Journal of Coastal Research	SI 64	95 - 99	ICS2011 (Proceedings)	Poland	ISSN 0749-0208

Application of a Statistical Method to investigate patterns of beach evolution in the vicinity of a seawall

J. M. Horrillo-Caraballo† and D.E. Reeve†
†School of Marine Science and Engineering,
University of Plymouth, Reynolds Building
Plymouth PL4 8AA, UK
Jose.Horrillo-Caraballo@plymouth.ac.uk
Dominic. Reeve@plymouth.ac.uk

ABSTRACT

Horrillo-Caraballo, J.M. and Reeve, D.E., 2011. Application of a statistical method to investigate patterns of beach evolution in the vicinity of a seawall. Journal of Coastal Research, SI 64 (Proceedings of the 11th International Coastal Symposium), 95 – 99. Szczecin, Poland, ISSN 0749-0208

In this paper we describe the application of Canonical Correlation Analysis (CCA) to an historical dataset of seabed elevations within a coastal segment of the English south eastern coast. The study site is located along the frontage of Walcott (Norfolk, East coast of England). The beaches along this segment of coast are mainly composite, mixed sand and shingle. The dataset comprises detailed bathymetric surveys of beach profiles covering a period of over 17 years (~ 2 beach profile surveys every year). The structure of the dataset and the data handling methods are described. The application of the CCA method is discussed as well as the ability of the CCA to provide useful forecast of the beach profile at Walcott. Some results and interpretation of the analysis are presented. Beach profiles predictions agreed well with the measured profiles for up to five years of forecast at the site. The level of error using CCA is proportionate with that found with dynamical modelling. It is expected that this relationship will be used by coastal planners to predict profile evolution based on waves and storm surges.

ADDITIONAL INDEX WORDS: *Beach profile, Canonical Correlation Analysis, Empirical Orthogonal Functions, Forecast, Statistical models.*

INTRODUCTION

In recent years, as a result of the implementation of a strategic approach to flood and coastal management, the necessity of more robust methodologies for incorporating risk assessments within coastal engineering design has been clearly identified (Cooper and Hutchinson, 2002). Climate change will modify wave and water level conditions and for this reason affect the vulnerability of coastal defence structures (Sutherland and Gouldby, 2003). There are a large number of coastal structures whose stability depends upon the characteristics of the beach where they are located. From this perspective, it is necessary to understand how seawalls and beaches interact during storms and how beach profiles interact with waves, tides and storms. Without this, it is difficult to provide realistic estimates of interaction between the structure and the beach profiles and how they are likely to change over time. Sea walls are often part of a larger scheme, operating in combination with beach and other structures (e.g. beach nourishment, groynes, etc.) to control wave energy, thereby improving the resistance to coastal erosion or limiting wave overtopping. Therefore, coastal engineering has dedicated significant efforts to study the interaction between wave conditions and morphological changes on the beaches. This phenomenon is of great complexity because the morphology of a beach at a certain time is the result of several hydrodynamic processes that act over a vast range of space and temporal scales (Larson and Kraus, 1995).

In order to improve predictions of likely wave overtopping and thus flood risk, improvements in both wave forecasting and beach movement are required. Wave forecasting models have seen significant improvements over the last few decades, but predicting beach levels remains difficult. The importance of beaches for coastal flood risk is due to the process of wave breaking in shallow water. Breaking dissipates much of the energy in a wave and can materially reduce overtopping. Beach prediction models come in essentially two forms: a) formulae that describe an equilibrium shape as a result of fixed wave conditions (e.g. Dean *et al.*, 1993); b) detailed process models that can provide good results over the period of a storm but have difficulties in terms of accuracy and stability for longer term predictions.

As a consequence of the limitations mentioned for both types of models, recent studies have highlighted the necessity of use of robust statistical models (Różyński, 2003; Li *et al.*, 2005). This type of approach uses correlation techniques of series of data to identify relationships and structures of two simultaneous series of time. The objective of these techniques is to associate the changes contained in the data sets to physical processes in the area of study. In essence, this relies on a statistical analysis of historical data and a process of extrapolation. Given the growing amount of beach survey data and nearshore wave forecasts that are now almost routinely available, in this paper we examine the use of data-driven technique for predicting beach changes. Canonical Correlation Analysis (CCA) is one of these methods that has been used in areas such as meteorology and climatology (Glahn, 1968; Graham *et al.*, 1987).

This method has been used with data from the FRF (Field Research Facility in North Carolina, USA) by Larson *et al.* (2000)

in order to detect occurring patterns in the wave and profile data and to investigate the use of CCA to predict the profile response due to waves. Horrillo-Caraballo and Reeve (2008) extended this study to investigate the sensitivity of the quality of predictions on the choice of distribution function used to describe the wave heights. In addition, Rozynski (2003) used the CCA method to analyse the evolution patterns of multiple long-shore bars and the interactions among them.

In this paper, CCA techniques are applied to a series of historical beach profile survey of the Walcott area (east coast of England) covering a period of over 17 years. The data were analysed in order to determine the covariability between waves and profile response.

The coastline in Walcott is orientated approximately northwest - southeast and is prone to attack from the North Sea, especially when the wind blows from the northeast (Figure 1). The danger increase when this wind is combined with a North Sea surge, which can raise the still water level of the sea by up to 2 m above the predicted sea level. These conditions have caused disastrous flooding in the town, the worst in recent memory being in 1953. (HR Wallingford, 2002).

Figure 1. Location of the study area and photo from the seawall in Walcott (top right) and wave data location (WDP1).

Wave hindcasts and beach profiles from the Environment Agency (EA) were used to investigate the relationship between the waves and the profile response over a longer time period. The data covered beach profile surveys taken along Profile N3C8 (Figure 1) over a period of 17 years and spectral wave properties (significant wave height, peak period and wave direction) calculated at least every 1h during the same period. Wind measured at Weybourne, UK (Chini, pers. comm.) were used to estimate wave parameters using the so-called CERC formulas (USACE, 1984) The wave data point (WDP1), where wave parameters were calculated, is located at position 52°51'0.25"N, 1°30'59.40"E (Figure 1).The waves were hindcasted at 12mCD (Chart Datum) water depth and these data were used to derive the wave properties employed in the CCA.

Initially, the analysis focused on directly linking the profile response to the significant wave heights (H_s) and to wave steepness (H/L). In order to represent the significant wave heights and wave steepness, probability density functions (pdf) were derived for the two wave parameters. Secondly, the theory behind the CCA and some description of how it was applied in this study is given. The results of the CCA between the profiles and the two wave parameters are presented. A discussion and conclusive remarks are provided on the potential for using these results to predict profile evolution based on the waves.

CANONICAL CORRELATION ANALYSIS (CCA) METHOD

CCA was proposed by Hotelling (1936) and can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximised. CCA is used to investigate the intercorrelation between two sets of variables, whereas factor/principal component analysis or EOF (Empirical Orthogonal Functions) identifies the pattern of relationship within one set of data (Clark, 1975). CCA may be used to investigate the presence of any patterns that tend to occur simultaneously in two different data sets and what the correlation is between associated patterns (Graham *et al.*, 1987). If the two original data sets are denoted Y (wave parameter data matrix with size $nt \bullet ny$) and Z (profile data matrix with size $nt \bullet nz$), we denote new transformed variables U and V that have maximally correlated column vectors for the same index and zero correlation for differing indices. U and V are orthonormal.

The desired weights to transform Y into U (Graham *et al.*, 1987),

$$\left[\left(Y^T Y\right)^{-1}\left(Y^T Z\right)\left(Z^T Z\right)^{-1}\left(Z^T Y\right)-\mu^2 I\right]=0 \qquad (1)$$

where μ^2 denotes the eigenvalues (squared correlation between the corresponding temporal amplitudes of the canonical modes - the columns vectors in U and V - and the associated eigenvectors R), yields the transformation $U = YR$.

The same to transform Z into V,

$$\left[\left(Z^T Z\right)^{-1}\left(Z^T Y\right)\left(Y^T Y\right)^{-1}\left(Y^T Z\right)-\mu^2 I\right]=0 \qquad (2)$$

having the same μ^2 eigenvalues and the associated eigenvectors Q, yields the other transformation $V = ZQ$.

The spatial amplitudes (G and H) of the canonical mode are obtained as:

$$G = Y^T U \qquad (3)$$
$$H = Z^T V \qquad (4)$$

The original data set are expressed as:

$$Y = UG \qquad (5)$$
$$Z = VH \qquad (6)$$

Noise in the data can adversely affect the accuracy of these matrix manipulations. Thus some filtering of the data is often performed prior to the analysis. EOF is commonly used in the filtering. The data set (Y and Z) are expressed in terms of EOFs

$$Y = AE^T \qquad (7)$$
$$Z = BF^T \qquad (8)$$

where A and B contains the temporal EOF (or principal scores) and E and F the spatial EOF for Y and Z respectively. Filtering is performed simply by truncating the expansion of EOFs.

The regression matrix is originated in the CCA analysis and relates the profiles to the wave proper-ties based on the correlation between the dominant patterns in the profile and wave data. This means that the regression matrix can be used to predict the profile response, if the wave properties are known (Larson *et al.*, 2000).

The associated profile data matrix Z_p, having a wave matrix Y_p (measured or simulated) is given by:

$$Z_p = Y_p \psi \quad (9)$$

where

$$\psi = GSF^T \quad (10)$$

$$S = U^T B \quad (11)$$

The necessary matrix algebra required in these cases is now available in FORTRAN® or MATLAB® routines.

SURVEY DATA

Beach Profile Data

The mixed shingle-sand beaches of Walcott (Figure 1) have been monitored since 1991 along 18 shore normal profiles, as part of a long-term beach management survey programme by North Norfolk District Council (HR Wallingford, 2002). The surveys were carried out in the summer and winter months so that the seasonal variations in beach morphology can be examined. In this study, Profile N3C8 has been used for this study, due to that this profile takes into account the seawall. Figure 2 displays the subset of measured profiles along Profile N3C8 included in this study that were used in the CCA (25 profiles shown).

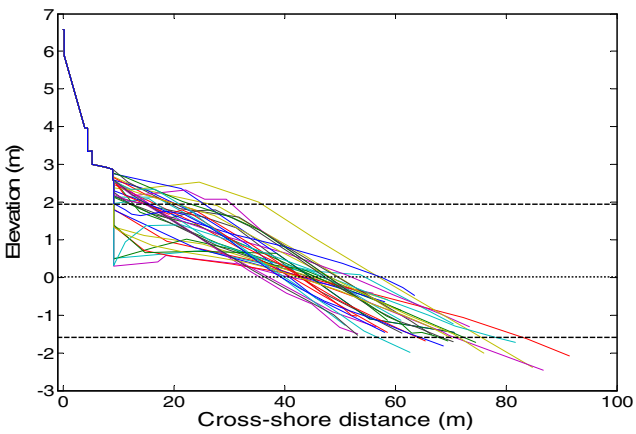

Figure 2. Time series of beach profile N3C8.

Only surveys that extended from the dune region out to a water depth of approximately the Mean Low Water Level (MLWL) were included in the analysis. Measured levels are referred to the UK Reference System - Ordnance Survey Datum Newlyn (ODN).

Wave Data

The wave dataset available for this area consisted of significant wave height (H_{m0}), peak spectral wave period (T_p) and wave direction (θ) obtained from Nicolas Chini (Chini, pers. comm.). The waves were obtained every 1h. The hindcasted wave point is located in approximately 12mCD water depth and 1.3km offshore. The predominant wave directions are from the north, north-east, and east and the largest surges are associated with winds from the north-west and north. The tidal ranges are 3.54m and 1.72m for spring and neap respectively. Wave conditions obtained from the hindcast will be used to calculate the empirical probability density functions that were used in the CCA analysis.

METHODOLOGY

Decomposition of the Data

The data used for this study have been detailed in the section above. The data obtained from Walcott have been manipulated in order to acquire time series with the same sampling rate for profiles and wave conditions. The profile dates were fixed in accordance with the Walcott dataset. In general terms, around 25 beach profiles were used in the CCA and around 10 beach profiles were used for comparison with the predictions on the basis of the regression matrix.

The beach profile is highly dependent on the prevailing wave conditions. Therefore, the wave conditions between two selected profiles were used to characterise the changes in each subsequent beach profile. Nonetheless, there are about 155209 wave data records for Walcott to characterise around 20 beach profiles. At each profile taken for this study there are many more wave observations than beach profile measurements. With the aim of generate two series of equal length, required to apply the CCA method, the wave conditions are compiled in probability density functions (pdfs). A parametric form of pdf was proposed by Larson and Kraus (1995). However, more recent studies (Horrillo-Caraballo and Reeve, 2008; 2010) have suggested that a better performance is obtained using an empirical distribution. The empirical distribution is a cumulative probability distribution function that concentrates probability 1/n at each of the n numbers in a sample. A combined pdf (p_n) may then be derived by superimposing the individual pdfs available for the measurement period between surveys,

$$p_n(H) = \frac{1}{n}\sum_{i=1}^{n} I(H_i \le H) \quad (12)$$

where H is the wave height, n the number of individual wave measurements between surveys and i an index. The superposition carried out in Equation (12) implies that all the individual pdfs derived from the wave measurements (H_s) have the same weight. The composite pdfs used in the CCA analysis were discretised into about 200 intervals, sufficient to provide a good resolution over the probability interval of significance.

Figure 3 shows the corresponding composite pdf for the offshore significant wave height valid for the time period between surveys and obtained by adding over a large number of Empirical pdfs according to Equation (12).

Data for the area of Walcott is quite scarce in comparison with other field data sets (e.g. Duck); even that profiles have been measured from the eighties, there is not enough wave data to correlate waves and profile from this time. Wave data is only available from the early nineteen-nineties and profiles in this site are measured on average twice a year. At Walcott, the higher probabilities are located in the area of smaller waves (Figure 3). The Empirical distribution is considered to be a reliable descriptor

of the wave height distribution in deep waters and in shallow waters (Horrillo-Caraballo and Reeve, 2008; 2010).

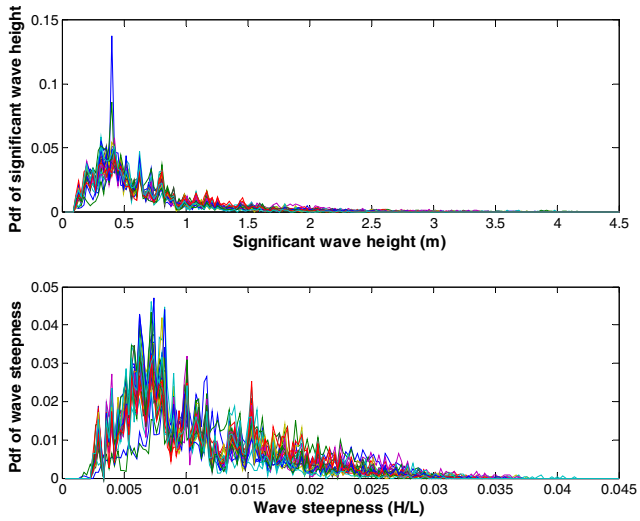

Figure 3. Time series of empirical pdfs of significant wave height (top panel) and time series of empirical pdfs of wave steepness (bottom panel) at Walcott

In this paper, the offshore distributions have been used to correlate the waves and the beach profile using CCA and the results are discussed below.

Analysis of the Data

For the purpose of comparison, the CCA is applied to a profile at Walcott datasets using 13 years of records and then performing predictions 5 years, until 2008.

Two different CCA analysis were performed (one for each of the wave parameters – significant wave height and wave steepness) using beach profiles, significant wave heights and peak period from 12/08/1991 to 01/08/2003 (~ 25 profiles and 25 empirical pdf distributions) to determine the regression matrices and then predictions from 19/01/2004 to 15/09/2008 (~10 profiles) were computed using the hindcasted wave and period data.

The approach used here with the different profiles is to investigate: the variation in profile predictions using different regression matrices (empirical pdfs with significant wave height and wave steepness)

For each case, EOF expansion and truncation were used in order to pre-filter the data sets before performing the CCA analysis (e.g., Clark, 1975). The "rule of thumb" of North *et al.* (1982) was applied to determine the appropriate truncation point in the EOF expansion.

RESULTS

The analyses were performed as described above. The motivation for the tests was to ascertain how the choose of pdfs might affect the quality of predictions. Such information could have direct application in the planning and collection of beach monitoring measurements for coastal management and prediction of the condition (erosion or accretion) of the coastline.

The first three spatial EOFs obtained from the profile data sets at Profile N3C8 explained 97% of the variation in the data(E_1: 76%; E_2:16%; E_3:5%). For the empirical significant wave pdfs, the first three spatial EOFs explained 54% of the variation in the data (E_1: 33%; E_2:12%; E_3:9%) and for the empirical wave steepness distribution, 43% of the variation is explained (E_1: 24%; E_2:10%; E_3:9%). In all cases, the time mean was subtracted before analysis in all data sets. For the two CCA analyses, the EOFs describing the profile shape are quite complex, but the EOF shapes can be used to quantify mean properties of the influence of the berm over the profile (Larson *et al.*, 2000). Moreover, the temporal EOFs can be used to determine trends of profile changes and oscillatory cycles. CCA was then performed at profile N3C8 as described in the section above. The regression matrices obtained were used to forecast the beach profiles using the wave data set for the prediction period. The wave data used goes from 02/08/2003 - 15/09/2008. The wave data was treated using the EOF method and then replaced in Equation 9 with the regression matrix obtained with the CCA analysis. The beach profile is obtained replacing the two in Equation 9.

The RMSE (Root Mean Square Error) obtained between the measured beach profiles and the predicted profiles are used to measure the quality of the predictions. Values of RMSE are shown as a measure of the error between prediction and measurement as a function of cross-shore position, averaged over the forecast window.

Figure 4 shows the time averaged RMSE for the two cases; the CCA was performed using different modes (shown in the figure legend). Overall, the best performances for Profile N3C8 are: the use of two canonical modes if the empirical wave significant pdfs are applied and also two canonical modes if the wave steepness pdfs are used. The general behaviour of the different modes in the two cases are very similar, the RMSE increases where major changes in the elevations are present. This is due largely to the movement and change of position of the berm in the beach profile.

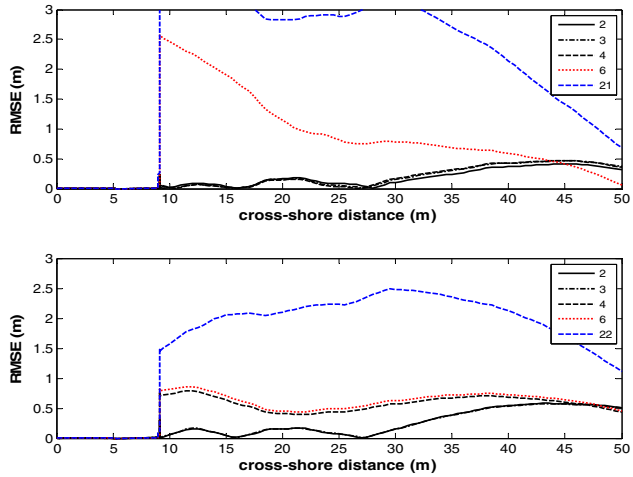

Figure 4. Time averaged RMSE for predictions made with the number of CCA modes indicated by the label key, using empirical wave significant pdfs (top panel) and using empirical wave steepness pdfs (bottom panel).

The major changes are present between 27m and 50m chainage in the profile and the behaviour of the error is very similar using 2 or 3 modes. CCA performed with 2 modes obtained the minimum

RMSE. The error in the profile shows similar behaviour in their cross shore distance, with this it is possible to assume that the wave significant pdf and the wave steepness present similar

performance. In this case, it is possible to use any of the two methods to predict the profiles.

Figure 5 shows the measured beach profile at the profile N3C8 for February 15th 2006 and the prediction made from January 31st 2006 using the two different empirical pdfs distributions. Satisfactory agreement is obtained in the area where the profile shows evidence of considerable change. In the offshore area, the error is over predicted. In contrast, at the beginning of the profile (between 8m and 25m) there is an under prediction of the beach levels.

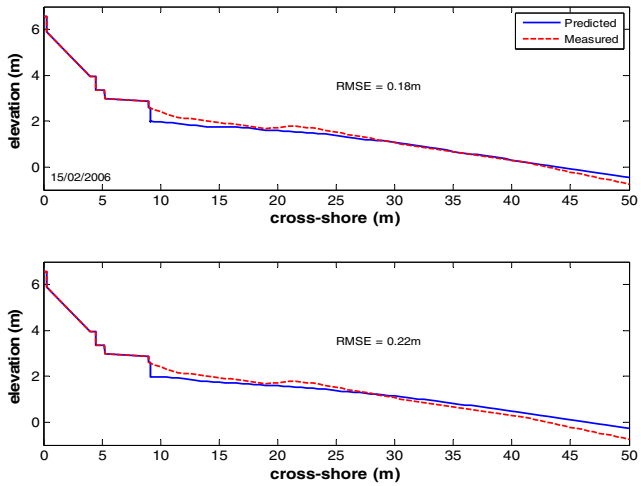

Figure 5. Measured and predicted profiles at Walcott (date 15/02/2006), using empirical wave significant pdfs (top panel) and using empirical wave steepness pdfs (bottom panel).

CONCLUSIONS

Coastal flood risk due to wave overtopping is closely linked to the health of beaches fronting the sea defences and hinterland. A full beach will dissipate much of the incoming wave energy before it reaches the shoreline, and is a very effective part of the defence. Predicting the evolution of beaches remains difficult. However, good improvements in wave forecasting have been made in recent years. Wave forecasts and measurements are, in general, much more easily available than the corresponding information on beaches levels. Finding a reliable link between wave conditions and beach levels would yield an economic means of forecasting beach levels, and thus a crucial factor in assessing coastal flood risk.

In this paper we have applied CCA at Walcott, at which there are coincident records of beach profiles and wave conditions for more than a decade, in order to investigate the correlation between wave climate and beach profile response in the vicinity of a seawall.

The results of the present study indicate that data-driven statistical analysis, such as CCA, are useful for analysing profile response to waves if there is strong correlation between the two variables (beach profiles and wave height distributions). Used in conjunction with EOF to reduce noise of the data, CCA was found to be well suited for identifying combined patterns of variation in the wave and profile data. Also, CCA showed potential as a method for forecasting profile response using different wave parameters for the empirical distributions.

LITERATURE CITED

Clark, D., 1975. *Understanding Canonical Correlation Analysis. Concepts and techniques in modern geography, No.3*. Norwich, UK: Geo Abstracts Ltd.

Cooper, N.J. and Hutchinson, J. 2002. Strategic approach to flood and coastal management in England. *Proceedings 28th International Conference on Coastal Engineering*. (Cardiff, Wales, UK. ASCE), pp. 3485-3497.

Dean, R.G.; Healy, T.R. and Dommerholt, A.P., 1993. A 'blind-folded' test of equilibrium beach profile concepts with New Zealand data. *Marine Geology,* 109 (3-4), 253-266.

Glahn, H.R., 1968. Canonical correlation analysis and its relationship to discriminant analysis and multiple regression. *Journal of Atmospheric Sciences*, 25, 23–31.

Graham, N.E.; Michaelsen, J. and Barnett, T.P., 1987. An investigation of the El Niño-southern oscillation cycle with statistical models: 1. Predictor field characteristics. *Journal of Geophysical Research*, 92 (C13), 14251–14270.

Horrillo-Caraballo, J.M. and Reeve, D.E., 2008. An investigation of the link between beach morphology and wave climate at Duck, North Carolina, USA. *Journal of Flood Risk Management*, 1 (2), 110-122.

Horrillo-Caraballo, J.M. and Reeve, D.E., 2010. An investigation of the performance of a data-driven model on sand and shingle beaches. *Marine Geology*, 274 (1-4), 120-134.

Hotelling. H., 1936. Relations between two sets of variates. *Biometrika*, 28, 312–377.

HR Wallingford, 2002. *Overstrand to Walcott Strategy Study. Hydrodynamics. Part II: Technical Support Information*. Report EX 4692.

Larson, M. and Kraus, N.C., 1995. Prediction of cross-shore sediment transport at different spatial and temporal scales. *Marine Geology*, 126, 111-127.

Larson, M.; Capobianco, M. and Hanson, H., 2000. Relationship between beach profiles and waves at Duck, North Carolina, determined by canonical correlation analysis. *Marine Geology*, 163, 275-288.

Li, Y.; Lark, M. and Reeve, D.E., 2005. The multi-scale variability of beach profiles at Duck, N. C., USA, 1981-2003. *Coastal Engineering,* 52 (12), 1133-1153.

North, G.R.; Bell, T.L.; Cahalan, R.F. and Moeng, F.J., 1982. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110, 699-706.

Różyński, G., 2003. Data-driven modelling of multiple long-shore bars and their interaction. *Coastal Engineering*, 48, 151-170.

USACE, 1984. *Shore Protection Manual*. U.S. Army Waterways Experiment Station. Coastal Engineering Research Center, U.S. Government Printing Office, Washington D.C.

Sutherland, J. and Gouldby, B., 2003. Vulnerability of coastal defences to climate changes. *Proceedings of the Institution of Civil engineers, Water and Maritime engineering*, 156, 137-145.

ACKNOWLEDGEMENT

This research was supported under the Flood Risk Management Research Consortium (FRMRC) via the UK Engineering and Physical Sciences Research Council grants No. GR/S76304/01 and EP/F020511/1. In addition, the support of the European Commission through FP7.2009-1, Contract 244104 - THESEUS ("Innovative technologies for safer European coasts in a changing climate"), is gratefully acknowledged. The authors would also like to thank to Prof. Peter Stansby and Dr. Nicolas Chini from the University of Manchester for the wave and beach profile data.