied network have been build for a period of 8 day and night periods, giving 16 time steps. In those scenarios, locks are not used during night periods, since no navigation is allowed. The penalty of being outside of the navigation rectangle has been set to an arbitrary large value  $g = 100000$ . To simplify the observation of the results, all transfer points have the same usage costs.

## 4.2 Operating scenarios

Several operating scenarios have been defined to test the proposed approach. The first scenario corresponds to an ideal scenario, the two reaches start at their NNL. In the second scenario, the first reach is initially near its LNL while the second reach is close to its HNL. For the third scenario, the initial conditions are swapped, with the first reach close to its HNL while the second reach is near its LNL. Finally, the fourth and last scenario is similar to the first scenario, with the two reaches starting close to their NNL. However, unlike the other scenarios, it is under probable perturbation. A perturbation is likely to affect the first reach and, if nothing is done beforehand would make the first reach overflow if it is close to its NNL. It is important to note that the following simulations consider continuous volumes, to be closer to real functioning of the network, while the planning uses discretized volumes.

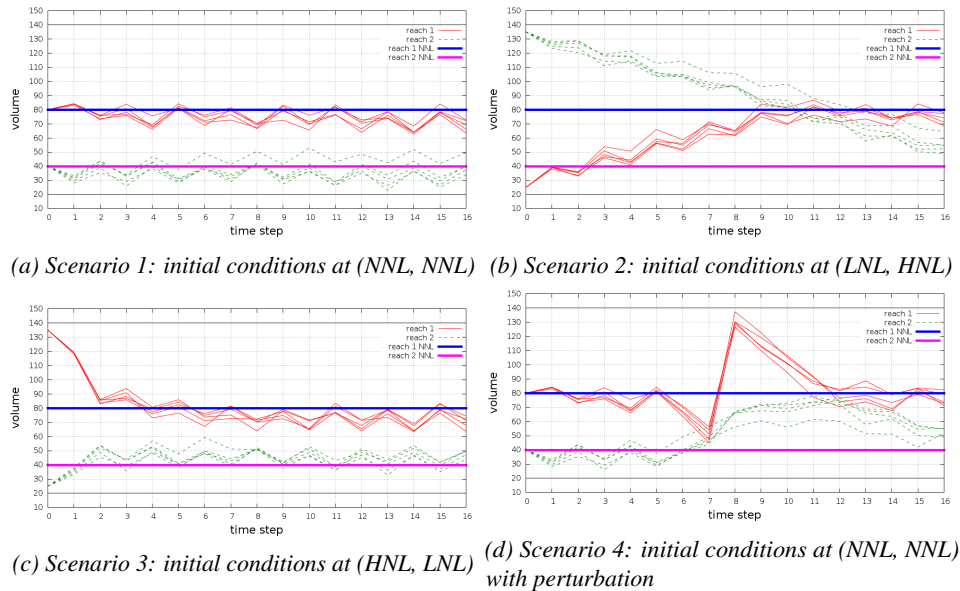


Figure 5: Different scenarios.

Since the policy actions are intervals, the scenarios have been simulated by randomly drawing a value from each interval of the action, instead of choosing the average or the best value of the interval. The goal is to have a better perception of the quality of the interval selected by the policy. Because the volumes transferred are chosen randomly, five simulations were made for each scenario. This help visualizing the consequences of the random selection of transferred volumes of the used policy. The results are shown on Figure 5a, 5b, 5c and 5d. Each figure shows the 5 simulations of a scenario.

An optimal planning of water transfer in the first scenario can maintain the NNL in both reaches. The policy resulting from the proposed approach is able to be relatively close to the real optimal planning. The differences are due to a broad discretization of the model in intervals. Smaller intervals would reduce the oscillation around the NNL. It is important to note that for this scenario, the policy produced resulting from the proposed approach is able to maintain the NNL of both reaches when choosing the best values of the action.

The reaction of the policy to some events that put the two reach on the limits of their navigation rectangle is shown on Figure 5b and 5c (respectively second and third scenario). In both case, the planning leads to both reaches being back to their NNL at the end of the time periods. The speed of the recovery is dependent on the configuration of the network. It is indeed easier for a reach to drain itself in the downstream reach when it needs to be filled than the opposite scenario.

In the last scenario, heavy rain might affect the first reach between time steps 7 and 8. Its magnitude is such that if the reach is close to its NNL before the event, the reach would overflow. The policy produced is able to anticipate this event. To prevent the overflow, the first reach drains itself in the second reach few time step before the rain, while preserving both navigation rectangles.

### 4.3 Limitations in scaling up

A naive implementation of the transition function consists in creating a  $|S|^2 \times |A|$  matrix.

Using relation 4 and relation 5, it is possible to write:

$$|S| = |\{0, \dots, 16\}| \times |[0, r1_{out}]| \times |[0, r2_{out}]| = 17 \times 9 \times 9 = 1377$$

$$|A| = \prod_{i,j \in [0,2]} |A_{i,j}| = 1 \times 6 \times 5 \times 1 \times 2 \times 1 = 60$$

So the transition function would contain  $1377^2 \times 60 = 1.137\,677\,40 \times 10^8$  values. Assuming 8 bytes to store each value, approximately 0.91 GB of memory space would be required to store the transition function.

However, most transitions between states are impossible, due to the temporality of the states. A state at time  $t$  can only reach states at time  $t + 1$ . For this kind of configuration, a sparse matrix could drastically reduce the number of stored values. A sparse transition matrix would store only the indices and the values of non-null transitions.

When computing the optimal policy of the MDP, several results were observed: the policies gave correct results, building the transition function is one of the longest operation and finally the algorithm converge really fast. Due to time being acyclic, choosing the action of state at the last time step depends only on the immediate reward, so the optimal actions for those states are found at the first iteration. Recursively, choosing the action of states at time step depends only on the choice of actions in the future ( $\forall t' > t$ ). So, the number of iterations to found the optimal policy is bound by the number of time steps, here 17. Since the number of time steps is bound, there is no need to use the discount factor  $\gamma$  (see equation 2) to ensure convergence.  $\gamma = 1$  can then be used for the best results.

As the size of the modeled network increases and/or the size of intervals decreases, the proposed approach rapidly encounter memory limitation problems. By construction, the number of states and actions grows exponentially in the number of reaches in the network. However, real applications contain a large amount of reaches. As an example, the inland waterway network in the north of France contains 50 reaches. Therefore the proposed approach could not be used. It would be physically impossible to build the set of states required by this approach.

## 5. Existing approaches dedicated to scaling

To bypass the difficulties due to spatial limitations, extensions of MDPs has been defined in the literature such as factored modeling or decomposing the MDP in sub-MDP. The approaches discussed are limited to those that plan over the entire state space. In this article, approximations such as Monte-Carlo tree search (Kocsis and Szepesvári, 2006) where the produced policy applies only to the most probable states are not presented.

### 5.1 Factored MDP

The factored MDP approach, introduced in (Boutilier et al., 1995), aims to represent in a compact way the transition and reward functions. For this purpose, states are represented by an assignation of variables. Every variables may have no influence on the value of a specific variable at the next time step. The idea behind factored MDPs is to explode the state space to group similar part of states in the transition and reward functions.

Algebraic decision diagrams are used in (Hoey et al., 1999) to represent in a more compact way transition, reward and value functions by capturing the regularity in the respective function.

Factored MDPs are used to solve multi-agent planning in (Guestrin et al., 2001). The value function is approximated as a linear combination of localized value functions. This method exploits both state and action space structure. It allows to solve problems with over  $10^{28}$  states and a  $10^9$  actions. Furthermore, agents need a coordination graph to decide their action on runtime. However, this approach assumes that agents interact only with a few of other agents. This limitation is bypassed by a rule-based approach introduced in (Guestrin et al., 2002). The rule-based approach does not require the rules to be mutually exclusive unlike tree or ADD representations.

In the case study, the state space (resp. action space) results in the cartesian product of the state space of each reach (resp. action space of each transfer point). Furthermore, the state of a reach usually does not depend directly on the state of nearby reaches, only their action matters. Indeed, when a reach receives water, the volume of the source is not used to determine its new level. This is due to the assumption that moving water is always possible, reaches are never full nor empty. Action leading to such states would be forbidden in the model. Reaches states are independent and it would be possible to exploit this property to factorize the MDP.

### 5.2 Decomposed MDP

The decomposition of a MDP permits to decrease the complexity of the policy computation by building a hierarchy between local problems and a global solution (Boutilier et al., 1999; Dean and hong Lin, 1995). It is particularly efficient in spatial problems as it is based on the topological aspect of transitions.

Decomposed MDP splits the state space into sub-MDPs in a way that the union of sub-MDPs cover the total MDP. Two approaches to solve decomposed MDP has been proposed in (Dean and hong Lin, 1995). The first approach consists in solving each sub-MDP iteratively until a stable point is reached. A stable point means that all the sub-policies are coherent among each other. This solution does not guarantee a speed-up of the policy computation but guarantee the optimality of the total policy (the union of sub-policies).

The second hierarchical approach consists in defining a set of parameters for each sub-MDPs in order to compute a finite collection of sub-policies. Then, a high level global MDPs is defined on “macro-states” to select the appropriates sub-policies to apply in order to have a coherent global behavior. This solution permits to control the policy computation but does not guarantee the optimality of the solution. The quality of the total policy takes benefit from rich parameter definition but, with a cost on the computation resources.

In most real problems, decomposed MDP could reduce policy computation (with or without optimal guarantee) but requires to compute a partition in the state space (Parr, 1998; Sabbadin, 2002). If there is no evident



decomposition, partitioning is a very hard (Bichot and Siarry, 2011) problem, and could penalize a decomposition approach.

A waterway MDP is not easily decomposable, even with each state representing a snapshot of the entire networks. However, an option exists in considering several levels of deterioration of navigation conditions. Each sub-MDP, matching a different deterioration level of the network, will produce a policy of supervision oriented towards the normal navigation conditions. For example, with a waterway MDP decomposed in three sub-MDP: normal, flood and drought, it is probable that the produced policy would keep the system in normal states (close to *NNL*) with sparse dependencies between the three sub-MDPs. This way, solving first the normal sub-MDP and then the two deteriorated sub-MDP should speedup the policy computation. However, decomposition will not impact the transition function construction and storage.

### 5.3 Distributed MDP

Distributed MDPs appear as an *ad hoc* solution to solve cooperative problems on agent based modeling. Such an approach is developed to solve decentralized MDP (Chades et al., 2002; Nair et al., 2005). This is a framework where the policy has to be distributed over agents and performed in a decentralized way. Each agent is responsible in computing its own policy considering its part of a mission. Communication protocols allow the agents to adapt their policies to reach a common interest. Distributed MDP is used for example in robotic missions to deal with traveling salesmen coordination (Lozenguez et al., 2015).

This approach combines factorization and decomposition ideas. The total MDP is split in several sub-MDPs by partitioning the set of variables defining the states and actions. Each sub-MDP is responsible for a subset of problem variables and ignores the others. In an agent based modeling, each sub-MDP will match the agent capabilities in the group (individual perceptions and actions).

An iterative approach is used to solve distributed MDP while each iteration will modify each sub-MDP structure (transition and/or reward values). The computation stops when policies are stable for all the sub-MDPs (agents). The state space explored to compute the policy could be significantly reduced. This should speedup the computation. However, they are no guarantees on the optimality of the solution.

A waterway network would be easily distributable as the transfer points are already distributed over a territory. An agent in the agent based model would be responsible for the control of one or few connected transfer points. A coordination mechanism based on reaches observed by two or more agents would be used. Distributed MDP seems very promising, for this problematic, since they can significantly decrease the computation complexity and allow flexible network definition. However, the result when using such an approach remains uncertain considering that distributed MDP solving is a young approach with mostly *ad hoc* contribution to specific application and no generic framework established yet.

## 6. Conclusion

In this paper, a MDP based approach is introduced to optimize the water management in inland waterway networks with a global view and planning on a given horizon. This approach aims to reduce the impact of drought and flood that may be increased by climate change in the next years.

Using MDPs, it is possible to model the dynamic and the uncertainties of such a system to optimize the navigation conditions. However, this model is quickly limited by the size of the state space and thus possible solutions to circumvent this limitation are presented: factored MDP, decomposed MDP and distributed MDP. Factored and distributed MDPs take advantage of the correlation between the variables in the states and actions definitions.

Future works, would continue to explore distributed MDP. Even though this approach does not guarantee the optimal solution, a better control of the required computation resources is expected and is required to handle networks of any size. The impact of the discretization would also be studied. Furthermore, to reduce the size of the action intervals, chained resolution might be envisaged. By doing another modeling with the action space limited to the actions allowed by the previous policy, the size of action intervals could be reduced. Finally, using this approach on a real network would allow comparison with expert in the management of the inland waterway network.

## 7. References

- Arkell, B. and Darch, G., 2006. Impact of climate change on London's transport network. *Proceedings of the ICE - Municipal Engineer*, 159:231–237.
- Bates, B., Kundzewicz, Z., Wu, S., and Palutikof, J., 2008. Climate change and water. *Technical report, Intergovernmental Panel on Climate Change, Geneva*.
- Bellman, R., 1957. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(4):679–684. ISSN 0022-2518.
- Beuthe, M., Jourquin, B., Urbain, N., Bruinsma, F., Lingemann, I., Ubbels, B., and Heumen, E. V., 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*, 54:387 – 401.
- Beuthe, M., Jourquin, B., Urbain, N., Lingemann, I., and Ubbels, B., 2014. Climate change impacts on transport on the Rhine and Danube: A multimodal approach. *Transportation Research Part D: Transport and Environment*, 27:6 – 11.
- Bichot, C.-E. and Siarry, P., 2011. *Graph Partitioning*. Wiley-ISTE.
- Boutilier, C., Dean, T., and Hanks, S., 1999. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94.
- Boutilier, C., Dearden, R., Goldszmidt, M., and others, 1995. Exploiting structure in policy construction. In *IJCAI*, volume 14, pages 1104–1113.
- Brand, C., Tran, M., and Anable, J., 2012. The UK transport carbon model: An integrated life cycle approach to explore low carbon futures. *Energy Policy*, 41:107–124. ISSN 0301-4215. doi:10.1016/j.enpol.2010.08.019.
- Chades, I., Scherrer, B., and 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*, pages 57–62. ACM, Madrid, Spain. ISBN 1-58113-445-2. doi:10.1145/508791.508804.
- Dean, T. and hong Lin, S., 1995. 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.
- Duviella, E., Nouasse, H., Doniec, A., and 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*.
- Guestrin, C., Koller, D., and Parr, R., 2001. Multiagent Planning with Factored MDPs. In *NIPS*, volume 1, pages 1523–1530.
- Guestrin, C., Venkataraman, S., and Koller, D., 2002. Context-specific multiagent coordination and planning with factored MDPs. In *AAAI/IAAI*, pages 253–259.
- Hoey, J., St-Aubin, R., Hu, A., and Boutilier, C., 1999. SPUDD: Stochastic planning using decision diagrams. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 279–288. Morgan



- Kaufmann Publishers Inc.
- IWAC, 2009. Climate change mitigation and adaptation. Implications for inland waterways in England and Wales. *Report*.
- Kocsis, L. and Szepesvári, C., 2006. Bandit Based Monte-Carlo Planning. In Fürnkranz, J., Scheffer, T., and Spiliopoulou, M., editors, *Machine Learning: ECML 2006*, volume 4212 of *Lecture Notes in Computer Science*, pages 282–293. Springer Berlin Heidelberg. ISBN 978-3-540-45375-8.
- Lozenguez, G., Adouane, L., Beynier, A., Mouaddib, A.-I., and Martinet, P., 2015. Punctual versus continuous auction coordination for multi-robot and multi-task topological navigation. *Autonomous Robots*, pages 1–15.
- Mallidis, I., Dekker, R., and Vlachos, D., 2012. The impact of greening on supply chain design and cost: a case for a developing region. *Journal of Transport Geography*, 22:118–128. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2011.12.007.
- Mihic, S., Golusin, M., and Mihajlovic, M., 2011. Policy and promotion of sustainable inland waterway transport in Europe - Danube River. *Renewable and Sustainable Energy Reviews*, 15(4):1801–1809. ISSN 1364-0321. doi:10.1016/j.rser.2010.11.033.
- Nair, R., Varakantham, P., Tambe, M., and Yokoo, M., 2005. Networked Distributed POMDPs: A Synthesis of Distributed Constraint Optimization and POMDPs. In *National Conference on Artificial Intelligence*, page 7.
- Nouasse, H., Doniec, A., Duviella, E., and Chuquet, K., 2016. Efficient management of inland navigation reaches equipped with lift pumps in a climate change context. *4<sup>th</sup> IAHR Europe Congress, Liege, Belgium 27-29 July*.
- Nouasse, H., Rajaoarisoa, L., Doniec, A., Duviella, E., Chuquet, K., Chiron, P., and Archimede, B., 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*, pages 1–6. IEEE.
- Pachauri, R. K., Allen, M., Barros, V., Broome, J., Cramer, W., Christ, R., Church, J., Clarke, L., Dahe, Q., Dasgupta, P., and others, 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*, pages 422–430.
- Puterman, M. L., 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. ISBN 0-471-61977-9.
- Sabbadin, R., 2002. Graph partitioning techniques for Markov Decision Processes decomposition. In *15th European Conference on Artificial Intelligence*, pages 670–674.
- Wanders, N. and Wada, Y., 2015. Human and climate impacts on the 21<sup>st</sup> century hydrological drought. *Journal of Hydrology*, 526:208–220.

