Application of Noise to Avoid Overfitting in TCAD Augmented Machine Learning

Sophia Susan Raju
Electrical Engineering
San Jose State University
San Jose, USA
sophiasusan.raju@sjsu.edu

Boyan Wang
Center for Power Electronics Systems
Virginia Polytechnic Institute and State University
Blacksburg, VA, U.S.A
wangboyan@vt.edu

Kashyap Mehta
Electrical Engineering
San Jose State University
San Jose, USA
kashyap.mehta@sjsu.edu

Ming Xiao
Center for Power Electronics Systems
Virginia Polytechnic Institute and State University
Blacksburg, VA, U.S.A
mxiao@vt.edu

Yuhao Zhang
Center for Power Electronics Systems
Virginia Polytechnic Institute and State University
Blacksburg, VA, U.S.A
yhzhang@vt.edu

Hiu-Yung Wong*
Electrical Engineering
San Jose State University
San Jose, CA, U.S.A
hiuyung.wong@sjsu.edu

***Abstract*— In this paper, we propose and study the use of noise to avoid the overfitting issue in Technology Computer-Aided Design-augmented machine learning (TCAD-ML). TCAD-ML uses TCAD to generate sufficient data to train ML models for defect detection and reverse engineering by taking electrical characteristics such as Current-Voltage, *IV*, and Capacitance-Voltage, *CV*, curves as inputs. For example, the model can be used to deduce the epitaxial thicknesses of a *p-i-n* diode or the ambient temperature of a Schottky diode being measured, based on a given *IV* curve. The models developed by TCAD-ML usually have overfitting issues when it is applied to experimental *IV* curves or *IV* curves generated with different TCAD setup. To avoid this issue, white Gaussian noise is added to the TCAD generated curves before ML. We show that by choosing the noise level properly, overfitting can be avoided. This is demonstrated successfully by using the TCAD-ML model to predict 1) the epitaxial thicknesses of a set of TCAD silicon diode *IV*'s generated with different settings (extra doping variations) than the settings in the training data and 2) the ambient temperature of experimental *IV*'s of Ga_2O_3 Schottky diode. Moreover, domain expertise is not required in the ML process.**

Keywords—Gallium Oxide, Machine Learning, Noise, Schottky Barrier Diode, TCAD

I. Introduction

TCAD augmented machine learning (TCAD-ML) uses TCAD to generate sufficient data to train the ML model for various purposes [1][2]. For example, the model can be used to deduce the physical parameters, such as the epitaxial thicknesses or the ambient temperature of a diode, based on a given *IV* curve. However, models developed by TCAD-ML are usually overfitted and cannot be used to predict the physical parameters of TCAD *IV* curves generated with different settings [3] or experimental *IV* curves [4]. This is because the machine cannot understand that the variation in the *IV* curves can be caused by factors it has not seen in the training.

It is expected that, with domain expertise (i.e. extracting V_{bi}, I_{on}, I_{off}, subthreshold swing from the IV curves before applying ML[5][6]), overfitting problem may be reduced. However, this tremendously reduces the benefit of using ML as domain expertise is required to pre-process the data before ML. Thus, the algorithms developed and the experience garnered cannot be transferred directly to other problems. The overfitting issue can also be understood as the fact that the machine trained by TCAD *IV* is "over-confident" and tries to attribute any variations in the *IV* only to the parameter variations it learned in the training dataset. As a result, when there are new variation sources (such as *IV* generated with different settings in TCAD or experimental data with noise and other unintentional variations), the machine fails to predict the correct physical parameters. To avoid this problem, Principal Components Analysis and Autoencoder are used in [4] and [3], respectively, while domain expertise is not required.

In this paper, we further propose to apply noise to the *IV* generated by TCAD to avoid overfitting without using domain expertise. By adding noise to the TCAD generated IV curves, the machine trained will understand that variations in *IV* curves can be caused by some hidden variations and will not be over-fitted. This is demonstrated successfully by using the TCAD-ML model to predict 1) the epitaxial thicknesses of the given TCAD *IV*'s of Silicon diode and 2) the ambient temperature of experimental *IV*'s of Ga_2O_3 Schottky diode in which a better performance in predicting experimental temperature than PCA in [4] is achieved.

II. Prediction of Si *p-i-n* Diode Epitaxial Thicknesses

Three sets of Si *p-i-n* diode *IV*'s were generated using TCAD simulations. Sentaurus Process [7] simulation is used to create the structures and Sentaurus Device [8] is used for electrical simulations. Essential physical models are turned on, including Fermi-Dirac statistics, doping dependent, and high field saturation models for carrier mobilities, Schottky-Reed-Hall Recombination (SRH), and non-local Band to Band tunneling (BTBT). 80-bit ExtendedPrecision is used to avoid noisy reversed curves. Poisson, electron, and hole continuity equations are solved self-consistently.

Set1 is for ML model training and has ~2000 1-D *p-i-n* diodes with layer thicknesses being varied independently and uniformly but with fixed doping concentrations (Table I). The ML model is trained to deduce the intrinsic layer (t_i) and total n- and p-layer thicknesses ($t_p + t_n$) based on any given *IV* curve. *Set2* and *Set3* have constant thicknesses but the doping concentrations are varied randomly. They are used for testing the ability of the *Set1* trained ML model of deducing the

*Corresponding Author: hiuyung.wong@sjsu.edu

TABLE I. *P-I-N* DIODES SIMULATION SETUP FOR SET1, SET2 AND SET3.

Data	Thicknesses (nm)			Doping (cm^{-3})		
	n	i	p	n	i	p
Set1	150 -250	5 -16	150 -250	10^{20}	10^{17}	10^{20}
Set2	200	10	200	5×10^{18} - 5×10^{20}	1×10^{13} - 1×10^{17}	5×10^{18} - 5×10^{20}
Set3	350	15	250	5×10^{18} - 5×10^{20}	1×10^{13} - 1×10^{17}	5×10^{18} - 5×10^{20}

thicknesses when unseen variations (in this case, doping concentrations) are introduced. This is to emulate the experimental data which typically contains multiple variation sources. Note that *Set2* has thicknesses equal to the means of *Set1*. If a machine trained by *Set1* can predict the thicknesses in *Set2* well, there is a possibility that the machine just uses the means it learned in *Set1* instead of correctly predicting the thicknesses in *Set2*. Therefore, *Set3* which has thicknesses different from the means in *Set1* is created to confirm the machine's prediction ability.

80% of *Set1* data is used for training (using linear regression) and 20% is used for validation to develop *Model1*. Fig. 1 shows that it deduces the thicknesses of the validation data very well. However, when this model is used to deduce the thicknesses of the *IV*'s from *Set2* (with t_i = 10nm and $t_n + t_p$ = 400nm), it produces spurious results (e.g. negative thicknesses or t_i > 2000nm). This is because of overfitting as the machine "thinks" that all variations in the *IV* curves are caused by the thickness variations alone. To explain the *IV* variation caused by doping in *Set2*, it deduces the spurious results. To avoid overfitting, we propose to add noise to the *IV* curves in *Set1* for training, so that the machine will be able to isolate the features that are affected by the thicknesses by "understanding" that there are variations (due to noise) that cannot be explained by using thicknesses alone. White Gaussian noise is thus added to the *IV* curves in *Set1* with a certain signal-to-noise ratio (SNR) in dB. Fig. 2 shows an example of the final *IV* after noises are added. When SNR is larger, the noise signal is smaller.

Fig. 3 shows that with noise (SNR = 12.5dB, *Model2*), the validation result in *Set1* becomes worse in the prediction of $t_n + t_p$ (the machine is less confident or less overfitted). However, spurious results in predicting the thicknesses in *Set2* and *Set3* can be completely avoided. The prediction of *Set2*

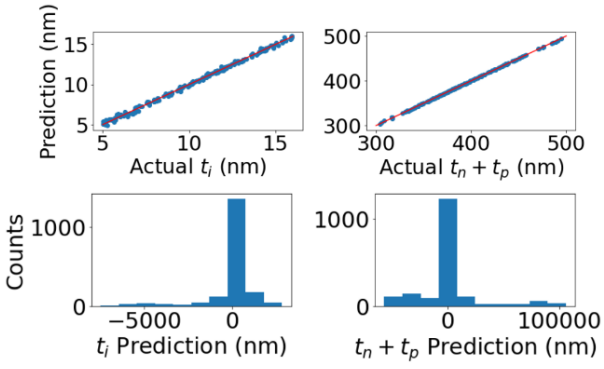

Fig. 1: *Model1* without noise. Top: validation result in *Set1* (scatter plots). Bottom: Prediction of *Set2* thicknesses (histograms).

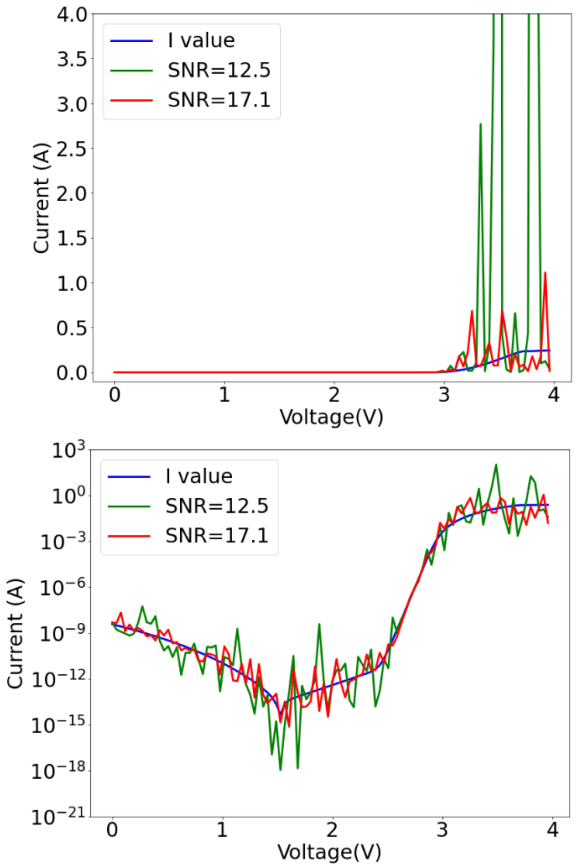

Fig. 2: *IV* curve from Dataset1 with (red and green) and without (blue) noise. Top: Current in linear scale. Bottom: Current in logarithmic scale.

and *Set3* t_i is fairly good that the prediction average is only about 2nm different from the actual values. However, the prediction of $t_n + t_p$ is not as good as desired. Prediction of *Set2* and *Set3* are 107nm and 236nm of difference from the actual values, respectively.

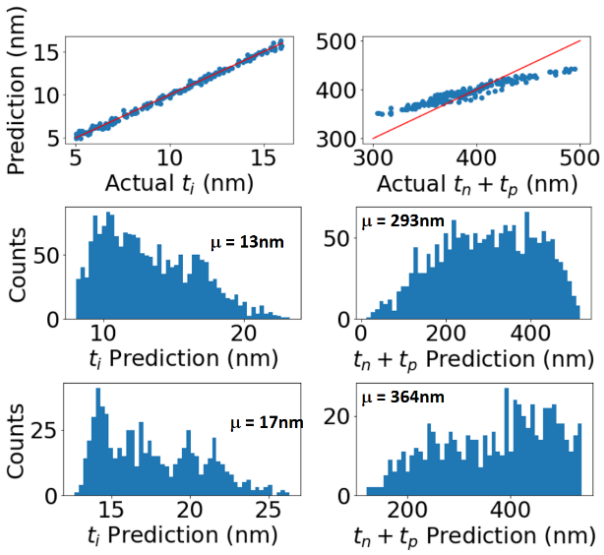

Fig. 3: *Model1* with SNR = 12.5dB performance. Top: validation result in *Set1* (scatter plots). Middle: *Set2* prediction histograms. Bottom: *Set3* prediction histograms

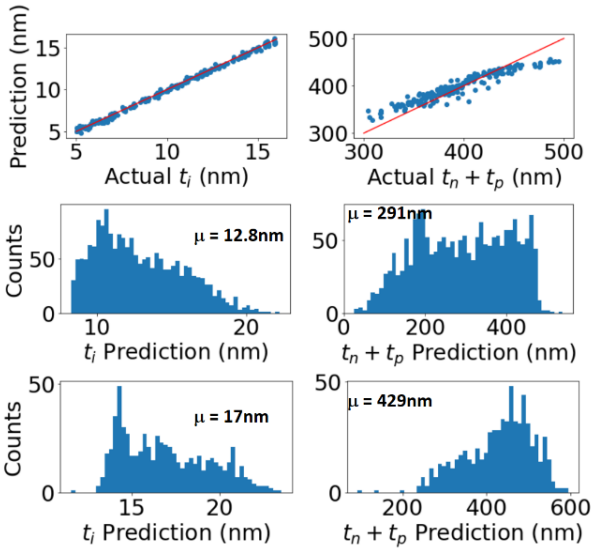

Fig. 4: *Model3* with SNR = 17.1 dB performance. Top: validation result in *Set1* (scatter plots). Middle: *Set2* prediction histograms. Bottom: *Set3* prediction histograms.

To achieve a better prediction, smaller noise (SNR = 17.1dB) is applied but with linear regression of 3rd order polynomial (*Model3*). Fig. 4 shows that it gives better validation results in *Set1* than *Model2* (therefore, it is more confident). It gives similar results in t_i prediction as *Model2*. But it improves the prediction of *Set3* $t_n + t_p$ by reducing the error to 171nm. Therefore, it is important to choose the noise level carefully to achieve the best performance. More importantly, the prediction means of t_i and $t_n + t_p$ changes by about 5nm and 140nm, respectively, from *Set2* and *Set3*, which is very similar to the actual changes (5nm for t_i and 200nm for $t_n + t_p$). This confirms that the trained machine is able to predict the thicknesses well and is not just using *Set1*'s means.

III. Prediction of Ga_2O_3 Diode Ambient Temperature

Ga_2O_3 is an emerging ultra-wide bandgap semiconductor and have been attracting more attention recently [9]-[11]. Potentially, it can also be used as a high-temperature sensor. Fig. 5 shows the structure. The drift region of the Schottky diode has large variations in drift layer thickness (t_D) and doping concentration (N_D) due to the immature technology of Ga_2O_3. Moreover, special chemical treatment was applied before anode metal deposition to achieve different effective workfunction (*WF*). The experimental *IV*'s are measured at 3

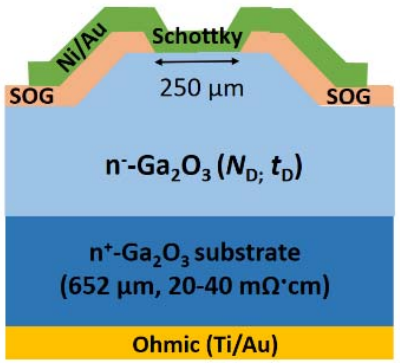

Fig. 5: Structure of the Ga_2O_3 Schottky Diode fabricated.

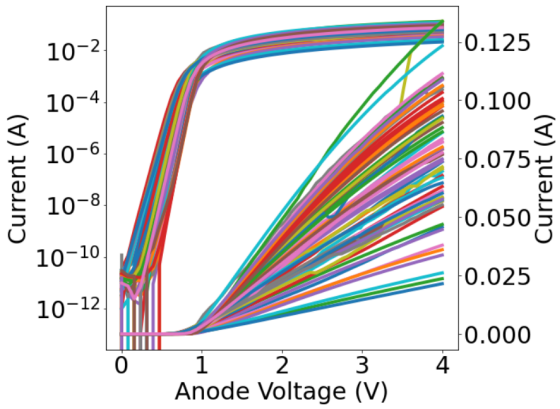

Fig. 6: 67 experimental *IV's* of Ga_2O_3 Schottky Diodes with variations in N_D, t_D, and effective workfunction measured at 3 different ambient temperatures. Left: log scale. Right: linear scale

different ambient temperatures. Fig. 6 shows 67 measured *IV*'s. Note that the experimental *IV*'s have a lot of non-idealities including the measurement noise and contact resistance variations.

TCAD models and parameters are calibrated to the experimental Ga_2O_3 Schottky diodes in [4]. Essential models such as dopants incomplete ionization and Philips Unified Mobility Model (PhuMob) [12] are calibrated to the literature [13]. A few normal experimental *IV's* are then used to calibrate TCAD *IV*'s from 300K to 510K. Self-heating simulation is turned on and solved self-consistently with Poisson equation and electron and hole continuity equations. The calibration details can be found in [4].

Afterward, 10,000 *IV*'s are generated in TCAD with variations in ambient temperature, *WF*, t_D, and N_D using a server with 2 Fourteen-Core Intel Xeon Processor E5-2690 v4 2.60GHz with hyperthreading. A machine is trained using similar settings in *Model3* to predict these variations with the TCAD *IV*'s as inputs. The machine is then used to deduce the ambient temperature of experimental *IV*'s.

As showed in Fig. 7, the model deduces spurious ambient temperature on the experimental *IV*'s and the predicted

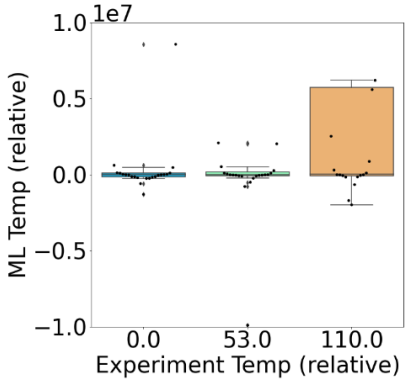

Fig. 7: Prediction of measurement ambient temperature using the machine trained with TCAD *IV*'s (*no noise)*.

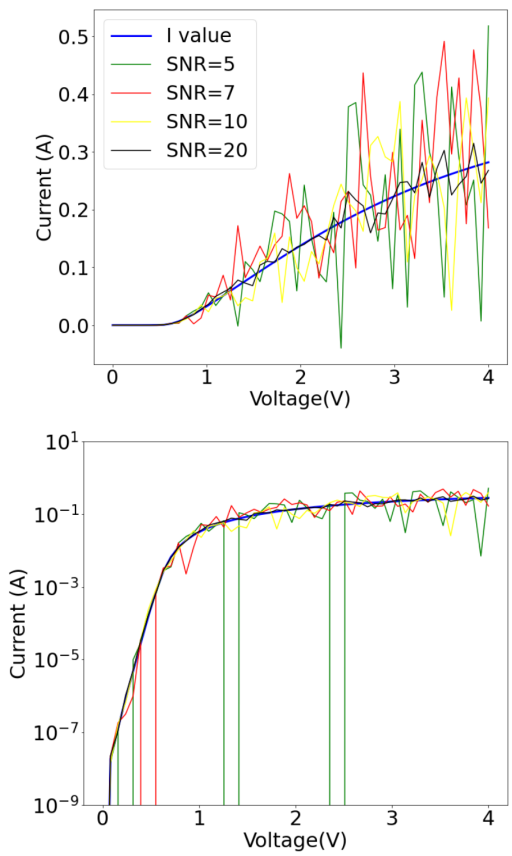

Fig. 8: TCAD Ga_2O_3 Schottky Barrier Diode *IV* curves with and without noise. Top: Current in linear scale. Bottom: Current in logarithmic scale.

temperatures can reach $\pm 10^6$K when noise is not added to the TCAD *IV*'s.

Therefore, noise (SNR = 7dB) is added to the TCAD *IV* curves before ML as shown in Fig. 8. With the noise and using *Model3*, the machine can predict the ambient temperature variation range very well (Fig. 9), despite the non-idealities in the experimental *IV*'s and the presence of unseen variations. The result is also better than the range predicted by using PCA in [4].

CONCLUSION

Noise addition to TCAD *IV*'s is proposed to avoid overfitting in TCAD-augmented machine learning. This is demonstrated successfully in two cases. One is to deduce the layer thicknesses of Si *p-i-n* diode *IV*'s generated by TCAD with known thicknesses and extra variation in doping. It is statistically shown that noise addition can minimize overfitting. The second case is to deduce the ambient temperature based on experimental Ga_2O_3 Schottky diode *IV*'s. This demonstrates the practical application of TCAD-augmented ML without domain expertise by adding white Gaussian noise.

ACKNOWLEDGMENT

This project was supported by San Jose State University College of Engineering New Faculty Start-up Fund. Hiu Yung Wong thanks Synopsys, Inc. for the donation of TCAD licenses.

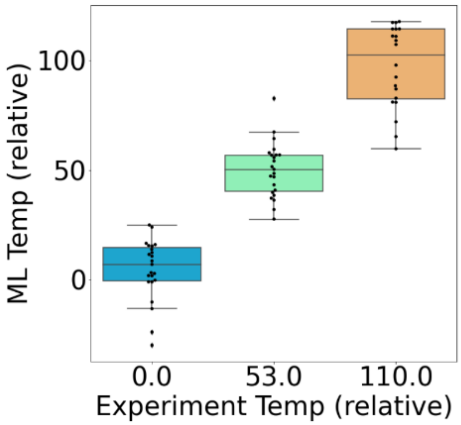

Fig. 9: Prediction of measurement ambient temperature using the machine trained with TCAD *IV*'s (*SNR = 7dB)*.

REFERENCES

[1] Y. S. Bankapalli and H. Y. Wong, "TCAD Augmented Machine Learning for Semiconductor Device Failure Troubleshooting and Reverse Engineering," 2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Udine, Italy, 2019, pp. 21-24, doi: 10.1109/SISPAD.2019.8870467.

[2] C. Teo, K. L. Low, V. Narang and A. V. Thean, "TCAD-Enabled Machine Learning Defect Prediction to Accelerate Advanced Semiconductor Device Failure Analysis," 2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Udine, Italy, 2019, pp. 17-20, doi: 10.1109/SISPAD.2019.8870440.

[3] K. Mehta, S. S. Raju, M. Xiao, B. Wang, Y. Zhang and H. Y. Wong, "Improvement of TCAD Augmented Machine Learning Using Autoencoder for Semiconductor Variation Identification and Inverse Design," in IEEE Access, vol. 8, pp. 143519-143529, 2020, doi: 10.1109/ACCESS.2020.3014470.

[4] Hiu Yung Wong, Ming Xiao, Boyan Wang, Yan Ka Chiu, Xiaodong Yan, Jiahui Ma, Kohei Sasaki, Han Wang, and Yuhao Zhang, "TCAD-Machine Learning Framework for Device Variation and Operating Temperature Analysis with Experimental Demonstration", submitted to IEEE Journal of Electron Devices Society (J-EDS).

[5] H. Carrillo-Nuñez, N. Dimitrova, A. Asenov and V. Georgiev, "Machine Learning Approach for Predicting the Effect of Statistical Variability in Si Junctionless Nanowire Transistors," in IEEE Electron Device Letters, vol. 40, no. 9, pp. 1366-1369, Sept. 2019, doi: 10.1109/LED.2019.2931839.

[6] J. Chen et al., "Powernet: SOI Lateral Power Device Breakdown Prediction With Deep Neural Networks," in IEEE Access, vol. 8, pp. 25372-25382, 2020, doi: 10.1109/ACCESS.2020.2970966.

[7] Sentaurus™ Process User Guide Version O-2018.06, June 2018.

[8] Sentaurus™ Device User Guide Version O-2018.06, June 2018.

[9] M. Higashiwaki, and G. H. Jessen, "Guest Editorial: The dawn of gallium oxide microelectronics," *Appl. Phys. Lett.*, vol. 112, no. 6, p. 060401, Feb 2018. doi: 10.1063/1.5017845.

[10] S. J. Pearton, Jiancheng Yang, Patrick H. Cary IV, F. Ren, Jihyun Kim, Marko J. Tadjer, and Michael A. Mastro, "A review of Ga_2O_3 materials, processing, and devices," *Appl. Phys. Rev.*, vol. 5, no. 1, p. 011301, Jan. 2018. doi: 10.1063/1.5006941.

[11] B. Wang, M. Xiao, X. Yan, H. Y. Wong, J. Ma, K. Sasaki, H. Wang, and Y. Zhang, "High-voltage vertical Ga_2O_3 power rectifiers operational at high temperatures up to 600K," *Appl. Phys. Lett.* 115, 263503 (2019); https://doi.org/10.1063/1.5132818.

[12] D. B. M. Klaassen, "A unified mobility model for device simulation—I. Model equations and concentration dependence," *Solid-State Electronics*, Vol. 35, No. 7, pp.953-959, 1992. doi: 10.1016/0038-1101(92)90325-7.

[13] K. Goto, K. Konishi, H. Murakami, Y. Kumagai, B. Monemar, M. Higashiwaki, A. Kuramata, and S. Yamakoshi, "Halide vapor phase epitaxy of Si doped β-Ga2O3 and its electrical properties," *Thin Solid Films* 666 (2018) 182–184. doi: 10.1016/j.tsf.2018.09.006.