An investigation of the suitability of Artificial Neural Networks for the prediction of core and local skin temperatures when trained with a large and gender-balanced database

K Michael[a], MDP Garcia-Souto[b] and P Dabnichki[c]

[a] Blizard Institute, Queen Mary, University of London

The Blizard Building, 4 Newark Street, London E1 2AT, UK.

[b] Medical Physics and Biomedical Engineering, University College London

Malet Place Engineering Building - Gower Street, London WC1E 6BT, UK

[c] School of Engineering, RMIT University

GPO Box 2476, Melbourne, Victoria, Australia, 3001

p.garciasouto@ucl.ac.uk --- peter.dabnichki@rmit.edu.au

Contact information:
Dr. Pilar Garcia Souto
p.garciasouto@ucl.ac.uk
Medical Physics and Biomedical Engineering, University College London
Malet Place Engineering Building - Gower Street – London - WC1E 6BT - UK
Phone: +44 (0)20 7679 0454

Abstract

Neural networks have been proven to successfully predict the results of complex non-linear problems in a variety of research fields, including medical research. Yet there is paucity of models utilising intelligent systems in the field of thermoregulation. They are under-utilized for predicting seemingly random physiological responses and in particular never used to predict local skin temperatures; or core temperature with a large dataset. In fact, most predictive models in this field (non-artificial intelligence based) focused on predicting body temperature and average skin temperature using relatively small gender-unbalanced databases or data from thermal dummies due to a lack of larger datasets.

This paper aimed to address these limitations by applying Artificial Intelligence to create predictive models of core body temperature and local skin temperature (specifically at forehead, chest, upper arms, abdomen, knees and calves) while using a large and gender-balanced experimental database collected in office-type situations.

A range of Neural Networks were developed for each local temperature, with topologies of 1-2 hidden layers and up to 20 neurons per layer, using Bayesian and the Levemberg-Marquardt back-propagation algorithms, and using various sets of input parameters (2520 NNs for each of the local skin temperatures and 1760 for the core temperature, i.e. a total of 19400 NNs). All topologies and configurations were assessed and the most suited recommended. The recommended Neural Networks trained well, with no sign of over-fitting, and with good performance when predicting unseen data. The recommended Neural Network for each case was compared with previously reported multi-linear models. Core temperature was avoided as a parameter for

local skin temperatures as it is impractical for non-contact monitoring systems and does not significantly improve the precision despite it is the most stable parameter. The recommended NNs substantially improve the predictions in comparison to previous approaches. NN for core temperature has an R-value of 0.87 (81% increase), and a precision of ±0.46°C for an 80%CI which is acceptable for non-clinical applications. NNs for local skin temperatures had R-values of 0.85-0.93 for forehead, chest, abdomen, calves, knees and hands, last two being the strongest (increase of 72% for abdomen, 63% for chest, and 32% for calves and forehead). The precision was best for forehead, chest and calves, with about ±1.2°C, which similar to the precision of existent *average skin temperature models* even though the average value is more stable.

Keywords

Core temperature; Skin temperature; Neural Network; Artificial Intelligence; Model

List of Abbreviations

ADALINE – Adaptive Linear Element

AI – Artificial Intelligence

ANN/NN – Artificial Neural Network(s)

BMI – Body Mass Index (body mass of the subject divided by the square of the height in m, i.e. kg/m^2)

Epoch – a full data cycle, i.e. each time the network is presented with a new training pattern

FNN – Fuzzy Neural Network(s)

Logsig – The logistic sigmoid Function

MSE – Mean Squared Error

Nodes – the interconnections between hidden layers in the Neural Network (equivalent to neuron connections)

PE – Percentage Error (%)

POAH – Pre-Optic Anterior Hypothalamus

R^2 – Coefficient of determination (determines the goodness of fit)

RSE – Relative Standard Error (%)

SD – Standard Deviation

SE – Standard Error

1. Introduction

The understanding of thermoregulation is essential for both our survival and daily comfort. It is essential to prevent heat strokes [1], minimise discomfort in large places [2] and reduce the spread of infections [1]. Thermoregulation also influence our susceptibility to diseases, since body temperature rise leads to the destruction of immune system proteins when reaching 41-2 °C, leaving it compromised [3]. Nowadays it is

used to adapt environmental temperatures to the occupant's needs in indoor spaces (e.g. public spaces, large venues, or aeroplanes) while aiming for reduced energy consumption [4].

Core body temperature regulation originates within the Pre-Optic Anterior Hypothalamic (POAH) region of the brain [5]. It is a complex system that works to maintain the core temperature in a narrow range of values (36 – 37 °C) and whose process is affected by individual characteristics (e.g. gender) and external factors (e.g. climate, exposure to the sun). The regulatory system has a great range of adaptability; even in extreme environmental conditions our body temperature is not easily changed [6], fluctuating just enough to adapt to the external conditions, our daily schedule and protecting our body from overheating [7]. Our thermal physiology is 'asymmetric', which adds to the complexity, with a survival range of 21 °C (thermal minimum) to 42 °C (thermal maximum) [6], although it is rare to go anywhere near these extreme during our lives.

1.1. Existing models

Body temperature regulation models need to consider a number of external factors such as room temperature, wind velocity, humidity in addition to regulatory responses like sweating, shivering and heat distribution through the body [8]. Two different approaches to model body temperature are found in the literature

(1) numerical or mathematical models where the temperature distribution is derived at discrete temperature points based on experimental data, or as a set of differential equations (nodes), respectively; [9] and

(2) physical models which simulate the heat production and thermal transport processes [9].

Numerical models for mean skin temperature were derived by Nielsen and Nielsen [8]. They studied the number and location of temperature points required for such prediction and found that 13 points deliver a good representation of skin temperature distribution for clothed subjects, but noted that more points are needed in cool environments. However, when measuring at 13 points is not possible/practical, 4 specific points (forehead, hand, lower back, lower leg) provide a good estimation of the mean skin temperature. Currently most experimental studies use Hardy and Dubois seven-point average which was found to be the minimum necessary for an accurate prediction of mean skin temperature [10].

One of the most significant and influential physical models for thermoregulation mapping is Stolwijk's (1966) 25-node model, originally a theoretical study on the thermoregulatory responses of the body divided into two systems: the passive and the controlling. The passive system deals with the heat exchanges within the body and the differential equations that govern it. The controlling system relates to the regulatory response aiming to keep the core temperature stable [9]. Stolwijk's 25-node model has since been used as the basis for new mathematical thermoregulation models of growing complexity and accuracy, e.g. accounting for the evaporation of sweat [11,12]; improving the blood flow description in the model by simulating the human arterial system [13]; developing further the controlling or active system to simulate responses such as shivering, sweating and blood vessel motion to address a shortcoming of the original Stolwijk model [14]; increasing the number and changing the shape (cylinders, spheres) of the nodes [12,15] as well as simplifying the model (decreasing temperature points) to see how it performs. Furthermore, models have been created that consider individual body characteristics [9], as well as to improve thermal comfort [14,16].

Some physical models use thermal manikins to simulate thermoregulation responses and local and overall heat exchange of the body, particularly within the automobile and textile industry as the thermal response can be quickly assessed in inhomogeneous environments (e.g. vehicles, planes, etc.) or with different clothing. However, these models only predict the effect of surrounding environments on the body rather than its thermoregulation. Alternative hybrid models have been investigated [14,17], for instance Nilsson and Holmer [17] coupled a thermal manikin with a computer simulation (CFD calculations) to predict thermal response in

different work environments, which seem to be successful but highlighted the need for more advanced computer simulations.

Despite the existence of a number of theoretical and computational models, very few have been tested or even developed using more than a few subjects, many are based on thermal manikins (who are in fact hardware models), rendering their findings less reliable. Mehnert et al. [18] and Garcia-Souto and Dabnichki [4] addressed this issue and developed statistical models for mean skin temperature and core and local skin temperature respectively based on independent large sets of experimental data. This improves the coefficients of the mathematical and numerical models and therefore their outcome [18]. Mehnert and Garcia-Souto and Dabnichki's investigations agreed with Fiala's et al. [16] observation of a non-linear relationship between regulatory responses (sweating, shivering, and blood vessel motion) and core and skin temperature, either local or global, respectively. Core temperature was found essential in predicting mean skin temperature in high temperatures [18]. Garcia-Souto and Dabnichki observed also that core temperature remained stable for low room temperatures while it increased significantly for higher temperatures; local skin temperature was found to have a multi-linear relation with room temperatures in agreement with Mehnert et al. [18] and Fiala et al. [16]; in low room temperatures forehead, chest and upper arm temperature remained constant, while for higher room temperatures the same happen for the abdomen and knees suggesting that local skin temperature stabilizes after reaching a certain threshold temperature; the largest variation of local skin temperature was at the limbs as expected while chest temperature remained the most stable for clothed subjects due to clothing insulation [8,18]. Garcia-Souto and Dabnichki also investigated the effect of gender on thermoregulation, which is still the sole large scale study of large scale despite strong anecdotal evidence of gender dependent thermal comfort requirements.

1.2. Neural networks approach

Over the last 2 decades researchers have been turning to AI for forecast and predictions in a variety of fields [19] such as forecasting peak demand electrical load [20], or short-term load forecasting for summer time [21], heating gas consumption [22], cash demand in ATMs [23], hourly forecast of ambient temperature in Spain [24], and predicting sea surface temperatures in the North Indian Ocean [25]. Also, hybrid models have been created where NNs have been coupled with fuzzy logic to forecast rainfall based on weather pattern [26] or forecasting daily Malaysian electricity load demand [27].

AI has also been widely applied to the prediction of medical outcomes [28], e.g. for the classification of patient's health status in relation to kidney disease [29], prediction of protein enzymatic class [30], ventilation control for thermal comfort and optimization of energy consumption in public buildings [2,31]. Some studies also presented the efficiency of the predictive NNs in comparison to the more traditional logistic regression [32] for re-hospitalisation of patients with stroke [33] or living after hip fracture [28]. Examples more closely related with this paper are prediction of body temperature changes and other responses as indicators or heat strain [34] and prediction of local and overall clothing thermal sensation [35]. Sanders & Lau [34] used a multi-layer ADALINE NN trained with external factors as well as temperature measurements and compared the results with a heat strain predictive model. Luo et al. [35] used a Fuzzy Neural Network (FNN) trained with local skin and core temperature measurements and their change rate and compared the model to data from a simulation system for human clothing. Both Sanders & Lau [34] and Luo et al [35] have shown that the use of NNs provide accurate results while helping to reduce the rigorous lab experiments and size of the data base needed to study such complex problems. The use of fuzzy logic can decrease data requirements further since the network is trained using inference rules from already established relationships within thermoregulation. However neither paper used a sufficiently large database to validate the use of AI.

1.3. Motivation

The ability to predict the core and skin temperature based on environmental parameters and basic descriptors of the individual is of great interest, e.g. for the fields of thermoregulation and climate control. Various models such as descriptive statistical models already exist, but have a limited success due to the numerous parameters affecting temperature values, and the complicated and unknown inter-dependencies among these factors. Most models are also based on small experimental datasets while large and gender-balanced databases are needed. Besides the potential benefits of using Artificial Intelligence to overcome in part the complexity of this problem, its application in the field of thermoregulation is still quite limited.

Our approach, and main significance of the study, was to identify an appropriate computational technique that does not rely on inferred relations. We selected a well-established method in AI (Neural Networks of various topologies, with Bayesian and Levemberg-Marquardt back-propagation algorithms) for the prediction of 'local skin temperatures' and 'core temperature' based on 'environmental conditions' (room temperature, humidity) and 'personal characteristics' (gender, age, BMI, clothing), and investigated its suitability in some detail. In the process we used a large gender-balanced database to train and make a detailed assessment of these Neural Networks' performances, and compared them with earlier non-AI approaches of analogous inputs and outputs [4] using the same database as to get an informed conclusion. Additional NNs with a reduced number of inputs were also assessed to study the robustness of the prediction under possible simplifications. For each of the temperatures we wished to predict, the optimal NN topology and training algorithm that maximizes the predicting performance were identified and reported.

2. Methods

2.1. Data

The data used was previously collected by Garcia-Souto and Dabnichki [4] from 159 healthy volunteers following a study protocol approved by the University of London Ethical Committee.

Temperature measurements were taken from 13 different skin areas on the body as shown in **Figure 1**. Skin temperature was simply measured on the skin when exposed, and reaching under the cloths when covered. In addition to these the core (tympanic) temperature was measured and the gender, age, BMI and layers of clothing at various locations were recorded.
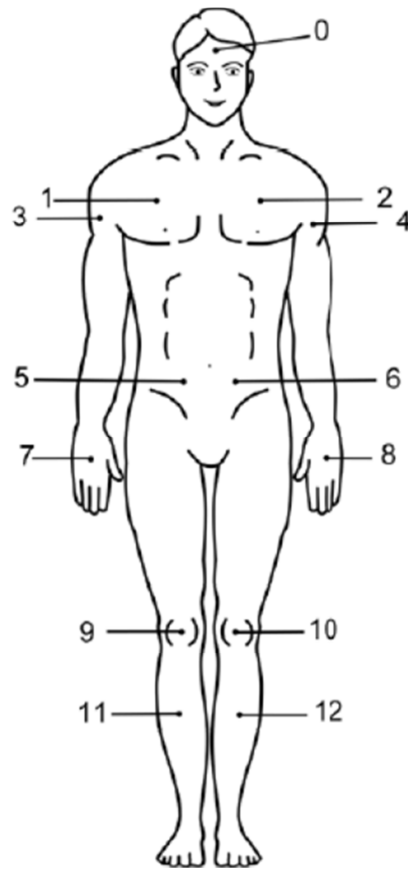
Figure 1: Measuring points

2.2. Analysis

Thermoregulation responses of the human body are not fully deterministic; hence AI techniques provide a suitable option for a thermoregulation model. Different techniques such as the Bayesian approach, Markov Chains, Expert Systems, evolutionary algorithms and artificial neural networks are potentially viable and were considered in this work. Approaches such as expert systems and evolutionary algorithms require the definition of rules to fit data, and such rules do not exist for thermoregulation where some stochastic inferences may affect the response. The use of NNs was selected instead for their black box nature and their ability to 'learn' these inference rules from the data. NNs are also very well suited for dealing with noisy data that might arise from an experiment [36].

Therefore it was decided to develop eight different Artificial Neural Networks to predict core and local skin temperature at seven sites (forehead, chest, upper arms, abdomen, knees and calves) assuming temperature symmetry and using individual and environmental information. Specifically, for each desired output we searched for the best combination of topology among a range of options, training algorithm and set of inputs to accurately predict each temperature but without overfitting the networks. This process resulted on an informed recommendation of 8 single NNs, each one to predict one of the desired temperatures (core and each of the 7 local skin temperatures).

The experimental data was normalised as NNs work better with smaller numbers [36]. The min-max and standardized normalisation (equation 1) were tested and the latter was selected.

$$X' = \frac{(X - \mu)}{\sigma} \qquad \text{Equation 1}$$

The experimental data (which includes input data X and response variable or output Y) was separated into two cohorts. 90% of the data was used for establishing the optimal training conditions and then training, validation and testing with the optimal topology (that deeming the lowest error as defined in section 2.2.2). The remaining 'unseen' 10% of the data were used to assess the obtained networks, by simulating the network generated predictions. This partition is independent to any partitions done within the various training algorithms used.

Various NNs were created, in MATLAB [37] using the input parameters proposed by Garcia-Souto and Dabnichki [4]. The accuracy of these NNs was assessed by studying:

a. the performance during training,
b. the accuracy of their prediction in the 'unseen' data, and
c. the normality of their error distribution using descriptive statistics.

Additionally, alternative NNs were created where the effect of different input parameters was investigated (these are further discussed in section 2.2.1), comparing accuracy of results and computational costs. These NNs were created and assessed following the same procedure as the 'original' NNs.

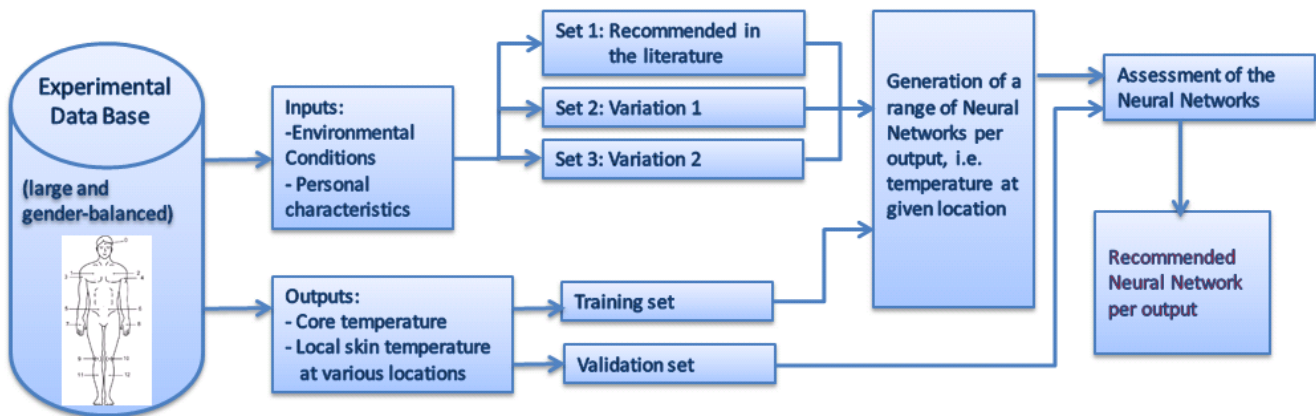The development process of the recommended NN is summarized in **Figure 2**.

Figure 2: Development process of the recommended NNs for the prediction of core and local skin temperatures followed in this paper

2.2.1.   Parameters

The parameters used for each original network (8 in total to predict *core temperature, T0, T1, T3, T5, T7, T9,* and *T11*) were initially selected to reflect earlier models by Garcia-Souto and Dabnichki (2013) (**Table 1**). Additional NNs were developed using different sets of inputs according to the following two variations to investigate its effect:

- removal of *core temperature* on the prediction of all local skin temperatures (locations T0-T11) as it is hard to obtain; and
- add *Average Clothing,* i.e. the average number of clothing layers covering the body (equation 2), substituting *Local Clothing*, i.e. number of layers covering the specific location, for the prediction of skin temperature at typically covered locations (T1, T3, T5, T9 and T11); or as an extra parameter for the prediction of forehead (T0) , hands (T7), and core (Tc) temperature.

These alternative NNs are summarized in **Table 2**.

$$Average\ Clothing = \frac{T0_{clothing\ (cl)} + T1_{cl} + T3_{cl} + T5_{cl} + T7_{cl} + T9_{cl} + T11_{cl}}{7} \qquad \text{Equation 2}$$

Output	Input Parameters	No. of Parameters
Core	Room Temperature, Humidity, Age, BMI, Gender	5
Forehead (T0)	Room Temperature, Humidity, Core, Age, BMI, Gender	6
Chest (T1)	Room Temperature, Humidity, Core, Age, BMI, Gender, Local Clothing *	7
Upper Arms (T3)	Room Temperature, Humidity, Core, Age, Gender, Local Clothing	6
Abdomen (T5)	Room Temperature, Humidity, Core, Age, BMI, Gender, Local Clothing	7
Hands (T7)	Room Temperature, Humidity, Core, Age, BMI, Gender	6
Knees (T9)	Room Temperature, Humidity, Core, BMI, Gender, Local Clothing	6
Calves (T11)	Room Temperature, Humidity, Core, Age, BMI, Gender, Local Clothing	7

Local Clothing refers to the number of layers covering the specific location

Table 1: List of input parameters used for the training of the 'original' NN for each location.

Output		Input Parameters	No. of Parameters
Core		Room Temperature, Humidity, Age, BMI, Gender, ***Average Clothing****	6
Forehead (T0)	V1	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender	5
	V2	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, ***Average Clothing***	6
Chest (T1)	V1	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, Clothing Layers	6
	V2	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, ~~Clothing Layers~~, ***Average Clothing***	6
Upper Arms (T3)	V1	Room Temperature, Humidity, ~~Core~~, Age, Gender, Clothing Layers	5
	V2	Room Temperature, Humidity, ~~Core~~, Age, Gender, ~~Clothing Layers~~, ***Average Clothing***	5
Abdomen (T5)	V1	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, Clothing Layers	6
	V2	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, ~~Clothing Layers~~, ***Average Clothing***	6
Hands (T7)	V1	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender	5
	V2	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, ***Average Clothing***	6
Knees (T9)	V1	Room Temperature, Humidity, ~~Core~~, BMI, Gender, Clothing Layers	5
	V2	Room Temperature, Humidity, ~~Core~~, BMI, Gender, ~~Clothing Layers~~, ***Average Clothing***	5
Calves (T11)	V1	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, Clothing Layers	6
	V2	Room Temperature, Humidity, ~~Core~~, Age, BMI, Gender, ~~Clothing Layers~~, ***Average Clothing***	6

Average Clothing refers to the average number of layers covering the body as calculated by equation 2

Table 2: List of input parameters used for the training of the alternative NN for each location. Strikethrough: removed parameters. Bold: Added parameter, in cases in exchange for *Local Clothing*.

2.2.2. Studied topologies

Neural networks can have different topologies, defined by the number of hidden layers and the number of nodes. Hidden layers transfer relationships through the network in a non-linear manner using the transfer function and using the nodes to map this relationship into and out of one layer to the next [36]. Hence choosing the right number of hidden layers and nodes is important in minimising error, i.e. less hidden layers and the network will not train appropriately, more and it will overfit. Overfitting occurs when the NN memorises data rather than establishing a relationship between the input data and output. This causes the network to not generalise well, and perform poorly when introduced to new data [36].

Various topologies for the NN were investigated in this study (all combinations of 1 or 2 hidden layers and 1 to 20 nodes, i.e. 420 different topologies) using the logistic-sigmoid (or logsig) activation function between layers –this relates the output parameters with the input parameters of the NN-, and their performance evaluated. Each NN was trained recursively (using various epochs), assessing its performance and updating the weights for the 'logsig' activation function after each epoch until the fitting error measured as MSE (equation 3) was minimized. Using the areas of low error valley, the optimal number of hidden layers and nodes were established for each network based on the point of least MSE, hereby the optimal topology. This simple criteria previously used in the literature [36] was considered sufficient at this stage, although there are other more advanced alternatives [38–40]The networks of optimal topology were further studied, as described in the following section.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \qquad\qquad \text{Equation 3}$$

2.2.3. Training Algorithm

A network with its optimal topology (identified in the previous section) was generated for each temperature location (core and each of the local skin temperatures) and tested against two training algorithms, both based on the back-propagation algorithm. The selected algorithms were the Levemberg – Marquardt back-propagation, which is the fastest and default back-propagation algorithm for NN training, and the Bayesian back-propagation, which is safer and better at generalisation.

The Levemberg – Marquardt back-propagation algorithm uses the Levemberg – Marquardt optimization to update the weights and bias within the network. Its goal is to minimize the MSE between the output of the network and the target outputs, and therefore the algorithm continues the training iteratively until the MSE is reached. When using this algorithm it is best practice to divide the data available for training into 3 further cohorts, called algorithm training, algorithm testing and algorithm validation set respectively. This approach was used to control the number of epochs that take place during training (i.e. control the training iterations required), as too few will fail to train the network and too many would cause the network to overfit the data. However, when data provided to the network is divided in this fashion, the network monitors the training error with respect to the validation and testing set instead of just minimising it at any cost (see Figure 3). The training continues until the error between the two sets (training and validation) and the validation error start to increase (overfitting). In that moment the training stops and the best performing weights (minimum validation error) are restored. In this paper, the data available for training (90% of our entire data) was divided into the following sets i) algorithm training - 70% (63% of all data); ,ii) algorithm testing - 15% (13.5% of all data) and iii) algorithm validation 15% - (13.5% of all data).

The Bayesian back-propagation updates the weights while the NN is trained according to Levemberg – Marquardt algorithm, while minimizing the square error simultaneously until it reaches an optimal combination for the network to perform and generalise. This algorithm has already a built-in validation function that monitor the process of minimising the error and therefore the original division of the data into a training set (90% in this study) and test set (10%) is sufficient.
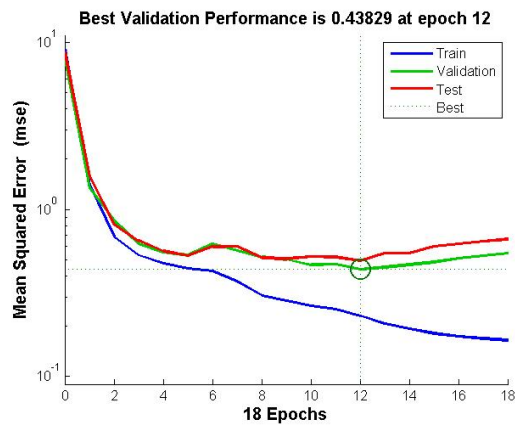

Figure 3. Example of network training using
Levemberg – Marquardt back-propagation
algorithm with training, validation, and testing sets.

Overfitting is always an issue to be considered. There are three main ways of improving generalisation and reducing overfitting when dealing with a small dataset: regularization, early stopping, and cross validation [41] However, it is important to note that there is little to no chance of overfitting when dealing with a large dataset i.e. when the number of parameters in a network is much smaller than the number of data points, as the network is exposed to sufficient variation to generalise well [36] which is the presented case.

2.2.4. Assessment of the neural networks

The performance of the NNs was first assessed by computing the error between predicted values (outputs of the networks) and target values for the training and test set of the data. This helped to determine the prediction accuracy and observe if the error is clustered around zero i.e. smaller prediction errors.

The performance of the NN during training was assessed using the MSE (equation 3), which was extracted from the network for the overall training, and specifically for the training, validation and test stage of the algorithm. A similar and low MSE in all stages of the algorithm indicates good training and consistency.

The performance of the NN with 'unseen' data was assessed by means of the average prediction Percentage Error (PE) (equation 4), which compares the predicted and actual real values, and using the remaining 10% of our data, which was 'unseen' data to the network.

The error distribution was then studied for the 'unseen' data. Ideally the error distribution would be normal, and therefore the skewness and kurtosis values were investigated. They respectively indicate how symmetric and how peaked/flat the distribution is , and have values of zero and three respectively for a normal distribution [41,42]. However, a higher kurtosis is expected since the original temperature data do not follow a normal distribution cause by the interconnected nature of the temperature points that makes them not truly random.

$$P.E. = \frac{|Actual - Predicted|}{Actual} \times 100\%$$

3. Results

3.1. Topology Results – original NNs

Various topologies for the NN were investigated in this study (1-2 hidden layers and 1-20 nodes, i.e. 420 topologies) as detailed in section 2.2.2. The MSE was computed and studied for all combinations, **Figure 4** being the example for one location. All locations presented large valleys of minimum error (MSE) when the number of hidden layers was two. The NN topologies with lower MSE for each of the temperature locations were selected as 'optimal' and are described in **Table A-1** and labelled as *v0*. The 2 layers had in the range of 10-16 and 8-14 nodes respectively.
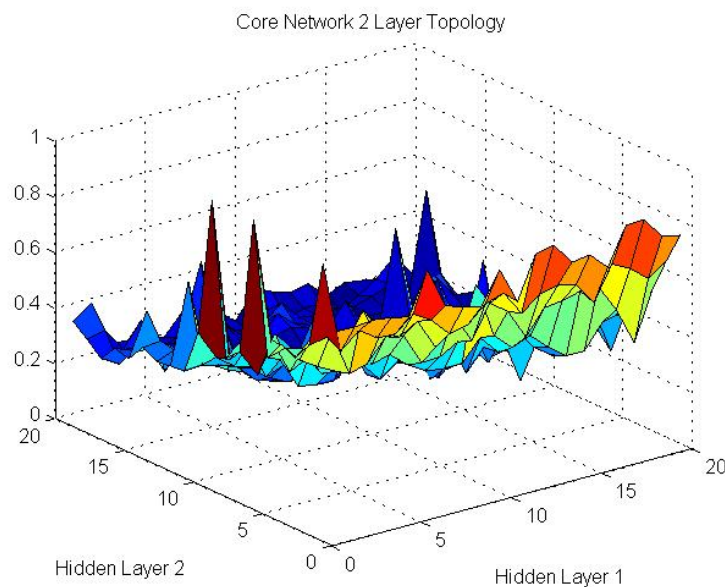

Figure 4: Example of a topology graph. Layer 1 [i] and 2 [j] show the number of nodes for the corresponding hidden layer plotted against the MSE. Blue areas of the graph indicate valleys of minimum error, which is where the combination of nodes with the lowest MSE will lie. Areas of increasing error values are represented with green, yellow, orange and red colours, the later representing the highest error.

3.2. Developed NNs

Neural Networks were created for the prediction of core and local skin temperature when using as inputs the parameters suggested by Garcia-Souto and Dabnichki [4]. Two different instances of neural networks using the identified optimal topology were generated for the each of the temperature locations (core and each local skin temperature) following two different training algorithms as described in section 2.2.3, i.e. the Bayesian back-propagation and the Levemberg – Marquardt back-propagation.

Alternative networks were developed for core and each local skin temperature in the same fashion, this time changing the parameters used as inputs. The variations were either (a) *core temperature* was removed as input from the original input set, or (b) the *average clothing* (**equation 2**) was used for forehead, hands and

core or instead of *local clothing* for chest, upper arms, abdomen, knees and calves. The summary of inputs used for these alternative NNs is presented in **Table 2**.

Therefore, a total of 19,400 NNs were created for the present study, broken down as follows: 2520 NNs for each of the seven skin locations (= 420 topologies * 2 training algorithms * 3 sets of input parameters), and 1760 NNs for core temperature (= 420 topologies * 2 training algorithms * 2 sets of input parameters). After initial assessment, one NN was selected for each of the combinations of set of inputs and desired output parameter by choosing the topology with the lowest MSE in each case, i.e. 23 NNs.

Each of these 23 pre-selected NNs were studied in depth (see **appendix A**) and compared. Their performance during training (overall and with respect to the training, validation and testing sub-sets) is summarized in **Table A-1** (original NNs labelled as *v0*, alternative NNs labelled as *v1* and *v2*). These NNs were then evaluated using the *'unseen data'*. Average Percentage Error (PE) and the goodness of fit value (R-value) are presented in **Table A-2**, while descriptive statistics and 80% and 90% confidence intervals for the error associated with the prediction of *'unseen data'* are given in **Table A-3**. Regression graph of the predicted vs. target temperatures from simulating the network with the *'unseen data' are presented* in **Figure A-1**. Boxplots are provided for the prediction errors observed both with the *'training data'* and the *'unseen data'* in **Figure A-2** for comparison.

3.3. Comparison of pre-selected NNs with existing models, and recommendations

The various pre-selected NNs reported in this paper were compared. A summary of the coefficient of fit (R-value) between the prediction of 'unseen' data and actual data is given in **Table 3**. The use of a different set of input parameters while constructing the NNs was assessed for each of the desired outputs. The recommended NNs are highlighted in **Table 3**, described in **Table 4**, and provided in the online **Appendix B**. Graphical representation of the recommended NNs' performance is given in **Figure 5**. The use of the recommended NNs was compared with the statistical models presented in the literature [4] and achieved improvement is reported in **Table 5**. No other studies providing a model for local skin temperature were found, although local skin temperature has been used as input for the prediction of other physiological parameters (e.g. heat loss, mean skin temperature, mean body temperature, etc.). Some other models within the field are given in **Table 6** for comparison.

Model	Tc	T0	T1	T3	T5	T7	T9	T11
NNs/Original set	0.846	0.843	0.832	0.795	0.816	0.854	0.873	0.779
NNs/Original set – {T Core}	--	0.728	**0.866**	0.690	**0.875**	0.889	**0.896**	0.720
NNs/Original set + {Average Clothing}	**0.870**	--	--	--	--	--	--	--
NNs/Original set – {T Core} + {Average Clothing }	--	**0.845**	--	--	--	**0.931**	--	--
NNs/Original set – {T Core} – {Local Clothing} + {Average Clothing}	--	--	0.690	**0.766**	0.790	--	0.891	**0.845**

Table 3: Comparison of the R-values for the NNs of optimal topology when using each of the given sets of input parameters. Bold number indicates recommended model for the location.

Networks	Core	Forehead (T0)	Chest (T1)	Upper Arms (T3)	Abdomen (T5)	Hands (T7)	Knees (T9)	Calves (T11)
Input parameters	Room temp; Humidity; Age; BMI; Gender; Avg clothing.	Room temp; Humidity; Age; BMI; Gender; Avg clothing	Room temp; Humidity; Age; BMI; Gender; Local clothing	Room temp; Humidity; Age; BMI; Gender; Avg clothing.	Room temp; Humidity; Age; BMI; Gender; Local clothing	Room temp; Humidity; Age; BMI; Gender; Avg clothing.	Room temp; Humidity; BMI; Gender; Avg clothing.	Room temp; Humidity; Age; BMI; Gender; Avg clothing.
Topology: 2 layers with [i, j] hidden neurons	[12,14]	[15,13]	[14,15]	[14,14]	[14,9]	[14,16]	[12,17]	[12,9]
Training Algorithm	L – M	L-M	L-M	L-M	L-M	L-M	L-M	L-M
Linear regression	Y=0.9X + 4.9	Y=0.7X+9.8	Y=0.8X+7	Y=0.6X+13	Y=0.8X+8.2	Y=0.9X+2.5	Y=0.9X+3.8	Y=0.7X+9
Precision*	0.46	1.3	1.2	1.5	1.8	2.5	2.0	1.2
Ground Truth*	0.44	1.5	1.3	1.8	2.1	2.6	2.4	1.5
R-value	0.870	0.845	0.866	0.766	0.875	0.931	0.896	0.845

Table 4: Description of the recommended Neural Networks for the prediction of core temperature and local skin temperature at various locations. The logistic-sigmoid (or logsig) activation function was used between layers.

* Precision and Ground Truth values are the approx. range needed to enclose 80%CI of the predicted values from the target temperatures and the measured values respectively for the unseen data.

Model	Tc	T0	T1	T3	T5	T7	T9	T11
Multi-linear/ Original set [Garcia-Souto & Dabnichki 2013]	0.48	0.64	0.53	0.68	0.51	0.76	0.79	0.64
Improvement between recommended NN model and multi-linear model.	81%	32%	63%	13%	72%	23%	13%	32%

Table 5: Comparison of efficiency between statistical body temperature models from the literature [4] and those reported in this paper using the recommended NNs.
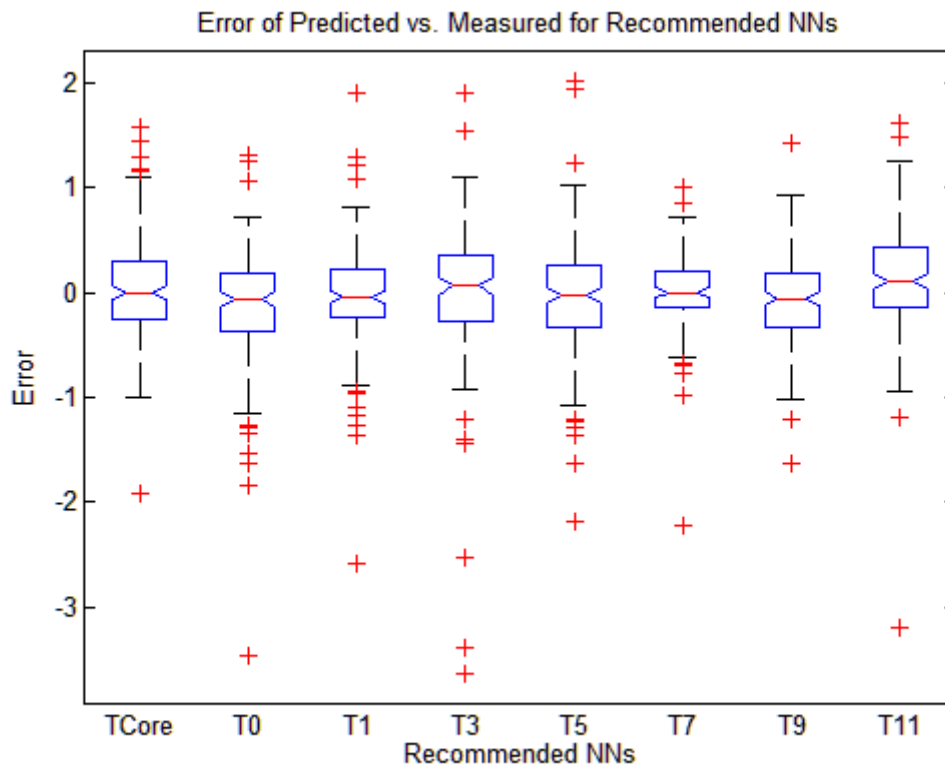
Figure 5: Performance of the recommended NNs. Error values are calculated as the predicted minus actual value for 'unseen data'. Box represents mean, 25 and 75% quartiles; whisker length represents 1.5 times interquartile range; red crosses represent outliers.

Study	Experimental protocol	Predicted parameter	Input parameters	Predicted vs Measured	Model performance
[Nielsen, 1984]	10 clothed, resting subjects; T room = 10°C; various formulas tested against formula by Olesen et al. (1972)	Mean Skin Temperature	1 – 12 local temperature points	<u>Pre-existing models:</u> R^2 = 0.28 (4 points) – 0.90 (12 points) <u>New models:</u> R^2 = 0.571 (1 point) - R^2 = 1.0 (12 points)	<u>Pre-existing models (examples):</u> *Mitchell and Wyndham was first with 90% within +/- 0.5 °C *Burton (1935) last with 23% within +/- 0.5 °C <u>New models:</u> 7 – 12 measuring points had 100% within +/- 0.5 °C <u>Conclusion:</u> No formula including less than 7 points gave a reasonable accurate results
[Mehnert, 2000]	Large dataset; Nude vs clothed. Environmental conditions: Radiation = {normal or high}; Humidity = {normal or high}; several	Mean Skin Temperature	T air; T radiant; partial vapour pressure; air velocity; T rectal	<u>Nude:</u> T_{obs} = 1.02* T_{pred} – 0.65 R ∈ (0.76, 0.89) <u>Clothed:</u> T_{obs} = 0.99* T_{pred} –	<u>Nude:</u> 83.3% data points within +/- 1°C 95.7% data points within +/- 1.5°C <u>Clothed:</u> 81.8% data points within +/- 1°C

				0.33 R ∈ (0.67, 0.80)	94.8% data points within +/- 1.5°C
[Sanders, 1999]	9 subjects; walking on treadmill; two simulated tropical environ; soldiers trained in heat; with uniform and carrying equipment; Neural Networks	T core	Nude weight, height; length of acclimatization; insulation; permeability; T room; relative humidity; wind speed; total metabolic rate	R = 0.19 at 30°C and 60% RH R = 0.34 at 40 °C and 30% RH	*Specificity and sensitivity indicate that NN had low predicted precision. *Regression analysis indicates positive but weak correlation. *Concluded that the use of NNs for predicting human physiological responses is worth further investigation.
[Lenhardt, 2006]	44 subjects; undergoing bypass vs healthy volunteers; T core ∈ [18,36.5]°C.	Mean Body Temperature (MBT) using Burton's formula	Tcore; T skin average	$Y = 0.94 \cdot X + 2.15$ $R^2=0.98$	87% estimates within ±0.5°C, 92% estimates within ±0.7°C
[Jay, 2007]	60 clothed, healthy subjects; cycling 60-90 minutes – until steady state T core;T room = {24, 30}°C, Humidity ={30, 60}%; various T core estimation methods	Change of MBT after exercise (ΔMBT)	Δ T_{rectal} ; Δ $T_{skin\ average}$	Model 1: R = 0.45 after 90 mins. Model 2: R ∈ (0.33, 0.75) depending on exercise time and T core estimation method.	Model 2: Mean percentage error with 95% CI in brackets, e.g. +15.0% (+27.2, +2.8) to −13.7% (−4.2, −23.3) after 90 min

Table 6: Other predicting models examples in the field of thermoregulation

4. Discussion

4.1. General study of the created NNs

As already outlined topologies with1 or 2 hidden layers and 1-20 nodes per layer were considered, and the optimal topology identified based on the point of least MSE. The number of hidden layers and nodes per layer was not increased any further as overfitting was observed as expected as the studied data has relatively few input parameters. Data proved to follow strongly non-linear relationship with poor performance with 1 hidden layer. Two hidden layers performed better as expected because of the data volume used for training (approximately 6 parameters and 1451 entries). The 'optimal topology' was consistent across the various locations, and characterized by 2 hidden layers and an average of 12 nodes per layer.

Two training back-propagation algorithms were tested, i.e. the Bayesian and the Levemberg – Marquardt. Both algorithms showed very similar performance (MSE) in every instance, however the Levemberg –

Marquardt algorithm was finally used for all networks as its training time was shorter (~10s for L-M versus ~1min for Bayesian).

The taken precautions and the large data set and few input parameters allowed to prevent over-fitting of the NNs. Nevertheless the NNs were generated using validation datasets to monitor the error during training and avoid over-fitting, i.e. creating neural network that generalize well.

4.2. Original NNs

Initially eight NNs were created to predict core temperature, and local skin temperature at the forehead, chest, upper arms, abdomen, knees and calves based on input parameters previously selected by Garcia-Souto and Dabnichki [4] (summarized in **Table 1**). Cross validation demonstrated a good performance (R-values between 0.78 and 0.87) (**Table A-2** labelled as *v0*) indicating that *'unseen data'* was accurately predicted throughout the networks. Similar performance results were observed during training, yet slightly poorer as expected, deeming it successful. There was no indication of overfitting.

The error distributions were approximately symmetrical around the zero error value for all networks. A small skewness value was observed for core, T0, T1, T3, T7 and T9 temperature locations (between -0.5 and 0.5) and slightly larger for abdomen and calves (but only up to -1.15) (**Table A-3, Figure A-2**). Kurtosis values were relatively higher than those of a normal distribution (e.g. between 5.8 and 6.3 for chest, calves and abdomen skin temperature). This is positive as it indicates that the peak around zero is higher and therefore the error for the majority of the predictions is smaller. High kurtosis could have been a sign of over-fitting but it was discarded following the assessment of the MSE valleys. The standard deviation and 80% confidence intervals observed in the error histograms were relatively small, with a minimum for forehead and knees (SD=0.47°C, 80% CI=1.2°C) and a maximum of for the upper arms skin temperature (SD=0.66°C, 80% CI=1.6°C).

These results are typical for thermoregulation studies rather than reflecting limitations of the NNs. The error observed is partly due to the diversity of temperatures at a specific collection point, which varies depending on the person, environmental conditions and also personal circumstances such as clothing and material and others of psychological and physiological nature. The influence of these factors changes with the location, and are cross-related, making some locations more predictable than others. The descriptive statistics for the experimental data showed that the type of distribution and variation observed on the outputs of each network reflected the experimental data. It was then concluded that the networks were trained successfully and generate accurate predictions.

The obtained skin temperature errors in the predictions are highly satisfactory as the data itself has a large variability and the localised body response is to some degree stochastic. The error for core temperature (SD= 0.55°C) is more than satisfactory for non-clinical applications showing that the models would be able to discern between high and low core temperatures. Surprisingly NN for knees had a similarly good performance as the forehead (limbs are periphery and temperature is prone to higher degree of variation as the prime target of the thermoregulation system is constant core temperature). This is due to the range of habitation comfort level of room temperature. Therefore this location could be considered for human temperature monitoring and as an indicator in body temperature prediction models.

4.3. Alternative NNs

Secondly, alternative NNs were created where the set of inputs was slightly modified from the original proposition, as summarized in **Table 2**. Similarly to previously created NNs, the ability to predict *'unseen data'*

was good, with R-values in the range of 0.69-0.90 when removing the *core temperature* as an input, and 0.69-0.93 when using *Average Clothing* in addition to the initial set of inputs or instead of the *Local Clothing* (**Table A-2** labelled as *v1* and *v2*).The error distributions were also approximately symmetrical around the zero error value for all networks and presented high Kurtosis and slightly negative Skewness values (**Table A-3**, **Figure A-2**), in similar fashion to the previously discussed NNs.

4.4. Comparison of models

Firstly it is important to note that human body temperature, and especially local skin temperature, depends on a number of parameters, with many interdependencies making it difficult to model to a (high) level of accuracy achieved in other applications. There is a great variability on temperature across the groups and intra-individual ones even when subjected to identical climate conditions, with SD of 0.37°C for core temperature, and SD of local skin temperature ranging from 0.7-0.8°C at forehead to 1.4-1.8°C for hands, these values being applicable to both the adult and toddlers population [43]. Furthermore, for practical reasons the factors that were considered within the models in this paper are only *environmental temperature, environmental humidity, gender, age, clothing and BMI.* These are "easy" to ascertain, are appropriate for indoor environments and render the models more usable in climate control systems, e.g. in offices or long-haul flights. However there are many other relevant factors that were not considered but could explain in part the variability, e.g. metabolic rate, sweating rate, level of acclimation to the environment, level of fitness, stress, etc. Therefore the models presented here, in the same way as other alternatives in the literature, for *core temperature* or *local skin temperature* are limited by the nature of the very parameter we aim to predict. For example, the models developed by Mehnert et al [18] for prediction of *average skin temperature* have R of 0.86 for nude people, and 0.77 for clothed people, with a ± 1.0°C 83.3 and 81.8% of the data points respectively. This is a similar performance to what the models presented here have achieved even though the average skin temperature is more consistent than local skin temperature. Another example is the models developed by Sanders et al [34] for prediction of core temperature for volunteers in a treadmill, with R of 0.19 and 0.34 at 30 and 40°C respectively. Therefore, the accuracy of the models presented in this paper are modest as expected, and yet better than earlier approaches, hence they are of value to a better understanding of thermoregulation and for the development of indoor climate control systems.

The original set of inputs for the NNs developed within this paper were 6-7 parameters accounting for personal characteristics and external factors that were easy and practical to collect. Core temperature was included as an input for the prediction of local skin temperature as it improves the accuracy [14]. However, core temperature is more difficult to measure and therefore was eliminated of the alternative NN models to make them more practical. The prediction of the '*unseen data*' (see R-values in **Table 3**) decreased as expected for skin temperature at forehead, upper arms, and calves a 14, 13 and 7.6% respectively. However, the prediction for chest, abdomen, hands and knees slightly improved (R-values increased between 2.6 and 7.2%), indicating that core temperature values do not provide a significant addition to the set of input parameters when predicting these skin temperatures, for the studied environmental conditions. The 80%CI only becomes larger for forehead (14.8% increase) but otherwise it remains constant or decreases (up to 13.2% at chest location), making the models more precise.

The NN for core temperature improved when *Average Clothing* was included as an additional input. The R-value increased just slightly from 0.85 to 0.87. However the SD and the 80% confidence interval decreased significantly, about 10 and 21% respectively from the original NN, to values of 0.50 and 1.15°C. The addition of *Average Clothing* to the NNs for forehead and hand skin temperature where the *core temperature* was not included had also a very positive effect. R-value increased 16 and 4.7% respectively, while the 80%CI decreased to 3.8 and 31% respectively, making the models more accurate and precise. *Local Clothing* was included as an input parameter in the original NNs for chest, upper arms, abdomen, knees and calves. When

Average Clothing was used instead, the efficiency of the NNs only improved for upper arms and calves (11 and 17% increase for R-values and 6 and 11% decrease of 80%CI respectively). For the other locations the R-value remained fairly identical but the 80%CI became larger.

Two approaches for the creating of body temperature models were compared: multi-linear regression models vs NNs. When using the same original set of inputs, the models generated with NNs produce significantly (p-value 0.0003 at 1% significance level) more accurate predictions, with a mean R value of 0.63 for the multi-linear models and 0.83 for the equivalent NNs. The greater improvement is observed for core temperature, with an increase from 0.48 to 0.85, i.e. an improvement of about 75% over the initial value. The ranking of locations with the best and worse predictions is very similar, with the best models being for knees (T9) and hands (T7).

The recommended NNs, i.e. with the recommended set of parameters, have a very significant increase on R-value with respect to the multi-linear regression model: 81% for core temperature, 72% for abdomen skin temperature, 63% for chest temperature, 32% for calves and forehead temperature and 13% for upper arms and knees skin temperature (**Table 3**). Therefore, the use of NNs for the prediction of body core and local skin temperature are demonstrated to be more suitable and they are recommended. The 80%CI values for the predictions by the recommended NNs and for the experimental 'unseen' data (ground truth) are very similar and slightly smaller for the predicted values. This indicates that the models are sensible and perhaps a bit conservative.

5. Conclusion

This paper demonstrated that core body and local skin temperatures (forehead, chest, upper arms, abdomen, knees and calves) can be successfully predicted by means of Neural Networks for both genders and in a wide range of every-day indoor situations. This paper has therefore opened the possibility of using AI in the modelling of thermoregulation and potentially for climatic control of shared public spaces.

Some of the main results and characteristics of the ultimately selected NN per location are:
- NNs were trained and tested using a large and gender-balanced experimental database on real office-type situations, covering a wide range of environmental temperatures (18 to 31°C) and humidity values (22 to 60%) (previously reported by Garcia-Souto and Dabnichki [4]).
- NNs have 2 hidden layers and 8-16 nodes per layer, and were trained with the Levemberg – Marquardt back-propagation algorithm.
- The input parameters included *room temperature*, and *humidity* in all cases, and additionally all or a subset of the following: *age*, *BMI*, *gender* and *clothing* (either local or average).
- NNs presented very good results when assessed with 'unseen' data: R=0.87 for core temperature and in the range of 0.85 and 0.93 for local skin temperature at forehead, chest, abdomen, calves, knees and hands, with the latter 2 holding the strongest predictions. Upper arms holds the lower performance (R = 0.77).
- The estimated precision for core temperature was 0.46°C at an 80%CI, which is quite acceptable for non-clinical applications. For skin temperature the precision was best at forehead, chest and calves (±1.2°C, similar to the precision achieved by Mehnert et al [18] for *average skin temperature*) and still acceptable for the other local temperatures (between ±1.5 and ±2.5°C).

In comparison with previous models [4]:
- *Core temperature* has been withdrawn as an input for the prediction of local skin temperatures as it is impractical to measure and advantageously it is a non-contributor to the NN output.

- *Average Clothing* as input is beneficial for the prediction of core, forehead and hands temperature, and should be used instead of the *Local Clothing* for upper arms and calves.
- Accuracy using NNs are significantly better than the previously-reported multi-linear models [4] using the same dataset and same (and even less) input parameters: 81% raise of R-value for core temperature and in the range of 13-72% for local skin temperatures, particularly improving the core, forehead, chest, abdomen and calves temperature predictions.
- NNs provide better precision but still error is observed, partly due to the diversity of temperatures characteristic of the human thermoregulation, which cannot be overcome by any model.

Acknowledgements

Appendix A – Performance Results for all pre-selected NNs

Results presented in this appendix correspond to the 23 pre-selected Neural Networks for the prediction of core and local skin temperature that were studied in depth out of the 19,400 NNs that were created. From these, the NNs which used as inputs the parameters in Garcia-Souto and Dabnichki [4] (summarized in **Table 1**) were named as 'original' and indicated by "v0". Those NNs using a variation of the set of input parameters (according to **Table 2**) were named as 'modifications' and indicated by "v1" and "v2".

The performance of the NNs during training is summarized in **Table A-1**. The NNs evaluation using *'unseen data'* is presented in **Table A-2** (e.g. average Percentage Error, and the goodness of fit value) and in **Table A-3** (e.g. descriptive statistics and confidence intervals of associated error). Regression graph of the predicted vs. target temperatures from simulating the network with the *'unseen data' are presented* in **Figure A-1**. Boxplots are provided for the prediction errors observed both with the *'training data'* and the *'unseen data'* in **Figure A-2** for comparison.

Networks		Core	Forehead (T0)	Chest (T1)	Upper Arms (T3)	Abdomen (T5)	Hands (T7)	Knees (T9)	Calves (T11)
No.Neurons $[i, j]$	V0	$[12, 13]$	$[13, 9]$	$[16, 10]$	$[10, 13]$	$[13, 14]$	$[11, 8]$	$[15, 14]$	$[12, 10]$
	V1	$[12, 14]$	$[14, 15]$	$[14, 15]$	$[15, 14]$	$[14, 9]$	$[12, 14]$	$[12, 17]$	$[15, 12]$
	V2	---	$[15, 13]$	$[14, 16]$	$[14, 14]$	$[15, 15]$	$[14, 16]$	$[14, 12]$	$[12, 9]$
Network Perf.	V0	0.178	0.245	0.291	0.347	0.219	0.204	0.181	0.298
	V1	0.141	0.355	0.261	0.386	0.165	0.121	0.165	0.274
	V2	---	0.249	0.292	0.309	0.250	0.103	0.141	0.318
Training Perf.	V0	0.163	0.207	0.231	0.326	0.190	0.199	0.148	0.250
	V1	0.115	0.294	0.201	0.376	0.118	0.0885	0.143	0.202
	V2	---	0.181	0.228	0.261	0.211	0.0842	0.125	0.279
Validation Perf.	V0	0.219	0.329	0.518	0.402	0.240	0.225	0.228	0.416
	V1	0.208	0.591	0.459	0.317	0.287	0.150	0.239	0.359
	V2	---	0.373	0.500	0.398	0.342	0.140	0.180	0.340
Testing Perf.	V0	0.264	0.333	0.340	0.393	0.219	0.206	0.291	0.405
	V1	0.183	0.404	0.341	0.502	0.262	0.242	0.195	0.528
	V2	---	0.439	0.380	0.445	0.340	0.154	0.174	0.482
R-value (Net. Targets vs. Net. Outputs)	V0	0.906	0.873	0.843	0.804	0.884	0.891	0.908	0.836
	V1	0.927	0.798	0.861	0.784	0.911	0.936	0.915	0.855
	V2	---	0.869	0.838	0.829	0.867	0.947	0.927	0.826

Table A-1: Description and performance during training (i.e. with *'training data'* only) of the NN selected for each body location according to the models specified in **Table 1** (original NNs indicated as v0) and **Table 2** (alternative NNs indicated as v1 and v2). The training algorithm was in all cases the Levemberg – Marquardt back-propagation.

| Networks		Core	Forehead (T0)	Chest (T1)	Upper Arms (T3)	Abdomen (T5)	Hands (T7)	Knees (T9)	Calves (T11)
Prediction Avg. P.E	V0	0.408	1.17	1.32	2.11	2.11	2.77	2.39	1.78
	V1	0.363	1.61	1.11	2.16	1.82	2.50	2.19	1.72
	V2	----	1.32	1.60	1.98	1.96	1.68	2.24	1.49
R-value (Unseen Targets vs. Predictions)	V0	0.846	0.843	0.832	0.795	0.816	0.854	0.873	0.779
	V1	0.870	0.728	0.866	0.690	0.875	0.889	0.896	0.720
	V2	----	0.845	0.690	0.766	0.790	0.931	0.891	0.845
Linear Regression	V0	Y=0.7x+9.3	Y=0.8x +6.3	Y=0.8x+7.4	Y=0.7x+10	Y=0.7x+8.6	Y=0.7x+8.7	Y=0.8x+6.3	Y=0.6x+12
	V1	Y=0.9x+4.9	Y= 0.6x+15	Y=0.8x+7	Y=0.6x+14	Y =0.8x+8.2	Y=0.8x+5.7	Y=0.9x+3.8	Y=0.6x+12
	V2	---	Y=0.7x+9.8	Y=0.6x+15	Y=0.6x+13	Y=0.7x+11	Y=0.9x+2.5	Y=0.8x+4.8	Y=0.7x+9

Table A-2: Description and performance with *'unseen data'* of the NN selected for each body location according to the models specified in **Table 1** (original NNs indicated as v0) and **Table 2** (alternative NNs indicated as v1 and v2).

Networks		Core	Forehead (T0)	Chest (T1)	Upper Arms (T3)	Abdomen (T5)	Hands (T7)	Knees (T9)	Calves (T11)
Average error (°C)	V0	0.051	0.0335	0.0966	0.0483	-0.0058	0.006	-0.077	-0.908
	V1	0.0384	-0.0579	-0.0273	-0.0118	-0.0630	-0.0086	-0.0548	0.0120
	V2	---	-0.133	-0.0882	0.0134	-0.0225	-0.0031	0.0589	0.109
Kurtosis	V0	3.49	4.0	5.82	5.30	6.31	3.96	4.36	6.13
	V1	4.61	9.44	8.08	7.01	5.79	7.42	4.18	9.10
	V2	---	10.1	15.6	11.6	8.66	11.8	7.06	10.8
Skewness	V0	0.124	0.247	-0.220	0.504	-1.15	-0.149	0.352	-0.874
	V1	0.228	-1.88	-0.589	-0.325	-0.0927	0.899	-0.0237	-0.616
	V2	---	-1.62	-2.34	-1.89	-1.39	-1.50	0.817	-1.26
SD (°C)	V0	0.553	0.469	0.571	0.663	0.614	0.538	0.467	0.648
	V1	0.498	0.756	0.498	0.732	0.552	0.488	0.436	0.673
	V2	---	0.570	0.802	0.680	0.602	0.355	0.461	0.542
80% CI	V0	1.46	1.15	1.29	1.56	1.49	1.32	1.19	1.37
	V1	1.15	1.32	1.12	1.51	1.28	1.15	1.06	1.37
	V2	---	1.27	1.44	1.42	1.46	0.791	1.08	1.22
90% CI	V0	1.86	1.67	1.85	2.04	1.99	1.65	1.59	1.95
	V1	1.72	2.39	1.45	2.46	1.62	1.62	1.43	1.89
	V2	---	1.77	2.08	1.66	1.87	1.08	1.39	1.61

Table A-3: Descriptive statistics and confidence intervals of the errors between targets and predictions values of the *'unseen data'* generated by the NN selected for each body location according to the models specified in **Table 1** (original NNs indicated as v0) and **Table 2** (alternative NNs indicated as v1 and v2).

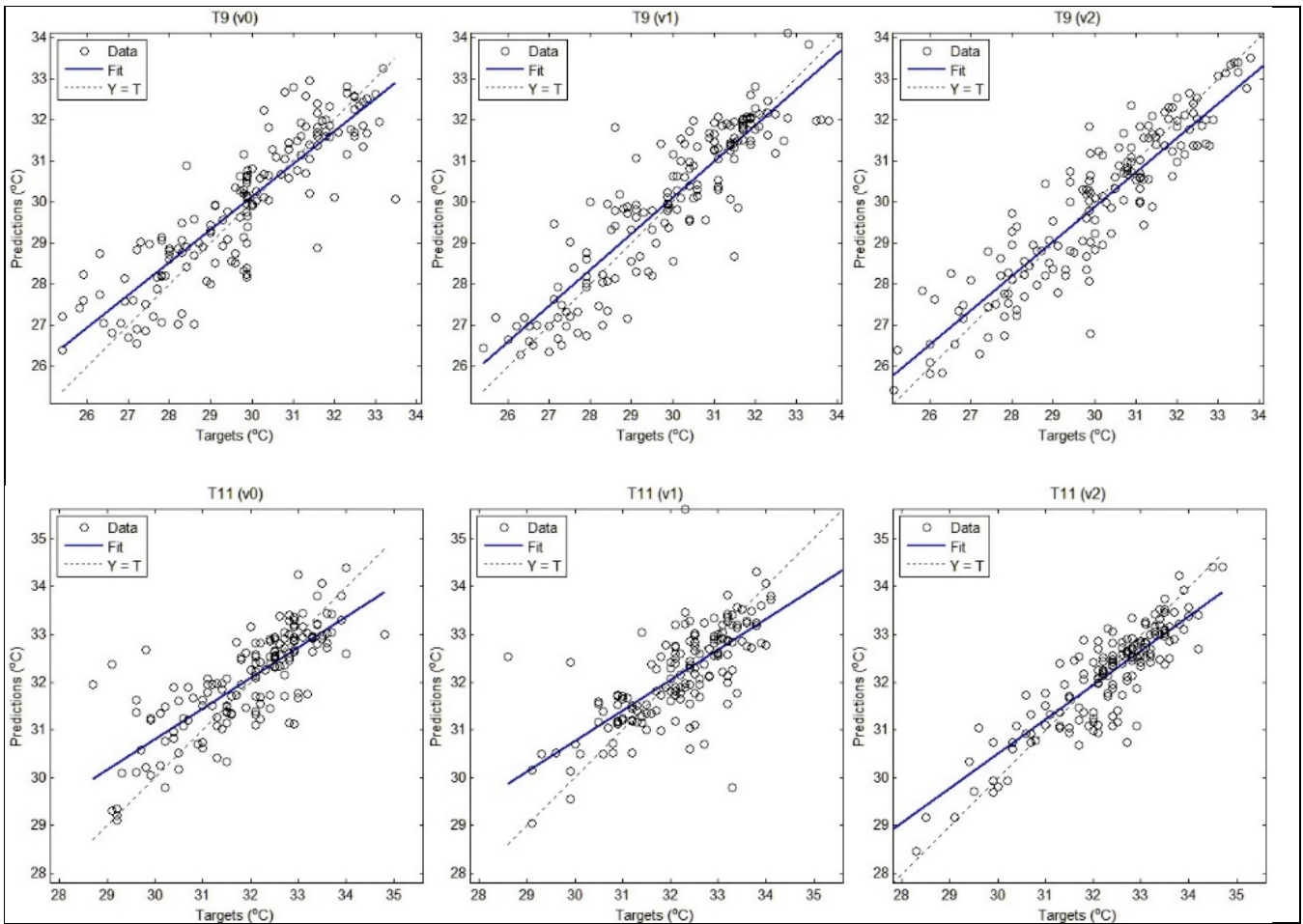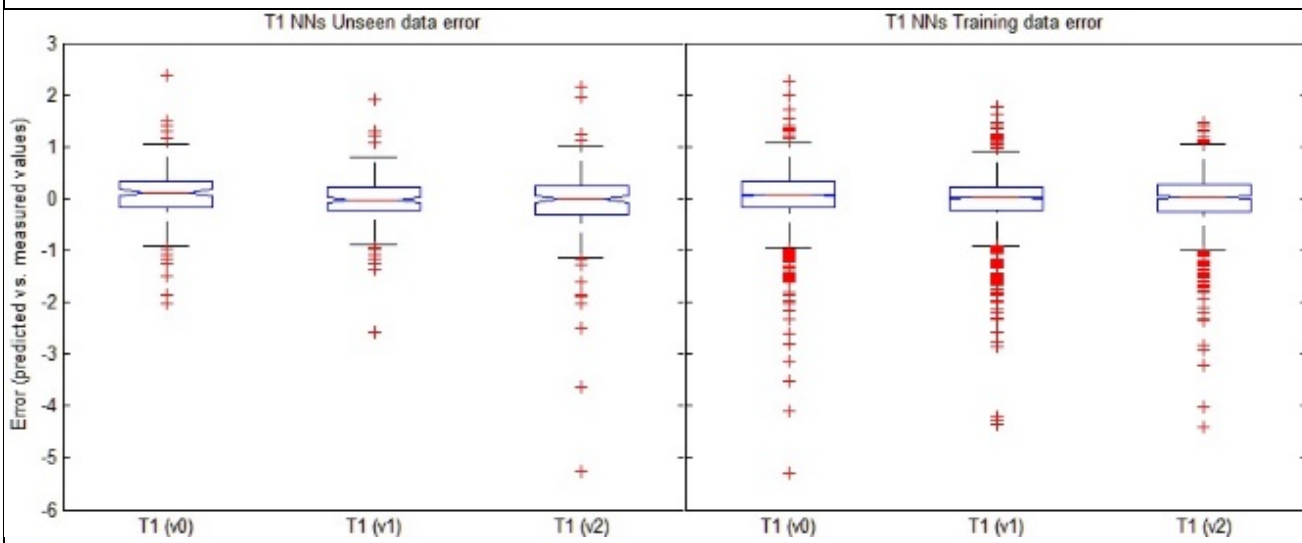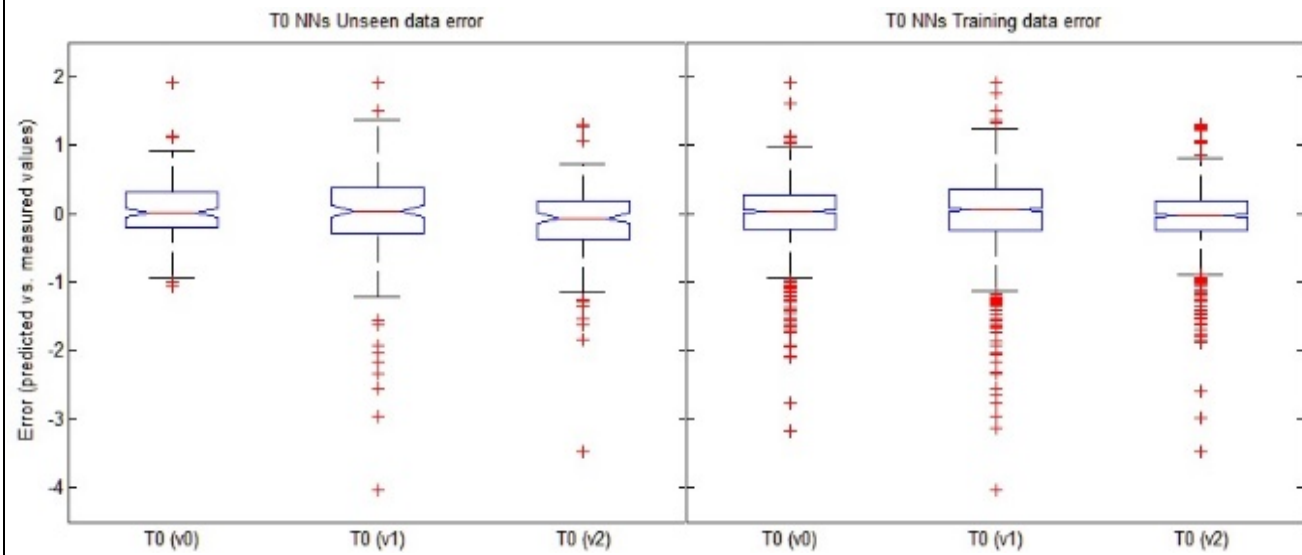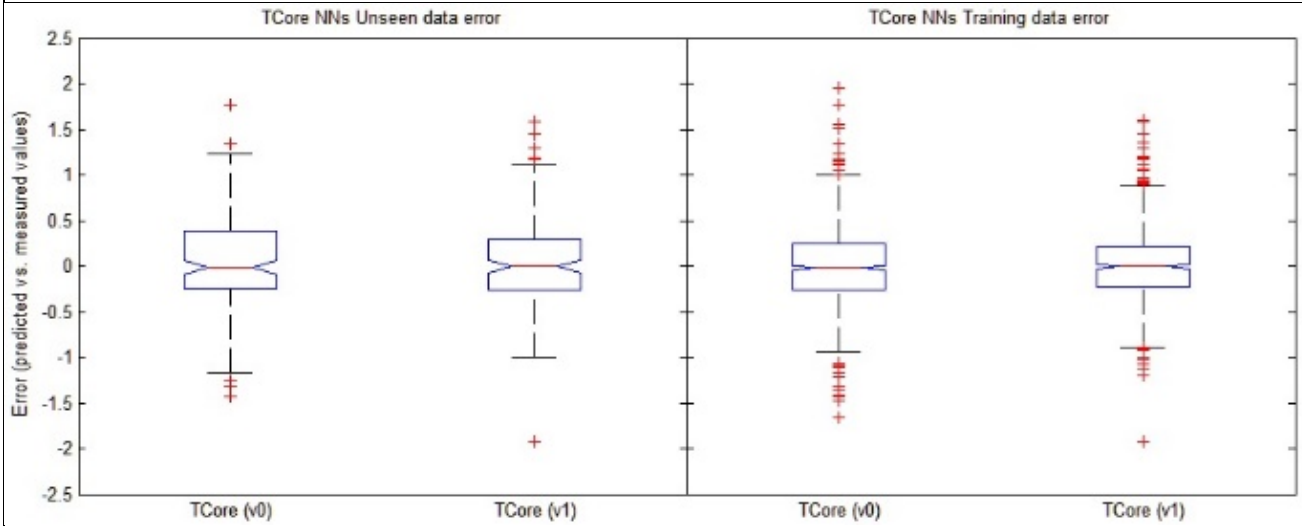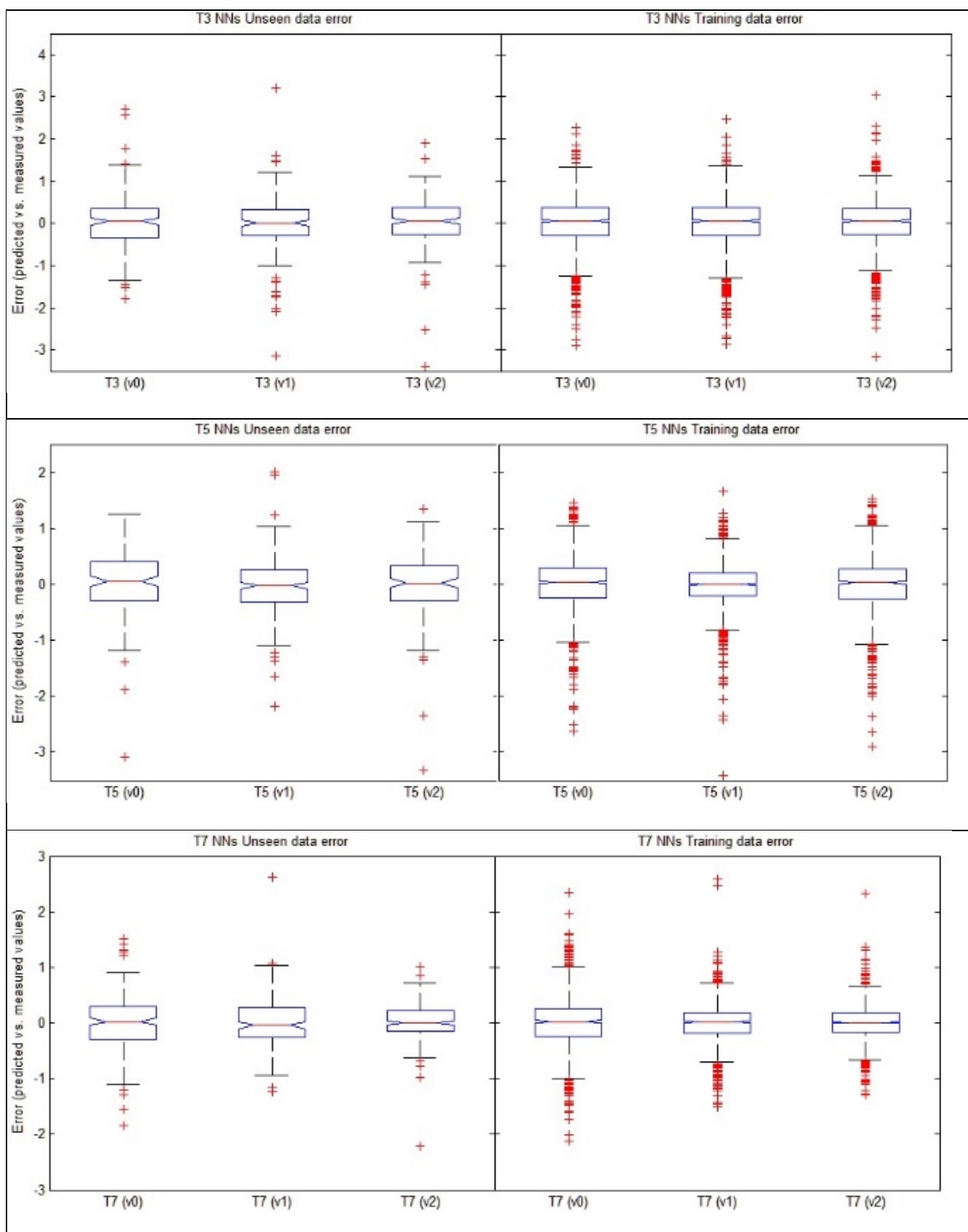**Unseen targets vs. Predictions for developed NN**
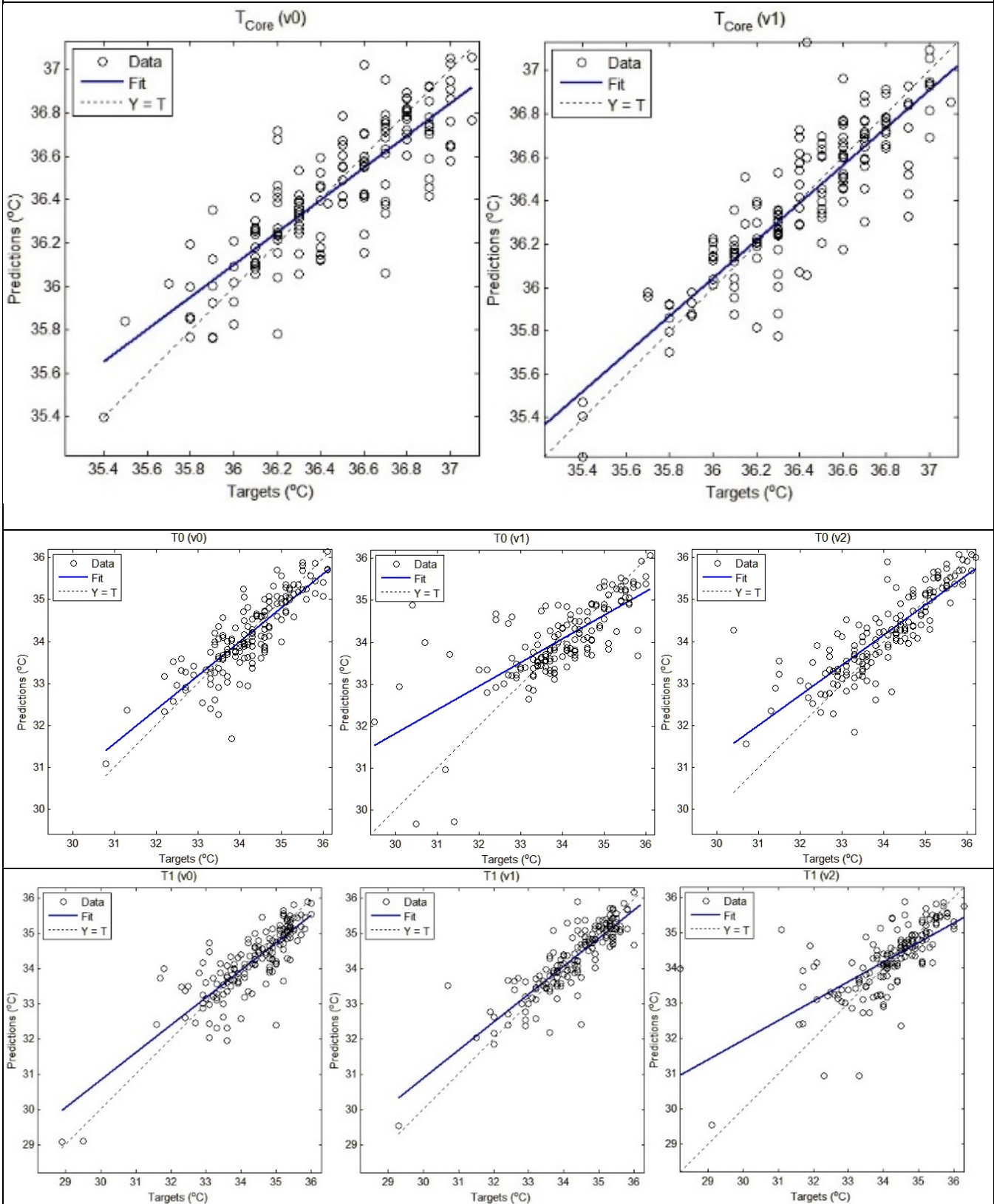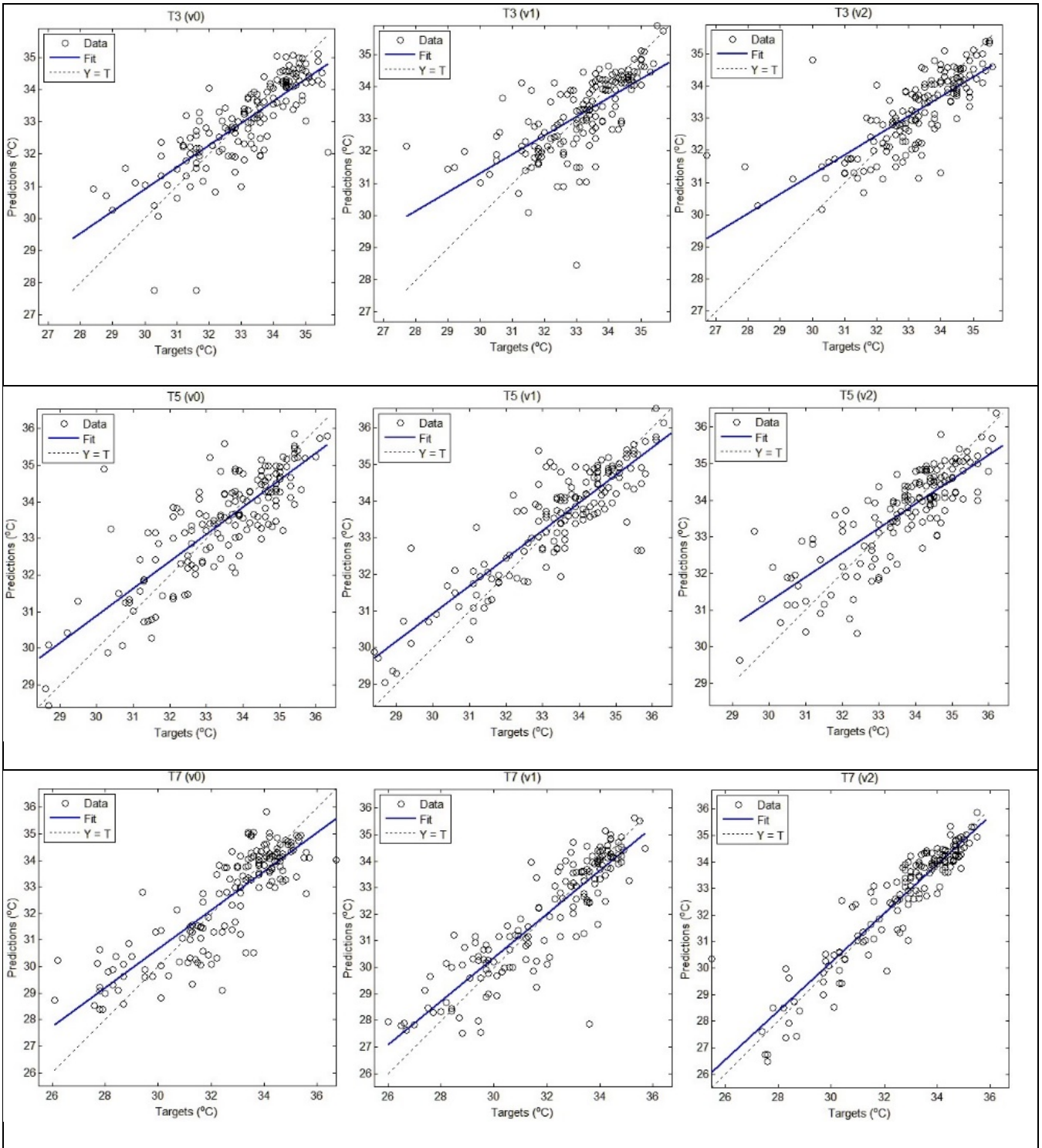
Figure A-1: Regression of the *unseen target data* against the predictions given by each developed NN. Regression equation and R-value for each fit is given in **Table A-2**. Original NNs are indicated as v0, while alternatives to these NNs are indicated as v1 and v2.
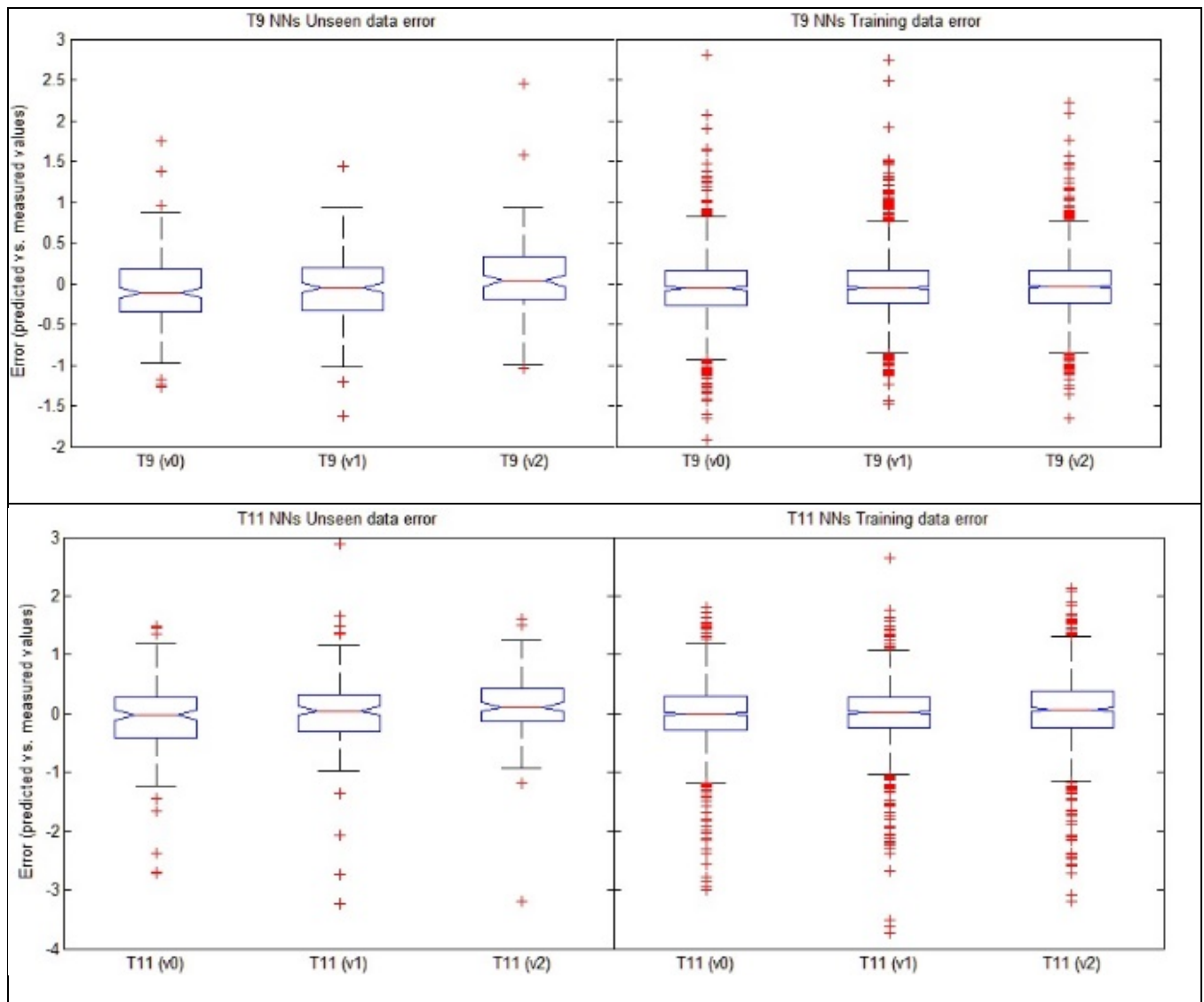
Performance of the developed NN

Figure A-2: Performance of the created NNs per location. Error values are calculated as the predicted minus actual value for 'training data' and 'unseen data' and presented separately for comparison. Box represents mean, 25 and 75% quartiles; whisker length represents 1.5 times interquartile range; red crosses represent outliers. Original NNs are indicated as v0, while alternatives to these NNs are indicated as v1 and v2.

Appendix B – Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.asoc.2016.11.006.

References

[1] M. Kosaka, M. Yamane, R. Ogai, T. Kato, N. Ohnishi, E. Simon, Human body temperature regulation in extremely stressful environment: Epidemiology and pathophysiology of heat stroke, J. Therm. Biol. 29 (2004) 495–501. doi:10.1016/j.jtherbio.2004.08.019.

[2] P.M. Ferreira, A.E. Ruano, S. Silva, E.Z.E. Concelcao, Neural networks based predictive control for thermal comfort and energy savings in public buildings, Energy Build. 55 (2012) 238–251. doi:10.1016/j.enbuild.2012.08.002.

[3] Y. Ogura, H. Naito, S. Akin, N. Ichinoseki-Sekine, M. Kurosaka, R. Kakigi, et al., Elevation of body temperature is an essential factor for exercise-increased extracellular heat shock protein 72 level in rat plasma., Am. J. Physiol. Regul. Integr. Comp. Physiol. 294 (2008) R1600–R1607. doi:10.1152/ajpregu.00581.2007.

[4] M.D.P. Garcia-Souto, P. Dabnichki, Skin temperature distribution and thermoregulatory response during prolonged seating, Build. Environ. 69 (2013) 14–21. doi:10.1016/j.buildenv.2013.07.006.

[5] H.T. Hammel, D.C. Jackson, J. a Stolwijk, J.D. Hardy, S.B. Stromme, Temperature Regulation By Hypothalamic Proportional Control With an Adjustable Set Point., J. Appl. Physiol. 18 (1963) 1146–1154.

[6] A.A. Romanovsky, Thermoregulation: some concepts have changed. Functional architecture of the thermoregulatory system., Am. J. Physiol. Regul. Integr. Comp. Physiol. 292 (2007) R37–46. doi:10.1152/ajpregu.00668.2006.

[7] O. Shido, N. Sugimoto, M. Tanabe, S. Sakurada, Core temperature and sweating onset in humans acclimated to heat given at a fixed daily time., Am. J. Physiol. 276 (1999) R1095–101. http://www.ncbi.nlm.nih.gov/pubmed/10198390.

[8] R. Nielsen, B. Nielsen, Measurement of mean skin temperature of clothed persons in cool environments, Eur. J. Appl. Physiol. Occup. Physiol. 53 (1984) 231–236. doi:10.1007/BF00776595.

[9] S. Takada, H. Kobayashi, T. Matsushita, Thermal model of human body fitted with individual characteristics of body temperature regulation, Build. Environ. 44 (2009) 463–470. doi:10.1016/j.buildenv.2008.04.007.

[10] A. Munir, S. Takada, T. Matsushita, Re-evaluation of Stolwijk's 25-node human thermal model under thermal-transient conditions: Prediction of skin temperature in low-activity conditions, Build. Environ. 44 (2009) 1777–1787. doi:10.1016/j.buildenv.2008.11.016.

[11] S. Takada, S. Hokoi, N. Kawakami, M. Kudo, Effect of Sweat Accumulation in Clothing on Transient Thermophysiological Response of Human Body To the Environment, (n.d.) 1–8.

[12] C. Huizenga, H. Zhang, T. Duan, E. Arens, An Improved Multinode Model of Human Physiology and Thermal Comfort, 9 (2001) 129–139.

[13] M. Salloum, N. Ghaddar, K. Ghali, A new transient bioheat model of the human body and its integration to clothing models, Int. J. Therm. Sci. 46 (2007) 371–384. doi:10.1016/j.ijthermalsci.2006.06.017.

[14] D. Fiala, J.K. Lomas, M. Stohrer, Computer prediction of human thermoregulatory and temperature responses to a wide range of environmental conditions, Int. J. Biometeorol. 45 (n.d.) 143–159. doi:10.1007/s004840100099.

[15] S.I. Tanabe, K. Kobayashi, J. Nakano, Y. Ozeki, M. Konishi, Evaluation of thermal comfort using combined multi-node thermoregulation (65MN) and radiation models and computational fluid dynamics (CFD), Energy Build. 34 (2002) 637–646. doi:10.1016/S0378-7788(02)00014-2.

[16]   D. Fiala, K.J. Lomas, M. Stohrer, A computer model of human thermoregulation for a wide range of environmental conditions: the passive system, J. Appl. Physiol. 87 (1999) 1957–1972. http://jap.physiology.org/content/87/5/1957.abstract.

[17]   H.O. Nilsson, I. Holmér, Comfort climate evaluation with thermal manikin methods and computer simulation models., Indoor Air. 13 (2003) 28–37. doi:10.1034/j.1600-0668.2003.01113.x.

[18]   P. Mehnert, J. Malchaire, B. Kampmann, a Piette, B. Griefahn, H. Gebhardt, Prediction of the average skin temperature in warm and hot environments., Eur. J. Appl. Physiol. 82 (2000) 52–60. doi:10.1007/s004210050651.

[19]   G.P. Zhang, E.B. Patuwo, H. Michael Y., Forecasting with artificial neural networks: The state of the art, Int. J. Forecast. 14 (1998) 35–62. doi:10.1016/S0169-2070(97)00044-7.

[20]   J.L. Grant, Short-Term Peak Demand Forecasting Using an Artificial Neural Network with Controlled Peak Demand Through Intelligent Electrical Loading, (2014).

[21]   D. Niu, H. Shi, D.D. Wu, Short-term load forecasting using bayesian neural networks learned by Hybrid Monte Carlo algorithm, Appl. Soft Comput. 12 (2012) 1822–1827. doi:10.1016/j.asoc.2011.07.001.

[22]   M. Macas, F. Moretti, F. Lauro, S. Pizzuti, A. Fonti, G. Comodi, et al., Importance of feature selection for recurrent neural network based forecasting of building thermal comfort, (n.d.) 1–10.

[23]   K. Venkatesh, V. Ravi, A. Prinzie, D. Van Den Poel, Cash demand forecasting in ATMs by clustering and neural networks, Eur. J. Oper. Res. 232 (2014) 383–392. doi:10.1016/j.ejor.2013.07.027.

[24]   F. Almonacid, P. Perez-Higueras, P. Rodrigo, L. Hontoria, Generation of ambient temperature hourly time series for some Spanish locations by artificial neural networks, Renew. Energy. 51 (2013) 285–291. doi:10.1016/j.renene.2012.09.022.

[25]   K. Patil, M.C. Deo, S. Ghosh, M. Ravichandran, Predicting Sea Surface Temperatures in the North Indian Ocean with Nonlinear Autoregressive Neural Networks, Int. J. Oceanogr. 2013 (2013) 1–11. doi:10.1155/2013/302479.

[26]   M.C. Valverde, E. Araujo, H. Campos Velho, Neural network and fuzzy logic statistical downscaling of atmospheric circulation-type specific weather pattern for rainfall forecasting, Appl. Soft Comput. J. 22 (2014) 681–694. doi:10.1016/j.asoc.2014.02.025.

[27]   R. Efendi, Z. Ismail, M.M. Deris, A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand, Appl. Soft Comput. J. 28 (2014) 422–430. doi:10.1016/j.asoc.2014.11.043.

[28]   K.J. Ottenbacher, R.T. Linn, P.M. Smith, S.B. Illig, M. Mancuso, C. V. Granger, Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture, Ann. Epidemiol. 14 (2004) 551–559. doi:10.1016/j.annepidem.2003.10.005.

[29]   T. Di Noia, V.C. Ostuni, F. Pesce, G. Binetti, D. Naso, F.P. Schena, et al., An end stage kidney disease predictor based on an artificial neural networks ensemble, Expert Syst. Appl. 40 (2013) 4438–4445. doi:10.1016/j.eswa.2013.01.046.

[30]   V. Volpato, A. Adelfio, G. Pollastri, Accurate prediction of protein enzymatic class by N-to-1 Neural Networks., BMC Bioinformatics. 14 Suppl 1 (2013) S11. doi:10.1186/1471-2105-14-S1-S11.

[31]   A. Kialashaki, J.R. Reisel, Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks, Appl. Energy. 108 (2013) 271–280. doi:10.1016/j.apenergy.2013.03.034.

[32]   J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, J. Clin. Epidemiol. 49 (1996) 1225–1231. doi:10.1016/S0895-4356(96)00002-9.

[33]   K.J. Ottenbacher, P.M. Smith, S.B. Illig, R.T. Linn, R.C. Fiedler, C. V Granger, Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke., J. Clin. Epidemiol. 54 (2001) 1159–1165. doi:10.1016/S0895-4356(01)00395-X.

[34]   P. Sanders, W.M. Lau, Neural Network Training on Human Body Core Temperature Data, DSTO Aeronautical and Maritime Research Laboratory, Melbourne, 1999. http://dspace.dsto.defence.gov.au/dspace/handle/1947/9033.

[35]   X. Luo, W. Hou, Y. Li, Z. Wang, A fuzzy neural network model for predicting clothing thermal comfort, Comput. Math. with Appl. 53 (2007) 1840–1846. doi:10.1016/j.camwa.2006.10.035.

[36]   T.M. Mitchell, Machine Learning, 1997. doi:10.1145/242224.242229.

[37]   MathWorks, Matlab 8.1, (2013).

[38]   A.K. Seghouane, S.I. Amari, The AIC Criterion and Symmetrizing the Kullback–Leibler Divergence, IEEE Trans. Neural Networks. 18 (2007) 97–106. doi:10.1109/TNN.2006.882813.

[39]   J.I. Arribas, J. Cid-Sueiro, A model selection algorithm for a posteriori probability estimation with neural networks, IEEE Trans. Neural Networks. 16 (2005) 799–809. doi:10.1109/TNN.2005.849826.

[40]   R. Andonie, L. Sasu, Fuzzy ARTMAP with input relevances, IEEE Trans. Neural Networks. 17 (2006) 929–941. doi:10.1109/TNN.2006.875988.

[41]   MathWorks, Matlab and Simulink for Technical Computing, (2015). http://uk.mathworks.com/index.html?s_tid=gn_logo (accessed October 20, 2015).

[42]   J.K. Patel, C.B. Read, Handbook of the Normal Distribution, Second Edition, Taylor & Francis, 1996. https://books.google.co.uk/books?id=zoVLF0VF9UYC.

[43]   M.D.P. Garcia-Souto, P. Dabnichki, Core and local skin temperature: 3–24 months old toddlers and comparison to adults, Build. Environ. 104 (2016) 286–295. doi:10.1016/j.buildenv.2016.04.016.