Wave Height Quantification using Land Based Seismic Data with Grammatical Evolution

Sarah Donne, Miguel Nicolau, Christopher Bean and Michael O'Neill

Abstract— **Accurate, real time, continuous ocean wave height measurements are required for the initialisation of ocean wave forecast models, model hindcasting, and climate studies. These measurements are usually obtained using in situ ocean buoys or by satellite altimetry, but are sometimes incomplete due to instrument failure or routine network upgrades. In such situations, a reliable gap filling technique is desirable to provide a continuous and accurate ocean wave field record. Recorded on a land based seismic network are continuous seismic signals known as microseisms. These microseisms are generated by the interactions of ocean waves and will be used in the estimation of ocean wave heights. Grammatical Evolution is applied in this study to generate symbolic models that best estimate ocean wave height from terrestrial seismic data, and the best model is validated against an Artificial Neural Network. Both models are tested over a five month period of 2013, and an analysis of the results obtained indicates that the approach is robust and that it is possible to estimate ocean wave heights from land based seismic data.**

I. Introduction

Significant wave height (SWH) is a commonly used ocean wave parameter, regarded as a good representation of the ocean wave state for a given time period. Accurate and continuous records of SWH are important as they are often used when initialising ocean wave forecast models or in generating accurate hindcasts, both necessary to a broad range of industries. Current methods suffer from spatial and temporal resolution issues as well as discontinuities due to equipment failure and maintenance.

The data used in the development of the models presented here are from Ireland and the Northeast Atlantic Ocean. Microseism data is recorded on land based seismic stations throughout Ireland and SWH is recorded offshore at buoy K4 located off the Northwest coast, (see Fig.1 for locations). Given the record wave heights that have been measured off the coast of Ireland in recent years there is particular interest in assessing the wave record for climate studies and energy studies looking to harness this energy. These studies require a continuous, high resolution dataset and the method presented here is a step towards this.

In this study, the Grammatical Evolution system [1] was used to generate predictors of wave height quantification, using land-based seismic data. The obtained models were validated against a Neural Network approach. Both systems generate accurate wave height models, with Grammatical Evolution generating compact symbolic models, which can be analysed and potentially further optimised. The results obtained show the potential of the application of evolutionary methods to this problem domain.

The structure of this paper is as follows. Section II provides an introduction to the domain of Wave Height Quantification, including an introduction to the use of microseisms as wave height predictors, and an overview of previous work. Section III presents the evolutionary system used, with Section IV presenting the experimental methodology of this study. Finally, Section V presents and analyses the results obtained, and Section VI concludes and presents future work directions.

II. Wave Height Quantification

Ocean gravity waves are surface waves that occur on the interface between oceans and the atmosphere. They are generated by wind blowing over stretches of water and can range in size from centimetres to tens of meters high. Typical ocean wave periods range from 1-30 sec. Ocean wavelengths range from 1.5m to 1.5km and wave velocity from 1.5m/s to 48m/s, with an 8 second ocean wave capable of travelling up to 45km/hr in deep waters. Deep water is defined as water depths greater than one half of the wavelength of the wave. They are directly influenced by changes in atmospheric conditions with the largest ocean waves corresponding to the strongest winds.

The Northeast Atlantic Ocean off the west coast of Ireland has one of the most energetic wave climates in the world with average annual significant wave heights (SWH), the average of the highest one third of waves recorded, between 2.5-3m. This average is significantly higher in winter months, with the current ocean buoy network in Ireland recording the largest SWH of 20.4 m to date in December 2011. Many industries, from fishing to coastal engineering and climate studies, rely on accurate sea state condition forecasts, in particular when the environment can be very rough and changeable. In terms of climate studies, analysis often needs to be done on the longest time series available. In order to provide accurate and reliable wavefield measurements and marine forecasts, continuous, high-resolution measurements of the sea state are needed.

At present ocean wave heights are measured using in situ ocean buoys, passively using satellite altimetry, and numerically using ocean wave models. Ocean buoys provide a good temporal resolution, off the west coast of Ireland, with SWH being transmitted once per hour, but have a poor spatial resolution with up to hundreds of kilometres between buoys (Fig 1). Also, due to the harsh conditions in the NE Atlantic, these buoys can often experience technical difficulties which

Sarah Donne and Christopher Bean are with the UCD School of Geological Sciences, University College Dublin, Ireland (email: Sarah.Donne@ucdconnect.ie, chris.bean@ucd.ie).

Miguel Nicolau and Michael O'Neill are with the Complex and Adaptive Systems Laboratory (CASL), University College Dublin, Ireland (email: {Miguel.Nicolau, M.ONeill}@ucd.ie).

can take time to repair, leading to data gaps and missing wave height information. Satellite altimetry on the other hand provides excellent spatial resolution covering entire areas of the ocean at a time, with accuracy of a few centimetres and homogeneous measurements with 1 measurement per second every 7 km in all weather conditions [2]. This method however, suffers from extremely poor temporal resolution for measuring ocean wave heights in real time, with the return period passes ranging from 10 days to 35 days [2], [3]. Satellite altimetry can be used to track large storm cells as they evolve and move along a storm path since they occur over long time periods, but cannot be used as a real time ocean wave height monitoring tool.

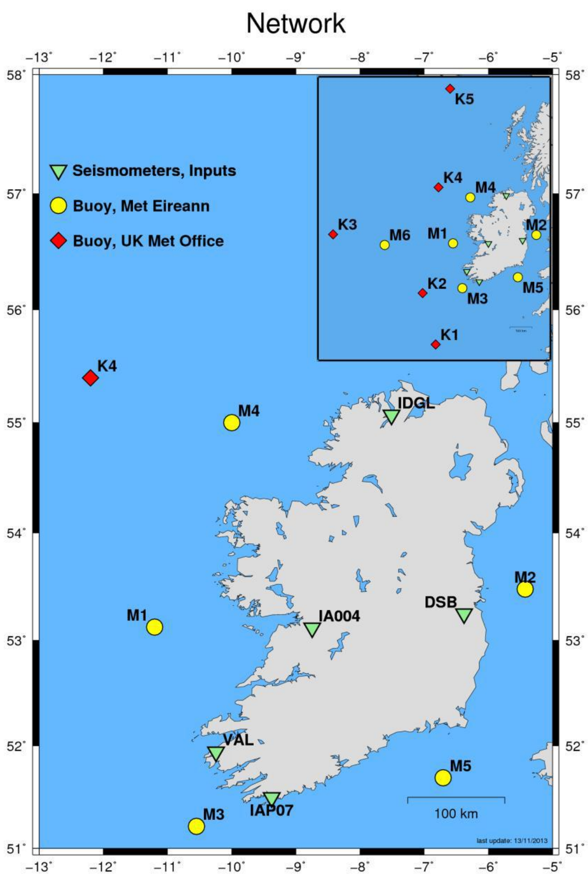

Fig. 1. Buoy and seismometer networks in Ireland.

There have been significant improvements in numerical wave models in recent years with the introduction of coastal reflection and associated effects [4]. As with any numerical model, there are associated errors due to incomplete understanding of the physics of the medium.

Taking the benefits and limitations of the current methods for measuring ocean wave height into account, we attempt to develop a method for estimating ocean wave heights at multiple locations using seismic waves, recorded on land, that are directly associated to the ocean wavefield.

A. Seismic Data

Ocean gravity waves exert a vertical pressure profile in the water column beneath them. If this wave is a propagating or travelling wave, its pressure profile drops off exponentially with depth [5]. For such a wave, if the water depth is shallow enough, less than half the wavelength of the ocean wave, the pressure exerted vertically downwards by the wave can transfer energy into the subsurface. This usually occurs as the wave interacts with the sloping seafloor near coastlines as waves break [5]. When energy is transferred into the subsurface it generates very small, continuous, ground vibrations, known as microseisms.

In the case of a standing or partial-standing wave, where two opposing propagating ocean waves interact with almost the same frequency, the pressure profile of the standing ocean wave does not decay with depth, due to second order non-linear effects [5], and is constant from sea surface to sea floor. The energy transferred into the subsurface by the pressure fluctuations generated by a (partial-) standing wave is much greater and the resulting microseisms are much stronger and are the main focus of this study. This transfer of energy is independent of depth.

When generated by a propagating wave, the resulting microseisms are known as primary or single frequency microseisms, and they have the same period as the causative ocean wave. The typical period of primary microseisms is between 10-20 seconds. Secondary or double-frequency microseisms are generated by standing waves and have approximately half the period of the causative ocean wave, with a typical range of 3-10 seconds. Microseisms travel primarily as Rayleigh waves, with a retrograde ellipsoidal motion and a propagation speed of 3000 ms^{-1} [5], and are detectable at seismic stations on land and can be observed throughout the world. The signals vary in space and time as the ocean wavefield conditions are constantly changing, driven by atmospheric forcings.

Inversion of secondary microseism amplitudes for ocean wave height estimation has been done before but never from an evolutionary approach. Both Bromirski [6] and Ardhuin et al. [7] determined ocean wave heights at a buoy located off the Californian coast using seismic data from a single inland station quite successfully. Empirical methods were used [6] for this inversion which required time series information for the buoy including directional properties of the ocean wave field and microseism source location.

B. Previous Work

Evolutionary approaches have been used before in wave studies. Artificial Neural Networks have been used for real time wave forecasting [8], improving wave predictions [9], the hindcasting of storm waves [10] and estimating ocean wave parameters from wave spectra [11]. Understanding the interrelationship between wave parameters has also been investigated [12]. Genetic programming has been used in real time wave forecasting [13] and the estimation of ocean wave heights, with wind information as the input [14] which was quite successful over the 50 day period presented.

More recently a model of microseism noise generation and propagation was introduced to determine possible source locations and, using empirical methods to retrieve the ocean wavefield frequency spectrum from seismic noise spectrum it is possible to estimate SWH at that buoy location [7].

These methods require more information than is currently available to us, and so an evolutionary approach has been adopted which was entirely data driven, with source location undefined and limited buoy data (one SWH measurement per hour), with no access to time series information. The success of the method thus depends entirely on the correlation and consistency of the data collected and used for training and testing.

The methods applied in this study attempt to estimate SWH at a buoy located off the Northwest coast of Ireland, using seismic data from multiple seismic stations distributed throughout Ireland (Fig. 1).

III. Evolutionary Approach

Grammatical Evolution (GE) [15], [1] is a grammar-based form of Genetic Programming (GP) [16], [17], which specifies the syntax of potential solutions in a grammar; this grammar is used to map evolved genotypic strings into syntactically correct, functional phenotype strings.

GE exhibits similar performance to GP on the symbolic regression domain [1], while its grammar provides extra control over the syntax of evolved programs, both in terms of biases [18], [19] and data-structures used. This allows GE to be applied to a multitude of domains, such as the evolution of interpolating models of CO^2 flux [20], Financial Modelling [21], horse gait optimisation [22], and optimisation of controllers for video-games [23].

One of the main advantages of using symbolic manipulation techniques is that they create human-readable structures. These can then be further analysed (and potentially optimised) by domain experts, thus providing an advantage over other black-box machine learning methods.

A. Mapping Process

GE is typically applied in a modular process. A variable-length Genetic Algorithm (GA) creates/evolves binary or decimal strings; these strings then go through a mapping process, using a grammar, which creates syntactically correct solutions. These solutions can then be evaluated in the problem domain, and their performance measure returned to the GA.

Grammars in GE are typically represented in a Backus-Naur Format (BNF). To illustrate the mapping process, consider the BNF grammar shown in Fig. 2. This grammar is composed of three *non-terminal* symbols (<e>, <o> and <v>) and six *terminal* symbols (+, −, *, /, x and 1.0). Given a genotype string composed of the sequence of integers (4, 5, 8, 4, 3, 1, 9, 7), a program (phenotype) can be constructed, which respects the syntax specified in the grammar.

```
<e> ::= <e> <o> <e>
      | <v>
<o> ::= + | - | * | /
<v> ::= x | 1.0
```

Fig. 2. Example GE Grammar.

This process works by using each integer to choose productions from the grammar, mapping a given start symbol (typically, the first non-terminal symbol defined in the grammar) to a sequence of terminal symbols. In this example, the first integer (4) is used to choose one of the two productions of the start symbol <e>, through the formula $4\%2 = 0$, i.e. the remainder of the division (modulus) of the integer value by the number of productions of the symbol <e>. This means that the first (0^{th}) production is chosen, transforming <e> into <e> <o> <e>, which becomes the mapping string under construction.

The following integer (5) is then used with the leftmost unmapped symbol in the mapping string, so through the formula $5\%2 = 1$ the first symbol <e> is replaced by <v>, and thus the mapping string becomes <v> <o> <e>.

The mapping process continues in this fashion, so in the next step the mapping string becomes x <o> <e> through the formula $8\%2 = 0$, then x + <e> through $4\%4 = 0$, and x + <v> through $3\%2 = 1$. Finally, the remaining non-terminal symbol is mapped with $1\%2 = 1$, and the final expression becomes x + 1.0, which can then be evaluated.

IV. Experiments

A. Input Data

Microseism theory states that the amplitude of displacement is proportional to the square of the causative ocean wave [5]; in this case, the height of the (partial-) standing wave. There is a clear relationship between the amplitude of the microseisms recorded on land and ocean wave heights, as seen in Fig. 3 (top and bottom panels).

In order to show the relationship between seismic data and buoy data, the seismic data is averaged in the same way as the raw buoy data, before being transmitted and stored as the hourly SWH. The hourly microseism data is then bounded between an upper and lower limit, to exclude effects due to anthropogenic and non-oceanic processes. The data is then median filtered over three points. Each input is then scaled by a factor of 1×10^6, making the relationship between the two datasets easier to see, as illustrated in Fig. 3 (middle and bottom panels).

Inputs to the evolutionary methods are in the form of hourly microseism amplitude, from five seismic stations distributed across Ireland (VAL, IAP07, IA004, DSB and IDGL, see Fig. 1), and the output is in the form of significant wave height recorded at buoy K4. The period of data collection was from 1^{st} September 2011 to 30^{th} June 2013, approximately 638 days; however, due to discontinuous data at times, from both the seismometers and the buoy, approximately 610 days (14640 samples) of data are used.

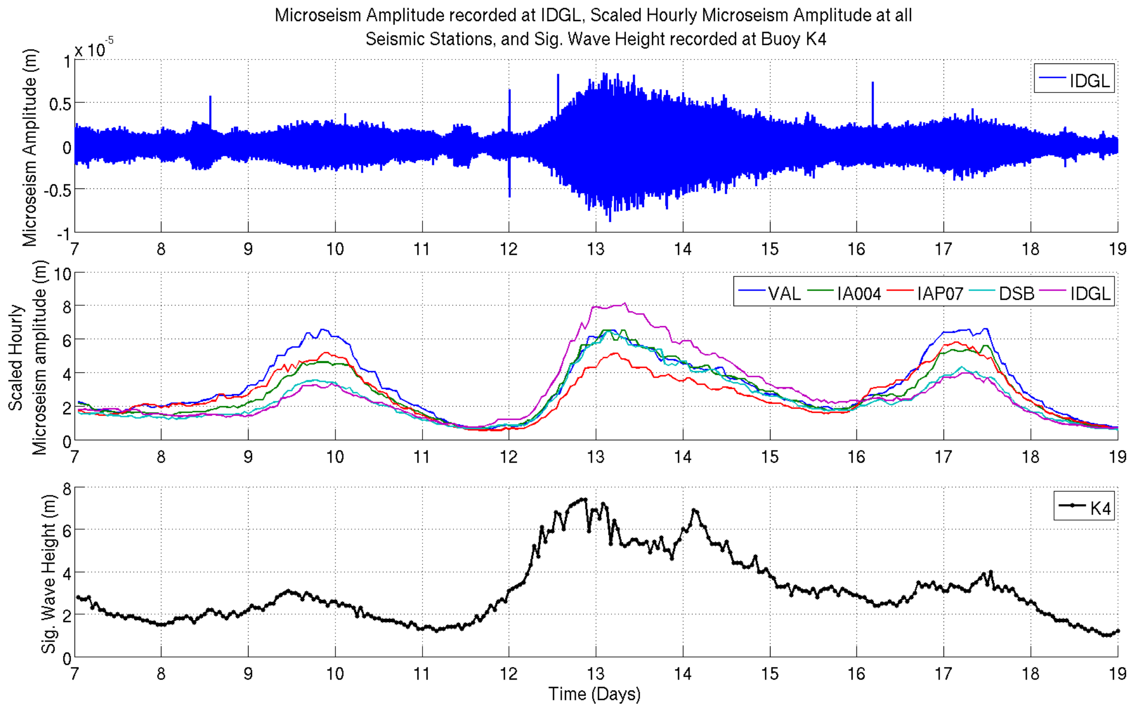

Fig. 3. Top panel: microseism amplitude recorded on land at station IDGL. Middle panel: resampled and scaled seismic amplitudes from all five buoys. Bottom panel: SWH measured at UK buoy K4. Data collection was from 8^{th} to 19^{th} May 2012.

SWH data is seasonal, in that in winter and summer months, the wave climate is expected to be at its most and least energetic (respectively). As such, the day of the year was also used as an input variable. In order to reduce the linear cumulative numerical weight of this variable, it was transformed into a fuzzy value (as seen in previous studies [20]); Fig. 4 shows the transformation used.

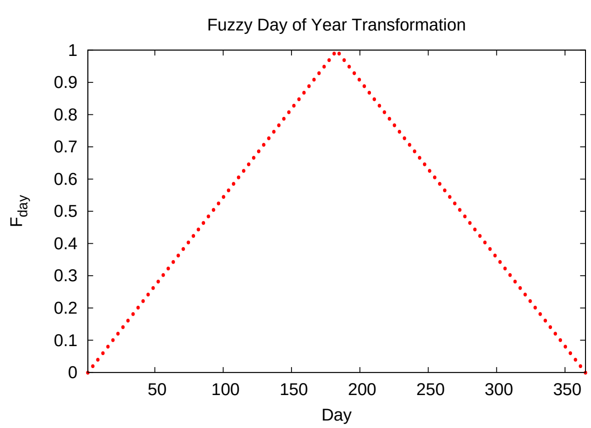

Fig. 4. Fuzzy day of year transformation.

B. Evolutionary Setup

Table I shows the parameters used for this study; these are standard GE parameters [1], and no attempt was made to optimise them. The initial genotypic strings were created using a reverse mapping process called *Sensible Initialisation* [24]; artificial non-mapping sections (*tails*) were appended to these (as per previous studies [25]). A "fair" tournament selection was used, where every individual participates at least in one tournament event. Finally, genetic operators were applied only to mapping regions of chromosomes.

TABLE I
EVOLUTIONARY SETUP

Population Size	500
Generations	50
Derivation-tree Max Depth (for initialisation)	10
Tail Ratio (for initialisation)	50%
Selection Tournament Size	1%
Elitism (for generational replacement)	10%
Crossover Ratio	50%
Average Mutation Events per Individual	1

All five buoy inputs, along with the day of year fuzzy transformation, were used in the GE grammar (see Fig. 5). Note that this grammar was automatically transformed, such that the bias of choice between recursive and non-recursive grammar productions became the same [19], [25]. The resulting grammar specifies the exact same syntax. Also notice the use of a reduced number of non-terminal symbols, as this has been shown [18] to improve the performance of GE.

```
<e> ::= + <e> <e>
      | - <e> <e>
      | * <e> <e>
      | / <e> <e>
      | VAL[i]
      | IAP07[i]
      | IA004[i]
      | DSB[i]
      | IDGL[i]
      | FuzzyDay[i]
      | <d><d>.<d>
<d> ::= 0 | 1 | 2 | 3 | 4
      | 5 | 6 | 7 | 8 | 9
```

Fig. 5. BNF grammar used for the experiments.

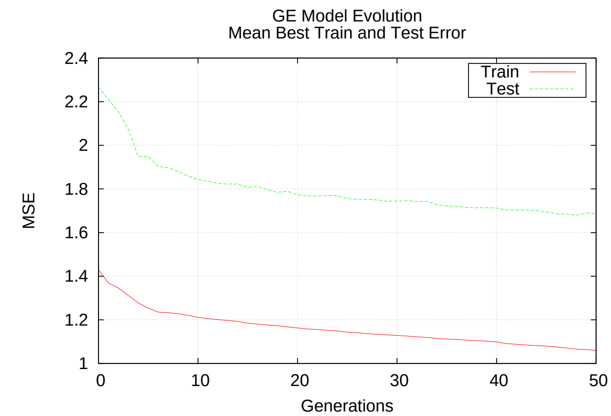

Fig. 6. Training and test performance for the mean best individual per generation (averaged across 50 runs).

C. Control Experiment

In order to validate the results obtained with GE, a Supervised Feed-Forward Backpropagation Artificial Neural Network (ANN) was used in this study. A three-layered neural network was applied, consisting of an input layer, a single hidden layer, and an output layer. The Levenberg-Marquardt optimisation algorithm was chosen for training. There were five inputs: amplitude at the specified seismic stations throughout Ireland, and a single output, SWH at buoy K4.

D. Measuring Performance

Both methods used a training period from 1^{st} September 2011 to the 31^{st} January 2013, using the Mean Squared Error (MSE) as a fitness measure. Once models were derived for this period, they were tested in a period of unseen data, from the 1^{st} February until the 30^{th} June 2013, comprising of 3387 datapoints. This allowed us to measure the degree of learning of the applied methods, and to monitor potential overfitting of the models to the training period.

V. Results & Analysis

100 independent runs were performed with GE; Fig. 6 plots the mean best individual per generation, both in seen (training) and unseen (test) data. It shows that on average, GE runs continuously reduce the training error, with a good correlation to test error. Although there are clear signs of overfitting occurring, the main trend is of test performance improvement.

For the ANN, training stopped after 12 iterations with a best validation performance of 1.0143 at epoch 6 and a gradient of 0.01712 at epoch 12 with 6 validation checks. Table II shows error measures (MSE and RMSE) and correlation value and coefficients for both methods. It shows that the ANN exhibits slightly better training performance, at the expense of test performance, where the best GE model is slightly better. The differences are quite small, thus validating both approaches.

Fig. 7 plots the SWH estimates of GE and the ANN, for the test period. It shows that both models agree quite well with the data measured at buoy K4. The estimates produced follow the overall trend of the ocean wavefield climate for this test period, and produce correlation coefficients of $r = 0.7895$ and $r = 0.7714$ for GE and the ANN, respectively.

When analysed more closely, there are times at which both models significantly under- or overestimate the SWH recorded. The training data contains wave heights up to 12-13m so the underestimations in the second half of March, April and May are not due to this. For the 15m SWH in February, the ANN was never trained on such large waveheights and given the limited extrapolation capabilities of ANNs it is possible that this is a result of insufficient training; the GE model does slightly better in this aspect. There is a notable overestimation of SWH at the beginning of March, with estimations following the trend of the wave record but overestimating by approximately 1.5m over the entire wave event. Also, if we look at the end of May and June we can see that the estimate is not in phase with the measured data.

These approaches are entirely data driven, and so input data dictates the quality of the estimate produced. For times when we know that the network is sufficiently trained, in this case on waveheights up to between 12 and 13m, and the estimate is an under- or overestimation of the measure SWH we must look at the input test data. It is possible that the SWHs recorded at the ocean buoy, the output, are not the exact ones which generated the secondary microseisms, the inputs, due to the spatio-temporal variability of the microseism sources. In this case neither model will be capable of resolving the SWH at the buoy because they are not directly related to the inputs. It is for this reason that data selection is so important. Another example can be seen in June 2013 when the estimates produced are out of phase with the target wave heights. Again, this is a result of the causative ocean waves and the ocean buoys not being spatially local. Until more is understood about the exact cause of these delays, it t

TABLE II

ERROR AND CORRELATION MEASURES FOR THE GE AND ANN APPROACHES

	Train Period		Test Period				
	MSE	RMSE	MSE	RMSE	r	m	c
GE	0.8577	0.9261	1.1289	1.0625	0.7895	0.6307	1.0985
ANN	0.732	0.8586	1.2040	1.0973	0.7714	0.6559	1.1476

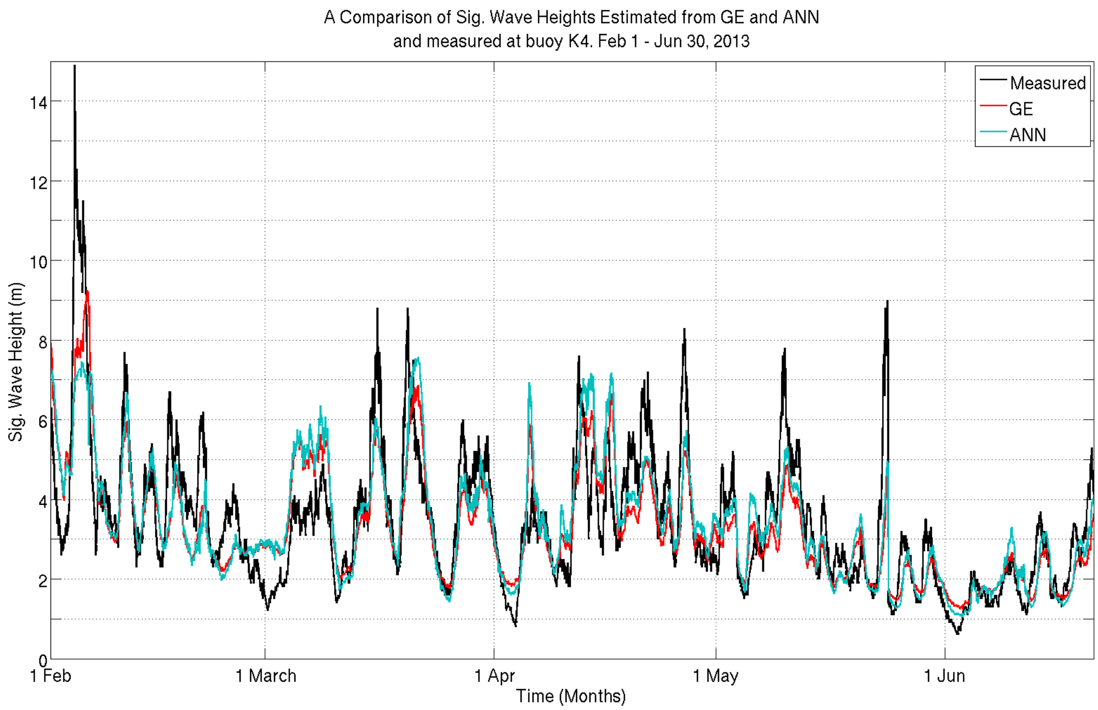

Fig. 7. Measured and estimated SWH with the GE and ANN approaches.

will limit the predictive capability and real time estimation of ocean waves.

When there is no delay between the ocean wave height and microseism amplitude, and their correlation is high, the resulting estimates are in good agreement with the measured wave heights. This can be seen for much of the month of March, at all wave heights, with a correlation coefficient from March 12th to 31st of $r = 0.8240$ for GE, and of $r = 0.8446$ for the ANN. Given these high correlation values, it is clear that it is possible to use microseism amplitudes to estimate ocean wave heights from an evolutionary approach.

One of the advantages of using symbolic combination models, such as GE, is that they are capable of producing human-readable models, which can be further analysed: equation 1 shows the best evolved model. It is interesting to see that only data from the IDGL, VAL and IAP07 sensors is used; from Fig. 1, one can see that these are respectively the northernmost and southernmost sensors, showing that their different amplitudes and distances to the source location are exploited by the model, with IDGL (the northernmost sensor, and closest to the source) being the principal variable. Also interesting is the use of the fuzzy day of year input, to dampen the summer estimates.

$$SWH_p = \frac{2 + IDGL + \frac{VAL}{IAP07} - F_{time}}{1.8} \quad (1)$$

Fig. 8 plots the correlation between the measured and estimated SWH for both methods. There appears to be saturation in the estimates produced by the ANN, with no estimation greater than approximately 7.5m produced. This could be due to the training phase of the network development. The frequency of the larger SWH is much lower than smaller SWH and could result in the network considering SWH beyond a particular range, outliers, applying a bias to a damped estimate. Ideally the network would be trained equally across the entire range of output values but this is not always possible.

The GE model also exhibits a slight saturation, but around the 9m value. This saturation is however entirely data driven, highlighting once again the importance of the quality of the source data.

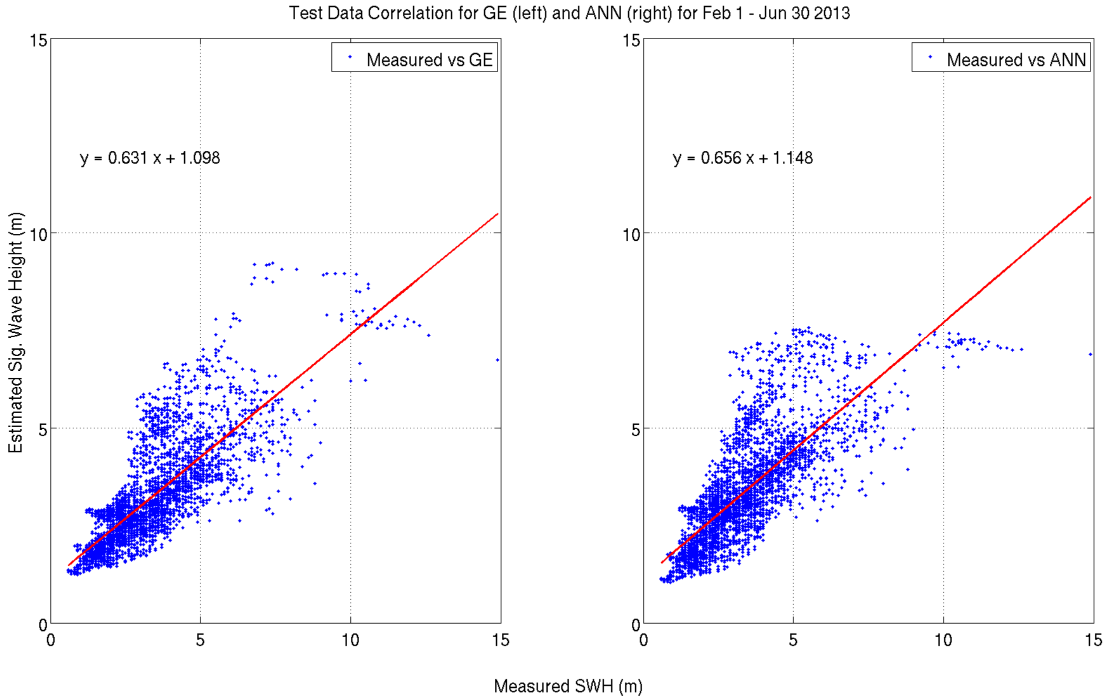

Fig. 8. Correlation plots for GE (left) and the ANN (right), for the test period.

VI. Conclusions

One of the main benefits of the approach presented is the ease of the data collection. It is approximately one order of magnitude cheaper to maintain a terrestrial network of microseism sensors than it is to maintain a network of in situ ocean buoys, and if there is a problem with one of the stations, it is much easier to fix or replace.

Considering the limited information available on microseism source location, the estimations obtained through GE and ANN models are quite good. There are clear improvements in the correlation coefficient between estimation and measured SWH when the correlation between test inputs and outputs are high, and there is no phase delay between them. If a delay exists, the ability of the network to estimate real time SWH is hindered. As information becomes available on source location of the microseisms, only time periods when microseisms are being generated by SWH recorded at or near a buoy will be used for testing of the model. In practice, a model is required for each of the buoy locations (Fig.1), in order to improve full wavefield estimations. Once trained, information regarding source location will be considered to determine which model is suitable and which ocean waves may be estimated at that time period.

There are many possible avenues of future work to explore. The grammar used with GE is specifically constrained to generate coarse constants, in order to minimise constant overfitting and improve learning; the constants in the generated models can thus be further optimised (no effort has been done to achieve this, in this study).

The evolutionary process with GE still showed signs of improving, beyond the pre-determined number of generations. It will be interesting to investigate how far can evolution go, while keeping the resulting models compact and understandable. Also, to avoid overfitting occurring, a three-set methodology (train, validation and test) will be used.

Future work will also include a similar model development for each of the buoy locations, and utilisation of microseism source location to determine suitability of these models in real time, for accurate, continuous SWH estimates in the Northeast Atlantic.

One of the main incentives towards using a land based monitoring system for the estimation of ocean wave parameters can be clearly seen this winter 2013, with storms in the Northeast Atlantic, so severe that they caused significant damage to the ocean buoy infrastructure. Now, rather than a gap in the ocean wavefield record for these buoy locations, one can apply the method presented here and reconstruct with confidence the ocean wave height parameters at that time.

References

[1] M. O'Neill and C. Ryan, *Grammatical Evolution - Evolutionary Automatic Programming in an Arbitrary Language*, ser. Genetic Programming. Kluwer Academic, 2003, vol. 4.

[2] B. D. Tapley, D. P. Chambers, C. K. Shum, R. J. Eanes, and J. C. Ries, "Accuracy assessment of the large-scale dynamic ocean topography from topex/poseidon altimetry," *Journal of Geophysical Research*, vol. 99, no. C12, pp. 24,605–24,617, 1994.

[3] M. Roca, S. Laxon, and C. Zelli, "The envisat ra-2 instrument design and tracking performance," *IEEE Transactions On Geoscience And Remote Sensing*, vol. 47, no. 10, pp. 3489–3506, 2009.

[4] F. Ardhuin, E. Stutzmann, M. Schimmel, and A. Mangeney, "Ocean wave sources of seismic noise," *Journal of Geophysical Research*, vol. 116, no. C9, pp. n/a–n/a, 2011.

[5] M. S. Longuet-Higgins, "A theory of the origin of microseisms," *Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences*, vol. 243, no. 857, pp. 1–35, 1950.

[6] P. D. Bromirski, R. E. Flick, and N. Graham, "Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the ne pacific," *Journal of Geophysical Research*, vol. 104, no. C9, pp. n/a–n/a, 1999.

[7] F. Ardhuin, A. Balanche, E. Stutzmann, and M. Obrebski, "From seismic noise to ocean wave parameters: General methods and validation," *Journal of Geophysical Research*, vol. 117, no. C5, pp. n/a–n/a, 2011.

[8] P. Jain and M. C. Deo, "Real-time wave forecasts off the western indian coast," *Applied Ocean Research*, vol. 29, no. 1-2, pp. 72–79, 2007.

[9] O. Makarynskyy, "Improving wave predictions with artificial neural networks," *Ocean Engineering*, vol. 31, no. 5-6, pp. 709–724, 2004.

[10] S. Rao and S. Mandal, "Hindcasting of storm waves using neural networks," *Ocean Engineering*, vol. 32, pp. 667–684, 2005.

[11] S. Mandal, S. Rao, and D. H. Raju, "Ocean wave parameters estimation using backpropagation neural networks," *Marine Structures*, vol. 18, pp. 301–318, 2005.

[12] J. D. Agrawal and M. C. Deo, "Wave parameter estimation using neural networks," *Marine Structures*, vol. 17, pp. 536–550, 2004.

[13] S. Gaur and M. C. Deo, "Real-time wave forecasting using genetic programming," *Ocean Engineering*, vol. 35, no. 11-12, pp. 1166–1172, 2008.

[14] S. P. Nitsure, S. N. Londhe, and K. C. Khare, "Application of genetic programming for estimation of ocean wave heights," in *World Congress on Nature Biologically Inspired Computing, NaBIC 2009, Coimbatore, India, December 9-11, 2009, Proceedings*. IEEE Press, 2009, pp. 1520–1523.

[15] C. Ryan, J. Collins, and M. O'Neill, "Grammatical evolution: Evolving programs for an arbitrary language," in *Genetic Programming, First European Workshop, EuroGP 1998, Paris, France, April 14-15, 1998, Proceedings*, ser. Lecture Notes in Computer Science, W. Banzhaf, R. Poli, M. Schoenauer, and T. C. Fogarty, Eds., vol. 1391. Springer, 1998, pp. 83–95.

[16] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.

[17] R. I. McKay, N. X. Hoai, P. A. Whigham, Y. Shan, and M. O'Neill, "Grammar-based genetic programming - a survey," *Genetic Programming and Evolvable Machines*, vol. 11, no. 3-4, pp. 365–396, 2010.

[18] M. Nicolau, "Automatic grammar complexity reduction in grammatical evolution," in *Genetic and Evolutionary Computation - GECCO 2004, Genetic and Evolutionary Computation Conference, Seattle, WA, USA, June 26-30, 2004, Workshops, Proceedings*, R. Poli, S. Cagnoni, M. Keijzer, E. Costa, F. Pereira, G. Raidl, S. C. Upton, D. E. Goldberg, H. Lipson, E. de Jong, J. R. Koza, H. Suzuki, H. Sawai, I. Parmee, M. Pelikan, K. Sastry, D. Thierens, W. Stolzmann, P. L. Lanzi, S. W. Wilson, M. O'Neill, C. Ryan, T. Yu, J. F. Miller, I. Garibay, G. Holifield, A. S. Wu, T. Riopka, M. M. Meysenburg, A. W. Wright, N. Richter, J. H. Moore, M. D. Ritchie, L. Davis, R. Roy, and M. Jakiela, Eds., 2004.

[19] R. Harper, "GE, explosive grammars and the lasting legacy of bad initialisation," in *IEEE Congress on Evolutionary Computation, CEC 2010, Barcelona, Spain, July 18-23, 2010, Proceedings*. IEEE Press, 2010, pp. 2602–2609.

[20] M. Nicolau, M. Saunders, M. O'Neill, B. Osborne, and A. Brabazon, "Evolving interpolating models of net ecosystem co2 exchange using grammatical evolution," in *Genetic Programming, 15th European Conference, EuroGP 2012, Malaga, Spain, April 11-13, 2012, Proceedings*, ser. Lecture Notes in Computer Science, A. Moraglio, S. Silva, K. Krawiec, P. Machado, and C. Cotta, Eds., vol. 7244. Springer, 2012, pp. 134–145.

[21] A. Brabazon and M. O'Neill, *Biologically Inspired Algorithms for Financial Modelling*. Springer, 2006.

[22] J. E. Murphy, M. O'Neill, and H. Carr, "Exploring grammatical evolution for horse gait optimisation," in *Genetic Programming, 12th European Conference, EuroGP 2009, Tübingen, Germany, April 15-17, 2009, Proceedings*, ser. Lecture Notes in Computer Science, L. Vanneschi, S. Gustafson, A. Moraglio, I. D. Falco, and M. Ebner, Eds., vol. 5481. Springer, 2009, pp. 183–194.

[23] D. Perez, M. Nicolau, M. O'Neill, and A. Brabazon, "Evolving behaviour trees for the mario ai competition using grammatical evolution," in *Applications of Evolutionary Computation - EvoApplications 2011: EvoCOMNET, EvoFIN, EvoHOT, EvoMUSART, EvoSTIM, and EvoTRANSLOG, Torino, Italy, April 27-29, 2011, Proceedings, Part I*, ser. Lecture Notes in Computer Science, C. Di Chio, S. Cagnoni, C. Cotta, M. Ebner, A. Ekárt, A. I. Esparcia-Alcázar, J. J. Merelo, F. Neri, M. Preuss, H. Richter, J. Togelius, and G. N. Yannakakis, Eds., vol. 6624. Springer, 2011, pp. 123–132, best paper award.

[24] C. Ryan and A. Azad, "Sensible initialisation in grammatical evolution," in *Genetic and Evolutionary Computation Conference (GECCO) Workshops*, Erick Cantú-Paz et al., Ed. AAAI, 2003.

[25] M. Nicolau, M. O'Neill, and A. Brabazon, "Termination in grammatical evolution: Grammar design, wrapping, and tails," in *IEEE Congress on Evolutionary Computation, CEC 2012, Brisbane, Australia, June 10-15, 2012, Proceedings*. IEEE Press, 2012.