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

Jason Zutty, Rodd Talebi, James Rick, Christopher Valenta, Domenic Carr, Gregory Rohling
925 Dalney St NW
Atlanta, GA 30332
USA

Email: jason.zutty@gtri.gatech.edu, rodd.talebi@gtri.gatech.edu, james.rick@gtri.gatech.edu, chris.valenta@gtri.gatech.edu, domenic.carr@gtri.gatech.edu, greg.rohling@gtri.gatech.edu

ABSTRACT

This paper describes a methodology to optimize underwater target detection using a waveform-resolved bathymetric airborne LIDAR leveraging evolutionary algorithms and genetic programming. A radiometric and geometric bathymetric LIDAR simulator is implemented to generate millions of test cases from which genetic programming techniques search for the optimal signal processing techniques and their parameters given an optimization goal.

In recent years, the field of automated machine learning (autoML) has quickly attracted a significant amount of attention both in academia and industry. The driving force is to reduce the amount of human intervention required to process data and to create models for classification and prediction, a tedious and arbitrary process for a data scientist that often does not result in a global optimum with respect to accuracy or other metrics.

Our entry into the field is EMADE, the Evolutionary Multi-objective Algorithm Design Engine, which affords several benefits not found in other autoML solutions including the ability to stack machine learning models, process time-series data using dozens of signal-processing techniques, and efficiently evaluate algorithms on multiple objectives.

We created a processing architecture that involves using the state-of-the-art interest point method as a trigger for a machine learning pipeline. We evolved solutions using our EMADE framework; in cases where the previous method was successful, we achieved a 13.8% reduction in over-prediction and a 68.2% reduction in under-prediction error; where the previous algorithm was not successful in detecting a peak from the seafloor, we were able to detect the seafloor with an average of 0.359 meters in over-prediction and 0.357 meters in under-prediction errors. These values are still more accurate than the interest point method when it was able to detect the seafloor.

INTRODUCTION

Waveform-resolved LIDAR systems generate 3D visualizations by extracting targets from digitized waveforms corresponding to the backscattered optical signal received from an environment. Coupled with highly accurate position sensors, these systems transmit millions of pulses to create 3D point clouds.

Currently, researchers tend to explore a finite set of digital signal processing techniques; each algorithm is tested and individually optimized to produce its best results. The outputs of these algorithms are then compared against some metric. Unfortunately, this process has been predominantly manual and operated on a small set of data.

Additionally, most traditional algorithms rely on peak locations; they are sensitive to noise and geometric stretching from system or environmental sources. These problems are exacerbated in shallow and deep waters, scenarios where alternative methods might lend themselves much better.

Our solution is to simulate a large high-fidelity LIDAR dataset, and train a data pipeline of three individual models with the EMADE framework (Evolutionary Multi-objective Algorithm Design Engine) [1] [2].

EMADE is able to tune and combine various digital signal processing methods into a single new and novel algorithm. We introduce to it many of the published methods for conditioning and operating LIDAR data, such as wavelet processing, supersampling techniques, knowledge from multiple, spatially co-located pulses, and waveform decomposition methods [3] [4] [5] [6] [7] [8] [9] [10] [11].

Further, EMADE provides the ability to evaluate the validity of a model with multiple objectives. This lends itself kindly to bathymetry applications: trade-offs between under-prediction and over-prediction, and probability of miss.

INTEREST POINT METHOD

Bathymetric LIDAR has a well-documented algorithm from the literature known as the Interest Point Detection algorithm. This algorithm consists of three steps. The first is to condition the waveform using a Savitzky-Golay polynomial filter. Second, we extract the peaks from the waveform. Finally, we use the peaks to conduct an informed search, where we process small windows preceding the peaks for inflection points by studying the zero crossings of the second derivative. We compute the optical path length (OPL) by finding the difference in samples between the two inflection points representing the sea surface reflection and seafloor reflection. The inflection points are used instead of the peak locations because the peaks are more susceptible to stretching, saturation, and noise effects than the location of the inflection points. We translate the OPL directly to depth by multiplying by the angle the beam enters the water.

SIMULATING DATASETS

Machine learning techniques require a vast amount of labeled data to correctly train and validate derived algorithms. Unfortunately, bathymetric LIDAR data is often locked up as proprietary data or in proprietary data formats. Furthermore, this data is often raw and unlabeled with what exactly is being imaged.

To address these challenges, this work improved a previously designed bathymetric LIDAR simulator [12] that validated against real-world data. Leveraging the improved simulation capabilities, a robust waveform and point cloud data set was produced by varying system, environmental, and operational characteristics to boost availability of labeled data. Figure 1 shows a simulated waveform compared to a bathymetric waveform collected by a LIDAR system in Tampa, Florida. The similarity between the simulated and collected waveforms validates the usage of simulated data for supervised machine learning.

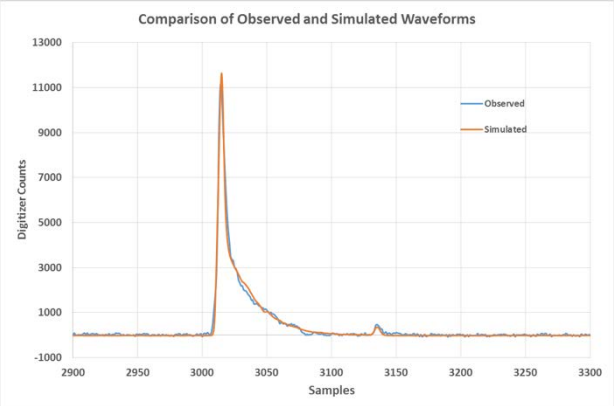

Figure 1. Validation of simulated waveforms with measured values from bathymetric LIDAR.

In this paper, using the parameter distributions in Table 1, we simulate environmental conditions of the South China Sea and a LIDAR system similar to the one used to validate the simulated waveforms. We chose the South China Sea for its challenging characteristics such as its high diffuse attenuation coefficient.

We constructed two datasets from the parameters in Table 1, one to be used during the optimization, and the other to be held out for validation after completion of the optimization. Each dataset comprises 10,000 waveforms randomly sampled from the given distributions.

Table 1. Parameters and Their Distributions for a Robust Bathymetric Dataset. Emulating characteristic of a LIDAR System and the South China Sea location.

Parameter	Distribution	Units
Energy Per Pulse	N(30,0.9)	W
Full Width Half Max	N(1.7, 0.133)	ns
Off Nadir Angle	N(20, 0.067)	deg
Height Above Sea Level	N(400, 10)	m
Filter Spectral Width	N(1.4, 0.00467)	nm
Scan Angle	U(0, 360)	deg
Water Depth	U(0.25, 55)	m
Latitude	U(4, 25)	deg
Longitude	U(104, 124)	deg
Wind Speed	U(0, 10)	m/s
PMT Bias Voltage	N(550, 9.167)	V
Detector Low Pass	N(614, 2.047)	MHz
Beta Pi	U(1e-3, 3e-3)	1/(m*sr)
Beta Pi Dev	U(8e-5, 4e-4)	1/(m*sr)
Rho Seafloor	U(0.01, 0.25)	
Tilt Seafloor	U(-20, 14)	deg
Diffuse Attenuation	LU(0.06, 10)	1/m

The training and validation datasets were broken up by several factors. First, the 10,000 instances were filtered down by where the seafloor should be visible: when the product □□ ∗ □□□□□$h < 4$. After this filtering, 2,298 detectable instances remained in the dataset. Next, to ensure a fair comparison between the interest point method and evolved solutions, the remaining data was split based on whether or not the interest point method was able to detect at least two peaks in the waveform successfully. Therefore, the interest point method can compute a range for all waveforms remaining in the dataset. If instead, we include failed cases, error quantification becomes a

challenging and somewhat arbitrary process, e.g., is an algorithm with higher error but more detected seafloors better than one with lower error but fewer detections? The combination of visibility and interest point requirements reduced our training dataset from 10,000 simulated instances to 1,783, leaving 515 that should be detectable, but where the interest method did not return an estimate.

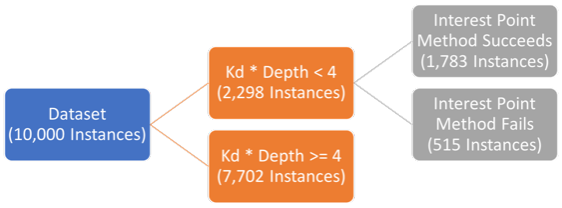

Figure 2. Dataset partitioning for optimization.

```
if interest point method predicts bottom then:
    return prediction using algorithm type 1
elif algorithm type 2 predicts K_D*Depth<4:
    return prediction using algorithm type 3
else:
    return "No bottom detected"
```

Listing 1. Pseudo-code for algorithm processing pipeline.

METHODS

The pipeline we introduce allows for the inclusion of both domain knowledge and evolved algorithms from genetic programming; it comprises three separate algorithms (Listing 1).

1. If interest point detection (Figure 3) can give an initial estimate for the seafloor, then implement an evolved algorithm to build in more accuracy.

2. If interest point detection fails to make an estimate, run an evolved algorithm to classify whether there is enough quality in the data to detect the seafloor using another method.

3. If the result from step 2 is “detectable,” execute an evolved algorithm to make an estimate.

The implementation of each algorithm in this pipeline emulates the same processing steps we used to filter the simulated dataset. The described pipeline allows for one-to-one comparison with the interest point method where the interest point method succeeds, and pure where the interest point method does not succeed.

We use EMADE to develop each of the three algorithms in the pipeline. To ensure success in this domain, we equip EMADE with state-of-the-art techniques in the LIDAR signal processing domain. Each of the following techniques is implemented as a building block in EMADE, along with a more extensive library of signal processing and machine learning methods, combine to construct and optimize algorithms:

- Richardson-Lucy Deconvolution
- Gaussian Decomposition
- Matched Filtering
- B-Spline
- Supersampling
- Gaussian Fitting
- Lognormal Fitting

In addition to implementing building blocks from the literature, we also developed seed algorithms. Figure 3 shows an EMADE genome of the interest point algorithm.

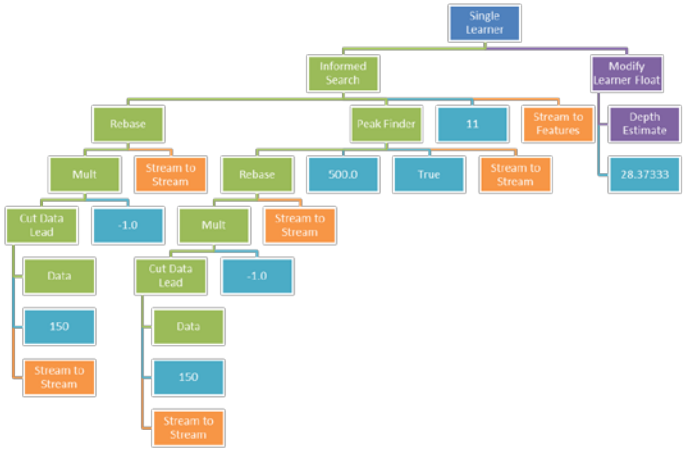

Figure 3. Interest Point Algorithm represented as a genome in EMADE.

ALGORITHM 1 – REFINING INTEREST POINT DETECTION

Our first optimization was on the data where the interest point method currently succeeds. We cross-folded the 1,783 instances in the dataset into five 80/20 train/test splits. EMADE uses these splits to protect against overtraining. EMADE scores each algorithm based on its average across all five splits. For each evaluation of an algorithm, the machine learning models are fit on the training portion, while being scored against the testing portion.

We seeded the first optimization with the interest point algorithm along with a hand-constructed algorithm that combines one of the LIDAR building blocks (Supersampling) along with a spectral representation and a gradient boosting regression machine learner. Figure 4 shows this hybrid seed algorithm.

Figure 4. A seed algorithm that combines a processing algorithm from literature and machine learning.

EMADE ran on a two-tiered dataset. The first tier was a small subset of one cross-folded pair that we used as a litmus test to determine if the algorithm would be successful. While an evaluation on the entire five folds takes an average of 1359.93 seconds per individual, the smaller dataset takes only an average of 64.5 seconds per individual. EMADE evaluated 11,902 individuals on the smaller dataset without needing to promote them to the second tier, while choosing to fully evaluate 2,455 algorithms on all five folds. In processing time, EMADE saved over four thousand CPU-hours of computation at the expense of 43.99 CPU-hours of litmus testing the successful individuals. EMADE minimized 8 objectives for this optimization:

- RMS Error
- Probability of Miss
- Percent Error
- Over Prediction Error
- Under Prediction Error
- False Negative Bottom
- Number of Invalid Individuals
- Number of elements in tree

EMADE used an initial population size and elite pool size of 512 individuals. The initial population contained 2 seeded and 512 randomly initialized algorithms. Each generation, EMADE produced 300 offspring. Table 2 shows the mating and mutation methods used along with the probabilities of application for each method.

Table 2. EMADE Mating and Mutation Rates.

Operation	Operation Type	Probability
Crossover	Mating	0.5
Crossover Ephemeral	Mating	0.5
Headless Chicken Crossover	Mating	0.1
Headless Chicken Ephemeral Crossover	Mating	0.1
Insertion	Mutation	0.05
Insert Modify	Mutation	0.05
Ephemeral	Mutation	0.25
Uniform	Mutation	0.05
Shrink	Mutation	0.05

We harvested results after EMADE ran for 57 generations and evaluated over 14,000 individuals. Figure 5 shows the non-dominated frontier on two objectives. Across all 8 objectives, 73 individuals were non-dominated solutions. One of EMADE's strengths is its ability to search this high-dimensional trade-off space to create algorithms that can be used under a number of different circumstances depending on an a posterior decision vector of objective importance. Figure 6 and Figure 7 show the gold and cyan algorithms from Figure 5 respectively.

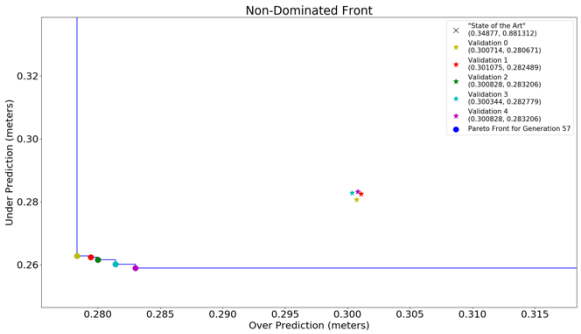

Figure 5. Non-Dominated frontier after 57 generations with respect to under prediction and over prediction errors. The stars show the performance of each non-dominated algorithm on the held-out validation set.

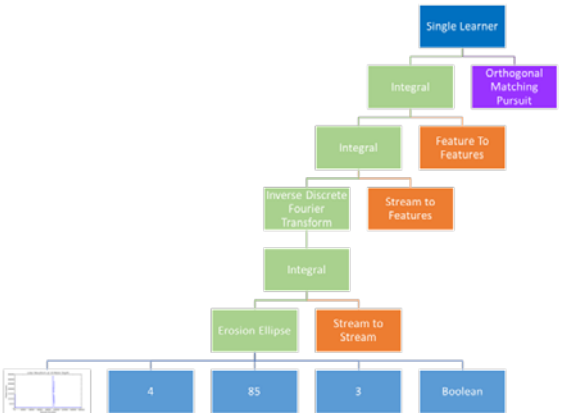

Figure 6. An EMADE evolved solution with validated over prediction error of 0.30 meters and validated under prediction error of 0.28 meters.

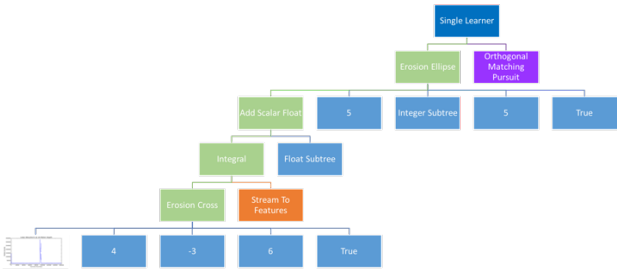

Figure 7. An EMADE evolved solution with validated over prediction error of 0.30 meters and validated under prediction error of 0.28 meters.

We further analyzed the algorithm in Figure 6 by evaluating the variance of optical path length estimation against water depth for a K_D of 0.14 1/m. Figure 8 shows the comparison of the evolved solution and the interest point method, with the EMADE solution strongly outperforming the state of the art.

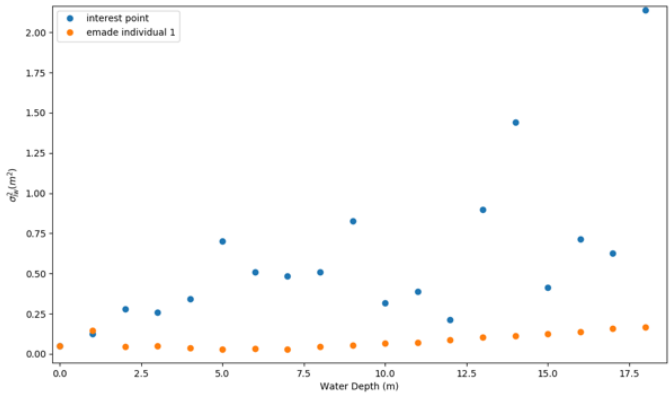

Figure 8. Variance vs Depth for evolved solution and interest point method.

Both evolved solutions in Figure 6 and Figure 7 behave in a similar fashion. They evolved around building blocks in the EMADE toolbox from computer vision, erosion ellipse and erosion cross. Both of these methods specialize in the sharpening of images, but here they instead are able to convert a LIDAR return into a rectangle function. The rectangle starts at the surface and ends at the sea floor. The orthogonal matching pursuit in each algorithm correlates the area of these boxes to a depth. Because the boxes have fixed heights from these methods, the area is directly proportional to the depth.

ALGORITHM 2 – CLASSIFICATION OF DETECTABILITY

If the interest point method fails to make a prediction, it is usually for one of two reasons. Either it is because the peak from the seafloor is indistinguishable from the peak of the sea surface due to how shallow the seafloor is, or it is too small to be distinguished from noise due to how deep the seafloor is. In both of these cases we would like to know prior to processing, if we should be able to see a bottom due to the murkiness of the water. Our second optimization focused on classifying waveforms based on the community accepted $K_D * Depth < 4$ metric.

For this problem, we constructed a dataset by randomly sampling 500 instances where $K_D * Depth \geq 4$ along with 500 instances where $K_D * Depth < 4$. We constructed our dataset this way so that our classes of 0 (≥ 4) and 1 (< 4) would be balanced. We again used a two-tiered structure, the first tier being 200 waveforms, the second being 800. We split the 200 waveforms in to 150 for training and 50 for testing. We cross-folded the 800 waveforms in to five 80%/20% partitions.

This optimization used three objectives, false positive rate, false negative rate, and the number of the elements in the genome. A false positive in this context is a classification that a waveform should be predictable but is not. A false negative is a classification that a waveform should not be predictable when in fact it is.

Figure 9 shows a classification algorithm from the optimization. This algorithm had a false positive rate of 10% and a false negative rate of 13.8%. The overall accuracy of the classifier is 88.1%.

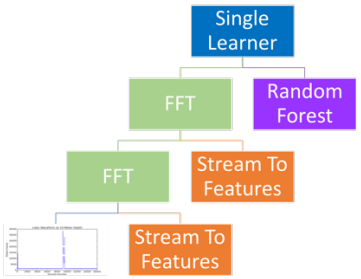

Figure 9. Binary classifier for predicting if $K_D * Depth < 4$.

ALGORITHM 3 – PREDICTIONS WITHOUT PEAKS

The final optimization we ran was the case where the interest point algorithm failed to detect at least two peaks in the waveform, and thus unable to make a depth prediction. For the third optimization, we cross-folded the 515 instances of the 10,000 where $K_D * Depth < 4$ and the interest point failed using five folds with an 80%/20% split.

Figure 10**Error! Reference source not found.** shows our non-dominated frontier after 60 generations. Note we do not show performance of the state-of-the-art interest point algorithm here, because this is the data on which it failed to predict and thus could not be scored. Figure 11 shows the algorithm represented in cyan in Figure 10. Note that similar to the algorithm we evolved in Optimization 1, this algorithm uses an erosion function, but also combines this with a spectral representation and a support vector machine to regress the depth. The performance of this algorithm validated to have an average of 0.359 meters of over prediction and an average of 0.357 meters of under prediction. While performance of these algorithms is lower than that of those from Optimization 1, these are inherently more challenging waveforms without a defined second peak.

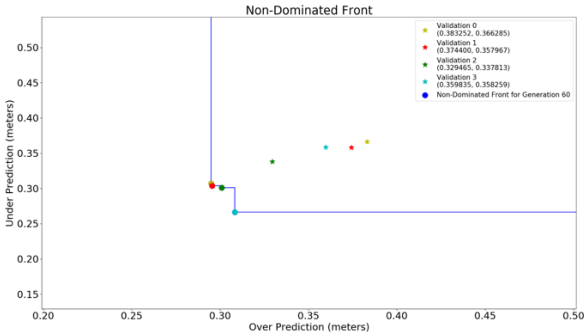

Figure 10. Non-dominated frontier for predicting depth where interest point method fails.

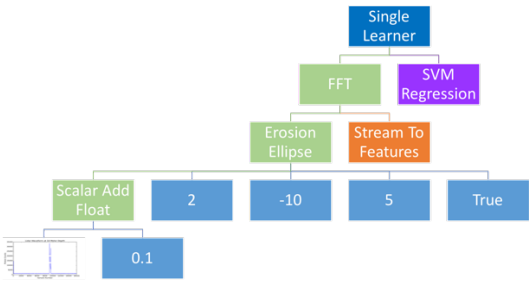

Figure 11. Evolved algorithm for predicting depth when interest point method fails, but the binary classifier from Optimization 2 predicts.

CONCLUSIONS

This problem represents the first time-domain regression problem solved using a genetic programming based autoML framework. The evolved solution is a simpler and more elegant solution than the one that has been used in industry for decades. The three algorithms developed in this paper afford a significant performance improvement over the interest point method. Where the interest point method succeeded before, the evolved solution validated to 0.30 meters of average over-prediction and 0.281 meters of under-prediction. On the same dataset, the interest point method obtained an average over-prediction of 0.348 meters and an average under-prediction of 0.881 meters. Thus, the evolved algorithm achieved a 13.8% reduction in over-prediction and a 68.1% reduction in under-prediction errors. Furthermore, we showed that not only did the evolved solution have better average performance, but the variance vs. depth is also significantly better.

In addition to improving performance where the current algorithm can detect a seafloor return, we also evolved a solution that enables ranging where it was previously impossible, with a validated performance of 0.38 meters over-prediction and 0.366 meters under-prediction. The evolved solution's performance where the interest point method fails compared to where the interest point method succeeds is only 9.2% worse in over-prediction, and 58.5% better in under-prediction.

Finally, through the evolution of the binary classifier in conditions where the interest point method fails, we achieve an 88.1% accuracy in our ability to predict with this evolved solution. Prior to this evolved classifier, the interest point method was failing to predict on 515 out of 2,298 instances, representing an accuracy of 77.6%. Thus, the evolved classifier picks up 46% of possible improvement, i.e.

$$\frac{new-previous}{1-previous}.$$

Overall, this paper shows the power of autoML coupled with high-fidelity simulation to outperform human-developed state-of-the-art. Our most substantial performance gains came from non-ideal signals that the interest point method encountered, such as saturation causing peaks to disappear (and thus making the detection of interest points fail), or spurious noise peaks causing false seafloor returns resulting in significant under-prediction. The EMADE evolved solution, found that the best way around these problems was to look at the overall energy level of the signal until it fell below a certain threshold, ignoring the spurious or saturated peaks that may fall between the sea surface and seafloor.

ACKNOWLEDGEMENTS.

The authors would like to thank GTRI for supporting this effort on IRAD.

REFERENCES.

[1] J. Zutty, "Creating Human-Competitive Algorithms Using Multiple Objective Vector Based Genetic Programming," Georgia Institute of Technology, Atlanta, 2016.

[2] G. Rohling, "Multiple Objective Evolutionary Algorithms for Independent, Computationally Expensive Objective Evaluations," Georgia Institute of Technology, Atlanta, 2004.

[3] T. Allouis, J.-S. Bailly, Y. Pastol and C. Le Roux, "Comparison of LiDAR waveform processing methods for very shallow water bathymetry using Raman, near-infrared and green signals," *Earth Surface Processes and Landforms*, pp. 640-650, 2010.

[4] J. Castorena and C. D. Creusere, "Sampling of Time-Resolved Full-Waveform LIDAR Signals at Sub-Nyquist Rates," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 3791-3802, 2015.

[5] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. P. Deseilligny and W. Puech, "Processing Full-waveform LIDAR Data: Modelling Raw Signals," in *ISPRS Workshop on Laser Scanning*, Finland, 2007.

[6] B. Jutzi and U. Stilla, "Range determination with waveform recording laser systems using a Wiener Filter," *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 95-107, 2006.

[7] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art," *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 1-16, 2009.

[8] Z. Pan, C. Glennie, P. Hartzell, J. C. Fernadez-Diaz, C. Legleiter and B. Overstreet, "Performance Assessment of High Resolution Airborne Full Waveform LiDAR for Shallow River Bathymetry," *Remote Sensingn*, pp. 5133-5159, 2015.

[9] C. E. Parrish, I. Jeong, R. D. Nowak and R. B. Smith, "Empirical Comparison of Full-Waveform Lidar Algorithms: Range Extraction and Discrimination Performance," *Photogrammetric Engineering and Remote Sensing*, pp. 825-838, 2011.

[10] J. Wu, J. Van Aardt and G. P. Asner, "A Comparison of Signal Deconvolution Algorithms Based on Small-Footprint LiDAR Waveform Simulation," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 2402-2414, 2011.

[11] J. Wu, J. Van Aardt, J. McGlinchy and G. P. Asner, "A Robust Signal Preprocessing Chain for Small-Footprint Waveform LiDAR," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 3242-3255, 2012.

[12] D. Carr and G. Tuell, "Estimating Field-of-View Loss in Bathymetric Lidar: Application to Large-Scale Simulations," *Applied Optics*, pp. 4716-4721, 2014.