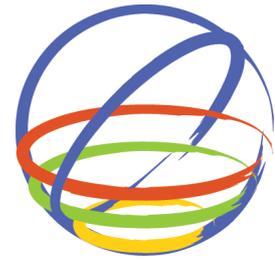


# Long-Wave Runup On A Plane Beach Behind A Conical Island



**Themistoklis S. Stefanakis & Frédéric Dias**

*Ecole Normale Supérieure de Cachan, France & University College Dublin, Ireland*

**Nicolas Vayatis**

*Ecole Normale Supérieure de Cachan, France*

**15 WCEE**  
LISBOA 2012

**Serge Guillas**

*University College London, United Kingdom*

## SUMMARY:

In recent years several tsunamis have hit coastal areas around the world. In some of these areas where small offshore islands were supposed to offer protection from wind and waves, communities were developed. However, post-tsunami survey data has shown that in some cases the runup in these areas was significantly higher than in neighboring locations. Two dimensional numerical simulations using the Nonlinear Shallow Water Equations were employed in order to investigate this phenomenon. The experimental setup for the bathymetry consists of a conical island and a plane beach behind the island, while the incoming wave has a generalized solitary waveform. The problem geometry is dominated by five physical parameters, namely the island slope, the beach slope, the water depth, the distance between the island and the plane beach and the wavelength. In the present study we compare some linear regression-based approaches used in machine learning in combination with kernels that map nonlinear dependencies into higher dimensional linear ones, in order to best describe the runup amplification on the area of the beach behind the island with respect to the runup on a lateral location on the beach, not directly affected by the presence of the island. A notion of sequential experimental design is given and we present some preliminary results.

*Keywords: tsunamis, runup, regression, experimental design*

## 1. INTRODUCTION

During the last decade, the world has encountered the deadly consequences of two of the most severe tsunamis ever recorded, namely the December 2004 event in Indonesia (Liu et al. 2005; Titov et al. 2005) and the most recent March 2011 Japanese tsunami. Of course the aforementioned catastrophes are not the only ones. There is a long list of such events dating back to the tsunami generated by the eruption of Santorini's (Thera) volcano in 1600 BC. Increased public attention to tsunamis during the last decade has raised awareness and preparedness, which is the only effective countermeasure and has saved lives, like during the Chile March 2010 tsunami (Peachey 2010).

Tsunami science has the role of better understanding the generation, propagation and runup of the killer waves and by doing so, tsunamologists should provide coastal communities early warnings and education. Since the 1950's great advancement is observed in the study of tsunami runup on a plane beach, with memorable studies such as those by Carrier & Greenspan (1958), Keller & Keller (1964), Synolakis (1987), Tadepalli & Synolakis (1994), Brocchini & Peregrine (1996), Didenkulova & Pelinovsky (2008) and Antuono & Brocchini (2010). All of the aforementioned articles deal with the mathematical description of runup on plane beaches for several waveforms. The catastrophe in Babi island (Yeh et al. 1993, 1994) raised scientists' attention on tsunami runup on islands and the studies that followed, with laboratory experiments (Briggs et al. 1994) as well as analytical models (Kanoglu & Synolakis 1998), showed that long waves may cause extensive runup on the lee side of the island. Before that, there were a few studies (Homma 1950, Longuet-Higgins 1967, Vastano & Reid 1967) that gave some insight into this phenomenon but never dealt with the runup calculation. What is made clear by the aforementioned studies is the fact that long waves do not behave like wind generated

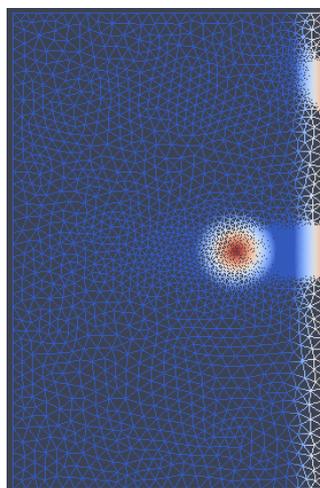
waves and that small islands that would act as natural barriers, transform into amplifiers of wave energy in areas believed to be protected and where coastal communities thrive.

In recent years, the developments in computer science and the increase of computational power in combination with the smaller associated cost compared to laboratory experiments, has led scientists to more and more rely on numerical simulations. However, each numerical simulation has a computational cost, which increases with model complexity and spatiotemporal resolution. Therefore, a series of experiments that have a specific objective, such as maximization/minimization of an output, should be carefully designed in order to reach the desired conclusion with the least number of experiments. Thus, finding the  $\operatorname{argmax}_x f(x)$  where  $f(x)$  is the output of the experiment depending on the inputs  $x$ , is not trivial since we do not know the analytical expression of  $f(x)$  and therefore it should be approximated iteratively. For this reason, machine learning algorithms can be applied in order to build a statistical model  $\hat{f}(x)$  of the experiment which will improve at each step until it reaches a point where it can confidentially model the experiment. Building such a statistical model (emulator) has further advantages, the most important one being the ability to use it instead of the actual simulator since it is much less computationally demanding to evaluate and thus can be applied very rapidly, especially in cases where someone needs a quick forecast. Moreover, an emulator can clarify the relations between the several inputs and the model output. Recently, Sarri et al. (2012) developed an emulator for landslide-generated tsunamis based on the theoretical model of Sammarco & Renzi (2008).

The present study aims to elucidate the tsunami runup amplification on a plane beach behind a small conical island with respect to an adjacent lateral point on the beach not directly influenced by the presence of the island. This is achieved with the use of numerical simulations of the Nonlinear Shallow Water Equations. In addition, several regression type emulators in combination with kernels are built and compared. A notion of sequential experimental design is given and some preliminary results are presented.

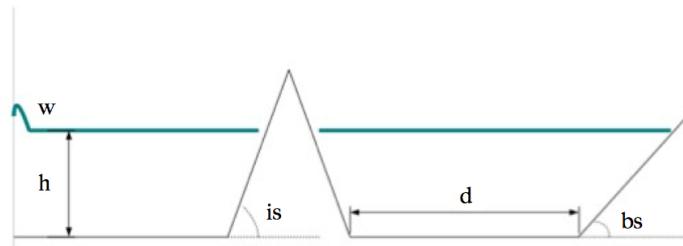
## 2. SETUP AND NUMERICAL SIMULATIONS

The numerical simulations were performed using VOLNA (Dutykh et al. 2011), which solves the Nonlinear Shallow Water Equations. VOLNA can handle the whole life cycle of a tsunami from the generation to the run-up. VOLNA uses a Finite Volume Characteristic Flux scheme with a MUSCL type of reconstruction for higher order terms (Kolgan 1972, 1975) and a third order Runge-Kutta time discretization. The code uses an unstructured triangular mesh, which can handle arbitrary bathymetric profiles and can also be refined in areas of interest. The mesh resolution that we used varied from 500m at the seaward boundary to 2m at the areas where we measured runup (Fig. 2.1.).



**Figure 2.1.** The unstructured triangular grid. Colours represent the bathymetry

The bathymetry consists of a conical island sitting on a flat bottom and a plane beach behind the island. The height of the crest of the island above still water level is always fixed at 100m. The distance between the seaward boundary and the toe of the island is also fixed at 7600m. A single wave profile is prescribed as forcing at the seaward boundary, having the form  $\eta_0(t) = 1.5 \operatorname{sech}^2(\omega t - 2.6)$ . We use this formulation because we want to avoid the solitary wave link between the water depth and the wave amplitude as is discussed in Madsen et al. (2008) and in Madsen & Schäffer (2010). The problem is governed by 5 physical variables (Fig. 2.2. & Table 2.1.), namely the island slope, the plane beach slope, the water depth, the distance between the island and the beach and the prescribed incident wavelength which is controlled by  $\omega$ .

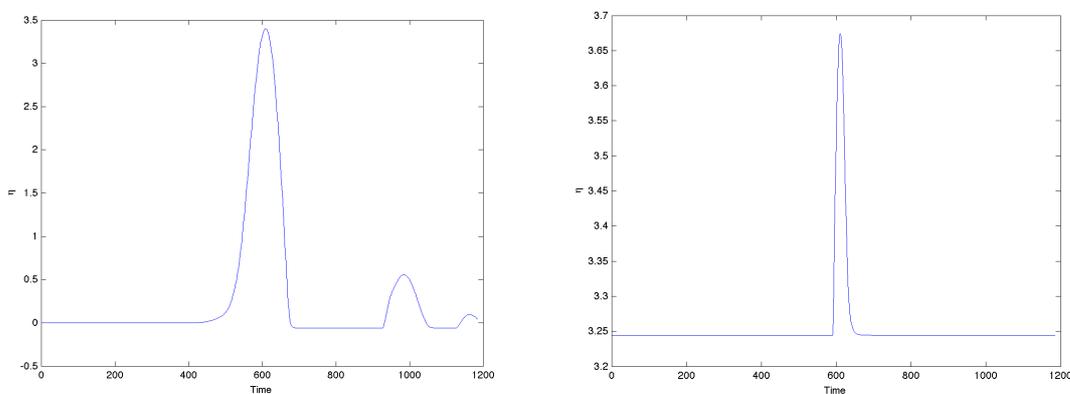


**Figure 2.2.** Schematic of the problem geometry and the governing physical parameters.

**Table 2.1.** Physical variable ranges

is	0.05 – 0.2
bs	0.05 – 0.2
d	0 – 5000m
h	100 – 1000m
$\omega$	0.01 – 0.1 rad/s

The runup was measured on the beach exactly behind the island and on a lateral location on the beach, which was not directly affected by the presence of the island. The runup measurements were made with the aid of 11 equally spaced virtual wave gauges at each location. The actual horizontal spacing of the wave gauges was dependent on the beach slope. The minimum height of the gauges was the still water level and the maximum height was selected to be 5.5m above the undisturbed water surface. The runup never exceeded this height in any of the simulations. The maximum runup is defined as the maximum recorded waveheight at the highest wave gauge. When the wave did not reach the height of a gauge, then that gauge did not record any signal. Two sample gauge recordings are shown in Fig 2.3.



**Figure 2.3.** Sample wave gauge recordings for the same event. The left is from a gauge positioned at the still water level, while the right is positioned 3.25m above the water surface.

In order to fill the input parameter space we need to choose the input points in such a way that maximal information is obtained with a moderate number of points. For this reason, we used a Latin Hypercube Sampling (McKay et al. 1979) with maximization of the minimum distance between points. When using the Latin Hypercube Sampling (LHS) of a function of  $M$  variables, the range of each variable is divided into  $N$  equally probable, non-overlapping intervals. Then one value from each interval is randomly selected for every variable. Finally, a random combination of  $N$  values for  $M$  variables is formed. The maximization of the minimum distance between points is added as an extra constraint. The LHS is found to lead to better predictions than regular grids when used with multivariate emulators (Urban & Fricker 2010). However, here we have the choice to either randomly or smartly select points in the LHS. The smart selection of points is the subject of the sequential experimental design discussed in Section 5. In order to accurately cover the input space, we ran 200 simulations.

### 3. STATISTICAL MODEL

The statistical model will be used as an emulator of the numerical code having as input the  $M$ -dimensional vector of the physical variables and as output the scalar value of the runup amplification. In statistics, regression analysis refers to techniques for modeling the functional relationship between an output and one or more input variables. We are focusing on a statistical model such as  $Y = f(X) + \varepsilon$ , where  $X$  is a matrix containing the input variables and  $Y$  is a vector containing the outputs.  $\varepsilon$  is the perturbation or “noise”, a real random variable, independent of  $X$ , which follows the normal distribution rule  $\varepsilon \sim N(0, \sigma^2 I_N)$ . The performance of these regression models in practice depends on the form of the data considered, and how it relates to the regression approach being used. For the description of the runup amplification problem we tested and compared three different statistical models.

#### 3.1. Ridge Regression

All of them have as a base the ridge regression, which is a regularized linear regression and is more robust than the classic linear regression since it allows for non-singular inverted input matrices. Since ridge regression is a linear model,  $f(X) = X\beta$ , where  $X$  is the design matrix with size  $N \times M$ , with  $N$  being the sample size and  $M$  the dimension of the problem (number of input variables). Therefore, one has to find the optimal  $\hat{\beta}$  that minimizes the least squares error:

$$\hat{\beta}(\lambda) = \operatorname{argmin}_{\beta} (\|Y - X\beta\|^2 + \lambda \|\beta\|^2) \quad (3.1)$$

where  $\lambda$  is the penalty coefficient. When  $\lambda = 0$  the above becomes a standard linear regression. The solution to the above equation is:

$$\hat{\beta}(\lambda) = (X^T X + \lambda I)^{-1} X^T Y \quad (3.2)$$

where  $I$  is the  $M \times M$  identity matrix. The aim of the ridge regression is to be functional without any condition on  $X^T X$ .

#### 3.2. Learning, Validation and Test

One classical technique in Statistics is to partition a sample of data into complementary subsets: performing the parameter ( $\beta$ ) estimation on one subset (called the training set), validating the smoothing parameter ( $\lambda$ ) on another subset (called the validation set), and testing the whole model on the last subset (called the testing set). The most commonly used partition of the initial data set is 60%-20%-20% corresponding to each of the previously mentioned subsets, respectively. The procedure is the following:

- We start computing  $\hat{\beta}$  over  $X_{\text{train}}$  and  $Y_{\text{train}}$ . This process optimizes the model parameters to make the model fit the training data as well as possible.
- Then we validate  $\lambda$  by looking at the smallest error over  $X_{\text{valid}}$  and  $Y_{\text{valid}}$ .  
(The previous two procedures are followed for each regression model respectively)
- Finally, we compute the mean error over  $X_{\text{test}}$  and  $Y_{\text{test}}$  to compare the models.

Sometimes, when we do not possess enough data, the validation and the test are inconsistent. A solution is to mix the base enough, so that each data point has been in both validation and testing sets, and to average the errors. This procedure is called *cross-validation*. There exist several types of cross-validation, two of them being the:

- *k*-fold cross-validation: The original sample is partitioned into *k* subsamples. Then 60% of the *k* subsamples are retained as the training data, 20% as the validation data and 20% as the test data. We repeat the cross-validation *k* times in order for each of the *k* subsamples to be used exactly once as the validation and the test data. The advantage of this method over repeated random sub-sampling is that all observations are used for training, validation and test.
- Leave-one-out cross-validation: This cross-validation retains only a single observation for the validation data and a single observation for the test data. All the remaining observations are the training data. Therefore we repeat the leave-one-out cross-validation, as many times as the number of observations, which may be expensive from a computational point of view. However, this type of cross-validation could be proved useful when the number of observations is limited.

### 3.3. Nonlinear Model: Kernelized Ridge Regression

Because ridge regression is linear, a way to deal with nonlinear dependencies between the input variables is to use kernels. The Kernel trick (Cristianini & Shawe-Taylor 2000) is a way of mapping observations from a general input space into a high dimensional feature space, in the hope that the observations will gain meaningful linear structure in the feature space. The trick is to use learning algorithms that only require dot products between the vectors in the feature space in order to avoid the explicit mapping, and choose the mapping such that these high-dimensional dot products can be computed within the original space, by means of a kernel function  $K(.,.)$  :

$$K(x_1, x_2) = \varphi(x_1)^T \varphi(x_2) \quad (3.3)$$

where  $\varphi : \mathbb{R}^M \rightarrow \mathbb{R}^Q$  and  $Q$  is the dimension of the feature space, which can go to infinity. Thus the new regression model becomes

$$f(x) = w^T \varphi(x) \quad (3.4)$$

Now set the change of parameters

$$w = \Phi^T \beta, \quad \text{where } \Phi = (\varphi(X_1), \dots, \varphi(X_N)) \quad (3.5)$$

and then the form of the regression model is the following :

$$f(x) = \sum_{i=1}^N \beta K(x, X_i) \quad (3.6)$$

The parameter  $\beta$  is estimated by the ridge regression solution as

$$\hat{\beta} = (K + \lambda I)^{-1} Y \quad (3.7)$$

We can use several types of kernels, the efficiency of which depends on the form of the data. For the current study we tested the following two types:

- Polynomial Kernels of the form  $K(X, X') = (X^T X')^d$ , where  $d$  is the degree of the polynomial.

- Gaussian Kernels of the form  $K(X, X') = \exp\left(\frac{\|X-X'\|^2}{2\sigma}\right)$

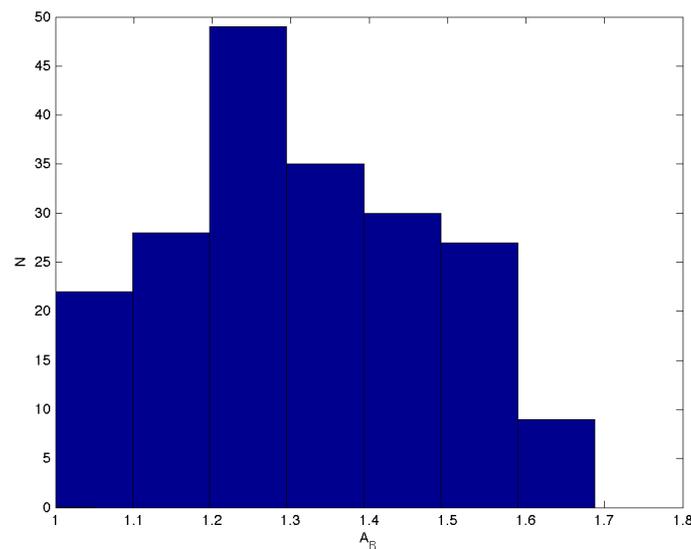
A validation may be performed to define the optimal degree of the polynomial kernel or the optimal standard deviation  $\sigma$  of the Gaussian kernel.

#### 4. RUNUP AMPLIFICATION

After running 200 simulations with the input variables in the intervals presented in Table 2.1, the runup amplification on the plane beach behind the island with respect to a lateral point on the beach, for which the influence of the island was minimal, was found to be  $1 < A_R < 1.69$  (Fig. 4.1.). The fact that the runup amplification is always greater than one clearly indicates that the island cannot offer any kind of protection from tsunamis. On the contrary, the energy of the tsunami is focused on the lee side of the island leading to increased runup and consequently inundation on the area of the plane beach directly behind the island.

**Table 4.1.** Mean Square Error for each type of model and cross-validation

Method / Cross-Validation	5 - fold	Leave-One-Out
Without Kernel	0.0988	0.1198
Polynomial Kernel	0.0086	0.0164
Gaussian Kernel	0.0081	0.0148



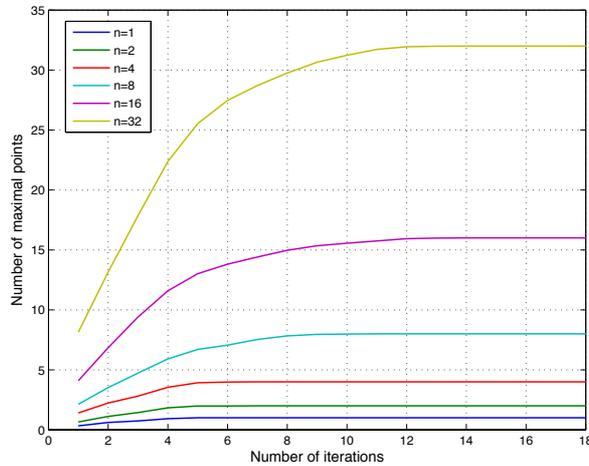
**Figure 4.1.** Histogram of the runup amplification on the point of the plane beach directly behind the island.

We built three emulators, namely a bare ridge regression, a ridge regression with a polynomial kernel and a ridge regression with a Gaussian kernel. By comparing the mean square error of all three of them in the testing set we found that the best performing model is the ridge regression with the Gaussian kernel (Table 4.1.). The use of kernels improves the performance of the problem by an order of magnitude. This implies that the runup amplification problem is highly nonlinear. The maximum runup amplification predicted on the whole sample by the best performing emulator is  $A_R = 1.641$ .

#### 5. SEQUENTIAL EXPERIMENTAL DESIGN

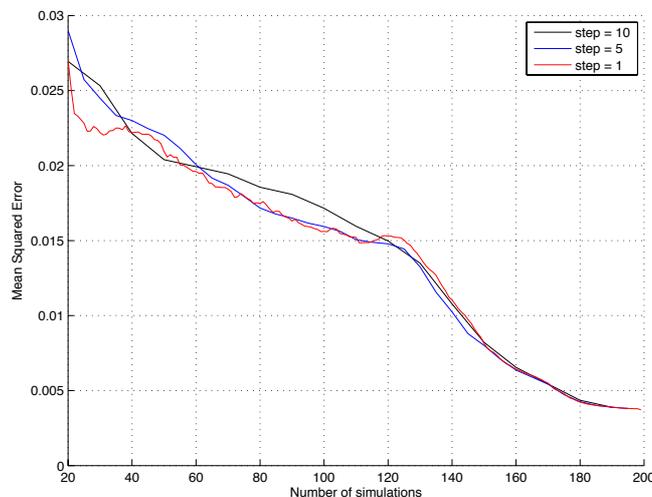
Running a large number of simulations without any “guidance” in order to find the combination of values of the input variables that maximizes a scalar output might be computationally expensive and

inefficient. One needs to learn from each simulation that he runs and use this knowledge to guide him to the desired output with the least number of experiments. This is the aim of the sequential experimental design. At each step the extra information gathered by each experiment refines the statistical model. We use this emulator to predict which unlabelled point will bring us more information, according to our criteria, and therefore we should examine next.



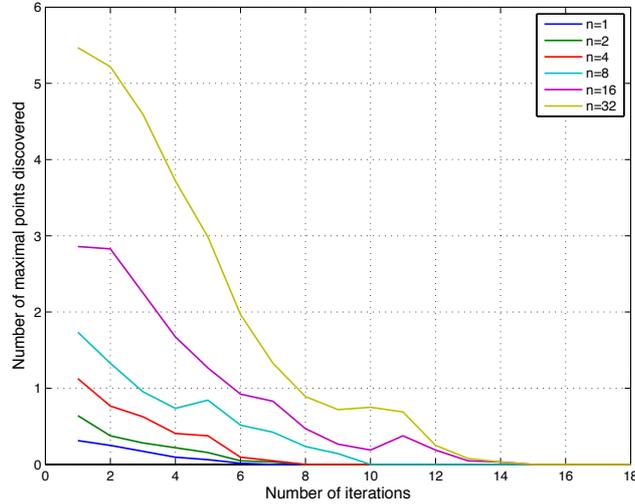
**Figure 5.1.** Average number of total maximal points discovered at each iteration, when  $q = 10$ .

We started by randomly selecting 20 points and we used as a model the ridge regression with the Gaussian kernel. Then, by using this emulator, we made predictions on the remaining 180 points and afterwards we selected to examine the  $q$  points that gave the highest prediction. The parameters  $\lambda$  and  $\sigma$  of the model can be selected at each step by cross-validation. The same procedure can be repeated until all simulations are ran. One should provide a stopping criterion, but we leave that for a future study. In Fig. 5.1. we present the mean over 64 random initializations of the total number of maximal points we have discovered at each iteration when  $q = 10$ . By  $n$  maximal points we mean the  $n$  points with the highest runup amplification. In the same plot we can observe that the maximum can be discovered after 60 experiments, while for the 32 higher runup amplifications we need 130 experiments. The step size  $q$  does not affect the speed of convergence, as can be seen in Fig. 5.2. where we plot the evolution of the mean squared error.



**Figure 5.2.** The evolution of the mean squared error.

In Fig. 5.3. we plot the average over 128 random initializations of the number of maximal points discovered in each iteration. We can see that more points are found in the first iterations than in the subsequent ones. However, the  $n = 16$  and  $n = 32$  curves present a smaller peak around the 12<sup>th</sup> and the 11<sup>th</sup> iteration respectively. This signifies the existence of a smaller, narrower regime of high runup amplification.



**Figure 5.3.** Average number of maximal points discovered in each iteration.

## 6. DISCUSSION

In this study we ran 200 simulations in order to see the effect of a small conical island in the vicinity of a plane beach on the tsunami runup on the beach behind the island. We observed that the island always causes amplification to the runup on the beach behind it, when compared to another point on the beach, which is not affected by the presence of the island. Therefore, coastal communities built behind small offshore islands should be informed that in the case of a tsunami, the island will not offer any protection, but on the contrary, it will amplify the runup. The situation may become even worse if resonant wave interactions occur (Stefanakis et al. 2011). Three different emulators based on ridge regression were built and the best performing model was found to involve Gaussian kernels, which signifies that there are nonlinear relations between the input variables and the runup amplification. Other statistical models, such as trees, should also be applied and compared. The advantage of trees is that their results are easily interpretable. We also implemented a sequential experimental design that allows us to find the maximum output with significantly less number of simulations. However, the experimental design was applied after the total number of simulations was conducted and the results of this technique were compared with a prior knowledge of the maximum. Hence, it is important to find a stopping criterion, which can be applied without knowing a priori the behaviour of the function at all the points. Moreover, in this study we only looked at the runup amplification, whereas an improved approach would be to look at both the actual value of the runup and the runup amplification, which would transform the statistical problem to multivariate optimization. Last but not least, future research on the subject should take into account both physical and numerical variables such as grid size, CFL condition and wave gauge spacing.

## AKNOWLEDGEMENT

We would like to acknowledge David Buffoni, Alexandre Robicquet, Chloé Pasin, Maxime Stauffert and Emile Contal, students at the Centre des Mathématiques et Leurs Applications of ENS Cachan for their valuable contributions to the statistical analysis.

## REFERENCES

- Antuono, M. and Brocchini, M. (2010). Solving the nonlinear shallow- water equations in physical space. *Journal of Fluid Mechanics*. **643**, 207–232.
- Briggs, M.J., C.E. Synolakis, G.S. Harkins and D.R. Green (1995): Laboratory experiments of tsunami runup on circular island. *Pure and Applied Geophysics*. **144:3/4**, 569–593.
- Brocchini, M. and Peregrine, D.H. (1996). Integral flow properties of the swash zone and averaging. *Journal of Fluid Mechanics*. **317**, 241–273.
- Carrier, G. F. and Greenspan, H. P. (1958). Water waves of finite amplitude on a sloping beach. *Journal of Fluid Mechanics*. **4**, 97 – 109.
- Cristianini, N. and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
- Didenkulova, I. and Pelinovsky, E. (2008). Run-up of long waves on a beach: the influence of the incident wave form. *Oceanology*. **48:1**, 1–6.
- Dutykh, D., Poncet, R. and Dias, F. (2011). The VOLNA code for the numerical modeling of tsunami waves: Generation, propagation and inundation. *European Journal of Mechanics B/Fluids*. **30:6**, 598–615.
- Homma, S. (1950). On behaviour of seismic sea waves around the circular island. *Geophysics Magazine*. **21**, 199–208.
- Kanoglu, U. and Synolakis, C. E. (1998). Long wave runup on piecewise linear topographies. *Journal of Fluid Mechanics*. **374**, 1–28.
- Keller, J and Keller, H (1964). Water wave run-up on a beach. *ONR Research Report Contract No. NONR-3828(00)*.
- Kolgan, N.E. (1972). Application of the minimum-derivative principle in the construction of finite-difference schemes for numerical analysis of discontinuous solutions in gas dynamics. *Uchen. Zap. TsAGI*. **3:6**, 68–77.
- Kolgan, N.E. (1975). Finite-difference schemes for computation of three dimensional solutions of gas dynamics and calculation of a flow over a body under an angle of attack. *Uchen. Zap. TsAGI*. **6:2**, 1–6.
- Liu, P. L.-F., Lynett, P., Fernando, H., Jaffe, B., Fritz, H., Higman, B., Morton, R., Goff, J., and Synolakis, C. (2005). Observations by the international tsunami team in Sri Lanka. *Science*. **308**, 1595.
- Longuet-Higgins M.S. (1967). On the trapping of wave energy round islands. *Journal of Fluid Mechanics*. **29**, 781–821.
- Madsen, P. A., Fuhrman, D. R. and Schäffer, H. A. (2008). On the solitary wave paradigm for tsunamis. *Journal of Geophysical Research*. **113:C12012**, 1–22.
- Madsen, P.A. and Schäffer, H.A. (2010). Analytical solutions for tsunami runup on a plane beach: single waves, N-waves and transient waves. *Journal of Fluid Mechanics*. **645**, 27–57.
- McKay, M.D., Beckman, R.J. and Conover W.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. **21**, 239–245.
- Peachey, P. (2010). How 12-year-old girl saved her Chilean island from catastrophe. The Independent. Independent Print Limited, London, U.K.
- Sammarco, P. and Renzi, E. (2008). Landslide tsunamis propagating along a plane beach. *Journal of Fluid Mechanics*. **598**, 107–119.
- Sarri, A., Guillas, S. and Dias, F. (2012). Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. *Natural Hazards and Earth System Sciences*. (in print)
- Stefanakis, T.S., Dias, F. and Dutykh, D. (2011). Local Run-up Amplification by Resonant Wave Interactions. *Physical Review Letters*. **107**, 124502.
- Synolakis, C.E. (1987). The runup of solitary waves. *Journal of Fluid Mechanics*. **185**, 523–545.
- Tadepalli, S. and Synolakis, C.E. (1994). The run-up of N-waves on sloping beaches. *Proceedings of the Royal Society A*. **445**, 99–112.
- Titov, V., Rabinovich, A.B., Mofjeld, H.O., Thomson, R.E. and Gonzalez, F.I. (2005). The global reach of the 26 December 2004 Sumatra tsunami, *Science* **309**, 2045–2048.
- Urban, N.M. and Fricker, T.E. (2010). A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Computational Geosciences*. **36**, 746–755.
- Vastano, A.C. and Reid, R.O. (1967). Tsunami response for island: Verification of a numerical procedure. *Journal of Marine Research*. **25**, 129–139.
- Yeh, H., Imamura, F., Synolakis, C.E., Tsuji, Y., Liu, P.L.-F. and Shi, S. (1993). The Flores Island tsunamis. *EOS Transactions*. **369**, 371–373.
- Yeh H., Liu, P.L.-F., Briggs M.J. and Synolakis, C.E. (1994). Tsunami Catastrophe in Babi Island. *Nature*. **372**, 6503.