Modeling Renewable Energy Transitions as a Markov Decision Process:

A Q-Learning approach to minimizing fossil-fuel reliance while meeting energy demands

Xander Hnasko, Luis Botin, and Jonathan Lee
CS238 Final Project

November 18, 2024

1 High Level Overview

1.1 Motivation

This project started from an interest in Germany's *Energiewende* policy inititative in 2016 — a commitment to fully transition the country to renewable energy by the end of the year. Energiewende aimed to eliminate reliance on fossil fuels, largely through the building and use of solar and wind infrastructure. Unfortunately, unprecedented weather patterns (less sun and wind than expected) meant that Germany was running a significant energy deficit, particularly in winter months. To meet demands, Germany was forced to reactivate old coal-fire plants it had shut down under the program. The upfront emissions-costs associated with reactivating old and inefficient coal plants are greater than the emissions-costs of simply leaving coal plants running. Thus, Germany's overall emissions for the year actually *increased*, despite the transition to renewable energy sources. Energiewende highlights the complexity of energy transition plans, as well as the stochastic nature of weather patterns, and the inherent uncertainty on renewable energy production. This begs the question: "how can we minimize our reliance on fossil fuel energy while always meeting energy demands?"

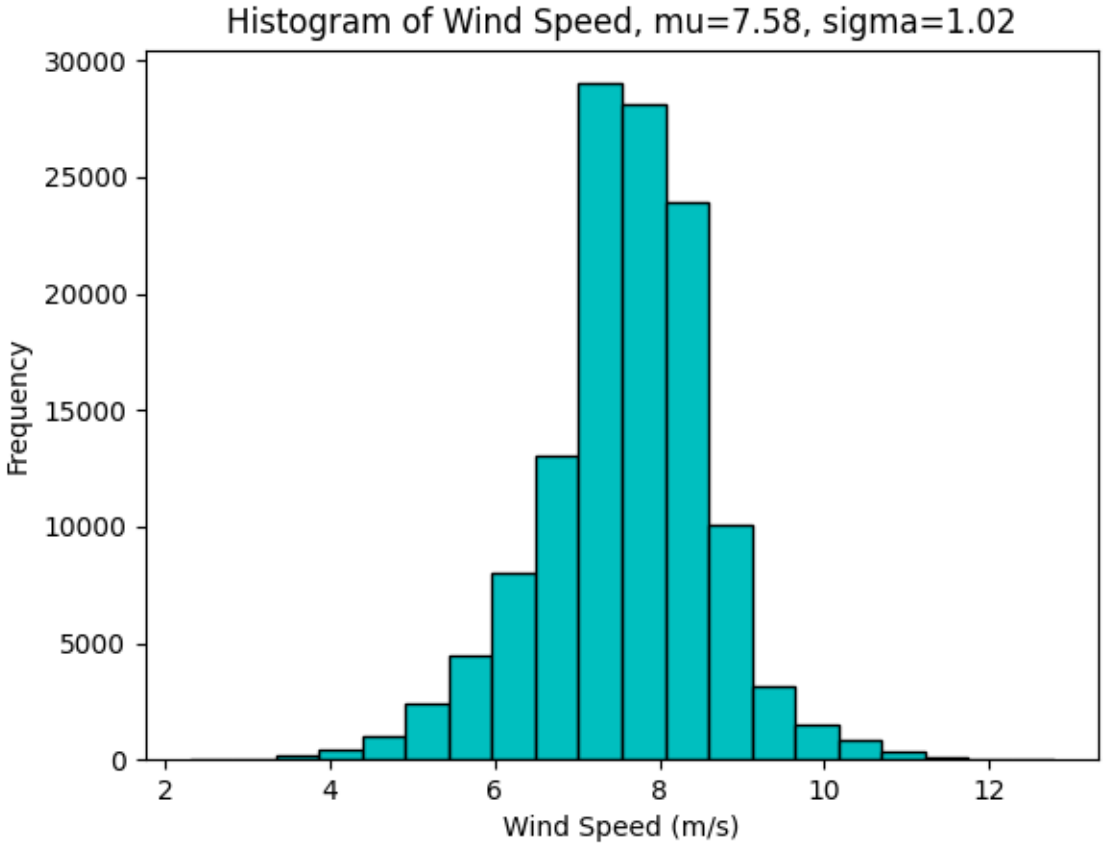
1.2 Goals

This project aims to model our reliance on fossil fuels as a Markov Decision Process. We have chosen to focus specifically on wind power. At each state, we can chose either to increase or decrease our reliance on fossil fuels to meet energy demands. It is assumed that energy production from fossil fuels is both infinite in potential and guaranteed. That is, we can meet energy demands with probability 1 using only fossil fuels. However, since the energy production of renewables is uncertain, we cannot guarantee that renewables alone will always meet energy demands, as was the case for Germany in 2016. We also assume that the infrastructure for renewables remains constant, so the theoretical ceiling of production

is entirely a function of weather patterns (specifically, wind). Our aim is to optimize the proportion of renewable to fossil fuel usage in the short term, seeking to minimize our reliance on fossil fuels while never failing to meet energy demands. Thus, there is a slight positive reward associated with decreasing fossil fuel usage while still meeting demands, a negative reward associated with increasing fossil fuel reliance while still meeting demands, and an extremely negative reward associated with failing to meet energy demands under any circumstance (Energiewende).

2 Data

We are using the Techno-Economic Summary and Index dataset from the National Renewable Energy Laboratory. It contains summarized statistics for 120,000 wind plants within the continental United States. Each row in the dataset contains the location of the wind plant along with a variety of other information, but we specifically care about `wind_speed` and `capacity_factor`, which are used in the formula to measure the power output in watts of the wind turbine (more on this in 3.1). This allows us to construct our state space for our MDP. We have processed and cleaned the data for our needs. A visualization of the distribution of wind speeds across the continental United States is included below.



3 MDP

3.1 State Space

$$s_i = (\varphi_i, w_i, O(w_i), d_i)$$

where:

- $\varphi_i \stackrel{\text{def}}{=}$ Proportion of energy demand accounted for by fossil fuels at state i, where $0 \leq \varphi \leq 1$.
- $w_i \stackrel{\text{def}}{=}$ Wind speed at state i in m/s.
- $O(w_i) \stackrel{\text{def}}{=}$ Power output of the wind turbines (MW), calculated using the formula:

$$O(w_i) = 0.5 \cdot C_p \cdot \rho \cdot \pi \cdot R^2 \cdot w_i^3,$$

  - C_p: Coefficient of performance (efficiency factor).
  - ρ: Air density in kg/m^3, left constant at U.S average of 1.225 kg/m^3.
  - R: Blade radius in meters.
  - w_i: Wind speed in m/s.
- d_i: Energy demand in MW

The state space S captures the system's key variables: fossil fuel reliance (φ), wind speed (w), and wind turbine power output ($O(w_i)$).

3.2 Action Space

$$a_i \in \{\varphi_{(+)}, \varphi_{(-)}, \varphi_0\}$$

where:

- $\varphi_{(+)} \stackrel{\text{def}}{=}$ increase φ_i, the proportion of energy demands met by fossil fuels, by a fixed amount, Δ, where $0 \leq \varphi_i + \Delta \leq 1$
- $\varphi_{(-)} \stackrel{\text{def}}{=}$ decrease φ_i by Δ, where $0 \leq \varphi_i - \Delta \leq 1$
- $\varphi_0 \stackrel{\text{def}}{=}$ leave φ_i unchanged

3.3 Transition Model

3.3.1 Wind Speed Transition ($P(w'|w)$)

The next wind speed w' is modeled as a Gaussian distribution $P(w'|w) \sim \mathcal{N}(\mu'_w, \sigma^2(w))$. The mean μ'_w for the next wind speed is computed as a combination of:

$$\lambda \cdot w + (1 - \lambda) \cdot \mu_w$$

where:

- w: Current wind speed.
- μ_w: Historical mean wind speed (computed from dataset)
- $\lambda = \exp(-\alpha \cdot t)$: Weighting parameter that controls how strongly the transition depends on the current wind speed versus reverting to the historical mean. α controls the rate of decay (reversion to mean) over time and t represents the time steps since last wind speed update.

The variance $\sigma^2(\Delta w)$ is the sample variance of observed wind speed changes in the dataset:

$$\sigma^2(\Delta w) = \hat{\sigma}^2_{\Delta w},$$

which we compute as:

$$\hat{\sigma}^2_{\Delta w} = \frac{1}{n-1} \sum_{i=1}^{n} (\Delta w_i - \bar{\Delta w})^2.$$

We take the variance of the change in wind speed (Δw) instead of the actual wind speeds (w) since this captures short-term, step by step fluctuations, while w would represent long-term variability and may overestimate changes, leading to unrealistic jumps in the model.

Putting everything together, the wind speed transition probability is:

$$P(w'|w) = \frac{1}{\sqrt{2\pi\sigma^2(\Delta w)}} \exp\left(-\frac{(w' - (\lambda \cdot w + (1-\lambda) \cdot \mu_w))^2}{2\sigma^2(\Delta w)} \right)$$

This Gaussian model makes sure that wind speed transitions are localized. If the current wind speed is $w = 10$ m/s, and the distribution is $\mathcal{N}(10, 1)$, most next wind speeds w' will likely fall between 9 and 11, with smaller probabilities for more distant values. The weighted mean $\mu'_w = \lambda \cdot w + (1-\lambda) \cdot \mu_w$ then ensures that wind speed transitions tend toward the historical mean μ_w over time, preventing wind speeds from drifting too high or too low indefinitely. λ decays over time ($\lambda = \exp(-\alpha \cdot t)$, which means that the influence of the historical mean increases with time.

3.3.2 Energy Demand Transition ($P(d'|d)$)

Similarly, we model the next energy demand d' as a Gaussian distribution $P(d'|d) \sim \mathcal{N}(\mu'_d, \sigma^2_d)$ where the mean μ'_d for the next energy demand is computed as:

$$\lambda \cdot d + (1-\lambda) \cdot \mu_d,$$

where:

- d: Current energy demand.
- μ_d: Historical mean energy demand (computed from dataset).
- $\lambda = \exp(-\alpha \cdot t)$: Controls how strongly the transition depends on the current energy demand versus reverting to the historical mean.

The variance σ_d^2 is the sample variance of observed energy demand changes in the dataset:

$$\sigma_d^2 = \hat{\sigma}_d^2,$$

which we compute as:

$$\hat{\sigma}_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\Delta d_i - \bar{\Delta d})^2.$$

We get the following energy demand transition probability:

$$P(d'|d) = \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{(d' - (\lambda \cdot d + (1-\lambda) \cdot \mu_d))^2}{2\sigma_d^2} \right),$$

This Gaussian model ensures that energy demand transitions, too, are localized. If the current energy demand is $d = 100\,\text{MW}$, and the distribution is $\mathcal{N}(100, 25)$, most next energy demands d' will likely fall between 95 and 105, with smaller probabilities for more distant values. The weighted mean μ_d' ensures that energy demand transitions tend toward the historical mean μ_d over time.

3.3.3 Full Transition Probability

Combining the above, we arrive at the full transition probability:

$$T(s'|s, a) = P(w'|w) \cdot P(d'|d),$$

with O' and φ' computed deterministically from w', d', and a.

3.4 Reward Function

The reward function encourages reducing fossil fuel use (φ) while penalizing unmet energy demand:

$$R(s, a, s') = \begin{cases} +10 \cdot (1 - \Delta\varphi) & \text{if } O(w') + \varphi' \cdot d' \geq d', \\ -50 + 10 \cdot (-\Delta\varphi) & \text{if } O(w') + \varphi' \cdot d' < d'. \end{cases}$$

where:

- $\Delta\varphi \stackrel{\text{def}}{=} \varphi' - \varphi$: Change in the proportion of energy demand met by fossil fuels after the transition.
- $O(w') \stackrel{\text{def}}{=} 0.5 \cdot C_p \cdot \rho \cdot \pi \cdot R^2 \cdot (w')^3$: Power output after the transition, where w' is the new wind speed.
- $d \stackrel{\text{def}}{=}$ Energy demand

Cases:

- If the total energy supply ($O(w') + \phi' \cdot d'$) meets or exceeds the demand (d'), agent is rewarded based on the reduction in fossil fuel reliance ($-\Delta\phi$).
- If demand is not met ($O(w') + \phi' \cdot d' < d'$), a significant penalty (-50) is applied, though reductions in ϕ still mitigate this penalty.

Deriving the Optimal Policy using Q-Learning

We solve the MDP and derive the optimal policy via Q-learning. The action-value function $Q(s, a)$ represents the expected cumulative reward we obtain by taking action a in state s and then following the optimal policy. The optimal policy $\pi^*(s)$ is derived as:

$$\pi^*(s) = \arg\max_a Q^*(s, a).$$

The Q-learning algorithm iteratively updates the Q-values based on observed transitions (s, a, r, s') using the update rule seen in class:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left(r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right),$$

In the context of our renewable energy transition problem, the Q-learning algorithm iteratively updates $Q(s, a)$ and derives the optimal policy $\pi^*(s)$. As we iterate, we learn the best actions a to reduce fossil fuel reliance φ while also ensuring demand is met ($O(w)+\varphi' \cdot d' \geq d'$), taking into account stochastic transitions in wind speed w and energy demand d to ultimately optimize a reward structure that balances environmental goals with meeting demand.

3.5 Outstanding Questions

The only area we are still uncertain about is the right way to model demand. As mentioned in 3.1, we need to include a current energy demand d_i. Once we find an initial way to model d_i, coding the rest of the MDP is relatively straightforward - this is our biggest bottleneck at the moment. The simplest option we considered was to simulate demand using Gaussian transitions. We can assume demand transitions smoothly around a historical or current average, μ_d with variance σ_d^2 which we can compute from our dataset. Thus, we can sample from the distribution using $P(d'|d) \sim \mathcal{N}(\mu_d', \sigma_d^2)$ (see 3.3.2). We also considered looking for a dataset that can show historical demand data for wind energy. An example of some of the datasets we are finding can be found here. But if the simulation approach seems reasonable to the teaching team, we will likely go with that method for the sake of time. Would greatly appreciate your help in sanity checking our proposed method!