Weight Initialization and Lamarckian Inheritance & Distributed Island Repopulation on Neuroevolution

Zimeng Lyu, AbdElRahman ElSaid, Joshua Karns, Mohamed Mkaouer, and Travis Desell

Rochester Institute of Technology, Rochester, NY 14623, USA

An Experimental Study of Weight Initialization and Lamarckian Inheritance on Neuroevolution

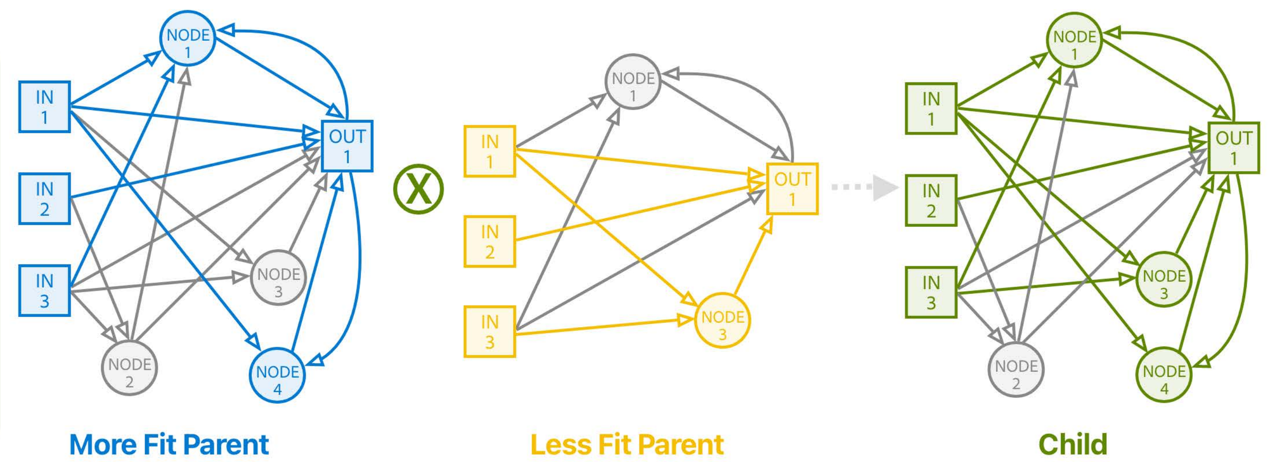
Explore the difference between the state-of-the-art Xavier and Kaiming methods, and novel Lamarckian weight inheritance for weight initialization during crossover and mutation operations.

Crossover

If both parents have the same component, the weight is generated by stochastic line search. Given a random number $r \sim \mathcal{U}[-0.5, 1.5]$, a child's weight w_c is set to:

$$w_c = r(w_{p2} - w_{p1}) + w_{p1}$$

Where w_{p1} is the weight of more fit parent, w_{p2} is the weight of the less fit parent.



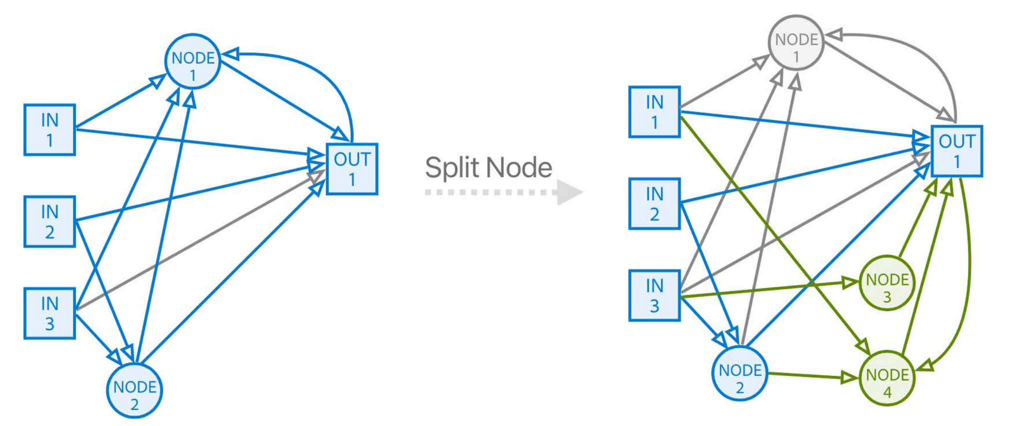
Mutation

Mutation types: *add edge / recurrent edge / node, enable edge / node, disable edge / node, split /merge node, clone.*

Mutations such as *add edge / node, split / merge node* produces new edges or nodes. The weight $\mathcal{W}$ for those new components is calculated by:

$$\mathcal{W} \sim \mathcal{N}(\mu_p, \sigma^2_p)$$

Where μ_p and σ^2_p are mean and variance from parent p. Other weights are directly copied from parent.

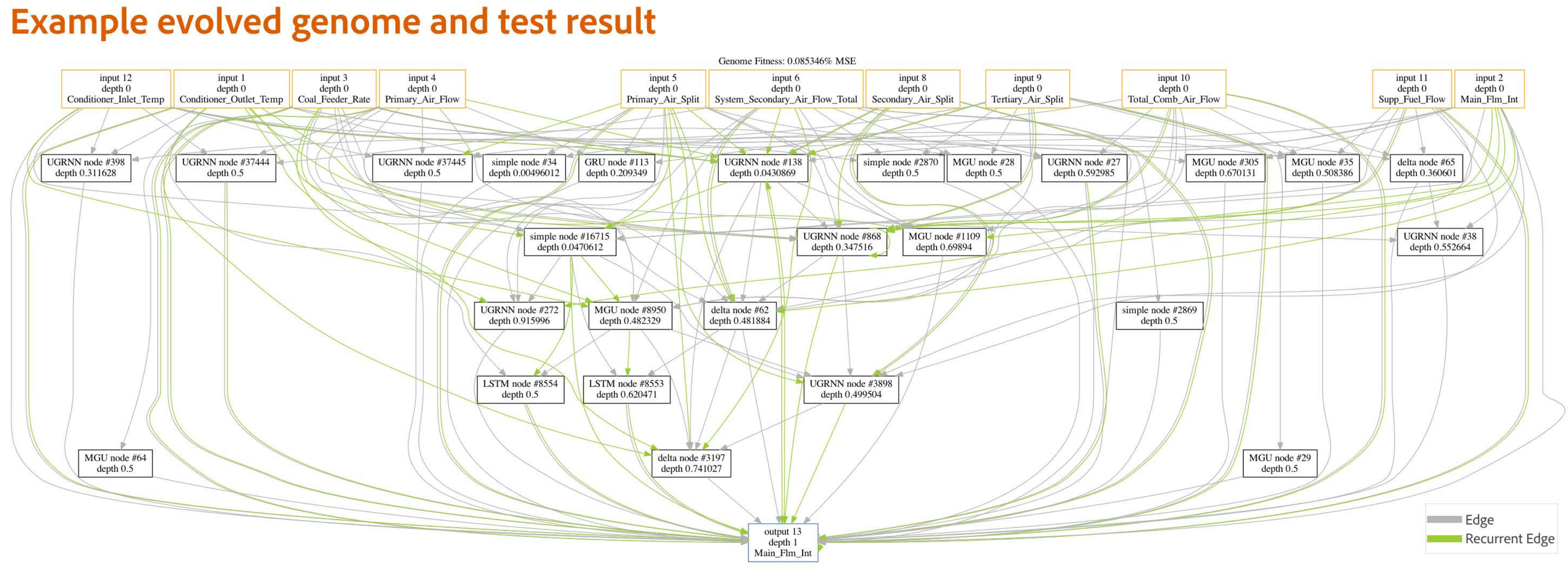


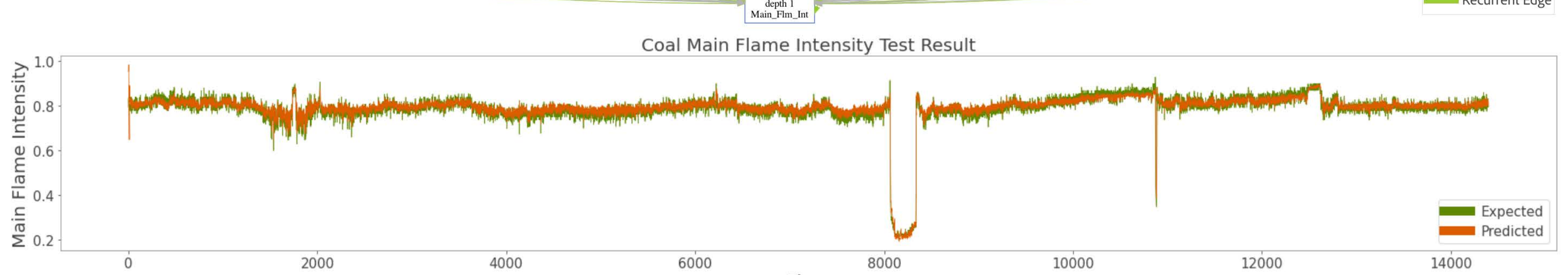
Experiments

Test a combination of different weight initialization and Lamarckian weight inheritance method on coal fired power plant, flight, and Wind turbine engine dataset:

- *Initial population: Xavier, Kaiming, and uniform random U ~ [-0.5, 0.5]*
- *Crossover: Xavier, Kaiming, uniform random, and Lamarckian*
- *Mutation: Xavier, Kaiming, uniform random, and Lamarckian*

Example evolved genome and test result



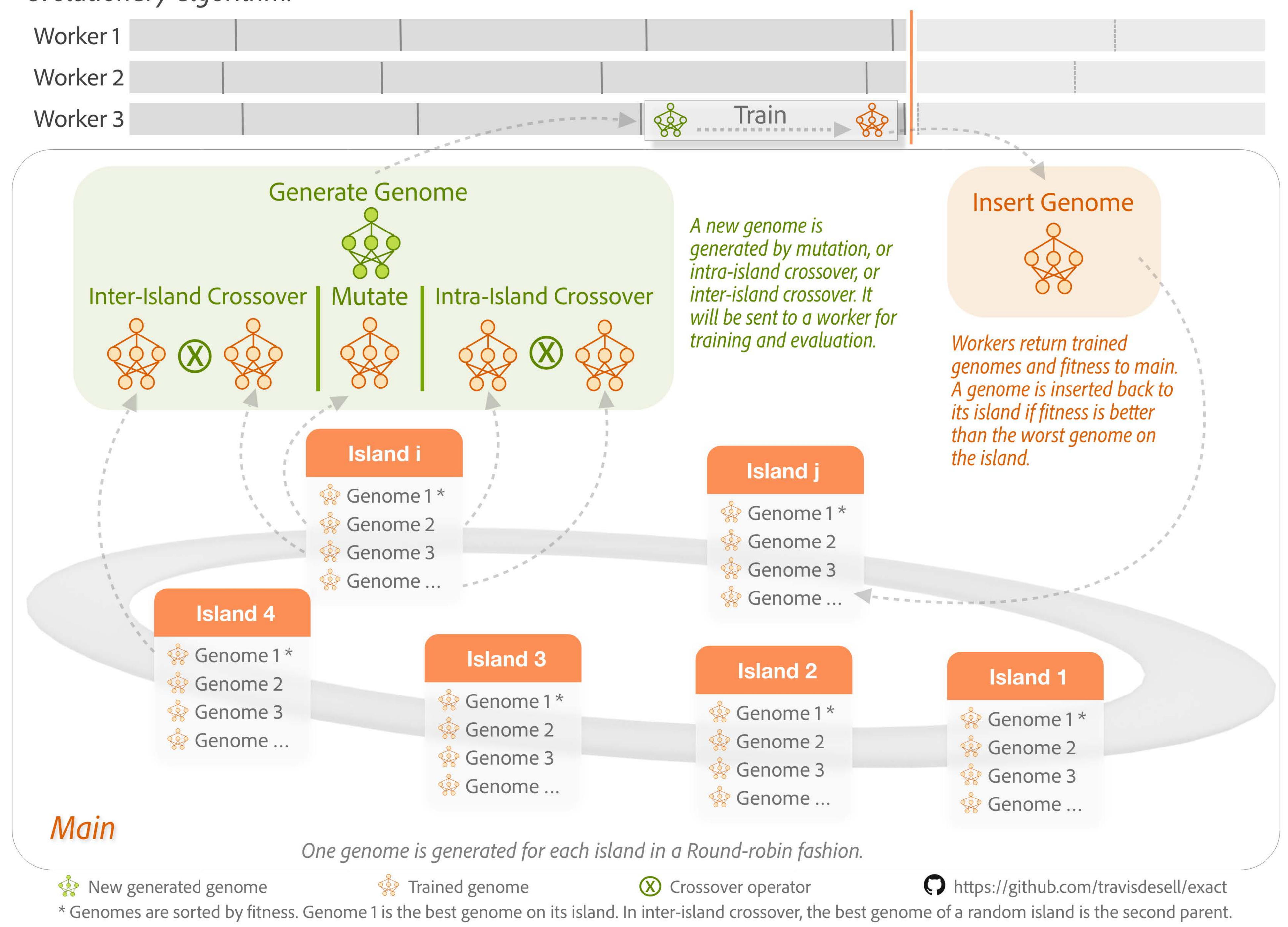


Conclusion

- Xavier and Kaiming methods outperform uniform random weight initialization method.
- All of the best performing test cases use Lamarckian on crossover or mutation or both.
- The *Mann-Whitney U* tests show statistical significance between Lamarckian method and non-Lamarckian methods with $\alpha = 0.05$.
- Utilizing Lamarckian weight inheritance could reduce the number of BP epochs for training.
- Evolved best genomes are reinitialized with Xavier or Kaiming weight initialization and retrained for 3000 epochs, the genomes with evolved weight still outperform the ones with trained weight.

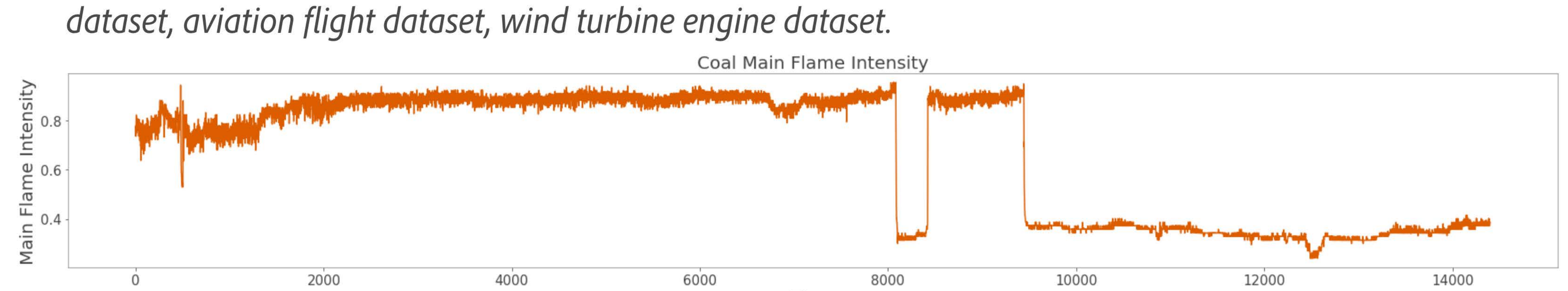
EXAMM Asynchronous Distributed Neuroevolution Strategy

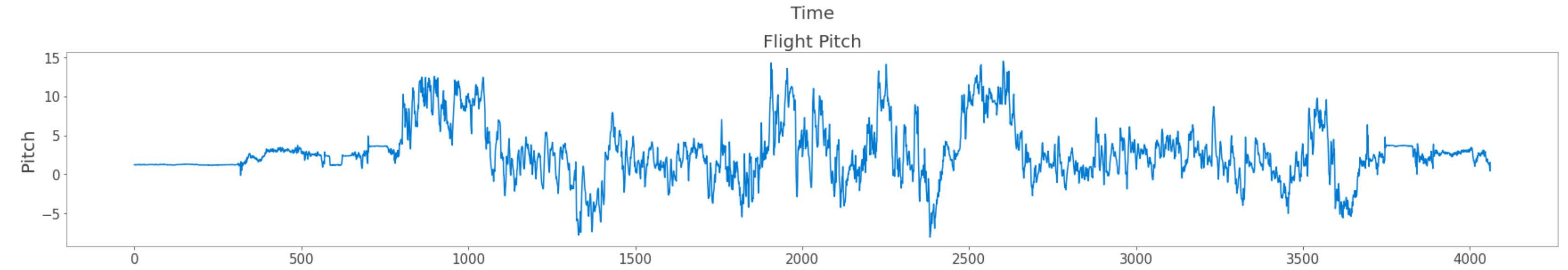
Evolutionary eXploration of Augmenting Memory Models (EXAMM) neuroevolution algorithm, is capable of evolving RNNs with a variety of modern memory cells (e.g., LSTM, GRU, MGU, UGRNN and Delta-RNN cells) as well as recurrent connections with varying time skips through a high performance island based distributed evolutionary algorithm.

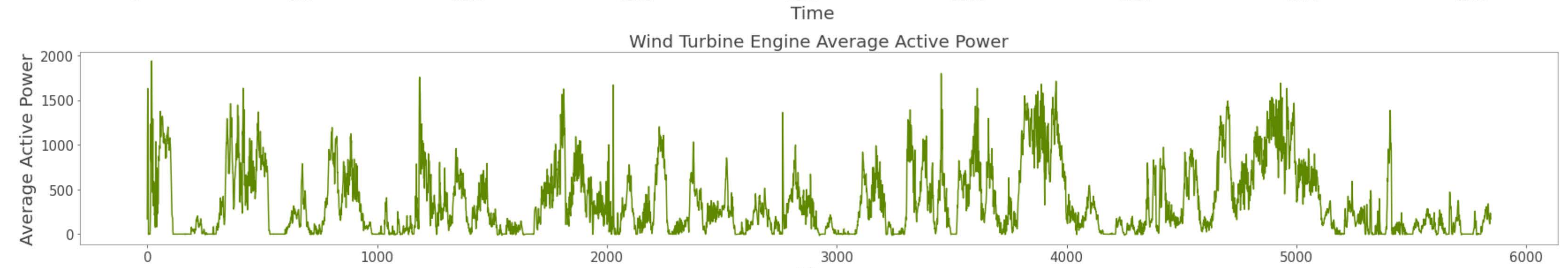


Datasets

Three real world, large-scale, multivariant, non-seasonal time series datasets: coal-fired power plant dataset, aviation flight dataset, wind turbine engine dataset.







Improving Distributed Neuroevolution Using Island Extinction and Repopulation

In distributed Evolutionary Algorithms (EAs), islands may experience stagnation, which prevents their convergence towards better solutions and can result in wasted computation.

Island extinction and repopulation strategy: all members of the worst performing island are erased periodically and repopulated with mutated versions of the global best RNN.

This island based strategy is additionally compared to NEAT's (NeuroEvolution of Augmenting Topologies) speciation strategy.

With statistical significance, results show that in addition to being more scalable, this island extinction and repopulation strategy evolves better global best genomes than both EXAMM's original island based strategy and NEAT's speciation strategy.

Repopulation

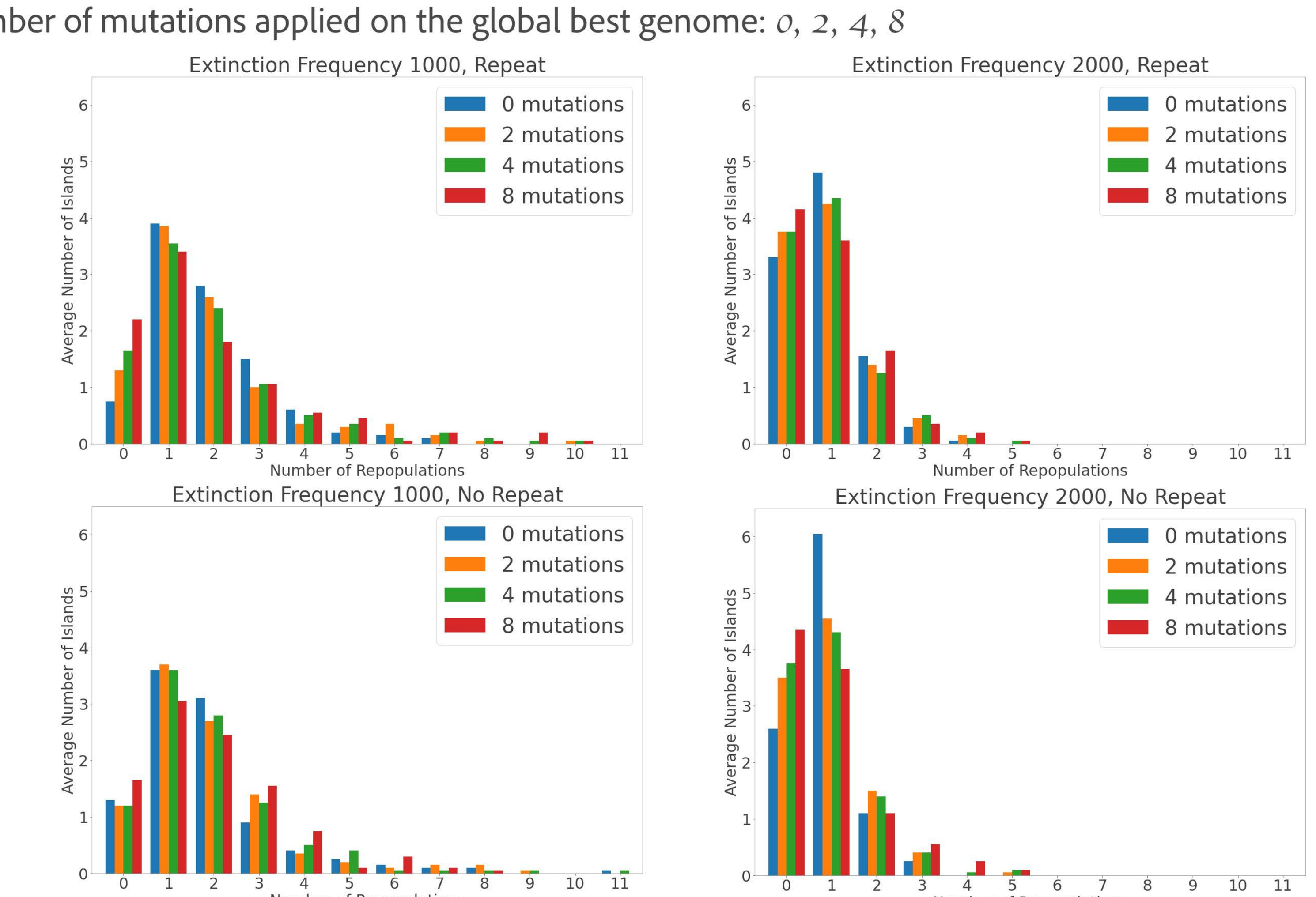
Originally EXAMM only has *initialization* and *filled* status.
The island repopulation strategy has *initialization*, *filled*, and *repopulation* status.
After all islands are filled, EXAMM will periodically repopulate the worst performing island. This worst island is completely erased and then repopulated with new genomes. Each of those new genomes is a child of the global best genome that's mutated m times.

Experiments

The Island extinction and repopulation strategy is tested against EXAMM baseline and NEAT on coal and aviation flight datasets.

A combination of following hyperparameters are tested:

- If an island could be repopulated repeatedly: *repeated repopulation* or *no repeat repopulation*
- Extinction and repopulation frequency: 1000 or 2000 genomes
- Number of mutations applied on the global best genome: $0, 2, 4, 8$



Conclusion

- The island extinction and repopulation strategy outperforms EXAMM baseline and NEAT.
- The *Mann-Whitney U* test shows statistical significance between EXAMM baseline and most of the test cases with $\alpha = 0.05$.
- Extinction frequency at 1000 generally provides better performance because erasing and repopulating islands more frequently could speed up the evolution.
- *No repeated* repopulation has slightly better results because newly repopulated islands has more time to evolve before being evaluated and repopulated again.
- Applying more than one mutations could speed up the evolutionary process, but having too many mutations at once might bring excessive variety to the new island population.

RIT Rochester Institute of Technology