Journal of Modern Applied Statistical Methods

Volume 19 | Issue 1 Article 9

6-8-2021

On the Level of Precision of a Heterogeneous Transfer Function in a Statistical Neural Network Model

Christopher Godwin Udomboso
University of Ibadan, cg.udomboso@gmail.com

Recommended Citation
Udomboso, Christopher Godwin (2021) "On the Level of Precision of a Heterogeneous Transfer Function in a Statistical Neural Network Model," *Journal of Modern Applied Statistical Methods*: Vol. 19 : Iss. 1 , Article 9.
DOI: 10.22237/jmasm/1608553560

On the Level of Precision of a Heterogeneous Transfer Function in a Statistical Neural Network Model

Cover Page Footnote

Appreciations goes to Prof. Godwin Nwazu Amahia (Department of Statistics, University of Ibadan, Ibadan, Nigeria), Prof. Isaac Kwame Dontwi (Department of Mathematical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana), and Dr Monday Edeki (Department of Mathematics, Covenant University, Ota, Nigeria), for their critical input into this study.

Journal of Modern Applied Statistical Methods
May 2020, Vol. 19, No. 1, eP2984.
doi: 10.22237/jmasm/1608553560

ISSN 1538 − 9472

On the Level of Precision of a Heterogeneous Transfer Function in a Statistical Neural Network Model

Christopher Godwin Udomboso
University of Ibadan
Ibadan, Nigeria

A heterogeneous function of the statistical neural network is presented from two transfer functions: symmetric saturated linear and hyperbolic tangent sigmoid. The precision of the derived heterogeneous model over their respective homogeneous forms are established, both at increased sample sizes hidden neurons. Results further show the sensitivity of the heterogeneous model to increase in hidden neurons.

Keywords: Neural network, transfer function, hidden neuron, precision

Introduction

The Multi-Layer Perceptron (MLP) is the most commonly used type of ANN. The reason for this stems from the fact that the MLP model is concentrated in the weights and the Transfer Functions (TFs) of its neurons. The TFs used in MLP networks are sometimes complex and can approximate complex problems in a fair number of neurons and layers, but are also not easily interpretable. Most neurons in an MLP network use the same TFs as the sigmoids and hyperbolic tangent, which also limit the model flexibility and can lead to large error. This has been observed in user reluctance to accept the model or even a complete rejection of modeling results. These observed limitations have been due to the fact that they are homogeneous functions. Thus, it is highly desirable to make neural network models more comprehensive, and to automatically determine the appropriate complexity of the model to avoid large error.

Udomboso (2013) reported on network analysis using homogeneous transfer functions in empirical studies. Tayfur (2002), Gan et al. (2005), Adepoju et al.

doi: 10.22237/jmasm/1608553560 | Accepted: July 10, 2019; Published: June 8, 2021.
Correspondence: Christopher Godwin Udomboso, cg.udomboso@gmail.com

(2007), Toprak and Cigizoglu (2008), Omole et al. (2009), Adewole et al. (2011), Ibeh et al. (2012), and Ashigwuike (2012) used the sigmoid transfer function, and Akinwale et al (2009) compared logistic and hyperbolic tangent transfer functions, Adeyiga et al. (2011) used the tangential transfer function (that is, family of tangents functions), and Udomboso and Amahia (2011), as well as Falode and Udomboso (2016) used the symmetric saturated linear transfer function in modeling rainfall as well as gas production, utilization and flaring respectively. Udomboso and Saliu (2016) used a bootstrap approach to build inference for the Statistical Neural Network with application to the Naira-Dollar exchange rate efficiency.

The use of heterogeneous functions in one network may give better results. Resop (2006) suggested that studies may be done on improving transfer functions in order to improve network models. Hence, if the limitation of homogeneous TFs in an MLP network are removed, and use instead a combination of some transfer functions with various complexities within the same network, the knowledge extraction algorithm could become minimal, which has the potential of becoming more comprehensible than homogeneous TFs. One important goal is to maintain the level of precision of the model as compared to existing knowledge extraction methods from neural networks which generally compromise the level of precision for higher comprehensibility. A heterogeneous TF aims to improve the complexity fitting and comprehensibility of the most popular type of MLP – the homogeneous TF feed-forward network (FFN). Therefore, this work is aimed at creating a heterogeneous neural network that is comprehensible, capable of modeling a wide range of problems, and at least comparable to current MLP in terms of precision and generalization, as a follow-up to Udomboso (2013) that started considering heterogeneous functions involving the convolution of linear functions as well as functions from the exponential family.

Materials and Methods

The form of the homogeneous model of the statistical neural network used is due to Anders (1996), and is given by

$$y = \mathrm{f}\left(\mathbf{X}, \mathbf{w}\right) + u, \tag{1}$$

where y is the dependent variable; $\mathbf{X} = (x_0 \equiv 1, x_1, \ldots, x_I)$ is a vector of independent variables; $\mathbf{w} = (\alpha, \beta, \gamma)$ is the network weight: α is the weight of the input unit, β is the weight of the hidden unit, and γ is the weight of the output unit; and u is the stochastic term that is normally distributed (that is, $u \sim N(0, \sigma^2 I_n)$).

From equation (1), f(**X**, **w**) is the artificial neural network function, written in the form

$$\mathrm{f}\left(\mathbf{X},\mathbf{w}\right)=\alpha\mathbf{X}+\sum_{h=1}^{H}\beta_h\,\mathrm{g}\left(\sum_{i=0}^{I}\gamma_{hi}x_i\right), \tag{2}$$

where g(·) is the transfer function

From Udomboso (2013), the proposed convoluted form of the artificial neural network function is given as

$$\mathrm{f}\left(\mathbf{X},\mathbf{w}\right)=\alpha\mathbf{X}+\sum_{h=1}^{H}\beta_h\left[\mathrm{g}_1\left(\sum_{i=0}^{I}\gamma_{hi}x_i\right)\mathrm{g}_2\left(\sum_{i=0}^{I}\gamma_{hi}x_i\right)\right], \tag{3}$$

which results in the proposed statistical neural network model

$$\mathrm{f}\left(\mathbf{X},\mathbf{w}\right)=\alpha\mathbf{X}+\sum_{h=1}^{H}\beta_h\left[\mathrm{g}_1\left(\sum_{i=0}^{I}\gamma_{hi}x_i\right)\mathrm{g}_2\left(\sum_{i=0}^{I}\gamma_{hi}x_i\right)\right]+u_iu_j\,, \tag{4}$$

where y is the dependent variable, $\mathbf{X} = (x_0 \equiv 1, x_1,\ldots,x_I)$ is a vector of independent variables; $\mathbf{w} = (\alpha, \beta, \gamma)$ is the network weight: α is the weight of the input unit, β is the weight of the hidden unit, and γ is the weight of the output unit; u_i and u_j are the stochastic term that is normally distributed (that is, $u_\mathrm{i} \sim N(0, \sigma^2 I_n)$); and $g_1(\cdot)$ and $g_2(\cdot)$ are the transfer functions.

First, let $g_1(\cdot)$ = Symmetric Saturated Linear function (SATLINS), defined as

$$\mathrm{SATLINS}=\mathrm{g}_1\left(\cdot\right)=\mathrm{f}_1\left(x\right)=\begin{cases}-1, & x<-1\\ x, & -1\le x\le 1\\ 1, & x>1\end{cases} \tag{5}$$

For this transfer function, the network model can thus be written as

$$y=\alpha_0+\alpha_1x_1+\sum_{h=1}^{H}\beta_h\left(\gamma_{h0}+\gamma_{h1}x_1\right)+u \tag{6}$$

for 2 variables and

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \sum_{h=1}^{H} \beta_h \left(\gamma_{h0} + \gamma_{h1} x_1 + \gamma_{h2} x_2\right) + u \tag{7}$$

for 3 variables.

Let $g_2(\cdot)$ = Hyperbolic Tangent Sigmoid function (TANSIG), defined as

$$\text{TANSIG} = \mathrm{g}_2(\cdot) = \mathrm{f}_2(x) = \frac{2}{1 - e^{-2x}} - 1 . \tag{8}$$

For this transfer function, the network model can thus be written as

$$y = \alpha_0 + \alpha_1 x_1 + \sum_{h=1}^{H} \beta_h \left(\frac{2}{1 - e^{-2(\gamma_{h0} + \gamma_{h1} x_1)}} - 1 \right) + u \tag{9}$$

for 2 variables and

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \sum_{h=1}^{H} \beta_h \left(\frac{2}{1 - e^{-2(\gamma_{h0} + \gamma_{h1} x_1 + \gamma_{h2} x_2)}} - 1 \right) + u \tag{10}$$

for 3 variables.

The Symmetric Saturating Linear and Hyperbolic Tangent Sigmoid Functions

(i) For $x < -1$, $f_1(x) = -1$. This implies that $f_1(x - y) = -1$. Let

$$\mathrm{f}(x) = \mathrm{f}_1(x) \otimes \mathrm{f}_3(x) = \int_{-r}^{x} \mathrm{f}_1(x - y)\,\mathrm{f}_3(y)\,dy \tag{11}$$

$$= \int_{-r}^{x} \left(1 - \frac{2}{1 - e^{-2y}} \right) dy$$

$$= (x + r) - 2\int_{-r}^{x} \left(1 + e^{-2y} + e^{-4y} + \ldots\right) dy$$

$$= \sum_{p=1}^{\infty} \frac{e^{-2px}}{p} - \sum_{p=1}^{\infty} \frac{e^{2pr}}{p} - (x + r) \tag{12}$$

(ii) For $-1 \leq x \leq 1$, $f_1(x) = x$. This implies that $f_1(x-y) = x - y$. Let

$$\begin{aligned} f(x) &= f_1(x) \otimes f_3(x) = \int_{-1}^{x} f_1(x-y) f_3(y)\,dy, \quad -1 \leq x \leq 1 \\ &= \int_{-1}^{x} (x-y)\left(\frac{2}{1-e^{-2y}} - 1\right) dy \end{aligned} \tag{13}$$

The integral

$$I = \int_{-1}^{x} \frac{2y}{1-e^{-2y}}\,dy$$

decreases rapidly for any interval of y. Hence, $I = 0$ and equation (13) becomes

$$f(x) = f_1(x) \otimes f_3(x) = 2x\int_{-1}^{x} 1 - e^{-2y}\,dy + \int_{-1}^{x} (x-y)\,dy \tag{14}$$

$$= x\left[2y - \sum_{p=1}^{\infty} \frac{e^{-2py}}{p}\right]_{-1}^{x} + \frac{y^2}{2} + x + \frac{1}{2}$$

$$= \left(2x^2 + 3x + \frac{1}{2}\right) - x\left(\sum_{p=1}^{x} \frac{e^{-2px}}{p} - \sum_{p=1}^{x} \frac{e^{-2p}}{p}\right) \tag{15}$$

(iii) For $x > 1$, $f_1(x) = 1$. This implies that $f_1(x-y) = 1$. Let

$$f(x) = f_1(x) \otimes f_3(x) = \int_{1}^{x} f_1(x-y) f_3(y)\,dy$$

Therefore,

$$f(x) = f_1(x) \otimes f_3(x) = \int_{1}^{x} 1\left(\frac{2}{1-e^{-2y}} - 1\right) dy \tag{16}$$

$$= 2\int_1^x \left(1 + e^{-2y} + e^{-4y} + \dots\right) dy - \int_1^x dy$$

$$= \left[2y - \sum_{p=1}^{\infty} \frac{e^{-2py}}{p} \right]_1^x - (x-1)$$

$$= (x-1) - \sum_{p=1}^{\infty} \frac{e^{-2px}}{p} + \sum_{p=1}^{\infty} \frac{e^{-2p}}{p} \tag{17}$$

The summary of the derived function is given as

$$\mathrm{g}_1\left(\sum\nolimits_{i=1}^{I} \gamma_{hi} x_i\right) \mathrm{g}_2\left(\sum\nolimits_{i=1}^{I} \gamma_{hi} x_i\right) = \mathrm{f}(x)$$

$$= \begin{cases} \sum_{p=1}^{\infty} \frac{e^{-2px}}{p} - \sum_{p=1}^{\infty} \frac{e^{2pr}}{p} - (x+r), & x < -1 \\ \left(2x^2 + 3x + \frac{1}{2}\right) - x\left(\sum_{p=1}^{x} \frac{e^{-2px}}{p} - \sum_{p=1}^{x} \frac{e^{-2p}}{p}\right), & -1 \le x \le 1 \\ (x-1) - \sum_{p=1}^{\infty} \frac{e^{-2px}}{p} - \sum\nolimits_{p=1}^{\infty} \frac{e^{-2p}}{p}, & x > 1 \end{cases} \tag{18}$$

Equation (18) is the derived transfer function for the Symmetric Saturated Linear transfer function and the Hyperbolic Tangent transfer function (SATLINS*TANSIG).

For this derived transfer function, the network model can thus be written as

$$y = \alpha_0 + \alpha_1 x_1 + \sum_{h=1}^{H} \beta_h \left(\left(2\left(\gamma_{h0} + \gamma_{h1} x_1\right)^2 + 3\left(\gamma_{h0} + \gamma_{h1} x_1\right) + \frac{1}{2} \right) - \left(\gamma_{h0} + \gamma_{h1} x_1\right) \left(\sum_{p=1}^{n} \frac{e^{-2p\left(\gamma_{h0} + \gamma_{h1} x_1\right)}}{p} - \sum_{p=1}^{n} \frac{e^{-2p}}{p} \right) \right) + u_1^2 \tag{19}$$

for 2 variables and

$$y=\alpha_0+\alpha_1x_1+\alpha_2x_2+\sum_{h=1}^{H}\beta_h\Big(\Big(2\left(\gamma_{h0}+\gamma_{h1}x_1+\gamma_{h2}x_2\right)^2 +3\left(\gamma_{h0}+\gamma_{h1}x_1+\gamma_{h2}x_2\right)+{}^1\!/_2\Big) -\left(\gamma_{h0}+\gamma_{h1}x_1+\gamma_{h2}x_2\right)\left(\sum_{p=1}^{n}\frac{e^{-2p\left(\gamma_{h0}+\gamma_{h1}x_1+\gamma_{h2}x_2\right)}}{p}-\sum_{p=1}^{n}\frac{e^{-2p}}{p}\right)\Big)+u_1u_2 \tag{20}$$

for 3 variables.

The criteria used to compare the models in this study include

(i) Mean Square Error (MSE):

$$\text{MSE}\left(\hat{\mathbf{w}},\theta\right)=\text{E}\left\{\left\|\hat{\theta}-\hat{\mathbf{w}}\right\|^2\right\}, \tag{21}$$

(ii) Network Information Criterion (*NIC*):

$$NIC=\text{MSE}\left[\mathbf{w}^*,\hat{\mathbf{w}}\right]+\frac{1}{m}\text{tr}\left\{\left[\frac{\partial}{\partial\hat{\mathbf{w}}}\text{var}\left(\mathbf{w}^*-\hat{\mathbf{w}}\right)\right]^2\left[\frac{\partial^2}{\partial\hat{\mathbf{w}}^2}\text{E}\left(\mathbf{w}^*-\hat{\mathbf{w}}\right)^2\right]^{-1}\right\} \tag{22}$$

(iii) Adjusted Network Information Criterion (*ANIC*), developed by Udomboso et al. (2016) as decision criterion when sample sizes vary in the network:

$$ANIC=NIC+\frac{np-p^2+n}{2\left(n-p\right)} \tag{23}$$

where **w** = (α, β, γ) is the network weight: α is the weight of the input unit, β is the weight of the hidden unit, and γ is the weight of the output unit; θ is the true parameter; and p is the number of parameters under estimation in the network. The *ANIC* is a correction for the biased *NIC*.

Results

The data used for the analyses in this research were simulated and split into two case – 2 variables and 3 variables. The results are based on the prediction and model selection criteria at different levels of sample sizes and hidden neurons, respectively. The sample sizes include 10, 20, 40, 60, 80, 100, 125, 150, 175, 200, 250, 300, and 400, while the hidden neurons include 2, 5, 10, 20, 40, 60, 80, and 100. Two primary transfer functions, as well as a derived transfer function arising from the convolution of the transfer functions, were used, namely:

(i) Symmetric Saturated Linear transfer function (SATLINS)
(ii) Hyperbolic Tangent Sigmoid transfer function (TANSIG)
(iii) Symmetric Saturated Linear and Hyperbolic Tangent Sigmoid transfer function (SATLINS*TANSIG)

Analyses of Results based on Sample Sizes

The analyses in this section are, respectively, discussed under the 2- and 3-variable cases. Model selection criteria were based on the MSE, *NIC*, and *ANIC*, respectively.

Table 1 contains results for the model fit across samples from the 2-variable case. The comparison of the models based on the MSE showed that SATLINS had local minima at sample sizes 10, 100, and 250, while TANSIG at sample sizes 10 and 200. Local minima of MSE were recorded with SATLINS*TANSIG at 10, 100, and 250. Results based on *NIC* showed that SATLINS local minima at sample sizes 20, 80, 150, and 250. Local minima at TANSIG were at sample sizes 10, 40, 100, 150, 250, and 400, while at SATLINS*TANSIG, the local minima occur at sample sizes 10, 60, 100 and 250. Moreover, results based on *ANIC* showed SATLINS to have local minima at sample size 250 only; local minima with TANSIG were noticed at sample sizes 20, 60, 200, and 300. On the other hand, local minima for SATLINS*TANSIG were at sample sizes 150, 250, and 300.

In the case of Table 2, it is the model fit across samples from the 3-variable case. The results showed that with the MSE, SATLINS had local minima at sample sizes 20, 80, 125, and 200. In the case of TANSIG, local minima occur at sample sizes 20, 80, 125, and 250, while for SATLINS*TANSIG, local minima were seen at sample sizes 20, 60, 125, 175, and 250. Using the *NIC*, local minima were noticed for SATLINS at sample sizes 40, 80, and 200, while for TANSIG, records of local minima were seen at sample sizes 40, 80, 125, 175, and 250. As for

SATLINS*TANSIG, the local minima were at sample sizes 10, 60, 125, and 250. Furthermore, with *ANIC*, SATLINS had local minima at sample sizes 80, 125, and 200, TANSIG at sample sizes 10, 40, 150, and 300. But for SATLINS*TANSIG, results of local minima occurred at sample sizes 40, 250, and 300.

Table 1. Model selection across samples (2 variables)

	Mean Square Error			Network Information Criteria		
Sample size	SATLINS	TANSIG	SATLINS* TANSIG	SATLINS	TANSIG	SATLINS* TANSIG
10	6.95E-06	8.06E-06	8.23E-06	3.81E-03	3.07E-03	4.92E-03
20	3.92E-05	3.42E-05	3.50E-05	2.56E-03	1.20E-02	1.25E-02
40	1.53E-04	1.08E-04	1.38E-04	2.39E-02	1.65E-03	5.59E-03
60	2.62E-04	2.31E-04	1.94E-04	2.13E-03	4.71E-03	9.78E-04
80	1.77E-04	2.88E-04	3.40E-04	1.73E-04	2.28E-03	1.26E-03
100	1.74E-04	3.49E-04	2.91E-04	6.93E-04	2.99E-04	2.70E-04
125	3.20E-04	3.58E-04	4.74E-04	1.31E-03	1.13E-02	1.84E-03
150	8.62E-04	9.00E-04	8.25E-04	1.06E-03	1.12E-03	1.92E-03
175	1.12E-03	1.31E-03	1.40E-03	4.39E-03	3.84E-03	5.01E-03
200	1.53E-03	1.19E-04	9.93E-04	3.91E-03	2.39E-03	3.91E-03
250	6.99E-04	1.20E-04	6.31E-04	1.18E-03	1.71E-03	7.23E-04
300	2.30E-03	3.02E-03	2.77E-03	3.15E-03	5.25E-03	4.13E-03
400	6.54E-03	3.68E-03	4.20E-03	6.79E-03	4.45E-03	4.31E-03

	Adjusted Network Information Criteria		
Sample size	SATLINS	TANSIG	SATLINS* TANSIG
10	1.622	1.563	1.526
20	1.558	1.519	1.526
40	1.550	1.520	1.515
60	1.515	1.509	1.512
80	1.513	1.510	1.509
100	1.511	1.519	1.507
125	1.509	1.516	1.506
150	1.507	1.508	1.504
175	1.505	1.506	1.507
200	1.505	1.505	1.506
250	1.504	1.508	1.505
300	1.505	1.506	1.596
400	1.508	1.610	1.571

Table 2. Model selection across samples (3 variables)

	Mean Square Error			Network Information Criteria		
Sample size	SATLINS	TANSIG	SATLINS* TANSIG	SATLINS	TANSIG	SATLINS* TANSIG
10	1.85E-02	1.81E-02	2.04E-02	4.68E-01	3.12E-01	5.17E-02
20	1.46E-02	1.38E-02	1.37E-02	3.06E-02	1.10E-01	7.84E-02
40	1.65E-02	1.44E-02	1.44E-02	1.96E-02	2.16E-02	6.01E-02
60	1.63E-02	1.53E-02	1.42E-02	3.63E-02	5.37E-02	1.98E-02
80	1.14E-02	1.39E-02	2.17E-02	2.10E-02	1.60E-02	2.01E-02
100	1.68E-02	1.82E-02	1.49E-02	5.61E-02	1.89E-02	2.82E-02
125	9.93E-03	1.23E-02	1.38E-02	8.99E-03	1.49E-02	1.93E-02
150	1.66E-02	1.36E-02	1.44E-02	1.66E-02	2.13E-02	2.06E-02
175	1.23E-02	1.73E-02	1.26E-02	1.54E-02	1.73E-02	1.80E-02
200	1.17E-02	1.79E-02	1.53E-02	1.39E-02	2.54E-02	1.76E-02
250	1.93E-02	1.11E-02	1.00E-02	2.03E-02	1.65E-02	1.43E-02
300	2.03E-02	1.60E-02	1.77E-02	2.30E-02	2.06E-02	1.92E-02
400	3.78E-02	3.66E-02	2.59E-02	4.36E-02	4.89E-02	1.38E-01

	Adjusted Network Information Criteria		
Sample size	SATLINS	TANSIG	SATLINS* TANSIG
10	2.117	2.008	2.185
20	2.108	2.185	2.136
40	2.057	2.038	2.009
60	2.027	2.075	2.041
80	2.041	2.034	2.037
100	2.068	2.034	2.031
125	2.021	2.025	2.025
150	2.023	2.016	2.022
175	2.022	2.022	2.017
200	2.019	2.023	2.014
250	2.023	2.016	2.009
300	2.023	2.014	2.019
400	2.035	2.022	1.882

Analyses of Results based on Hidden Neurons

The analyses are also discussed under the 2- and 3-variable cases, respectively. The criteria for model selection used here include the MSE and the *NIC*. The *ANIC* is not used in this section since sample sizes are not involved in the analyses.

As for Table 3, it is the model fits across the hidden neurons from the 2-variable case. Results showed that with MSE, SATLINS had local minima at hidden neurons 5 and 20, while TANSIG had local minima at hidden neurons 10, 60, and 100. In the case of SATLINS*TANSIG, local minima occurred at hidden neurons 10, 60, and 100. Using *NIC*, it is shown that SATLINS had local minima at hidden neuron 20 only, TANSIG at hidden neurons 20 and 80. As for SATLINS*TANSIG, local minima occurred at hidden neurons 20, 60, and 100.

The results in Table 4 are for the model fits across hidden neurons from the 3-variable case. It is noticed that for MSE, SATLINS had local minima at hidden neuron 20 and 100, while TANSIG at hidden neurons 60 and 100. In the case of SATLINS*TANSIG, local minima occurred at hidden neurons 10, 40, and 100. As for *NIC*, SATLINS had local minima at hidden neurons 20 and 100, TANSIG at hidden neurons 5, 20, and 80, while for SATLINS*TANSIG, local minima were noticed at hidden neurons 10, 60, and 100.

Table 3. Model selection across hidden neurons (2 variables)

	Mean Square Error			**Network Information Criteria**		
Hidden neurons	**SATLINS**	**TANSIG**	**SATLINS* TANSIG**	**SATLINS**	**TANSIG**	**SATLINS* TANSIG**
2	2.22E-03	2.32E-03	2.46E-03	2.02E-02	1.42E-02	1.63E-02
5	1.25E-03	7.84E-04	1.16E-03	2.77E-03	1.67E-03	3.28E-03
10	1.40E-03	3.75E-04	6.65E-04	2.24E-03	1.77E-03	2.42E-03
20	4.00E-04	7.15E-04	7.99E-04	8.70E-04	1.64E-03	1.55E-03
40	4.17E-04	1.03E-03	6.64E-04	1.01E-03	1.87E-03	2.09E-03
60	6.70E-04	7.79E-04	5.46E-04	1.53E-03	2.54E-03	1.19E-03
80	1.18E-03	9.29E-04	8.06E-04	2.21E-03	1.94E-03	1.57E-03
100	1.18E-03	8.66E-04	4.74E-04	3.03E-03	7.61E-03	7.20E-04

Table 4. Model selection across hidden neurons (3 variables)

	Mean Square Error			**Network Information Criteria**		
Hidden neurons	**SATLINS**	**TANSIG**	**SATLINS* TANSIG**	**SATLINS**	**TANSIG**	**SATLINS* TANSIG**
2	7.34E-02	7.76E-02	7.14E-02	3.15E-01	2.83E-01	1.20E-01
5	2.46E-02	2.02E-02	2.25E-02	2.47E-02	1.38E-02	2.17E-02
10	6.98E-03	7.24E-03	5.51E-03	1.49E-02	3.23E-02	1.10E-02
20	3.80E-03	6.98E-03	6.29E-03	1.17E-02	1.23E-02	4.72E-02
40	4.96E-03	6.37E-03	3.14E-03	2.60E-02	2.44E-02	2.34E-02
60	5.10E-03	4.94E-03	4.81E-03	4.26E-02	2.72E-02	2.29E-02
80	1.10E-02	8.20E-03	7.90E-03	2.50E-02	1.79E-02	4.74E-02
100	8.35E-03	2.95E-03	6.95E-03	1.60E-02	1.84E-02	1.66E-02

Table 5. Mean performance of the model selection across samples

		Mean Square Error			Network Information Criteria		
		SATLINS	TANSIG	SATLINS* TANSIG	SATLINS	TANSIG	SATLINS* TANSIG
2-var	Mean	1.09E-03	8.10E-04	9.46E-04	4.24E-03	4.16E-03	3.64E-03
	SD	1.77E-03	1.19E-03	1.23E-03	6.18E-03	3.63E-03	3.23E-03
3-var	Mean	1.71E-02	1.68E-02	1.61E-02	5.95E-02	5.36E-02	3.89E-02
	SD	7.01E-03	6.37E-03	4.29E-03	1.23E-01	8.20E-02	3.59E-02

		Adjusted Network Information Criteria		
		SATLINS	TANSIG	SATLINS* TANSIG
2-var	Mean	1.52E+00	1.52E+00	1.52E+00
	SD	3.42E-02	3.02E-02	2.86E-02
3-var	Mean	2.05E+00	2.04E+00	2.03E+00
	SD	3.35E-02	4.68E-02	7.00E-02

Table 6. Mean performance of the model selection across hidden neurons

		Mean Square Error			Network Information Criteria		
		SATLINS	TANSIG	SATLINS* TANSIG	SATLINS	TANSIG	SATLINS* TANSIG
2-var	Mean	1.09E-03	9.75E-04	9.47E-04	4.23E-03	4.16E-03	3.64E-03
	SD	6.00E-04	5.77E-04	6.46E-04	6.50E-03	4.53E-03	5.18E-03
3-var	Mean	1.73E-02	1.68E-02	1.61E-02	5.95E-02	5.37E-02	3.88E-02
	SD	2.36E-02	2.51E-02	2.32E-02	1.04E-01	9.29E-02	3.54E-02

Compiled in Table 5 is the mean performance of the model fits across samples. Results showed for the 2-variable case, with the MSE, TANSIG had the best mean performance (8.10E-4, 1.19E-3). Results obtained from *NIC* showed SATLINS*TANSIG having best mean performance (3.64E-3, 3.23E-3), and from *ANIC*, the SATLINS*TANSIG also showed best performance (1.522E+0, 2.86E-2). In the case of the 3-Variable case, for MSE, SATLINS*TANSIG had the best mean performance (1.61E-2, 4.29E-3). Similarly, for *NIC* and *ANIC*, SATLINS*TANSIG had the best mean performances, (3.89E-2, 3.59E-2) and (2.03E+0, 7.00E-2).

Table 6 contains the mean performance of the model fits across hidden neurons. From the 2-variable case, the results from the MSE showed that SATLINS*TANSIG had the best mean performance (9.47E-4, 6.46E-4), and also with the *NIC*, SATLINS*TANSIG had the best mean performance (3.64E-3, 5.18E-3). In the case of 3 variables, from the MSE result, SATLINS*TANSIG had the best mean performance (1.61E-2, 2.32E-2) and, in the same vein, from the *NIC*, SATLINS*TANSIG had the best mean performance (3.88E-2, 3.54E-2).

Conclusion

The heterogeneous model of the Statistical Neural Network had higher precision overall in comparison with the homogeneous models of the Statistical Neural Network from which it was derived. Specifically, in both 2- and 3-variable cases in relation to sample sizes, SATLINS*TANSIG was shown to have a better performance in relation to the homogeneous models of SATLINS and TANSIG, respectively. This is shown in the Adjusted Network information being sensitive to increase in sample size, indicated by the several local minima in the analyses. Likewise, in the case of hidden neurons, it is shown in the 2-variable case, SATLINS*TANSIG was sensitive to higher neurons, in comparison to SATLINS and TANSIG, respectively. In the case of 3 variables, SATLINS*TANSIG was more sensitive to the hidden neurons in relation to SATLINS and TANSIG. This is also indicated by the several local minima in the analyses. Therefore, in a training a neural network model, large sample sizes and hidden neurons would be necessary if precision of a model is of importance.

References

Adepoju, G. A., Ogunjuyigbe, S. O. A., & Alawode, K. O. (2007). Application of neural network to load forecasting in Nigerian electrical power system. *The Pacific Journal of Science and Technology, 8*(1), 68-72.

Adewole, A. P., Akinwale, A. T., & Akintomide, A. B. (2011). Artificial neural network model for forecasting foreign exchange rate. *World of Computer Science and Information Technology Journal, 1*(3), 110-118.

Adeyiga, J. A., Ezike, J. O. J., Omotosho, A., & Amakulor, W. (2011). A neural network based model for detecting irregularities in e-banking transactions. *African Journal of Computer and ICTs, 4*(3.2), 7-14.

Akinwale, A. T, Arogundade, O. T., & Adekoya, A. F. (2009). Translated Nigeria stock market prices using artificial neural network for effective prediction. *Journal of Theoretical and Applied Information Technology, 9*(1), 36-

43. Retrieved from http://www.jatit.org/volumes/research-papers/Vol9No1/6Vol9No1.pdf

Anders, U. (1996). Statistical model building for neural networks. In *6th AFIR colloquium* (pp. 963-979). Nunberg, Germany. Retrieved from https://www.actuaries.org/AFIR/colloquia/Nuernberg/Anders.pdf

Ashigwuike, C. E. (2012). Estimation of solar power generation in some Nigerian cities using artificial neural network. *Journal of Chemical, Biological and Physical Sciences, 2*(2), 929-936.

Falode, O. A., & Udomboso, C. G. (2016). Predictive modeling of gas production, utilization and flaring in Nigeria using TSRM and TSNN: A comparative approach. *Open Journal of Statistics, 6*(1), 194-207. doi: 10.4236/ojs.2016.61017

Gan, C., Limsombunchai, V., Clemes, M., & Weng, A. (2005). Consumer choice prediction – Artificial neural networks versus logistic models. *Journal of Social Sciences, 1*(4), 211-219.

Ibeh, G. F., Agbo, G. A., Rabia, S., & Chikwenze, A. R. (2012). Comparison of empirical and artificial neural network models for the correlation of monthly average global solar radiation with sunshine hours in Minna, Niger State, Nigeria. *International Journal of Physical Sciences, 7*(8), 1162-1165.

Omole, O., Falode, O. A., & Deng, A. D. (2009). Prediction of Nigerian crude oil viscosity using artificial neural network. *Petroleum & Coal, 51*(3), 181-188.

Resop, J. P. (2006). *A comparison of artificial neural networks and statistical regression with biological resources applications* [Unpublished Master's thesis]. University of Maryland, College Park. Retrieved from https://drum.lib.umd.edu/handle/1903/3901

Tayfur, G. (2002). Artificial neural networks for sheet sediment transport. *Hydrological Sciences Journal, 47*(6), 879-892. doi: 10.1080/02626660209492997

Toprak, Z. F., & Cigizoglu, H. K. (2008). Predicting longitudinal dispersion coefficient in natural streams by artificial neural networks. *Hydrological Processes, 22*(20), 4106-4129. doi: 10.1002/hyp.7012

Udomboso, C. G. (2013). On some properties of a hetereogeneous transfer function involving symmetric saturated linear (SATLINS) with hyperbolic tangent (TANH) transfer functions. *Journal of Modern Applied Statistical Methods, 12*(2), 427-435. doi: 10.22237/jmasm/1383279900

Udomboso, C. G., & Amahia, G. N. (2011). Comparative analysis of rainfall prediction using statistical neural network and classical linear regression model. *Journal of Modern Mathematics and Statistics, 5*(3), 66-70. Retrieved from http://docsdrive.com/pdfs/medwelljournals/jmmstat/2011/66-70.pdf

Udomboso, C. G., Amahia, G. N., & Dontwi, I. K. (2016). An adjusted network information criterion for model selection in statistical neural network models. *Journal of Modern Applied Statistical Methods*, *15*(2), 411-427. doi: 10.22237/jmasm/1478003040

Udomboso, C. G., & Saliu, F. U. (2016). On building inference for the statistical neural network with application to Naira-Dollar exchange rate efficiency – A bootstrap approach. *CBN Journal of Applied Statistics*, *7*(2), 123-136.