Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-851

Probabilistic modeling of limit pressure with depth: massive data calibration and bias control

Arnaud Finiasz, Balasubramanian Ilango
Fondasol, Nanterre, France, arnaud.finiasz@groupefondasol.com

Etienne de Rocquigny
Blaise Pascal Advisors, France, edr@pascal-advisors.com

ABSTRACT: In order to adopt a probabilistic approach for geotechnical design, a probabilistic distribution must be established for the soil strength parameters that govern the limit state. The strength parameters can be the cohesion and internal friction angle of the soil, or the results of in-situ tests where empirical design methods exist. In this paper, we will use the Ménard limit pressure, derived from pressuremeter tests, that can be used directly for foundation design. To establish the probabilistic distribution of the limit pressure, we infer multiple curves, each representing the evolution of a quantile with depth (from Q10 to Q90). The algorithm used to infer those curves is governed by various parameters: horizontal range, vertical smoothing, weighting law, quantile estimation method… In this paper, we describe the large-scale benchmarking process that was adopted to fine tune the algorithm used to infer those curves, on a nationwide sample of approximately 200 000 boreholes with pressuremeter logs.

KEYWORDS: Probabilistic modelling, quantile regression, numerical benchmarking.

1 INTRODUCTION

This paper is part of Fondasol's endeavor to integrate historical data into geotechnical design through a probabilistic approach. The premise here is that historical data, accumulated through previous site investigations, is abundant and should not be ignored in geotechnical design, be it to establish a preliminary ground model during desk study, or the final ground model for detailed design.

To provide an unbiased synthesis of the results of multiple site investigations, we have proposed (Finiasz et al. 2024) a graphical representation of the probabilistic distribution of soil strength parameters, in the form of multiple curves, each representing the evolution of a quantile with depth (from Q10 to Q90). The strength parameter we are using here is the Ménard limit pressure, derived from pressuremeter tests, that can be used directly for foundation design. This type of test is very common in France, but the methods exposed here could be applied to any discrete in-situ test, such as the SPT.

2 MONITORING USAGE OF THE ALGORITHM

Since the initial development of the algorithm, in late 2021, numerous geotechnical engineers have been using it on a regular basis within the framework of our probabilistic geotechnical suite. The general use case can be summarized as follows:

- During the initial desk study, the engineer selects a set of relevant pressuremeter boreholes from our database, within a given search radius, on a cartographic interface. This selection is based on topographical and geological considerations, and, depending on the local geological heterogeneity, we observed that the search radius can range from 250 m or less to more than 2000 m.
- Our algorithm can then infer a probabilistic distribution of the limit pressure (P_{lm}) with depth, represented by 4 curves, each corresponding to the evolution of a quantile with depth. The quantiles that are represented are the first and last decile (Q10 & Q90) and two quartiles (Q25 & Q75).
- This probabilistic distribution is used to establish a preliminary ground model. Additional site investigations are then conducted to validate, or adjust, this preliminary ground model.
- The engineer can then establish a design ground model considering both the results of the site investigation and the prior knowledge from the likely soil variability modeled through a probability distribution.

Figure 1. below illustrates the probabilistic distribution of the Ménard limit pressure, compared to the results from the site investigation for an industrial project in Le Havre (Seine Maritime, France). The 4 curves corresponding to the empirical quantiles Q10, Q25, Q75 and Q90 determine 3 surfaces:

- A central dark grey zone corresponding to the two middle quartiles, where most likely values can be found (used for example when calculating the sleeve friction contribution to a pile's bearing capacity).
- A lower bound light grey zone, between Q10 and Q25 where characteristic values are likely to be selected, say for shallow foundation design.
- And an upper bound zone (also light grey), between Q75 and Q90, for situations where higher values need to be considered, for example when assessing the drivability of a pile through a compact layer.

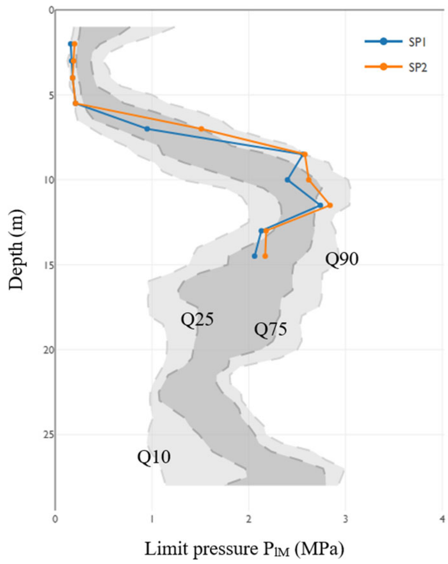

Figure 1. Probabilistic distribution of the limit pressure, compared to the results of the site investigation, Le Havre.

In this example, the probabilistic distribution was constructed using data from 45 boreholes, from previous site investigations, carried out across the industrial site and its immediate vicinity. The geological context corresponds to alluvial deposits, with alternating dense and very dense sand layers below a 6 m cover of soft sandy clay. We can see that the 2 pressuremeter boreholes carried out for the project, SP1 and SP2, fit nicely within the expected distribution.

This is one example among many, as these tools are gradually getting adopted by most of our geotechnical staff. With the growing use of data driven preliminary ground models, we have been able to observe in most cases that the subsequent site investigations are consistent with the predetermined probabilistic distribution.

The next step is obviously a consolidated probabilistic distribution including both prior knowledge and the new data acquired through the dedicated site investigation. This distribution could then serve as the basis for a fully probabilistic geotechnical design, like for example deep foundation design.

But before jumping into a full probabilistic design, we feel it is necessary to regularly challenge the setup of the algorithm through benchmarks, both to ensure the reliability of the model and to better understand the individual impact of the different parameters governing the construction of the distribution. We did finally launch a very large benchmark, at the core of the present publication.

3 LARGE SCALE BENCHMARKING

As an introduction, it should be mentioned that the initial setup of our algorithm (back in 2021) was carried out through a relatively small-scale leave-one-out process. The test samples consisted of manually selected groups of boreholes (corresponding typically to 1 or 2 site investigation reports) and the learning sample included all other boreholes in the vicinity.

This initial benchmark first compared the performance of different models (KNN, indicator kriging…) and then allowed us to adjust the weighting laws of the selected model. One of the conclusions of this initial benchmark was that the most impactful factor on the quality of the predicted model was the meaningful selection of relevant boreholes from our database. This selection is still left in the hands of the geotechnical engineer.

3.1 *The data set*

In order to challenge the initial setup of the algorithm, we decided to launch a more automated benchmark on the full database. This meant forgoing the expert selection of boreholes by the geotechnical engineer but allowed us to test thousands of samples. The idea was to check for more subtle systematic bias that could not be discerned by local tests.

At the time of testing, our database comprised 101 500 site investigations with pressuremeter boreholes and a reliable positioning across the French territory (amounting to approximately 278 000 pressuremeter boreholes). Figure 2 shows the spatial distribution of these site investigations across France. We can see that the entire territory is covered, with obviously more focus on densely populated areas, as these are the areas where most projects are located.

A random target sample of 10 600 site investigations was selected with the only prerequisite that other investigations be present around those sites, in order to allow the construction of probabilistic distributions without using the boreholes from the target sample.

3.2 *The benchmarking process*

For each of the 10 600 site investigations of the target sample, probabilistic distributions have been computed from the rest of the database, with varying parameters, and results have been recorded for each set of parameters. The predicted distributions are then compared to the pressuremeter test results of the target sample.

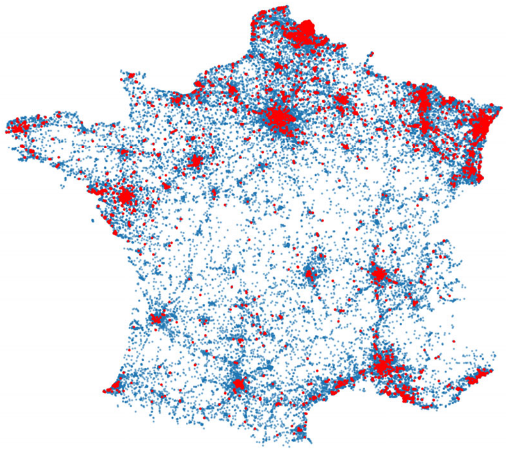

Figure 2. In blue, location of the 101 500 site investigations used for the benchmark and in red the target sample of 10 600 investigations

Two different metrics were originally considered to assess the reliability of the predicted distributions: quantile loss (Koenker & Bassett, 1978) and quantile coverage. If quantile loss is a popular function when training regression algorithms to predict certain quantiles, the resulting values were difficult to translate into a measurable error. Quantile coverage was therefore preferred, as the results were a lot easier to interpret.

Quantile coverage is simple to understand: it checks whether the predicted quantiles cover the expected proportion of the actual observations (the Ménard limit pressure values in the target sample). For each inferred quantile curve, the proportion of observations that are lower than the corresponding quantile is measured. For example, it is expected that approximately 10% of the pressuremeter tests will fall below the predicted Q10 curves.

For a given set of parameters (interpolation method, search radius, weighting laws…), the quantile curves are computed at the location of each of the 10 600 target site investigations. Then all the pressuremeter tests (approx. 190 000 units) from the pressuremeter boreholes (approx. 29 000 units) of the target sample are compared to the inferred curves.

The conformity of the predicted quantiles is expressed as the difference between the actual proportion of tests that are below the predicted quantiles, and the expected proportion. We will note ΔqX the global error for quantile X, defined as follows:

- $\Delta q10$ = actual proportion of pressuremeter tests that fall below the predicted Q10 curves – 10%
- $\Delta q25$ = actual proportion of pressuremeter tests that fall below the predicted Q25 curves – 25%
- $\Delta q75$ = actual proportion of pressuremeter tests that fall below the predicted Q75 curves – 75%
- $\Delta q90$ = actual proportion of pressuremeter tests that fall below the predicted Q90 curves – 90%

Example: a value $\Delta q10$ of +1 indicates that 11% of the tests in the target sample were below the predicted Q10, which means the Q10 is slightly over-estimated.

These global errors can then be plotted as a function of the search radius, the number of pressuremeter tests used to compute each curve, the number of boreholes, or the equivalent number of tests (including their individual weight in the formulas). For clarity, in this paper all representations of the global error will be expressed as a function of the number of pressuremeter boreholes used.

4 DETAILED RESULTS

The original set of parameters of our algorithm was first tested. It proved to be fairly accurate when predicting the Q10 and Q25 quantiles, which is important as design values for geotechnical design are usually selected on this side of the distribution, but significantly underestimated the Q75 and Q90.

This was in part due to the lower bound approach we had adopted regarding quantile interpolation, which didn't prove conservative for the higher quantiles where high values of limit pressure indicating stronger soils can be problematic for excavation works or pile driving.

However, it could also be the result of the peculiar distribution of the values of the limit pressure in the data set. The standard procedure for the Ménard pressuremeter test states that the test is stopped when the applied pressure reaches 5 MPa in the probe. Censored data thus results in a lot of values around 5 MPa, when the soil would actually be stronger. This is quite visible in Figure 3, where we have plotted the distribution of the full 1.8 million tests from the data set. We can also see that there are some high-pressure tests in the database, with values exceeding the 5 MPa threshold.

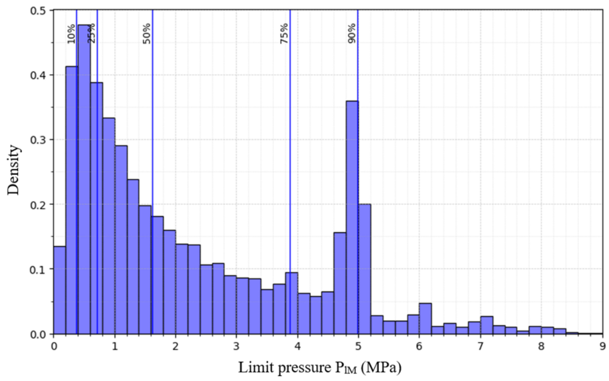

Figure 3. Distribution of the limit pressure values for the complete data set

4.1 *Quantile estimator*

To improve the reliability of our algorithm for the higher quantiles, without altering its behavior for the Q10 and Q25, we decided to test different interpolation methods.

When estimating quantiles from a set of independent observations for a random variable, there are number of available definitions in statistical computer packages and literature (Hyndman and Fan, 1996). Traditional estimators use linear interpolation between two consecutive values (within the ordered observations). We refer here to the Hyndman-Fan taxonomy for quantile estimators, as detailed in Table 1.

Table 1. The Hyndman-Fan taxonomy of quantile estimators.

Type	h	Equation
4	np	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$
5	$np + 1/2$	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$
6	$(n + 1)p$	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$
7	$(n - 1)p + 1$	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$
8	$(n + 1/3)p + 1/3$	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$
9	$(n + 1/4)p + 3/8$	$x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)\,(x_{(\lceil h\rceil)} - x_{(\lfloor h\rfloor)})$

Where:

h = calculated index for quantile p in a sample of size n
$\lfloor\cdot\rfloor$ = floor rounding
$\lceil\cdot\rceil$ = ceiling rounding

After some testing on synthetic data (which is not detailed in this paper), we selected estimators 4, 6 and 9 for the full-scale benchmark with 4 corresponding to the initial iteration of the algorithm.

We also considered adaptations of the Harrel-Davis estimator and its trimmed variant, to account for the different weight of the observations, as proposed by Akinshin (2023).

Figure 4 below shows the plots for global error as a function of the number of boreholes for the 4 predicted quantiles and the different quantile estimators (only the trimmed version of Harrell-Davis is represented).

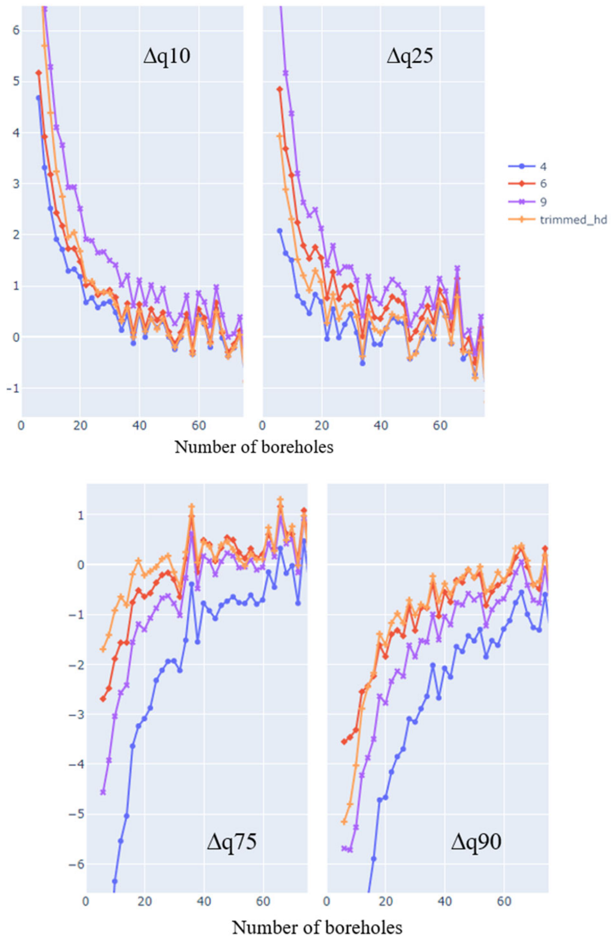

Figure 4. Global error as a function of the number of boreholes used to infer the quantile curves for 4 selected estimators

As mentioned earlier, the linear estimator 4 proved to be quite accurate when estimating the Q10, with an average error of less than 1% when using at least 20 boreholes, and as expected, even better results for the Q25 with half as many boreholes. Seeing as this is the setup that has been used by our geotechnical staff for the past couple of years, those results are quite reassuring. On the other hand, this estimator is off for the Q75 and Q90, even when many boreholes are considered.

For the higher quantiles (Q75 and Q90), the linear estimator 6 performs much better. For all estimators, the Q90 seems to be the most difficult to accurately predict, possibly due to the singularity of the distribution of the limit pressure around 5 MPa mentioned earlier.

The trimmed Harrel-Davis estimator seems to provide the best overall performance but is more complex and requires much higher calculation times (approx. 50 times more than linear estimators), which is not very practical for a production tool. A good compromise would appear to be using a linear estimator of type 4 for quantiles below the median, and switching to type 6 for quantiles above the median.

4.2 *Weighting with distance*

When monitoring the usage of the preliminary ground model by our engineers, it appeared that the selected search radius was not always directly correlated with the geological homogeneity of the studied area. In that context, we wanted to test if weighting with distance should be a set value (possibly varying regionally) or dynamically linked to the search radius.

Figure 5 below shows the average global error for 3 different functions of weight with distance: either a rapid decrease after around 500 m (in red), around 1000 m (in green) or a dynamic weight adjusted in proportion with the search radius (blue curve).

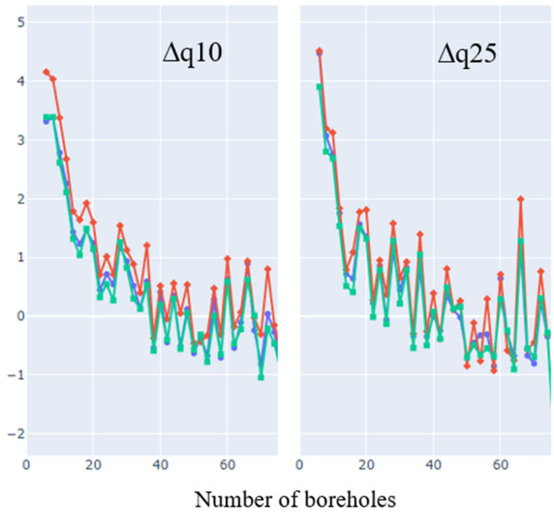

Figure 5. Global error for 3 different weighing laws

The graphs are provided for Q10 and Q25, but similar conclusions could be obtained from the higher quantiles. If we put aside the noise of the curves, there is no significant difference between the 3 options, meaning that there is no global bias towards one or another. This opens the door to regionally set weighting with distance, possibly adjusted based on geotechnical homogeneity.

4.3 *Vertical smoothing – weighting with depth*

When plotting the quantile curves with depth, one must account for the fact that pressuremeter tests are spaced every 1.0 to 1.5 m and sometimes fall in between these values for technical reasons onsite. Similarly to what is done with horizontal distance, decreasing weight with vertical distance (depth) is considered around the depth where the calculations are made.

This results in some smoothing of the curves, which makes nicer graphs, but tends to blur the transition between layers of contrasted strength.

Here we tested 3 values of vertical smoothing, from 0.6 which is pretty much the minimum that we can use considering the spacing between the pressuremeter tests, to 1.0 which gives much smoother graphs, but blurs the limit between different soil layers.

As can be seen on Figure 6, there is not much difference between the 3 values of smoothing, indicating that smoothing is not detrimental to the overall reliability of the predicted quantile curves.

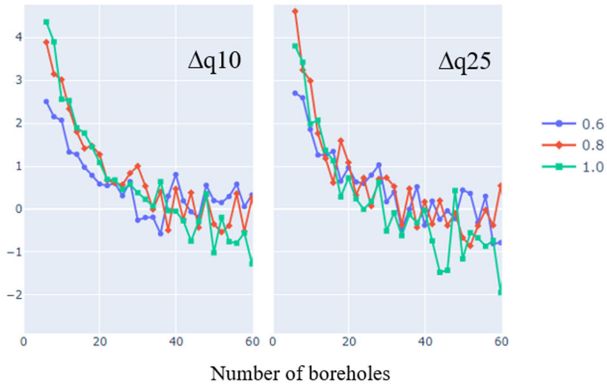

Figure 6. Global error as a function of the number of boreholes for different values of vertical smoothing.

4.4 *Measuring uncertainty*

As expected, all the graphs show that when estimating quantiles the average global error diminishes when the volume of data used to infer the curves increases.

Beyond this obvious observation, the results from the global benchmark will allow us to provide our geotechnical staff with quantitative indications on the reliability of the probabilistic distribution that they are using, based on the amount of available data around the project they are studying.

5 CONCLUSIONS

As exposed in this paper, a large-scale benchmark allowed us to better identify the parameters that influenced the reliability of our probabilistic distributions. We found that mixing two different interpolation methods allowed us to mostly remove any systematic bias. We also found that a lot of parameters could be adjusted within a wide range of values without adversely affecting the reliability of the distributions, opening the way to locally adjusted parameters. This will obviously need further testing, with regional data subsets, to account for the wide variety of geological contexts in France, keeping in mind that bias figures have been considered in global means, leaving open the question of comparative site-specific deviations.

Plotting the mean relative error versus the number of points used to infer the quantile curves also allows us to give an estimate of the increased uncertainties when using smaller samples. On average, we see that between 15 and 25 relevant borehole logs favor a very reliable probabilistic distribution, while again can depend a lot on local soil variability and should always rely on strong human expertise.

This presents a solid basis for future applications that will directly use these distributions of soil strength with depth in a probabilistic approach, for example to design deep foundations using the empirical methods available for the Ménard pressuremeter.

6 REFERENCES

Akinshin, A. 2023. Weighted quantile estimators. http://dx.doi.org/10.48550/arXiv.2304.07265

Finiasz, A., Hassaoui, J., Jacquard, C. & Rocquigny, E. 2024. Integrating local data into geotechnical design through a predictive probabilistic approach. *Proc. 18th European Conference on Soil Mechanics and Geotechnical Engineering*, Lisbon, 474-477. http://dx.doi.org/10.1201/9781003431749-65

Hyndman, R. J. & Fan, Y. 1996. Sample quantiles in statistical packages. *The American Statistician*, Vol. 50, 361-365. https://doi.org/10.2307/2684934

Koenker, R. & Bassett, G. 1978. Regression Quantiles. *Econometrica*, Vol. 46, 33–50. https://doi.org/10.2307/1913643