SPINNE: an app for human vertebral height estimation based on artificial neural networks

Vilas-Boas D[a], Wasterlain SN[a,b], d'Oliveira Coelho J[b,c], Navega D[b], Gonçalves D[a,b,d,*]

[a]Research Centre for Anthropology and Health (CIAS), Department of Life Sciences, University of Coimbra. Calçada Martim Freitas, 3000-456, Coimbra, Portugal.

[b]Laboratory of Forensic Anthropology, Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim Freitas, 3000-456, Coimbra, Portugal.

[c]Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford, OX2 6PN, United Kingdom.

[d]Archaeosciences Laboratory, Directorate General for Cultural Heritage (LARC/CIBIO/InBIO), Rua da Bica do Marquês 2, 1300-087 Lisboa, Portugal.

*Corresponding author:

davidmiguelgoncalves@gmail.com

Abstract

The absence or poor preservation of vertebrae often prevent the application of the anatomical method for stature estimation. The main objective of this paper was to develop a web app based on artificial neural network (ANN) models to estimate the vertebral height of absent or poorly preserved vertebrae from other vertebrae and thus enable the application of anatomical methods. Artificial neural models were developed based on the vertebral height of vertebrae C2 to S1 of a sample composed of 56 adult male and 69 adult female individuals. The skeletons belong to the Identified Skeletal Collection of the University of Coimbra and the ages at death of these individuals ranged from 22 to 58 years old. Statistical analysis and algorithmic development were performed with the R language, R Core Team (2018).

Intra- and inter-observer errors regarding the vertebral height were small for all vertebrae (<.45 mm). Significant models to estimate vertebral height were obtained for both sexes and for the sex-pooled group, although none with an R^2 higher than 0.48 and 0.34 for the C2 and the S1, respectively. The root mean square error (RMSE) regarding the predicted vertebral height and the observed vertebral height was almost always smaller than 1.0 mm while most R^2 values were higher than 0.6 although models with worse performances were obtained for some vertebrae located at the ends of the vertebral column (C3, L4, and L5). The ANN models have clear potential to predict vertebral height. This mathematical approach may be used to enable the application of the anatomical method for stature estimation when some vertebrae are absent or poorly preserved. The application of the ANN models can be carried out by using the new web based app SPINNE.

Key words: Biological Anthropology; Forensic Anthropology; anatomical method; missing bones; vertebral column.

1. Introduction

Stature is one of the four main parameters of the biological profile that can be estimated through human skeletal remains [1]. This biological parameter has relevance for forensic anthropology, since it helps narrowing the list of candidates for subsequent positive identification [2,3]. Stature is also used in the study of past populations, namely as an indicator of: i) developmental secular trends [4-8]; ii) evolutionary dynamics [9,10,11]; iii) nutritional stress [4,11,12]; and iv) health status and living conditions [4,12,13].

Essentially, stature can be estimated through skeletal remains via two different methods. Dwight [14] was the first to develop an anatomical method, which was afterwards adapted by other authors such as Fully [15] and Raxter et al. [16]. This method is based on a direct reconstruction of stature by adding the measurements of all bones contributing to it (cranium, vertebral bones, femur, tibia, talus and calcaneus). The resulting value is then added to a correction factor representing the soft tissues [15-17]. On the other hand, the mathematical method is based on linear regression equations that benefit from the significant correlation between stature and, more often, long bones [16-17]. This potential has been investigated for numerous bones [18-28].

In contrast to the anatomical method, the mathematical method has the advantage of being easier to apply and much less time-consuming. However, inter-individual variability in body proportions, namely between the length of the limbs and the height of the trunk, is not taken into account by this method [15,17]. Furthermore, secular changes have been documented for stature [5-6,8,29-33] as well as for limbs proportions [10,34-35]. These changes may manifest themselves in an allometric fashion [36] and therefore impair the application of regression equations that are not specific to the population under study [37].

The problem of population-specificity in stature estimation through regression equations has been known for quite some time now [20]. Stature is a multifactorial trait influenced by both genetic and environmental conditions [5,38-41]. This means that mathematical stature estimation must be preceded by age-at-death, sex and ancestry estimations [11]. Only then, regression equations may be chosen according to the characteristics of the subjects under study [42]. In archaeological collections, this problem has recently been dealt with by developing sample-specific mathematical formulae based on stature estimations obtained through the anatomical method [43-47].

The anatomical method is not affected by the problems listed above. It allows a more precise estimation and should therefore be preferentially used [15-17,48]. However, it is not free of disadvantages. Besides being more time-consuming, its application requires the presence of all bones that contribute to stature. Vertebrae represent most of these bones (24 out of 29), but due to their structure and morphology, they often preserve poorly after death [49-54]. Poor vertebral presentation may have major consequences because the absence or fragmentation of even a single bone may hinder the application of the anatomical method. Therefore, although this approach allows us to obtain more precise estimations, its implementation is much more difficult to achieve than the mathematical approach.

In this paper, we propose to combine both methods into one. Simply put, we investigated the potential of using artificial neural networks (ANN) to predict the height of each of the 24 vertebrae that contribute to stature by using the height of other vertebrae as independent variables. This approach would allow the estimation of the height of poorly preserved vertebrae using the height of better preserved vertebrae from the same skeleton. This study partially replicates and was inspired by previous work from Auerbach [55] who concluded that vertebral height can generally be predicted from contiguous vertebrae, either as a percentage or by using linear regression equations. According to this author, the option for one approach or another depends on the specific vertebra being predicted. The regression approach focused only on the C2, C3, C6, T1, T11, L1, and L5 vertebrae. Other authors before Auerbach [55] also tried similar approaches [17,56-57]. For instance, Sciulli et al. [57] estimated the height of missing vertebrae by averaging the heights of adjacent vertebrae. The predictions obtained through these methods had small standard errors, thus confirming their reliability. However, they can only be applied if vertebrae adjacent to the missing vertebrae are available. Expectantly, an artificial neural networks approach, based on the prediction of the height of a missing vertebra through more vertebrae rather than merely through the adjacent ones, offers more models thus increasing the applicability of method. Other advantages of ANN are that it minimizes the normalized root mean squared error for every output variable simultaneously and, contrary to regression methods such as the one used by Auerbach [55], it does not assume that the data are linear.

The objectives of this paper were: i) to investigate the correlation between the heights of all vertebrae contributing to stature in a sample composed of Portuguese individuals; ii) to develop ANN models that allow to predict the height of any missing or unpreserved vertebra; and iii) to develop an app that facilitates and simplifies the implementation of those models.

2. Material and Methods

Neural networks models were developed based on the collection of identified skeletons of the University of Coimbra, henceforth designated as CEIUC [58]. This sample was composed of vertebrae from 125 adult individuals with ages ranging from 22 to 58 years old (mean = 36.1; sd = 8.8). It included 69 females with a mean age of 36.2 years old (sd = 9.3) and 56 males with a mean age of 36.1 years old (sd = 8.3). All individuals were of Portuguese nationality but no further information regarding their ancestry was available.

No signs of exuberant traumas or pathologies that could eventually affect measurements were present in any of the vertebrae used in this research. As described in Raxter et al. [16], the maximum vertebral height was taken in mm with a digital calliper (resolution = .01 mm) for all vertebrae from the second cervical vertebra (C2) to the first sacral vertebra (S1). Measurements are illustrated in Figure 1. The intra- and inter-observer errors were calculated by using the technical error of measurement on a random subsample of 20 individuals [59-60]. The second session of measurements to assess intra-observer variation took place four weeks after the first one.

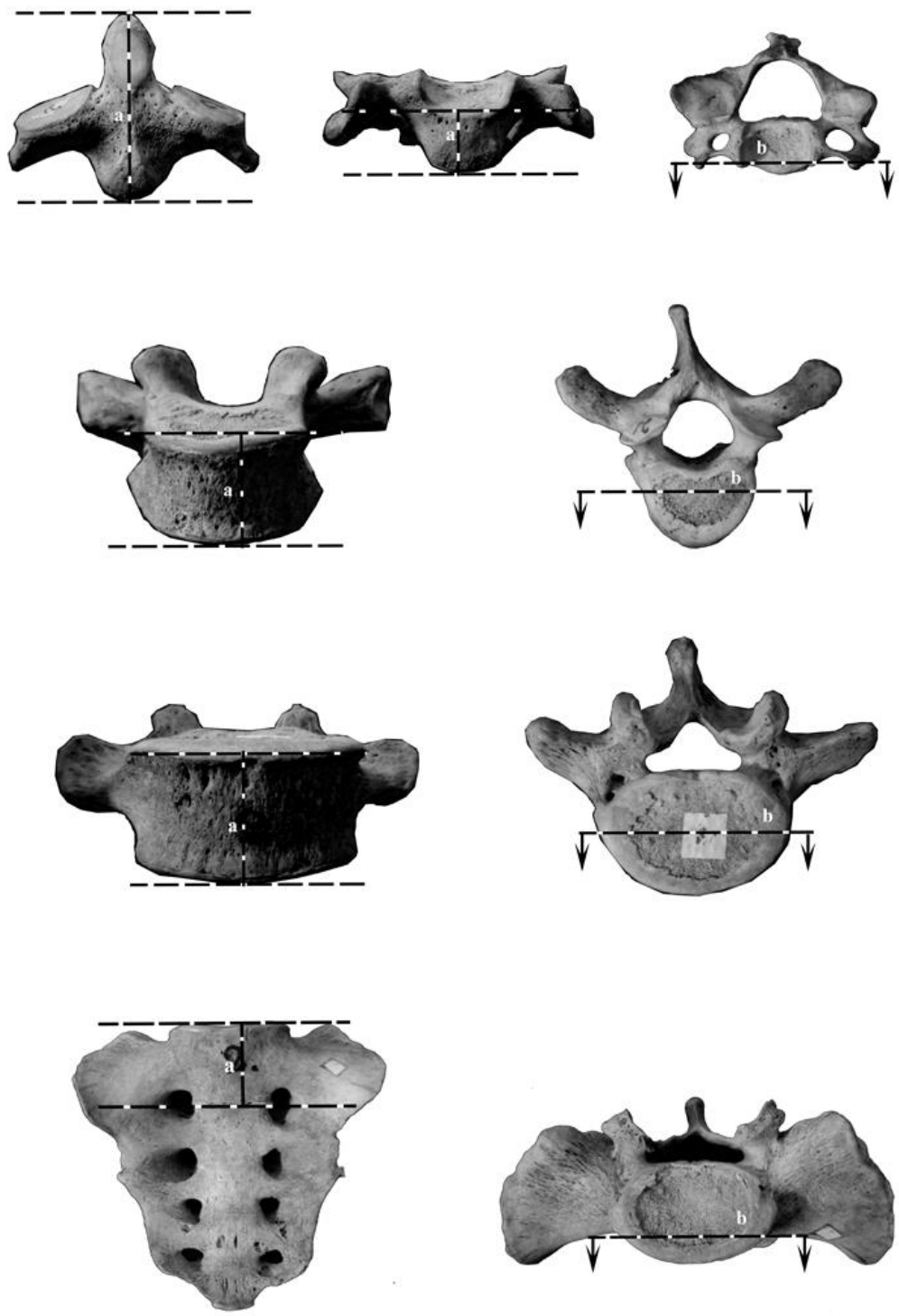

Figure 1 - Illustration of the locations at which the vertebral height measurement. (a) was taken. The antero-posterior location in superior view is also shown (b). From top to bottom: axis (C2) and another typical cervical vertebra; thoracic vertebra; lumbar vertebra; first sacral vertebra (S1). Adapted from Raxter et al. [16].

Mean differences in vertebral height – which is here defined as the summed height of all vertebrae – between both sexes were tested by computing student's t-tests for independent samples (the alpha level used in this study was 0.01). This test was also used to evaluate differences between two sex-pooled age groups (20-39 years old; 40-59 years-old). Depending on whether the assumptions for using parametric statistics were met, Pearson correlations (r) or Spearman's rank correlations (rho) were calculated to assess if associations between vertebral heights of all vertebrae involved in this study were statistically significant. Finally, ANN models to estimate the height of specific vertebrae were developed by using other vertebrae as independent variables. Models were generated separately for each sex as well as for pooled sexes.

Artificial neural network modelling was based on the generalized regression neural network architecture [61]. This type of network relies on a memory-based approach to statistical learning and like any neural network its estimates can converge to any type of regression surface, linear or nonlinear. The main difference of this neural model is that it can learn in one-pass (feed-forward) without relying on derivative-based optimization and backpropagation. A generalized regression neural network (GRNN) has a simple four layered structure: an input layer which process input data (i.e. vertebra height), a pattern layer which compares network input to each example (neuron) pre-stored in memory, a summation layer and an output layer. The last one operationalizes the final network output or regression estimate.

The network regression surface and estimates are formed, for each network input (i.e. individual), through the weighted mean of the output pre-stored in memory. The weights are dynamically computed by finding the distance (i.e. typically Euclidean) between the target input and the patterns stored in the network memory and transforming it to a similarity measure through the application of a radial basis function or a kernel function. The network output, $Y(X)'$, is mathematically defined as follows

$$Y(X)' = \frac{\sum_{i=1}^{n} YK(x, x')}{\sum_{i=1}^{n} K(x, x')}$$

where Y is the output pre-stored in the network memory, x is the target input, x' is the network inputs pre-stored in memory and $K(x, x')$ is a radial basis function or a kernel function. $K(x, x')$ is the squared exponential kernel function and it is mathematically defined as

$$K(x, x') = exp(-(\frac{D(x, x')}{\sigma})^2)$$

where $D(x,'x)$ is the euclidean distance between two multidimensional vectors with

$$D(x,'x) = (x - x')^T (x - x')$$

and $\sigma$is the radial basis function (or kernel) function width or smoothing parameter. This parameter controls the volume of information neighboring each artificial neuron (i.e. a vector of vertebrae heights pre-stored in memory). Training a GRNN model involves feeding the network memory with input and output data and finding an optimal value for the smoothing parameter σ. If the smoothing parameter is too low the network overfits by memorizing completely the training data because the "associative memory" of the network puts too much weight on the cases that are exactly equal to the input. If the smoothing parameter is too high the network underfits because each neuron in the network has a similar weight and the regression surface converge to the global mean of the output. This network architecture, in a modified form, has been previously applied in physical anthropology by Navega and collaborators [62] to model age-at-death estimation from femoral bone mineral density.

Formulating vertebral height estimation as a regression task comes with the challenge that at least 24 ANN models can be generated, one for each vertebra missing or poorly preserved. This computational challenge can be tackled by applying a multi-output regression framework. That is, if several vertebrae are missing or ill-preserved, their most likely height can be regressed simultaneously from the available vertebrae without constructing more than one predictive model. In the current paper we applied such approach by modifying the GRNN output architecture and cost function to allow for multiple output regression. Under such a scenario we used the averaged normalized root mean squared error (ANRMSE) to find the optimal value of the smoothing parameter with the cost function being formally written as

$$ANRMSE = \frac{1}{m}\sum_{i=1}^{m} \frac{\sqrt{\frac{\sum_{j=1}^{n} (Y_i - Y'_i)^2}{n}}}{IQR(Y_i)}$$

where m is the number of outputs, Y_i is the i-th output, Y'_i is the i-th estimated output and is interquartile of the i-th output. The prediction interval of each output is computed by applying the conformal prediction (CP) framework, a machine learning meta-algorithm that allows for robust heteroscedastic distribution-free error models. For the interested reader, we suggest

Shafer and Vovk [63] and Norinder *et al.* [64] as primers on conformal prediction. The value of network σ and the parameters necessary for robust conformal prediction were obtained by an internal leave-one-out cross-validation loop. The performance of the network models were assessed with an outer *K*–fold cross-validation loop (K = 5). Network optimization was performed using Brent's algorithm, a simple derivative-free optimization routine.

To assess the accuracy of the constructed neural models, the following standard metrics were computed: mean absolute error (MAE), root mean squared error and it normalized variant (RMSE and NRMSE) and R^2. In addition, bias was evaluated through the slope of the residuals of each estimated output on known output and efficiency was assessed by computing the average width of the prediction interval.

Missing values in the training data set were processed with a simple mean value imputation. This strategy was adopted because the number of missing values in the training dataset is low and the noise introduced by mean imputation is minimal when compared to constructing complex machine learning models from a reduced number of training observations which would be the case for certain arrangements of input–output parameterization using case-wise deletion of incomplete data instances.

3. Results

The intra-observer error was very small for all vertebrae. The absolute technical error of measurement ranged from .01 to .35 mm while the relative technical error of measurement ranged from .17% to 1.10%. The coefficient of reliability was always higher than .96. As for the inter-observer error, the absolute technical error of measurement ranged from .12 to .44 mm while the relative technical error of measurement ranged from .47% to 2.55%. In this case, the coefficient of reliability was always higher than .93. Therefore, vertebral height appears to be a very reliable, replicable and reproducible standard measurement.

Descriptive statistics for vertebral height from the C2 to the S1 in the CEIUC are given in Table 1. Both females and males presented the same trend. Specifically, the C2 presented the larger height and this measurement decreased until the C4. Afterwards, an uninterrupted increment in height was present from the C4 to the S1. Males presented statistically significant ($p < .001$) larger dimensions than females in all vertebrae and in the height of the vertebral column. The only exception was the L2 ($p = .01$). No significant difference at the .05 level was found between the sex-pooled 20-39 and the 40-59 age groups. Correlations among vertebrae were statistically significant in most cases (supporting information: Table S1). However, significant correlations between one vertebra and another tended to become less frequent as

223 the distance between the two, *i.e.* their relative position within the vertebral column,
224 augmented. Therefore, thoracic vertebrae tended to present more statistically significant
225 correlations with other vertebrae than cervical and lumbar vertebrae. In particular, the C2 and
226 the S1 were the vertebrae presenting less statistically significant correlations with other
227 vertebrae. A missing value analysis is given in Table 2. Overall at least one missing value was
228 detected for most the skeletal elements analysed, exceptions are the T5, T7, T8, L4 where no
229 missing values were detected. Male individuals presented at least one missing value for most
230 of the vertebrae. Missing values in female individuals were concentrated on the cervical spine.

231 Neural networks models with an R^2 that explained more than 50% of the variation of
232 vertebral height were obtained for all vertebrae except for the C2 and the S1. For the former,
233 the predictor with the largest correlation was T7 (.48), while for the latter it was T3 (.34) for
234 the latter. Examples of models based on vertebrae adjacent to the predicted vertebra are
235 given in Tables 3-5. In general, these were the better performing models, i.e. models with
236 lowest RMSE (< 1.0 mm) and highest R^2 (> 0.50) values. The models predicting the height of
237 thoracic vertebrae were the ones with better performances, i.e. with smaller NRMSE and Bias.
238 Unsurprisingly, the prediction of C2 and S1 were among the vertebrae presenting the
239 largest RMSE values, regardless of sex. Even in these cases, RMSE values remained low, usually
240 near to 2.5 mm although higher values were obtained in some models. Other vertebrae whose
241 prediction presented worse performance were usually at the ends of the vertebral column
242 (*e.g.* C3, L4, and L5). Among other predicted variables, RMSE values were almost always below
243 1.0 mm while most R^2 values were higher than 0.6. Results did not vary much after the
244 application of most ANN models, whenever adjacency was not used as a criterion. Examples of
245 poor and good performance of two models regressing C2 and T7, respectively, are given in
246 Figures 2 to 4.
247 Reporting coefficients, even for only a fraction of the models, is impracticable.
248 Therefore, an app was developed to facilitate the application of the ANN models resulting
249 from this research. It has been named SPINNE, which stands for Spine Proportion through
250 Implementation of Neural Networks Estimation and can be accessed at
251 osteomics.com/SPINNE. In some cases, R^2 presented negative values which in the case of our
252 neural network architecture is indicative that the regression surface is converging to the mean
253 of the target output(s). In such cases, Bias will also have a value near 1, which means that due
254 to the ill-defined regression surface the predicted value severely over- or under-estimates the
255 true value. That occurred for sex-pooled models developed from smaller samples that were
256 more susceptible to outliers.

257

258 Table 1 - Descriptive statistics for the vertebral height of each vertebra according to sex in the
259 collection of identified skeletons of the University of Coimbra. The significance of Student's t-
260 tests regarding sexual mean differences are also presented.

Measurement	Female					Male				
	n	mean	sd	Min.	Max.	*n*	mean	sd	Min.	Max.
C2*	60	35.24	2.39	28.89	40.57	54	38.25	2.69	33.06	45.12
C3*	65	12.71	0.85	10.06	15.15	56	14.23	1.14	11.74	17.24
C4*	67	12.27	0.93	9.78	14.76	55	13.72	0.97	11.85	15.87
C5*	67	11.79	0.85	8.94	13.25	55	13.04	1.08	10.80	15.32
C6*	65	11.96	0.83	9.10	13.46	56	13.25	1.05	11.08	17.27
C7*	66	13.48	0.93	10.95	15.62	53	14.95	1.00	12.72	16.96
T1*	66	15.13	1.08	12.49	17.94	55	16.74	0.99	13.82	18.64
T2*	69	16.89	0.90	14.65	19.52	55	18.36	0.95	15.82	20.19
T3*	69	17.02	0.81	14.29	19.68	55	18.49	1.06	15.77	20.48
T4*	69	17.35	0.88	14.78	19.24	55	18.92	1.12	16.16	21.26
T5*	69	17.91	0.95	15.34	20.14	56	19.55	1.06	16.51	21.30
T6*	69	18.36	0.99	16.24	20.74	55	20.40	1.08	17.40	22.42
T7*	69	18.93	0.97	16.57	21.01	56	20.51	0.94	18.48	22.25
T8*	69	19.29	1.02	16.21	21.50	56	20.76	0.92	18.84	22.75
T9*	69	19.88	1.16	16.96	22.76	55	21.30	0.94	19.43	22.95
T10*	69	20.93	1.31	17.18	23.44	55	22.29	1.13	19.43	24.81
T11*	69	21.58	1.35	18.66	23.96	54	23.00	1.26	20.35	25.69
T12*	69	23.17	1.52	19.70	26.58	55	24.33	1.36	20.57	27.27
L1*	69	24.93	1.55	21.51	28.29	54	26.09	1.22	22.86	29.26
L2*	69	26.27	1.59	23.29	29.66	55	26.93	1.37	23.94	29.71
L3*	69	26.99	1.55	23.69	31.03	54	27.90	1.39	25.08	31.04
L4*	69	26.87	1.87	20.24	31.07	56	28.37	1.54	25.66	32.30
L5*	67	27.24	1.65	24.15	31.13	56	28.39	1.45	25.85	31.80
S1*	67	31.25	1.89	25.89	34.76	54	32.96	2.05	28.66	37.62

261 *significant at the .01 level.

262 Table 2: Missing value analysis of vertebrae height.

Vertebra Height	Pooled			Males			Female		
	n	Missing (n)	Missing (%)	n	Missing (n)	Missing (%)	n	Missing (n)	Missing (%)
C2	114	11	8.80	54	2	3.57	60	9	13.04
C3	121	4	3.20	56	0	0.00	65	4	5.80
C4	122	3	2.40	55	1	1.79	67	2	2.90
C5	122	3	2.40	55	1	1.79	67	2	2.90
C6	121	4	3.20	56	0	0.00	65	4	5.80
C7	119	6	4.80	53	3	5.36	66	3	4.35
T1	121	4	3.20	55	1	1.79	66	3	4.35
T2	124	1	0.80	55	1	1.79	69	0	0.00
T3	124	1	0.80	55	1	1.79	69	0	0.00
T4	124	1	0.80	55	1	1.79	69	0	0.00
T5	125	0	0.00	56	0	0.00	69	0	0.00
T6	124	1	0.80	55	1	1.79	69	0	0.00
T7	125	0	0.00	56	0	0.00	69	0	0.00
T8	125	0	0.00	56	0	0.00	69	0	0.00
T9	124	1	0.80	55	1	1.79	69	0	0.00
T10	124	1	0.80	55	1	1.79	69	0	0.00
T11	123	2	1.60	54	2	3.57	69	0	0.00
T12	124	1	0.80	55	1	1.79	69	0	0.00
L1	123	2	1.60	54	2	3.57	69	0	0.00
L2	124	1	0.80	55	1	1.79	69	0	0.00
L3	123	2	1.60	54	2	3.57	69	0	0.00
L4	125	0	0.00	56	0	0.00	69	0	0.00
L5	123	2	1.60	56	0	0.00	67	2	2.90
S1	121	4	3.20	54	2	3.57	67	2	2.90

263

264

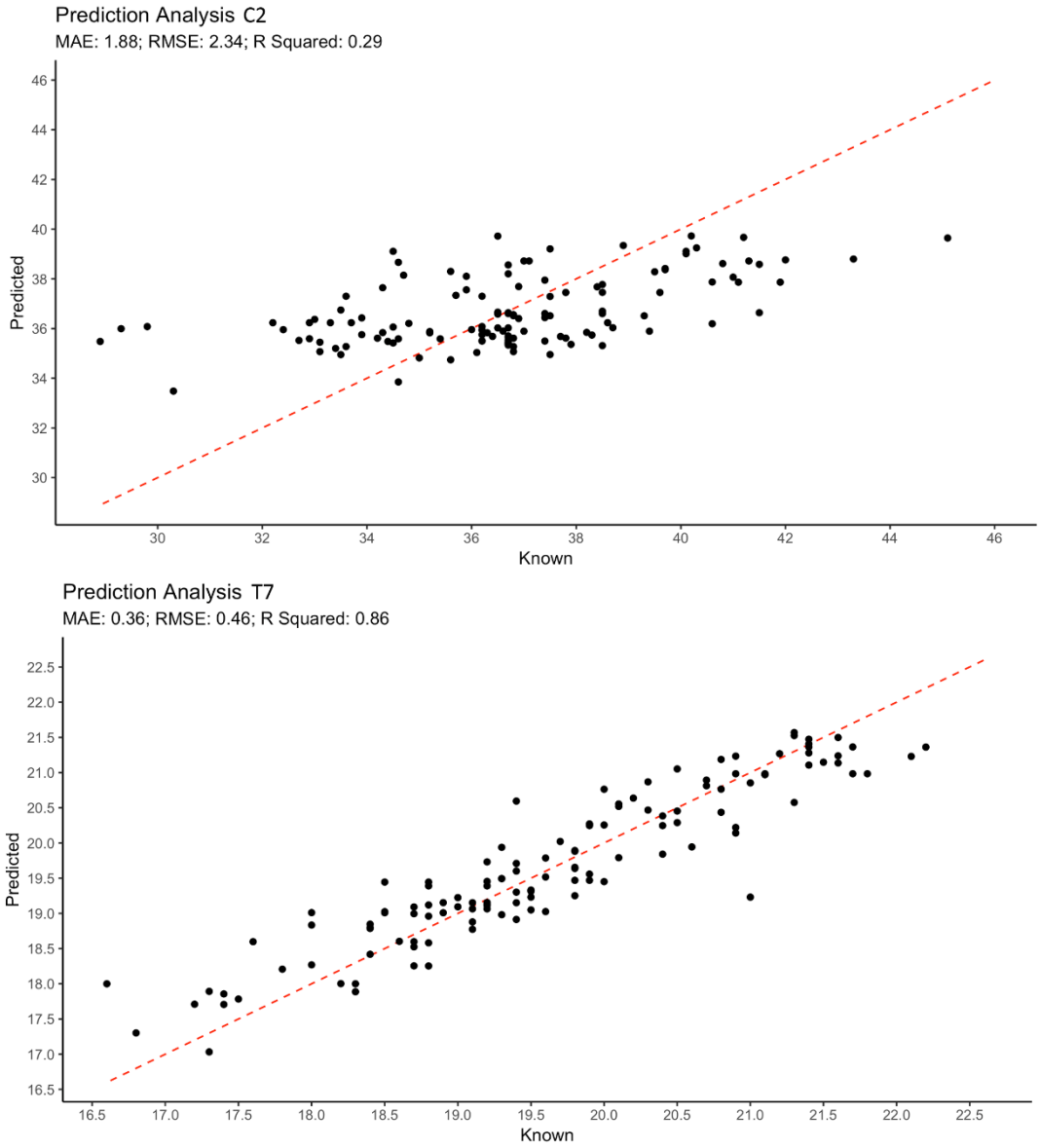

Figure 2 – Top: prediction analysis of C2 vertebra regressed from C3, C4, C5 and C6 vertebrae, used as an example of a model with poor performance; bottom: prediction analysis of T7 vertebra regressed from T5, T6, T8, and T9 vertebrae, used as an example of a model with good performance.

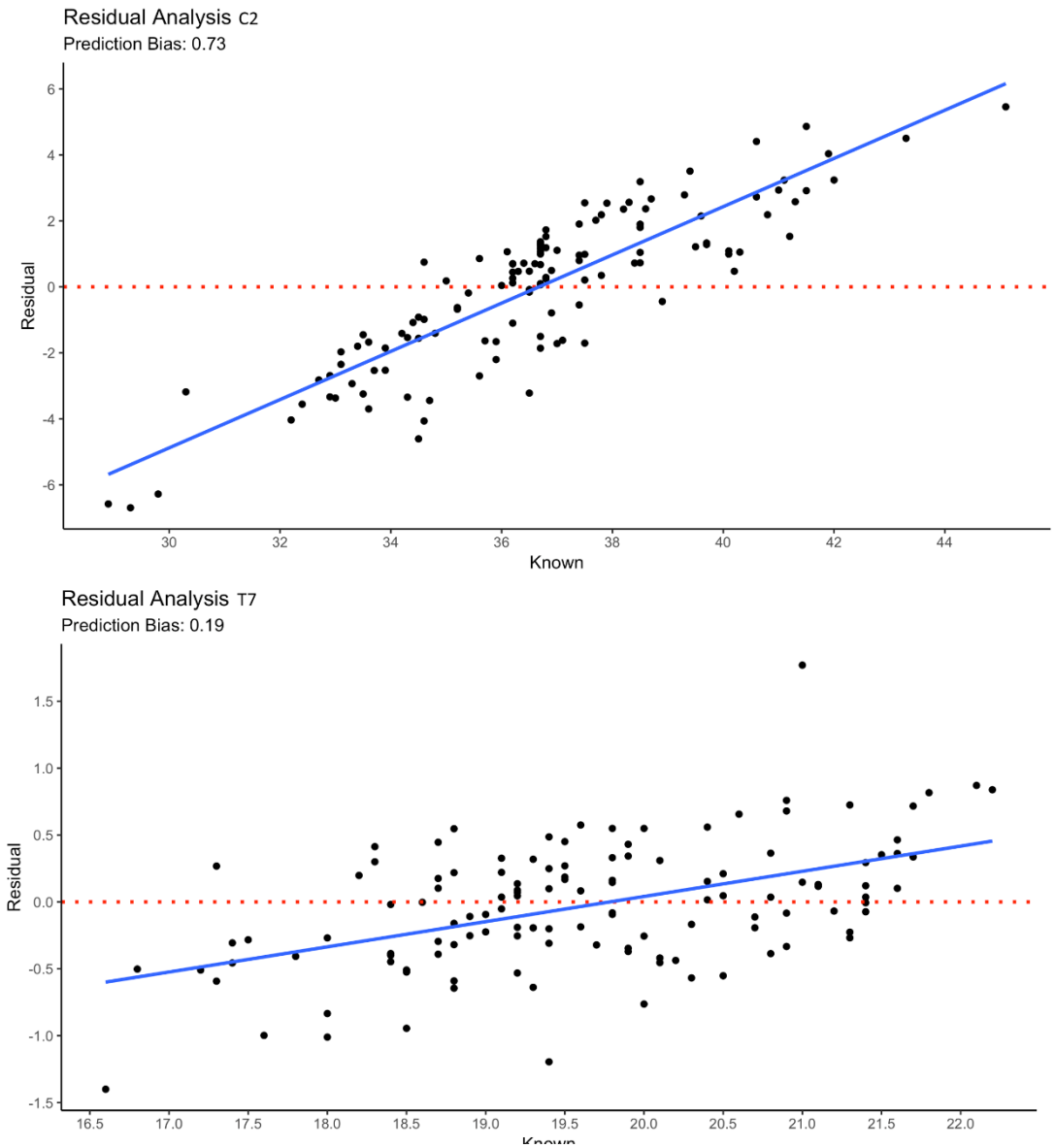

Figure 3 - Top: residual analysis of C2 vertebra regressed from C3, C4, C5, and C6 vertebrae, used as an example of a model with poor performance; bottom: residual analysis of T7 vertebra regressed from T5, T6, T8, and T9 vertebrae, used as an example of a model with good performance.

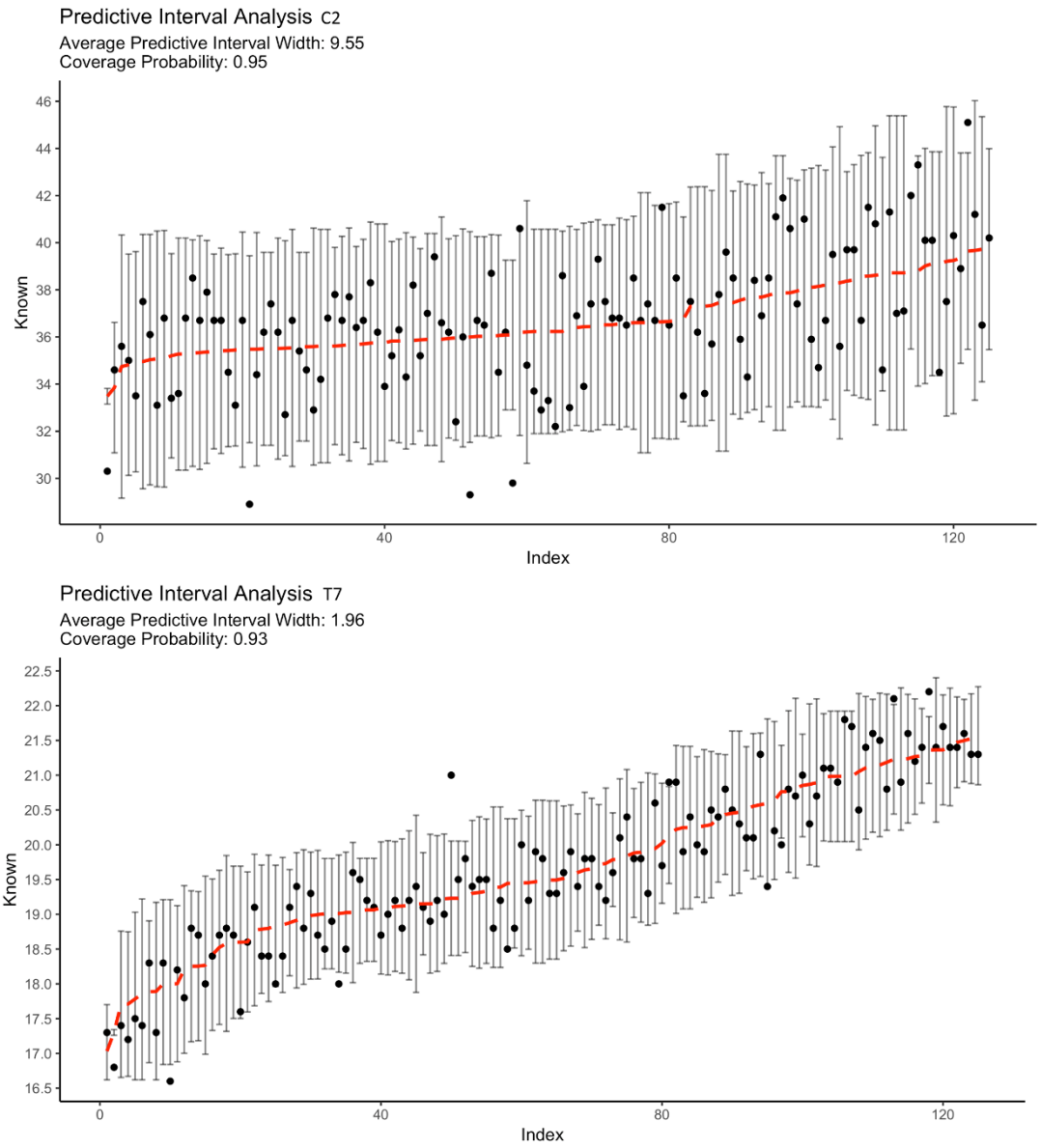

Figure 4 - Top: predictive interval analysis of C2 vertebra regressed from C3, C4, C5, and C6 vertebrae, used as an example of a model with poor performance; bottom: predictive interval analysis of T7 vertebra regressed from T5, T6, T8, and T9 vertebrae, used as an example of a model with good performance.

Table 3 – Artificial neural networks models performance for the sex-pooled sample.

Estimated	Estimators	MAE	RMSE	NRMSE	R^2	Bias	Coverage	PIW
C2	S1, C3	1.84	2.29	0.62	0.33	0.69	0.94	8.97
C3	C2, C4	0.68	0.85	0.57	0.52	0.53	0.94	3.89
C4	C3, C5	0.53	0.69	0.43	0.65	0.42	0.95	2.82
C5	C4, C6	0.51	0.64	0.42	0.68	0.37	0.94	2.41
C6	C5, C7	0.52	0.73	0.61	0.57	0.44	0.93	2.72
C7	C6, T1	0.46	0.59	0.36	0.75	0.31	0.93	2.37
T1	C7, T2	0.48	0.63	0.34	0.76	0.29	0.96	2.67
T2	T1, T3	0.43	0.53	0.32	0.8	0.25	0.95	2.09
T3	T2, T4	0.38	0.47	0.29	0.84	0.22	0.94	1.96
T4	T3, T5	0.44	0.53	0.29	0.82	0.23	0.94	2.07
T5	T4, T6	0.39	0.5	0.29	0.85	0.2	0.94	2.11
T6	T5, T7	0.46	0.59	0.29	0.83	0.22	0.94	2.32
T7	T6, T8	0.37	0.47	0.26	0.85	0.19	0.93	1.88
T8	T7, T9	0.43	0.54	0.29	0.8	0.23	0.96	2.35
T9	T8, T10	0.41	0.55	0.32	0.81	0.24	0.94	2.49
T10	T9, T11	0.52	0.65	0.32	0.78	0.24	0.94	2.57
T11	T10, T12	0.56	0.71	0.41	0.77	0.27	0.92	2.50
T12	T11, L1	0.63	0.82	0.39	0.72	0.34	0.91	3.50
L1	T12, L2	0.61	0.79	0.4	0.73	0.32	0.95	3.21
L2	L1, L3	0.71	0.89	0.44	0.65	0.38	0.91	3.43
L3	L2, L4	0.63	0.79	0.43	0.73	0.28	0.93	3.52
L4	L3, L5	0.75	1.05	0.37	0.69	0.38	0.94	4.22
L5	L4, S1	0.78	1.00	0.41	0.63	0.44	0.94	4.48
S1	L5, C2	1.37	1.78	0.67	0.27	0.77	0.93	7.23

MAE, mean absolute error; RMSE, root mean squared error; NRMSE, normalized root mean squared error (IQR normalization); R^2, R Squared (Explained Variance); PIW, Predictive Interval Width (Mean).

Table 4 – Artificial neural networks models performance for the male sample.

Estimated	Estimators	MAE	RMSE	NRMSE	R^2	Bias	Coverage	PIW
C2	S1, C3	2.13	2.51	0.68	0.08	0.87	0.91	9.15
C3	C2, C4	0.73	0.93	0.7	0.33	0.74	0.93	4.11
C4	C3, C5	0.49	0.66	0.49	0.51	0.56	0.93	2.83
C5	C4, C6	0.52	0.64	0.4	0.63	0.4	0.96	2.89
C6	C5, C7	0.51	0.77	0.58	0.46	0.58	0.93	2.88
C7	C6, T1	0.47	0.59	0.39	0.62	0.38	0.93	2.46
T1	C7, T2	0.51	0.7	0.55	0.49	0.58	0.95	2.87
T2	T1, T3	0.49	0.6	0.57	0.59	0.5	0.93	2.35
T3	T2, T4	0.46	0.58	0.43	0.7	0.34	0.91	2.58
T4	T3, T5	0.5	0.61	0.48	0.7	0.36	0.91	2.43
T5	T4, T6	0.46	0.57	0.38	0.71	0.34	0.95	2.19
T6	T5, T7	0.53	0.64	0.51	0.63	0.46	0.95	2.73
T7	T6, T8	0.39	0.48	0.33	0.74	0.3	0.93	2.12
T8	T7, T9	0.45	0.59	0.53	0.58	0.5	0.93	2.41
T9	T8, T10	0.48	0.61	0.51	0.56	0.46	0.91	2.77
T10	T9, T11	0.61	0.79	0.55	0.5	0.57	0.93	3.6
T11	T10, T12	0.64	0.81	0.42	0.57	0.52	0.93	3.95
T12	T11, L1	0.81	1.01	0.7	0.43	0.63	0.95	4.41
L1	T12, L2	0.56	0.75	0.46	0.6	0.46	0.93	3.55
L2	L1, L3	0.67	0.81	0.49	0.64	0.42	0.98	3.8
L3	L2, L4	0.7	0.87	0.52	0.59	0.47	0.93	3.55
L4	L3, L5	0.77	0.96	0.46	0.6	0.48	0.91	4.09
L5	L4, S1	0.78	0.98	0.53	0.53	0.53	0.93	4.67
S1	L5, C2	1.6	2.01	0.75	-0.02	0.97	0.95	8.03

MAE, mean absolute error; RMSE, root mean squared error; NRMSE, normalized root mean squared error (IQR normalization); R^2, R Squared (Explained Variance); PIW, Predictive Interval Width (Mean).

306 Table 5 – Artificial Neural Networks models performance for the female sample.

Estimated	Estimators	MAE	RMSE	NRMSE	R^2	Bias	Coverage	PIW
C2	S1, C3	1.61	2.04	0.71	0.14	0.89	0.94	8.04
C3	C2, C4	0.54	0.7	0.66	0.27	0.78	0.94	3.73
C4	C3, C5	0.55	0.68	0.7	0.44	0.62	0.93	2.79
C5	C4, C6	0.4	0.53	0.43	0.6	0.51	0.94	2
C6	C5, C7	0.47	0.61	0.51	0.41	0.65	0.93	2.42
C7	C6, T1	0.46	0.59	0.55	0.58	0.48	0.91	2.19
T1	C7, T2	0.58	0.75	0.58	0.49	0.6	0.91	2.75
T2	T1, T3	0.44	0.56	0.44	0.6	0.47	0.93	2.19
T3	T2, T4	0.36	0.49	0.53	0.63	0.46	0.94	2.09
T4	T3, T5	0.4	0.5	0.4	0.68	0.43	0.91	2.01
T5	T4, T6	0.37	0.49	0.39	0.73	0.36	0.91	1.79
T6	T5, T7	0.41	0.5	0.37	0.74	0.3	0.88	1.82
T7	T6, T8	0.35	0.46	0.39	0.77	0.27	0.91	1.91
T8	T7, T9	0.45	0.55	0.53	0.7	0.36	0.96	2.38
T9	T8, T10	0.42	0.55	0.35	0.77	0.3	0.94	2.58
T10	T9, T11	0.44	0.55	0.35	0.82	0.26	0.9	1.88
T11	T10, T12	0.43	0.54	0.25	0.84	0.2	0.87	2
T12	T11, L1	0.53	0.68	0.34	0.8	0.28	0.9	2.54
L1	T12, L2	0.64	0.85	0.41	0.7	0.39	0.93	3.4
L2	L1, L3	0.73	0.94	0.4	0.64	0.44	0.93	3.53
L3	L2, L4	0.72	0.89	0.41	0.66	0.42	0.93	5.29
L4	L3, L5	0.77	1.12	0.47	0.63	0.45	0.91	4.06
L5	L4, S1	0.85	1.08	0.46	0.55	0.54	0.9	4.26
S1	L5, C2	1.25	1.64	0.91	0.21	0.77	0.94	7.11

307 MAE, mean absolute error; RMSE, root mean squared error; NRMSE, normalized root mean squared error (IQR normalization); R^2,
308 R Squared (Explained Variance); PIW, Predictive Interval Width (Mean).

309

310

311

4. Discussion

Our results demonstrated that it is possible to predict the height of missing or poorly preserved vertebrae through ANN models based on the height of other vertebrae, although the precision of the estimates varied quite considerably according to each specific vertebra. In most cases, however, mean differences between the predictions and the observations were smaller than 1.0 mm. There seems to be a lot of potential in using this technique to fill eventual gaps that may arise whenever the application of an anatomical method of stature estimation is attempted. Therefore, the use of an anatomico-mathematical method, such as the one recommended by Auerbach [55], appears to be a reliable option.

Predictions tended to worsen as the position of the vertebrae acting as independent variables became further away from the predicted vertebra. This may affect the performance of the regression models approach since poorly preserved vertebrae often tend to be accompanied by contiguous vertebrae presenting the same conditions. For instance, cervical vertebrae are often in worse conditions than lumbar vertebrae [50,54]. Whenever several contiguous vertebrae are missing or badly preserved, this basically leaves three choices. The examiner may: i) attempt making the prediction via a more distant vertebra; ii) use a preserved vertebra to predict the height of its contiguous vertebra and then use this prediction to predict the next vertebra in line, and so on until the heights of all vertebrae are predicted; iii) or predict the total height of the vertebral column, a procedure also made available by the SPINNE app.

An interesting realization obtained after comparing our results with the ones from Auerbach [55] is that, in both investigations, the best performing models tended to include vertebrae that are adjacent to the predicted vertebra. However, the better performing models presented in both papers generally had different configurations in terms of their independent variables. This suggests that variability is expected from one population to another and that the prediction potential of each vertebra may vary among populations. Nonetheless, eventual sample-specificity of our models should not be a major problem as long as the size proportion that each vertebrae occupies along the spine is similar across populations.

Since sexual differences have been found regarding vertebral height, the successful application of this method should be enhanced if the sex of the skeleton under examination is known, since sex-specific models can be applied. However, differences between sex-pooled and sex-specific models were rather small.

One obstacle to the application of regression models to estimate vertebral height is age since these variables maintain a negative correlation [65-66]. Although pathological

degenerative processes tend to have a greater impact on the elderly, they can occur much earlier, usually during the fourth decade of life [67]. No statistically significant mean age differences were present for the two age groups in our sample since it was composed of adult individuals who were less than 60 years old at time of death and presented no exuberant pathological conditions of the spine. If older individuals or pathological vertebrae had been included in the sample, larger prediction errors would surely have occurred.

Another potential obstacle is that the models here presented may be population-specific. This problem may be inflated if the skeleton under examination presents many absent or unpreserved vertebrae. Until further testing on samples from different populations are performed, it is advisable to keep this in mind whenever deciding if the SPINNE app should be applied to a specific skeleton or not. Auerbach [55] did not find large differences when comparing the accuracies and precisions of the anatomical, mathematical and anatomico-mathematical methods to estimate stature but the latter tended to overestimate stature. However, Auerbach [55] was only able to base his observations on comparisons with the anatomically estimated stature so comparisons are not completely enlightening. Nonetheless, the anatomico-mathematical approach appears to have clear potential to contribute for stature estimation of partial skeleton remains which are frequently found in both archaeological and forensic settings [43-47]. The main advantage of our method over others such as the ones from Sciulli et al. [57] and Auerbach [55] is its flexibility. It is not dependent on the presence and preservation of vertebrae that are adjacent to the missing vertebra. The large amount of models made available by the SPINNE app turns it more flexible to the skeletal preservation observed in a case-by-case basis thus widening its applicability.

The potential number of ANN models to predict vertebral height obtained in this research is numerous - these can be accessed by using our web app SPINNE. Although we have illustrated very simple models exploiting vertebral adjacency in this paper, a large number of possible configurations of input - output can be generated using a multi-output regression framework with neural networks. It is possible to regress the missing value of two vertebrae from the remaining 22 or vice-versa. This flexibility however should be explored with caution, it might not be reasonable to create a neural network to predict the height of C2 and C3 from the metric parameters of the lumbar spine.

Through SPINNE it is possible to explore data and, primarily, estimate vertebral heights. To estimate missing vertebral heights, users should access the "estimating missing values" tab and: 1) Select the "sex of subject" if you wish to use this option or leave it as “unknown” otherwise; 2) In "input variables", select the vertebra or vertebrae for which vertebral height is known; 3) Insert their heights in "measurements"; and 4) "Calculate missing

values". The results give the predictions for the missing vertebrae heights as well as for the partial total column. The latter is obtained by summing all individual vertebrae predictions and is not a regression *per se* in itself. The performance metrics of each model (e.g. adjusted R^2; RMSE) can be used as a decision support system regarding the interpretation of the results. However, choosing a model is a somewhat subjective decision that depends on the objective one wishes to fulfil. For example, if SPINNE is used within a forensic case, then the expert should be especially cautious bearing in mind that performance metrics are merely part of the decision support system.

5. Conclusions

The use of regression models to estimate vertebral height has value not only as a contribution for stature estimation, but it may also be helpful to assess the minimum number of individuals. For example, more than one individual may be identified if vertebrae in commingled remains are metrically incompatible with one another. The neural network architecture implemented in this paper can also be applied to solve imputation of numerical predictors or nonlinear regression in other contexts of skeletal analysis and for that reason source code of the algorithm is available as an R package at https://github.com/dsnavega. SPINNE is also available as an R package (*grnnet* - Generalized Regression Neural Network) at https://github.com/dsnavega and https://github.com/Delvis.

Acknowledgements

The authors acknowledge the following financial support by the Portuguese Foundation for Science and Technology and COMPETE 2020 program [SFRH/BPD/84268/2012; SFRH/BD/122306/2016; SFRH/BD/99676/2014; PTDC/IVC-ANT/1201/2014 & POCI-01-0145-FEDER-016766; PEst-OE/SADG/UI0283/2013].

References

[1] M.Y. Işcan, M. Steyn, The human skeleton in forensic medicine, Charles C. Thomas, Springfield, 2013.

[2] R.J. Wilson, N.P. Herrmann, L.M. Jantz, Evaluation of stature estimation from the database for forensic anthropology, J. Forensic Sci. 55 (2010) 684-689.

[3] T.D. White, M.T. Black, P.A. Folkens, Human osteology, Academic Press, Oxford, 2012.

[4] I.M. Schmidt, M.H. Jørgensen, K.F. Michaelsen, Height of conscripts in Europe: is postneonatal mortality a predictor? Ann. Hum. Biol. 22 (1995) 57–67.

[5] C. Padez, Secular trend in stature in the Portuguese population (1904-2000), Ann. Hum. Biol. 30 (2003) 262–278.

[6] C. Padez, Secular trend in Portugal, J. Hum. Ecol. 22 (2007) 15–22.

[7] M. Giannecchini, J. Moggi-Cecchi, Stature in archaeological samples from central Italy: methodological issues and diachronic changes, Am. J. Phys. Anthropol. 135 (2008) 284–292.

[8] H.F.V. Cardoso, J.E.A. Gomes, Trends in adult stature of peoples who inhabited the modern Portuguese territory from the Mesolithic to the late 20th century, Int. J. Osteoarchaeol. 19 (2009) 711–725.

[9] W.L. Jungers, Lucy's length: stature reconstruction in *Australopithecus afarensis* (A.L.288-1) with implications for other small-bodied hominids, Am. J. Phys. Anthropol. 76 (1988) 227–231.

[10] T.W. Holliday, Brachial and crural indices of European late upper Paleolithic and Mesolithic humans, J. Hum. Evol. 36 (1999) 549–566.

[11] K.M. Moore, A.H. Ross, Stature estimation, in: E.A. DiGangi, M.K. Moore (Eds.), Research Methods in Human Skeletal Biology, Academic Press, San Diego, 2013, pp. 151-176.

[12] J.M. Tanner, Growth as a measure of the nutritional and hygienic status of a population, Horm. Res. 38 (Suppl) (1992) 106-115.

[13] A. Fernihough, M.E. McGovern, Physical stature decline and the health status of the elderly population in England, Econ. Hum. Biol. 16 (2015) 30–44.

[14] T. Dwight, Methods of estimating the height from parts of the skeleton, Medical Record, 46 (1894) 293–296.

[15] M. G. Fully, Une nouvelle méthode de détermination de la taille, Annales de Médecine Légale et de Criminologie 35 (1956) 266–273.

[16] M.H. Raxter, B.M. Auerbach, C.B. Ruff, Revision of the Fully technique for estimating statures, Am. J. Phys. Anthropol. 130 (2006) 374–384.

[17] J.K. Lundy, The mathematical versus anatomical methods of stature estimate from long bones, Amer. J. Forensic Med. Pathol. 6 (1985) 73–75.

[18] K. Pearson, Mathematical contribution to the theory of evolution: on the reconstruction of the stature of prehistoric races, Philosophical Transactions of the Royal Society of London 192 (1899) 169 -244.

[19] A. Telkkä, On the prediction of human stature from long bones, Acta Anatomica 9 (1950) 103-117.

[20] M. Trotter, G.C. Gleser, Estimation of stature from long bones of American whites and Negroes, Am. J. Phys. Anthropol. 10 (1952) 463–514.

[21] G. Olivier, C. Aaron, G. Fully, G. Tissier, New estimation of stature and cranial capacity in modern man, J. Hum. Evol. 7 (1978) 513-518.

[22] G.L. Tibbetts, Estimation of stature from the vertebral column in American Blacks, J. Forensic Sci. 26 (1981) 715–723.

[23] L. Meadows, R.L. Jantz, Estimation of stature from metacarpal lengths, J. Forensic Sci. 37 (1992) 147–154.

[24] M.C. Mendonça, Estimation of height from the length of long bones in a Portuguese adult population, Am. J. Phys. Anthropol. 112 (2000) 39–48.

[25] M.A. Bidmos, Adult stature reconstruction from the calcaneus of South Africans of European descent, J. Clin. Forensic Med. 13 (2006) 247–252.

[26] M.A. Bidmos, Metatarsals in the estimation of stature in South Africans, J. Forensic Leg. Med. 15 (2008) 505-509.

[27] C.L. Giroux, D.J. Wescott, Stature estimation based in dimensions of the bony pelvis and proximal femur, J. Forensic Sci. 53 (2008) 65-68.

[28] C. Cordeiro, J. Muñoz-Barús, S. Wasterlain, E. Cunha, D.N. Vieira, Predicting adult stature from metatarsal length in a Portuguese population, Forensic Sci. Int. 193 (2009) 131.e1-131.e4.

[29] M. Trotter, G.C. Gleser, Trends in stature of American whites and negroes born between 1840 and 1924, Am. J. Phys. Anthropol. 9 (1951) 427-440.

[30] L.G. Sandberg, R.H. Steckel, Heights and economic history: the Swedish case, Ann. Hum. Biol. 14 (1987) 101-110.

[31] K. Silventoinen, E. Lahelma, O. Lundberg, O. Rahkonen, Body height, birth cohort and social background in Finland and Sweden, Eur. J. Public Health 11 (2001) 124-129.

[32] K. Staub, F.J. Rühli, U. Woitek, C. Pfister, The average height of 18- and 19-year-old conscripts (N=458,322) in Switzerland from 1992 to 2009, and the secular height trend since 1878, Swiss Medical Weekly 141 (2011) w13238.

[33] J. Myburgh, K. Staub, F.J. Rühli, J.R. Smith, M. Steyn, Secular trends in stature of late 20th century white South Africans and two European populations, HOMO – J. Comp. Human Biol 68(6) (2017) 433-439.

[34] L.M. Jantz, R.L. Jantz, Secular change in long bone length and proportion in the United States, 1800 – 1970, Am. J. Phys. Anthropol. 110 (1999) 57–67.

[35] T.W. Holliday, C.B. Ruff, Relative variation in human proximal and distal limb segment lengths, Am. J. Phys. Anthropol. 116 (2001) 26–33.

[36] L. Meadows, R.L. Jantz, Allometric secular change in the long bones from the 1800s to the present, J. Forensic Sci. 40 (1995) 762–767.

[37] I. Sierp, M. Henneberg, Reconstruction of body height from the skeleton: Testing a dozen different methods for consistency of their results, Anthropologischer Anzeiger 73 (2016) 7-21.

[38] S. Macgregor, B.K. Cornes, N.G. Martin, P.M. Visscher, Bias, precision and heritability of self-reported and clinically measured height in Australian twins, Hum. Genet. 120 (2006) 571–580.

[39] M. Perola, S. Sammalisto, T. Hiekkalinna, N.G. Martin, P.M. Visscher, G.W. Montgomery, B. Benyamin, J.R. Harris, D. Boomsma, G. Willemsen, J.-J. Hottenga, K. Christensen, K.O. Kyvik, T.I.A. Sørensen, N.L. Pedersen, P.K.E. Magnusson, T.D. Spector, E. Widen, K. Silventoinen, J. Kaprio, A. Palotie, L. Peltonen, GenomEUtwin Project, Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci, PLoS Genetics 3 (2007) e97.

[40] M.E. Danubio, E. Sanna, Secular changes in human biological variables in Western countries: an updated review and synthesis, J. Anthropol. Sci. 86 (2008) 91–112.

[41] M.N. Weedon, T.M. Frayling, Reaching new heights: insights into the genetics of human stature, Trends in Genetics 24 (2008) 595–603.

[42] B.J. Adams, N.P. Herrmann, Estimation of living stature from selected anthropometric (soft tissue) measurements: applications for forensic anthropology, J. Forensic Sci. 54 (2009) 753-760.

[43] M.H. Raxter, C.B. Ruff, A. Azab, M. Erfan, M. Soliman, A. El-Sawaf, Stature estimation in ancient Egyptians: A new technique based on anatomical reconstruction of stature, Am. J. Phys. Anthropol. 136 (2008) 147-155.

[44] B.M. Auerbach, C.B. Ruff, Stature estimation formulae for indigenous North American populations. Am. J. Phys. Anthropol. 141 (2010) 190-207.

[45] E. Pomeroy, J.T. Stock, Estimation of stature and body mass from the skeleton among coastal and mid-altitude Andean populations, Am. J. Phys. Anthropol. 147 (2012) 264-279.

[46] C.B. Ruff, B.M. Holt, M. Niskanen, V. Sladék, M. Berner, E. Garofalo, H.M. Garvin, M. Hora, H. Maijanen, S. Niinimäki, K. Salo, E. Schuplerová, D. Tompkins, Stature and body mass estimation from skeletal remains in the European Holocene, Am. J. Phys. Anthropol. 148 (2012) 601-617.

[47] V. Sládek, J. Macháček, C.B. Ruff, E. Schuplerová, R. Prichystalová, M. Hora, Population-specific stature estimation from long bones in the Early Medieval Pohansko (Czech Republic, Am. J. Phys. Anthropol. 158 (2015) 312-324.

[48] D. Brits, P.R. Manger, M.A. Bidmos, Assessing the use of the anatomical method for the estimation of sub-adult stature in Black South Africans, Forensic Sci. Int. 283 (2018) 221.e1-221.e9.

[49] J. Henderson, Factors determining the state of preservation of human remains, in: A. Boddington, A.N. Garland, R.C. Janaway (Eds.), Death Decay and Reconstruction: Approaches to Archaeology and Forensic Science, Manchester University Press, Manchester, 1987, pp. 43-54.

[50] S. Mays, Taphonomic factors in a human skeletal assemblage, Circaea 9 (1992) 54–58.

[51] D.W. Von Endt, D.J. Ortner, Experimental effects of bone size and temperature on bone diagenesis, J. Archaeol. Sci. 11 (1984) 247–253.

[52] T. Waldron, The relative survival of the human skeleton: implications for paleopathology, in: A. Boddington, A.N. Garland, R.C. Janaway (Eds.), Death Decay and Reconstruction:

Approaches to Archaeology and Forensic Science, Manchester University Press, Manchester, Manchester, 1987, pp. 43-54.

[53] C.M. Nielsen-Marsh, R.E. Hedges, Patterns of diagenesis in bone I: The effects of site environments, J. Archaeol. Sci. 27 (2000) 1139–1150.

[54] S. Bello, P. Andrews, The intrinsic pattern of preservation of human skeletons and its influence on the interpretation of funerary behaviours, in: R. Gowland, C. Knüsel (Eds.), The Social Archaeology of Funerary Remains, Oxbow Books, Oxford, pp. 1-13.

[55] B.M. Auerbach, Methods for estimating missing human skeletal element osteometric dimensions employed in the revised Fully technique for estimating stature, Am. J. Phys. Anthropol. 145 (2011) 67-80.

[56] M.G. Fully, H. Pineau, Détermination de la stature au moyen du squelette, Annales de Médecine Légale, 40 (1960) 145–154.

[57] P.W. Sciulli, K.N. Schneider, M.C. Mahaney, Stature estimation in prehistoric Native Americans of Ohio, Am. J. Phys. Anthropol. 83 (1990) 275–280.

[58] E. Cunha, S. Wasterlain, The Coimbra identified osteological collections, in: G. Grupe G, J. Peters (Eds.), Skeletal series and their socio-economic context. Documenta Archaeobiologiae 5, Verlag Marie Leidorf GmbH, Rahden/Westf, 2007, pp. 23-33.

[59] S.J. Ulijaszek, D.A. Kerr, Anthropometric measurement error and the assessment of nutritional status, Br. J. Nutr. 82 (1999) 165-177.

[60] T.A. Perini, G.L. Oliveira, J.S. Ornellas, F.P. Oliveira, Cálculo do erro técnico de medição em antropometria, Revista Brasileira de Medicina do Esporte 11 (2005) 86-90.

[61] D.F. Specht, A general regression neural network, IEEE Transactions on Neural Networks 2(6) (1991) 568-576.

[62] D. Navega, J. d'Oliveira Coelho, E. Cunha, F. Curate, DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks, J. Forensic Sci. 63 (2018) 497–503.

[63] G. Shafer, V. Vovk, A tutorial on conformal prediction, J. Mach. Learn. Res. 9 (2008) 371-421.

[64] U. Norinder, L. Carlsson, S. Boyer, M. Eklund, Introducing conformal prediction in predictive modelling: A transparent and flexible alternative to applicability domain determination, J. Chem. Inf. Model. 54 (2014) 1596-1603.

[65] M.F. Ericksen, Some aspects of aging in lumbar spine, Am. J. Phys. Anthropol. 45 (1976) 575-580.

[66] A.P. Herrmann, K. Brixen, J. Andresen, L. Mosekilde, Reference values for vertebral heights in scandinavian females and males, Acta Radiol 34 (1993) 48–52.

[67] A. Aufderheide, C. Rodríguez-Martín, The Cambridge encyclopedia of human paleopathology, Cambridge University Press, Cambridge, 1998.