

Statistical supervised learning with engineering data: A case study of low frequency noise measured on semiconductor device

Revision for *The International Journal of Advanced Manufacturing Technology*

January 2022

Dear Editor,

Please find attached the revision of our paper, “Statistical supervised learning with engineering data: A case study of low frequency noise measured on semiconductor device” by M.L. Gámiz, A. Kalén, R. Nozal-Canãdas and R. Raya-Miranda, following your suggestion and the comments from the reviewers.

We hope you will find suitable our revision.

Yours sincerely,

M.L. Gámiz

Editor

Based on the advice received, I feel your manuscript could be accepted for publication should you be prepared to incorporate minor revisions. When preparing your revised manuscript, you are asked to carefully consider the reviewer comments which are attached, and submit a list of responses to the comments. Your list of responses should be uploaded as a file in addition to your revised manuscript.

Response:

Thank you for reviewing our paper and the valuable feedback from the Reviewers. We have taken into account all the suggestions and have answered appropriately to their comments and requests and we hope that the revised version of the paper is satisfactory.

Reviewer 1

This paper studies the relationship between noise and voltage from the perspective of statistical learning, and finally establishes a model that can quantify the influence of threshold voltage on the noise signal. There are few researches in this direction at present, and it has research value. This research is of great help to the machine learning technology in the field of electronic engineering.

Response. Thank you for your valuable report.

Reviewer 2:

After the revision, it seems to me that the contribution is sufficient. The authors propose a new approach for the analysis of potential correlations between spectral noise current and threshold voltage from common on-wafer MOSFETs. The manuscript entitled "Statistical supervised learning with engineering data: A case study of low frequency noise measured on semiconductor devices" has been investigated in detail. The study seems very valuable. The topic addressed in the manuscript is potentially interesting and the manuscript contains some practical meanings, however, there are some issues which should be addressed by the authors:

Response

Thank you very much for your comments regarding our paper. We have prepared a revised version of the paper following your comments and suggestions, below we list your comments in italics, and our responses in ordinary text.

There are some revisions need for the article to be eligible for possible publication.

1. The paper is in need of substantial editing for grammatical errors especially in: a. In page 7 the paragraph "suggested by figure ??", and the same for the "left and side of the figure ..." Number of figure is needed. b. In figure 3 the y axis represent what ? the same for the figure 6,7,8,9,10,12,13.

Response

We have added the number of the figure and additional information about the y axis in the different figures.

2. In section 1, The justification of the proposed method is needed based on model was designed based on statistical supervised learning?

Response

The following paragraph has been added in Section 1 (Introduction) to justify the adequacy of our methodology both as a statistical tool and a supervised learning algorithm.

In this paper we use modern statistical tools that works under very weak assumptions, mainly smoothing techniques, the backfitting algorithm and the bootstrap. Kernel smoothing is one of the most popular statistical tools for nonparametric regression. Basically the method predicts the output to be the weighted average of the inputs of all training subjects. The backfitting algorithm is widely used to approximate the additive components in multiple regression problems such as the one formulated in this paper. It has longly proven very good performance in practical applications and despite its iterative nature, which makes it more difficult to derive theoretical results, consistency and asymptotic properties have been obtained under weak conditions (see for example [14]). We use this algorithm as a efficient tool to solve our regression problem. In this sense, our method can be considered a supervised learning algorithm in the usual classification of methods in machine learning,

according to which a supervised algorithm, broadly speaking, involves building a statistical model for predicting, or estimating, an output based on one or more inputs (see [13]).

[13] James, J., Witten, D., Hastie, H., and Tibshirani, R. (2021), *An Introduction to Statistical Learning with Applications in R*, Springer.

[14] Mammen, E., Linton, O. and Nielsen, J.P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, *Annals of Statistics*, **27** (5), 1443–1490.

And what are some of the advantages and disadvantages of bootstrapping?

Response

The following paragraph has been added in the Introduction of the paper:

The bootstrap is a widely applicable and extremely powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method. As a simple example, the bootstrap can be used to estimate the standard errors of the coefficients from a linear regression fit. The power of the bootstrap lies in the fact that it can be easily applied to a wide range of statistical learning methods, including some for which a measure of variability is otherwise difficult to obtain as it results in our case.

Also we would like to emphasize that Sizer Map consists of making inferences about the derivative of the function $m(x)$, using asymptotic confidence intervals. The bootstrap methodology is used to avoid calculating the standard error from an explicit expression. What is done instead in this paper is to generate B bootstrap samples from the original one and calculate their standard error.

3. The manuscript not contain a flowchart or graphical representation of a process of optimization. This pictorial representation can give step-by-step solution of the given problem, and What is the convergence criteria for program terminated?

Response

Following your suggestion, we have added a flowchart at the end of Section 3.2 of the manuscript aimed to represent the steps of the algorithm in sequential order.

The convergence criterion is given in step 4 of algorithm 1 (page 13), specifically it is written:

Step 4. Convergence.

Repeat Steps 2-3 until convergence, that is, until the difference between two consecutive

estimations is small enough. More specifically, stop at the r -th iteration when

$$\frac{\sum_{i=1}^n \left(\hat{m}^{(r)}(x_i) - \hat{m}^{(r-1)}(x_i) \right)^2}{\sum_{i=1}^n \hat{m}^{(r-1)}(x_i)^2 + 10^{-4}} < 10^{-4}$$

4. How to avoid the transit time noise or partition noise for the proposed approach ?

Response

To clarify this point we have added the following paragraph to the Conclusions:

The noise data analyzed in the manuscript correspond to MOSFET transistors at low frequencies. For these devices and in this frequency range, the main noise source is flicker noise or 1/f noise. For higher frequencies, thermal noise could also be important [1]. On the contrary, transit time noise and partition noise are usually noise sources more important for other types of devices and at higher frequencies. We expect that the contribution of transit time noise or partition noise to the data used in the present study is totally negligible [10].

[10] Gray, P. R., Hurst, P. J., Lewis, S. H., and Meyer, R. G. (2009). *Analysis and design of analog integrated circuits*. John Wiley & Sons.

5. In section 4; How to set the parameters of proposed model for better performance? The complexity of the proposed model and the model parameter uncertainty are not mentioned.

Response

We do not fully understand the first question in this comment of the reviewer.

With respect to the second question, we can understand model complexity as the number of parameters and/or features involved. Our model is formulated considering a parametric term affecting the factor (*Frequency*) and a non-parametric term for the covariate (*Voltage*). The result is a model simple enough to be easily interpretable and flexible enough to fit the data reasonably as can be deduced from the simulations presented in the paper. The complexity of the non-parametric part of the model is controlled by the so called bandwidth parameter. The smaller the bandwidth the greater the complexity of the model while bigger values of the bandwidth lead to smoother estimates which might miss important features in the data. Our approach based on SiZer methodology allows to circumvent the bandwidth selection problem by a scale-space analysis of the model that summarizes in a unique plot (a color map provided by SiZer methodology) all the relevant information contained in the data

Finally, to evaluate parameter uncertainty, we develop in Section 5 a testing procedure that involves the quantification of the estimation error among other things.

6. In section 5, the performances of the proposed method should be better analyzed, commented and visualized in the experimental section.

Response

The numerical results section has been reorganized to ease the reading of the paper and make it more understandable the proposed method. We have included additional comments about the graphs.

7. Conclusion can be more specific listing the efficiency when compared with other methods and the future scope of the work implemented on this paper.

Response

We can not compare the efficiency with other methods because, to our knowledge, there are no other researches in this direction at present. We believe that the insights in this paper can eventually help to improve the production process and further the reliability of MOSFET transistors.