INL/CON-23-73367-Revision-0

Quantifying Uncertainty of Deep Reinforcement Learning Based Decision Making for Operations and Maintenance of Nuclear Power Plant

July 2023

Changing the World's Energy Future

Ryan Matthew Spangler, Daniel Cole

INL/CON-23-73367-Revision-0

Quantifying Uncertainty of Deep Reinforcement Learning Based Decision Making for Operations and Maintenance of Nuclear Power Plant

Ryan Matthew Spangler, Daniel Cole

July 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-NE0008909

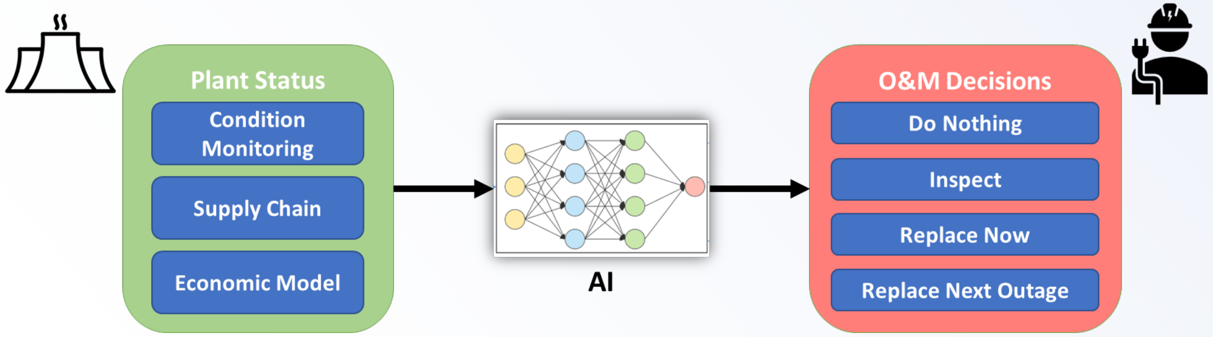
NPIC HMIT 2023

PSA 2023
18th International Probabilistic Safety Assessment and Analysis

Quantifying Uncertainty of Deep Reinforcement Learning-Based Decision Making for Operations and Maintenance of Nuclear Power Plant

Ryan Spangler

*Ph.D. Candidate – University of Pittsburgh**
Research Scientist – Idaho National Laboratory



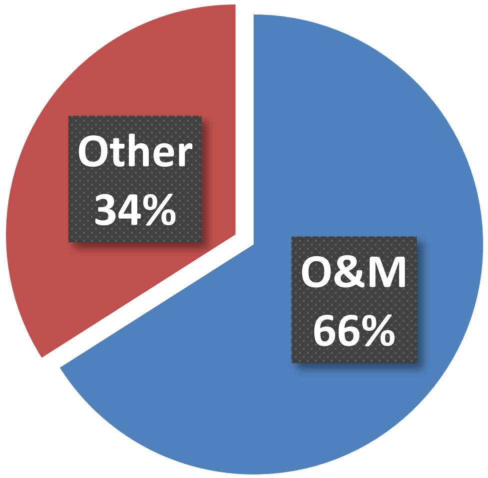
ANS

**this work completed while at University of Pittsburgh*

University of Pittsburgh

Motivation – Although nuclear power is reliable, it remains costly due to high operations and maintenance spending

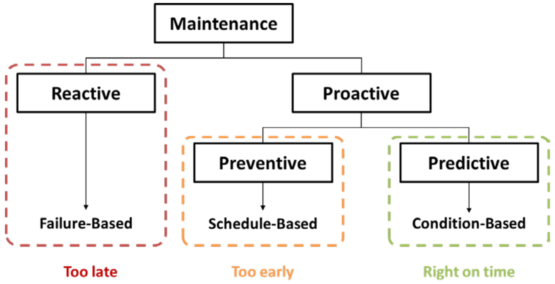
Total Yearly Operating Budget



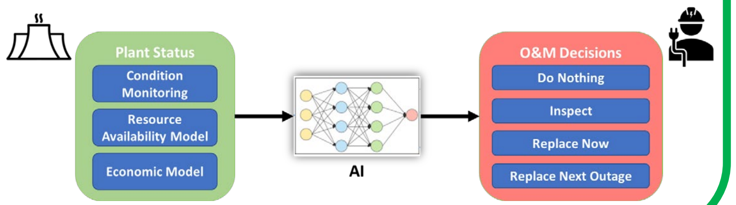
<u>High-Cost Asset Management</u>

- Unexpected maintenance/shutdown
- Overly conservative maintenance
- High staffing levels

Labor-Centric Preventive Maintenance



Risk-Informed Predictive Maintenance



"Information has no value at all unless it has the potential to change a decision"

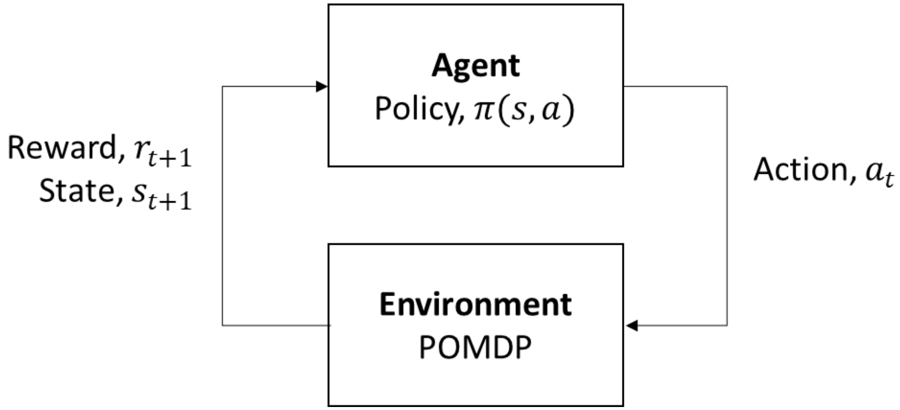
Sam L. Salvage, *The Flaw of Averages*

University of Pittsburgh

ANS

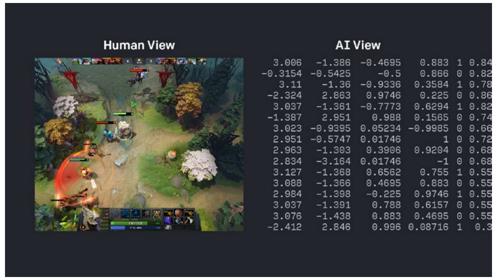
Approach – Using reinforcement learning techniques, we can evaluate multiple decisions for several components over time

Reinforcement Learning



$$\overbrace{Q(s,a) = r(s,a)}^{\text{Immediate reward}} + \underbrace{\gamma}_{\text{Discount factor}} \overbrace{\max_{a} Q(s',a)}^{\text{Max Q-value for next state}}$$

Dota 2 (OpenAI)



Go (Google DeepMind)

DRL Strengths:

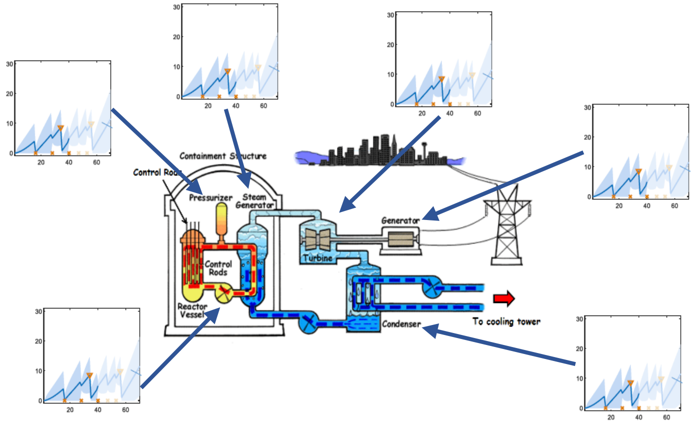
- **Long forecast horizons**
- **Uncertainty and partial observability**
- **Large action and decision spaces**
- **Multi-agent cooperation**

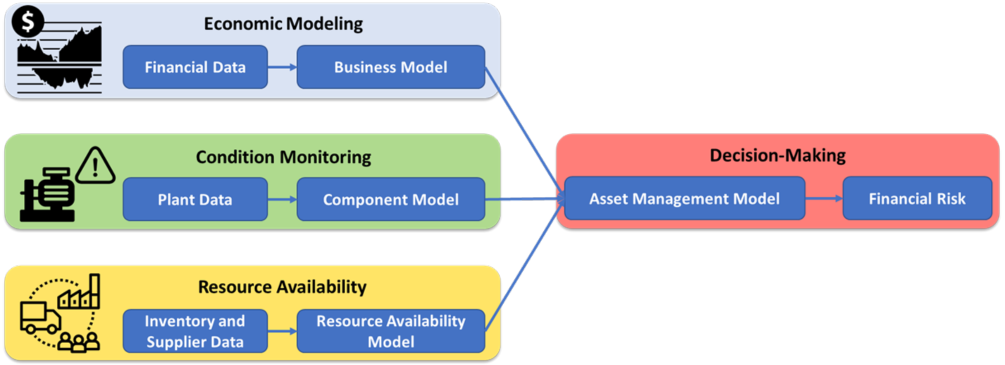
University of Pittsburgh

ANS

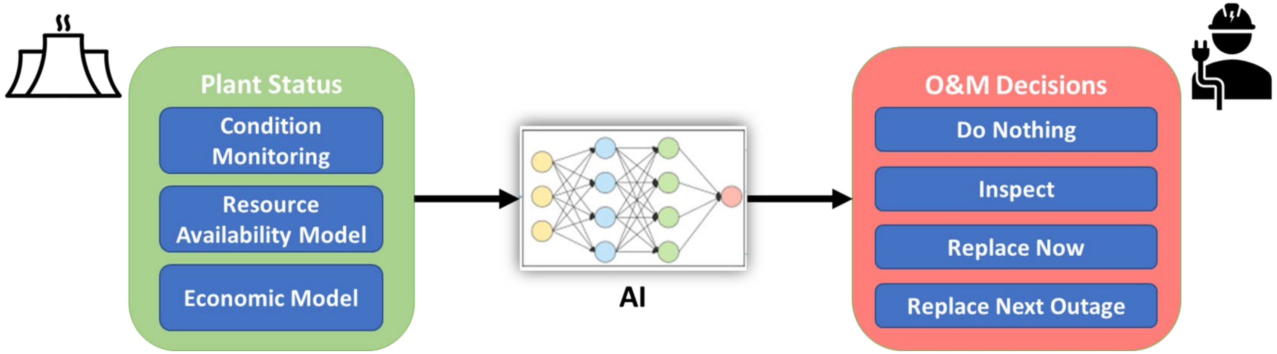
Approach – We need a framework for assessing risk and forecasting decision-making for better asset management

Reliability and Decision Modeling





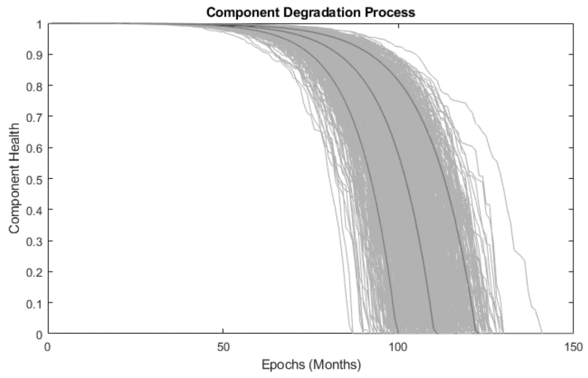
Decision Making

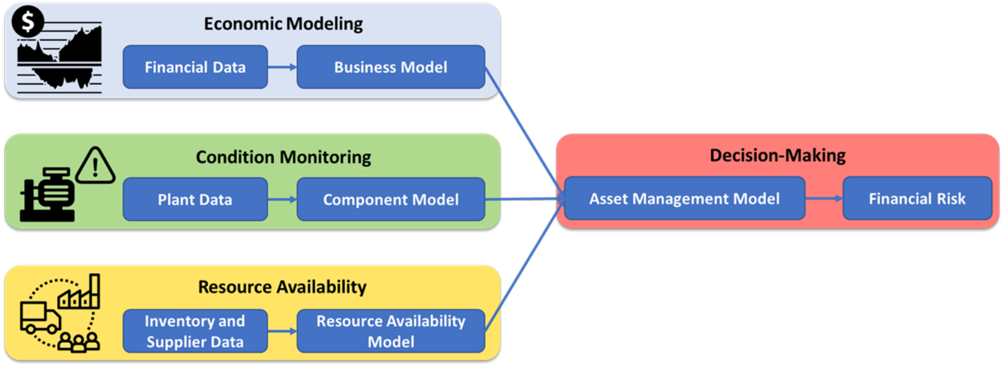


Approach – We need a framework for assessing risk and forecasting decision-making for better asset management

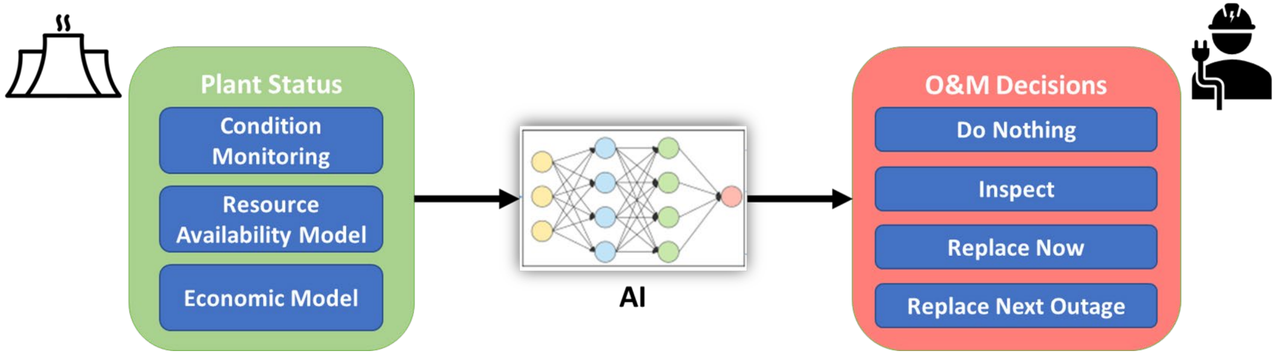
Reliability and Decision Modeling

Stochastic Degradation Process





Decision Making



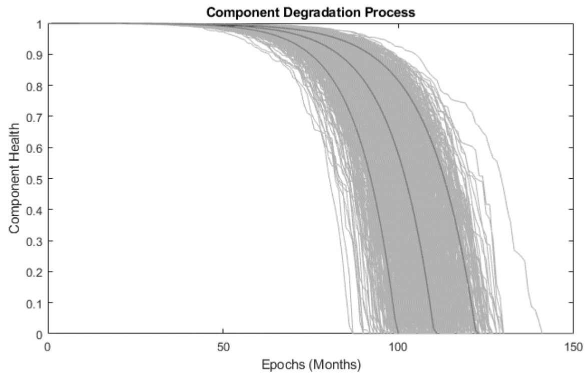
University of Pittsburgh

ANS

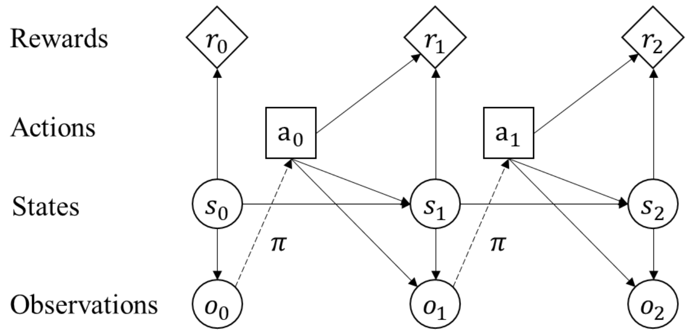
Approach – We need a framework for assessing risk and forecasting decision-making for better asset management

Reliability and Decision Modeling

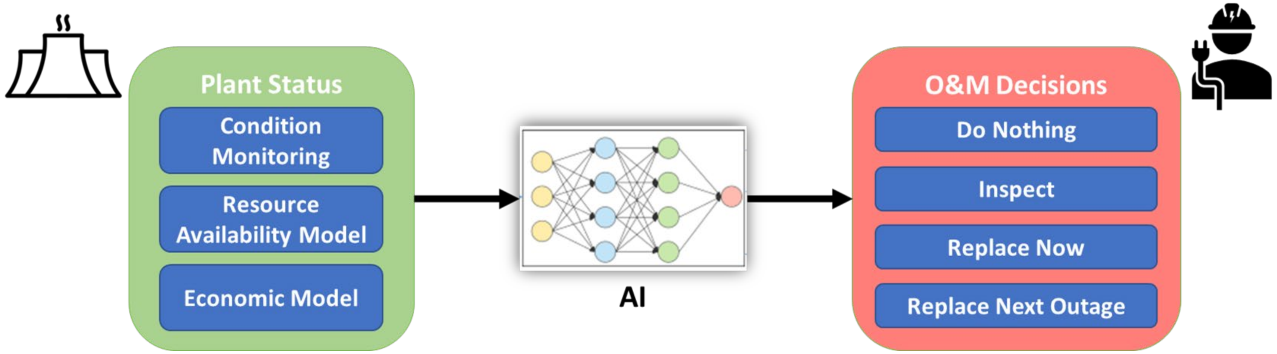
Stochastic Degradation Process



Partially Observable Markov Decision Process



Decision Making



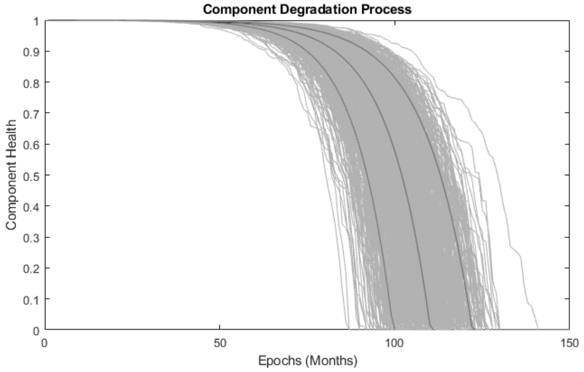
University of Pittsburgh

ANS

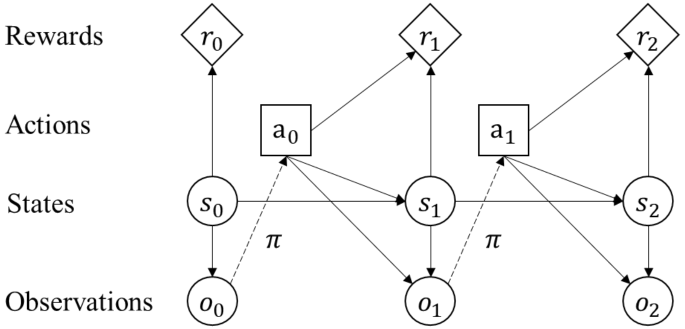
Approach – We need a framework for assessing risk and forecasting decision-making for better asset management

Reliability and Decision Modeling

Stochastic Degradation Process



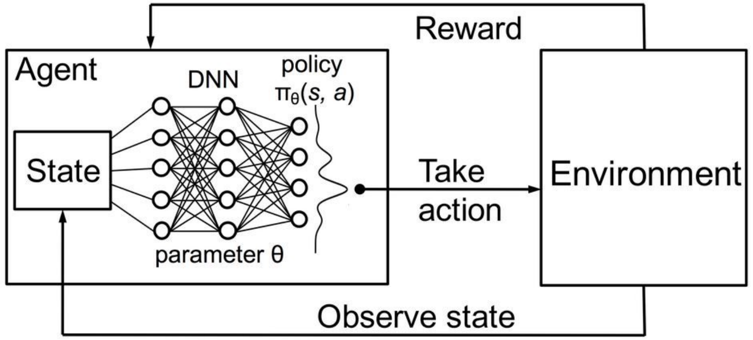
Partially Observable Markov Decision Process



Decision Making

Deep Reinforcement Learning

- Learns through trial and error
- Neural network maps states to actions
- Approximates optimal policy

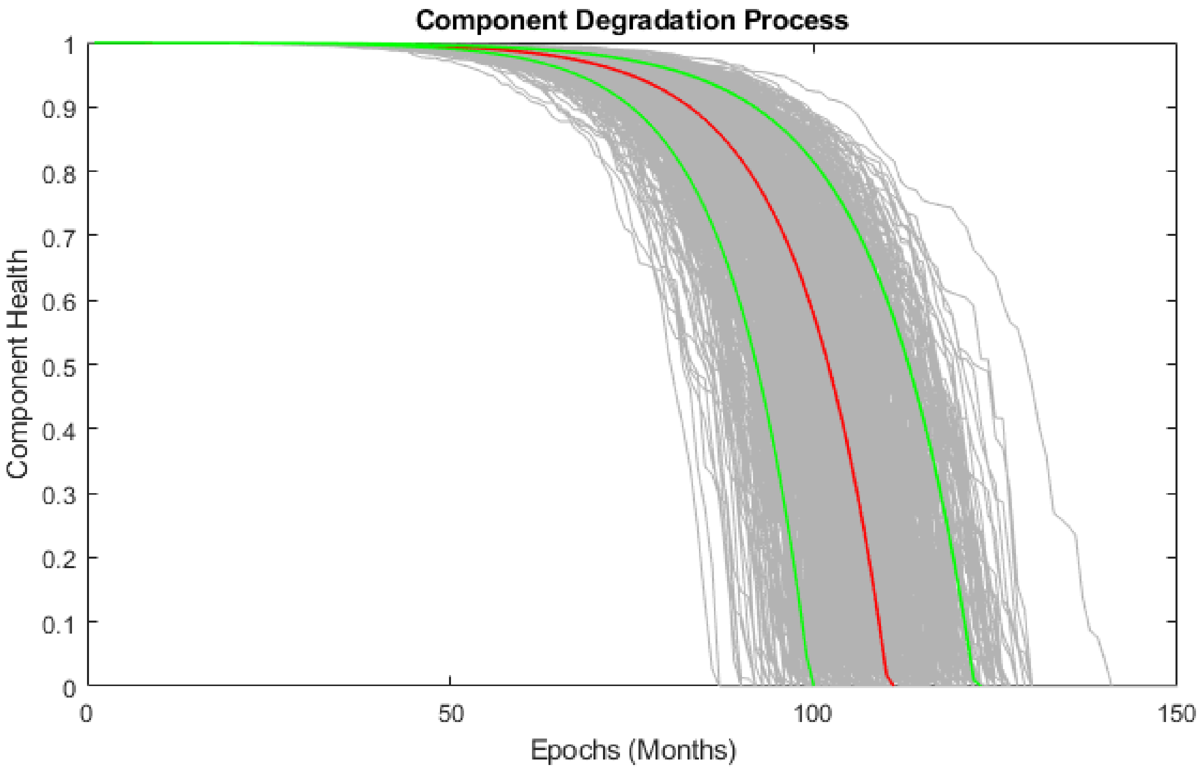


University of Pittsburgh

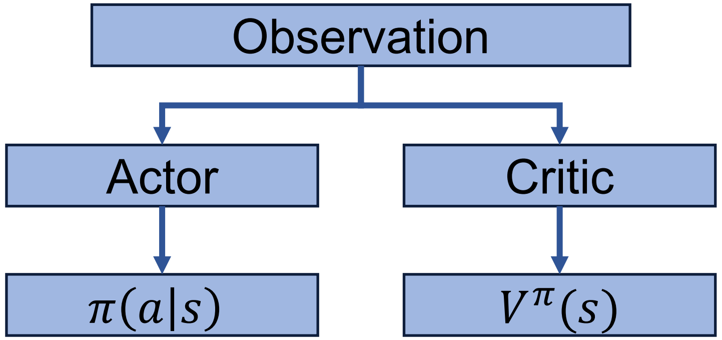
ANS

Environment – To test the reinforcement learning algorithm, we created an environment with one degrading component with inventory management

State:	(1) Component Health (2) Inventory (3) Outage Information
Maintenance Actions:	(1) Do nothing (2) Repair (3) Replace
Inv. Actions:	(1) Do nothing (2) Order spare
Costs:	Storage = -1 Repair = -5 Replace = -15 Unplanned shutdown = -100

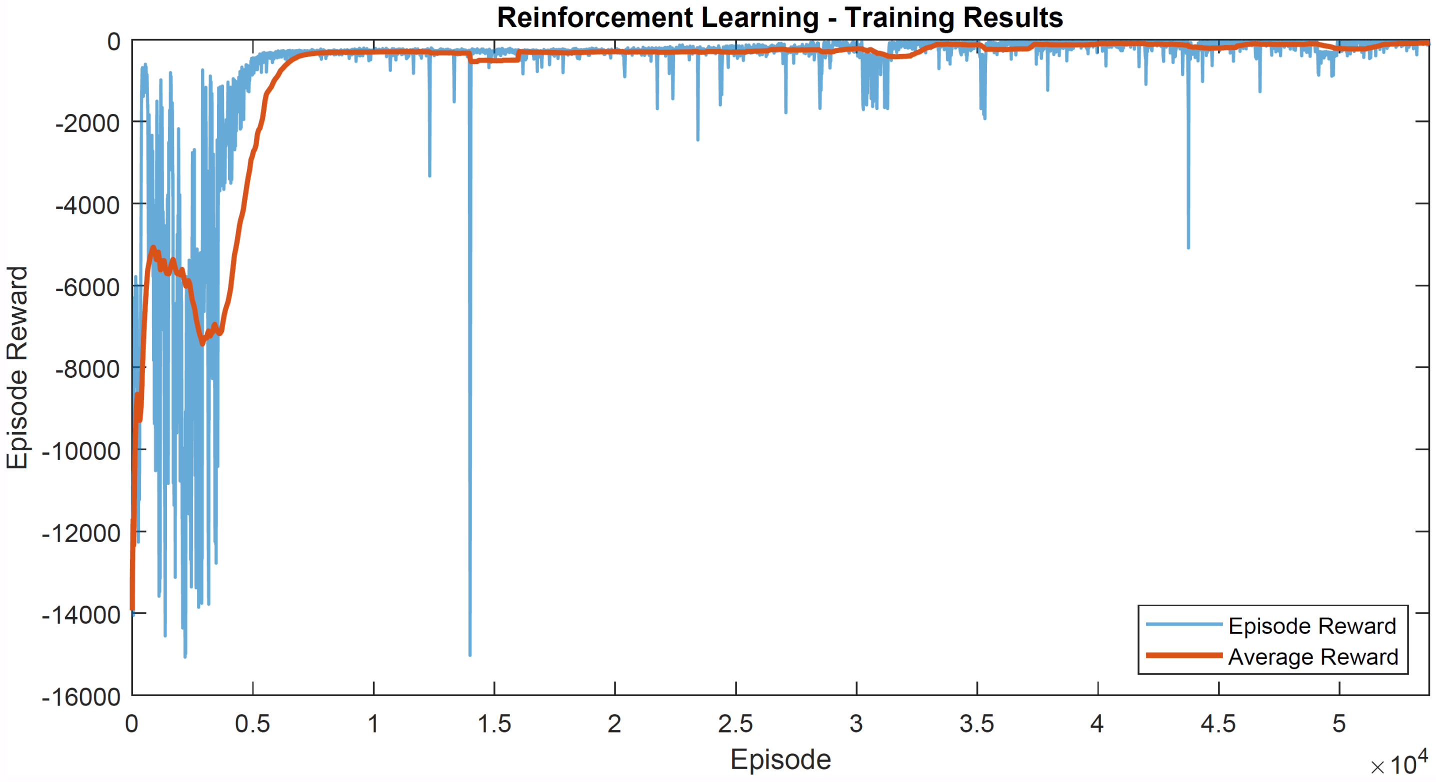


Training – The agent was successfully trained to make maintenance and inventory decisions, minimizing overall lifecycle costs

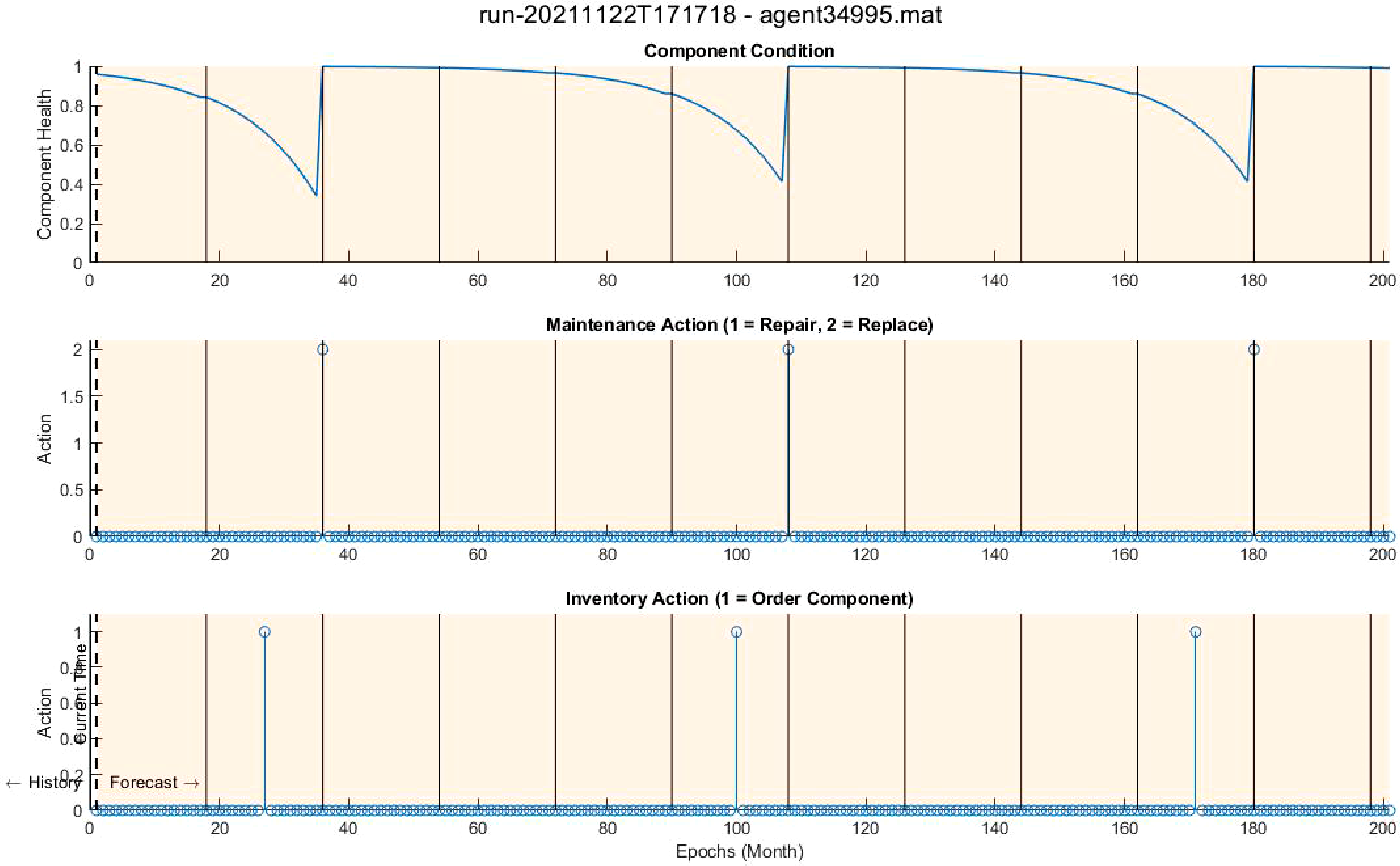


Hyperparameters

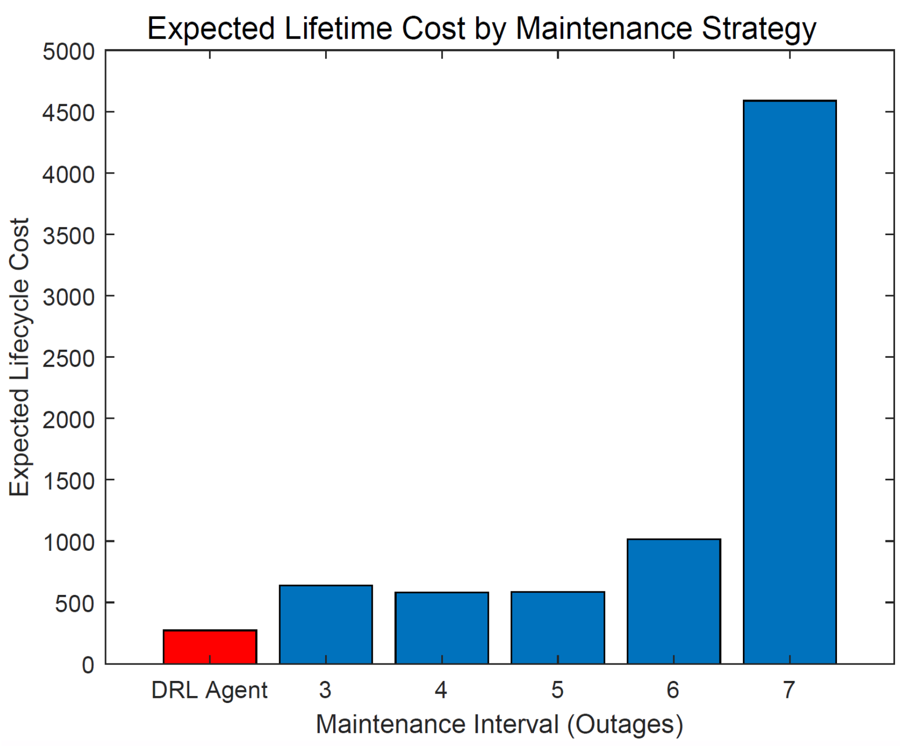
Parameter	Value
Algorithm	A2C
Network	Actor-Critic
Layers	2
Neurons/Layer	300
Discount Factor	0.999
Learning Rate	0.001
Optimizer	Adam

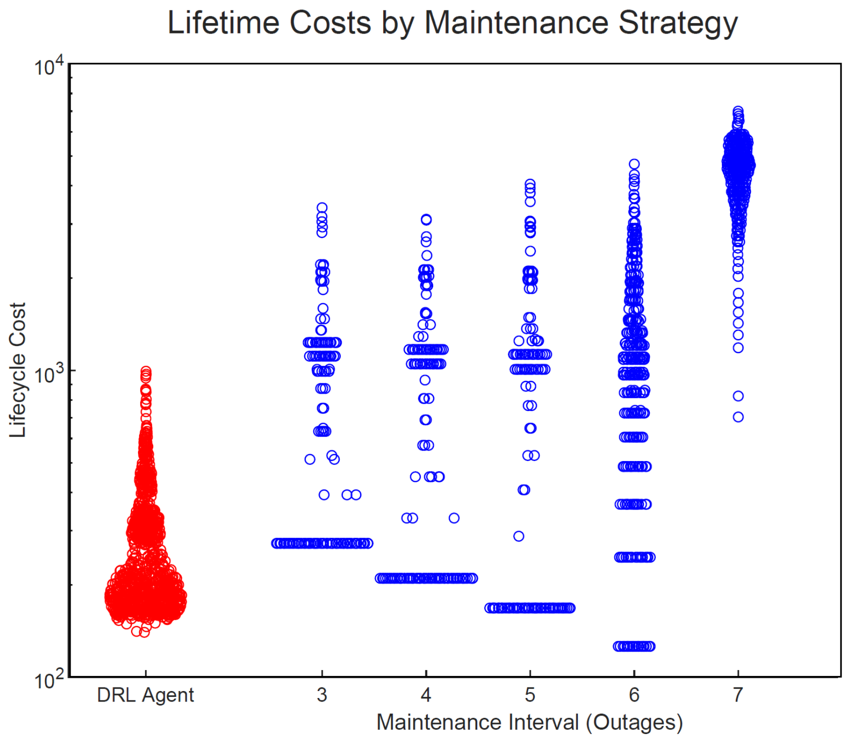


Results – Using the trained agent, we can evaluate its actions and effect on expected lifetime O&M costs



Results – The trained agent was simulated for 60 years and lowered lifecycle costs when compared to time-based strategies



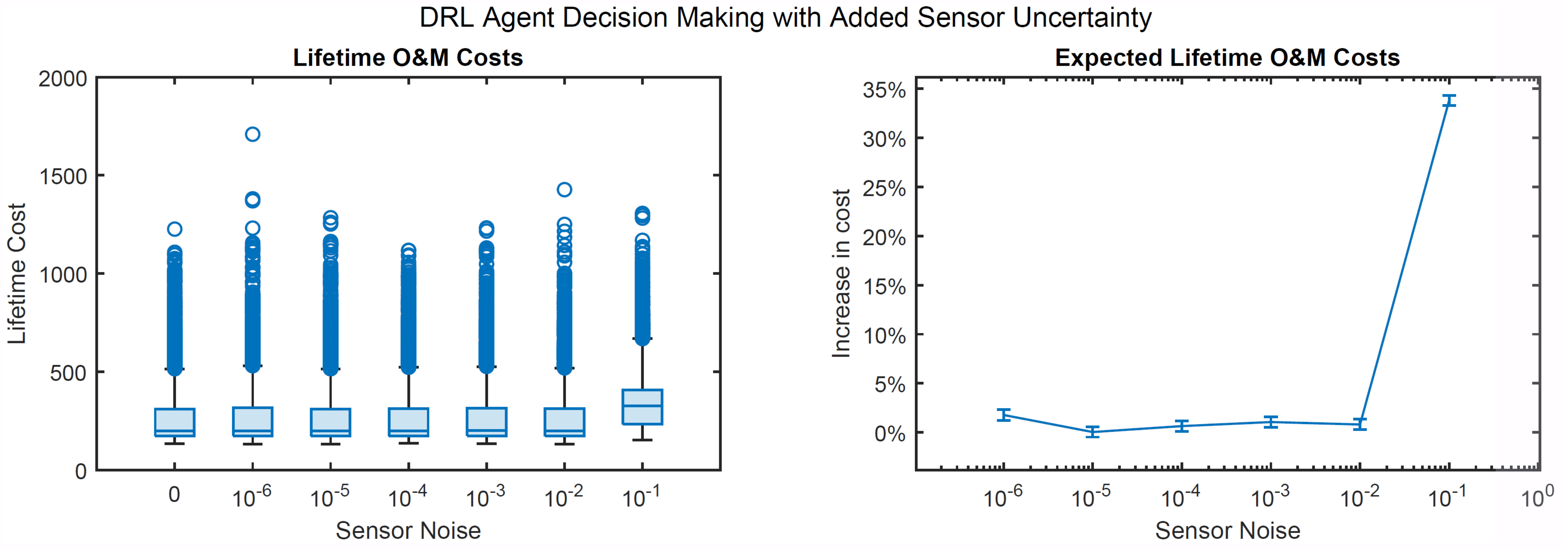


53% cost reduction in expected lifetime costs

University of Pittsburgh

ANS

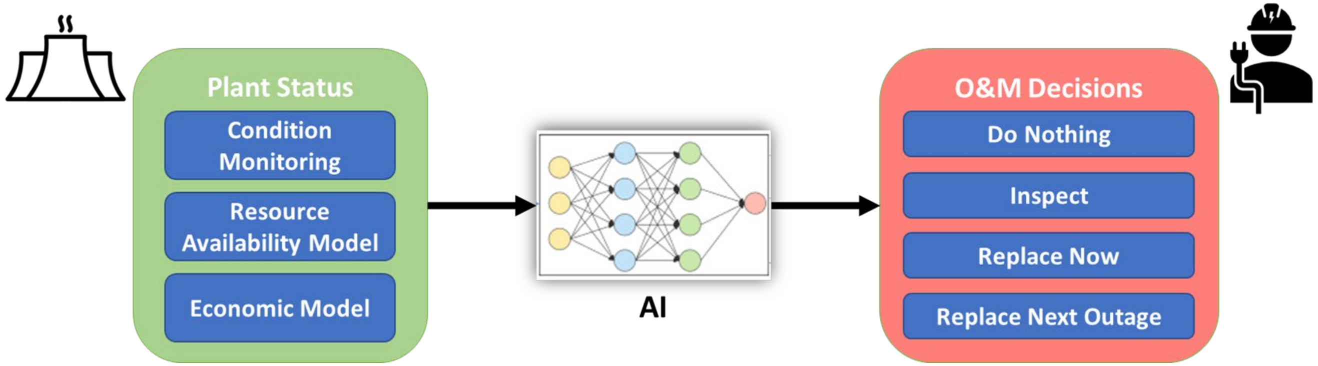
Uncertainty – By adding observation noise, we can test the agent's performance with partial observability and uncertainty



University of Pittsburgh

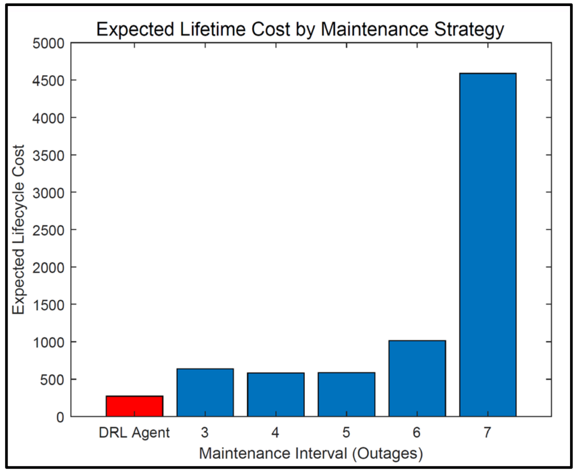
ANS

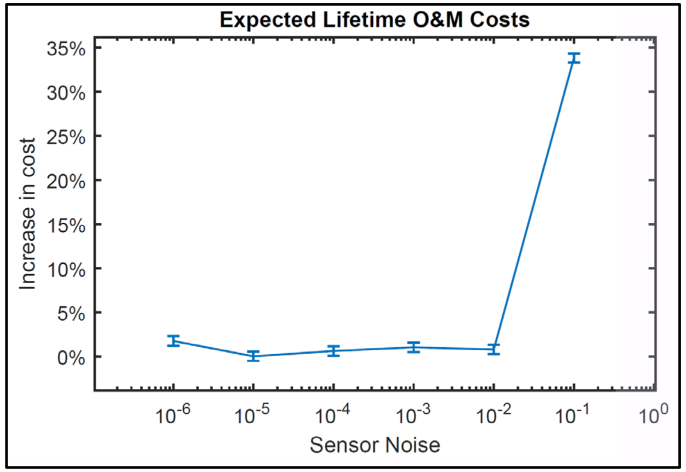
Conclusion – By using deep reinforcement learning, we can create optimal policies that are robust to uncertainty



Ryan Spangler
ryan.spangler@inl.gov

Daniel G. Cole
dgcole@pitt.edu





University of Pittsburgh

This work was supported through U.S. Department of Energy Office of Nuclear Energy's Nuclear Energy University Program DE-NE0008909 under the Nuclear Energy Enabling Technologies Advanced Sensors and Instrumentation Program.

ANS