Third Workshop on the Future of
Machine Learning in Geotechnics
Florence, Italy
October 15-17 2025

# EXPLOITING DATABASE INFLUENCE ON THE PREDICTION OF OFFSHORE PLATE ANCHOR UPLIFT CAPACITY THROUGH PCE-BASED METAMODELS

Alessio Mentani
Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM),
University of Bologna, Italy. E-mail: alessio.mentani2@unibo.it
Ritesh Gupta
Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR,
38000 Grenoble, France E-mail: ritesh.gupta@3sr-grenoble.fr
Anamitra Roy, Rene Kurniadi, Shiaohuey Chow
The University of Melbourne, Victoria (Australia). E-mail: anamitra.roy@unimelb.edu.au;
rene.kurniadi@student.unimelb.edu.au; shiaohuey.chow@unimelb.edu.au

*Keywords*: Offshore plate anchors, metamodel, polynomial chaos expansion.

## 1 Introduction

As the offshore renewable energy sector expands into deeper waters, the need for reliable and cost-effective anchoring systems becomes critical. Offshore platforms, including floating wind turbines and wave energy converters, require secure mooring. Among the anchoring solutions, plate anchors have emerged as a promising option due to their ease of installation and high capacity. Once embedded, plate anchors are primarily subjected to uplift loading. The prediction of their capacity can be approached using limit analysis theory (Merifield et al., 2006). However, analytical methods may not fully capture key aspects of anchor behavior. In particular, centrifuge tests have revealed that analytical solutions cannot fully represent complex phenomena such as the transition from embedded to shallow failure mechanisms (Hao et al., 2019).

To address these limitations, researchers increasingly rely on advanced finite element (FE) models, calibrated and validated against centrifuge tests (Roy et al., 2021a). These models incorporate sophisticated constitutive laws (Roy et al., 2021b), offering improved soil–structure interaction representation. However, such high-resolution models are computationally expensive and impractical for use in early-stage design or parametric studies. Similarly, extensive centrifuge campaigns remains logistically and financially demanding.

To bridge this gap, this study proposes the use of Polynomial Chaos Expansion (PCE) as a surrogate modeling technique to predict the uplift capacity of circular plate anchors in sand. The PCE metamodel establish a regression-like correlation between input parameters and the desired output. Here, the input-output database, accumulated through the aforementioned solutions, serves as training database for the PCE-based metamodel.

The scope of the study is to demonstrate how the PCE technique could be employed to capture the anchor response. A key contribution lies in enhancing the predictive capability of the PCE metamodel through two main strategies:

- strategically filtering the input-output database (e.g., outliers) based on input parameter distributions;
- integrating the underlying physics by clustering the database, particularly stratifying it by the anchor embedment.

## 2 Methodology

The plate anchor uplift capacity factor, $N_\gamma$, is investigated in this study. The anchor capacity was estimated from various experimental and finite element studies, which results are compiled into a dataset comprising three input parameters. A polynomial chaos expansion (PCE)-based metamodeling technique is employed (a) to generalize the input-output relationship for evaluating $N_\gamma$ at new input sets at negligible computational cost, and (b) to enable statistical and probabilistic analyses of the soil-structure interaction response. The PCE models are developed by incorporating the underlying physics through data clustering. Accuracy measures are then evaluated to assess the prediction quality of the metamodel.

### 2.1 Polynomial Chaos Expansion metamodel

The PCE metamodel is a spectral approximation function that represents the output of a model as a series of orthogonal polynomials in the input variables. The PCE surrogate $g(\boldsymbol{x})$ approximates a model response $f(\boldsymbol{x})$ as:

$$g(\boldsymbol{x}) = \sum_{\alpha \in A} c_\alpha \Psi_\alpha(\boldsymbol{x}) \quad (1)$$

where:

- $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ denotes a vector of $n$ independent input variables;
- $\Psi_\alpha(\boldsymbol{x})$ are multivariate orthogonal polynomials constructed as tensor products of univariate polynomials;
- $c_\alpha$ are the corresponding expansion coefficients;
- $A$ is the set of retained multi-indices defining the truncated basis.

Orthogonality is defined with respect to the probability distribution of the input variables considering the inner product in a Hilbert space. This governs the choice of the polynomial family (Sudret, 2008). In this study, a hyperbolic truncation scheme is adopted to construct the index set $A$, based on the condition:

$$\|\alpha\|_q = \left(\sum_{i=1}^{n} \alpha_i^q\right)^{\frac{1}{q}} \leq p \quad (2)$$

Where the hyperbolic parameter, $q$, ranges as $0 < q \leq 1$, and $p$ is the maximum polynomials degree. This truncation promotes sparsity by favouring low-order interaction terms among input variables.
The expansion coefficients $c_\alpha$ are estimated with a regression-based approach, specifically the Least Angle Regression (LARS) algorithm (Blatman and Sudret, 2010). LARS is a sparse regression method that efficiently selects the most relevant basis functions. Starting from an empty model, LARS iteratively adds basis that are most correlated with the residuals, adjusting coefficients to minimise residuals, while maintaining sparsity. This is particularly effective when candidate basis functions outnumber training data. Formally, given a training dataset of $N$ input-output pairs, the PCE coefficients $\boldsymbol{c} = \{c_\alpha\}$ are found by solving the regularized least squares problem:

$$c = \|\boldsymbol{F} - \boldsymbol{\Psi} c\|_2^2 + \lambda \|c\|_1 \quad (3)$$

where:

- $\boldsymbol{F} = [f(\boldsymbol{x}^{(1)}), \dots, f(\boldsymbol{x}^{(m)})]$ is the vector of model response;
- $\boldsymbol{\Psi}$ is the $N \times M$ matrix of evaluated polynomial basis functions, with $N$ being the size of the training dataset and $M$ the total number of candidate polynomials and entries $\Psi_{i,j} = \Psi_{\alpha_j}\left(x^{(i)}\right)$;
- $\lambda \geq 0$ is a regularization parameter controlling the trade-off between data fit and sparsity.

LARS efficiently approximates the solution path without requiring an exhaustive search, making it well-suited for high-dimensional surrogate modelling.

#### 2.1.1 Accuracy measure

The accuracy of the constructed PCE is estimated by using a leave-one-out (LOO) cross-validation method. It consists in building $N$ meta-models ($g^{PC\setminus i}$), each one created on a reduced experimental design ($N$-$i$) and comparing its prediction on the excluded point ($i$). The PCE overall performance is then analysed in terms of the accuracy score, $Q^2$, computed as:

$$Q^2 = 1 - \frac{\sum_{i=1}^{N}\left(f\left(x^{(i)}\right) - g^{PC\setminus i}\left(x^{(i)}\right)\right)^2}{\sum_{i=1}^{N}\left(f\left(x^{(i)}\right) - \bar{f}\right)^2} \quad (4)$$

Where $\bar{f}$ is the mean of the model response.

### 2.2 Database

The uplift capacity of the circular plate anchor is characterised by the anchor factor $N_\gamma = q_u/\sigma'_{v0}$, where $q_u$ is the peak anchor capacity, and $\sigma'_{v0}$ is the effective overburden stress at the anchor depth. To train the PCE model, a database comprising 128 data points was assembled from multiple sources. Specifically, the dataset includes 67 finite element analysis (FEA) results and 61 centrifuge experiments, covering a broad range of conditions.

An analysis of the available database led to the identification of the three most influential input variables governing the anchor's response: the relative density, $R_D$, of the surrounding sand; the anchor diameter, $D$; and the embedment ratio, $H/D$, where $H$ is the depth of installation.

Data from six different sources were considered, encompassing plate diameters ranging from 0.4 m to 5 m, relative densities between 30% and 90%, and embedment ratios from 1.05 to 15.

A summary of the dataset is provided in Table 1. Figure 1 illustrates the distribution of the three input variables against the anchor factor. The dataset contains a slightly greater number of data points for shallow embedment ratios, which correspond to the typical installation depths of practical applications. Additionally, many of the anchors with smaller diameters originate from centrifuge experiments, where small-scale models were tested under acceleration levels of up to 20*g*.

The FEA data were generated using the commercial FE software Abaqus under axisymmetric conditions, with soil behavior governed by a modified SANISAND model (Roy et al. 2021b).

Table 1. Summary of the compiled database: sources and range of variation of input and output variables.

| Reference | ID | Source | Size, $N$ | Diameter, $D$ (m) | Embedment ratio, $H/D$ (-) | Relative density, $R_D$ (%) | Capacity factor, $N_\gamma$ (-) |
|---|---|---|---|---|---|---|---|
| Roy *et al.* 2021a | DB1 | *exp* | 19 | 0.6 ÷ 2.0 | 1.8 ÷ 6.0 | 45 ÷ 75 | 2.4 ÷ 14.5 |
| Hao *et al.* 2019 | DB2 | *exp* | 21 | 0.4 | 2.0 ÷ 12.0 | 85 ÷ 92 | 5.8 ÷ 35.7 |
| Roy *et al.* 2021a | DB3 | *FEA* | 37 | 1.0 ÷ 5.0 | 1.0 ÷ 4.0 | 30 ÷ 85 | 2.6 ÷ 15.0 |
| Kurniadi *et al.* 2025 | DB4 | *FEA* | 30 | 0.4 | 2.0 ÷ 15.0 | 40 ÷ 85 | 3.9 ÷ 33.0 |
| Rasulo *et al.* 2017 | DB5 | *exp* | 13 | 0.25 | 1.8 ÷ 4.0 | 55 ÷ 60 | 4.8 ÷ 11.0 |
| Giampa *et al.* 2017 | DB6 | *exp* | 8 | 0.15 ÷ 0.4 | 1.8 ÷ 5.0 | 32 ÷ 45 | 4.1 ÷ 20.4 |

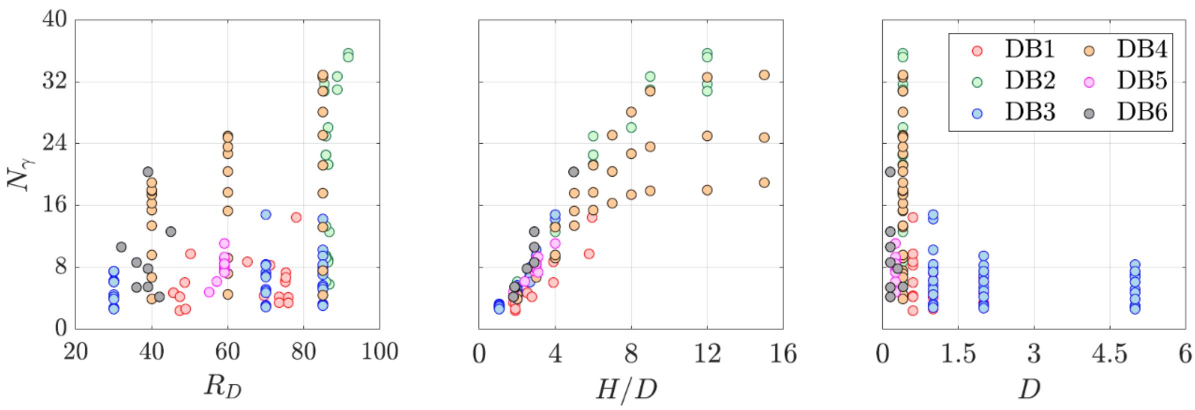

Figure 1. Distribution of the three input variables with respect to the anchor factor in the compiled database.

### 3 Results

The input database was first filtered to remove repeated input sets, retaining the mean value of $N_\gamma$. The resulting dataset comprises 112 datapoints, of which 11 corresponding to $D$ = 5m were excluded due to the large gap in the input space. Thus, 101 datapoints (Figure 2) were used for PCE metamodeling.

For the PCE development, uniformly distributed marginals of the input parameters were adopted for simplicity. The dataset was split into training and validation subsets. To achieve this, the database was sequenced by chronological sum of all input parameter values and then alternately assigned to the training and validation sets. The hyperbolic parameter, $q$, was varied from 0.5 to 1.0 in increments of 0.05, and the maximum polynomial degree, $p$, ranged from 1 to 20.

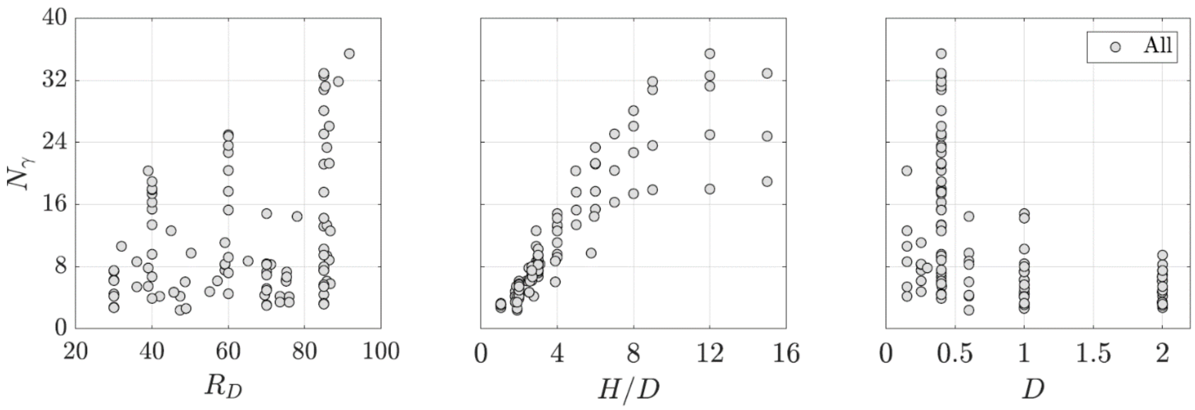


Figure 2. Filtered (for multiple entries) and cropped (for D>2) database for PCE metamodeling work

The results of PCE in terms of accuracy score, relative prediction error and distribution of $N_\gamma$ estimates are presented in Figure 3. Two datasets were used for validation: (*i*) the training dataset, to check for potential overfitting, and (*ii*) an independent validation dataset for the qualitative assessment of the PCE metamodel.
A high overall accuracy was achieved in both cases, with score of 0.95 when the validation dataset was used. Further, each predicted $N_\gamma$ was accessed for relative error distribution, with a few cases reaching errors up to 100%, but the majority remaining within ±30% range.
Finally, a monte-carlo (MC) sample of size $10^5$ is used to evaluate the probability density function of the capacity estimated by the PCE metamodel. The most likely estimated capacity factor peaks around 24, consistent with the trends observed for the training and validation datasets.

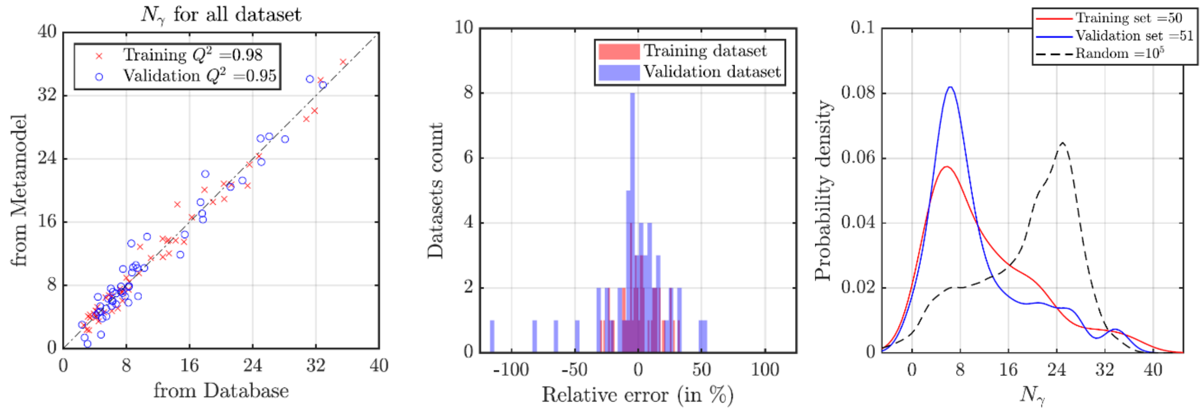


Figure 3. PCE metamodel predictions of $N_\gamma$ compared with true database estimates (left) relative error distribution (middle) and distribution of estimated $N_\gamma$ *(right)*

### *3.1 Clustering to improve PCE accuracy*

Despite the high accuracy scores obtained with the two considered datasets, cases with a large relative error limit the reliability and robustness of the PCE estimation. To address this issue, the input parameters were systematically clustered according to the anchor embedment ratio in order to separate shallow and deep failure mechanisms more effectively. Clustering based on *H*/*D* yields a clear and distinct separation of the output variable at a threshold of H/D = 2, as illustrated in Figure 4.

Two PCEs were then generated, with predictions for high *H*/*D* shown in Figure 5. Although the accuracy score remained unchanged, the relative error bounds reduce significantly, emphasizing the improvement in PCE metamodel through physics-based considerations. The overall $N_\gamma$ distribution is also reported for the random MC sample alongside the training and validation datasets. Notably, the absence of unrealistic negative $N\gamma$ further highlights the robustness of the cluster-based PCE.

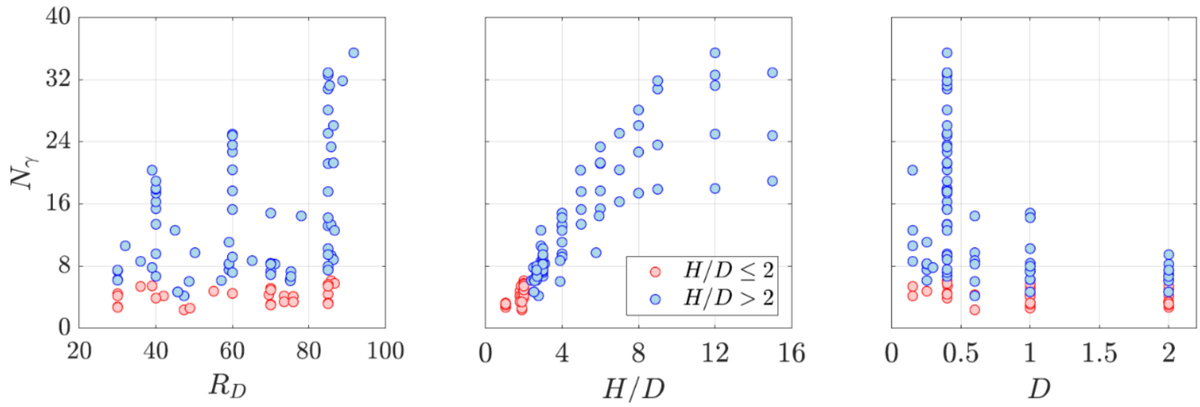

Figure 4. Clustered input space based on low and high embedment ratios

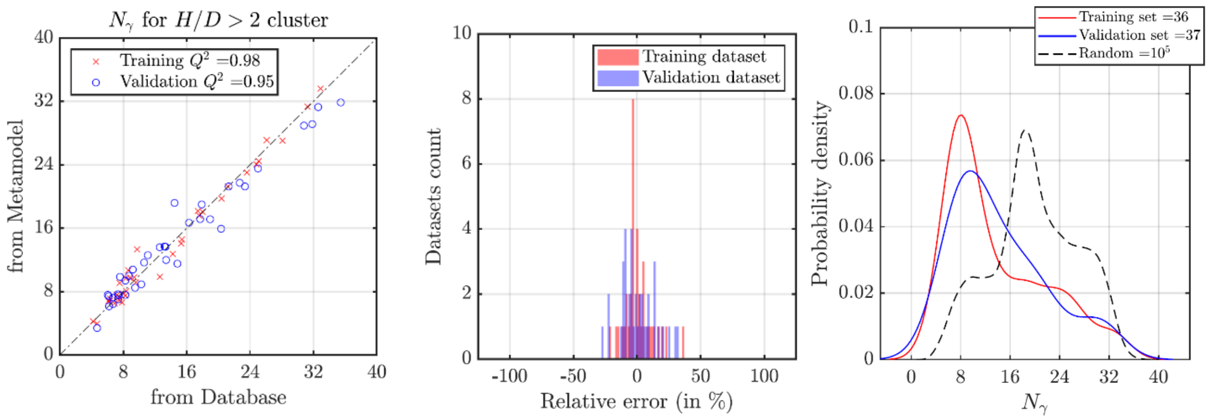

Figure 5. Cluster-based PCE metamodel (for high H/D) predictions of $N_\gamma$ compared with true database estimates (left) relative error distribution (middle) and distribution of estimated $N_\gamma$ *(right)*

## 4 Conclusions

This work presented a PCE-based metamodel for the prediction of the plate anchor uplift capacity factor, $N_\gamma$, from three input parameters describing the soil-anchor interaction mechanism. The main conclusions are as follows:

1. The PCE metamodel developed with a hyperbolic truncation strategy and a LARS algorithm as regression technique enables generalization of plate anchor capacity estimates from a limited database compiled from multiple sources, providing high accuracy score ($Q^2 > 0.95$). Most predicted values exhibit relative errors within ±30%, although isolated cases with errors exceeding 100% were observed.
2. Incorporating physics-based clustering, through separation into shallow ($H/D < 2$) and deep ($H/D > 2$) failure mechanisms, reduces the spread of relative errors. The maximum errors are limited to approximately ±30% with the majority of predictions concentrated within ±15%, while maintaining the same high global accuracy score.

The proposed PCE-based metamodel for the circular plate anchor in sand provides rapid, low-cost evaluations of uplift capacity. By reproducing the response trends embedded in costly centrifuge experiments and computationally demanding finite element simulations, the metamodel serves as an efficient surrogate.

Importantly, it is well suited for preliminary design stages, where site information is often sparse and highly uncertain. Within such contexts, the probabilistic framework enabled by PCE allows meaningful analysis of uncertain input data and supports informed early-stage design decisions.

## 5 References

Blatman G., & Sudret B. (2010). Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.*, 230: 2345-2367.

Giampa J.R., Bradshaw A.S., Schneider J.A. (2017). Influence of Dilation Angle on Drained Shallow Circular Anchor Uplift Capacity. *Int. J. Geomech.*, 17(2), 04016056.

Hao D., Wang D., O'Loughlin C., Gaudin C. (2019). Tensile monotonic capacity of helical anchors in sand: interaction between helices. *Can. Geotech. J.*, 56.

Kurniadi, R., Roy, A., Chow, S.H. & Cassidy, M.J. (2025). Drained uplift capacity of deeply embedded horizontal circular plate anchors in sand. *Submitted for publication.*

Rasulo M.P., Schneider J.A., Newgard J.T., Aubeny C.P. (2017). *Transition depths for deep circular anchors in saturated sand.* Proc. 8th Int. Conf. Offshore Site Investigation Geotechnics. 987-994.

Roy A., Chow S.H., O'Loughlin C.D., Randolph M.F. (2021a). Towards a simple and reliable method for calculating uplift capacity of plate anchors in sand. *Can. Geotech. J.*, 58(9), 1314-1333.

Roy A., Chow S.H., O'Loughlin C.D., Randolph M.F., Whyte S. (2021b). Use of a bounding surface model in predicting element tests and capacity in boundary value problems. *Can. Geotech. J.*, 58(6), 782-799.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab. Engng. Syst. Saf.* 93(7): 964–979