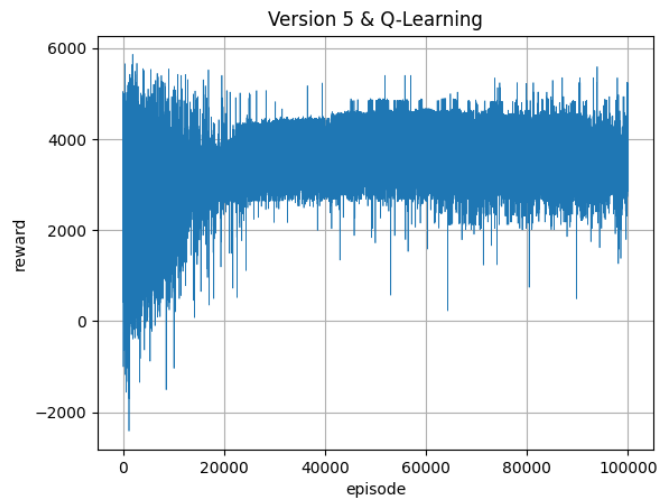
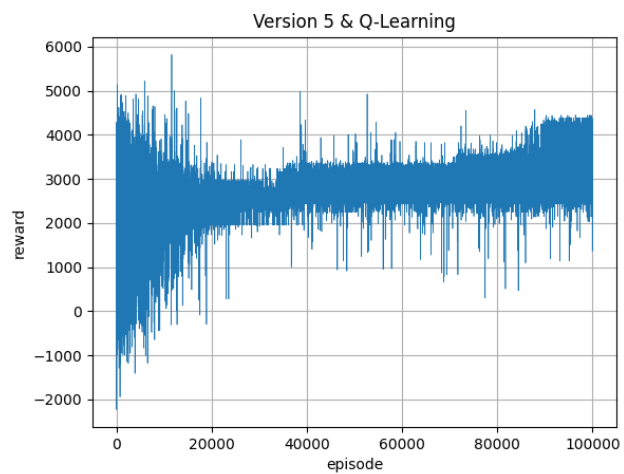


Random Seeds(Q-learning 100000 rounds)

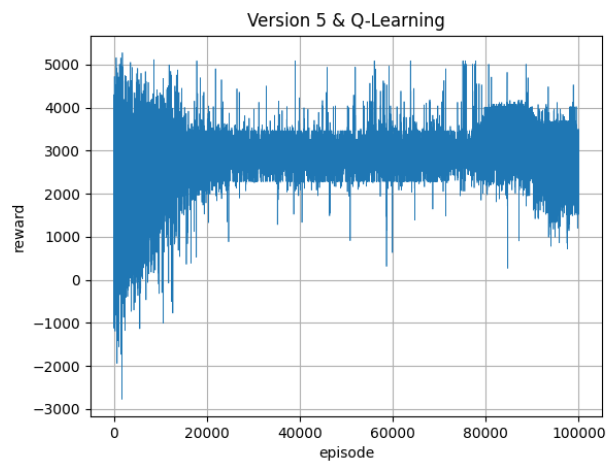
1. Random Seed = 10



2. Random Seed = 1



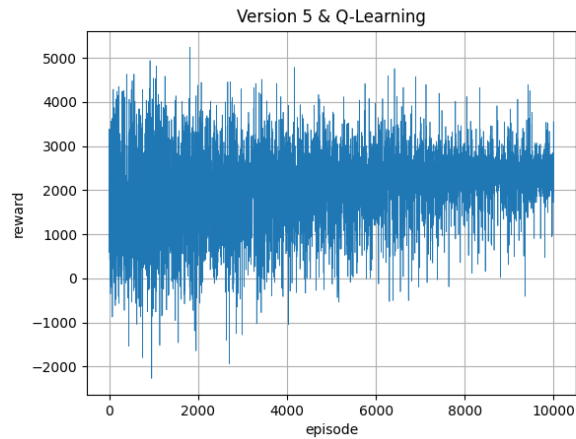
3. Random Seed =0



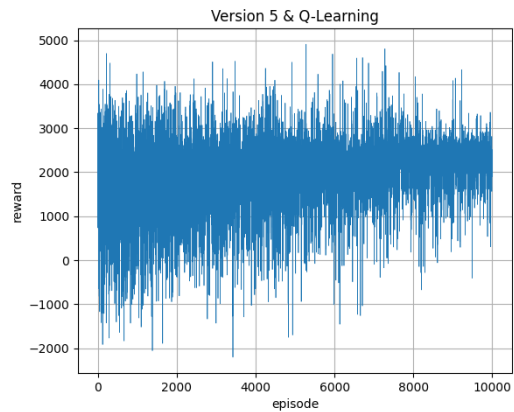
MINIMUM_REQ_GHG_10 And MINIMUM_REQ_TIMBER_1

(Q-learning 1000 rounds)

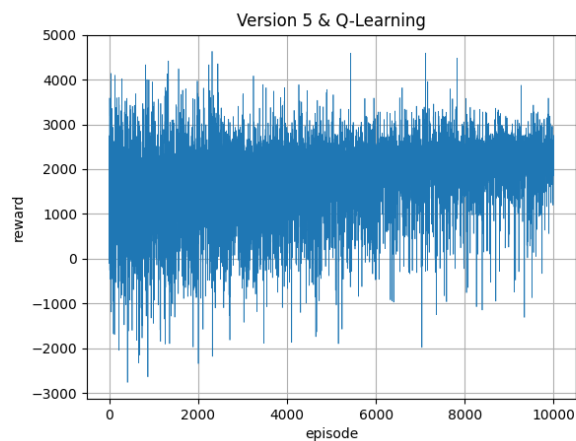
1. MINIMUM_REQ_GHG_10 = 2000 MINIMUM_REQ_TIMBER_1 = 25



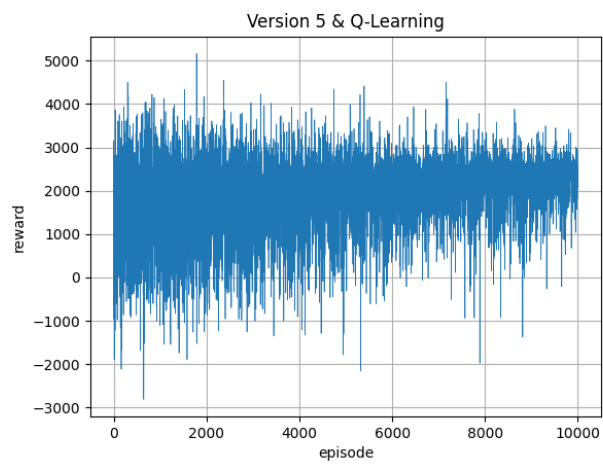
2. MINIMUM_REQ_GHG_10 = 2000 MINIMUM_REQ_TIMBER_1 = 50



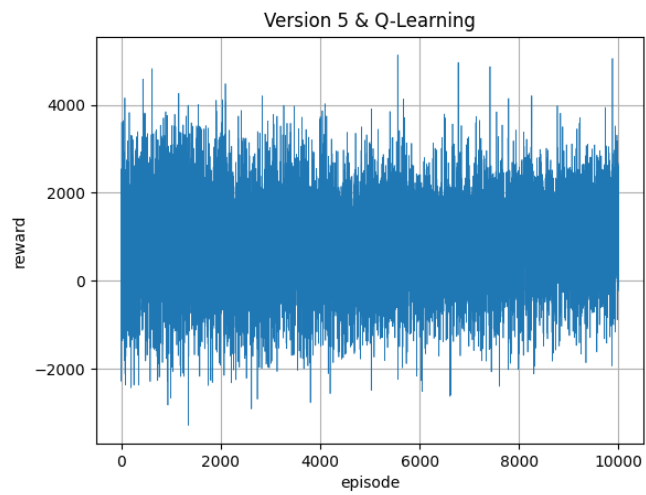
3. MINIMUM_REQ_GHG_10 = 2000 MINIMUM_REQ_TIMBER_1 = 75



4. MINIMUM_REQ_GHG_10 = 4000 MINIMUM_REQ_TIMBER_1 = 50



5. MINIMUM_REQ_GHG_10 = 40000 MINIMUM_REQ_TIMBER_1 = 50



Impact of seeds on RL algorithms in Version 5

Q-Learning:

Hyperparameters: Learning rate: 0.2, Discount factor: 1.0, 3. Exploration rate: 1.0, Exploration rate decay: 0.9999, Minimal exploration rate: 0.05, Value:CO2=1:1

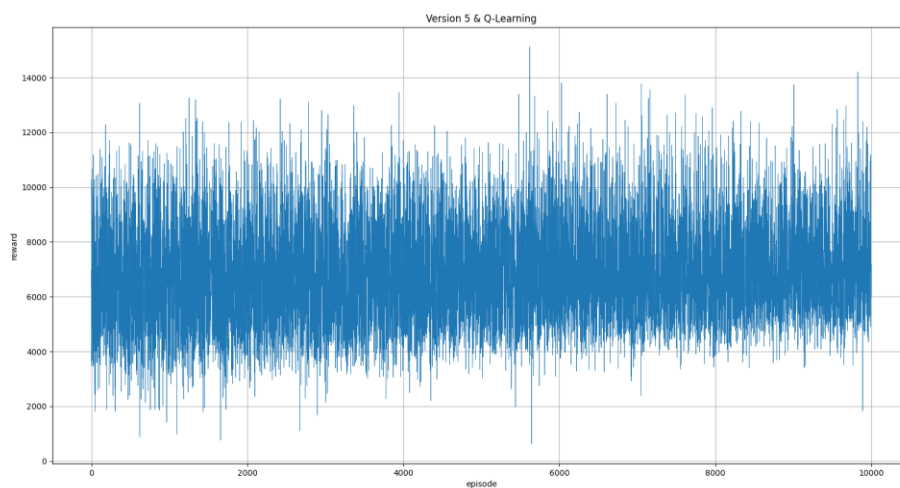
Seed: 0

Mean reward: 6848.7613

Standard deviation of reward: 1983.4275

Maximal reward: 15118.3904

Minimal reward: 632.0347



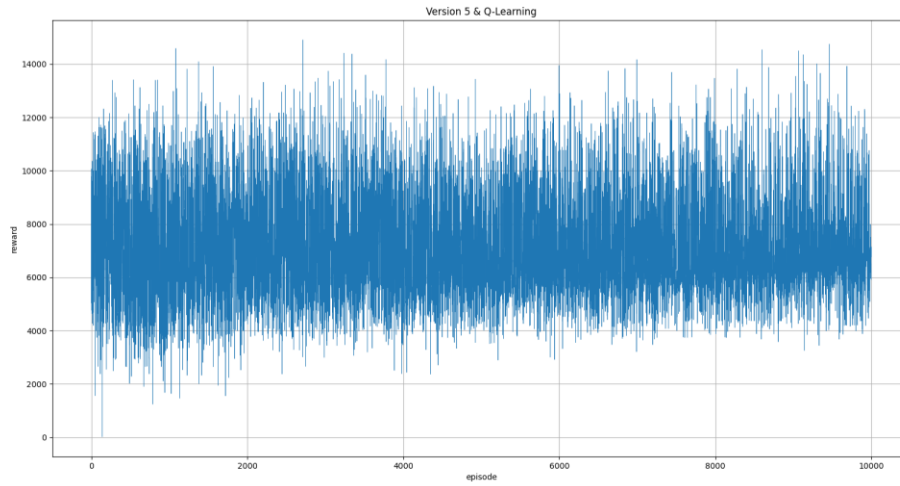
Seed: 1

Mean reward: 7100.9568

Standard deviation of Learning: 2120.9461

Maximal reward: 14904.4447

Minimal reward: 17.1992



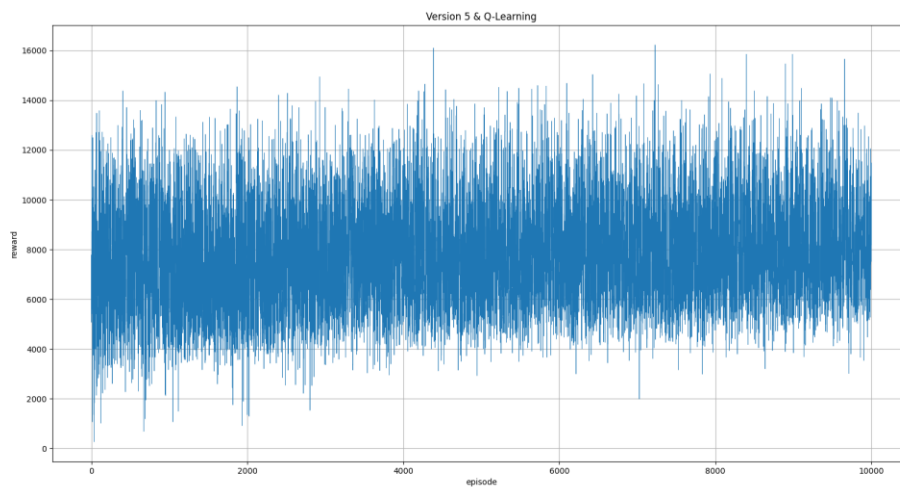
Seed: 2

Mean reward: 7761.5631

Standard deviation of Learning: 2323.6350

Maximal reward: 16228.2980

Minimal reward: 266.9626



Deep Q-learning:

Hyperparameters: Learning rate: $1e-5$, Discount factor: 1.0, Exploration rate: 0.9, Exploration rate decay: 0.9999, Minimal exploration rate: 0.05, Batch size: 64, Hidden layer: ReLU 100x100 ReLU, Optimizer: Adam, Replay buffer size: 5, Value:CO2=1:1

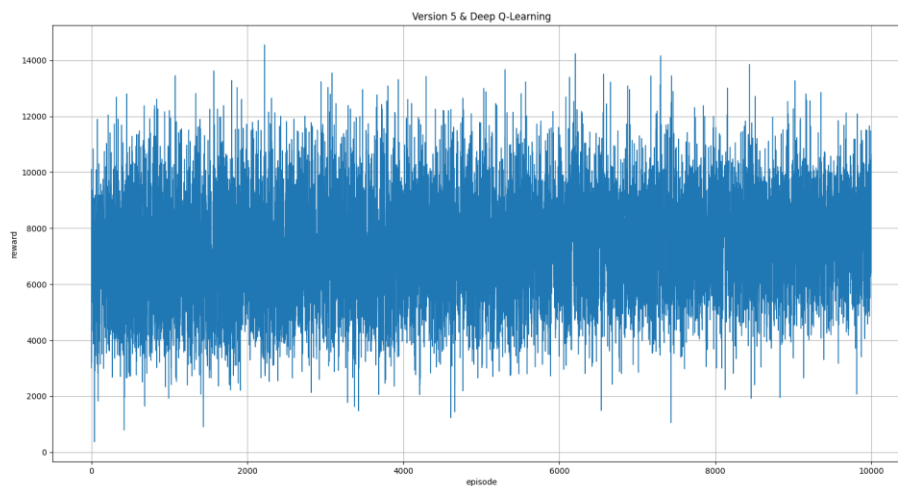
Seed 0:

Mean reward: 7369.8055

Standard deviation of Learning: 2052.5930

Maximal reward: 14551.7836

Minimal reward: 372.0220



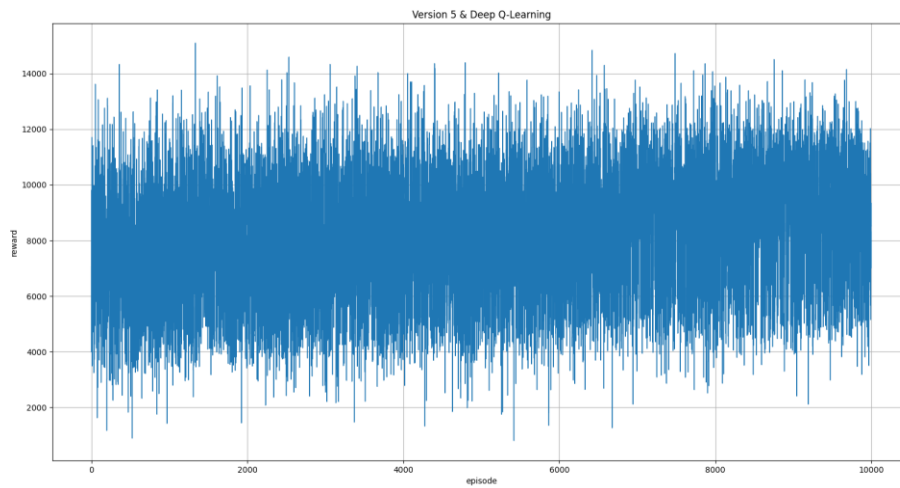
Seed: 1

Mean reward: 8239.6171

Standard deviation of Learning: 2478.5593

Maximal reward: 15098.3253

Minimal reward: 802.0636



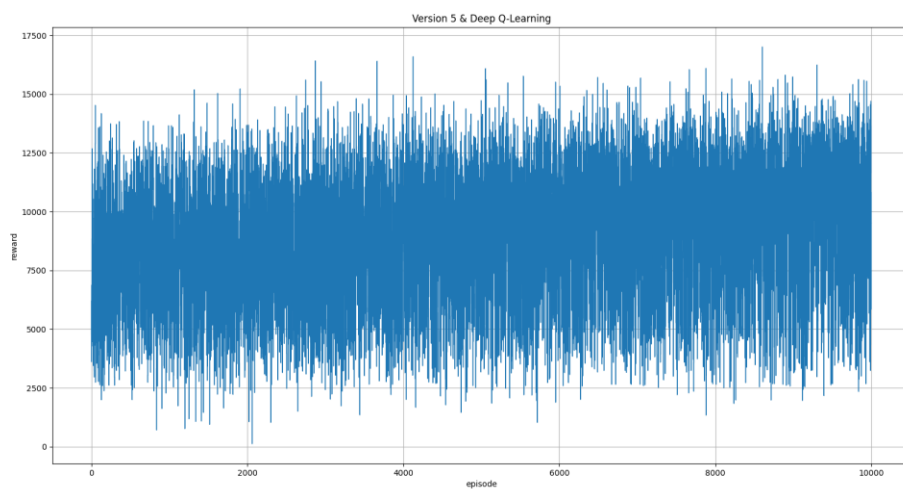
Seed: 2

Mean reward: 9068.6768

Standard deviation of Learning: 2978.8094

Maximal reward: 17007.1085

Minimal reward: 121.0351



Policy Gradient:

Hyperparameters: Learning rate: $1e-4$, Discount factor: 1.0, Batch size: 128,

Hidden layer: 256x64 sigmoid 64x256, Optimizer: Adam

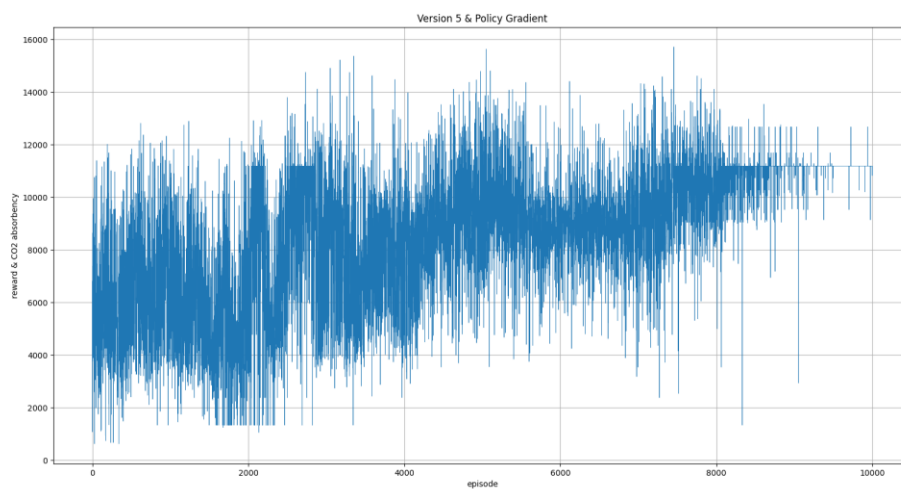
Seed: 0

Mean reward: 7843.2415

Standard deviation of reward: 2817.3521

Maximal reward: 15783.6520

Minimal reward: 886.3654



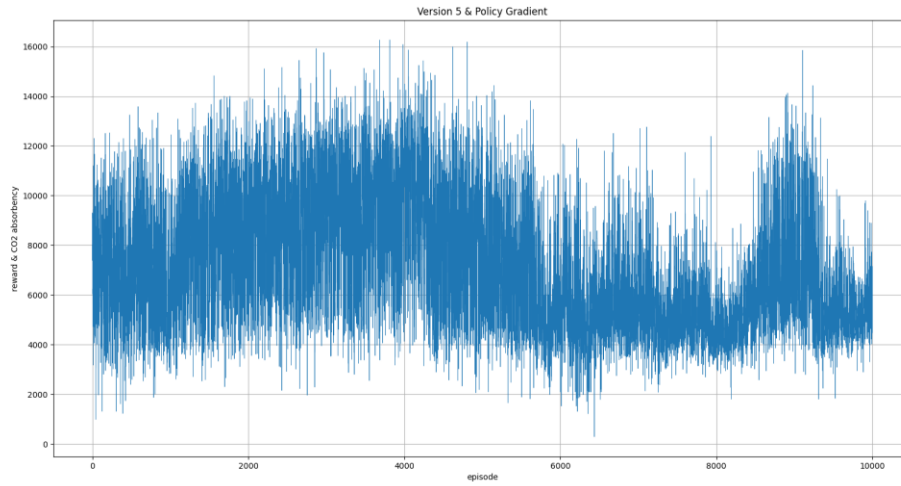
Seed: 1

Mean reward: 7097.7038

Standard deviation of reward: 2762.0682

Maximal reward: 16268.7128

Minimal reward: 285.7807



Seed: 2

Mean reward: 10484.1143

Standard deviation of Learning: 4375.0319

Maximal reward: 15975.7326

Minimal reward: 266.9626

