Georgia Tech Research Institute

BathEMADE: Evolutionary Multi-objective Algorithm Design Engine for Bathymetric LIDAR

Jason Zutty
Rodd Talebi
James Rick
Christopher Valenta
Domenic Carr
Gregory Rohling

How is a Machine Learning Algorithm Made?

- Involves a number of steps

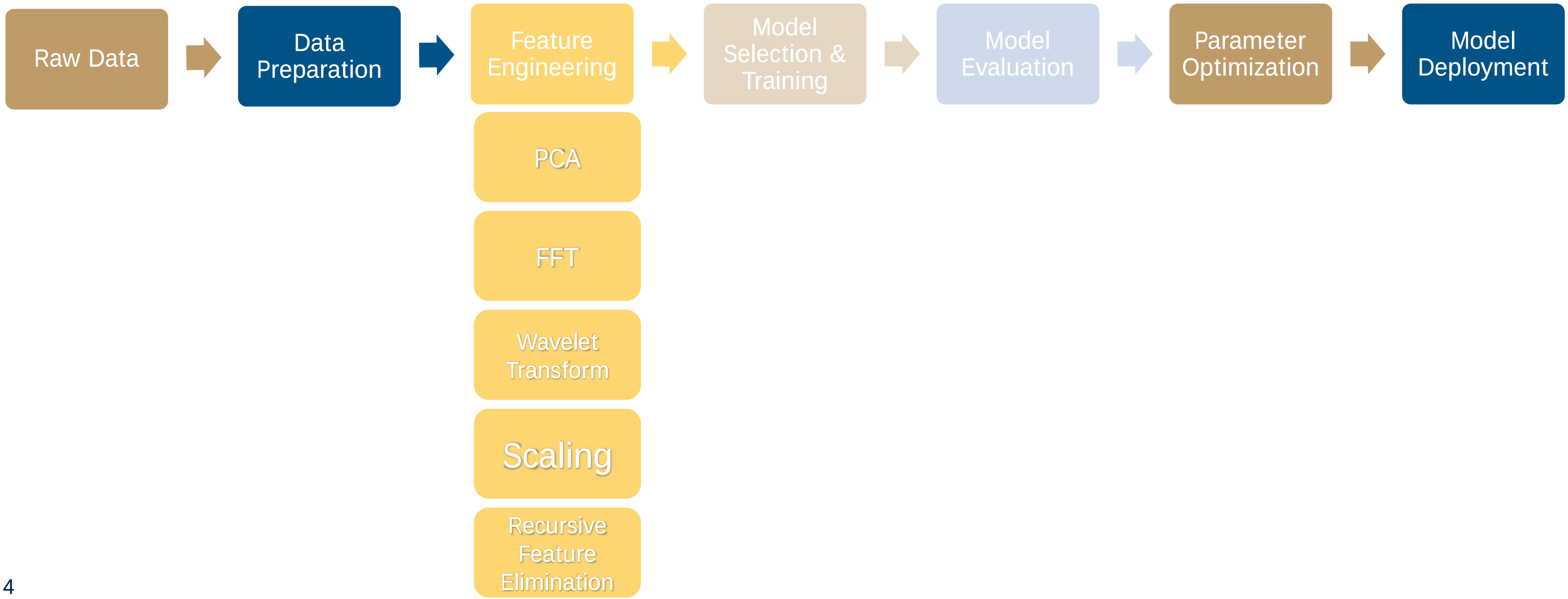
How is a Machine Learning Algorithm Made?

- Involves a number of steps

- Each step represents a choice of one method from a large library of options
- Each choice has a number of parameters that affect the performance
- Overall this is a **time consuming, iterative, and labor intensive process**

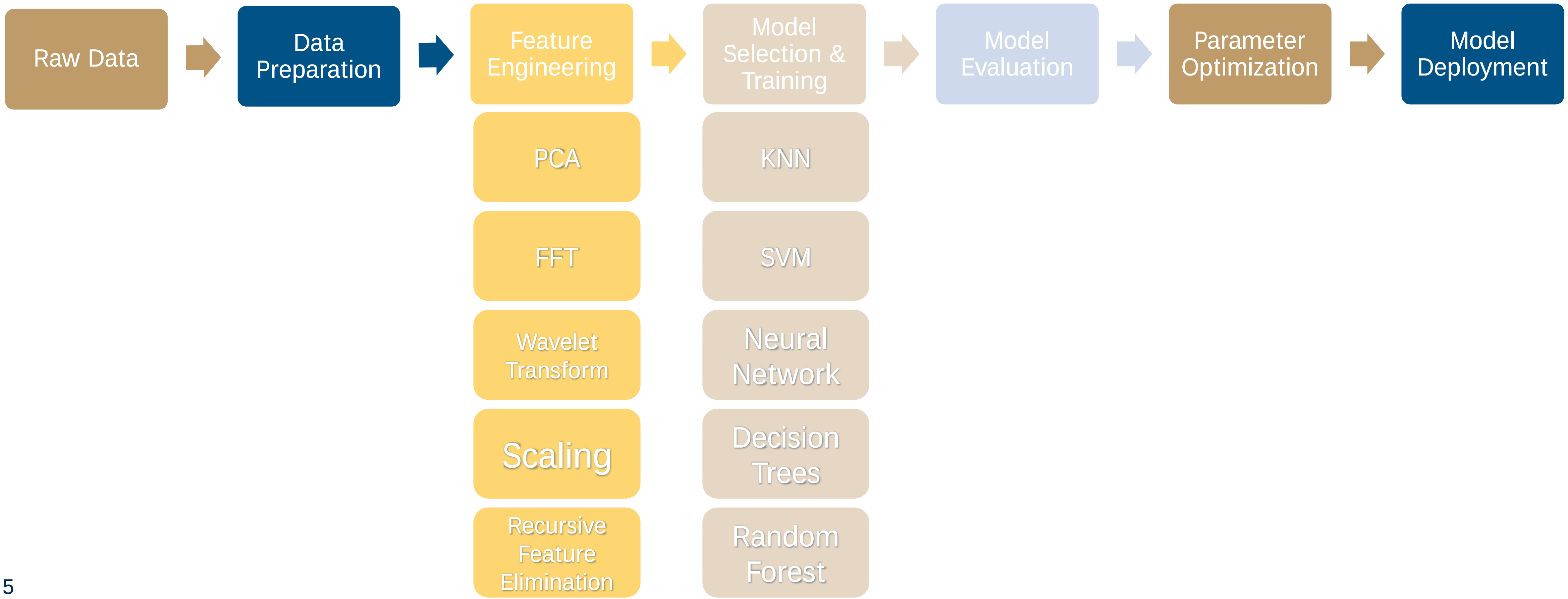
How is a Machine Learning Algorithm Made?

- Involves a number of steps



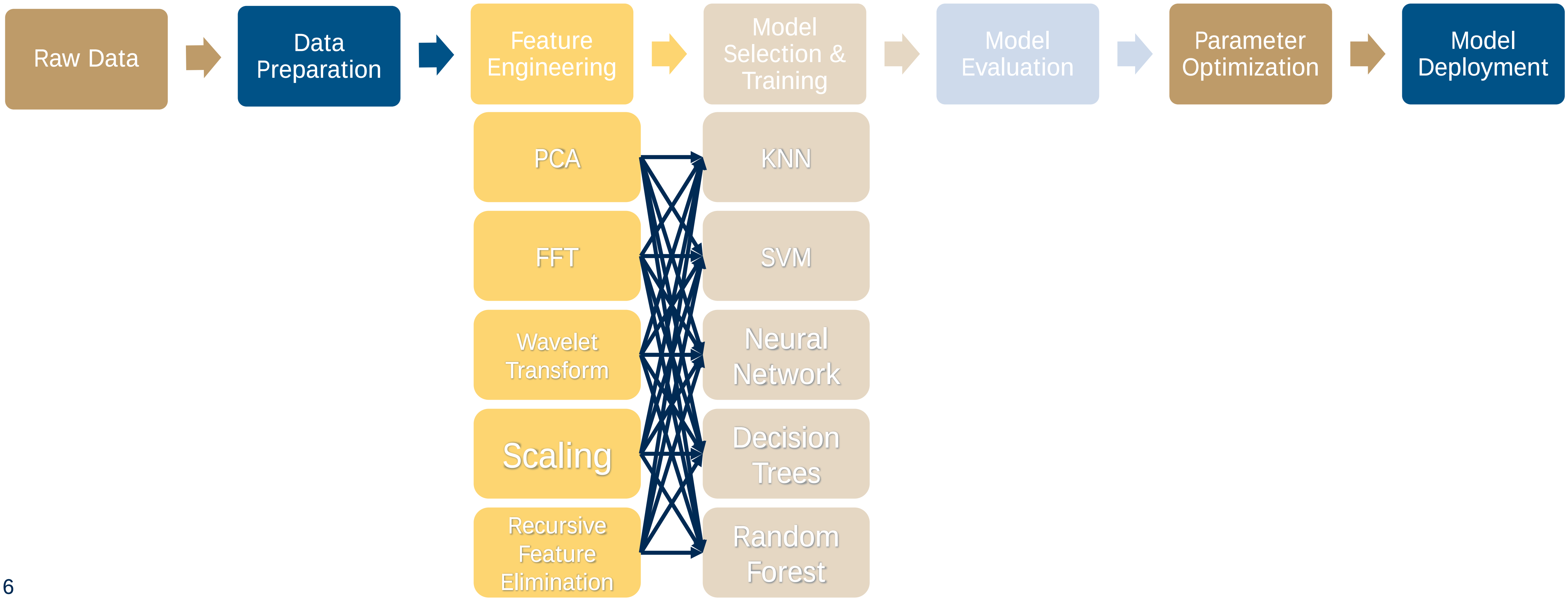
How is a Machine Learning Algorithm Made?

- Involves a number of steps



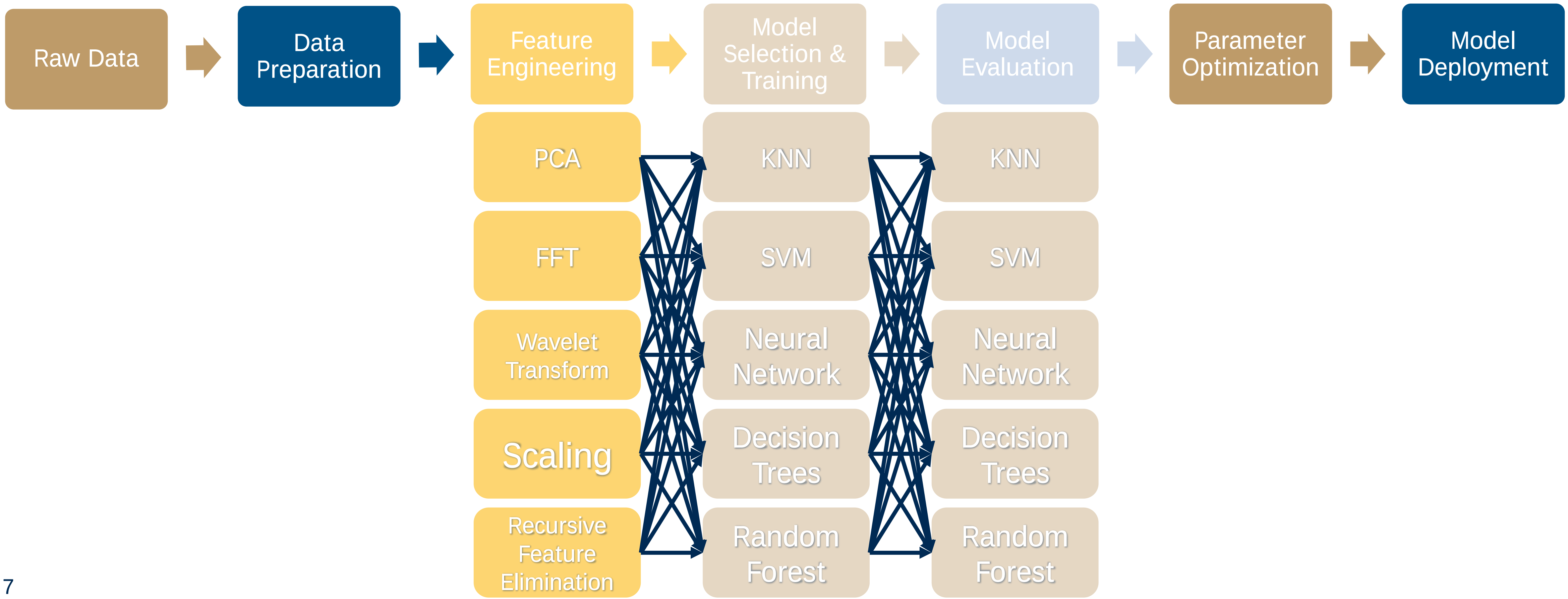
How is a Machine Learning Algorithm Made?

- Involves a number of steps



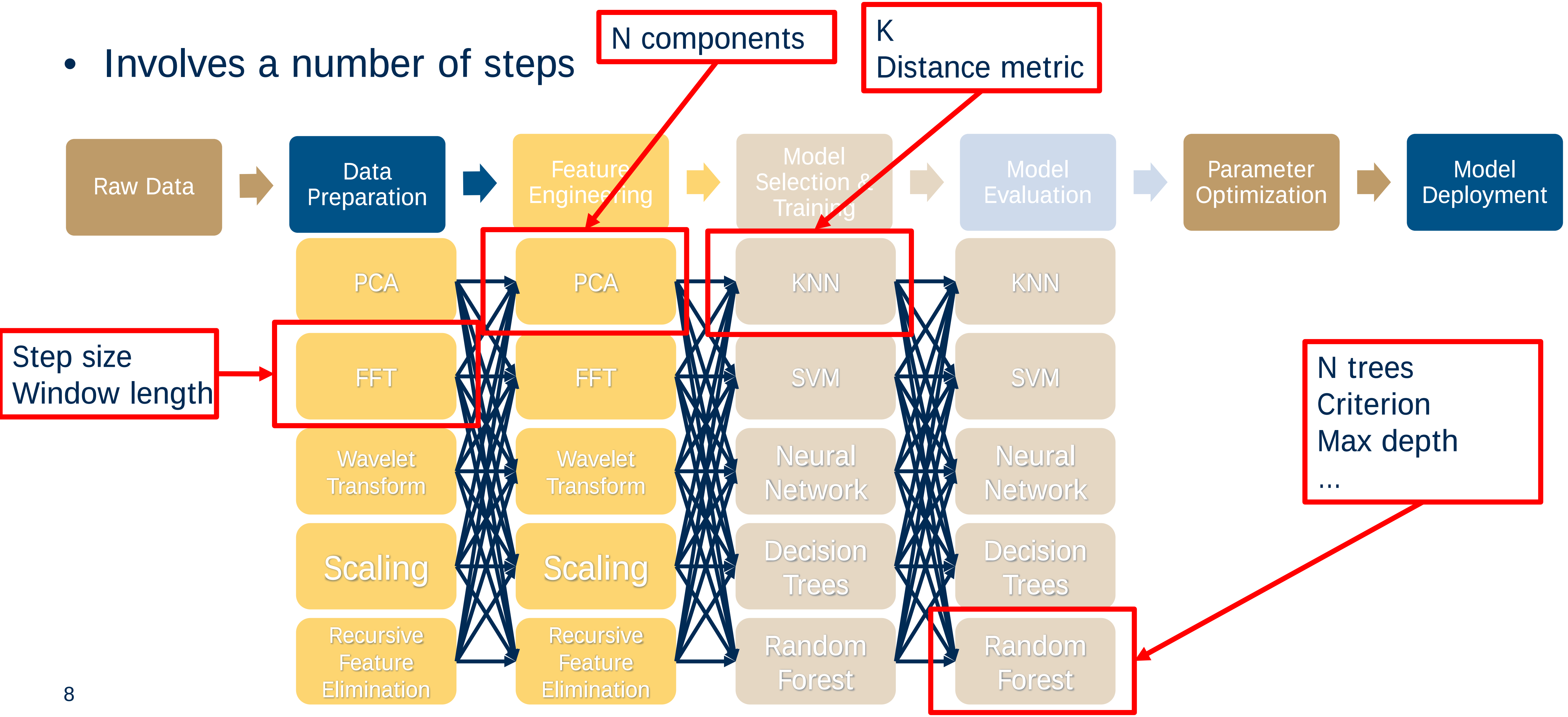
How is a Machine Learning Algorithm Made?

- Involves a number of steps



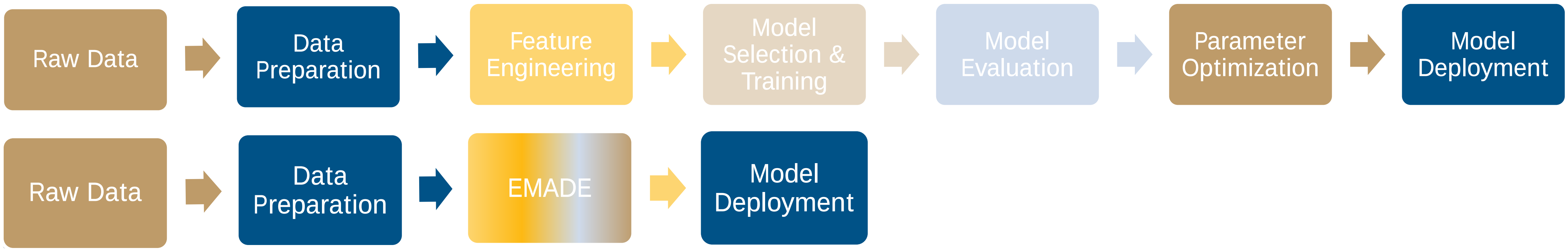
How is a Machine Learning Algorithm Made?

- Involves a number of steps

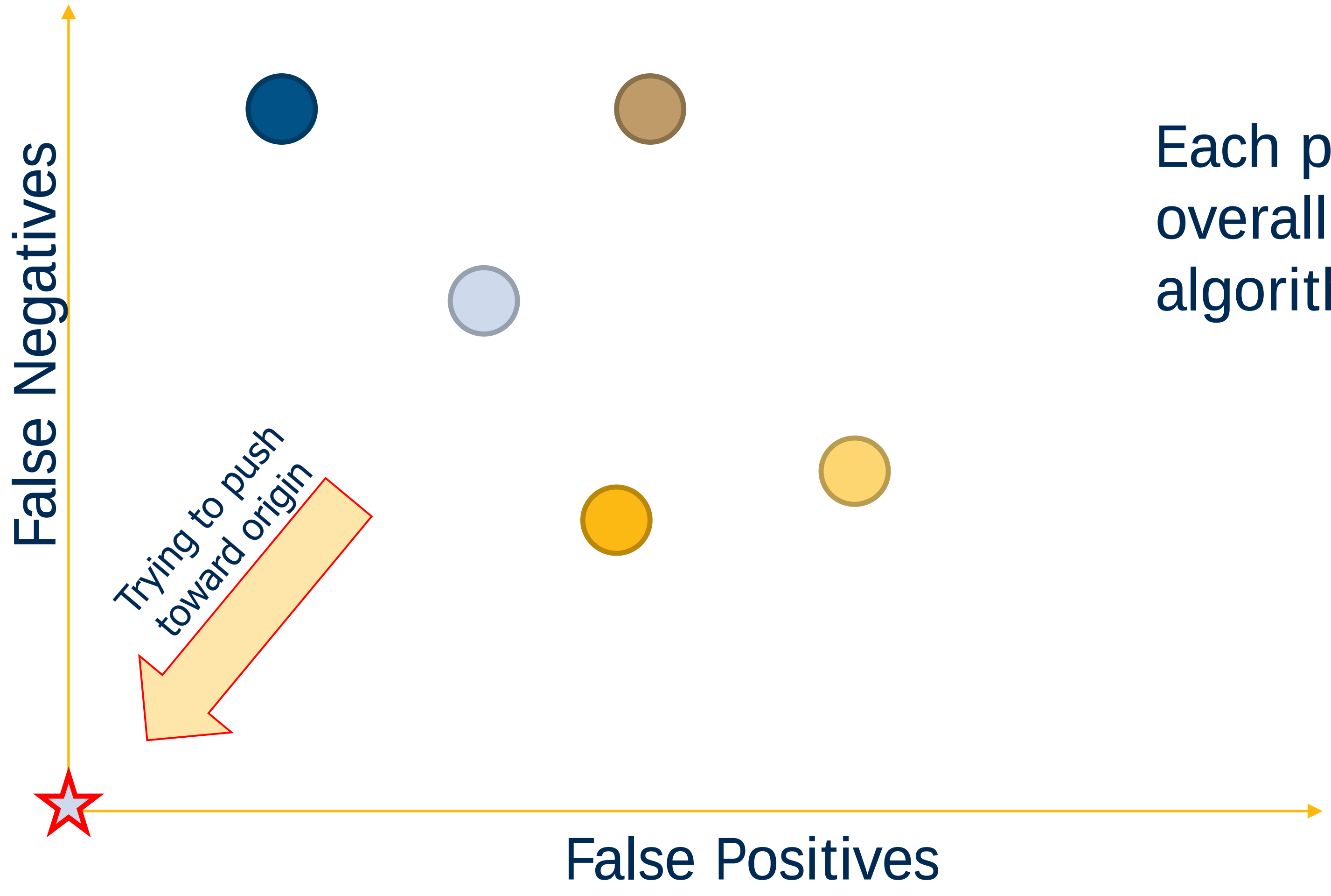


Evolutionary Multi-objective Algorithm Design Engine

- Synthesizes new algorithms from existing building blocks
- Evaluates hundreds of thousands of algorithms in the time a data scientist could try a dozen
- Simultaneously optimizes against multiple performance criteria
- Initialized by the state of the art

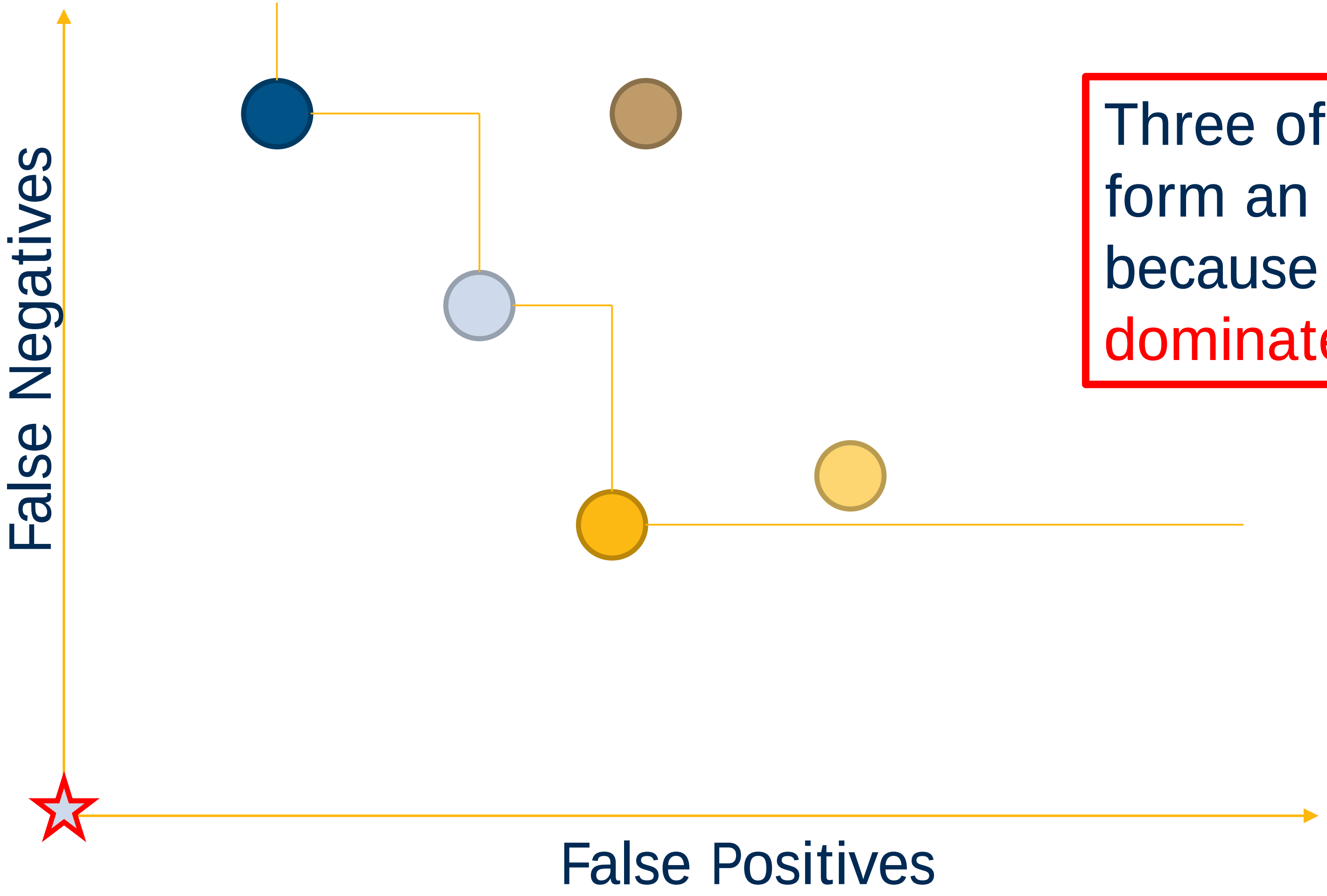


Scoring the Algorithms



Each point represents the overall performance of one algorithm

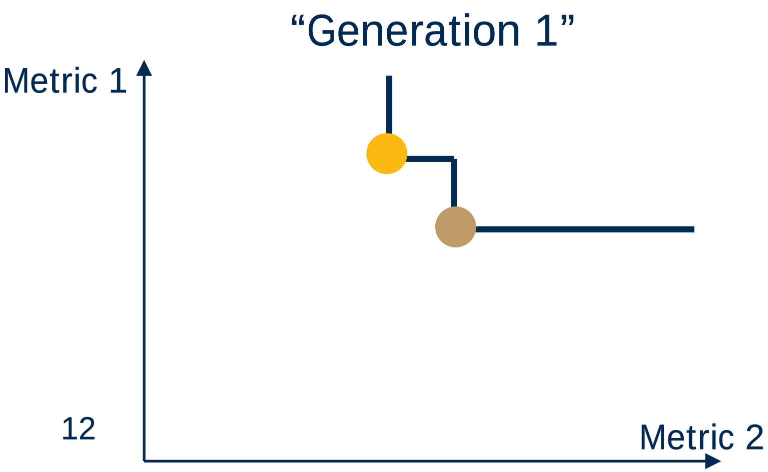
Scoring the Algorithms

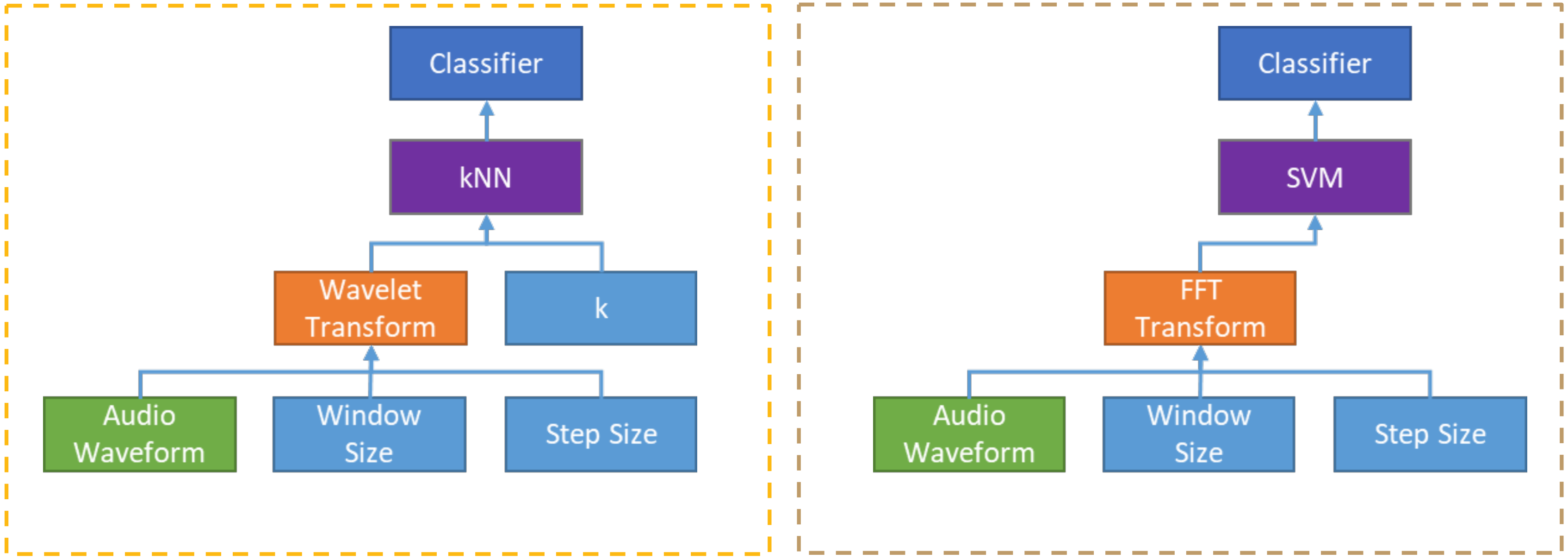


Three of these algorithms form an optimal set because they are non-dominated

Evolutionary Multi-objective Algorithm Design Engine

EMADE applies concepts from biology such as “survival of the fittest” and “DNA” to the world of software to **evolve** algorithms tailored to a specific dataset and problem



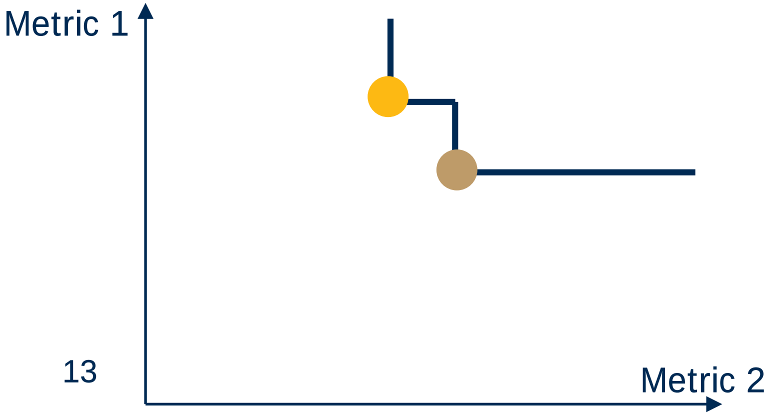


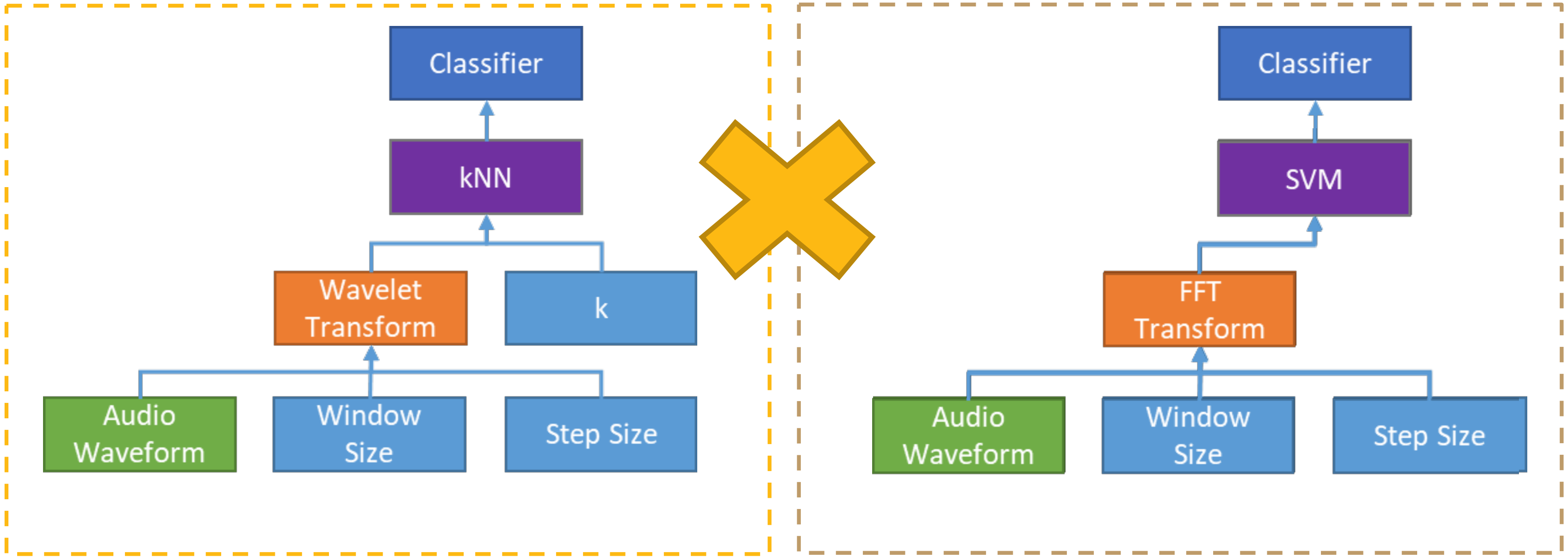
Composite Algorithms

Evolutionary Multi-objective Algorithm Design Engine

EMADE applies concepts from biology such as “survival of the fittest” and “DNA” to the world of software to **evolve** algorithms tailored to a specific dataset and problem

“Generation 1”

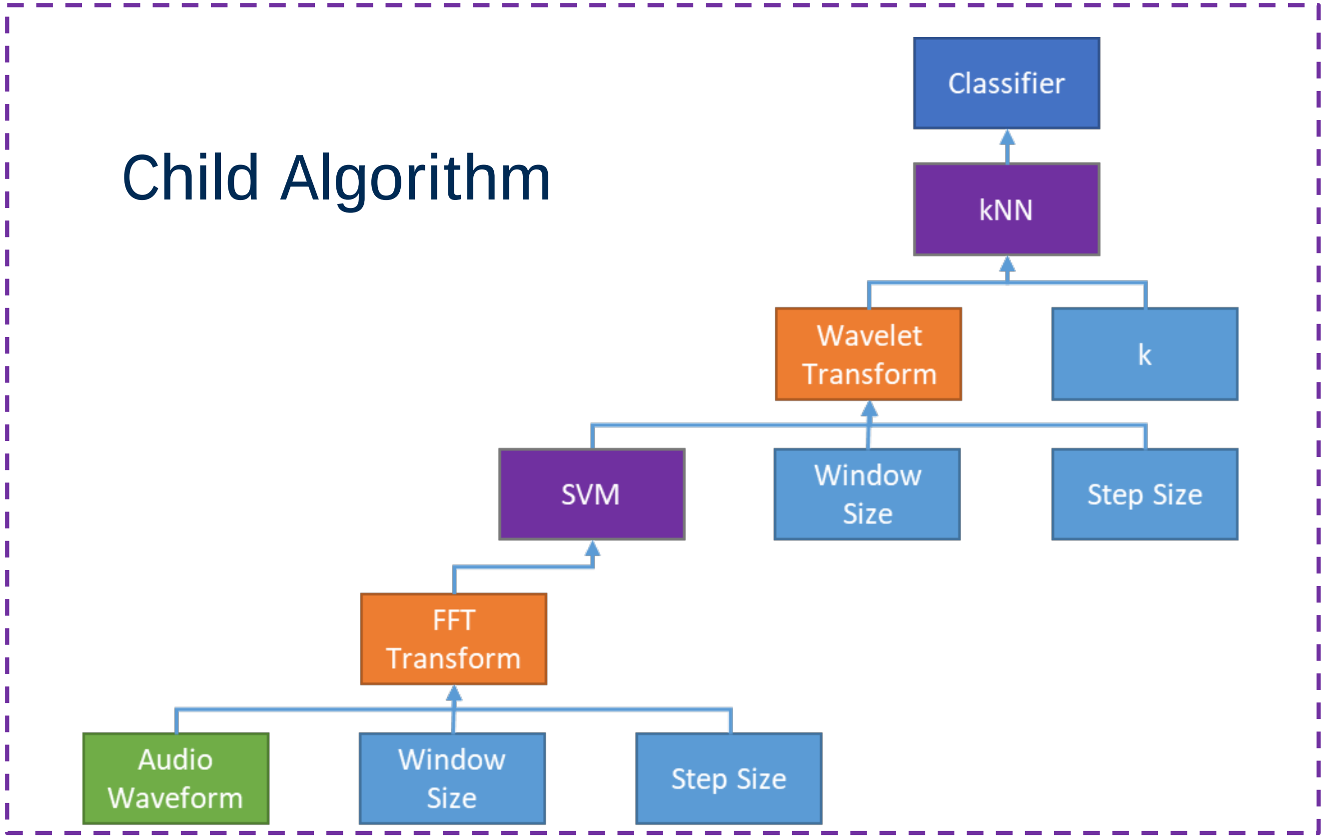


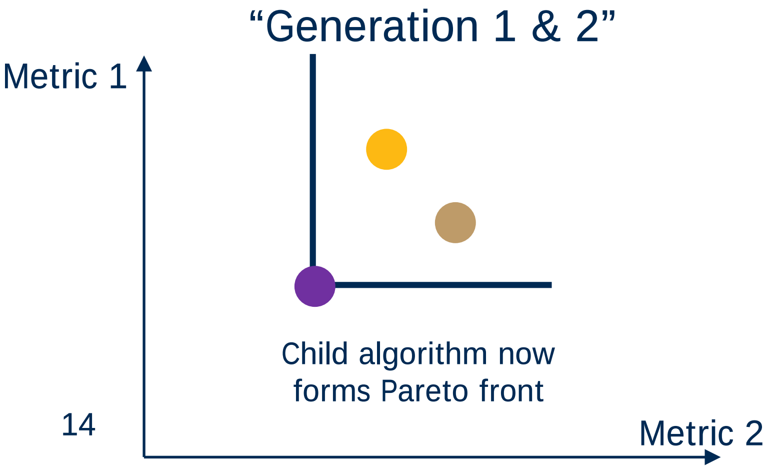


Crossover

Evolutionary Multi-objective Algorithm Design Engine

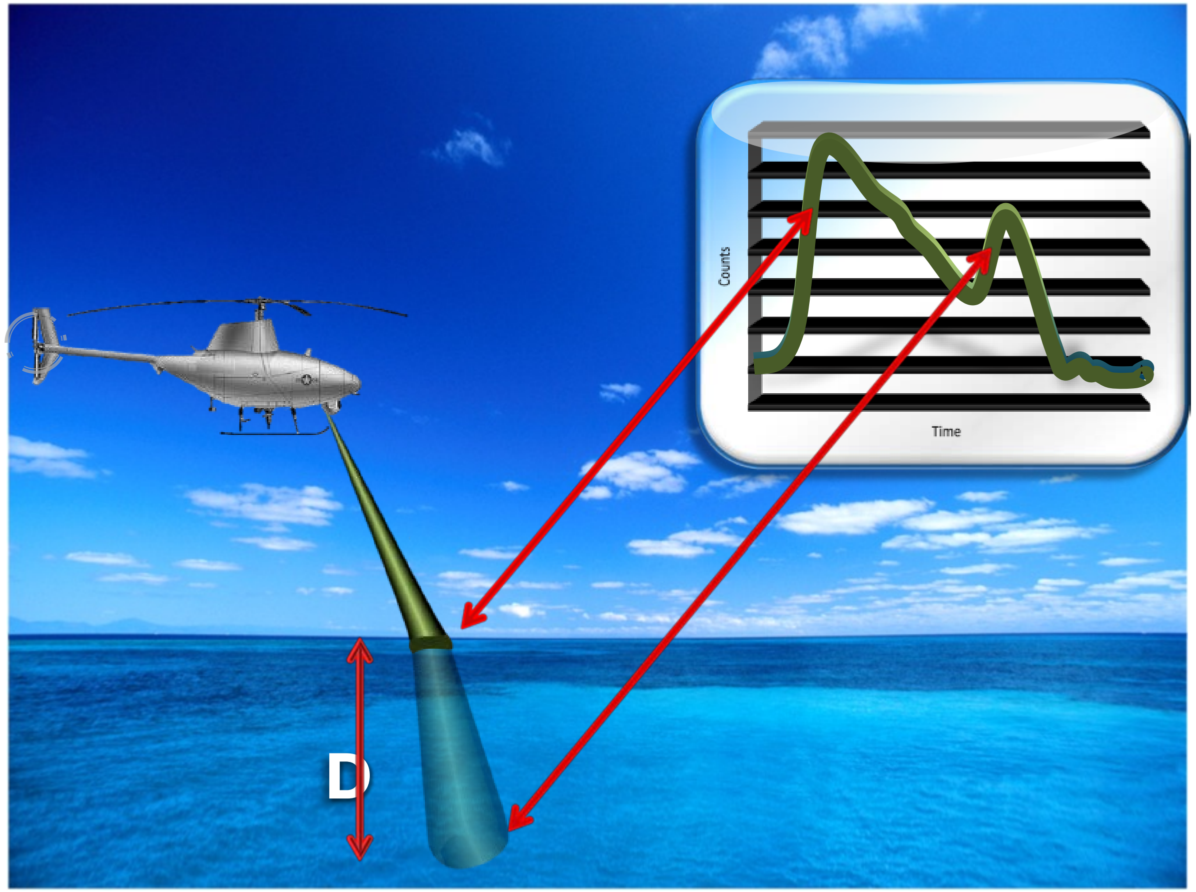
EMADE applies concepts from biology such as "survival of the fittest" and "DNA" to the world of software to **evolve** algorithms tailored to a specific dataset and problem





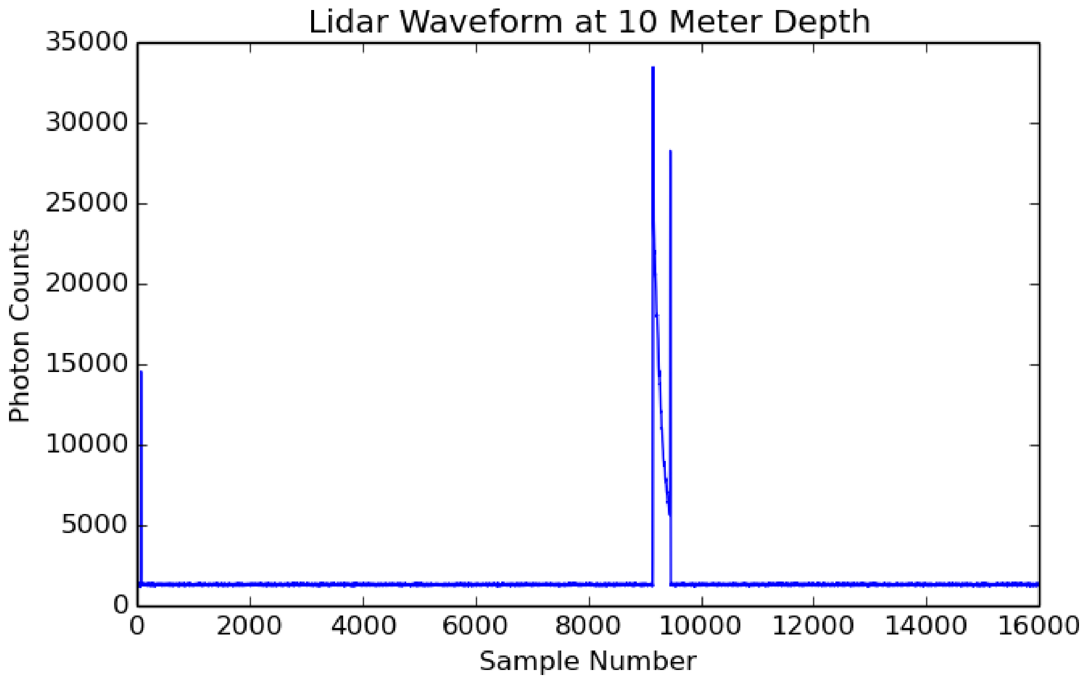
Demonstration of Capability: Bathymetric LIDAR

Task: Estimate the optical path length (OPL) in meters from the sea surface to the sea floor



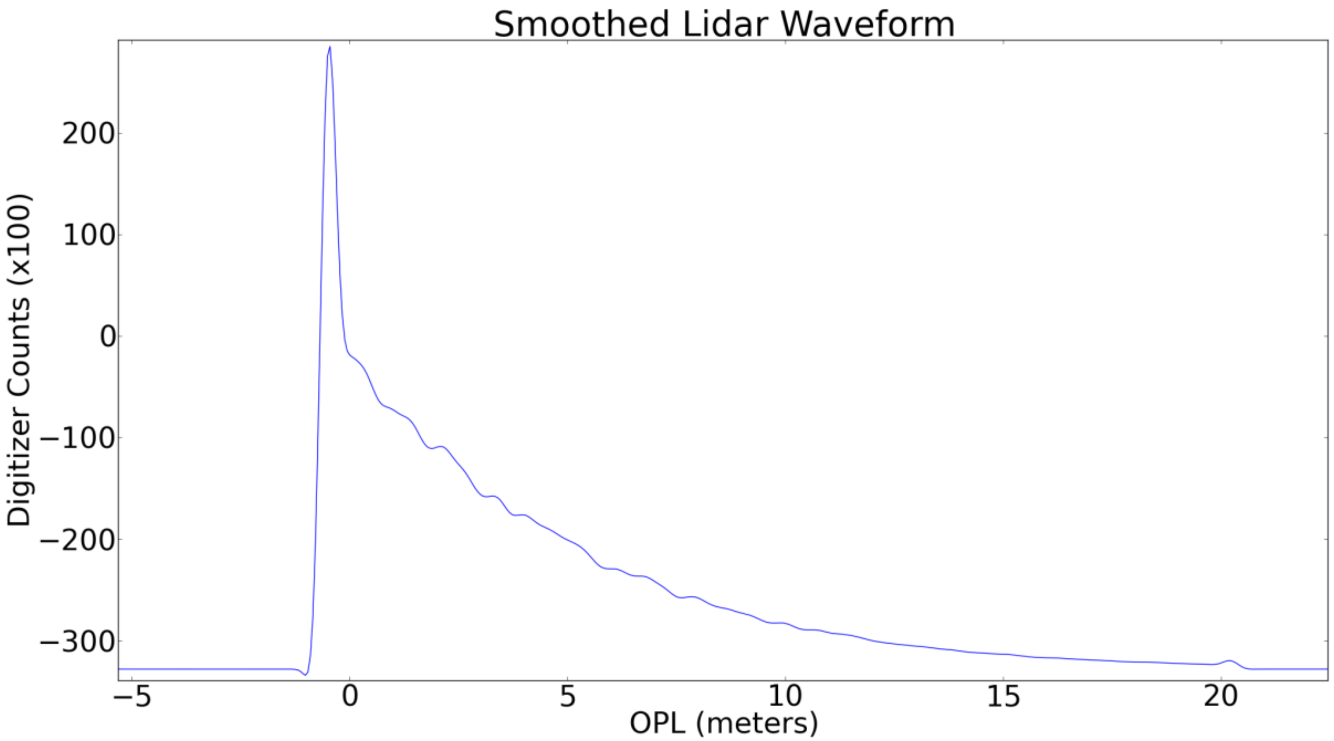
Interest Point Method [1]

- Industry standard for optical path length estimation
- Look for inflection points preceding peaks to compute optical path length in water
  - Peaks round out with reflections, noise can influence where peak is detected
- Three Steps
  1. Smooth LIDAR waveform
  2. Detect Peaks
  3. Perform informed search to find inflection points



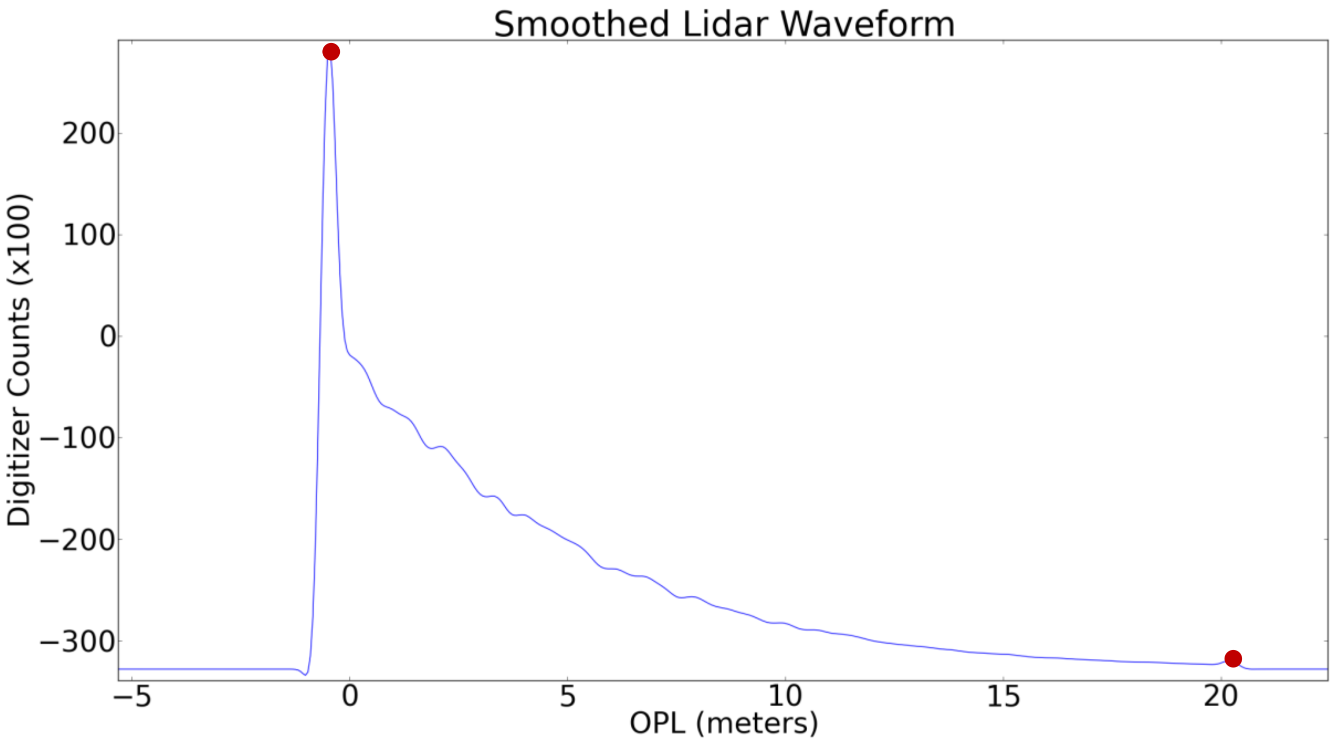
Smoothing the Waveform

Uses a Savitzky-Golay polynomial filter



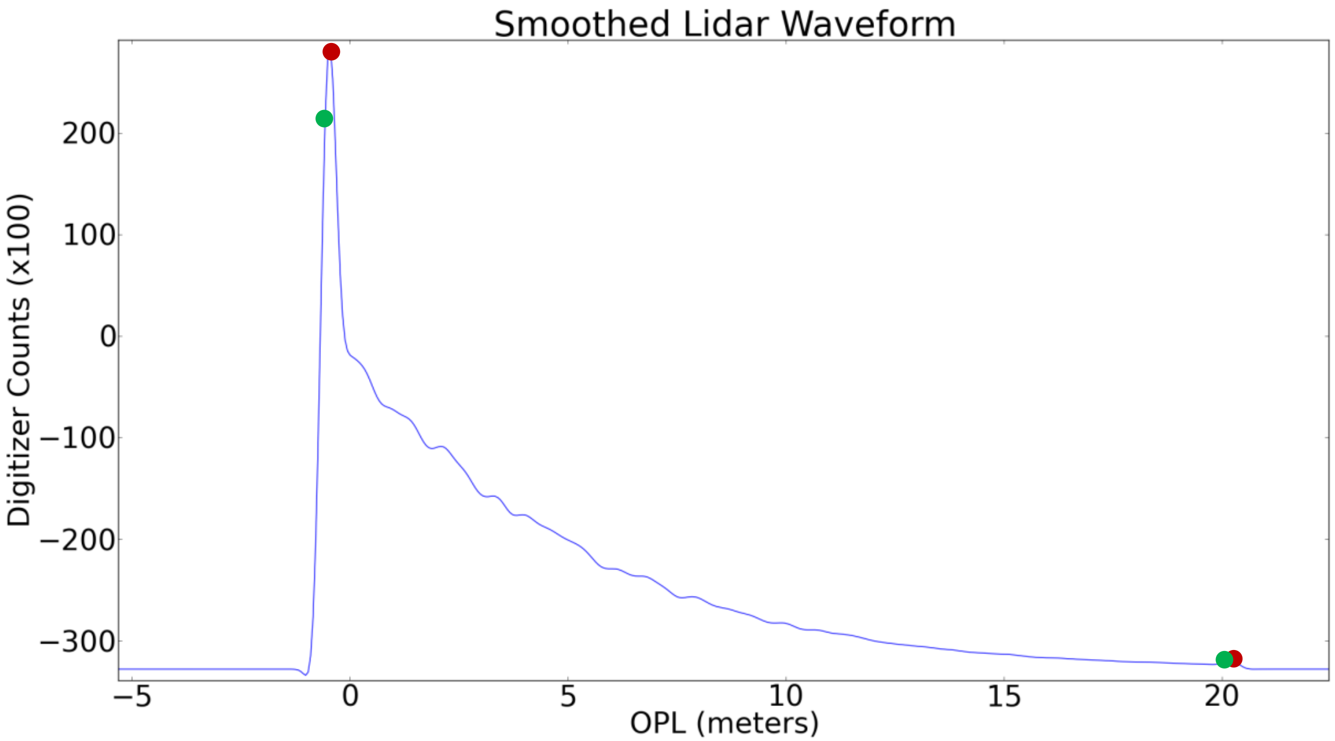
Peak Detector

Finds peaks based on relative strength of nearby points



Informed Search

Finds inflection points preceding the peaks using zero crossings

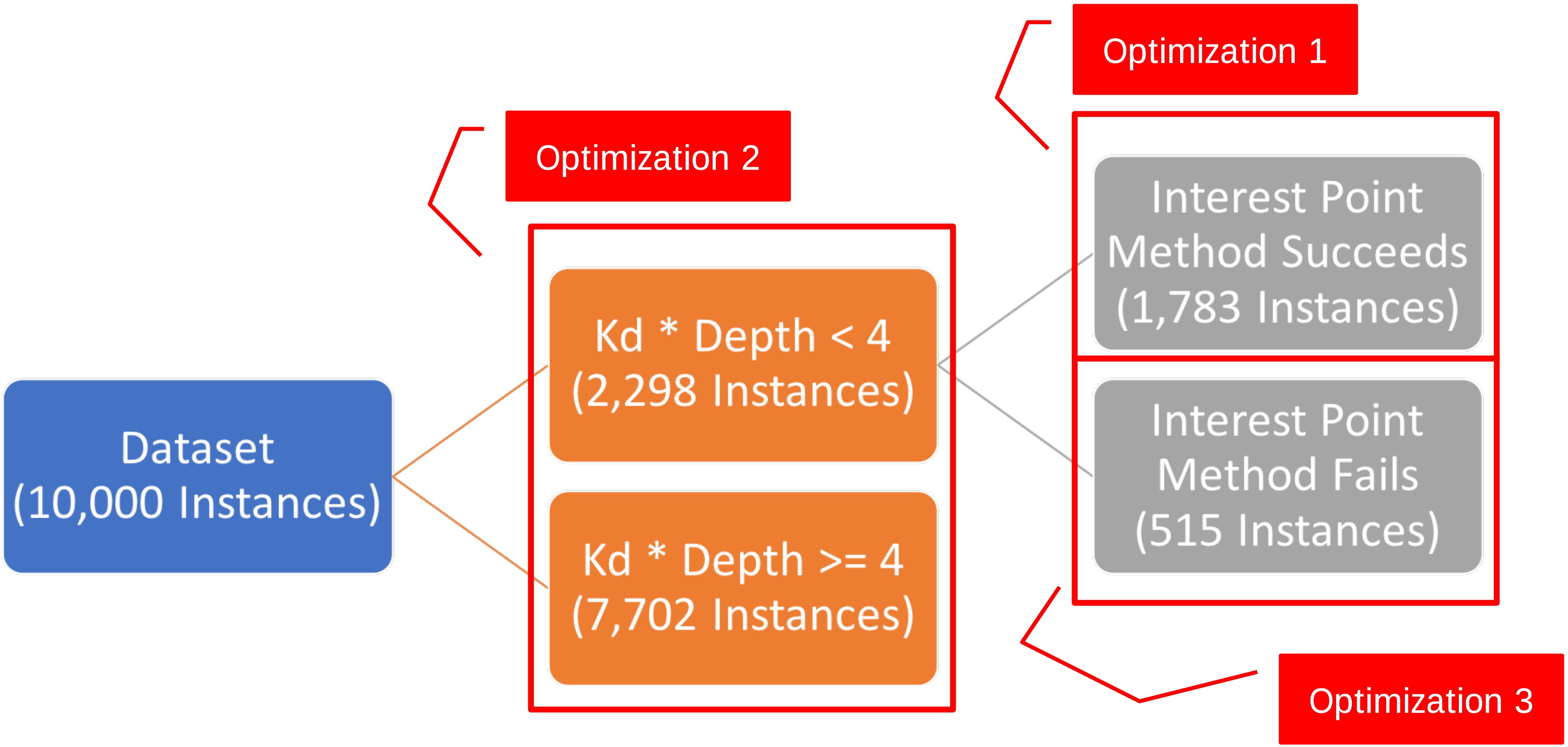


Creating a Robust Data Set

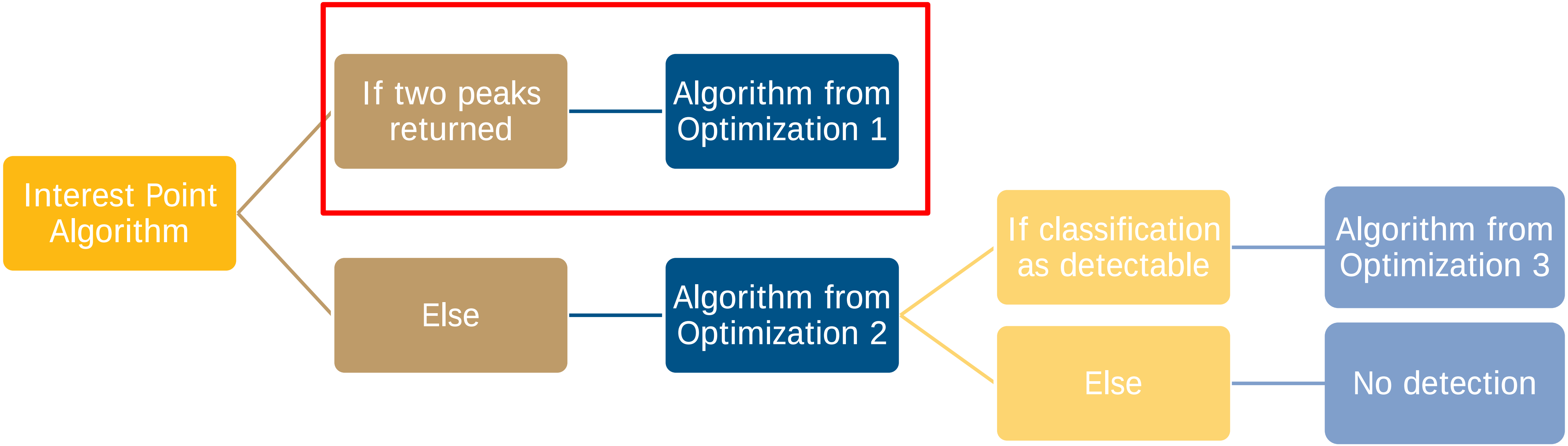
- Used simulator developed at GTRI EOSL as part of Dr. Domenic Carr's Dissertation [2]
- Captured variations in system and environment

Parameter	Distribution	Units
Average laser power	$\mathcal{N}(\mu = 30, \sigma = 0.9)$	W
Full-width half-max	$\mathcal{N}(\mu = 1.7, \sigma = 0.133)$	ns
Off nadir angle	$\mathcal{N}(\mu = 20, \sigma = 0.0673)$	deg
Filter spectral width	$\mathcal{N}(\mu = 1.4 \times 10^{-9}, \sigma = 0.00467)$	nm
Scan angle	$0.00197 * \mathcal{U}(\text{lower} = 0, \text{upper} = 183 \times 10^3)$	deg
PMT bias voltage	$\mathcal{N}(\mu = 550, \sigma = 9.167)$	V
Lowpass filter frequency	$\mathcal{N}(\mu = 6.14 \times 10^8, \sigma = 2.047)$	MHz
Latitude	$\mathcal{U}(\text{lower} = 4, \text{upper} = 25)$	deg
Longitude	$\mathcal{U}(\text{lower} = 104, \text{upper} = 124)$	deg
Water depth	$\mathcal{U}(\text{lower} = 0.25, \text{upper} = 55)$	m
Height above sea level	$\mathcal{N}(\mu = 400, \sigma = 10)$	m
Seafloor reflectance	$\mathcal{U}(\text{lower} = 0.01, \text{upper} = .25)$	
Seafloor tilt	$\mathcal{U}(\text{lower} = -20, \text{upper} = 14)$	deg
Wind speed	$\mathcal{U}(\text{lower} = 0, \text{upper} = 10)$	$\frac{m}{s}$
K_D	$\text{LogUniform}(\text{lower} = 0.06, \text{upper} = 10)$	$\frac{1}{m}$
β_π	$\mathcal{U}(\text{lower} = 0.001, \text{upper} = 0.003)$	$\frac{1}{\text{m} \cdot \text{sr}}$
σ_{β_π}	$\mathcal{U}(\text{lower} = 8 \times 10^{-5}, \text{upper} = 4 \times 10^{-4})$	$\frac{1}{\text{m} \cdot \text{sr}}$

Designing a Set Of Optimizations

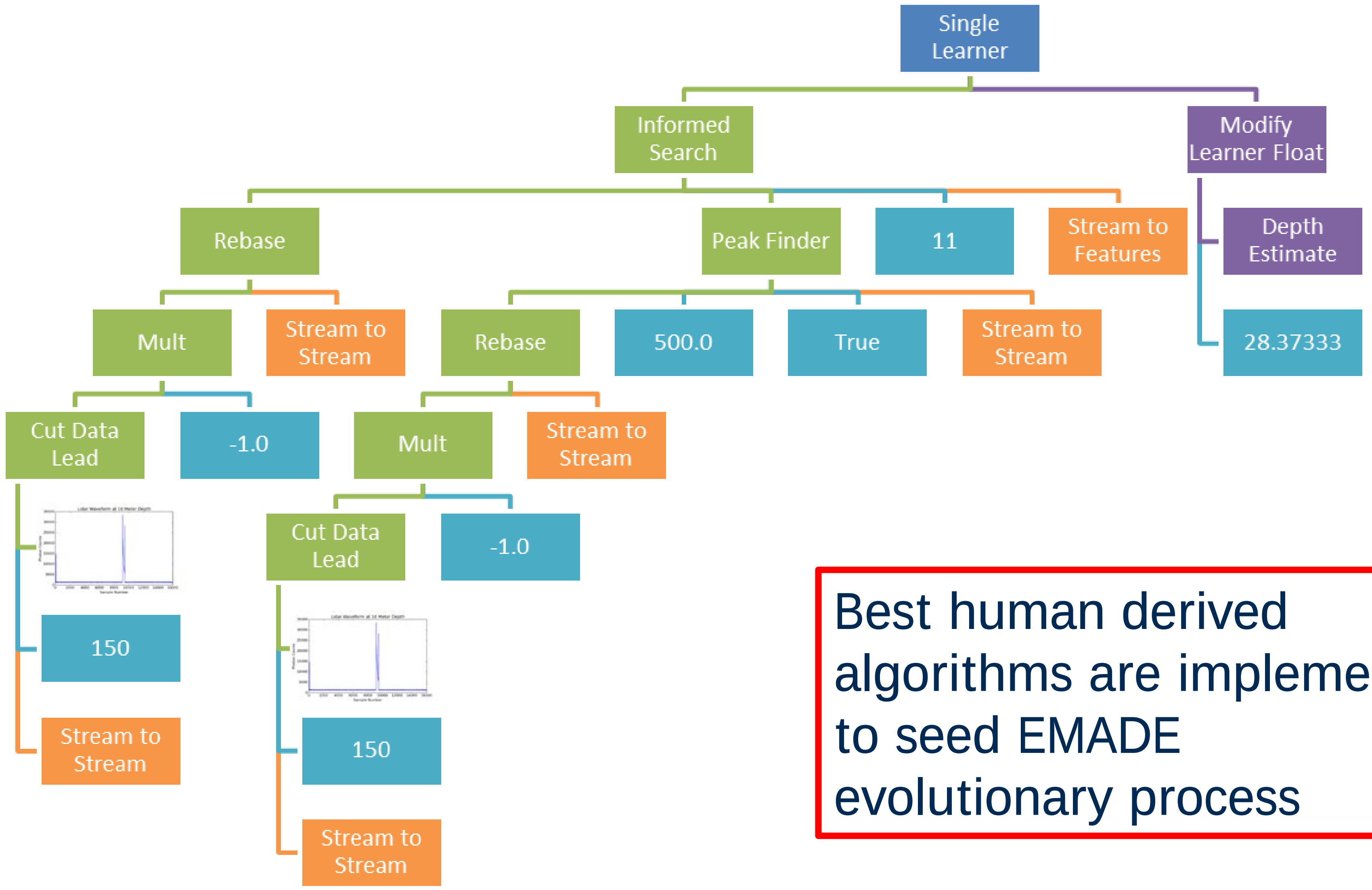


Designing a Set Of Optimizations



We will focus on Optimization 1

Seeding EMADE – Interest Point Detection



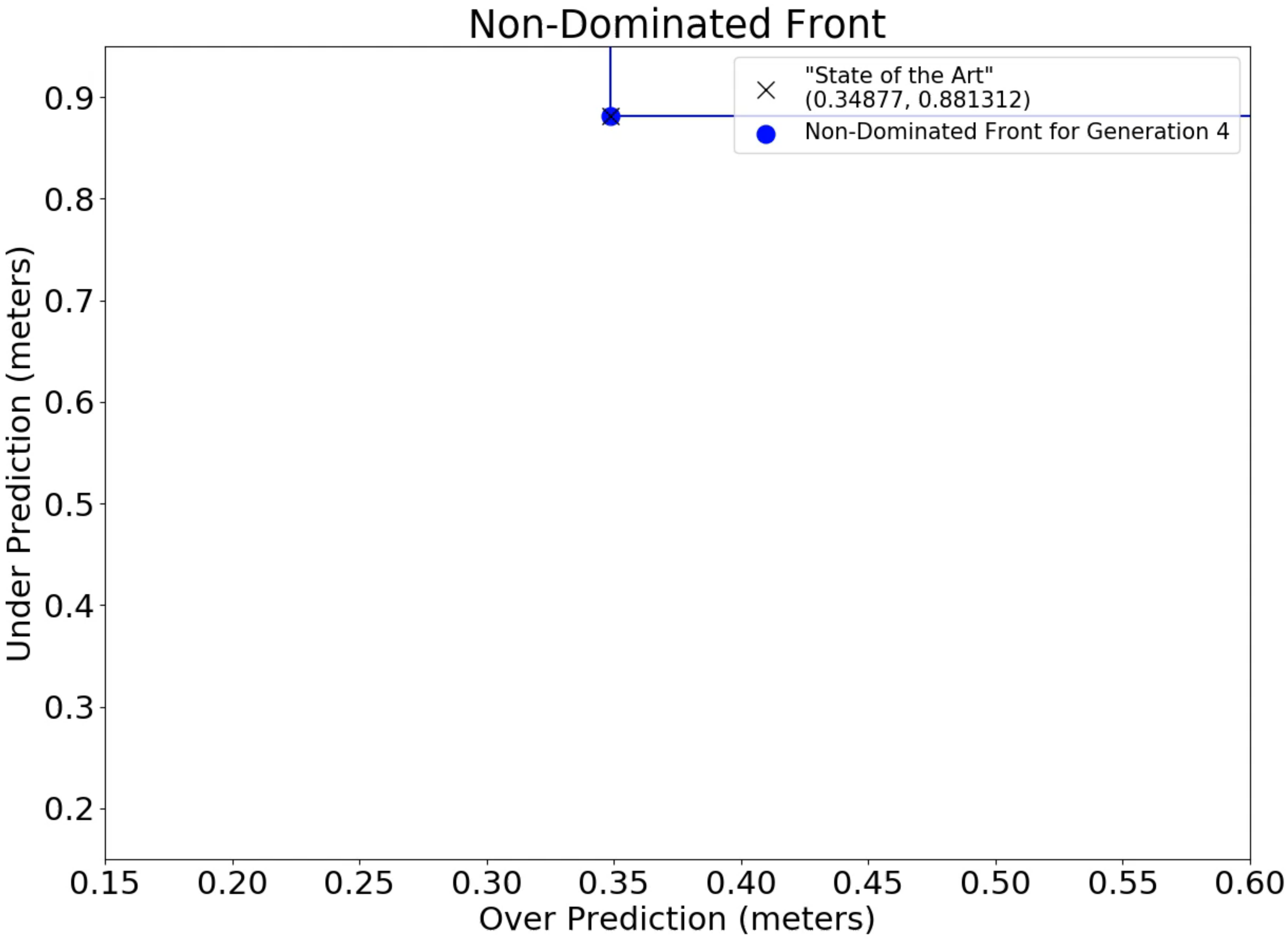
Best human derived algorithms are implemented to seed EMADE evolutionary process

Results

Evolutionary process creates solutions that out perform existing human derived algorithms

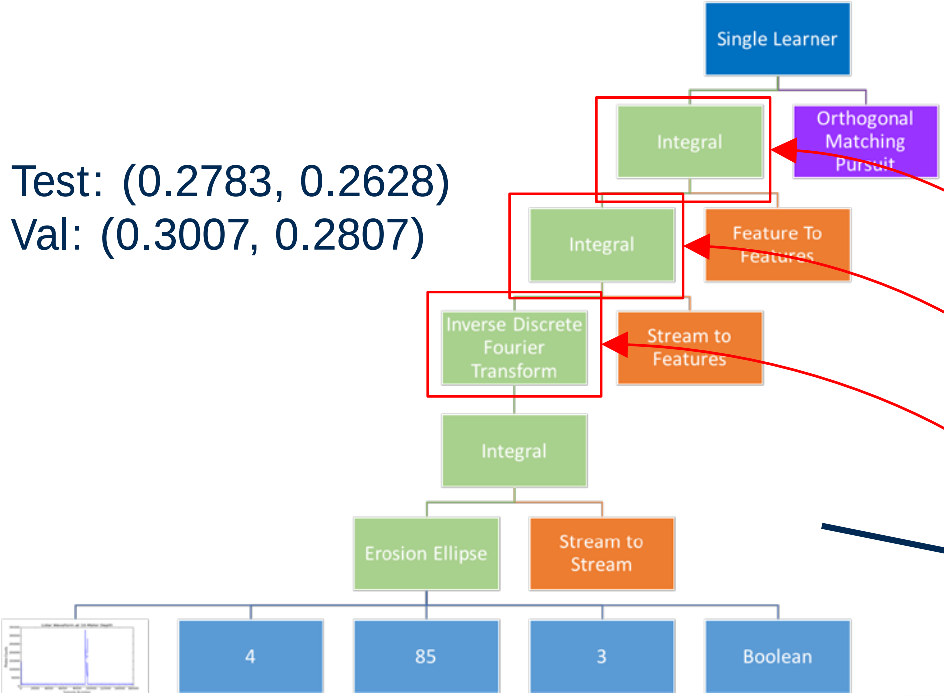
Validated after evolution:

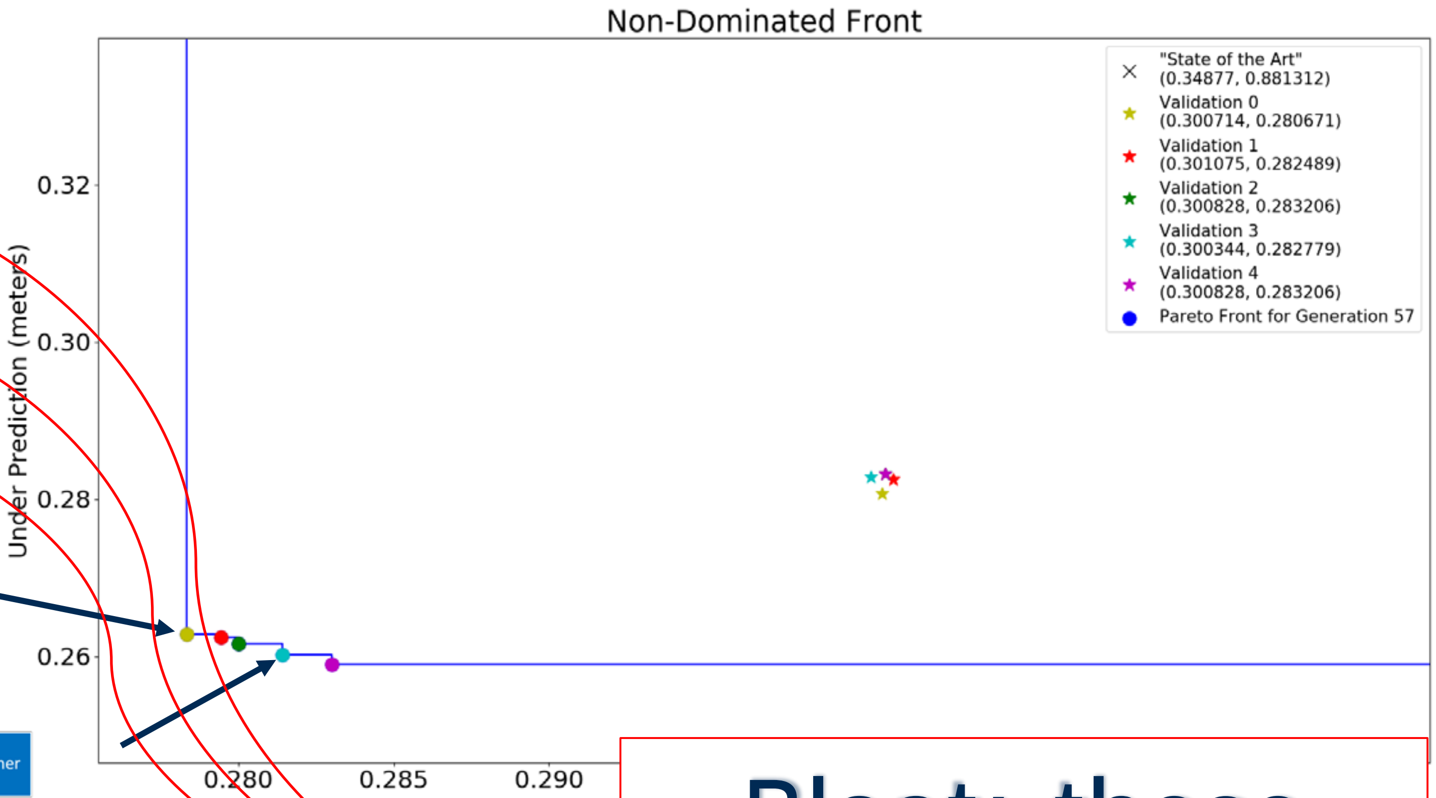
- Over prediction from 0.348 meters to 0.301 meters (13.5% Improvement)
- Under prediction from 0.881 meters to 0.281 meters (68.1% Improvement)



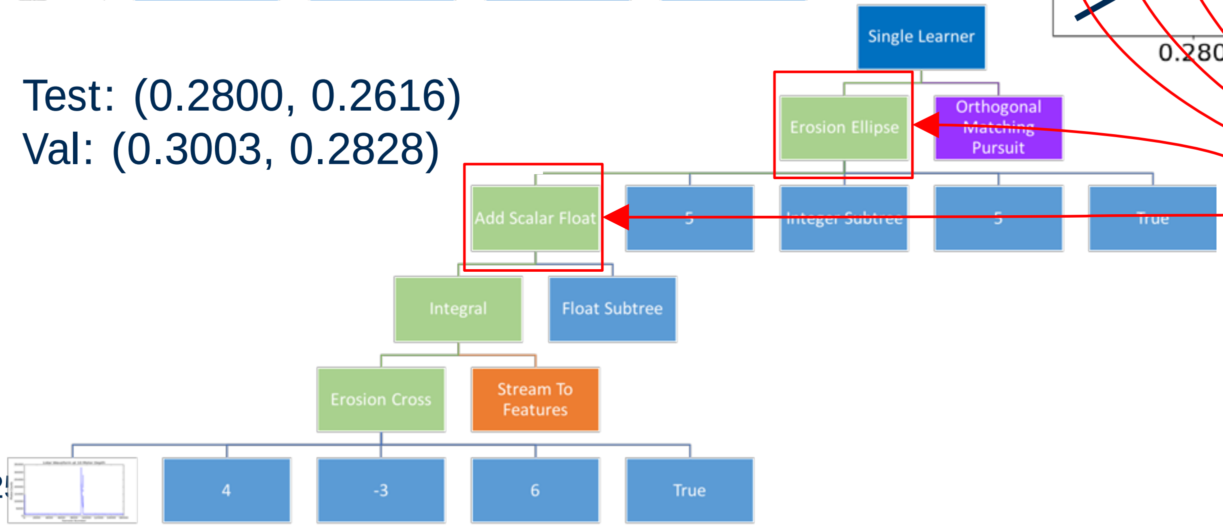
24

Test: (0.2783, 0.2628)
Val: (0.3007, 0.2807)





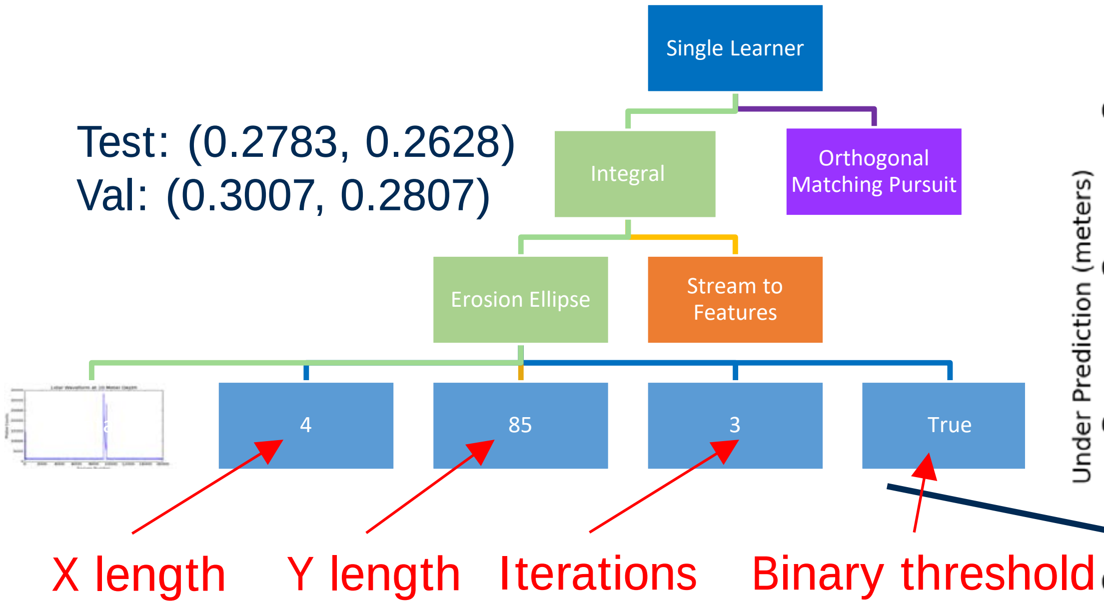
Test: (0.2800, 0.2616)
Val: (0.3003, 0.2828)

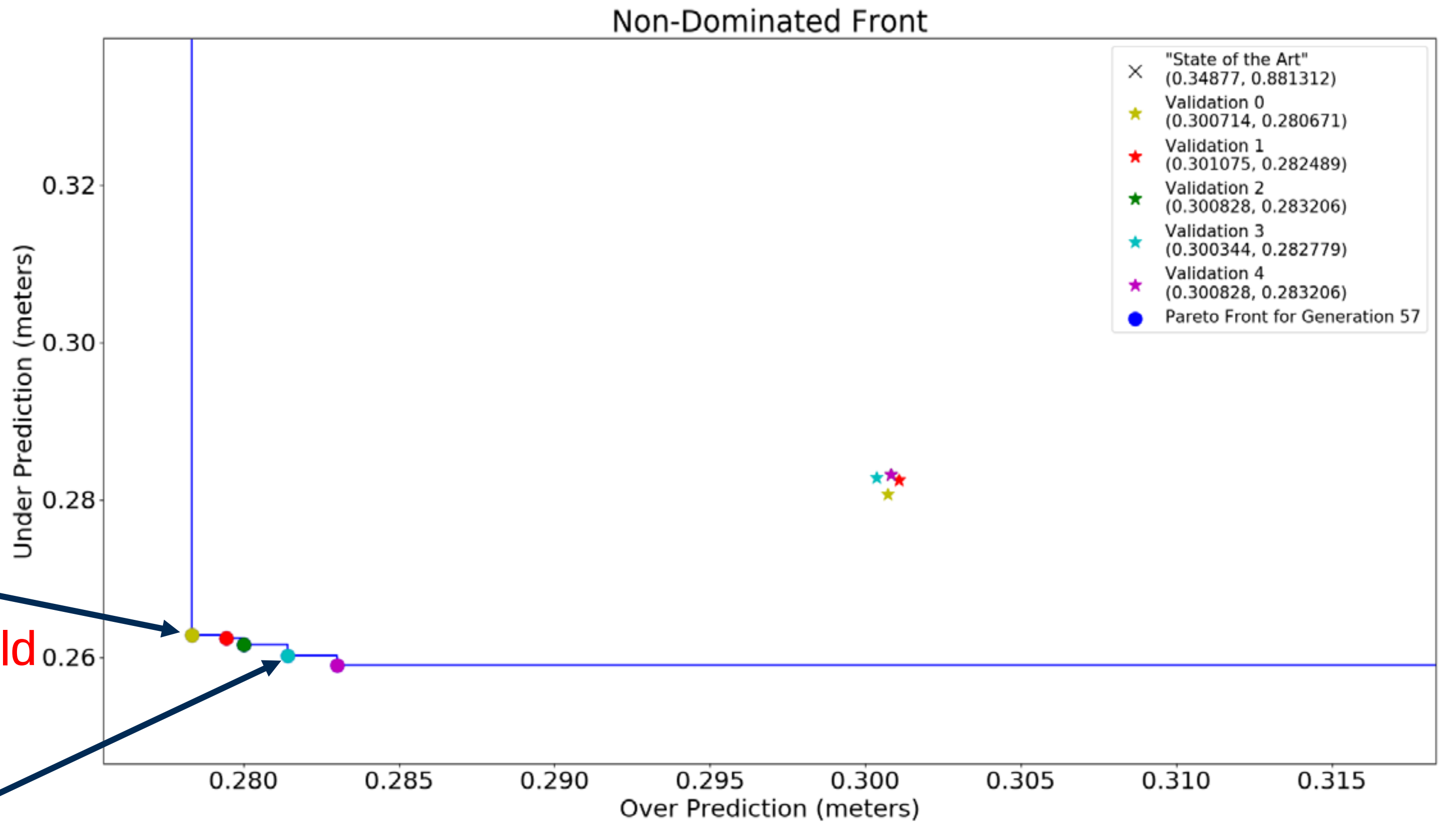


57 Gen
experim

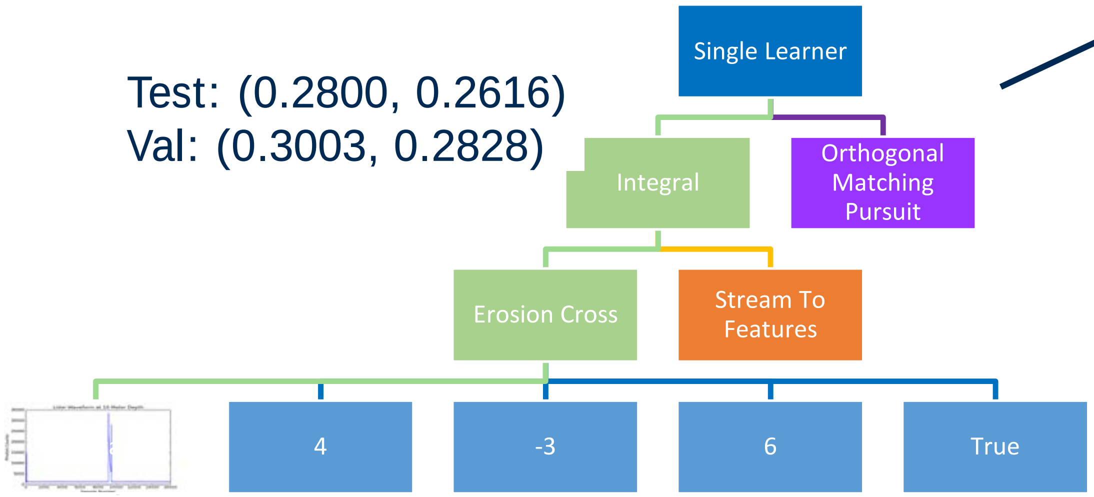
Bloat: these primitives have no effect

Test: (0.2783, 0.2628)
Val: (0.3007, 0.2807)



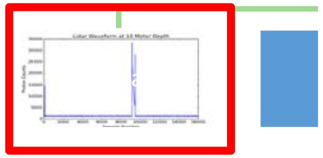


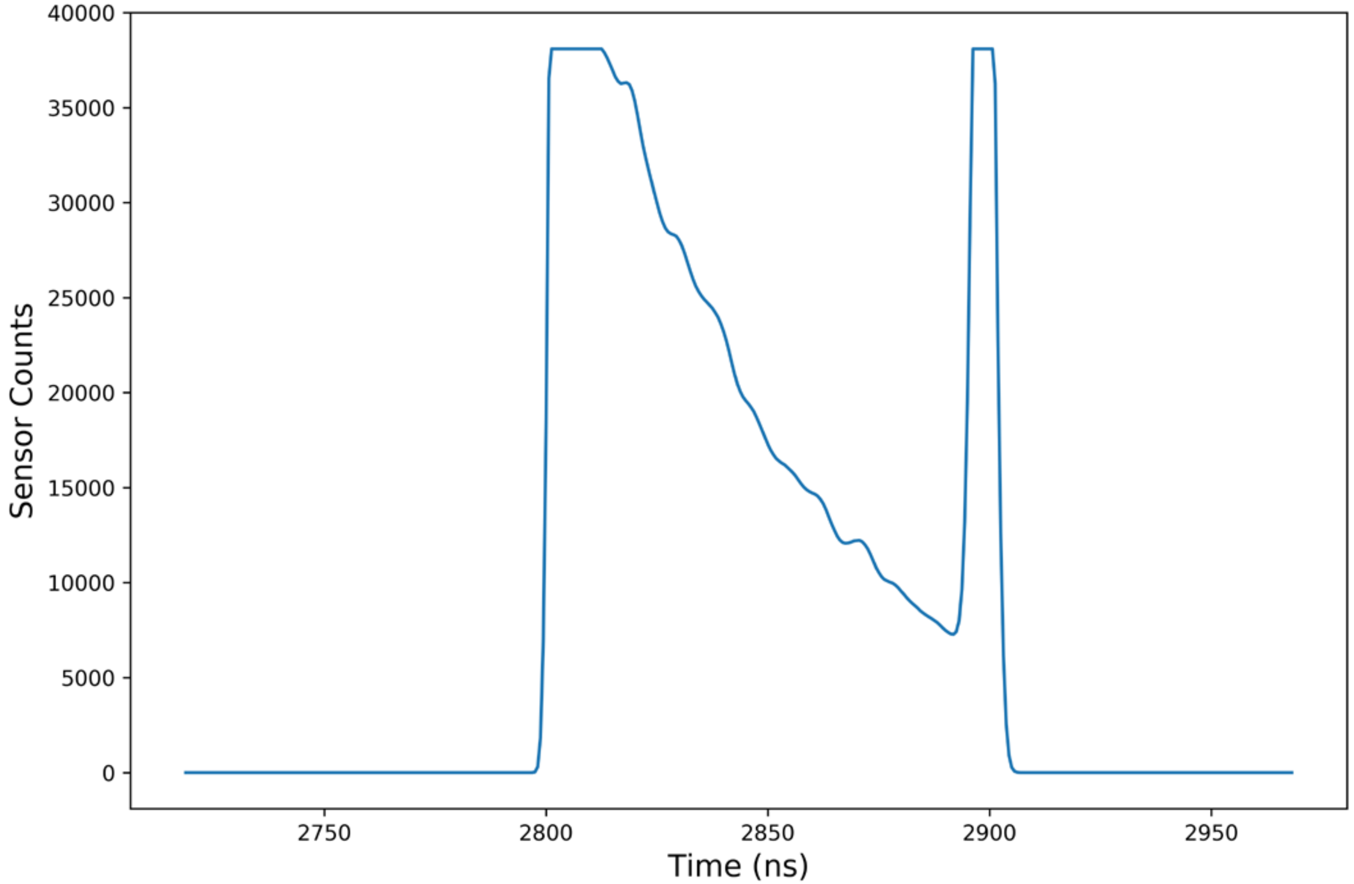
Test: (0.2800, 0.2616)
Val: (0.3003, 0.2828)

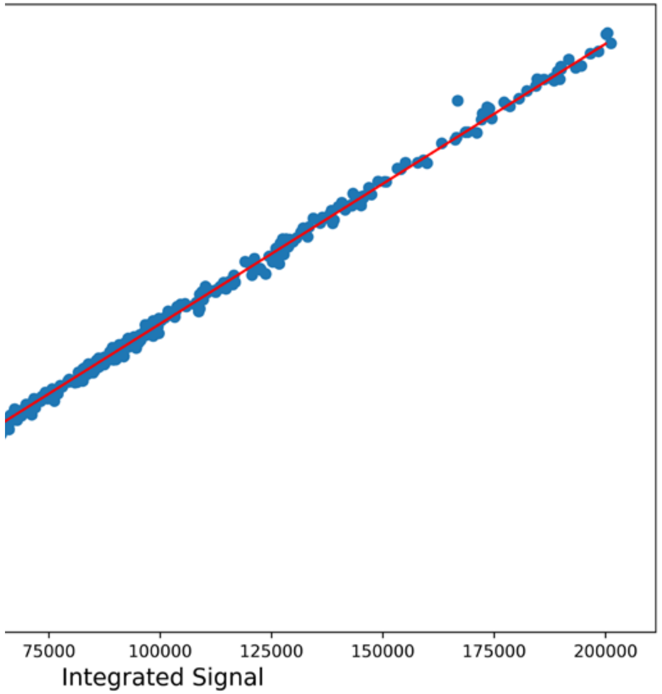


57 Generations represents ~17,000 experiments completed

Understanding the Algorithm

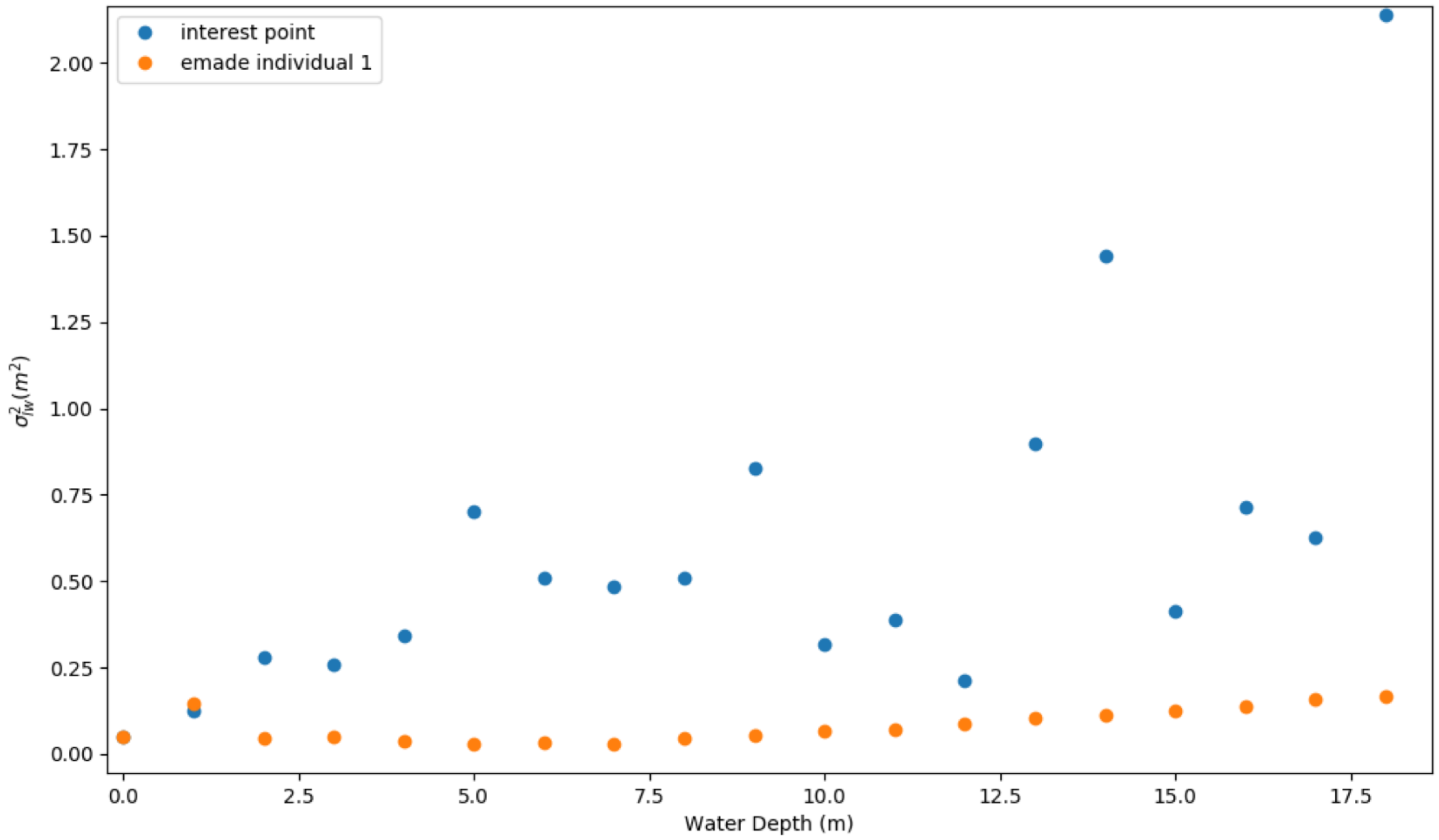






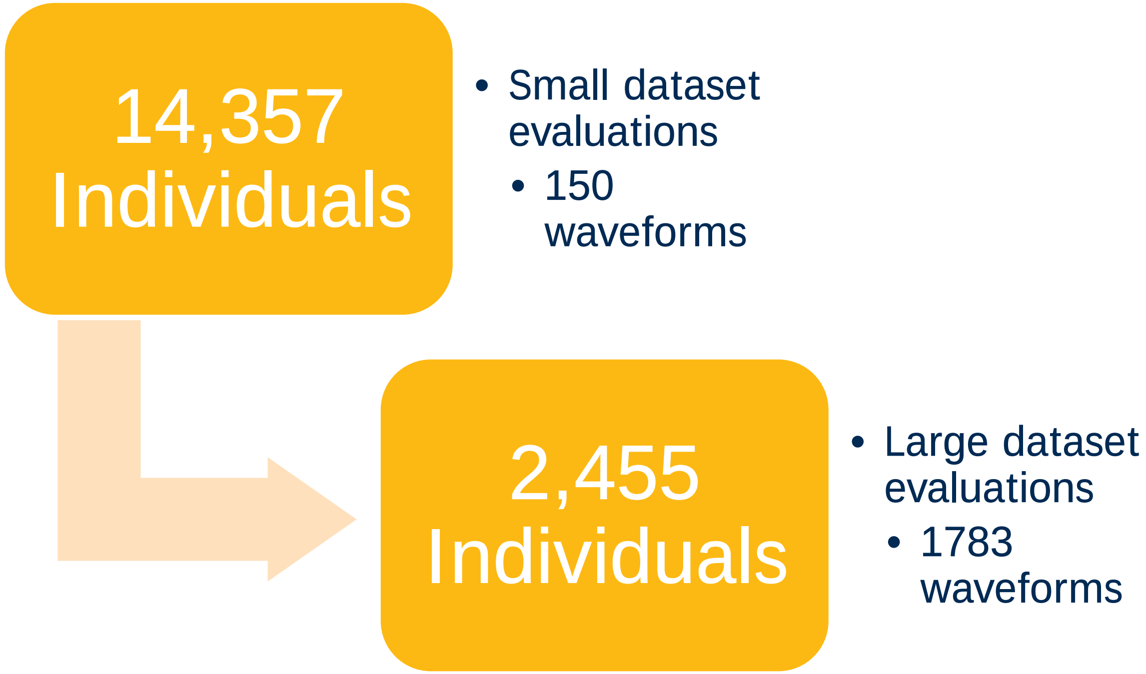
27

Comparison on Variance of Estimate vs Depth



Tiered Datasets – Savings on Two Tiers

- In total 14,357 individuals evaluated on the small dataset
  - Compute time averaged 64.5 seconds
- Only 2,455 matured to the large dataset
  - Compute time averaged 1,359.93 seconds
- Skipped 11,902 evaluations on the large dataset
  - Saved over 4400 CPU hours
  - EMADE ran for only 470 CPU hours
- Two evaluations on each of 2,455 individuals
  - Spent approximately 44 hours total on extra step of evaluation on small dataset



29

Conclusions

- EMADE improved on the state of the art by discovering a simpler technique.
- The optimal algorithms for time-domain signals relied on an OpenCV image-processing function.
- Scalable architecture allowed the 470 CPU-hours to be completed in little over a day on a cluster of computers.
- Full results for Optimization 2 and 3 are detailed in the paper.
  - Optimization 2 achieved prediction of bottom visibility with 88.1% accuracy, a 46% improvement over interest point alone.
  - Optimization 3 achieved errors of 0.38 meters over-prediction and 0.366 meters under-prediction where the interest point method fails.
- Automated algorithm design combines the best of human knowledge with the power of evolution to rapidly respond to new challenges.

References

1. A. Mathur, V. Ramnath, V. Feygels, E. Fuchs, J. Y. Park, and G. H. Tuell, “Predicted lidar ranging accuracy for czmil,” in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*, International Society for Optics and Photonics, vol. 7695, 2010, 76950Z.

2. D. A. Carr, “A study of the target detection capabilities of an airborne lidar bathymetry system,” PhD thesis, Georgia Institute of Technology, 2013.

Questions?