bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PREDICTING THE PRINCIPAL COMPONENTS OF CORTICAL MORPHOLOGICAL VARIABLES

MANUSCRIPT

V. B. B. Mello[1], **F. H. de Moraes**[1,2], **B. Mota**[1]
[1] Instituto de Física, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[2] Brain Connectivity Unit, Instituto D'Or de Pesquisa e Ensino (IDOR), Rio de Janeiro, Brazil
Contact: vbraga@pos.if.ufrj.br

March 6, 2023

ABSTRACT

The generating mechanism for the gyrification of the mammalian cerebral cortex remains a central open question in neuroscience. Although many models have been proposed over the years, very few were able to provide empirically testable predictions. In this paper, we assume a model in which the cortex folds for all species of mammals according to a simple mechanism of effective free energy minimization of a growing self-avoiding surface subjected to inhomogeneous bulk stresses, to derive a new set of summary morphological variables that capture the most salient aspects of cortical shape and size. In terms of these new variables, we seek to understand the variance present in two morphometric datasets: a human MRI harmonized multi-site dataset comprised by 3324 healthy controls (CTL) from 4 to 96 years old and a collection of different mammalian cortices with morphological measurements extracted manually. This is done using a standard Principal Component Analysis (PCA) of the cortical morphometric space. We prove there is a remarkable coincidence (typically less than $8°$) between the resulting principal components vectors in each datasets and the directions corresponding to the new variables. This shows that the new, theoretically-derived variables are a set of natural and independent morphometrics with which to express cortical shape and size.

Keywords: PCA, cortical folding model, cortical folding universal rule.

1 Introduction

Why the cerebral cortex folds in some mammals is a fundamental question in neuroscience that still lacks a broadly accepted mechanism. It is reasonable to think about the cortical folding as a characteristic fixed by natural selection because it allows a dramatic increases in cortical size and complexity, while preserving the cranial size, axonal elongation and metabolic cost [31, 32]. However, the evolutionary argument is not sufficient to unravel the full process that results in the surface folding. Empirical evidence points towards an integrative explanation [18], unifying bio-mechanical, genetic and cellular signaling mechanism.

From a purely bio-mechanical point of view, there is a variety of models proposed to explain the physics of cortical folding: buckling from a constrained expansion [33, 34], warping due to differential tangential expansion [28] or emerging from axonal elongation [13]. Despite the qualitative compatibility with the cortical morphology, all these theories failed to provide a set of empirical testable predictions to further investigate their validity.

Following the bio-mechanical approach, [22] proposed a folding model hypothesizing that the cortex is a crumpled elastic surface in thermal equilibrium (like a crumpled paper ball) [15]. The model is based on the physics of minimization of effective free energy of a growing self-avoiding surface subject to inhomogeneous bulk stresses. It predicts a power-law relationship between cortical thickness (T), exposed (A_E), and total area (A_T) (Equation 1)

$$T^{\frac{1}{2}} A_T \;=\; k A_E^{\alpha} \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for which the exponent α is a universal constant with a theoretical value of 1.25 and k is a dimensionless parameter related to the mechanical properties of cortical tissue. The value of k not specified by the model, but is apparently highly conserved across various mammalian orders. The most immediate consequence of this model is that the usual set of geometrical variables $\{T^2, A_T, A_E\}$ used to describe the cortex should not be independent, but rather correlated in a highly restrictive way. Representing these variables in the $\{\log T^2, \log A_T, \log A_E\}$ space so that products of power laws become linear combinations of the log-variables, one can define

$$K \equiv \log k \;=\; \underbrace{\frac{1}{4}\log T^2 + \log A_T - \frac{5}{4}\log A_E}_{a(x-x_0)+b(y-y_0)+c(z-z_0)\;=\;0} \tag{2}$$

In this log-space, Equation (1) can be re-interpreted as an assertion that actual cortices must lie at or near a $K = constant$ plane, called henceforth the theoretical cortical plane. Thus, as long as it is constrained to a plane perpendicular to the vector $\vec{K} = (1/4, 1, -5/4)$ in log-space, cortical morphology should otherwise be compatible with any set of values $\{T^2, A_T, A_E\}$. Taking a step further, Equation 1 is a power-law relation between the intrinsic and extrinsic values for cortical areas, measured in unites of squared cortical thickness, suggesting that the cortex is self-similar over some range of length scales, i.e, approximately fractal with fractal dimension $f_{\text{dim}} = 2\alpha$.

The predictions of this simple statistical physics model were experimentally tested in a variety of studies. Firstly, it was found that for 63 different mammalian species cortical folding scales universally according Equation 1 with $\alpha_{\text{exp}} = 1.31 \pm 0.01$ (when considering a heterogeneous, i.e, multiple acquisition protocols dataset) [22]. This work was extended in [40], where it was shown that not only the universal law also holds within a single species MRI dataset with $\alpha_{\text{exp}} = 1.25 \pm 0.01$ compatible with theoretical predictions, but also provided evidence that a deviation from the universal rule could be used to identify diseases. The discrepancy between the observed α and the theoretical prediction for the comparative neuroanatomy study needs further investigation to determine if it is due to a systematic effect on the heterogeneous dataset or more profound information about the model's limitations. Then, in [23], it was observed that the white matter volume and white/gray matter ratio in mammalian species are a consequence of the universal scaling of cortical folding [42]. The first evidence of the self-similarity of the cortical surface was presented in [41] with experimental verification of the universal scaling law across lobes of the same cortex. More recently, [37] provided the first proof by construction of a universal archetypal fractal primate cortical shape, of which all actual cortices are approximations of a specific range of scales.

As a consequence, and regardless of the validity of the initial model, it is clear that the set of geometrical variables $\{T^2, A_T, A_E\}$ is indeed highly correlated. In [38], it was explicitly shown that by representing these variables in the $\{\log T^2, \log A_T, \log A_E\}$ space one can define K (Equation 2).

In log space, healthy human cortices lie at or near the theoretical cortical plane, defined by points for which $K = constant$, in agreement with the crumpled cortex model. In an attempt to understand the differences between cortices at or near the cortical plane, [38] derived a set of orthogonal morphological variables, starting with the nearly invariant K ($\vec{K} = (1/4, 1, -5/4)$) that is used to define the plane itself. Since k is dimensionless, it is not affected by an isometric scaling $\log T^2 \to \log \lambda T^2$, $\log A_T \to \log \lambda A_T$ and $\log A_E \to \log \lambda A_E$, in which λ is a positive real number; thus one can define both a direction $\vec{I} = \{1, 1, 1\}$ that correspond to this isometric scaling, and a corresponding new variable $I = \log T^2 + \log A_T + \log A_E$ that captures all information about cortical size. The third and last vector to both $\vec{K}$ and $\vec{I}$ ($\vec{S} = \vec{I} \times \vec{K}$), and is associated with the dimensionless variable $S = -\frac{9}{4}\log T^2 + \frac{3}{2}\log A_T + \frac{3}{4}\log A_E$ that would captures most of the variation in shape across actual cortices. These new variables are plausibly a natural way to decompose morphology in terms of its size; and its variant, and invariant shape components. Empirically, these new variables were shown to better discriminate between typical and atypical structural changes, like aging and diseases [38].

The existence of the cortical plane is by itself a significant empirical result for neuroscience. In principle, cortical morphology could be biologically compatible with any set of values $\{\log T^2, \log A_T, \log A_E\}$. In practice, it has been shown to be restricted to a plane. From a more instrumental perspective, this result implies that: (i) using the usual morphological variables without taking into account their correlations would severely distort statistical inference; (ii) there is clearly a direction in the log-space with the least morphological variation between cortices that is very close to the direction defined by the vector $\vec{K} = (1/4, 1, -5/4)$, as predicted by the crumpled cortex theory; (iii) diseases that affect cortical morphology could be characterized by particular deviations from and along the cortical plane in relation to healthy controls. It is also important to emphasize that although the variables K, S and I were derived from a specific model, their use to express and analyze morphology over the cortical plane is model-agnostic.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: The normalized vectors found by PCA and their respective deviation from theoretical expectation for the MRI humans dataset (single species). The components are ordered by the Explained Variance Ratio (EVR).

Dataset	Direction	Theoretical	PCA	θ_{diff}	EVR
20-40 years	$\hat{v}$	$\hat{n}_1 = (0, 0.781, 0.625)$	$(0.056, 0.704, 0.707)$	$7.9°$	0.65
	$\hat{u}$	$\hat{n}_2 = (-1, 0, 0)$	$(-0.996, 0.089, -0.009)$	$6.9°$	0.33
	$\hat{w}$	$\hat{K} = (-0.154, -0.617, 0.771)$	$(-0.070, -0.704, 0.707)$	$7.5°$	0.02
20-60 years (deaged)	$\hat{v}$	$\hat{n}_1 = (0, 0.781, 0.625)$	$(0.077, 0.703, 0.707)$	$7.9°$	0.65
	$\hat{u}$	$\hat{n}_2 = (-1, 0, 0)$	$(-0.994, 0.110, -0.0004)$	$6.9°$	0.33
	$\hat{w}$	$\hat{K} = (-0.154, -0.617, 0.771)$	$(-0.078, -0.703, 0.707)$	$7.5°$	0.02

The objective of this work is to test and systematize the description of cortical morphology, contrasting the old and new variables in terms of their statistical interdependence. To test the independence between the new variable set $\{K, S, I\}$ we performed a standard Principal Component Analysis (PCA) to obtain the independent components of variance in the $\{\log T^2, \log A_T, \log A_E\}$ space, for two different cases: across multiple species and across healthy human subjects (obtained from [3, 22]). The principal components (PC) expected for the first case are the independent set $\{K, S, I\}$ [41]. For humans, in addition to a nearly invariant K, it is known that the average cortical thickness varies very little in comparison to A_T and A_E, a direct consequence of being the same species. With the extra nearly invariant direction $\log T^2$, one would expect that the component with the most variance should lie along the $\vec{n}_1 = (0, 5/4, 1)$ direction. The remaining PC can then be directly found by the cross product $\vec{n}_2 = \vec{K} \times \vec{n}_1 = (-1, 0, 0)$. The PC vectors obtained from the data were then directly compared to the theoretical vectors $\{\vec{K}, \vec{S}, \vec{I}\}$ and a new set of vectors $\{\vec{K}, \vec{n}_1, \vec{n}_2\}$ derived in this work to describe a single species.

To test the statistical robustness of our results, a simulated dataset was generated according to Equation 1 to address the sensitivity of the PCA technique to the dataset's true principal directions. Finally, our results were validated by an alternative method. Rewriting Equation 1 as

$$K = \log k = \beta \log T + \gamma \log A_T - \alpha \log A_E \tag{3}$$

one realizes that the model prediction are actually the exponents $\alpha = 5/4$, $\beta = 1/2$ and $\gamma = 1$, and that obtaining the least variance PC corresponds to performing a joint fit of these parameters. So, the same task was performed with a standard maximum likelihood approach combining information from both humans and multiple species datasets.

2 Results

2.1 Human MRI multi-site harmonized dataset

Performing the PCA in the human harmonized dataset [1, 3, 9, 19, 25, 30, 40, 41], one obtain the results summarized in Figure 1. We analyzed separately subjects in the 20-40 yo age range, who presumably have a fully developed cortex (in broad terms, after development but before significant aging), and subjects in the age range between 20-60 years, for whom age effects were mathematically removed (henceforth "de-aged") [3]. Table 1 summarizes the differences between the principal components and the theoretical expectation. These differences do not change substantially between the two groups (see Appendix 5).

2.2 Comparative Neuroanatomy

The original work by Mota & Herculano-Houzel included two primary datasets of mammals, analyzed separately here: Herculano-Houzel's own data (referred here as dataset A) [10, 11, 16, 24, 27, 35], and previously published cortical morphological measurements (dataset B) [2, 5, 12, 20, 26]. Regarding the comparative neuroanatomy dataset, the PCA result is shown in Figure 2. The deviation between the principal components and the theoretical predictions are summarized in Table 2.

The combination of both datasets is discussed in Appendix 5, showing that this PCA method can detect subtle systematic differences between different datasets and underscores the need to properly harmonize said datasets before performing joint analyzes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

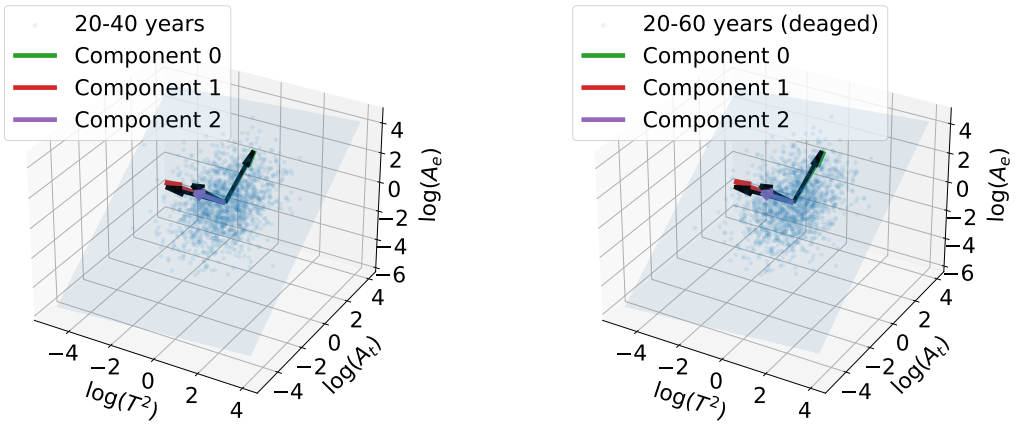

Figure 1: Cortical plane for the human MRI dataset and the vectors reconstructed by PCA. On the left panel, we considered only subjects in the age range of 20-40 years. On the right panel, the de-ageing procedure was applied, allowing the analysis of the full age range in which the model was experimentally validated (20 to 60 years old). The results of the two approaches are consistent. The arrows in black represent the principal components predicted by the model. The angular difference between theoretical components and the experimental ones are $\theta_{\text{diff}} < 8°$. The results are in remarkable agreement with the theoretical predictions.

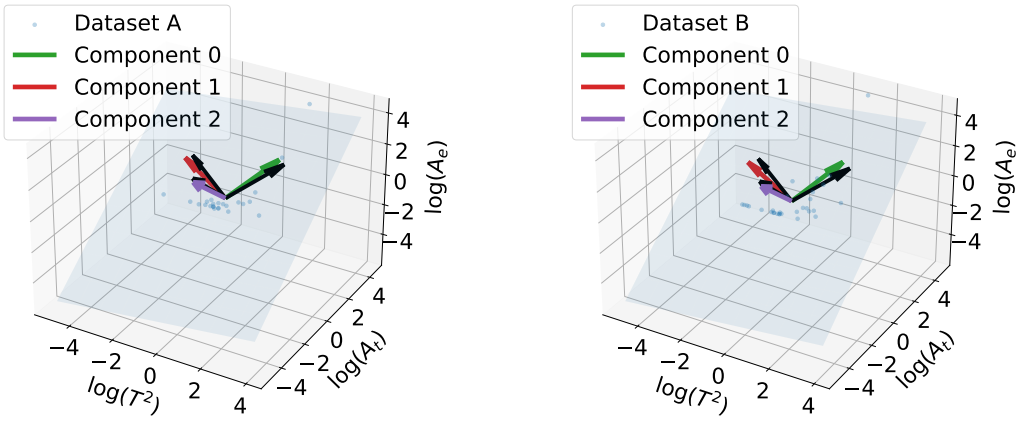

Figure 2: Cortical plane considering different mammalian species datasets and the vectors reconstructed by PCA. The arrows in black represent the principal components predicted by the model, the directions of the natural variables K, S, and I. To avoid systematic uncertainties from different datasets, we applied the analysis in datasets A (left panel) and B (right panel) separately. The angular difference between theoretical components and the experimental ones are $\theta_{\text{diff}} < 7°$. The results are remarkably in agreement with the predictions and consistent between distinct experiments, suggesting that these variables are indeed a natural mathematical language to describe the cortex, encapsulating orthogonal information about the morphology.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: The normalized vectors found by PCA and their respective deviation from theoretical expectation for comparative neuroanatomy. The components are ordered by the Explained Variance Ratio (EVR).

Dataset	Direction	Theoretical	PCA	θ_{diff}	EVR
A	$\hat{v}$	$\hat{I} = (0.577, 0.577, 0.577)$	$(0.490, 0.604, 0.628)$	$6.0°$	0.80
	$\hat{u}$	$\hat{S} = (-0.802, 0.534, 0.267)$	$(-0.863, 0.437, 0.253)$	$6.7°$	0.19
	$\hat{w}$	$\hat{K} = (-0.154, -0.617, 0.771)$	$(-0.121, -0.666, 0.736)$	$4.0°$	0.01
B	$\hat{v}$	$\hat{I} = (0.577, 0.577, 0.577)$	$(0.499, 0.602, 0.623)$	$5.4°$	0.81
	$\hat{u}$	$\hat{S} = (-0.802, 0.534, 0.267)$	$(-0.859, 0.439, 0.264)$	$6.4°$	0.18
	$\hat{w}$	$\hat{K} = (-0.154, -0.617, 0.771)$	$(-0.114, -0.666, 0.736)$	$4.1°$	0.01

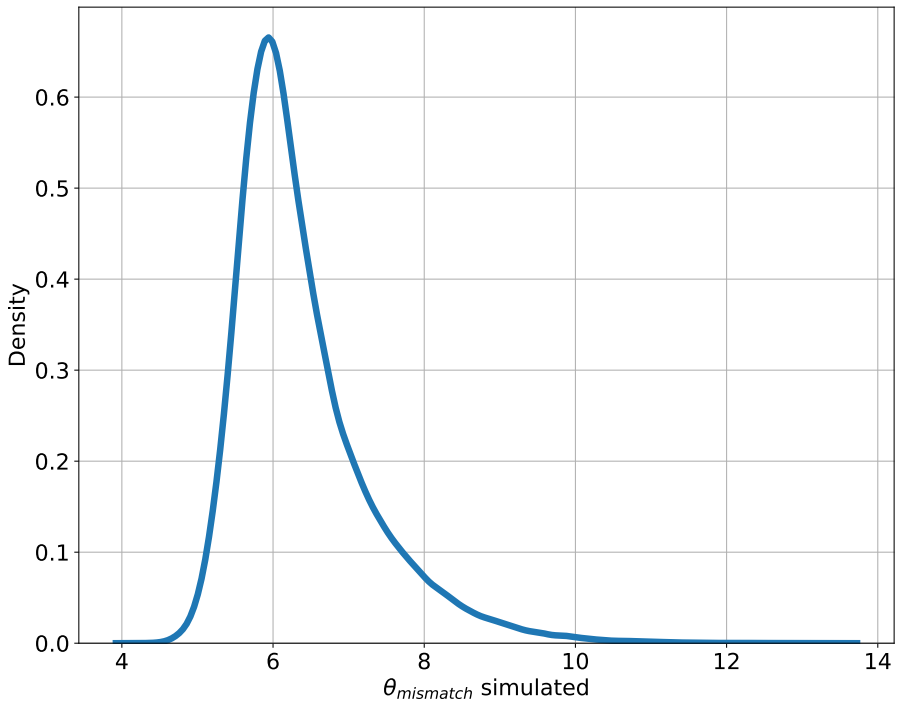

Figure 3: Angular difference θ_{diff} between the real principal component $\hat{K}$ and the one reconstructed by the PCA with the smallest explained variance. The deviation found is $\theta_{\text{diff}} = (6.2° + 2.5° - 0.9°)$.

2.3 Simulation of the PCA performance

In order to address the power of the PCA to recover the true principal components vectors in the comparative neuroanatomy datasets, we ran the same analysis using $N_{\text{sim}} = 10^5$ simulated datasets with known least variance PC direction. Figure 3 shows the distribution of the angle θ_{diff} measuring the difference between the theoretical direction and the ones reconstructed by the method. We found the deviation between the real $\hat{K}$ direction simulated and the one reconstructed by PCA was $\theta_{\text{diff}} = (6.2° + 2.5° - 0.9°)$ comparable to the values found when using real data.

2.4 Fitting the power-law

An alternative approach to the PCA is the direct determination of the exponents α, β and γ in Equation 3 using a standard maximum likelihood approach. The analysis starts performing a linear regression of the universal scaling relation in the particular case of lissencephalic mammals results in Figure 4. The angular coefficient $\frac{(\alpha-\gamma)}{\beta}$ gives prior information on the desired parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

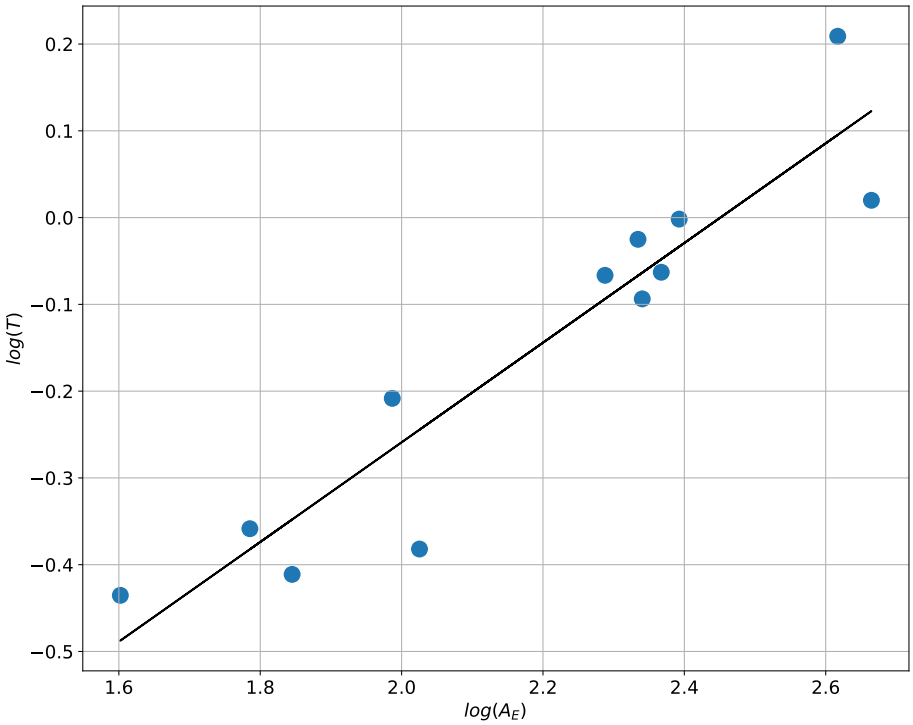

Figure 4: figure

Fitting the universal scaling relation in the particular case of lissencephalic mammals. The angular coefficient $\frac{(\alpha-\gamma)}{\beta} = (0.57 \pm 0.07)$ provides a constraint on the possible values for the exponents α, β and γ in Equation 3 and is in agreement with the theoretical predictions $\alpha = 5/4$, $\beta = 1/2$ and $\gamma = 1$.

One may note that the experimental angular coefficient (0.57 ± 0.07) is compatible with the theoretical predictions $((\alpha-\gamma)/\beta = (5/4-1)/(1/2) = 0.5)$, providing a constraint on the possible values for the exponents. Using this constraint, it is possible to fit the data from gyrencefalic species. Here the human MRI dataset was used because it has significantly more data. Figure 5 shows the log-likelihood as function of the exponents α and β.

3 Discussion

There is already some evidence that the cortical surface behaves as an elastic crumpled surface [22]. Within the framework of this theory, a set of orthogonal morphological variables $\{K, S, I\}$ was proposed as a more natural way of expressing a given cortical shape in the $\{\log T^2, \log A_T, \log A_E\}$ morphometric space [38]. The result presented here show that the directions defined by these new variables largely coincide with the principal components of variance for comparative neuroanatomy data of several dozen mammalian species from two independent datasets. Specifically, as expected, the $\hat{K}$ is very close to being the component with the least variance, reflecting the confining of data points at or close to the theoretical cortical plane. The new variables are thus shown to be almost uncorrelated with one another and can be properly regarded as a set of both natural and independent morphometrics, greatly enhancing their potential analytical usefulness.

An additional result derived from the comparative neuroanatomy dataset suggested by the analysis, and further discussed in Appendix 5, is the possible presence of subtle systematic differences between datasets A and B. This underscores the need to properly harmonize this data before performing joint analyzes and may explain the discrepancy between the observed $\alpha_{\text{exp}} = 1.31 \pm 0.01$ and the theoretical prediction obtained by [22].

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

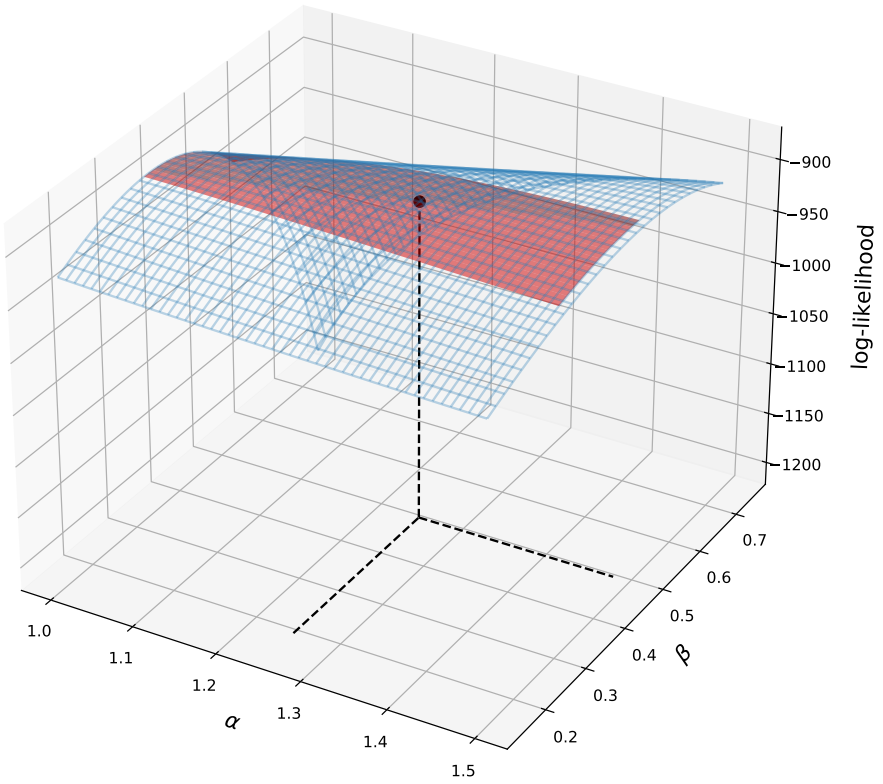

Figure 5: figure
Log-likelihood function with the constraint established by lissencephalic data. Instead of a function peaked around a localized region, it is observed a large range of equivalent peaks. The maximum of the likelihood function with 1 σ confidence region represented in red. Remarkably, the theoretical expectation (black dot) is in agreement with the values obtained by the regression.

In the human dataset we also observe a remarkable agreement between the directions $\{\hat{K}, \hat{n}_1, \hat{n}_2\}$, obtained by further imposing the single species constraint of restricted cortical thickness variation, and the principal components of variance. Again, the $\hat{K}$ is very close to being the component with the least variance, reflecting the confining of data points at or close to the theoretical cortical plane. This result suggests that the apparent correlation between S and I in humans is due to the fact that the subjects belong to the same species, a constraint not taken into account when deriving these variables. More than that, in terms of practical application, it means that although the average cortical thickness is nearly invariant within a species and could be used as a biomarker for structural changes in the cortex, it is not an optimal one, as this constraint is evolutionarily contingent. Thus, even for human data, it is valuable to express morphology in terms of the K, S and I variables, as suggested in [17,41], as these variable independently capture most of the global morphological variation in cortex. This approach contrasts with previous studies such as [4,6,8], and may in the end produce more robust biomarkers.

The degree of agreement between morphological variables and principal components was quantified using a simulated dataset, generated with principal components given by the exponents in Equation 1 and the same variance as present in the analyzed dataset. It was found that the deviation between the real $\hat{K}$ direction simulated and the one reconstructed by PCA was $\theta_{\text{diff}} = (6.2^\circ + 2.5^\circ - 0.9^\circ)$ (see Figure 3), a value consistent with the one observed in the experimental data.

A deeper explanation of the systematic discrepancy observed between the principal components, the theoretically predicted vectors, and those obtained using the PCA is given by the result of the maximum likelihood approach. The PCA cannot reconstruct the vectorial components with the required precision, reflecting exactly on the distribution

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: Summary of the subjects used in the human MRI dataset.

Datasets	Diagnostic	N (F)	Age (Age range) [years]
AHEAD [1]	CTL	100 (56)	42 ± 19 (24; 76)
AOMIC PIOP01 [30]	CTL	208 (120)	22 ± 1.8 (18; 26)
AOMIC PIOP02 [30]	CTL	224 (128)	22 ± 1.8 (18; 26)
HCP900r [9]	CTL	881 (494)	29 ± 3.6 (22; +36)
IXI-Guy's	CTL	314 (175)	51 ± 16 (20; 86)
IXI-HH	CTL	181 (94)	47 ± 17 (20; 82)
IXI-IOP	CTL	68 (44)	42 ± 17 (20; 86)
NKI [25]	CTL	168 (68)	34 ± 19 (4; 85)
OASIS [19]	CTL	312 (196)	45 ± 24 (18; 94)
TOTAL	**CTL**	**2456 (1762)**	**46 ± 23 (4; 96)**

shown in Figure 3. The explanation for this impossibility becomes clearer when one realizes that the PCA is a way to reconstruct all exponents in Equation 1 and 3. As shown in Figure 5, instead of a function peaked around a localized region, one obtains a large range of equivalent peaks that, although it is compatible with the theoretical values, fitting these values with high precision is not possible. However, the fact that the all exponents of the power-law were in agreement with the observation is an astonishing new positive evidence in favor of the crumpled cortex model.

The model originally proposed by Mota & Herculano-Houzel suggests that the cortical morphology approximates a real fractal with fractal dimension $f_{\text{dim}} = 2\alpha$. Different aspects of this model were tested and found to be in agreement with a series of experimental results. Taken together, these results help determine the regime of validity for this model, by mapping the cases in which the overall morphology of the cortical surface can be modeled using this theoretical framework, opening the possibility of biologically interpreting morphological phenomena as cortical folding patterns or even creating reliable simulations of the cortical structure and geometry. The biological interpretation of the cortical folding provided by the Mota & Herculano-Houzel model is a profound statement about its origins. Instead of being a new adaptive property specific to some clades, cortical gyrification is a consequence of the thermodynamic equilibrium of the cortical surface, a mechanism that must be universal to all mammals. This universality is reflected in the scaling of self-similar cortical morphology and its restriction to the cortical plane. Essentially, the difference between a gyrencefalic and a lissencefalic developmental end-states is simply the rate of increase of A_T relative to T^2. The cellular signaling mechanism and genetics regulate the cortical development processes that will result or not in a crumpled surface according to the evolution of each clade and how is the relation between A_T and T growing rates. But ultimately, the final surface will reach thermodynamic equilibrium and will result in a self-similar structure with fractal dimension $f_{\text{dim}} = 2\alpha$

This work makes an addition to the empirical evidence corroborating the crumpled cortex model, showing that not only the principal components of cortical morphological variables are in agreement with the theoretical prediction but also it suggests that the independent set of morphological variables K, S, and I are indeed a natural mathematical language to describe the cortex, encapsulating orthogonal information about the morphology. Also, by finding the direction $\hat{K}$ very close to being the component with the least variance in multiple datasets from different mammals, we observe the cortical plane as evidence of a conserved physical mechanism leading the cortical folding. Furthermore, this study shows the possibility of understanding cortical morphology using a unique mathematical description, advocating in favor of incorporating compared neuroanatomy insights in studies of human diseases. The variables discussed here to describe cortical morphology have an immediate practical application when considering disease discrimination and healthy aging studies. Regarding the global morphometry of the cortex, these variables gives the correct treatment to inherent correlation given by the cortical plane. Finally, we also provided a remarkable new evidence in favor of the crumpled cortex model by verifying that each theoretical exponent of the power-law is supported empirically.

4 Methods and Materials

4.1 Dataset

The morphological data used in this analysis were all previously published. Human participants and their respective MRI-derived metrics were either published by [40, 41] or from open-access databases, as AHEAD [1] and AOMIC (PIOP01 and PIOP02) [30] (human subjects are described in Table 3). The total number of human subjects MRI included in this work is 2456 healthy controls (CTL) from 4 to 96 years old.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The structural images from AHEAD, AOMIC ID1000, PIOP1, and PIOP2 were processed in FreeSurfer v6.0.0 [7] without manual intervention at the surfaces [21]. The FreeSurfer localGI pipeline generates the external surface and calculates each vertex's local Gyrification Index (localGI) [29]. Values of Cortical Thickness, Total Area, and Exposed Area were extracted with Cortical Folding Analysis Tool [39] for both hemispheres.

In order to perform a joint analysis, the morphological variables were harmonized following the prescription given in [3]. The data harmonization procedure is based on the time evolution of the basic morphological variables T, A_T, or A_E, modeled as an exponential decay with the age t_{age}. A joint linear fit was performed in the log-linear scale, maintaining the same angular coefficient among all different samples but different linear coefficients. The linear coefficients were then subtracted to harmonize all datasets.

The time evolution of the morphological variables itself can significantly impact the PCA results and the posterior interpretation. To avoid this problem, we used two techniques: (i) restrict the analysis to the age range between 20-40 years (adulthood) as it would be the best representation of a stationary fully developed cortex; (ii) take advantage of the harmonization linear regression, removing the age dependence using each subject age and the angular coefficient fitted, a procedure called here deageing. The latter allows to perform the analysis to the full age range (20-60 years) in which the model was experimentally validated [3]. In Appendix 5, one finds the replication of the analysis for all age ranges on the dataset (with and without the deageing).

Regarding the mammalian dataset from multiple clades, this work used the extracted morphological variables made available by the authors at [22]. This dataset also presents heterogeneity in acquisition, compiling Mota & Herculano-Houzel's group data (referred here as dataset A) [10, 11, 16, 24, 27, 35] with previously published cortical morphological measurements (dataset B) [2, 5, 12, 20, 26]. The details about each acquisition protocol can be found in the original publications. This variability can also significantly impact the PCA results due to systematic uncertainties from different experiments. To avoid this problem, we applied the analysis using datasets A and B separately. In Appendix 5, one finds the replication of the analysis for the datasets combined.

Principal Component Analysis

Principal Component Analysis (PCA) is a standard procedure in multivariate data analysis, usually used to dimensionality reduction in a dataset [14]. With a direct analogy to the calculation of a rigid body principal axes of inertia, it creates a new set of uncorrelated variables, the principal components, solving an eigenvalue/eigenvector problem for the covariance matrix of a distribution or set of points. In this work, the PCA was performed numerically using Python SciPy library [36]. Across this article, the vectors $\hat{v} = (v_1, v_2, v_3)$, $\hat{u} = (u_1, u_2, u_3)$ and $\hat{w} = (w_1, w_2, w_3)$ represent the normalized principal components ordered by the size of explained variance.

Considering the $\{\log T^2, \log A_T, \log A_E\}$ space, the universal scaling law strongly suggests that the vector $\vec{K} = (1/2, 1, 5/4)$ (Equation 2), must be the principal component with the lowest explained variance. It is a near-invariant quantity obtained by isolating k in Equation 1, thus defining the theoretical cortical plane, and in its initial formulation it was associated with conserved viscoelastic properties of cerebral cortical matter ($k \propto (f_S + p_{\text{CSF}})^{1/2}$ [15,22]). For humans, for whom in addition to a nearly invariant K one also finds that the average cortical thickness T varies little [41], one would expect only small variations along the $\log T^2$ direction. Thus, the component with the most variance should lie along the $\vec{n}_1 = (0, 5/4, 1)$ direction. The remaining PC is directly found by the cross product $\vec{n}_2 = \vec{K} \times \vec{n}_1 = (-1, 0, 0)$. The normalized vectors that will be used in comparison with the PCA are denote here as $\hat{n}_1 = (0, 0.781, 0.625)$, $\hat{n}_2 = (-1, 0, 0)$, and $\hat{K} = (-0.154, -0.617, 0.771)$.

For the comparative neuroanatomy datasets, there is substantial variation in cortical thickness and areas. Following the interpretation in [38], we predict that the principal components in this case are the normalized directions defined by the independent morphological variables K, S and I given by $\hat{I} = (0.577, 0.577, 0.577)$, $\hat{S} = (-0.802, 0.534, 0.267)$, and $\hat{K} = (-0.154, -0.617, 0.771)$. If the variables $\{K, S, I\}$ are actually a natural set of morphological variables, capturing different and orthogonal aspects of the cortex, their respective vectors should be the principal components in the $\{\log T^2, \log A_T, \log A_E\}$ space.

4.2 Statistical Significance of the PCA results

To better interpret the results above, we used Monte Carlo simulations as a tool to address the power of the analysis to recover the true principal components' vectors and address the statistical significance of the results. The data was generated over $N_{\text{sim}} = 10^5$ iterations using a Monte Carlo method to sample points near the theoretical cortical plane defined by Equation 2. The number of points generated and their variance was chosen to match the available data on the human MRI dataset. In this simulated dataset, we applied the same PCA procedure used to analyze the real data,

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

calculating the distribution for the angle θ_{diff} between the real principal component $\vec{K}$ and the component in the PCA with the least explained variance for each simulated dataset.

4.3 Fitting the power-law

Assuming a relation in the form of Equation 3, the comparison between PCA least variance component and the theoretical vector is equivalent to perform a joint fit of the exponents α, β and γ and test if the values are compatible with the theoretical ones, $\alpha = 5/4$, $\beta = 1/2$ and $\gamma = 1$. This alternative approach is done using a standard maximum likelihood approach. Firstly, we start fitting the universal scaling relation in the particular case of lissencephalic mammals for which the relation scales as $\log T = \gamma \log k + \frac{(\alpha-\gamma)}{\beta} \log A_E$. The angular coefficient $\frac{(\alpha-\gamma)}{\beta}$ work as a constraint or prior information that can be used fit the data from gyrencefalic species obtaining all parameters.

4.4 Data and code availability

All codes and processed datasets are available at `https://github.com/bragamello/metaBIO/tree/main/Universal_cortical_rule`. AHEAD and AOMIC (PIOP01 and PIOP02) data is available at `https://zenodo.org/record/5750619`.

5 Acknowledgments

This work was supported by Instituto Serrapilheira (grant Serra-1709-16981) and CNPq (PQ 2017 312837/2017-8). We are thankful to the research team and volunteers for participating in this research project. We especially thank Dr. Yujiang Wang for the careful reading and commenting on this manuscript, and for many invaluable suggestions for improvement. The authors have no competing interests.

References

[1] Anneke Alkemade, Martijn J Mulder, Josephine M Groot, Bethany R Isaacs, Nikita van Berendonk, Nicky Lute, Scott JS Isherwood, Pierre-Louis Bazin, and Birte U Forstmann. The Amsterdam Ultra-high field adult lifespan database (AHEAD): A freely available multimodal 7 Tesla submillimeter magnetic resonance imaging database. *NeuroImage*, 221:117200, November 2020.

[2] Frederico A.C. Azevedo, Ludmila R.B. Carvalho, Lea T. Grinberg, José Marcelo Farfel, Renata E.L. Ferretti, Renata E.P. Leite, Wilson Jacob Filho, Roberto Lent, and Suzana Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[3] Fernanda H. P. de Moraes, Victor B. B. Mello, Fernanda Tovar-Moll, and Bruno Mota. Establishing a baseline for human cortical folding morphological variables: A multisite study. *Frontiers in Neuroscience*, 16, 2022.

[4] Bradford C. Dickerson and David A. Wolk. Mri cortical thickness biomarker predicts ad-like csf and cognitive decline in normal adults. *Neurology*, 78(2):84–90, 2012.

[5] Hans Elias and David Schwartz. Cerebro-cortical surface areas, volumes, lengths of gyri and their interdependence in mammals, including man. *Z. Saugetierkd*, 36:147–163, 1971.

[6] Christine Fennema-Notestine, Donald J Hagler, Jr, Linda K McEvoy, Adam S Fleisher, Elaine H Wu, David S Karow, Anders M Dale, and Alzheimer's Disease Neuroimaging Initiative. Structural MRI biomarkers for preclinical and mild alzheimer's disease. *Hum Brain Mapp*, 30(10):3238–3253, October 2009.

[7] Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, August 2012.

[8] Ladan Ghazi-Saidi, Ryan R Walsh, Guogen Shan, and Sarah J Banks. Biomarkers of cognitive impairment: Brain cortical thickness, volumetrics, and cerebrospinal fluid. *Alzheimer Dis Assoc Disord*, 32(3):255–257, July 2018.

[9] Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013. Mapping the Connectome.

[10] Suzana Herculano-Houzel, Kamilla Avelino-de Souza, Kleber Neves, Jairo Porfírio, Débora Messeder, Larissa Mattos Feijó, José Maldonado, and Paul R. Manger. The elephant brain in numbers. *Frontiers in Neuroanatomy*, 8, 2014.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[11] Suzana Herculano-Houzel, Bruno Mota, Peiyan Wong, and Jon H. Kaas. Connectivity-driven white matter scaling and folding in primate cerebral cortex. *Proceedings of the National Academy of Sciences*, 107(44):19008–19013, 2010.

[12] M. A. Hofman. Size and shape of the cerebral cortex in mammals (part 1 of 2). *Brain, Behavior and Evolution*, 27(1):28–40, 1985.

[13] Maria A. Holland, Kyle E. Miller, and Ellen Kuhl. Emerging brain morphologies from axonal elongation. *Annals of Biomedical Engineering*, 43(7):1640–1653, Jul 2015.

[14] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[15] Yacov Kantor, Mehran Kardar, and David R. Nelson. Statistical mechanics of tethered surfaces. *Phys. Rev. Lett.*, 57:791–794, Aug 1986.

[16] Rodrigo S. Kazu, José Maldonado, Bruno Mota, Paul R. Manger, and Suzana Herculano-Houzel. Cellular scaling rules for the brain of artiodactyla include a highly folded cortex with few neurons. *Frontiers in Neuroanatomy*, 8, 2014.

[17] Karoline Leiberg, Christoforos Papasavvas, and Yujiang Wang. Local morphological measures confirm that folding within small partitions of the human cortex follows universal scaling law, 2021.

[18] Cristina Llinares-Benadero and Víctor Borrell. Deconstructing cortical folding: genetic, cellular and mechanical determinants. *Nature Reviews. Neuroscience*, 20(3):161–176, March 2019.

[19] Daniel S. Marcus, Anthony F. Fotenos, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*, 22(12):2677–2684, 12 2010.

[20] T M Mayhew, G L Mwamengele, V Dantzer, and S Williams. The gyrification of mammalian cerebral cortex: quantitative evidence of anisomorphic surface expansion during phylogenetic and ontogenetic development. *J Anat*, 188 (Pt 1)(Pt 1):53–58, feb 1996.

[21] Christopher S. McCarthy, Avinash Ramprashad, Carlie Thompson, Jo-Anna Botti, Ioana L. Coman, and Wendy R. Kates. A comparison of FreeSurfer-generated data with and without manual intervention. *Frontiers in Neuroscience*, 9, 2015.

[22] Bruno Mota and Suzana Herculano-Houzel. Cortical folding scales universally with surface area and thickness, not number of neurons. *Science (New York, N.Y.)*, 349(6243):74–77, July 2015.

[23] Bruno Mota, Sandra E. Dos Santos, Lissa Ventura-Antunes, Débora Jardim-Messeder, Kleber Neves, Rodrigo S. Kazu, Stephen Noctor, Kelly Lambert, Mads F. Bertelsen, Paul R. Manger, Chet C. Sherwood, Jon H. Kaas, and Suzana Herculano-Houzel. White matter volume and white/gray matter ratio in mammalian species as a consequence of the universal scaling of cortical folding. *Proceedings of the National Academy of Sciences*, 116(30):15253–15261, 2019.

[24] Kleber Neves, Fernanda Meireles Ferreira, Fernanda Tovar-Moll, Nadine Gravett, Nigel Bennett, Consolate Kaswera, Emmanuel Gilissen, Paul Manger, and Suzana Herculano-Houzel. Cellular scaling rules for the brain of afrotherians. *Frontiers in Neuroanatomy*, 8, 2014.

[25] Kate Nooner, Stanley Colcombe, Russell Tobe, Maarten Mennes, Melissa Benedict, Alexis Moreno, Laura Panek, Shaquanna Brown, Stephen Zavitz, Qingyang Li, Sharad Sikka, David Gutman, Saroja Bangaru, Rochelle Tziona Schlachter, Stephanie Kamiel, Ayesha Anwar, Caitlin Hinz, Michelle Kaplan, Anna Rachlin, Samantha Adelsberg, Brian Cheung, Ranjit Khanuja, Chaogan Yan, Cameron Craddock, Vincent Calhoun, William Courtney, Margaret King, Dylan Wood, Christine Cox, Clare Kelly, Adriana DiMartino, Eva Petkova, Philip Reiss, Nancy Duan, Dawn Thompsen, Bharat Biswal, Barbara Coffey, Matthew Hoptman, Daniel Javitt, Nunzio Pomara, John Sidtis, Harold Koplewicz, Francisco Castellanos, Bennett Leventhal, and Michael Milham. The nki-rockland sample: A model for accelerating the pace of discovery science in psychiatry. *Frontiers in Neuroscience*, 6, 2012.

[26] R. L. Reep and T. J. O'Shea. Regional brain morphometry and lissencephaly in the sirenia. *Brain, Behavior and Evolution*, 35(4):185–194, 1990.

[27] Pedro Ribeiro, Lissa Ventura-Antunes, Mariana Gabi, Bruno Mota, Lea Grinberg, José Marcelo Farfel, Renata Ferretti, Renata Leite, Wilson Jacob Filho, and Suzana Herculano-Houzel. The human cerebral cortex is neither one nor many: neuronal distribution reveals two quantitatively different zones in the gray matter, three in the white matter, and explains local variations in cortical folding. *Frontiers in Neuroanatomy*, 7, 2013.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[28] Lisa Ronan, Natalie Voets, Catarina Rua, Aaron Alexander-Bloch, Morgan Hough, Clare Mackay, Tim J. Crow, Anthony James, Jay N. Giedd, and Paul C. Fletcher. Differential tangential expansion as a mechanism for cortical gyrification. *Cerebral cortex (New York, N.Y. : 1991)*, 24(8):2219–2228, Aug 2014.

[29] M. Schaer, M. B. Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J. Thiran. A Surface-Based Approach to Quantify Local Cortical Gyrification. *IEEE Transactions on Medical Imaging*, 27(2):161–170, February 2008.

[30] Lukas Snoek, Maite M. van der Miesen, Tinka Beemsterboer, Andries van der Leij, Annemarie Eigenhuis, and H. Steven Scholte. The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. *Scientific Data*, 8(1):85, March 2021.

[31] Georg F. Striedter, Shyam Srinivasan, and Edwin S. Monuki. Cortical folding: When, where, how, and why? *Annual Review of Neuroscience*, 38(1):291–307, 2015. PMID: 25897870.

[32] George F. Striedter. *Principles of brain evolution.* Principles of brain evolution. Sinauer Associates, Sunderland, MA, US, 2005.

[33] Tuomas Tallinen, Jun Young Chung, John S. Biggins, and L. Mahadevan. Gyrification from constrained cortical expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35):12667–12672, Sep 2014.

[34] Tuomas Tallinen, Jun Young Chung, François Rousseau, Nadine Girard, Julien Lefèvre, and L. Mahadevan. On the growth and form of cortical convolutions. *Nature Physics*, 12(6):588–593, Jun 2016.

[35] Lissa Ventura-Antunes, Bruno Mota, and Suzana Herculano-Houzel. Different scaling of white matter volume, cortical connectivity, and gyrification across rodent and primate brains. *Frontiers in Neuroanatomy*, 7, 2013.

[36] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[37] Yujiang Wang, Karoline Leiberg, Nathan Kindred, Christopher R. Madan, Colline Poirier, Christopher I. Petkov, Peter N. Taylor, and Bruno C. C. Mota. Neuro-evolutionary evidence for a universal fractal primate brain shape, 2022.

[38] Yujiang Wang, Karoline Leiberg, Tobias Ludwig, Bethany Little, Joe H Necus, Gavin Winston, Sjoerd B Vos, Jane de Tisi, John S Duncan, Peter N Taylor, and Bruno Mota. Independent components of human brain morphology. *NeuroImage*, 226:117546, February 2021.

[39] Yujiang Wang, Tobias Ludwig, and Bruno Mota. Analysis pipeline to extract cortical morphology measures from Freesurfer for cortical folding analysis. https://zenodo.org/record/3608675, oct 2019.

[40] Yujiang Wang, Joe Necus, Marcus Kaiser, and Bruno Mota. Universality in human cortical folding in health and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 113(45):12820–12825, November 2016.

[41] Yujiang Wang, Joe Necus, Luis Peraza Rodriguez, Peter Neal Taylor, and Bruno Mota. Human cortical folding across regions within individual brains follows universal scaling law. *Communications Biology*, 2(1):191, May 2019.

[42] Kechen Zhang and Terrence J. Sejnowski. A universal scaling law between gray matter and white matter of cerebral cortex. *Proceedings of the National Academy of Sciences*, 97(10):5621–5626, 2000.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendix A

Table 4: The normalized vectors found by PCA and their respective deviation from theoretical expectation combining comparative neuroanatomy datasets A and B. The components are ordered by the Explained Variance Ratio (EVR).

Direction	Theoretical	PCA	θ_{diff}	EVR
$\hat{v}$	$\hat{I} = (0.577, 0.577, 0.577)$	$(0.417, 0.632, 0.653)$	$10.6.0°$	0.74
$\hat{u}$	$\hat{S} = (-0.802, 0.534, 0.267)$	$(-0.904, 0.364, 0.224)$	$11.7°$	0.25
$\hat{w}$	$\hat{K} = (-0.154, -0.617, 0.771)$	$(-0.096, -0.684, 0.723)$	$5.8°$	0.01

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

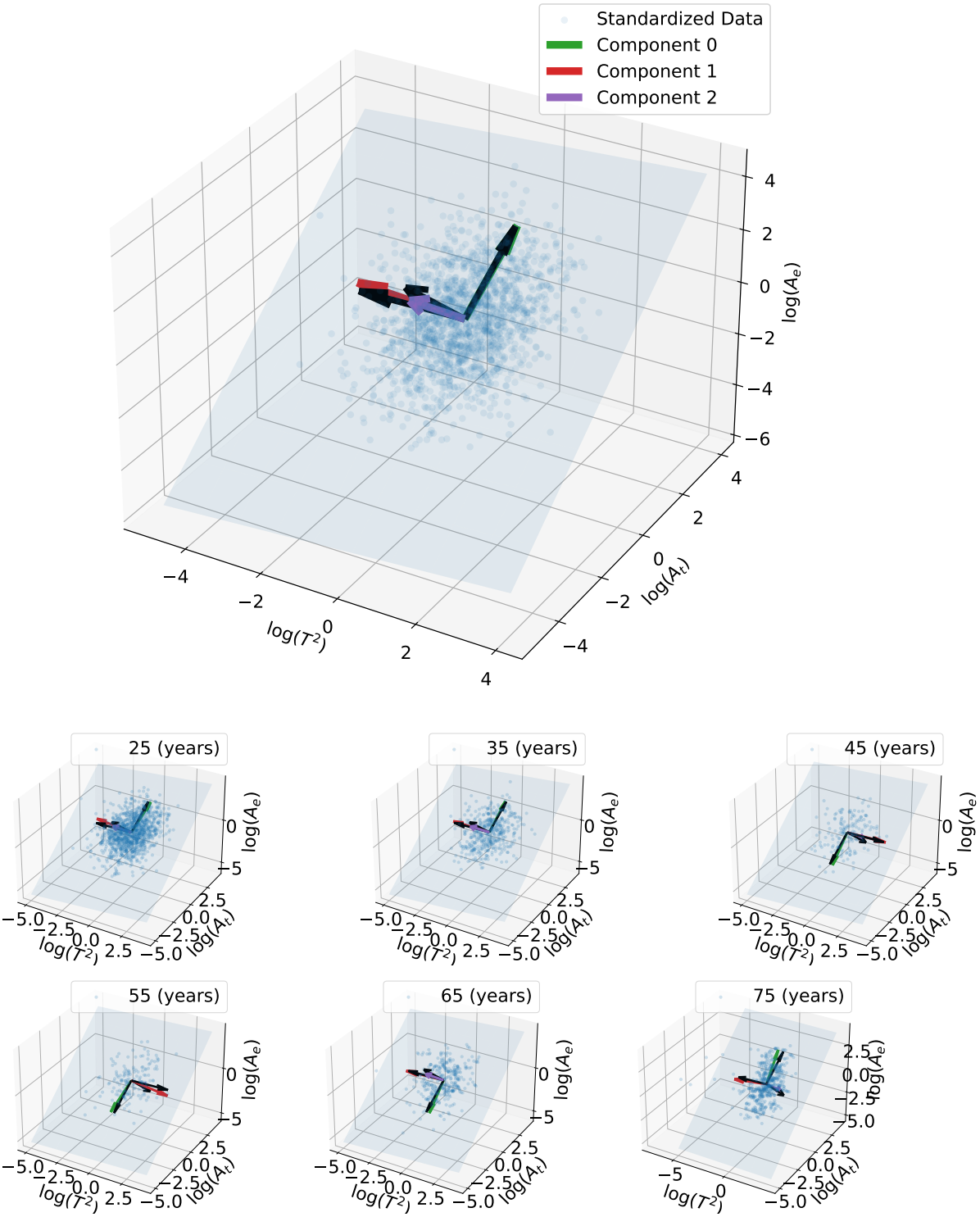

Figure 6: figure
Cortical plane for the human MRI harmonized dataset and the vectors reconstructed by PCA for different age ranges. The arrows in black represent the principal components predicted by the model. An angular difference of 180° on the vectors reflects the ambiguity of selecting the vector normal to the plane.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

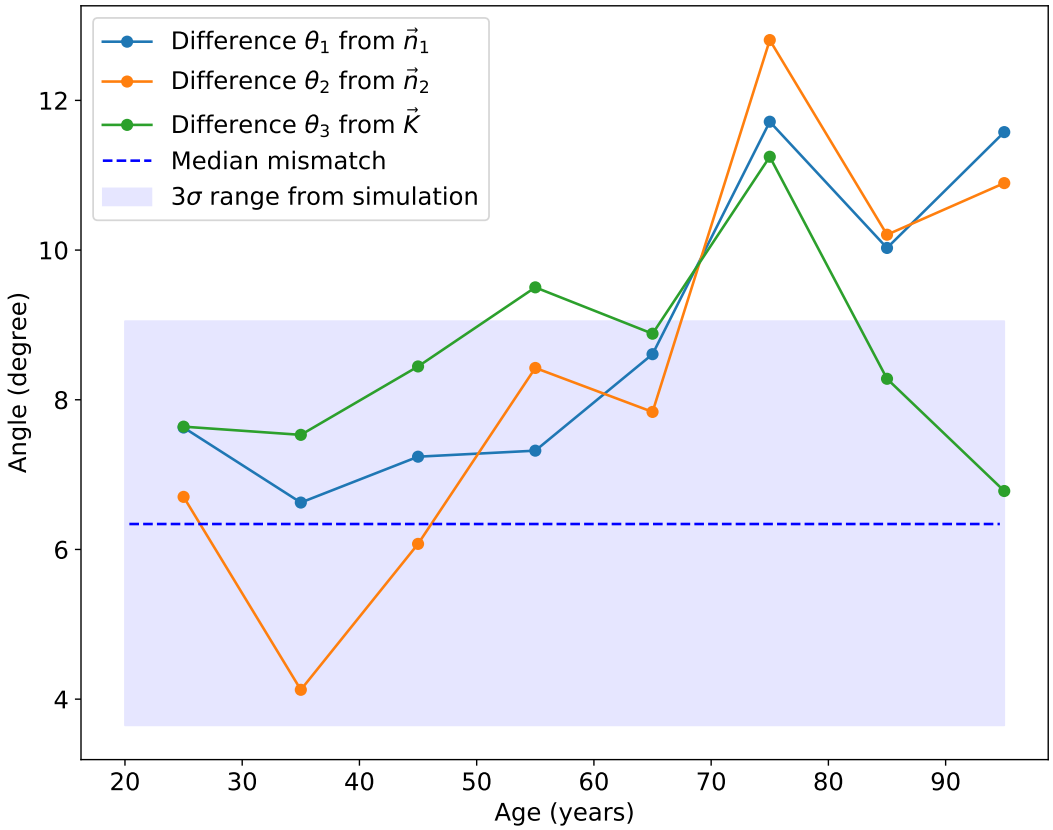

Figure 7: figure
Angular difference between the vectors reconstructed by PCA and the theoretical predictions across healthy aging. The 3σ region was obtained using the simulations. The results are remarkably in agreement with the predictions when considering ages below 65 years. Above this age, there is a clear change in the directions with the greatest variability. This deviation from the theoretical predictions correlated with age can be further exploited in future works in order to build a theoretical model for the morphology of healthy cortical aging.

380

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

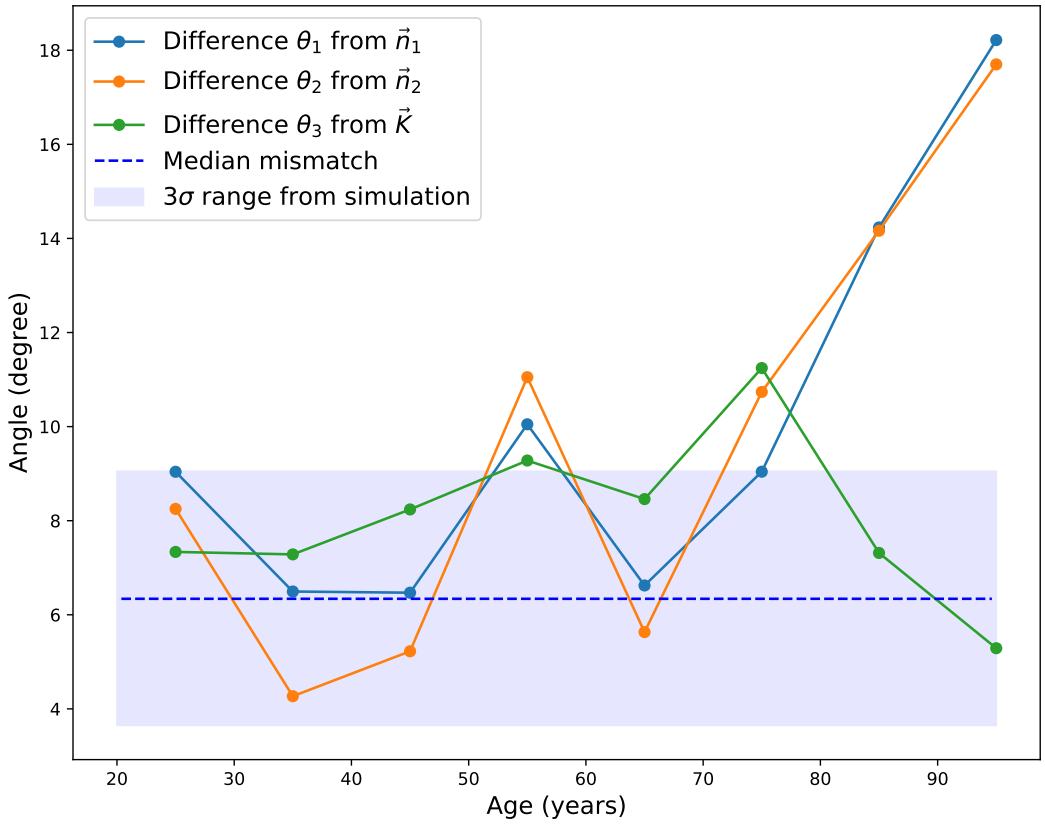

Figure 8: figure
Angular difference between the vectors reconstructed by PCA and the theoretical predictions across healthy aging using the de-aged dataset. The 3σ region was obtained using the simulations. The results are remarkably in agreement with the predictions when considering ages below 65 years. As expected, without the age dependency, the results are constant until the deviation from the model after 65 years. Above this age, there is a clear change in the directions with the greatest variability. This deviation from the theoretical predictions correlated with age will be further exploited in future works in order to build a theoretical model for the morphology of healthy cortical aging.

bioRxiv preprint doi: https://doi.org/10.1101/2022.07.07.499214; this version posted March 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendix B

The PCA applied to the combined mammalian dataset, as shown in Figure 9, presents a larger deviation from the theoretical prediction. The vectors found are summarized in table 4.

There is a statistically significant increase in the angular difference from $\hat{S}$ and $\hat{I}$. In comparison with the result presented in the main text, in which both datasets had consistent findings when analyzed separately, there is an indication of a relative unaccounted systematic error between the datasets. This illustrates how this PCA method can be used to detect subtle systematic differences between different datasets and underscores the need to properly harmonize said datasets before performing joint analyses. The suggestion of an unaccounted systematic error between datasets A and B should be further investigated by analyzing the experimental errors of data acquisition, having the potential to elucidate the experimental discrepancy between the reported index $\alpha = 1.32$ and the theoretical $\frac{5}{4}$.

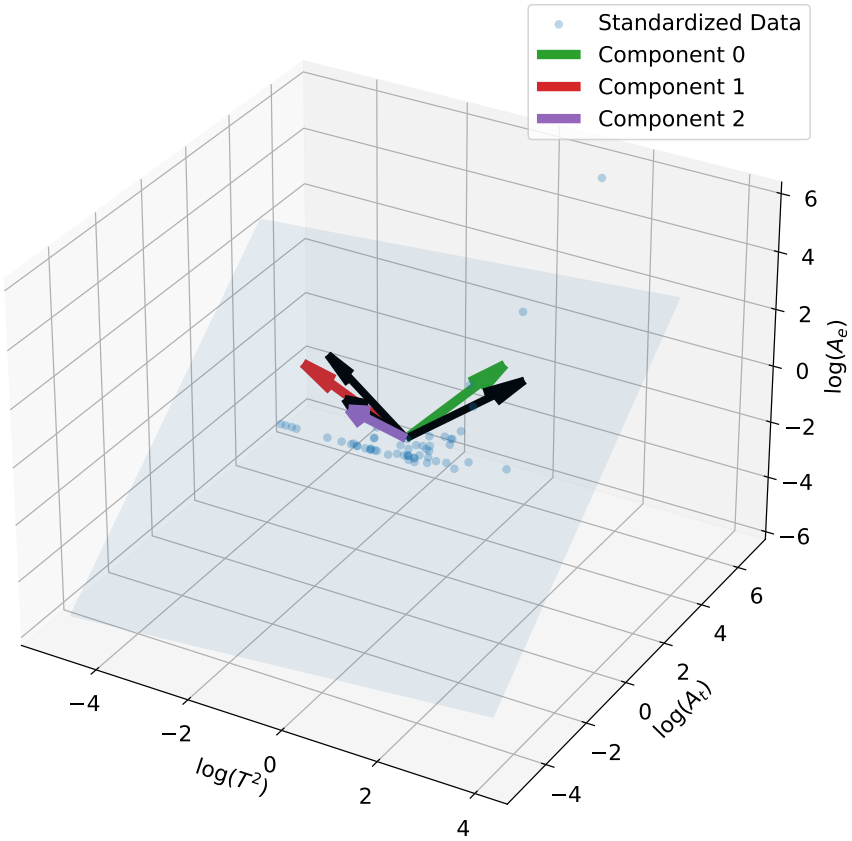

Figure 9: figure
Cortical plane considering different mammalian species datasets and the vectors reconstructed by PCA. The arrows in black represent the principal components predicted by the model, the directions of the natural variables K, S, and I. In this analysis, datasets A and B were used together. There is an increase in the angular difference between theoretical principal components $\hat{S}$ and $\hat{I}$ and the experimental ones. This result is interpreted as indicative of a relative uncounted systematic error between the datasets.