



**UCLouvain**

Institute of Mechanics,  
Materials and Civil Engineering

# Robust optimisation of the pathway towards a sustainable whole-energy system

A hierarchical multi-objective  
reinforcement-learning based approach

Doctoral dissertation presented by

**Xavier RIXHON**

in partial fulfillment of the requirements for  
the degree of Doctor in Engineering Sciences

December 2023

## **Thesis committee**

Pr. Francesco CONTINO (supervisor, UCLouvain)

Pr. Hervé JEANMART (supervisor, UCLouvain)

Dr. Stefano MORET (ETH Zurich)

Pr. Sylvain QUOILIN (ULiège)

---

---

© 2023  
Xavier Rixhon  
All Rights Reserved

---

---

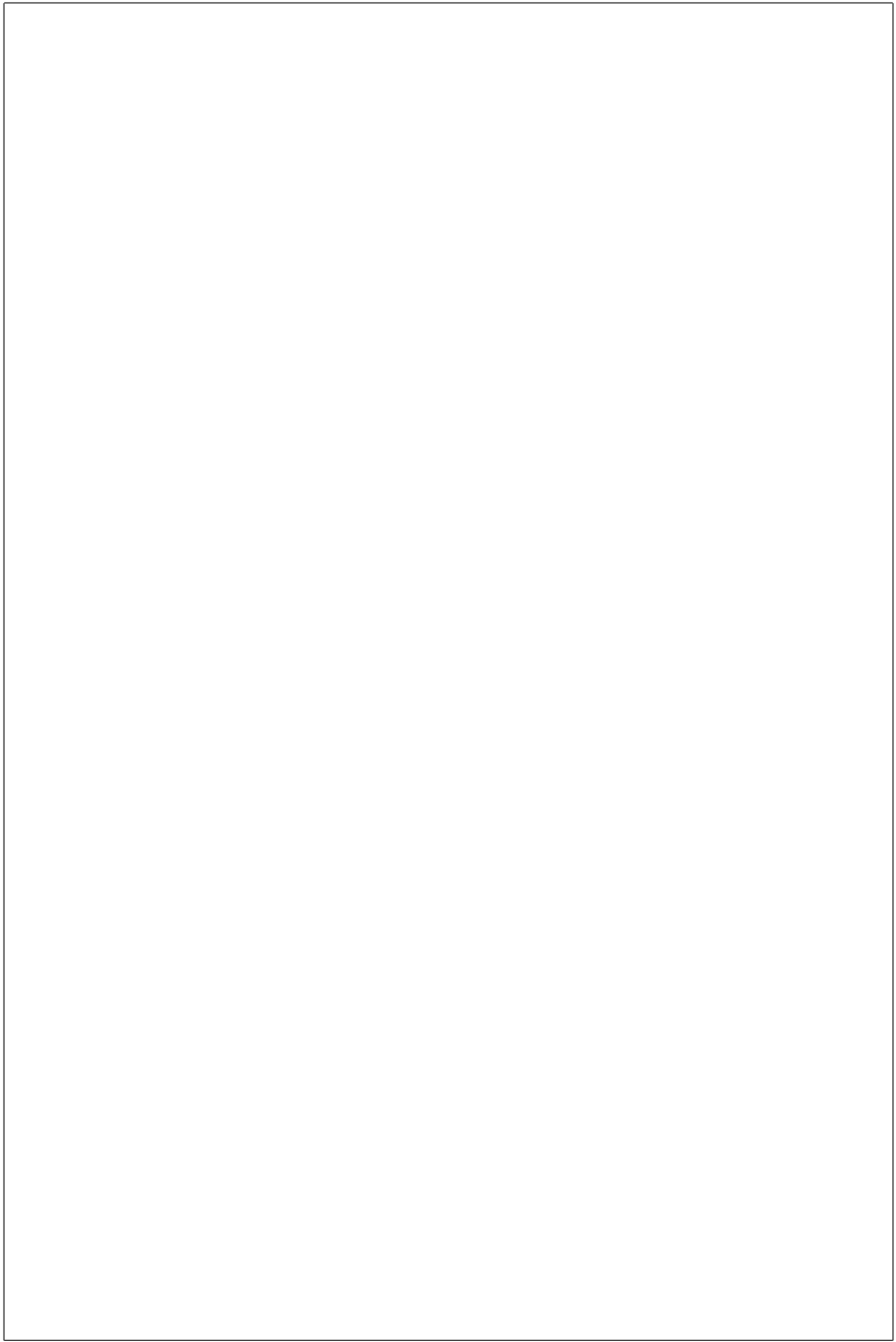
---

# Contents

<b>Symbols</b>	<b>iii</b>
<b>1 Reinforcement Learning CO<sub>2</sub>-policy investigation</b>	<b>1</b>
1.1 Definition of the actions, states and rewards . . . . .	3
1.2 Results - As in Smart Energy Systems Conference . . . . .	5
<b>Bibliography</b>	<b>7</b>

---

---



---

---

---

# Symbols

## Acronyms

<b>API</b>	application programming interface
<b>BECCS</b>	bioenergy with carbon capture and storage
<b>BEMS</b>	building energy management system
<b>BEV</b>	battery electric vehicle
<b>BTX</b>	benzene, toluene and xylene
<b>CAPEX</b>	capital expenditure
<b>CCGT</b>	combined cycle gas turbine
<b>CCS</b>	carbon capture and storage
<b>CHP</b>	combined heat and power
<b>CNG</b>	compressed natural gas
<b>DC</b>	direct current
<b>DHN</b>	district heating network
<b>ESOMs</b>	energy system optimisation models
<b>EnergyScope TD</b>	EnergyScope Typical Days
<b>EUD</b>	end-use demand
<b>FC</b>	fuel cell
<b>GDP</b>	gross domestic product
<b>GHG</b>	greenhouse gases
<b>GSA</b>	global sensitivity analysis
<b>GWP</b>	global warming potential
<b>HP</b>	heat pump
<b>HT</b>	high-temperature
<b>HVC</b>	high value chemicals
<b>IAMs</b>	integrated assessment models
<b>ICE</b>	internal combustion engine

<b>IEA</b>	International Energy Agency
<b>IPCC</b>	intergovernmental panel on climate change
<b>IQR</b>	interquartile range
<b>LCA</b>	life cycle assessment
<b>LCOE</b>	levelised cost of energy
<b>LFO</b>	light fuel oil
<b>LOO</b>	leave-one-out
<b>LPG</b>	liquefied petroleum gas
<b>LT</b>	low-temperature
<b>MDP</b>	Markov decision process
<b>MMSA</b>	Methanol Market Services Asia
<b>MTBE</b>	methyl tert-butyl ether
<b>MTO</b>	methanol-to-olefins
<b>NED</b>	non-energy demand
<b>NG</b>	fossil gas
<b>NN</b>	neural network
<b>NRE</b>	non-renewable energy
<b>NSC</b>	naphtha steam cracker
<b>OPEX</b>	operational expenditure
<b>PC</b>	principal component
<b>PCs</b>	principal components
<b>PCA</b>	Principal Component Analysis
<b>PCE</b>	Polynomial Chaos Expansion
<b>PDF</b>	probability density function
<b>PV</b>	photovoltaic
<b>RE</b>	renewable energy
<b>RL</b>	reinforcement learning
<b>SAC</b>	Soft Actor Critic
<b>SDGs</b>	Sustainable Development Goals
<b>SMR</b>	small modular reactor
<b>SVD</b>	singular value decomposition
<b>UQ</b>	Uncertainty Quantification
<b>VRES</b>	variable renewable energy sources

---

---

## List of publications

Limpens, G., **Rixhon, X.**, Contino, F., & Jeanmart, H. (2024). “*EnergyScope Pathway: An open-source model to optimise the energy transition pathways of a regional whole-energy system.*” In *Applied Energy*, (Vol. 358). URL: <https://doi.org/10.1016/j.apenergy.2023.122501>

**Rixhon, X.**, Limpens, G., Coppitters, D., Jeanmart, H., & Contino, F.(2022). “*The role of electrofuels under uncertainties for the Belgian energy transition.*” In *Energies* (Vol. 14). URL: <https://doi.org/10.3390/en14134027>

**Rixhon, X.**, Tonelli, D., Colla, M., Verleysen, K., Limpens, G., Jeanmart, H. ,& Contino, F.(2022). “*Integration of non-energy among the end-use demands of bottom-up whole-energy system models.*” In *Frontiers in Energy Research, Sec. Process and Energy Systems Engineering*, (Vol. 10). URL: <https://doi.org/10.3389/fenrg.2022.904777>

**Rixhon, X.**, Colla, M., Tonelli, D., Verleysen, K., Limpens, G., Jeanmart, H., & Contino, F.(2021). “*Comprehensive integration of the non-energy demand within a whole-energy system: Towards a defossilisation of the chemical industry in Belgium.*” In *proceedings of ECOS 2021 conference* (Vol. 34, p. 154).

**Rixhon, X.**, Limpens, G., Contino, F., & Jeanmart, H. (2021). “*Taxonomy of the fuels in a whole-energy system.*” In *Frontiers in Energy Research, Sec. Sustainable Energy Systems*, (Vol. 9). URL: <https://doi.org/10.3389/fenrg.2021.660073>

Limpens, G., Coppitters, D., **Rixhon, X.**, Contino, F., & Jeanmart, H. (2020). “*The impact of uncertainties on the Belgian energy system: application of the Polynomial Chaos Expansion to the EnergyScope model.*” In proceedings of ECOS 2020 conference (Vol. 33, p. 711).



---

---

## Chapter 1

# Reinforcement Learning CO<sub>2</sub>-policy investigation

*“For the things we have to learn before we can do them, we learn by doing them.”*

Aristotle, in *The Nicomachean Ethics*, IV<sup>th</sup> century BC

Uncertainties about the future along with a large variety of integrated assessment models (IAMs) yield to an even larger variety of greenhouse gases (GHG) emissions reduction pathways [1]. For instance, several studies [2, 3] advocate for actions to take in the near future, especially to keep on track with the 1.5°C (if not, 2°C) increase of global temperature by the end of the century. On the contrary, using their top-down model DICE, Nordhaus [4] states that immediate and drastic actions are not compulsory to meet the ambition of climate change mitigation. This is even more valid when models assess a myopic transition pathway subject, with limited foresight through the future with progressively unveiled uncertainties.

To address this issue, several approaches have been used. Among them, multi-stage stochastic programming is often put forward as a promising method. Stochastic programming formulates the problem as a mathematical program with probabilistic constraints or objective functions. These models explicitly consider the uncertainty by incorporating probability distributions for the uncertain parameters. The goal is to find an optimal decision that minimizes/maximizes the expected value of the objective function while satisfying the probabilistic constraints, modelled as a scenario tree. At each stage of the problem, here the transition pathway of a whole-energy system, the model has the possibility of recourse, i.e. to adapt the decisions made at earlier stages, in the response to unveiled uncertainties [5]. Using MARKAL model [6], Kanudia and

Loulou [7] assessed a multi-stage stochastic optimisation of the 5-year steps transition of Quebec between 1995 and 2035 accounting for high/low mitigation action plan and high/low growth scenarios. The authors found that hedging strategies, adapting with the future uncertainties, were outperforming the perfect foresight and deterministic optimisation of the different scenarios. However, stochastic programming is usually applied to limited number of uncertainties, i.e. up to 10, and relies on probability distribution that are often difficult to define properly. Increasing the number of these uncertainties in stochastic programming usually leads to a computational burden that limits the use of such a method in IAMs [1]. Based on the approach of Bertsimas and Sim [8], and similarly to Moret [9], Nicolas et al. [1] rather opted for the robust optimisation of the global pathway up to 2200 given different temperature deviation targets, i.e. 2 or 3°C by 2200 via the use of uncertainty budget,  $\Gamma$ , in the TIAM-World model [10]. Considering 9 climate parameters and their respective lower and upper bounds, the idea behind the uncertainty budget stems from the improbability of all parameters simultaneously reaching one of their two extreme values.

In the exploration of the myopic transition pathway under uncertainties, we decided to investigate the reinforcement learning (RL) approach to benefit from its policy optimisation mechanism. Indeed, policymaking for transitioning a whole-energy system can be viewed as an iterative process of learning from policy implementation efforts, involving ongoing analysis of energy policy challenges and experimenting with various solutions [11]. RL exhibits two main advantages: its effectiveness to handle uncertainties and the model-free approach where an accurate representation of the real world is not needed to optimize the policy [12]. Besides the environment, i.e. the myopic transition pathway of the whole-energy system via EnergyScope Pathway (see Chapter ??), the first part of this chapter presents the three key features of interaction between the agent, optimizing its policy, and the environment: actions, states and reward. Then, the results of this policy optimisation point out strategies to follow, i.e. *sweet spots*, in the transitions under uncertainties as well as *no-go zones* where the chances of succeeding the transition, i.e. respecting the CO<sub>2</sub>-budget, are very limited. Finally, these results are compared with references, i.e. the perfect foresight and the myopic optimisation of the transition under the same uncertainties but without the trained RL-agent that can support this transition thanks to its learned policy.

## Contributions

Applying the RL approach to the optimization of the myopic transition pathway of a whole-energy system presents several novelties. First of all, as introduced in Section

??, when applied to energy systems, RL is more dedicated either to smaller scale systems (e.g. building energy management system (BEMS), vehicles and energy devices) or to sector-specific, often the power sector, problems (e.g. dispatch problems, energy markets and grid) [12]. In our case, the sector-coupling, the long-term goal at the end of a multiple-steps transition and the number of uncertain parameters make this application new for RL.

Then, applying to this optimisation environment, i.e. EnergyScope myopic Pathway, rather than a simulation environment, allows building a hierarchical multi-objective optimisation framework. In this agent, while the objective of the environment remains the minimisation of the total “transition” cost (on the concerned limited time window), the agent optimises its strategy to respect the CO<sub>2</sub>-budget.

Finally, comparing the RL-based results with more conventional approaches, i.e. perfect foresight, if not myopic, optimisation without learning process, highlights the added-value brought by the optimised policy.

## 1.1 Definition of the actions, states and rewards

As already introduced in Section ??, the environment with which the RL-agent interacts is the optimisation of the transition pathway whole-energy system on a specific time window, e.g. 2020-2030 then 2025-2035 and so on, until 2040-2050 (see Figure ??). In a nutshell, starting from the initial state of the environment (i.e. the whole-energy system in 2020), the agent takes a set of actions that influence the environment. Then, the window 2020-2030 is optimised via EnergyScope. Some of the outputs of this optimisation feed the agent with either the new state of the system or the reward, i.e. telling the agent how good the actions were at the state he took it. Based on these two pieces of information, i.e. the new state and the reward, the agent takes another set of actions and the window 2025-2035 is optimised. This goes on until eventually reaching 2050. The main purpose of this section is to define the shape of the reward as well as the sets of actions and states.

### Reward

Properly defined the reward fed by the environment to the agent is crucial in RL for several reasons. If the reward is not properly defined, the agent may optimize for an unintended objective, leading to undesired or suboptimal behavior, i.e. the misalignment of the learning objective [13]. Even worse, it can lead to reward hacking/tampering where the agent exploits loopholes in the reward function to achieve higher rewards without actually performing the desired task [14]. On the contrary, a proper defini-

tion of the reward function increases the sample efficiency, i.e. requiring less episode to converge to the optimal policy. It also makes the policy more stable and able to withstand variations and uncertainties in the environment [15].

### Actions

Defining the levers of action, the core of the policy, to support the transition of a country-size whole-energy system is challenging, especially when accounting for political and socio-technical aspects [16]. In our work, focusing only on the techno-economic aspect, we assume that the actions taken by the agent are directly implemented and impacting the environment, without “misfire”. In other words, considering only the techno-economic lens, there is no moderation nor contest towards the agent’s actions, as the objective is to assess how far and when within the transition to push the different levers of action. Given the overall objective of the agent to succeed the transition, i.e. respecting the CO<sub>2</sub>-budget by 2050, we have defined the actions in this sense. The first action,  $\text{act}_{\text{gwp}} \in [0, 1]$ , aims at limiting the emissions at the representative year ending the concerned time window,  $\mathbf{GWP}_{\text{tot}}(y_{\text{end of the window}})$ , between the level of emissions in 2020, i.e.  $\mathbf{GWP}_{\text{tot}}(2020) = 123 \text{ Mt}_{\text{CO}_2, \text{eq}}$ , and carbon-neutrality:

$$\mathbf{GWP}_{\text{tot}}(y_{\text{end of the window}}) \leq \text{act}_{\text{gwp}} \cdot \mathbf{GWP}_{\text{tot}}(2020) \quad (1.1)$$

Out the total GHG emissions in Belgium in 2020, oil (i.e. so-called light fuel oil (LFO) in the model), on the one hand and, on the other hand, fossil gas, account for roughly 40% and 31%, respectively. Then, even though its use in 2020 is much more limited compared to the two formers, i.e. 28 TWh of solid fossil fuels (i.e. so-called COAL in the model) versus 159 and 142 TWh for oil and fossil gas, respectively, coal is a cheap, about 0.017€/GWh, and highly-emitting resource, 0.40 kt<sub>CO<sub>2</sub>,eq</sub>/GWh. For these reasons, three additional actions to support the strict limitation of overall emissions of the first action: limiting the consumption of these three fossil resources up to the level of consumption in 2020,  $\mathbf{Cons}_{\text{fossil gas}}(2020)$ ,  $\mathbf{Cons}_{\text{LFO}}(2020)$  and  $\mathbf{Cons}_{\text{coal}}(2020)$ , over the entire concerned time window, except the first one as this year is the initial condition of the time window and cannot be optimised any more:

$$\mathbf{Cons}_{\text{fossil gas}}(y) \leq \text{act}_{\text{fossil gas}} \cdot \mathbf{Cons}_{\text{fossil gas}}(2020) \quad \forall y \in \text{time window} \quad (1.2)$$

$$\mathbf{Cons}_{\text{LFO}}(y) \leq \text{act}_{\text{LFO}} \cdot \mathbf{Cons}_{\text{LFO}}(2020) \quad \forall y \in \text{time window} \quad (1.3)$$

$$\mathbf{Cons}_{\text{coal}}(y) \leq \text{act}_{\text{coal}} \cdot \mathbf{Cons}_{\text{coal}}(2020) \quad \forall y \in \text{time window} \quad (1.4)$$

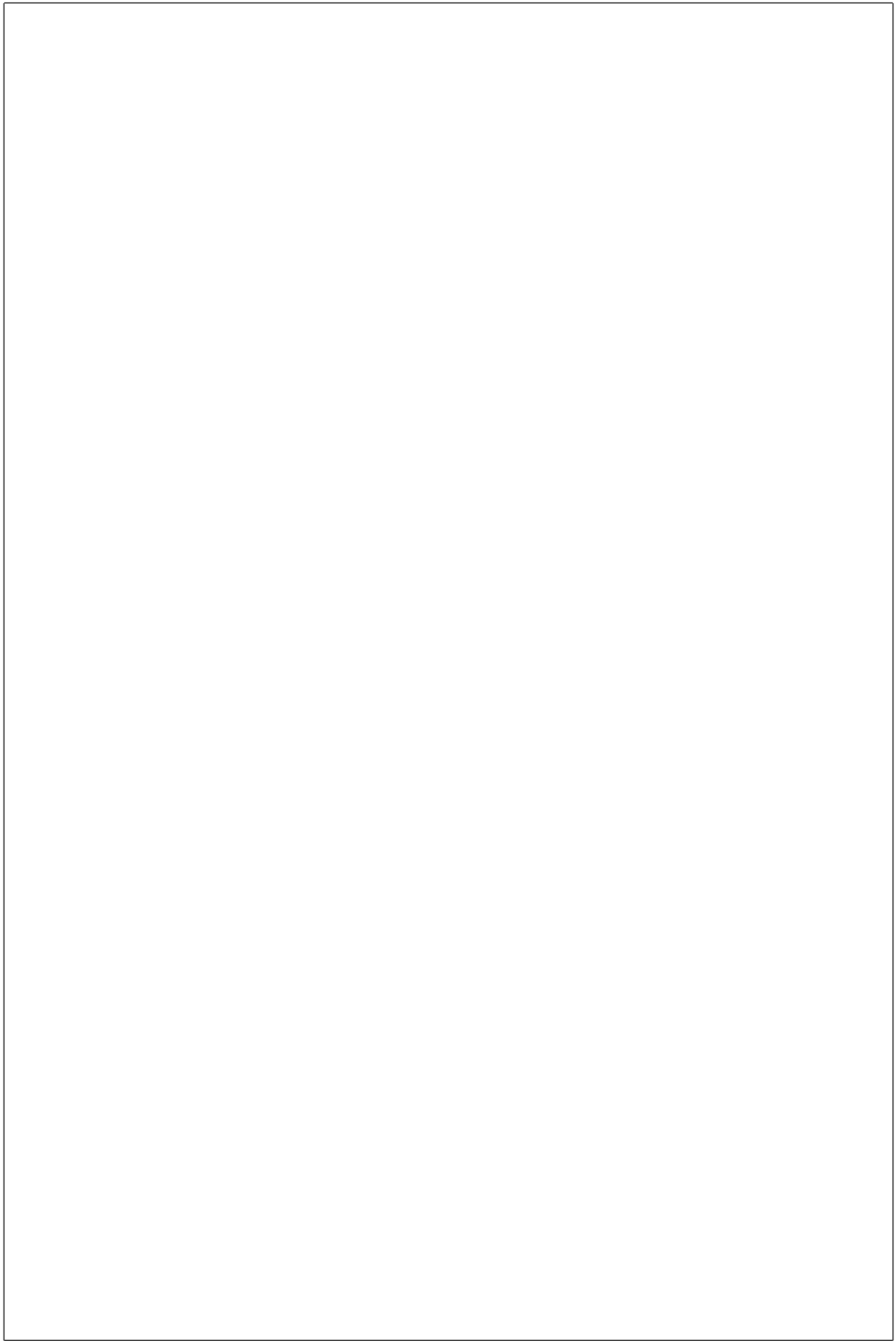
where  $\text{act}_{\text{fossil gas}}$ ,  $\text{act}_{\text{LFO}}$  and  $\text{act}_{\text{coal}}$  can take values between 0 and 1. These complete the action space of the agent,  $A \in \mathbb{R}_{[0,1]}^4$ .

**States****1.2 Results - As in Smart Energy Systems Conference**

Confirm here with what is said in [17]: Importantly, our results also show that near-term policy stringency is an important driver of cumulative Res-FFI-CO<sub>2</sub> in climate change mitigation scenarios. If strengthening of NDCs fails, Res-FFI-CO<sub>2</sub> will be even higher, not only because of additional near-term emissions, but also due to a decrease of economic mitigation potentials in the longer term caused by further carbon lock-in. Delaying the strengthening of mitigation action will increase the world's dependence on CDR for holding warming to well below 2°C, and is likely to push the 1.5°C target out of reach for this century.

---

---



---

---

---

## Bibliography

- [1] C. Nicolas, S. Tchung-Ming, O. Bahn, E. Delage, Robust Enough? Exploring Temperature-Constrained Energy Transition Pathways under Climate Uncertainty, *Energies* 14 (2021) 8595.
- [2] Intergovernmental Panel on Climate Change (IPCC), Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, Technical Report, IPCC, 2018.
- [3] W. Steffen, J. Rockström, K. Richardson, T. M. Lenton, C. Folke, D. Liverman, C. P. Summerhayes, A. D. Barnosky, S. E. Cornell, M. Crucifix, et al., Trajectories of the earth system in the anthropocene, *Proceedings of the National Academy of Sciences* 115 (2018) 8252–8259.
- [4] W. Nordhaus, *A question of balance: Weighing the options on global warming policies*, Yale University Press, 2014.
- [5] I. E. Grossmann, R. M. Apap, B. A. Calfa, P. García-Herreros, Q. Zhang, Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty, *Computers & Chemical Engineering* 91 (2016) 3–14.
- [6] L. G. Fishbone, H. Abilock, Markal, a linear-programming model for energy systems analysis: Technical description of the bnl version, *International journal of Energy research* 5 (1981) 353–375.
- [7] A. Kanudia, R. Loulou, Robust responses to climate change via stochastic MARKAL: The case of Québec, *European Journal of Operational Research* 106 (1998) 15–30.
- [8] D. Bertsimas, M. Sim, The price of robustness, *Operations research* 52 (2004) 35–53.
- [9] S. Moret, *Strategic energy planning under uncertainty*, Ph.D. thesis, EPFL, 2017.

- [10] R. Loulou, U. Remme, A. Kanudia, A. Lehtila, G. Goldstein, Documentation for the times model part ii, Energy technology systems analysis programme (2005).
- [11] M. Howlett, M. Ramesh, A. Perl, et al., Studying public policy: Policy cycles and policy subsystems, volume 3, Oxford university press Toronto, 1995.
- [12] A. Perera, P. Kamalaruban, Applications of reinforcement learning in energy systems, Renewable and Sustainable Energy Reviews 137 (2021) 110618.
- [13] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, D. Amodei, Deep reinforcement learning from human preferences, Advances in neural information processing systems 30 (2017).
- [14] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, D. Mané, Concrete problems in ai safety, arXiv preprint arXiv:1606.06565 (2016).
- [15] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, D. Meger, Deep reinforcement learning that matters, in: Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.
- [16] O. Castrejon-Campos, L. Aye, F. K. P. Hui, Making policy mixes more robust: An integrative and interdisciplinary approach for clean energy transitions, Energy Research & Social Science 64 (2020) 101425.
- [17] G. Luderer, Z. Vrontisi, C. Bertram, O. Y. Edelenbosch, R. C. Pietzcker, J. Rogelj, H. S. De Boer, L. Drouet, J. Emmerling, O. Fricko, et al., Residual fossil CO<sub>2</sub> emissions in 1.5–2 C pathways, Nature Climate Change 8 (2018) 626–633.