Evolutionary data-modelling of an innovative low reflective vertical quay

C. Altomare, X. Gironella and D. Laucelli

ABSTRACT

Vertical walls are commonly used as berthing structures. However, conventional vertical quays may have serious technical and environmental problems, as they reflect almost all the energy of the incident waves, thus affecting operational conditions and structural strength. These drawbacks can be overcome by the use of low reflective structures, but for some instances no theoretical equations exist to determine the relationship between the reflection coefficient and parameters that affect the structural response. Therefore, this study tries to fill this gap by examining the wave reflection of an absorbing gravity wall by means of evolutionary polynomial regression, a hybrid evolutionary modelling paradigm that combines the best features of conventional numerical regression and genetic programming. The method implements a multi-modelling approach in which a multi-objective genetic algorithm is used to get optimal models in terms of parsimony of mathematical expressions and fitting to data. A database of physical laboratory observations is used to predict the reflection as a function of a set of variables that characterize wave conditions and structure features. The proposed modelling paradigm proved to be a useful tool for data analysis and is able to find feasible explicit models featured by an appreciable generalization performance.

Key words | data-mining, evolutionary polynomial regression, low reflective vertical quay, wave reflection

C. Altomare (corresponding author)
X. Gironella
Laboratori d'Enginyeria Marítima,
Universitat Politécinica de Catalunya,
Carrer Jordi Girona 1–3,
UPC Campus Nord,
Edifici D1, 08034,
Barcelona,
Spain
E-mail: *corrado.altomare@upc.edu*

D. Laucelli
Department of Civil Engineering and Architecture,
Politecnico di Bari,
Via E. Orabona 4,
70125 – Bari,
Italy

NOMENCLATURE

A	dimensionless variable defined as $A = H_s/d - h_c$;
a_j, a_0	constant coefficients in the general EPR model structure;
avg(y_{exp})	average value of observations;
a, b	coefficients in Ahrens & Seelig (1980) formula;
C_r	reflection coefficient;
d	water depth;
D_{50}	nominal diameter of armour units;
E_i	incident wave energy;
E_r	reflected wave energy;
ES(j,k)	vector of exponents for the jth term, for the kth explanatory variable;
g	gravitational acceleration;
h_c	height of the base of the chamber with respect to sea bottom;
H_i	incident wave height;
H_r	reflected wave height;
H_s	significant wave height;
K	the number of armour layers;
I	dimensionless variable defined as $= \sqrt{\frac{d}{g \cdot T_p^2}}$;
L	wave length;
L_0	deepwater water length;
m	number of terms of the pseudo-polynomial expression;
N	number of data;
R	coefficient of correlation;
R	dimensionless reflection number (Davidson *et al.* 1996);
R^*	dimensionless variable defined as $R^* = \frac{d \cdot L_0^2}{H_s \cdot (0.5 D_{50})^2}$;

doi: 10.2166/hydro.2012.219

s_0	deepwater wave steepness defined as $s_0 = H_s/L_0$;
T	mean wave period;
T_p	wave peak period;
X	matrix of input variables in EPR;
X_k	kth input column vector in EPR;
$\hat{Y}$	matrix of estimated outputs;
$\hat{y}$	estimated value of the process;
y_{exp}	vector of experimental observations;
ξ	Iribarren number;
ξ_h	inverted Iribarren parameter, as in Hughes & Fowler (1995);
φ	slope of the seabed;
θ	slope angle of the rubble mound;
CoD	Coefficient of determination;
SSE	Sum of squared errors.

INTRODUCTION

Wave action on natural environment and engineering structures can cause different kinds of problems that need to be predicted to prevent possible unexpected effects. Analysing the response of coastal structures to wave action is a very challenging task for engineers; in particular, a detailed characterization of the effects of quays or breakwaters is a key point for design, involving technical, economic and environmental aspects. However, it is not easy to achieve a holistic representation of the involved hydrodynamic phenomena, and in particular of wave reflection. To this aim, prototype measurements and physical laboratory tests can provide information on coastal structures and their behaviour, but most of the time it is difficult to characterize the main aspects of such phenomena. Recently, the increasing availability of different kind of data and information allows the application of innovative data-mining techniques to study and predict wave action and its effects (Zhang *et al.* 2006; Zamani *et al.* 2009; Zanganeh *et al.* 2011; Bhattacharya *et al.* 2012. This work investigates the efficiency of a recently developed hybrid data-driven modelling technique, the evolutionary polynomial regression (EPR; Giustolisi & Savic 2006, 2009), to analyse the wave energy reflection (in terms of coefficient of reflection) of an innovative low reflective quay.

Wave reflection inside a harbour is an undesirable phenomenon, which may adversely affect operating conditions and the safety of people and the environment, e.g. increasing the wave agitation inside the seaport, making navigation and berthing difficult, or leading to overwhelming of the coastal structure. Therefore, a proper characterization of reflection due the coastal structures is a key issue in harbour design. The main cause of such problem is the conventional vertical quay, which reflects a large amount of the incident wave energy. A good solution can be the use of low reflective quays as they reduce wave reflection and decrease loads on the structures. At the same time, vertical dissipative quays work similarly to rubble mound slopes, having a lower environmental impact, i.e. concrete prefabricated structures instead of a huge volume of natural stones, a smaller area occupied by the quay resulting in an area less aggressive to the sea environment.

Low reflective quays are featured as porous or open structures, and have been studied experimentally over many years (Jarlan 1961, 1965; Ijima *et al.* 1976; Matteotti 1991; Fugazza & Natale 1992; Tanimoto & Takahashi 1994; Suh *et al.* 2006; Garrido *et al.* 2010; Taveira Pinto *et al.* 2011). Each of these studies, however, works well only on its own dataset, thus a general formulation (in terms of reflection coefficient) to relate the reflected wave energy to structural features and wave conditions does not exist. The reflection coefficient depends on the configuration of the quay, and thus it is very important to define its relationship with the parameters that affect the structural response in order to predict the behaviour of a low reflective vertical quay. Therefore, EPR will be used to find simple and easily interpretable mathematical model expressions of the phenomenon at stake, as well as to avoid a subjective choice of the input parameters that affect reflection. The multi-objective optimization paradigm embedded in the EPR technique helps in this direction, producing a wide range of possible optimal solutions to the problem as trade-off between accuracy and parsimony of the mathematical expressions (i.e. conflicting objectives; Giustolisi & Savic 2009). The EPR has already been applied successfully to other complex civil engineering problems, see for

example Javadi & Rezania (2009), Berardi *et al.* (2008), Doglioni *et al.* (2010) and Laucelli & Giustolisi (2011).

WAVE REFLECTION COEFFICIENT

A coastal structure reflects the incident waves and such reflection can be quantified by the bulk reflection coefficient, defined as:

$$C_r = \sqrt{\frac{E_r}{E_i}} = \frac{H_r}{H_i} \tag{1}$$

where H_r and H_i are the wave height of reflected and incident wave, respectively, and E_r and E_i are the related energies. The reflection depends both on the sea state and the structural layout; in particular, sloping structures behave differently than vertical structures and the structural response changes if the structure is overtopped or not.

Miche (1951) determined an empirical expression of the reflection coefficient for monochromatic waves on a smooth sloping structure as function of the slope angle, wave height and frequency. Ahrens & Seelig (1980), starting from Battjes (1974), improved the estimates of C_r for non-overtopped sloping structures using the following equation:

$$C_r = \frac{a\xi^2}{b + \xi^2} \tag{2}$$

where a and b are empirical coefficients and ξ is the Iribarren number (Battjes 1974), defined as:

$$\xi = \frac{\tan\theta}{\sqrt{H/L}} \tag{3}$$

where θ = the slope angle of the structure and H/L = the wave steepness, where L = wave length.

Ahrens & Seelig (1980) stated that wave reflection is ruled by H, L and θ for smooth breakwaters, while for rough porous structures there are additional variables to be considered, among them d = water depth and D_{50} = nominal diameter of the armour units. Therefore, using linear or nonlinear regression techniques, several authors analysed the values of a and b in Equation (2) for different kinds of coastal structure and ranges of ξ; they generally found that the reflection coefficient increases proportionally with ξ, this means that large wave periods induce high wave reflection.

Hughes & Fowler (1995) analysed wave reflection for sloping waterproof and rough structures, introducing an inverted Iribarren parameter ξ_h, producing the following expression for the reflection coefficient:

$$C_r = \frac{0.1415}{0.1415 + \xi_h^{0.804}} \tag{4}$$

with

$$\xi_h = \frac{\sqrt{\frac{d}{gT^2}}}{\tan\theta} \tag{5}$$

where d = the water depth, θ = slope angle of rubble mound, T = wave period and g = gravitational acceleration.

Equation (4) has a mathematical limit for $\xi_h \to 0$ that can be rarely reached for rubble mound structures, which typically have slopes lower that 1:1.5, and thus $\xi_h \to 0$ only for very long waves. However, for this reason, Hughes & Fowler's proposition is using Equation (4) within the range of the used input variables, i.e. seaward slope of 1:2 and $0.12 < \xi_h < 0.60$.

Based on full-scale experiments on rubble mound structures with slopes equal to 1:1.55 and 1:0.82, Davidson *et al.* (1996) introduced a dimensionless reflection number having the following expression:

$$R = \frac{dL_0^2 \tan\theta}{H_i D_{50}^2} \tag{6}$$

which has been found by means of a multiple regression analysis; L_0 = deepwater wave height and D_{50} = nominal diameter of the armour units. Then, Davidson *et al.* (1996) proposed a similar equation to Equation (4), using the dimensionless reflection number R instead of the Iribarren number:

$$C_r = \frac{0.635R^{0.5}}{41.2 + R^{0.5}} \tag{7}$$

where the wave length influence on the C_r is similar to the formulation of Ahrens & Seelig (1980), but the influence of wave height and slope angle is reduced.

For more complex structures (e.g. composite breakwaters, sloping top caissons, single perforated screens and perforated caissons), the reflection coefficient depends also on other structural characteristics, as presented by Allsop & Hettiarachchi (1988) for perforated caissons, usually showing non-linear relationship among C_r and the input variables, thus asking for the use of more complex methodologies in order to find reliable formulations.

This paper aims to face such a problem by analysing wave reflection of an absorbing gravity wall by means of the EPR, which is a hybrid evolutionary modelling paradigm combining the best features of conventional numerical regression and genetic programming (Giustolisi & Savic 2006, 2009; EPR-MOGA 2010).

THE EVOLUTIONARY POLYNOMIAL REGRESSION PARADIGM

Models are usually supported by physical knowledge of the phenomenon at stake. This fact led scientists to collect data and define the mathematical structure of the models, which were calibrated on measured data by numerical regression or alternative data-driven techniques. However, the generalization of experimental models outside their calibration range (i.e. different realization of the same phenomenon) can be eventually reduced, due to measured data often incomplete or affected by errors. Nevertheless, the availability of large and detailed databases, associated with increased computational capabilities, has motivated researchers to propose innovative techniques and methodologies to mine information from data.

Among these techniques, the EPR has been recently introduced in the Hydroinformatics Community as a hybrid data-driven technique, which combines the effectiveness of genetic programming with numerical regression for developing simple and easily interpretable mathematical model expressions (Giustolisi & Savic 2006). The EPR approach helps to avoid some drawbacks of other modelling approaches, such as physically based models and black-box data-driven models. In the first case, the construction of models can be difficult due to underlying mechanisms that may not be always fully understood, or to the need for much data, which is sometimes difficult to measure in the field. On the other hand, black-box models, for example artificial neural networks, are very effective in reproducing any database related to some observed phenomenon; however, inexpert users especially can experience some overwhelming problems such as model structure identification (i.e. selection of inputs and parameters estimation), over-fitting to training data, and the inability to exploit his/her physical insight about the phenomenon at stake into the learning process. The EPR can overcome these problems by means of an explicit model expression for the system under observation. One of the general model structures that EPR can manage is reported in the following:

$$\hat{\mathbf{Y}} = \left[a_0 + \sum_{j=1}^{m} a_j (\mathbf{X}_1)^{\mathbf{ES}(j,1)} \cdot \ldots \cdot (\mathbf{X}_k)^{\mathbf{ES}(j,k)} \right] \tag{8}$$

where m is the number of additive terms, a_j are numerical parameters to be estimated, $\mathbf{X}_i$ are candidate explanatory variables, $\mathbf{ES}(j,z)$ (with $z = 1,\ldots,2k$) is the exponent of the zth input within the jth term in Equation (8), and f is a function selected by the user among a set of possible alternatives (including no function selection). The exponents $\mathbf{ES}(j,z)$ are selected from a user-defined set of candidate values (which should include 0). In brief, the search for model structure is performed by exploring the combinatorial space of exponents to be assigned to each candidate input of Equation (8). Thus, although exponent values could be any real number, they are coded as integers during the search procedure. It is worth noting that, when an exponent is $= 0$, relevant input $\mathbf{X}_i$ is basically deselected from the resulting equation. This, in turn, reduces the complexity of final mathematical expressions.

The search for model mathematical structures (i.e. combinations of exponents) is performed through a population-based strategy that mimics the evolution of the individuals in nature. In particular it employs a genetic algorithm (Goldberg 1989) whose individuals are represented by the sets of exponents in Equation (8). It is worth noting that EPR mathematical structures, e.g. Equation (8), are linear with respect to their parameters although not necessarily linear in their

attributes (due to both exponents different from one and possible selection of function f). They are referred to as pseudo-polynomials as they are obtained by adding a number of monomial terms (i.e. argument of sum in Equation (8)). Model parameters are computed from data by solving a linear inverse problem in order to guarantee a two-way (i.e., unique) relationship between each model structure and its parameters. More details can be found in Giustolisi & Savic (2006).

MULTI-OBJECTIVE FRAMEWORK FOR EPR

In recent years, the EPR has been embedded in a multi-objective search paradigm (EPR-MOGA) (Giustolisi & Savic 2009; EPR-MOGA 2010) in order to develop multiple models by simultaneously optimizing fitness to training data and parsimony of resulting mathematical expressions. Generally, in regression-based modelling, fitness usually refers to a measure of how closely the regression expression fits the data points, but it is widely accepted that the best modelling approach is also the simplest that fits the purpose of the application, according to the so-called principle of parsimony: for a set of otherwise equivalent models of a given phenomenon, it is advisable to choose the simplest one to explain a set of data. This principle, often called Occam's razor, is attributed to the medieval philosopher William of Occam (or Ockham 1300–1349). Therefore EPR-MOGA searches for the best models as trade-offs of model accuracy and parsimony (Giustolisi & Savic 2006).

The multi-objective analysis in EPR-MOGA is achieved by means of a multi-objective genetic algorithm (Goldberg 1989), based on the Pareto dominance criterion (Pareto 1896), named OPTImized Multi-Objective Genetic Algorithm (OPTIMOGA) (Laucelli & Giustolisi 2011). The EPR-MOGA helps for a sudden understanding of existing patterns in data by comparing different optimal models and selecting the model that best fits the analysis purposes. Once the EPR-MOGA search options (candidate model attributes, candidate exponents for attributes, maximum number of parameters, etc.) has been defined, it returns a set of Pareto-optimal models considering three conflicting objective functions: (1) maximization of model accuracy; (2) minimization of the number of model coefficients; and (3) minimization of the number of model inputs; whereas the last two objectives are representatives of model parsimony.

From a regressive standpoint, the model search paradigm of EPR-MOGA has some significant beneficial features, not found in other data-driven techniques, such as: (i) a small number of constants to be estimated (i.e. helps avoiding over-fitting problems, especially for small datasets); (ii) a linear parameters estimation (assuring the unique solution is found when the inverse problem is well conditioned); (iii) an automatic model construction (avoiding the need to preselect the functional form and the number of parameters in the model); and (iv) a transparent form of the regression characteristics, which makes model selection easier, i.e. the multi-objective feature allows selection not only based on fitting statistics. As a consequence, once the Pareto set of optimal models is obtained, the analyst can evaluate the optimal models looking also at key aspects other than those encoded as objective functions, as for example: (i) the model structure with respect to the physical insight related to the problem; (ii) similarities of mathematical structures among EPR-MOGA Pareto set of models; (iii) recurrent groups of variables in different EPR-MOGA models; (iv) generalization performance of models as assessed in terms of both statistical analysis and mathematical parsimony; and (v) the reliability of experimental data used to build the model and/or final purpose of the model itself. This eventually results into a more robust selection. Figure 1 is explicative of the model selection process that can be achievable by the EPR-MOGA paradigm. For more details see Giustolisi & Savic (2009) and EPR-MOGA (2010).

In this work the authors used the EPR-MOGA modelling paradigm to analyse the wave reflection of an innovative vertical low reflective quay as described in the following sentence. Starting from a series of experimental data measured on a reduced scale model of the quay, the modelling strategy adopted here simply assumes that the user can define a roughly wide set of hypotheses (explanatory variables), which can be considered as influent on the phenomenon at stake, and the main features of the EPR pseudo-polynomial structure, thus leaving to the evolutionary multi-objective optimization paradigm the burden of the combinatorial search of the optimal model for C_r.

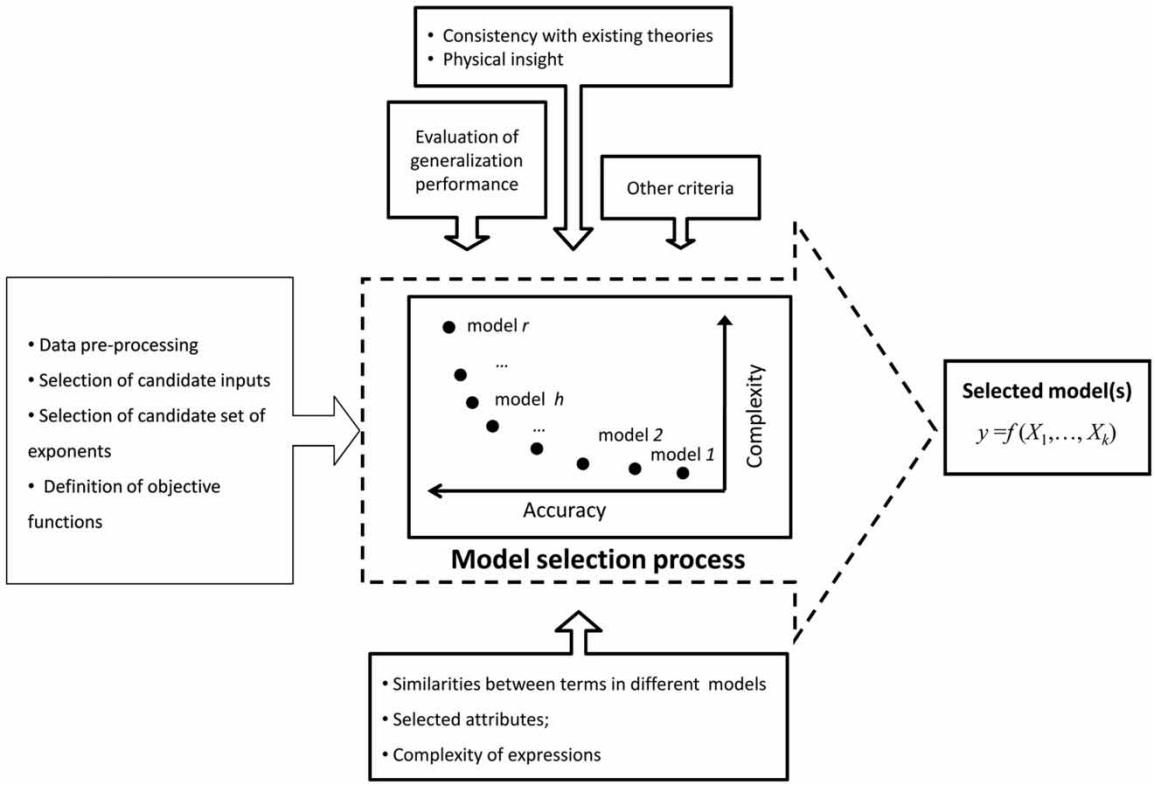

Figure 1 | Decision support framework for the model selection based on EPR-MOGA strategy.

CASE STUDY

The vertical low reflective quay here studied is characterized by a particular geometrical configuration, firstly introduced by Matteotti (1991). The wave reflection process is here tackled by means of the EPR-MOGA modelling paradigm, aiming to obtain a feasible expression for the reflection coefficient, useful for maritime structure design, physical and numerical modelling, as a function of the main involved parameters.

Figure 2 shows a schematic representation of the analysed quay. The innovative aspect of such structure is that the upper part is replaced by empty cells on its seaward side; such cells contain blocks of natural rock sloping down to the sea, thus opposing a very porous wall to the incident waves. This structure can be considered something between the completely permeable vertical face with an air chamber located in the rear side and a rock sloping breakwater.

Experiments at small scale (1:33) have been carried out at the Maritime Engineering Laboratory of the Technical University of Catalunya (CIIRC/LIM-UPC) in Barcelona (Spain), using two different porosities varying the dimension and distribution of stones that form the rubble mound inside the upper chamber. A theoretical prototype value, usually used in engineering applications for the rubble mound slope, has been considered and its equivalent porosity has been assessed (45%). Starting from the resulting distribution in stone diameter, the first part of experiments has been carried out using a porosity of about 32%, obtained by scaling the nominal diameter of stones in a geometrical way. The second part of the experiment has been carried out considering a correction to the porosity, to take into account the water turbulence within the stone layer if compared with a real scale case (Burcharth *et al.* 1999); in this case the used porosity was about 44%. Table 1 summarizes the porosities and the nominal diameters of stones used in the experiments.

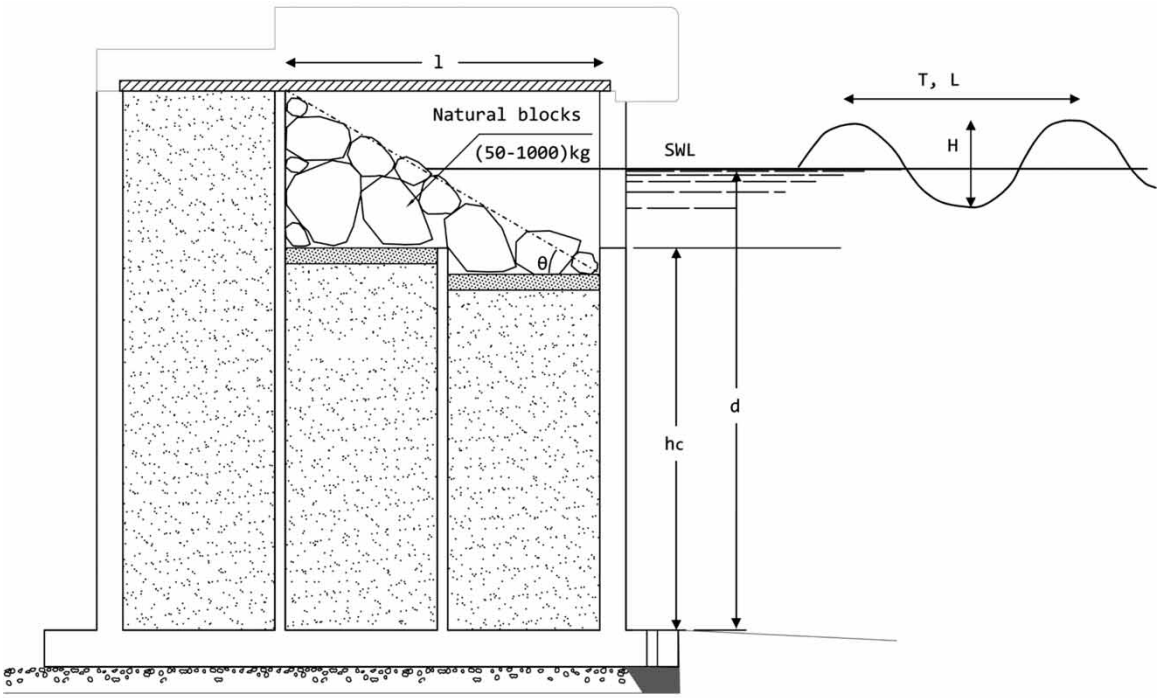

Figure 2 | Cross-section of the low reflective quay.

Table 1 | Porosity and nominal diameter used in laboratory tests (scale 1:33)

Porosity of rubble mound	D_{50} (m)	Number of tests
0.319	0.024	12
0.441	0.018	66

Figure 3 show a snapshot of the used prototype during experiments.

EXPERIMENTAL SETUP

The incident waves have been created using a small scale wave flume (named CIEMito), which is 18 m long, 0.38 m wide and 0.56 m high. It is equipped with a piston-type wave generator driven by a software code developed in LabView at CIIRC-LIM/UPC. The program allows generation of regular and random waves characterized by target spectrum, as well as target wave time series. Wave actions were measured using five resistive gauges with an accuracy of 1 mm, positioned in front of the caisson in order to get incident and reflected wave components.

As showed in Figure 4, the five gauges were positioned between the wave paddle and the structure: WG0 and WG1 are used to measure the transfer function, while WG2, WG3 and WG4 are used to measure the incident/reflected wave components at a distance of one wave length (at least) from the structure, in order to avoid non-linear effects. The position of WG0, WG1 and WG4 is fixed, while the other two gauges could be moved according to the wave length.

Figure 4 shows the distances from the paddle of the wave gauges, and the variable distances A and B, which assumed the value of 0.24–0.33 and 0.18–0.21–0.29, respectively.

The experiments conducted for the present study considered only irregular waves. Each irregular wave set included 300 waves and was characterized by JONSWAP spectra with a peak enhancement factor of 3.3. Table 2 reports the ranges of the input (wave height, peak period, wave length and steepness, water depth and main structural features) and output variables used for the wave tests.

The Goda & Suzuki (1976) and Mansard & Funke (1980) methods were used and compared to calculate the reflection

Figure 3 | Snapshot of the model during the experiments.

coefficient for each test. These methodologies are implemented within the WaveLab software of the Aalborg University (Andersen & Liu 2006). Table 3 reports all the experimental data as measured by means of the above described setup.

ANALYSIS OF THE EXPERIMENTAL RESULTS AND DATA SELECTION

The selection of the explanatory variables is an important issue in environmental modelling. Usually it is performed based on the general knowledge of the phenomenon, the physical insight of the analyst and it is influenced by the availability of measurements.

In this case, the laboratory experiments on the physical model in Figures 3 and 4 provided a number of data points for the input variables in Table 3. A preliminary analysis was thus conducted in order to understand the existing relationships among the input and output variables. Therefore, bearing in mind that the analysed structure has never been investigated before (i.e. no formulations are available) the starting point was to consider that wave reflection generally increases with wave period, as well as with wave length. This dependence can be expressed by different parameters, such as the Iribarren number (Battjes 1974), the Miche

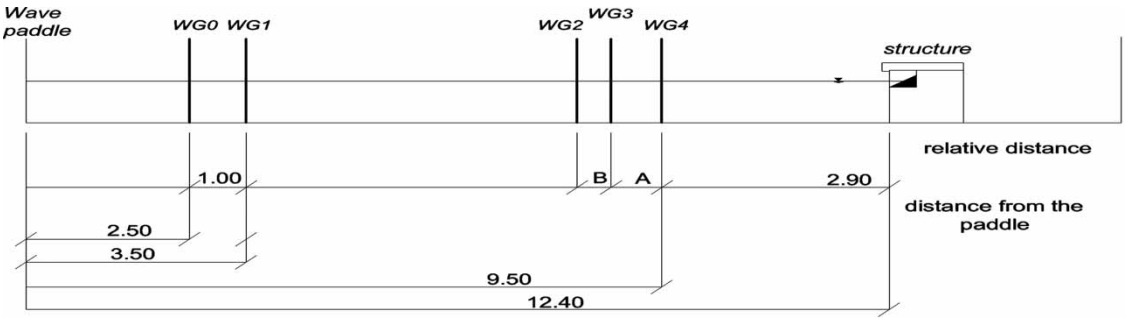

Figure 4 | Distribution of wave gauges along the flume.

Table 2 | Ranges of input variables for the irregular wave tests (model scale 1:33)

H_s (m)	T_p (s)	L_0 (m)	s_0	d (m)	h_c (m)	l (m)	D_{50} (m)
1.09–1.72	0.900–1.866	0.028–0.055	0.055–0.0412	0.275–0.293	0.245	0.18	0.018–0.024

Table 3 | Measured values for the input and output variables observed during the experiments (model scale 1:33)

C_r	s_0	L_0 (m)	H_s (m)	T_p (s)	d (m)	D_{50} (m)
Training set						
0.340	0.0212	1.56	0.033	1.000	0.293	0.024
0.364	0.0324	1.26	0.041	0.900	0.293	0.024
0.373	0.0263	1.56	0.041	1.000	0.293	0.024
0.389	0.0175	1.89	0.033	1.100	0.293	0.024
0.409	0.0147	2.25	0.033	1.200	0.293	0.024
0.425	0.0333	1.56	0.052	1.000	0.293	0.024
0.432	0.0141	2.19	0.031	1.185	0.293	0.018
0.433	0.0141	2.19	0.031	1.185	0.293	0.024
0.434	0.0217	1.89	0.041	1.100	0.293	0.024
0.451	0.0108	3.06	0.033	1.400	0.293	0.024
0.456	0.0163	2.19	0.036	1.185	0.293	0.024
0.456	0.0163	2.19	0.036	1.185	0.293	0.018
0.456	0.0261	1.26	0.033	0.900	0.275	0.024
0.456	0.0183	2.25	0.041	1.200	0.293	0.024
0.472	0.0131	2.54	0.033	1.276	0.293	0.018
0.474	0.0324	1.26	0.041	0.900	0.275	0.024
0.481	0.0134	3.06	0.041	1.400	0.293	0.024
0.485	0.0275	1.89	0.052	1.100	0.293	0.024
0.489	0.0155	2.54	0.039	1.276	0.293	0.024
0.489	0.0106	3.10	0.033	1.410	0.293	0.024
0.490	0.0156	2.64	0.041	1.300	0.293	0.024
0.500	0.0155	2.54	0.039	1.276	0.293	0.018
0.505	0.0412	1.26	0.052	0.900	0.275	0.024
0.509	0.0141	2.86	0.040	1.354	0.293	0.024
0.512	0.0212	1.56	0.033	1.000	0.275	0.024
0.513	0.0106	3.10	0.033	1.410	0.293	0.018
0.514	0.0231	2.25	0.052	1.200	0.293	0.024
0.514	0.0179	2.54	0.045	1.276	0.293	0.024
0.517	0.0141	2.86	0.040	1.354	0.293	0.018
0.519	0.0179	2.54	0.045	1.276	0.293	0.018
0.523	0.0170	3.06	0.052	1.400	0.293	0.024
0.529	0.0167	2.86	0.048	1.354	0.293	0.024
0.534	0.0083	3.99	0.033	1.600	0.293	0.024
0.540	0.0333	1.56	0.052	1.000	0.275	0.024
0.546	0.0197	2.64	0.052	1.300	0.293	0.024
0.546	0.0167	2.86	0.048	1.354	0.293	0.018
0.551	0.0117	3.51	0.041	1.500	0.293	0.024
0.558	0.0175	1.89	0.033	1.100	0.275	0.024
0.562	0.0191	2.86	0.055	1.354	0.293	0.024
0.562	0.0103	3.99	0.041	1.600	0.293	0.024
0.569	0.0191	2.86	0.055	1.354	0.293	0.018
0.571	0.0217	1.89	0.041	1.100	0.275	0.024
0.576	0.0147	2.25	0.033	1.200	0.275	0.024
0.577	0.0073	4.51	0.033	1.700	0.293	0.024
0.585	0.0275	1.89	0.052	1.100	0.275	0.024
0.590	0.0183	2.25	0.041	1.200	0.275	0.024
0.598	0.0148	3.51	0.052	1.500	0.293	0.024
0.601	0.0091	4.51	0.041	1.700	0.293	0.024
0.607	0.0130	3.99	0.052	1.600	0.293	0.024
0.607	0.0125	2.64	0.033	1.300	0.275	0.024
0.614	0.0156	2.64	0.041	1.300	0.275	0.024
0.614	0.0231	2.25	0.052	1.200	0.275	0.024
0.620	0.0059	5.43	0.032	1.866	0.293	0.024
0.624	0.0108	3.06	0.033	1.400	0.275	0.024
0.627	0.0134	3.06	0.041	1.400	0.275	0.024
0.629	0.0068	5.43	0.037	1.866	0.293	0.024
0.629	0.0055	5.09	0.028	1.807	0.293	0.018
0.632	0.0197	2.64	0.052	1.300	0.275	0.024
0.633	0.0094	3.51	0.033	1.500	0.275	0.024
0.634	0.0170	3.06	0.052	1.400	0.275	0.024
0.637	0.0068	5.43	0.037	1.866	0.293	0.018
0.639	0.0117	3.51	0.041	1.500	0.275	0.024
0.649	0.0083	3.99	0.033	1.600	0.275	0.024
0.650	0.0103	3.99	0.041	1.600	0.275	0.024
0.651	0.0148	3.51	0.052	1.500	0.275	0.024
0.652	0.0130	3.99	0.052	1.600	0.275	0.024
0.661	0.0091	4.51	0.041	1.700	0.275	0.024
0.661	0.0073	4.51	0.033	1.700	0.275	0.024
Test set						
0.329	0.0261	1.26	0.033	0.900	0.293	0.024
0.411	0.0412	1.26	0.052	0.900	0.293	0.024
0.448	0.0125	2.64	0.033	1.300	0.293	0.024
0.462	0.0131	2.54	0.033	1.276	0.293	0.024
0.515	0.0094	3.51	0.033	1.500	0.293	0.024
0.522	0.0263	1.56	0.041	1.000	0.275	0.024
0.601	0.0055	5.09	0.028	1.807	0.293	0.024
0.635	0.0059	5.43	0.032	1.866	0.293	0.024
0.640	0.0115	4.51	0.052	1.700	0.293	0.024
0.667	0.0115	4.51	0.052	1.700	0.275	0.024

number (Miche 1951) or by wave steepness s_0. In this case, however, as the slope of rubble mound within the upper chamber (see Figure 2) is a constant, the wave steepness could be representative of the wave reflection phenomenon, as it takes into account only the properties of the waves, together with the wave length L_0.

Figures 5–8 show the behaviour of the experimental values of C_r with respect to the observed hydraulic variables, including the ±10% confidence bands, showing that C_r increases with the wave length and wave period, and decreases with wave steepness in a less than linear way, while for wave height values the relationship is not clearly discernible (see Figure 8). For the sake of completeness, note that the captions of Figures 5–8 report also the coefficient of correlation (R) between C_r and the relevant variable in the figure.

Although the dependencies between C_r and L_0 and s_0 are almost clear, the aim of the present study is to provide a formulation for the wave reflection coefficient that is representative of the phenomenon possibly considering both the structural and hydraulic key variables involved.

This study takes the approach of Matteotti (1991) as starting point for further investigation, because it studied the wave reflection for a similar quay, but considering only monochromatic waves. Therefore, Matteotti conclusions could give good indications about the main variables affecting the response of such structure. In particular, that work shows the dependence of the reflection coefficient on the wave period, the wave steepness H/L and the dimensionless chamber width l/L, thus C_r can be expressed by the functional expression:

$$C_r = f\left(T, \frac{H}{L}, \frac{l}{L}\right) \tag{9}$$

where T is the mean wave period, H is the wave height, l is the inner length of the chamber and L is the wave length.

Bearing in mind that wave energy dissipation is mainly due to the rubble mound in the upper chamber, the functional expression in Equation (9) could be integrated by referring to formulations for sloping structure, such as those of Ahrens & Seelig (1980), Hughes & Fowler (1995) and Davidson *et al.* (1996). In these studies, the variability

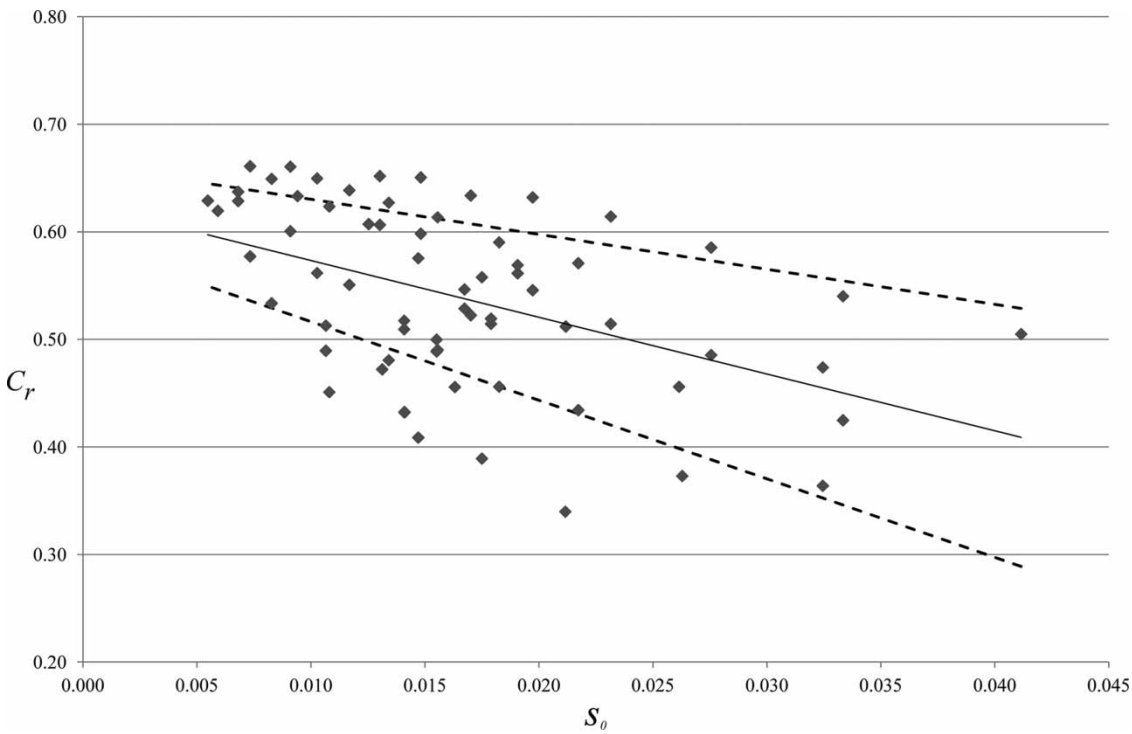

Figure 5 | Measured C_r vs wave steepness s_0 ($R = -0.472$).

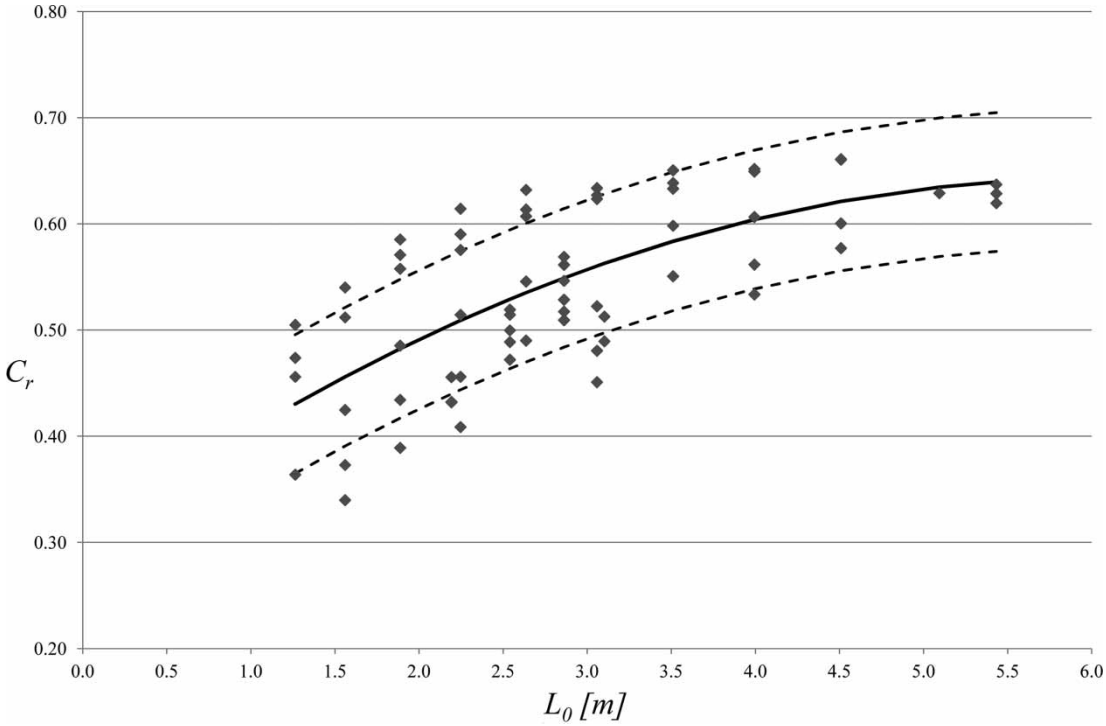

Figure 6 | Measured C_r versus deepwater wave length L_0 ($R = 0.676$).

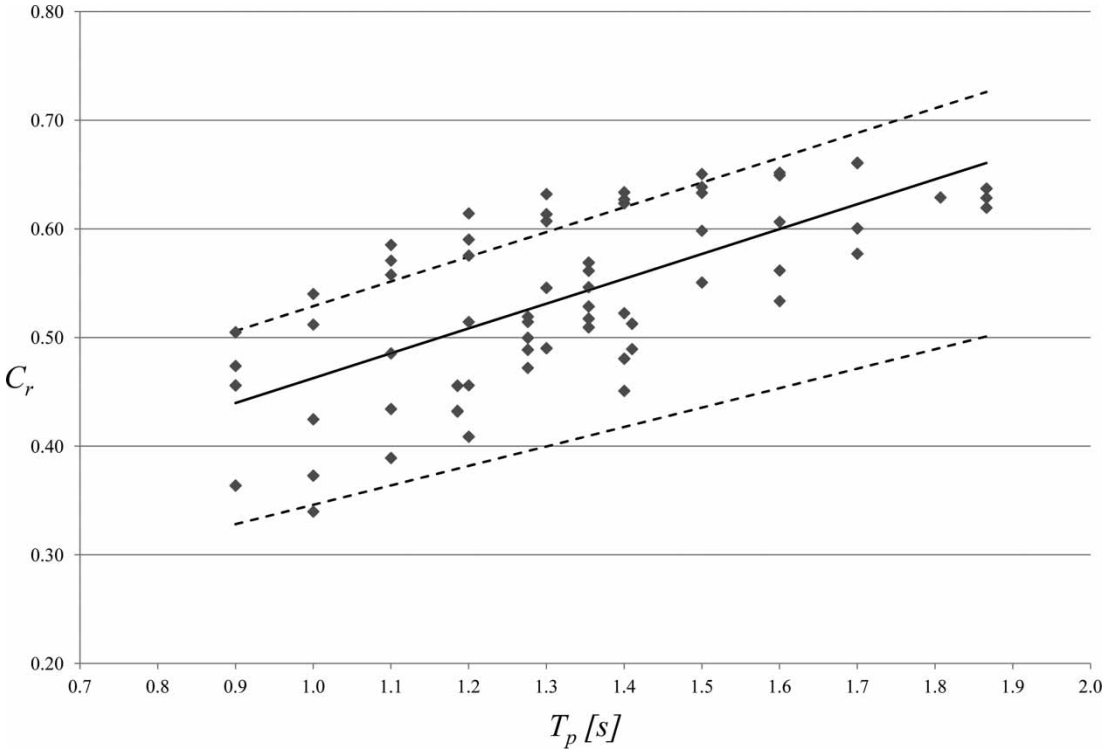

Figure 7 | Measured C_r versus wave peak period T_p ($R = 0.687$).

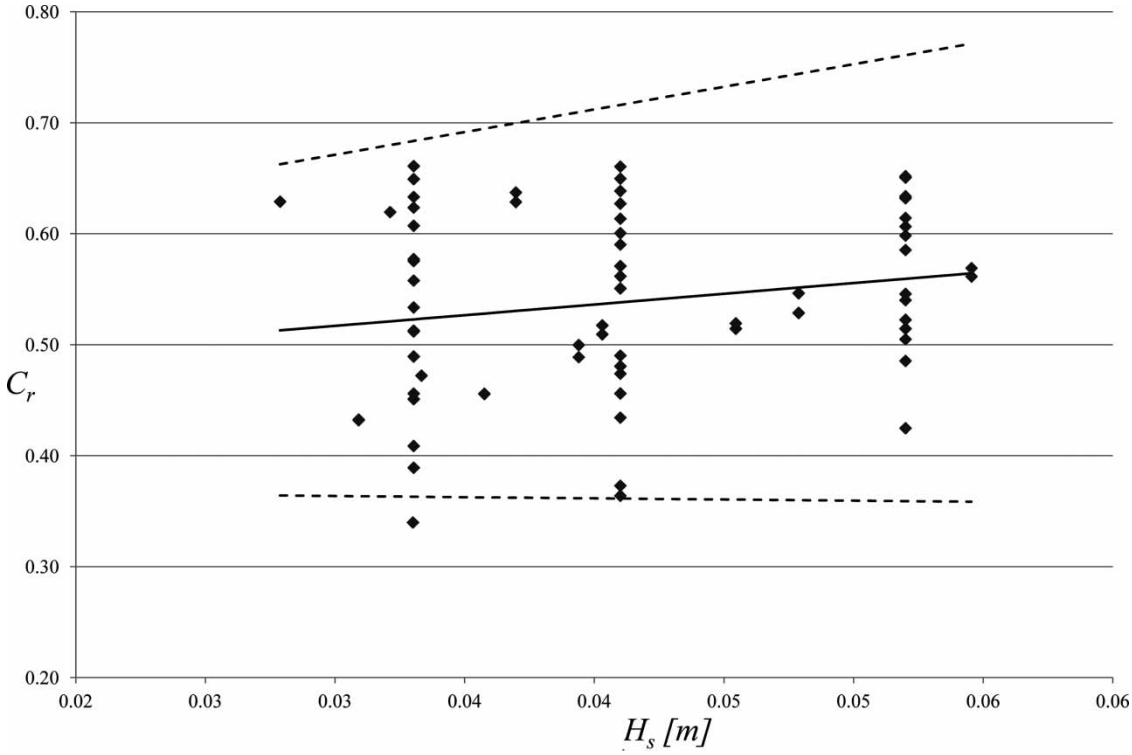

Figure 8 | Measured C_r versus wave height H_s ($R = 0.182$).

of the following variables is studied: water depth d, wave peak period T_p, significant wave height H_s, nominal diameter of the armour units D_{50}; the elevation of the chamber base respect to its bottom h_c (see Figure 2).

Therefore, starting from this research, other dimensionless variables can be identified, and Equation (9) can be written as follows:

$$C_r = f(I, R^*, A, s_0) \tag{10}$$

where $I = \sqrt{\dfrac{d}{g \cdot T_p^2}}$; $R^* = \dfrac{d \cdot L_0^2}{H_s \cdot (0.5 D_{50})^2}$; $A = \dfrac{H_s}{d - h_c}$.

The dimensionless group I is the inverted Iribarren number introduced by Hughes & Fowler (1995), see Equation (6), neglecting $\tan\theta$ (in this case equal to 0.58), as the slope angle of rubble mound is constant in this case, and its influence on wave reflection cannot be investigated. Several authors, as for example Chen *et al.* (2006), have shown that the reflection coefficient could be correlated with the dimensionless group I. Similarly, the dimensionless group R^* descends from the expression of the Iribarren number as given by Davidson *et al.* (1996), see Equation (7), again neglecting $\tan\theta$. The dimensionless group A, finally, is representative of the interaction between the significant wave height and the height of the upper chamber that is submerged under the water level.

EPR MODELLING RESULTS

Starting from the functional expression in Equation (10), which is physically sound with the previous studies on similar quays, the modelling strategy adopted herein simply assumes the set of explanatory variables in Equation (10), to be considered for modelling construction, and the main features of the EPR pseudo-polynomial structure, thus leaving the burden of the combinatorial search to the EPR-MOGA multi-objective optimization paradigm.

From a modelling standpoint, EPR-MOGA models were selected to have the type of pseudo-polynomial structure in Equation (8) (Giustolisi *et al.* 2007). In the specific case, the maximum assumed number of terms $m = 3$, the range

of exponents from which EPR will define the matrix **ES**(j,z) is [–2–2] with an increment interval of 0.1, in order to find a trade-off between data fitting and generalization capacity of the model, and a_0 was neglected, because the reflection is equal to 0 with no waves.

The train set used by EPR-MOGA during the evolutionary model search was composed of the first 68 data points in Table 3, while the test set, used for the evaluation of the returned models at the end of the run, and not used for model construction, is composed by the last 10 data points in Table 3.

The multi-objective optimization paradigm embedded in EPR-MOGA is named OPTIMOGA. Each individual (C_r formulation) is coded as a string of indices related to the model structure (i.e. the candidate exponents). OPTIMOGA employs a dynamical archive growing in size generation by generation. The archive is useful for storing the older non-dominated solutions, which were involved more times in the reproduction. The genetic operators used are multi-point crossover with a probability rate of 40% and mutation with a probability rate of 10%. The probability rates of crossover and mutation are not generally changed in OPTIMOGA (Laucelli & Giustolisi 2011). The three objective functions used herein are: (1) maximization of model accuracy; (2) minimization of the number of model coefficients; and (3) minimization of the number of model inputs.

At the end of the run, EPR-MOGA returns a set of 'optimal' model formulations (i.e. the Pareto front of models), which are the best in fitting training data (C_r, in this case) at different levels of parsimony (i.e. number of coefficients and dimensionless groups). Therefore, the selection of inputs and the definition of relationships between inputs and outputs can be accomplished from different points of view, as for example the fitting to the test data, evaluated by means of the coefficient of determination (CoD) as used by Giustolisi & Savic (2006) for EPR:

$$\mathrm{CoD} = 1 - \frac{\sum_N (\hat{y} - y_{\exp})^2}{\sum_N (y_{\exp} - \mathrm{avg}(y_{\exp}))^2} \tag{11}$$

where N is the number of data; avg($y_{\exp}$) is the average value of observations; $\hat{y}$ is the value predicted by the model and $y_{\exp}$ is the corresponding observation. Note that the CoD ranges within [0; 1], whereas models with high accuracy have a CoD close to 1, and that OPTIMOGA uses the sum of squared errors (SSE) during the evolutionary search of models in order to evaluate their fitness to data (Giustolisi & Savic 2009).

Other possible points of view to select the preferred model of C_r are the parsimony of the model and the presence of physically relevant groups of inputs, according to the physical insight of the user, being this last point as EPR-MOGA can return symbolic expressions, in spite of its data-driven nature.

Table 4 reports the 10 expressions of reflection coefficient as returned by the EPR-MOGA search procedure,

Table 4 | Model expressions returned by EPR-MOGA

				Training		Test	
Equation	**Model expression**	**No. of inputs**	**No. of terms**	**CoD**	**SSE**	**CoD**	**SSE**
(12)	$C_r = 0.157721\ I^{-0.6}$	1	1	0.53	3.03×10^{-3}	0.66	3.11×10^{-3}
(13)	$C_r = 0.097472\ (R^*/s_0)^{0.1}\ A^{0.4}$	3	1	0.90	6.23×10^{-4}	0.86	1.31×10^{-3}
(14)	$C_r = 0.024098\ (R^*/s_0)^{0.3}\ I\ A^{0.7}$	4	1	0.75	1.66×10^{-3}	0.84	1.49×10^{-3}
(15)	$C_r = 0.0188591\ I^{-1.3} + 0.2601\ A^{0.6}$	2	2	0.89	7.11×10^{-4}	0.90	9.32×10^{-4}
(16)	$C_r = 0.1986\ s_0^{-0.2}\ A^{0.5} + 0.0001124\ R^{*0.5}$	3	2	0.92	5.42×10^{-4}	0.90	9.32×10^{-4}
(17)	$C_r = 0.0010416\ s_0^{-1} + 0.12709\ R^{*0.1}\ A^{0.5}$	3	2	0.92	5.38×10^{-4}	0.90	9.11×10^{-4}
(18)	$C_r = 0.1121\ (R^*/s_0)^{0.1}\ I^{0.1}\ A^{0.5} + 1.7064\text{e-}005(R^*/s_0)^{0.3}\ I^{-1.1}\ A^{-0.4}$	4	2	0.91	5.46×10^{-4}	0.89	9.99×10^{-4}
(19)	$C_r = 0.1949\ s_0^{-0.2}\ A^{0.5} + 0.0002485\ R^{*0.1}\ I^{-0.3}$	4	2	0.92	5.35×10^{-4}	0.90	9.84×10^{-4}
(20)	$C_r = 0.19214\ s_0^{-0.2}\ A^{0.5} + 0.0003345\ R^{*0.4}\ I^{-0.4}\ s_0^{0.1}$	4	2	0.92	5.39×10^{-4}	0.90	9.92×10^{-4}
(21)	$C_r = 0.19873\ s_0^{-0.2}\ A^{0.5} + 0.00015258\ R^{*0.4}\ I^{-1.3}\ s_0^{0.4}\ A^{0.3}$	5	2	0.92	5.20×10^{-4}	0.91	9.16×10^{-4}

coupled with the values of the three objective functions used for the evolutionary search: the number of selected dimensionless variables (third column); the number of model constants (fourth column); the accuracy measures, CoD and SSE, on the training set (fifth and sixth columns); and the CoD and SSE evaluated on the test set (last columns).

It is worthy of note that the Pareto front has been obtained by the accuracy measures as evaluated on the training set, thus not considering the accuracy measures on the test set, which can be considered as measures of generalization ability of the candidate solutions (Table 4).

During the exploration of the solution space, the search of model coefficients a_j has been constrained to only positive values ($a_j > 0$), as described in Giustolisi *et al.* (2007). This is useful for two general reasons: (i) positive constants are generally consistent with the physical meaning of the models; and (ii) the alternation of negative and positive constants, in models developed from data, is sometimes useful to fit errors and generates over-fitting to data (i.e. poor generalization performance).

DISCUSSION

The aim of this section is to discuss EPR-MOGA results from the modelling point of view, emphasizing its main features (for example, the ability to provide symbolic expressions, ranking of formulae according to their accuracy versus parsimony, etc.). As evidenced by Table 4, the optimal solutions range from a very parsimonious model, the model in Equation (12), up to Equation (21) that conversely contains the higher number of explanatory variables. Equation (12) is the simplest one, but presents a very low accuracy with respect to the others.

From models obtained, as listed in Table 4, it is firstly possible to have an idea of the most used explanatory variables (i.e. dimensionless groups) for the phenomenon at stake. Therefore, it is possible to discuss the dimensionless group's importance:

- Nine equations (90% of the total) show $A = H_s/(d\text{–}h_c)$ with a positive exponent (i.e. direct dependence), except from Equation (19) where it appears also with negative exponent (at a second term).
- Eight equations (80%) show $R^* = dL_0^2/H_s(0.5D_{50})^2$ with a positive exponent (i.e. direct dependence).
- Eight equations (80%) show $s_0 = H_s/L_0$ with a negative exponent (i.e. inverse dependence), with the exception of Equations (20) and (21) where it appears with a positive exponent (at a second term).
- Finally, $I = (d/gT_p^2)^{0.5}$ appears in seven equations (70%) with exponents both positive and negative.

Equation (12) is the simplest model, with only one constant and one dimensionless group, but it presents a very low accuracy. Equations (13) and (14) present only one constant, but with a combination of more than one variable. Both Equations (13) and (14) and Equation (18) show a new group R^*/s_0. Equation (13) is a simple model but accurate, differently from Equation (14) where the presence of I gives an inverse proportionality with the wave period, which is not physically consistent, resulting in a lower CoD than Equation (13).

Models with more than one constant lead to a better accuracy than the simplest equations: generally one term is a function of the nominal diameter of the armour units D_{50}, contained in R^*, and the other term depends on the wave steepness and water depth.

A is present in several parsimonious models, showing an influence of the quantity ($d\text{–}h_c$) that represents the wet part of the chamber, thus responsible of wave energy dissipation due to the eddies among the armour units.

The exponents of R^* are always positive: this result is physically sound because it implies a direct dependence of C_r on the wave period and an inverse dependence on the nominal diameter. In fact a greater D_{50} means higher porosity, and therefore higher dissipations leading to reduction in wave reflection.

The wide variation of the exponent for I suggests that this parameter is not significant, thus excluding models where it appears (as showed by the low accuracy of the Equation (12) that is totally supported by I). This fact suggests that the influence of water depth itself, present in I, is not significant as well as the quantity ($d\text{–}h_c$) that substantially represents the influence of the chamber with respect to the vertical part of the quay wall.

The exponents of s_0 are generally negative showing a physical direct proportionality with wave period; the

Equations (20) and (21) show positive exponents for s_0, which is combined with other variables into a secondary term. Finally, this combination leads to the same dependence on the wave period.

The use of the selected non-dimensional groups shows different weighting for the variables than in classical analysis, if the hydraulic and geometrical quantities are exploded; for example the weighting of the wave length with respect to the wave height is different from for the expression of the Iribarren number.

It is necessary to try to make a choice among models in Table 4, thus mimicking what could happen in real life. When a researcher/operator is asked to define a good model to use in further applications, the selected model is Equation (13) as it represents a simple and accurate model, dependent on the main structural and hydraulic parameters.

Equation (13) written with the dimensional variables stays in the following form:

$$C_r = 0.11196 \frac{d^{0.1} L_0^{0.3} H_s^{0.2}}{D_{50}^{0.2} (d - h_c)^{0.4}} \quad (22)$$

From a physical point of view, C_r has to show a maximum asymptotic value of one for totally reflective structures (such as vertical walls): in contrast Equation (22) shows a rising shape with increasing values for wavelength and wave significant height, passing over one as maximum value. Therefore, the authors' proposition is to use Equation (22) only within its test range.

The presence of D_{50}, d and h_c confirms the conclusions of Davidson *et al.* (1996): the only use of classical non-dimensional parameters, such as the Iribarren number, leads to poor prediction of the reflection coefficient. Accordingly, the application of a more comprehensive modelling technique, such as EPR-MOGA, has lead us to obtain a formulation that best fits and interprets the data and, at the same time, resumes the complexity of the analysed phenomena in a simple and affordable model. Additionally, the selected model is fully consistent with the relationships shown in Figures 5–8: in fact even though T_p is not explicitly present in Equation (22), L_0 depends on the wave peak period, thus Equation (22) can be expressed as a function of either T_p or L_0.

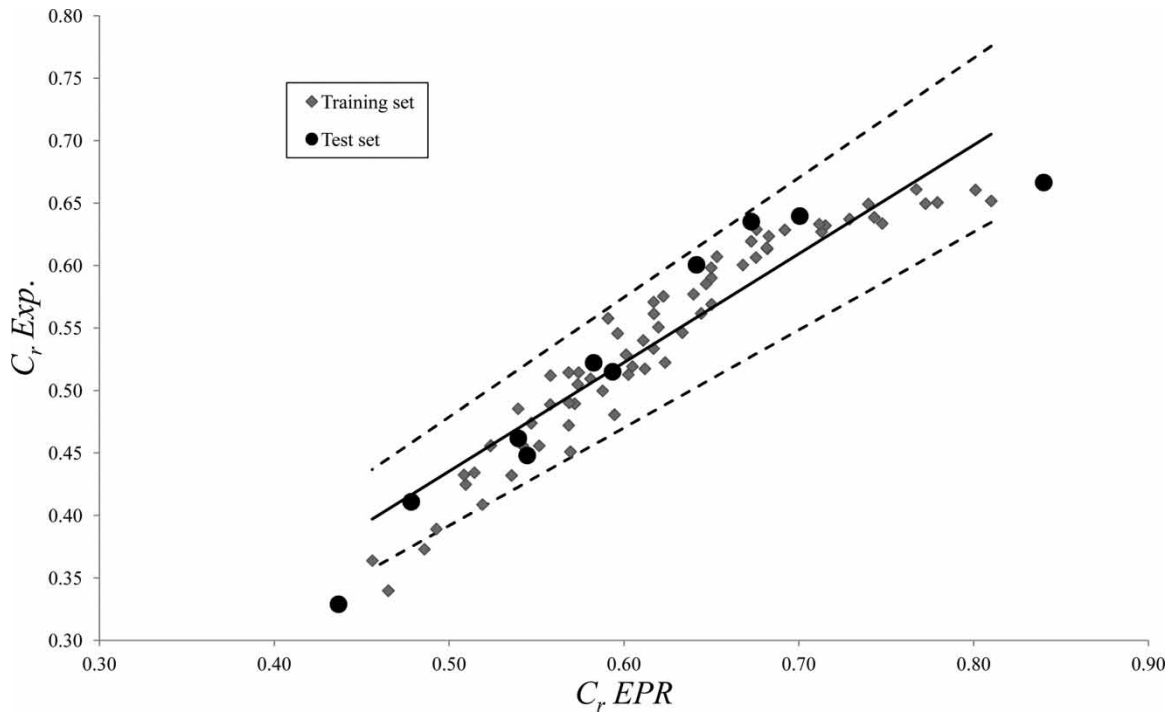

Figure 9 | Experimental data versus EPR-MOGA model prediction, Equation (22) (with dimensional inputs) or Equation (13) (with dimensionless inputs).

Figure 9 plots the experimental values of C_r versus the EPR-estimated values (including the ±10% confidence bands): the model fits very well the experimental data in the range 0.40–0.70, which can be reasonably considered as a possible range for the application of the selected model for the evaluation of the average C_r for this kind of quay structure.

CONCLUSIONS

The paper proposed an interesting application of an evolutionarily-based data-driven technique (the EPR-MOGA) to estimate wave reflection of dissipative quays, consisting in a vertical wall with an inner chamber, filled with rubble mound and exposed to irregular wave action. The characterization of the wave reflection of a quay is a very important issue in harbour design, because a proper evaluation of such a phenomenon can avoid structural and operational problems in port basins. Furthermore, the low environmental impact justifies the choice of this kind of solution: therefore it is necessary to know and describe in detail the structural response for a suitable design.

The application of EPR-MOGA, instead of other data-driven or numerical regression techniques, is justified due to the complexity of the phenomena and the high number of variables that could describe the process. This is because EPR-MOGA can perform a population-based multi-objective optimization in the space of solutions, returning a set of optimal solution as trade-offs between accuracy and parsimony of models.

In the proposed case study, an experimental set of data was measured on a laboratory-scale model (1:33) of the proposed low reflective quay, including all the most significant variables featuring the wave reflection. The application of EPR-MOGA as a modelling paradigm returns a set of solutions all showing physical consistency with the literature. In particular, the authors proposition is to use the model of Equation (13), which is a compact formula dependent on the main involved variables (i.e. water depth d, deep-water wave length L_0, significant wave height H_s, and nominal diameter of armour units D_{50}). However, the authors' advice is to employ this formulation inside the ranges for which it has been developed, considering the main dimensional variables summarized in Table 1 and carefully out of those ranges.

REFERENCES

Ahrens, J. P. & Seelig, W. N. 1980 *Wave Run-up on a Rip-rap Protected Dike*. US Army Corps of Engineers, Coastal Engineering Research Centre, Fort Belvoir, VA.

Allsop, N. W. H. & Hettiarachchi, S. S. L. 1988 Reflections from coastal structures. In: *Proceedings of the 21st International Conference on Coastal Engineering*, ASCE, Malaga, Spain, pp. 782–794.

Andersen, T. L. & Liu, Z. 2006 *WaveLab 2.921 Hydraulics and Coastal Engineering*. Aalborg University, Aalborg, Denmark, 2002–2006 (//hydrosoft.civil.aau.dk/wavelab).

Battjes, J. A. 1974 Surf similarity. In: *Proceedings of the Fourteenth International Conference on Coastal Engineering ASCE I*, Copenhagen, Denmark, pp. 466–480.

Berardi, L., Kapelan, Z., Giustolisi, O. & Savic, D. 2008 Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics* **10**, 113–126.

Bhattacharya, B., van Kessel, T. & Solomatine, D. P. 2012 Spatio-temporal prediction of suspended sediment concentration in the coastal zone using an artificial neural network and a numerical model. *Journal of Hydroinformatics* **14**, 574–584.

Burcharth, H. F., Liu, Z. & Troch, P. 1999 Scaling of core material in rubble mound breakwater model tests. In: *Proceedings of the Fifth International Conference on Coastal and Port Engineering in Developing Countries*, Cape Town, South Africa, pp. 1518–1528.

Chen, H., Tsai, C. & Chiu, J. 2006 Wave reflection from vertical breakwater with porous structure. *Ocean Engineering* **33**, 1705–1717.

Davidson, M. A., Bird, P. A. D., Bullock, G. N. & Huntley, D. A. 1996 A new non-dimensional number for the analysis of wave reflection from rubble mound breakwaters. *Coastal Engineering* **28**, 93–120.

Doglioni, A., Mancarella, D., Simeone, V. & Giustolisi, O. 2010 Inferring groundwater system dynamics from hydrological time-series data. *Hydrology Science Journal* **55**, 593–608.

EPR-MOGA 2010 A MS-Excel based tool for advanced data-driven modelling (www.hydroinformatics.it).

Fugazza, M. & Natale, L. 1992 Hydraulic design of perforated breakwaters. *Journal of Waterways, Port, Coastal and Ocean Engineering* **118**, 1–14.

Garrido, J., Ponce de León, D., Berruguete, A., Martínez, S., Manuel, J., Fort, L., Yagüe, D., González-Escrivá, J. & Medina, J. 2010 Study of reflection of new low-reflectivity quay wall caisson. In: *Proceedings of the 32nd International*

Conference on Coastal Engineering, Shanghai, China **1** (32), structures. 27.

Giustolisi, O. & Savic, D. A. 2006 A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics* **8**, 207–222.

Giustolisi, O. & Savic, D. A. 2009 Advances in data-driven analyses and modelling using EPR-MOGA. *Journal of Hydroinformatics* **11**, 225–236.

Giustolisi, O., Doglioni, A., Savic, D. A. & Webb, B. 2007 A multi-model approach to analysis of environmental phenomena. *Environmental Modeling Software* **22**, 674–682.

Goda, Y. & Suzuki, Y. 1976 Estimation of incident and reflected waves in random wave experiments. In: *Proceedings of the Fifteenth Coastal Engineering International Conference I*, Honolulu, Hawaii, pp. 828–845.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, London, UK.

Hughes, S. A. & Fowler, J. E. 1995 Estimating wave-induced kinematics at sloping structures. *Journal of Waterways, Port, Coastal and Ocean Engineering* **121**, 209–215.

Ijima, T., Tanaka, E. & Okuzono, H. 1976 Permeable seawall with reservoir and the use of warock. In: *Proceedings of the Fifteenth Coastal Engineering International Conference III*, ASCE, Honolulu, Hawaii, pp. 2623–2641.

Jarlan, G. E. 1961 A perforated vertical wall breakwater. *The Dock and Harbour Authority* **41**, 394–398.

Jarlan, G. E. 1965 *The Application of Acoustic Theory to the Reflective Properties of Coastal Engineering Structures Quart Bulletin*. National Research Council Canada, pp. 23–64.

Javadi, A. A. & Rezania, M. 2009 Application of artificial intelligence and data mining techniques in soil modelling. *Geomechanical and Engineering International Journal* **1**, 53–74.

Laucelli, D. & Giustolisi, O. 2011 Scour depth modelling by multi-objective evolutionary paradigm. *Journal of Environmental Modelling and Software* **26**, 498–509.

Mansard, E. & Funke, E. 1980 The measurement of incident and reflected spectra using a least squares method. In: *Proceedings of the Seventeenth Coastal Engineering International Conference*, ASCE, Sidney I, pp. 154–172.

Matteotti, G. 1991 The reflection coefficient of a wave dissipating quay wall. *Dock and Harbour Authority* **71**, 285–291.

Miche, M. 1951 Le pouvoir reflêchissant des ouvrages maritimes exposés à l'action de la houle. *Annals Ponts Chaussées* **121**, 285–319. [In French]

Pareto, V. 1896 *Cours D'Economie Politique* (volumes I and II). Rouge & Cic, Lausanne, Switzerland.

Suh, K., Park, J. K. & Park, W. S. 2006 Wave reflection from partially perforated wall caisson breakwaters. *Journal of Ocean Engineering* **33**, 264–280.

Tanimoto, K. & Takahashi, S. 1994 Design and construction of caisson breakwaters – the Japanese experience. *Coastal Engineering* **22**, 57–77.

Taveira Pinto, F. A., Rosa Santos, P. J., Veloso Gomes, F. F. M. & Guedes Lopes, H. 2011 Efficiency analysis to reflection of a new quay wall type. *Journal of Hydraulic Research* **49**, 539–546.

Zamani, A., Azimian, A., Heemink, A. & Solomatine, D. 2009 Wave height prediction at the Caspian Sea using a data-driven model and ensemble-based data assimilation methods. *Journal of Hydroinformatics* **11**, 154–164.

Zanganeh, M., Yeganeh-Bakhtiary, A. & Bakhtyar, R. 2011 Combined particle swarm optimization and fuzzy inference system model for estimation of current-induced scour beneath marine pipelines. *Journal of Hydroinformatics* **13**, 558–573.

Zhang, Z., Chi-Wai, L., Yok-Sheung, L. & Yiquan, Q. 2006 Incorporation of artificial neural networks and data assimilation techniques into a third-generation wind–wave model for wave forecasting. *Journal of Hydroinformatics* **7**, 65–76.

First received 18 July 2012; accepted in revised form 15 October 2012. Available online 27 December 2012