

# PREDICTION OF EQUILIBRIUM SUBAQUEOUS DUNE CHARACTERISTICS USING GENETIC ALGORITHMS

Doré Arnaud, University of Auckland, [arnaud.dore@auckland.ac.nz](mailto:arnaud.dore@auckland.ac.nz)  
Coco Giovanni, University of Auckland, [g.coco@auckland.ac.nz](mailto:g.coco@auckland.ac.nz)

## INTRODUCTION

Sand dunes are ubiquitous in natural subaqueous environments and may pose a significant risk in coastal setups for many domains such as the offshore industry or marine renewable energies (Vantorre et al., 2013 ; Barrie and Conway, 2014). Observations of bedform development show that an initially flatbed evolves through different phases: an incipient bedform phase, a growing phase and a stabilizing phase leading to a fully developed dune field. Finite-amplitude dune equilibrium is essentially controlled by the flow depth, the bed shear stress, and the inertia length of the sediment transport in suspension,  $L_{sat,s}$ , and for bedload at smaller depths,  $L_{sat,b}$  (Doré et al., 2023). Existing predictors fail to describe dune dimensions at equilibrium (height,  $H$ , and wavelength,  $\lambda$ ) (Bradley and Venditti, 2017).

Data availability and the poor description of complex morphodynamic phenomena has led to an increasing use of Machine learning in coastal science (Goldstein et al., 2019). In the present work we use Genetic algorithms (GAs) to perform regression analysis on dune laboratory data available in the literature. Our results showed that GAs outperform usual predictors against performance metrics.

## METHODOLOGY

We used a compiled dataset of dune dimensions and flow characteristics from both laboratory flume experiments and field observations from all published and unpublished sources known and accessible (Bradley and Venditti, 2017). The data was filtered for aberrant values and missing data, and important parameters were derived for the regression study (settling velocity,  $L_{sat,b}$ ,  $L_{sat,s}$ ).

We used GAs, which are based on the evolutionary idea of natural selection and genetics, according to the principle of “survival of the fittest”. GAs mimics species evolutionary processes, where possible solutions represent chromosomes. Reproduction operators are applied for each generation of algorithms to perform mutations and recombination (tournament selection, crossover, mutation). The interest of such evolutionary models is the gradual algorithm optimization, at each generation, leading to the fittest and simplest solution to the regression problem.

We employed a Maximum Dissimilarity Algorithm (MDA), which is a selection method, to create our training and test data sets. This method is particularly suitable for reduced data sets to avoid any form of bias. In our analysis, we tested other approaches and concluded that MDA outperformed random splitting and stratification methods. Random splitting and stratification techniques often resulted in over-fitting, as they were not

able to effectively capture the nuances and complexities of the data. By utilizing the MDA approach, we were able to improve the overall performance and reliability of the predictors. Finally, we carried a global sensitivity analysis (GSA) to help identify the key parameters controlling dune dimensions using the PAWN method (Pianosi and Wagener, 2015). PAWN provides a sensitivity index for each model parameter.

## RESULTS

By exploring the hyperparameter space through iterative searches, we aimed to discover the most optimal algorithms. Our main goal was to find a balance between model simplicity and performance. To achieve this, we experimented with various combinations of input features and operators employed in algorithm construction.

The resulting estimators significantly outperformed the existing theoretical formulas, as evidenced by Figure 1 and Figure 2. For instance, our estimators achieved R2 scores as high as 0.7, a stark contrast to the negative values yielded by the existing predictors. Furthermore, our estimators exhibited lower Root Mean Square Error (RMSE) values, typically below 0.4, in contrast to the values around 0.7 obtained by the previous predictors.

The complexity of the derived equations and the absence of a valid theoretical formula in the literature highlight the intricate nature of the physical processes that govern dune evolution. These nonlinear processes arise from the feedback interactions between flow, sediment transport, and bed morphology until a state of equilibrium is attained. In previous studies simpler estimators were achieved, *e. g.* for wave ripples, wave setup, or settling velocity, where the dynamics of the system is highly correlated to a small number of parameters.

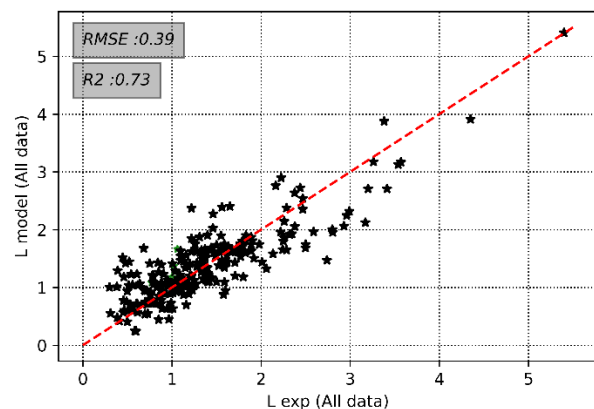


Figure 1 - Model-data correlation of equilibrium dune wavelengths using a genetic algorithm (flume data)

The PAWN global sensitivity analysis provided valuable insights into the contribution of each input parameter to the determination of dune dimensions at equilibrium. Figures 3 and 4 illustrate the weights assigned to each parameter, revealing their relative importance. This finding is consistent with the findings of Doré and Coco (2024), further validating the significance of these parameters in understanding dune dynamics.

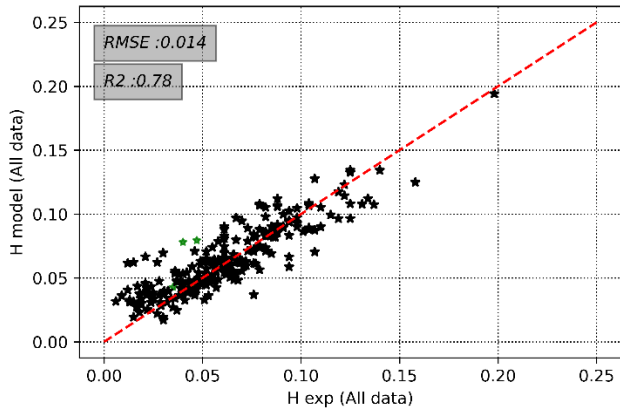


Figure 2 - Model-data correlation plot of equilibrium dune heights using a genetic algorithm (flume data)

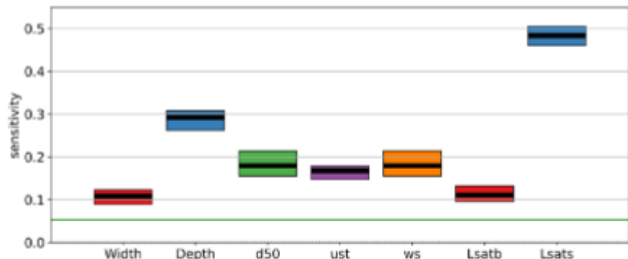


Figure 3 - PAWN sensitivity analysis for the dune wavelength (flume data)

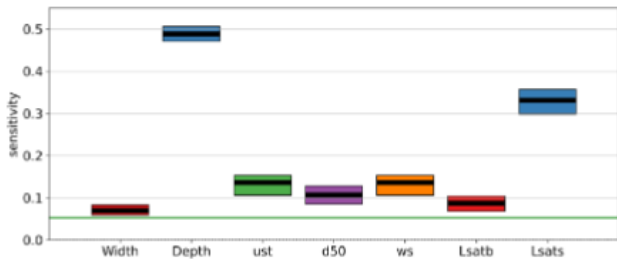


Figure 4 - PAWN sensitivity analysis for the dune height (flume data)

**CONCLUSIONS**

GP are evolutive algorithms that utilize Darwinian principles to transform generations of equations through recombination, mutation, and cloning. We fine-tuned the

hyperparameters of the genetic program to achieve an optimized fitness of predictor estimations versus laboratory measurement data. Equations remained intricate, and we conducted a global sensitivity analysis to evaluate the impact of input features on the solutions. The sensitivity analysis confirmed the instrumental role of the sediment transport inertial length and of the depth in setting equilibrium dune wavelength and height. The available data should be supplemented by further targeted measurement campaigns, to take better account of underrepresented flow regimes and the effects of channel width on the dimensions of the dunes at equilibrium. However, the predictors generated outperform currently available ones, and offer an easily implementable and operational alternative for prediction of dune geometry for steady-state flows.

**ACKNOWLEDGEMENTS**

This research is supported by H2020-MSCA-IF-2020 grant No.101021848.

**REFERENCES**

Barrie, J. V., & Conway, K. W. (2014). Seabed characterization for the development of marine renewable energy on the pacific margin of canada. *Continental Shelf Research*, 83 , 45-52.

Bradley, R., & Venditti, J. (2017). Reevaluating dune scaling relations. *Earth Science Reviews*, 165.

Doré A., G. Coco (2024). Numerical modelling study of finite amplitude dune saturation at low Froude numbers, under review, 2024.

Goldstein, E., Coco, G. & Plant, N. G (2019). A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Rev.* 194, 97-108.

Pianosi, F., Wagener, T. (2015), A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, *Environmental Modelling & Software*, 67, 1-11.

Vantorre, M., Lataire, E., Candries, M., van Doorn, J., & van Heel, D. (2013). An equivalent bottom for navigation above irregular bottoms. In Van lancker, v. and garlan, t. (eds), 2013. *marid 2013.*, pp. 301-308.