
Deep Reinforcement Learning to Minimize Long-Term Carbon Emissions and Cost in the Investment of Electricity Generation

Anonymous Author(s)

Affiliation

Address

email

Abstract

A transition from a high-carbon emitting electricity power system to one based on renewables would aid in the mitigation of climate change. Decarbonization of the electricity grid would allow for low-carbon heating, cooling and transport. Investments in renewable energy, however, must be made over a long time horizon, with various uncertainties in future electricity demand and costs.

To account for imperfect information of the future, we use the deep deterministic policy gradient (DDPG) deep reinforcement learning to optimize for a low-cost, low-carbon electricity supply using a modified version of the FTT:Power model. The DDPG algorithm is able to learn the optimum electricity mix through experience to achieve this between the years of 2017 and 2050. We find that a transition from fossil fuels and nuclear power to renewables, based upon wind, solar and wave would provide a cheap and low-carbon alternative.

1 Introduction

A transition from a high carbon electricity supply to a low-carbon system is central to avoiding catastrophic climate change [1]. A low-carbon electricity supply would aid in the decarbonization of the automotive and heating sectors. Such a transition must be made in a gradual approach to maintain grid reliability [2].

Renewable energy costs, such as solar and wind energy, have dropped recently, making them cost-competitive with fossil fuels. These drops in prices are projected to continue [3]. The future cost of generation, demand and fuel prices, however, remain uncertain over the long-term future. These uncertainties are risks which investors must analyze while making long-term investment decisions.

In this paper, we use the deep deterministic policy gradient (DDPG) reinforcement learning algorithm to simulate the behaviour of investors over a 33-year horizon, between 2017 and 2050 [4]. We projected until 2050 due to the fact that this is a frequent target for governments to reach zero carbon. The environment used was a modified version of the FTT:Power model [5]. FTT:Power is a global power systems model that uses logistic differential equations to simulate technology switching.

We modified the FTT:Power model to use the DDPG algorithm in place of the logistic differential equations to simulate investment decisions. In addition to this, we only simulated two countries: the United Kingdom and Ireland. We chose these due to our previous work on modelling the UK electricity mix over a similar horizon [1]. The DDPG algorithm allowed us to simulate the decisions made by investors under imperfect information over a 33-year period.

The reinforcement learning algorithm enabled us to model the behaviour of an investor without perfect knowledge of the future, with a view to reducing carbon emissions and the overall cost of the

34 system. The reinforcement learning agent is a single actor that invests in both the UK and Ireland.
 35 This work enabled us to see whether a low-carbon mix is possible over the next 33-years to avert
 36 climate change.

37 Oliveira *et al.* also use reinforcement learning for the capacity expansion problem [6]. They, however,
 38 focus on a 20-year time horizon. Kazempour *et al.* use a mixed integer linear programming approach
 39 to solve the generation investment problem [7]. In our work, we address a gap in the literature for the
 40 capacity expansion problem over a 33 year time period using deep reinforcement learning to reduce
 41 both carbon emissions and electricity price.

42 Through this work, it is possible to assess whether a low cost, low-carbon electricity mix is viable
 43 over the long-term using a deep reinforcement learning investment algorithm, as well as finding what
 44 this optimum mix should be. Our approach is in contrast to a mixed-integer linear programming
 45 problem, where full knowledge of the time-horizon is required. This work enables us to closely
 46 match the investment behaviour of rational agents, without knowledge of the future. It can help guide
 47 investors on which technologies to invest in over the long-term, as well as the proportions to invest in.

48 2 Model and methodology

49 In this paper, we used the Future Technology Transformations for the power sector model (FTT:Power)
 50 [5]. This model represents global power systems based on market competition, induced technological
 51 change and natural resource use and depletion. Induced technological change occurs through
 52 technological learning produced by cumulative investment and leads to nonlinear path dependent
 53 technological transitions [5]. The model uses a dynamic set of logistic differential equations for
 54 competition between technology options.

55 For this work we modified the FTT:Power model to use the deep reinforcement learning investment
 56 algorithm, DDPG. That is, the size of the investment made in each technology was made by the
 57 DDPG algorithm. In addition, we reduced the model to only consider the countries of Ireland and the
 58 UK. This enabled us to iterate through enough episodes for the reinforcement learning to converge to
 59 an optimal reward.

60 Reinforcement Learning

61 The investment decision-making process can be formulated as a Markov Decision Process (MDP) [8].
 62 In an MDP environment, an agent receives an observation about the state of their environment s_t ,
 63 chooses an action a_t and receives a reward r_t based upon their environment and action. Solving an
 64 MDP consists of maximizing the cumulative reward over the lifetime of the agent.

65 For our simulation environment, the agent makes a continuous investment decision for each energy
 66 technology, in each region and each year, starting from 2017 until 2050. Technology switching is
 67 modelled using a pairwise comparison of flows of market shares of different electricity generation
 68 capacity. That is, how much capacity flows from one technology to another.

69 The agent's observation space is a matrix consisting of the electricity produced by each technology,
 70 total capacity, total CO₂ emissions over the simulation, levelized cost of electricity (LCOE) including
 71 both taxes and without taxes, cumulative investment in each technology, investment in new capacity,
 72 carrier prices by commodity, fuel costs and carbon costs.

73 The reward r is defined as:

$$r = - \left(1000 \times \text{CO}_{2e} + \frac{LCOE}{1000} \right), \quad (1)$$

74 where CO_{2e} is equal to total CO₂ emissions over the simulation, and LCOE is equal to LCOE,
 75 excluding taxes. The scaling factors were used to place the LCOE and CO₂ on the same scale. The
 76 reward was multiplied by -1 due to the RL algorithm maximizing reward and our requirement to
 77 reduce both LCOE and CO₂ emissions.

78 RL approaches have been used to solve MDP [9]. In recent times, however, RL has been extended to
 79 incorporate Deep Reinforcement Learning (DRL). DRL relies on deep neural networks to overcome
 80 the problems of memory complexity and computation complexity [10].

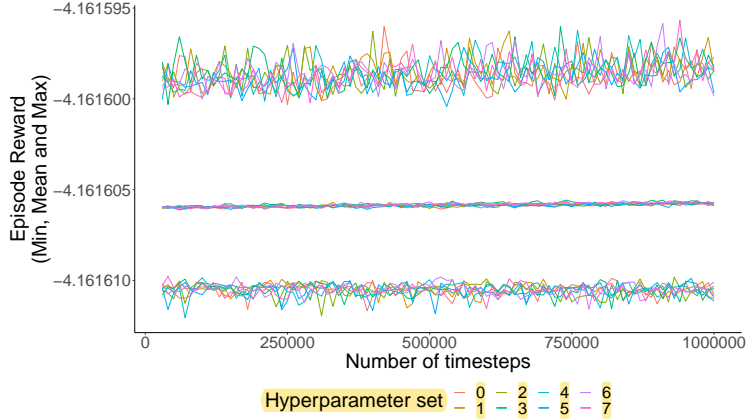


Figure 1: Training with different hyperparameters, displaying the minimum, mean and maximum rewards per episode.

We applied the deep deterministic policy gradient (DDPG) DRL algorithm [4] from the Ray RLlib package [11]. The DDPG algorithm is made up of an actor and critic network. We designed both of these to have two hidden layers, made up of 400 and 300 units per layer. The training batch size was set to 40,000. We chose these parameters due to them being the default implementation in Ray RLlib. We trialled a variety of different configurations for number of neurons per layer. To increase speed of computation, we reduced the simulation to run from 2007 to 2017. We chose this number as it would allow for a transition in electricity mix. However, we found that the approach worked well irrespective of parameter choice, as shown by Figure 1.

3 Results

Our results show that our investment agent is able to increase its reward over time, as shown in Figure 2. A total of $\sim 400,000$ steps were required to see a levelling off in reward. The total time taken to simulate $\sim 400,000$ steps was ~ 8 days. We stopped the training and simulation after this time due to diminishing returns and the cost of computation.

Figure 3 displays the results of reinforcement learning algorithm. Before the black vertical line (2017), the investments made are based upon historical data used by FTT:Power. The reinforcement learning algorithm starts to make investments after the black vertical line.

The historical electricity mix before 2017 is largely based on fossil fuels: coal, combined cycle gas turbine (CCGT) and oil. Additionally, nuclear is a major component of the electricity mix before 2009. After reinforcement learning optimizes for LCOE and carbon emissions, a rapid transition occurs from fossil fuel and nuclear to renewable energy.

This rapid transition occurs due to the reinforcement learning algorithm not taking into account the technical and timeframe constraints embedded in the unmodified FTT:Power model. However, whilst it is likely that whilst the transition speed is unrealistic, the electricity mix found by the reinforcement learning algorithm is likely to be optimal, according to the reward function defined in Equation 1.

The primary source of energy after the reinforcement learning algorithm begins is offshore, followed by onshore, solar photovoltaics (PV) and wave. As can be seen by Figure 4, the carbon emissions reduce significantly at the time that the reinforcement learning algorithm begins to control investments.

This mix of renewable electricity generation across Ireland and the UK allows for demand to be met during the quarterly time periods of the model. The demand scenario is shown in Figure 5, where the demand can be seen to closely match the electricity mix shown by Figure 3.

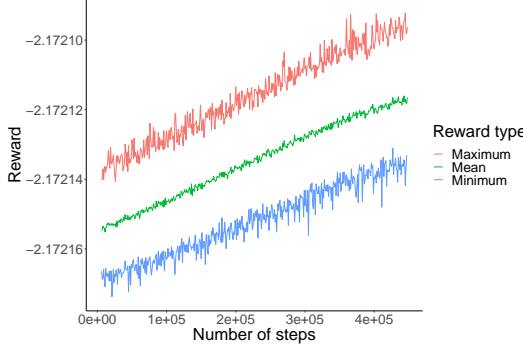


Figure 2: Mean, minimum and maximum rewards over run time.

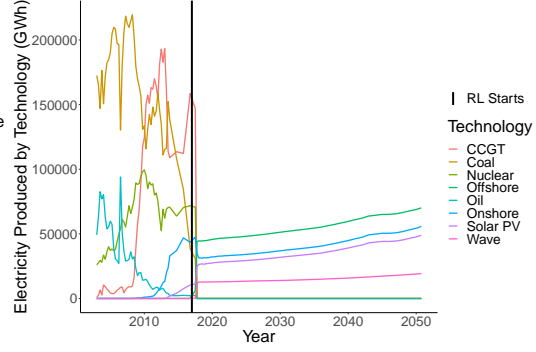


Figure 3: Electricity mix over time.

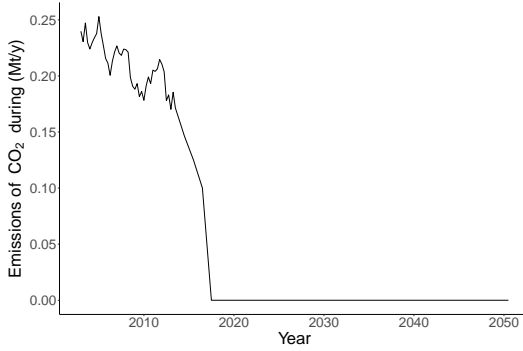


Figure 4: Carbon emissions.

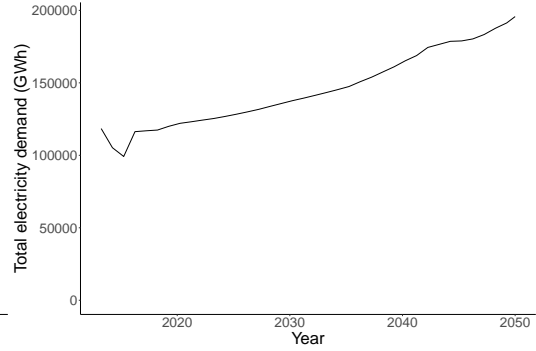


Figure 5: Demand scenario over simulation.

111 4 Discussion

112 A transition from a high carbon-emitting electricity grid to a low-carbon system is required. In order
 113 to achieve this, investments in electricity generators must be made whilst taking into account future
 114 uncertainty. In this paper, we have modelled a central agent which makes investment decisions in an
 115 uncertain environment to find an optimal low-cost, low-carbon electricity mix. To achieve this, we
 116 used the reinforcement learning algorithm, DDPG. The environment is modelled using FTT:Power.

117 Through this exercise, we are able to see the optimal electricity mix in the UK and Ireland. We found
 118 that a mixture of renewable sources such as wind, solar and wave power would meet demand at
 119 quarter year intervals, as well as providing a cost-effective and low-carbon system.

120 A limitation of this work is the fact that the investment algorithm does not take into account the
 121 technical and timeframe constraints of transitions between technologies. It is for this reason that
 122 the reinforcement learning algorithm is able to make such a rapid transition in 2017. However, we
 123 believe that the investment algorithm is able to find a general solution to the problem of investing in a
 124 cost-efficient and low-carbon system over a long time horizon.

125 In future work, we would like to increase the number of steps of the FTT:Power model to more
 126 adequately model the investment behaviour introduced by the reinforcement learning algorithm. A
 127 lower number of simulated time steps leads to an overestimation of the supply of renewables and
 128 underestimation of storage and dispatchable technologies [12]. In addition, an increase in the number
 129 of countries modelled would enable us to see a global picture of how different, interdependent regions
 130 may evolve in a new climate of a requirement of low-carbon emissions. This would require an
 131 exponentially longer runtime for the reinforcement learning algorithm to converge. This is due to
 132 the increased number of decisions that the reinforcement learning algorithm would need to make to
 133 account for the different countries. Finally, we would like to incorporate the technical and timeframe
 134 constraints for technology switching. This could be undertaken by modifying the reward function to
 135 ensure the transition remains within these constraints.

5 Acknowledgements

Acknowledgements redacted to maintain anonymity for double-blind.

References

- [1] A. J. M. Kell, M. Forshaw, and A. S. McGough, “Long-Term Electricity Market Agent Based Model Validation using Genetic Algorithm based Optimization,” *The Eleventh ACM International Conference on Future Energy Systems (e-Energy’20)*, 2020.
- [2] F. Kahrl, J. Williams, D. Jianhua, and H. Junfeng, “Challenges to China’s transition to a low carbon electricity system,” *Energy Policy*, vol. 39, no. 7, pp. 4032–4041, 2011.
- [3] IEA, “Projected Costs of Generating Electricity,” p. 215, 2015.
- [4] J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous learning control with deep reinforcement,” *ICLR*, 2016.
- [5] J. F. Mercure, “FTT:Power A global model of the power sector with induced technological change and natural resource depletion,” *Energy Policy*, vol. 48, pp. 799–811, 2012.
- [6] F. S. Oliveira and M. L. Costa, “Capacity expansion under uncertainty in an oligopoly using indirect reinforcement-learning,” *EJOR*, vol. 267, no. 3, pp. 1039–1050, 2018.
- [7] S. J. Kazempour, A. J. Conejo, and C. Ruiz, “Strategic generation investment using a complementarity approach,” *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 940–948, 2011.
- [8] M. L. Puterman, “Markov decision processes: discrete stochastic dynamic programming,” 2014.
- [9] R. S. Sutton and A. G. Barto, “An introduction to reinforcement learning,” *The MIT Press*, 2015.
- [10] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “A Brief Survey of Deep Reinforcement Learning,” *IEEE Signal Processing Magazine*, pp. 1–16, 2017.
- [11] E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J. E. Gonzalez, M. I. Jordan, and I. Stoica, “RLlib : Abstractions for Distributed Reinforcement Learning,” 2014.
- [12] S. Ludig, M. Haller, E. Schmid, and N. Bauer, “Fluctuating renewables in a long-term climate change mitigation strategy,” *Energy*, vol. 36, no. 11, pp. 6674–6685, 2011.