Morphometrical Data Analysis using Wavelets

C. M. Takemura [a,1], R. M. Cesar-Jr. [a,2], R. A. T. Arantes [b], L. da F. Costa [b], E. Hingst-Zaher [c], S. F. dos Reis [d], V. Bonato [d]

[a] *USP – IME, Rua do Matão, 1010 – Cidade Universitária CEP: 05508-090, São Paulo, SP, Brasil*

[b] *USP – IFSC, Av. Trabalhador Sãocarlense, 400 Caixa Postal 369, CEP: 13560-970, São Carlos, SP, Brasil*

[c] *USP – MZUSP, Av. Nazaré, 481 – Bairro do Ipiranga CEP: 04263-000, São Paulo, SP, Brasil*

[d] *UNICAMP – IB, Cidade Universitária "Zeferino Vaz" Barão Geraldo, Campinas, SP, Brasil*

Abstract

In this paper we present a new shape analysis approach using the well known wavelet transform and exploring shape representation by landmarks. Firstly, we will describe the approach used to transform the landmark representation to a signal representation. Then we show the relation of the derivatives of Gaussian wavelet transform applied to the signal to differential properties of the shape that it represents. Experimental results are presented using real data to show how it is possible to characterize shapes through multiscale and differential signal processing techniques in order to relate morphology variables with phylogenetics, environmental and sexual factors. The goal of this research is to develop an effective wavelet transform-based method to represent and classify multiple classes shapes given by landmarks.

Key words: wavelets, shape representation, morphometry
PACS: 87.23.-n Ecology and evolution, 87.57.Nk (Medical imaging) Image analysis
1991 MSC: 92B05 General biology and biomathematics, 92D15 Problems related to evolution, 92C05 Biophysics, 92C15 Developmental biology, pattern formation, 92C55 Biomedical imaging and signal processing, 44A35 Convolution, 42C40 Wavelets, 65T60 Wavelets

Email address: cesar@vision.ime.usp.br (R. M. Cesar-Jr.).
URL: http://www.ime.usp.br/~cesar/ (R. M. Cesar-Jr.).

[1] Supported by CAPES.
[2] R. M. Cesar-Jr is grateful to FAPESP (99/12765-2) and to CNPq (300722/98-2).

1 Introduction

Many works related to the study of evolution have been presented to the scientific community having as main objective the relation between phenotype and pre-determined time-space or genetic conditions. This kind of analysis concerns morphometry, which can be described as the study of biological shapes. Traditional morphometrics generally apply multivariate statistical methods to size or shape variables such as distances and angles[1, 2]. On the other hand, geometric morphometrics regard geometrical relationships between these measurements [3]. Note that morphometry is a multidisciplinary area related to shape analysis [4, 5]. Nevertheless there are few works that explore multiscale and differential methods well-known in shape analysis and image processing community, e.g. wavelets [6, 7, 8], a gap that is partially filled by the method introduced in this paper.

This paper belongs to the research field of computer vision and, more specifically, to shape analysis. Many problems in a wide variety of disciplines may be addressed in terms of computer vision concepts and methods and it is possible to reduce many of these problems to shape analysis [4], where visual information like color, texture and motion can be discarded. In fact, shape analysis has played a central role in perception and signal understanding [9, 5].

The wavelet transform is a particularly useful tool to analyze signals that are composed by non-stationary, local events that vary quickly in time because of

their local analysis property [10, 11]. Furthermore, its scaling properties have been extensively used to implement multiscale tools for signal analysis, thus explaining its popularity in a several practical problems [12, 13].

The goal of the present work is to review and generalize the work presented in [5] where 2D landmark data sets were analyzed through the wavelet transform. This problem is addressed by formulating a signal processing solution to 2D and 3D morphometry. We analize two different sets of data: (a) morphological traits of 10 species of American canids expressed by 2D landmarks, taken from the dorsal view of the skull; (b) phenotype of specimens of 5 species in a monophyletic lineage of rodents of the genus *Trinomys*, wich are represented by 3D landmark, taken from the skull of the analyzed specimens.

In the first analysis, the intraspecific variability in biological species is the main interest and differences among males and females are considered. Here we examine patterns of sexual dimorphism among the sexes in a monophyletic clade formed by the south-american canids. In this family, broadly distributed throughout the world, differences in size and shape between sexes are usually related to mating systems and social organization. The south american members are among the less studied of the dogs, and little is known about possible differences among individuals of the same species.

As far as the 3D data is concerned, the interest is to determine whether morphological descriptors can be matched to phylogenetic trees or the ecological differentiation that may result from colonization of different environments.

The expected result, from the evolutionary theory viewpoint, is some inference about an evolutive trace [14] between studied species through a comparative study concerned to the relation of wavelet coefficients to shape of these species [15, 16].

This work is organized as follows. In order to allow the use of shape analysis's tools in morphometry, section 2 describes how it is possible to extract wavelets representation of shapes described by landmarks. In the section 3, we show a process by which it is possible to use the wavelet transform to analyze 2D and 3D signals based on a useful geometrical interpretation. Experimental results obtained using simulated and real data will be presented in section 4. This paper is concluded with a discussion about the methodology and the successful biological results of the application wavelet-based analysis over the real data sets.

2 From Landmarks to Signals

Usually, landmarks represent curvature extreme or junctions between structures, in key regions of the shape [17, 18]. In our case the landmarks are selected in images, in this sense relating the biological information to the geometry of the studied structures.

The morphometric concept of homology [1] (e.g. a landmark l in shape a correspond to the same morphological structure that the landmark l in another shape b) allows the landmarks from all shapes to be ordered in a consistent

manner. The sequence of ordered landmarks can be viewed as the vertices of a polygonal line, being subsequently interpolated [5]. Note that it allows the comparation between the shapes using only landmark-related values.

Therefore, a landmark set L is interpolated generating a set of points denoted as $u(t) = (x(t), y(t))$ and $\vec{u}(t) = (x(t), y(t), z(t))$ in the 2D and 3D cases, respectively, to which the procedure explained in the next section can be applied.

3 Wavelet Representation

The wavelet transform is a multiscale tool that leads to useful geometrical characterization of the signal $u(t)$ by allowing the detection and analysis of transient events. The wavelet transform is defined as:

$$U[\psi, u](b, a) = U_\psi(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^*(\frac{t-b}{a}) u(t) dt \tag{1}$$

$$U_\psi(b, a) = \sqrt{a} \int_{-\infty}^{\infty} \Psi^*(af) U(f) e^{j2\pi fb} df \tag{2}$$

where $U_\psi(b, a)$ is the wavelet transform of $u(t)$, $a > 0$ is the scale parameter and b is the shift parameter of the mother wavelet ψ.

In the Fig. 1 we see the application of the wavelet transform to a complex signal corresponding to a polygonal line of a square. In Fig. 1(c) we assign each of a-lines of the matrix of wavelet coefficients to a internal contour of the polygon. Note that the internal representation of the wavelet transform shows

some relation of this transform to the curvature. The higher the curvature, the higher the related wavelet coefficient.

Differentiation operators from signal analysis, have been widely applied to the characterization of shapes in applications such as classification of neural cells [4]. In the next sections we will show how wavelets can be explored in order to extract meaningful differential shape descriptors.

3.1 Wavelet representation of 2D curves

In the 2D case, we follow the approach proposed in [13], in which the input to the wavelet transform is a complex signal representing the shape. Thus, the signal $u(t)$ can be defined as:

$$u(t) = x(t) + jy(t) \tag{3}$$

As we see in [5], the convolution theorem [4, 19] allows applying the wavelet transform with n-th derivative-of-Gaussian mother wavelet as an approximation of n-th derivative of $u(t)$. In the next section we introduce the corresponding methodology for 3D data.

3.2 Wavelet Representation of 3D curves

The adopted methodology for extraction of wavelet coefficients from 3D data was to consider each dimension of a 3D signal $\vec{u}(t) = (x(t), y(t), z(t))$ as a 1D

real-valued signal and to apply the wavelet transform to each one. The 3D wavelet representation $\vec{U}(b,a)$ is formed from those coefficients, i.e.

$$\vec{U}(b,a) = (X(b,a), Y(b,a), Z(b,a)) \tag{4}$$

3.3 Interpretation of 3D Wavelet Representation with derivatives-of-Gaussian-kernel

Recalling the principles studied in [5], the wavelet transform with n^{th} derivative of Gaussian $g^{(n)}$ of $u(t)$ (denotaded as $U(b,a)$, with $\psi = g^{(n)}$) is an estimative of the n^{th} derivative of $u(t)$. Then, for $\vec{u}(t) = (x(t), y(t), z(t))$ it is possible to calculate $X(b,a)$, $Y(b,a)$ and $Z(b,a)$ in order to estimate the n^{th} derivatives of $x(t)$, $y(t)$ and $z(t)$ respectively. Thus, $X(b,a)$, $Y(b,a)$ and $Z(b,a)$ are related to the partial derivatives of $\vec{u}$ and the total derivative of $\vec{u}$ in t can be estimated by the vectorial sum of these partial derivatives.

$$\vec{u}(t)^{(n)} \approx \vec{X}(b,a) + \vec{Y}(b,a) + \vec{Z}(b,a) \tag{5}$$

This means that $|\vec{U}(b,a)|$ is an estimative of $|u^{(n)}|$. Therefore, $|\vec{U}(b,a)|$, with $\psi = g^{(1)}$ represents the length of the tangent vector of $\vec{u}$ in t and $|\vec{U}(b,a)|$, with $\psi = g^{(2)}$ is a measure of orientation variation, i.e. curvature, of $\vec{u}$. The features used to classify the data set are listed in Table 1. We have adopted the second derivative of Gaussian as analyzing wavelet in the results reported in this paper.

$X(b,a)$, in each landmark
$Y(b,a)$, in each landmark
$Z(b,a)$, in each landmark
$\|\vec{U}(b,a)\|$, in each landmark
Mean value of $\|\vec{U}(b,a)\|$ in each scale
Variance of $\|\vec{U}(b,a)\|$ in each scale
Mean value of $\|\vec{U}(b,a)\|$ of each landmark
Variance of $\|\vec{U}(b,a)\|$ of each landmark

Table 1. Feature space.

4 Experimental Results

4.1 Simulated Data

In order to assess the performance of the introduced framework, simulated data sets with 3 and 4 different classes were created for the 2D(Fig. 2 (a)) and 3D (Fig. 2 (b)) cases, respectively. The procedure to obtain this data set was to add noise to a prototype model of each class (see examples in Figure 2), generating 10 samples per class.

Figure 3 presents the vectors obtained by the transform at the original landmarks (i.e. representing the corresponding second derivatives at each landmark).

Because of the wavelet redundancy characteristic a feature selection algorithm has been developed and applied to the wavelets features in order to select the best features that categorize the classes in our application.

In order to speed the feature selection algorithm, only the features related with the landmarks positions are used. Therefore, for each landmark, it is possible to select the aforementioned corresponding wavelet features. Thus, starting from the above explained features (Table 1, it is possible to identify which feature vectors perform better classification. The applied procedure is to test all 2×2 combinations of features by selecting sets for training (randomly chosen 1/3 of the input set) and testing (remaining 2/3) a minimum distance to prototype classifier. Training and testing procedures are repeated for each feature set until convergence of the classifier error rate.

The 3D data set was classified and the resulting feature space using the best set of 2 features (i.e. variance of the wavelets coefficients at scale 13 and the variance at landmark 8) is showed in Figure 4.

In the next sections we present results of the clustering of real data related to biological aspects: (a) sexual dimorphism, an important source of intraspecific variability in biological species; (b) correlation between morphological or molecular descriptors and the evolutive process, one of the main questions in morphometric analysis.

4.2 Morphology and sexual dimorphism

The methodology previously presented was also applied to a real data set to analize intraspecifc variability in biological species by taking differences

Code	Class	Specimens	Male	Female
1	*Atelocynus microtis*	13	7	6
2	*Cerdocyon thous*	24	12	12
3	*Chrysocyon branchiurus*	13	5	8
4	*Dusicyon australis*	2	1	1
5	*Lycalopex vetulus*	18	10	8
6	*Pseudalopex culpaeus*	30	15	15
7	*Pseudalopex gymnocercus*	30	15	15
8	*Pseudalopex sechurae*	10	5	5
9	*Speothos venaticus*	22	13	9
10	*Nyctereutes procyonoides*	11	5	6
	Total:	173	88	85

Table 2. Specimens of canids studied

among males and females into consideration. Therefore, we took samples of all 10 species belonging to *Mamalia, Canidae* group (See Table 2). The goal was to study differences in size and shape among males and females, mapping the results in two phylogenetic hypothesis already proposed for the canids, one molecular and one constructed using morphological characters. Samples consisted of digital images of the dorsal surface of the skull of prepared specimens taken in various natural history museums, and including, when possible, the same number of adult males and females. We digitized 14 2D landmarks in each one of the 173 images (Fig. 5).

We have explored computational shape analysis to characterize dimorphism using multiscale tools through the application of the second derivative of Gaussian as wavelet kernel. The wavelet features are used as input to an automated feature selection procedure that uses the Mahalanobis distances between male

and female classes of all combinations of 2 features to select the set that best characterizes dimorphism. Thus it is possible to construct a rank of discrimination levels of the features then allowing a biological interpretation of the results.

Fig. 7 (b) and 8 (b) we have a quantitative set of results, using *Atelocynus* as reference and sex as independent variable (1 represents females specimens and 2 represents male). Note that we got better discrimination level for 3 features.

The resutls demonstrate separation between males and females in each class, but it was not possible to define a set of features that separates males and females in all classes at the same time.

4.3 Morphology and filogeny

In order to validate the proposal of use of 3D wavelet representation, we analyze the relation of morphological and filogenetics patterns of *Trinomys*. The studied specimens consist of 5 classes (*T.yonenagae*, *T.paratus*, *T.s.denigratus*, *T.s.elegans*, *T.eliasi*) of *Trinomys* (see Table 3), and each shape has 51 3D landmarks (Fig. 9).

These morphological landmarks were primarily defined on the basis of topological features of the cranium [20]. Each cranium was placed parallel to the focal plane under a Pixera (Pixera Corporation, Los Gatos, California) digital camera system and the x-, y- and z-coordinates of each landmark were

obtained using PhotoModeler Lite software [21].

Thus, using a total of 613 features, the feature space in Figure 10 was selected. In Figure 10 (b) it is possible to see the dendrogram of the distances between the mean values of each pair of classes.

Code	Class	Specimens
1	*T. eliasi*	5
2	*T. paratus*	6
3	*T. s. denigratus*	9
4	*T. s. elegans*	5
5	*T. yonenagae*	16

Table 3. Studied species.

The projection of specimens of the 5 species in the feature space defined by the the variance of the wavelet coefficients at scale 63 and the wavelet coefficient at landmark 1 and scale 49 shows the distinctiveness of *T. yonenagae* relative to the others species. This result demonstrates that morphological descriptors of shape modeled by wavelets map onto patterns of ecological differentiation since *T. yonenagae* diverged uniquely in ecology and morphology from all others species *Trinomys* [22].

Contrary to species in this genus, which are ground dwellers in forest habitats, *T. yonenagae* inhabits fossil sand dunes in a restricted area in the left bank of the São Francisco river in the state of Bahia in north-eastern Brazil. The differentiation of *T. yonenagae* in this particular habitat has produced

a strikingly distinct morphology involving a combination of traits such as a light-coloured pelage, a well developed tail brush, large hind feet and cranial modifications, including an inflated bulla [22]; all of which are usually associated with evolution in arid environments.

The uniqueness of ecological life-history and morphological traits of *T. yonenagae* are not paralleled by molecular differentiation as inferred from mitochondrial cytochrome *b* gene sequences, since *T. yonenagae* is well nested within a lineage which includes *T. elegans*, *T. paratus* and *T. eliasi*. Thus, there is a disparity between divergences at the organismal (ecological and morphological) and molecular (mitochondrial DNA sequences) levels.

It is evident that the wavelets representation mapped variation in cranial shape among the species of *Trinomys* onto the ecological component of differentiation that arise form adaptation to distinct environments.

5 Concluding Remarks

In this paper we show how it is possible to characterize biological shapes represented by 2D and 3D landmarks through multiscale and differential signal processing techniques in order to relate ecological and sexual variables with the morphology of a species. Results demonstrate that the wavelet transform can be used as a powerful descriptor of variation in shapes and can be successfully applied in morphometrics studies. We claim that this is a generic technique that may be successfully applied to other types of 3D data, which belongs to

our ongoing work besides further development of the current approach.

References

[1] F.L. Bookstein. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge, 1992.

[2] L. R. Monteiro and S. F. dos Reis. *Princípios de morfometria geométrica*. Holos, Ribeirão Preto, 1999.

[3] E.P. Martins. Adaptation and the comparative method. *Tree*, 15(7):296–299, 2000.

[4] L.F. Costa and R.M. Cesar Jr. *Shape analysis and classification: theory and practice*. CRC Press, Boca Raton, 2001.

[5] C. M. Takemura and R. M. Cesar-Jr. Shape analysis and classification using landmarks: Polygonal wavelet transform. In *Proc. 15th European Conference on Artificial Intelligence ECAI2002*, 2002.

[6] P. E. Lestrel, R. M. Cesar-Jr, O. Takahashi, and E. Kanazawa. A fourier-wavelet representation of 2-d shapes: Sexual dimorphism in the japanese cranial base. *Anthropological Science*, 2003. in press.

[7] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE SP Magazine*, pages 14–38, 1991.

[8] J. P. Antoine. Wavelet analysis: a new tool in signal processing. *Physicalia Magazine*, 16:17–42, 1994.

[9] M. Leyton. A process-grammar for shape. *Artificial Intelligence*, 34:213–247, 1988.

[10] A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J.-F. Muzy. *Ondelettes, Multifractales et Turbulences: de l'ADN aux Croissances Cristallines*. Diderot Editeur, Arts et Sciences, Paris, 1995 (in French).

[11] K.R. Castleman. *Digital image processing*, pages 303–347. Chapter 14: Wavelet Transforms. Prentice-Hall, Englewood Cliffs, 1996.

[12] A. Grossmann. Wavelet transforms and edge detection. In M. Hazewinkel S. Albeverio, Ph. Blanchard and L. Streit (Eds.), editors, *Stochastic processes in physics and engineering*, Dordrecht, 1988. Reidel Publishing Company.

[13] J. P. Antoine, D. Barache, R. M. CesarJr., and L. F. Costa. Shape characterization with the wavelet transform. *Signal Processing*, 62(3):265–290, 1997.

[14] P. H. Harvey and M. D. Pagel. *The comparative method in evolutionary biology*. Oxford University Press, Oxford, 1991.

[15] S. F. dos Reis, L. R. Monteiro, and E. L. A. Monteiro-Filho. Skull shape and size divergence in dolphins of the genus *sotalia*: a tridimensional morphometric analysis. *Journal of Mammalogy*, 81(1):125–134, 2002.

[16] S. F. dos Reis, L. R. Monteiro, J. A. F. Diniz-Filho, and E. D. Araújo. Geometric estimates of heritability in biological shape. *Evolution*, 56(3):563–572, 2002.

[17] F. L. Bookstein, B. Chernoff, J. Humphries R. Elder, G. Smith, and R. Strauss. *Morphometric in evolutionary biology*. Academy of Natural Sciences of Philadelphia, 1985.

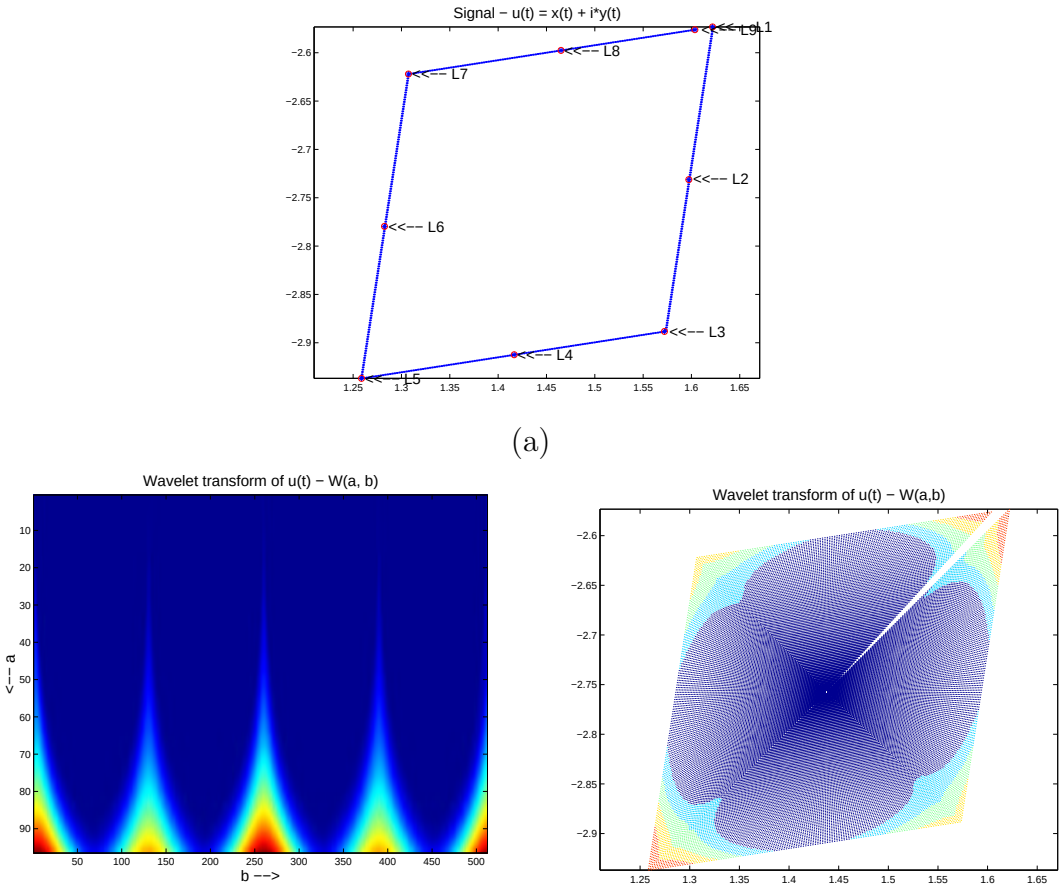
[18] I.L. Dryden and K.V. Mardia. *Statistical shape analysis*. John Wiley, Chichester, 1998.

[19] R. C. Gonzalez and R. E. Woods. *Processamento de imagens digitais*. Edgard Blücher LTDA., São Paulo, 2000.

[20] S. F. dos Reis, L. C. Duarte, L. R. Monteiro, and F. J. Von Zuben. Geografic variation in cranial morphology in *thricomys apereoides*(rodentia: Echimyidae). i.geometric descriptors and patterns of variation in shape. *Jornal of Mammalogy*, 83(2):333–344, 2002.

[21] S. Flach, J. Hargrave, S. Moore, G. Rempel, A. Valji, and N. Walford. Photomodeler lite(eos systems inc.), 1999.

[22] P. L. B. Rocha. *Proechimys yonenagae*, a new species of spiny rat (rodentia: Echimyidae). *Mamalia*, 59:537–549, 1995.

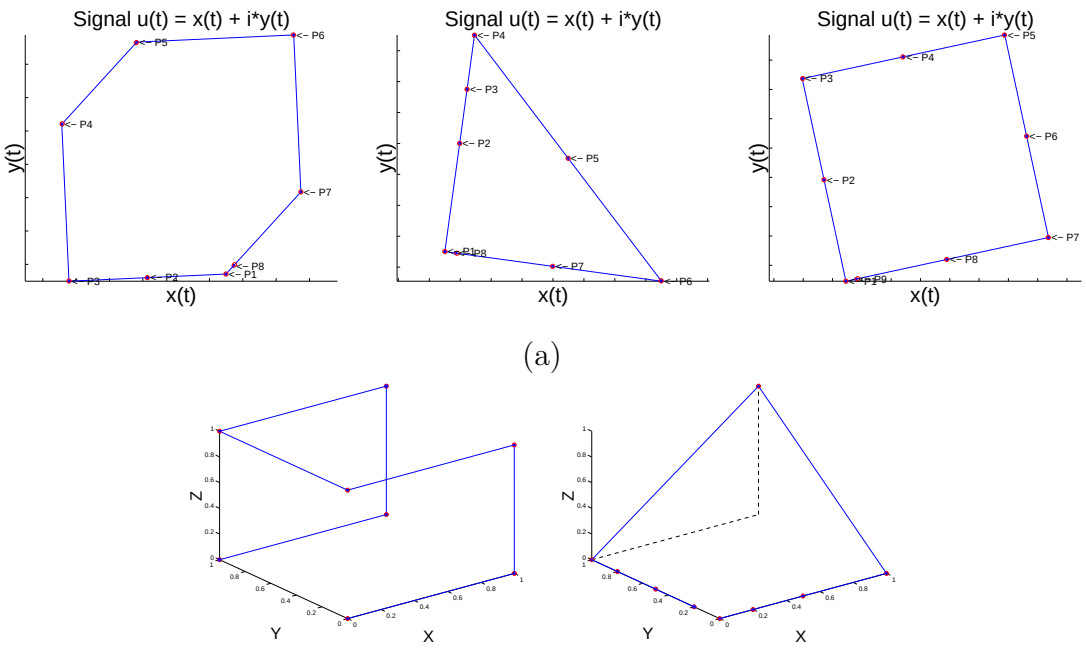


(a)

(b)

(c)

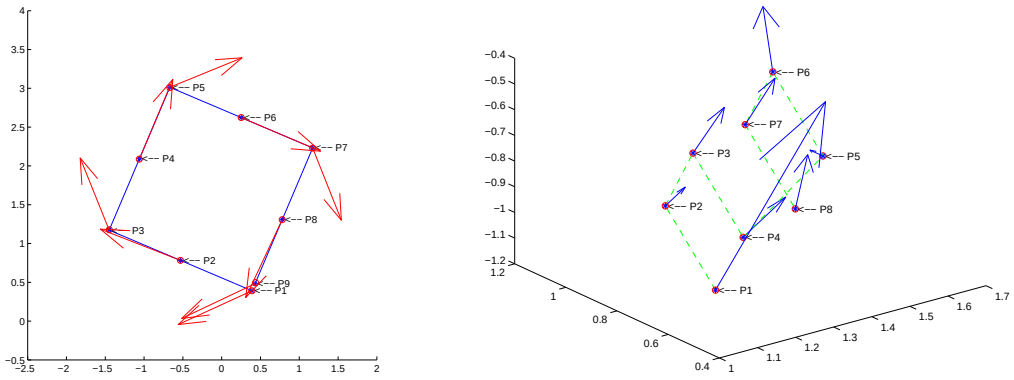
Fig. 1. (a) $u(t)$ complex signal; (b) Wavelet transform with 2nd derivative of gaussian as mother wavelet; (c) Internal representation of (b): relation of (b) and the curvature of $u(t)$



(a)

(b)

Fig. 2. (a)Models used to simulate the 2D data; (b)Two of the models used to simulate the 3D data.



(a) (b)

Fig. 3. (a) Estimative of the 2nd derivative of the signal formed by $x(t)$ and $y(t)$; (b) Estimative of the 2nd derivative of the signal formed by $x(t)$, $y(t)$ and $z(t)$.

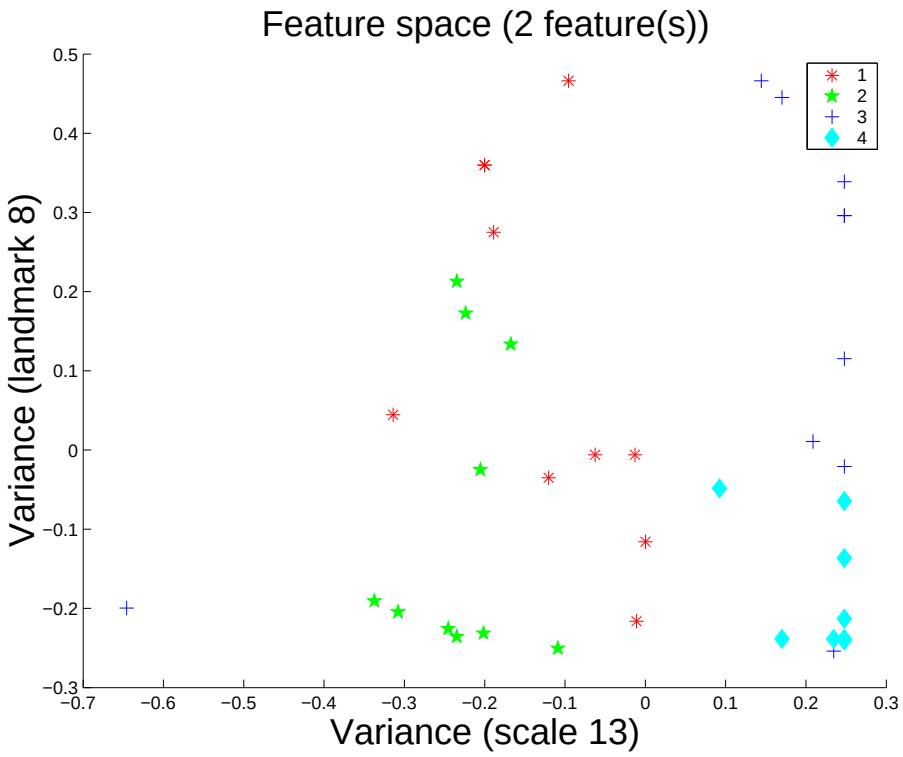
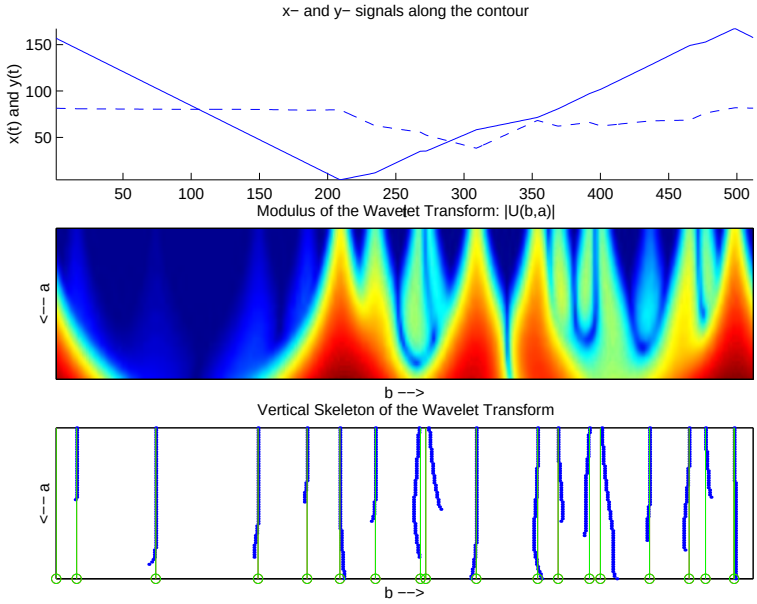


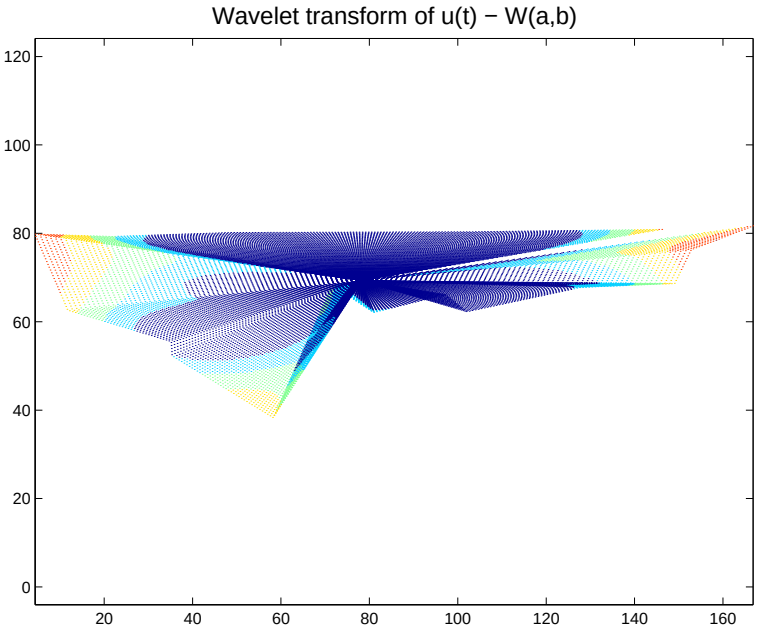
Fig. 4. Feature space of the 3D simulated data.

Fig. 5. Landmarks of canids

(a)

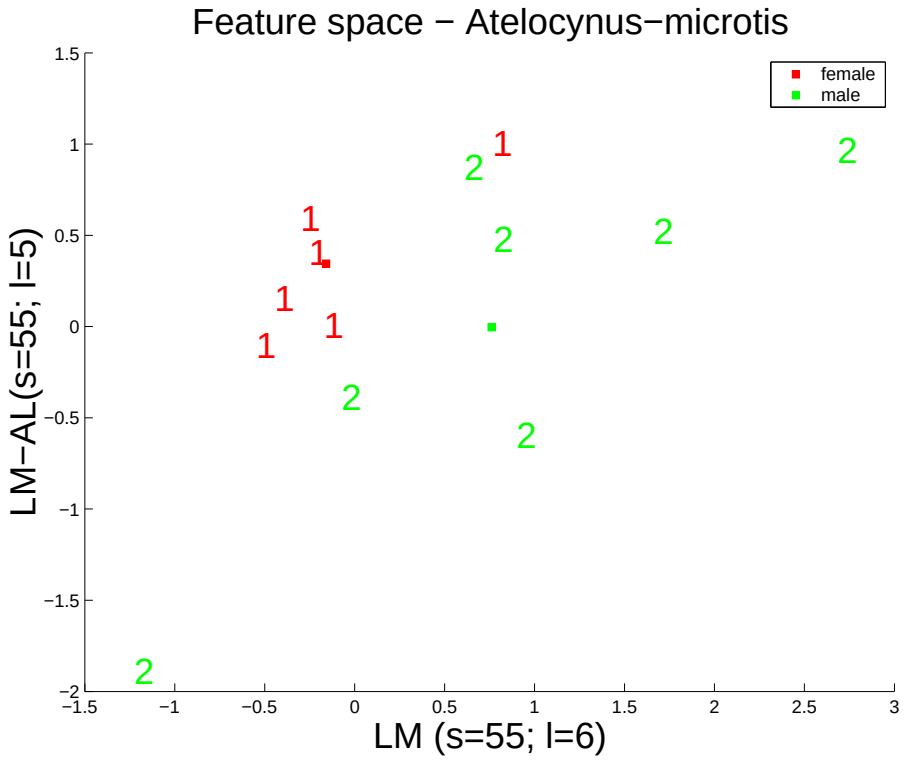


(b)

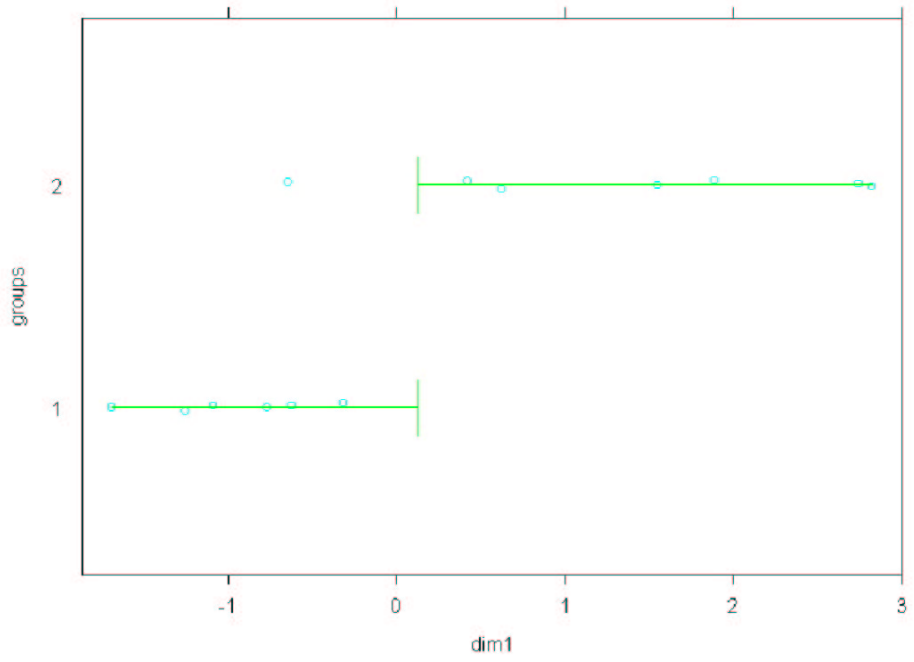


(c)

Fig. 6. (a) Dorsal view of one specimen of *Atelocynys microtis*; (b) Signal and wavelet based representations of (a); (c) Internal wavelet representation of (a).

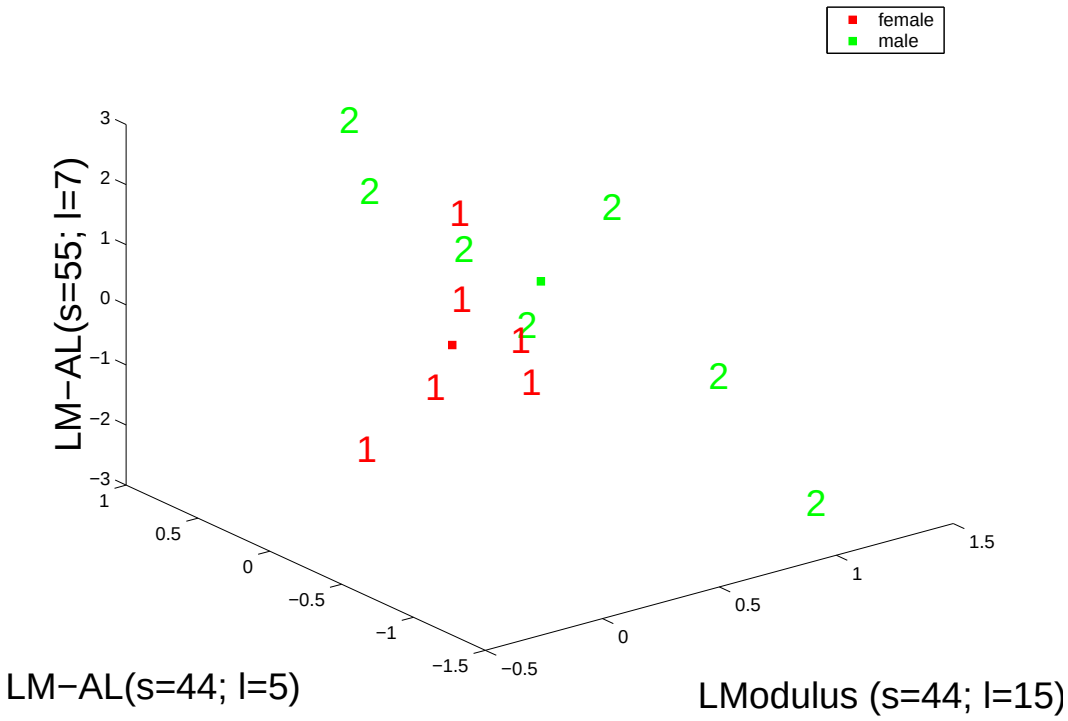


(a)

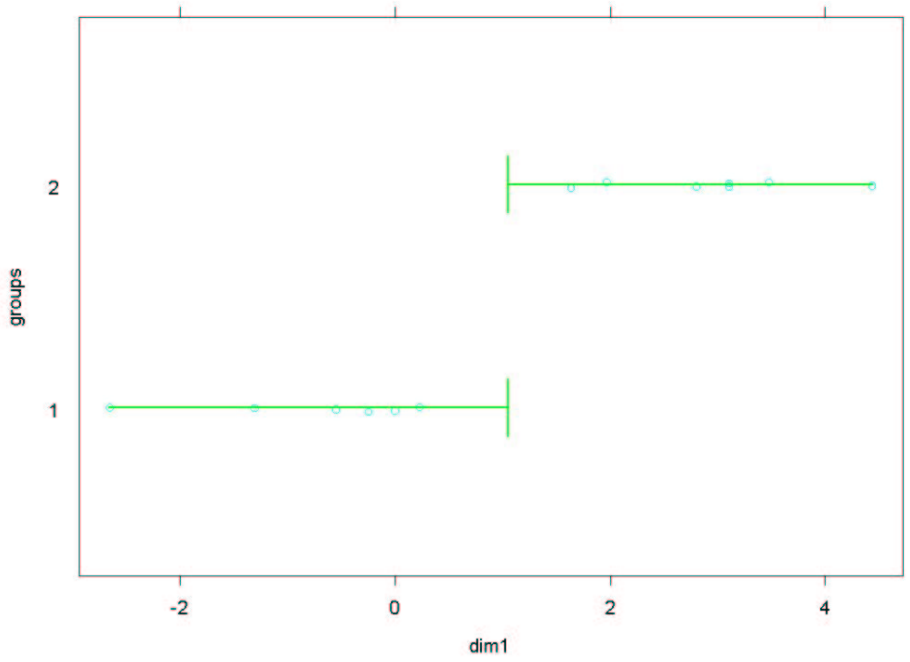


(b)

Fig. 7. (a) Feature space with 2 best discriminative features for Atelocynus microtis; (b) Discrimant analysis of (a).



(a)



(b)

Fig. 8. (a) Feature space with 3 best discriminative features for Atelocynus microtis; (b) Discriminat analysis of (a).

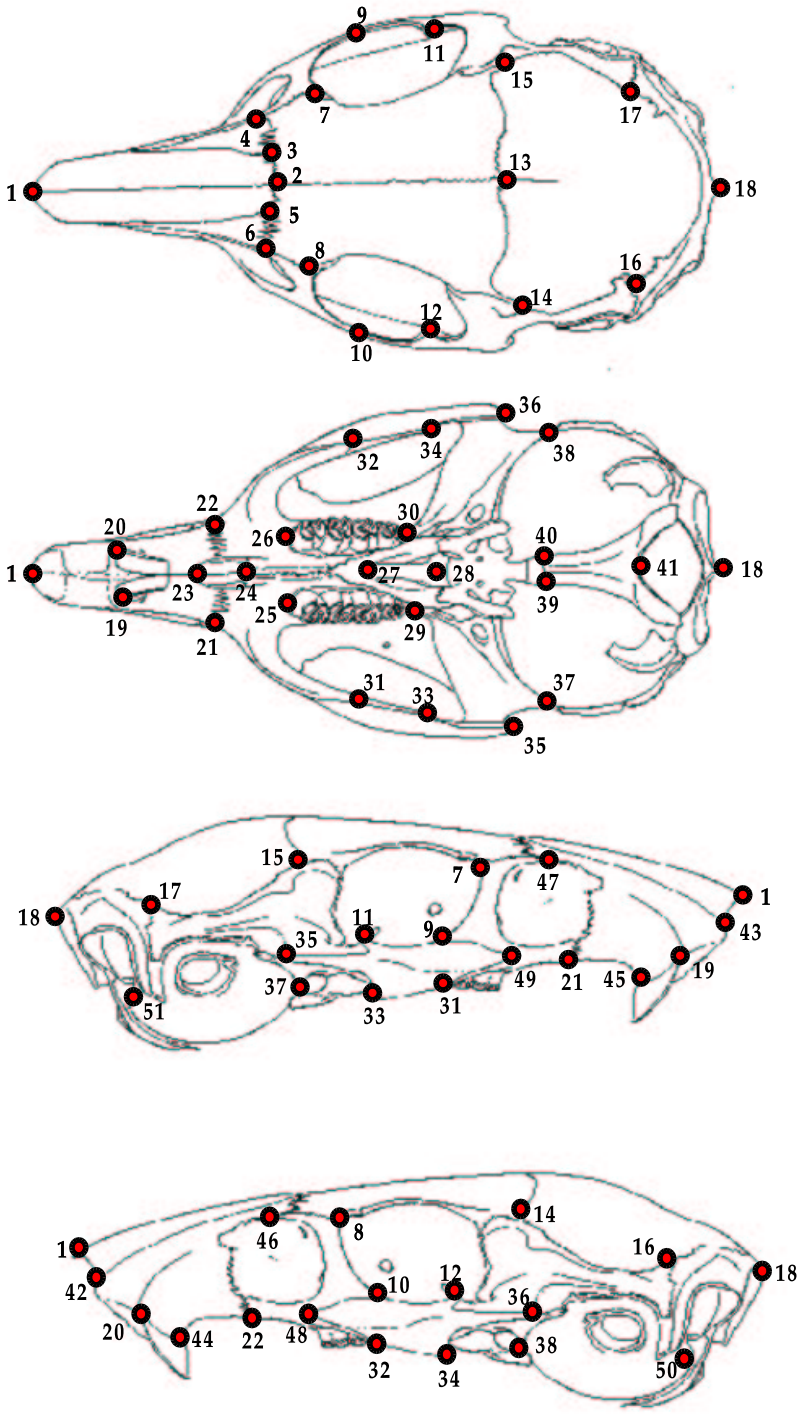
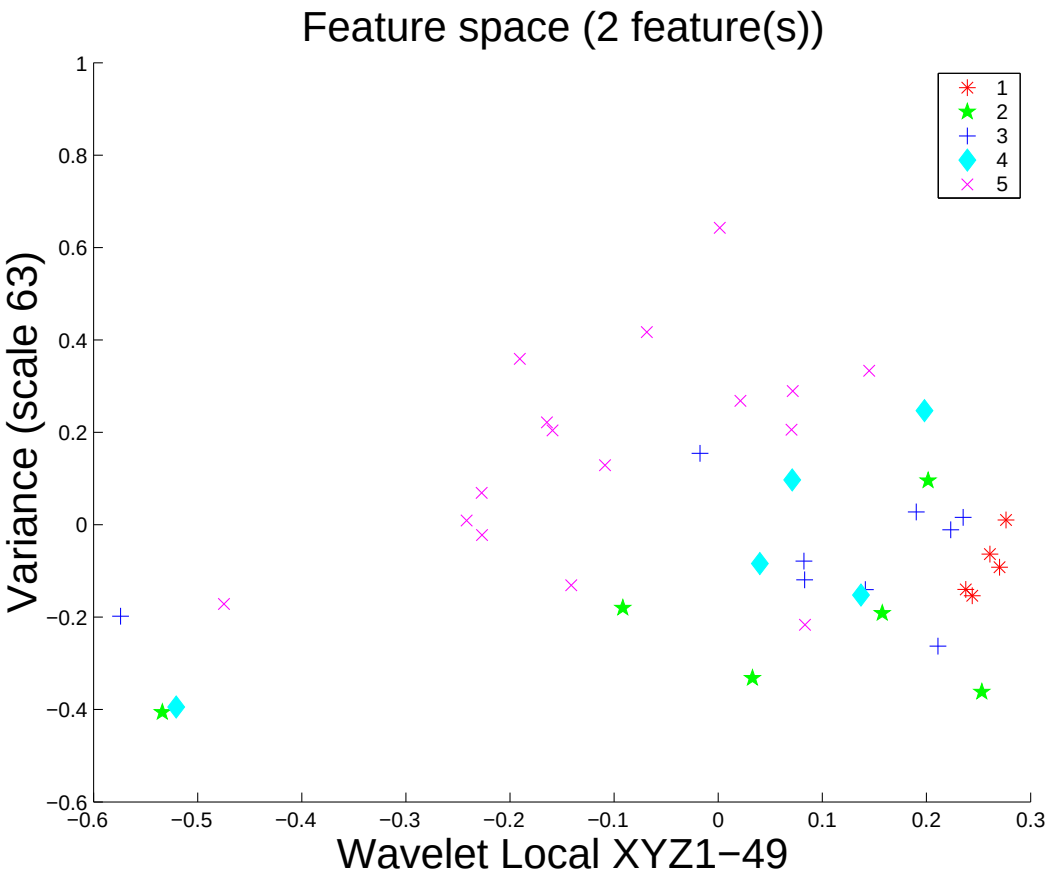
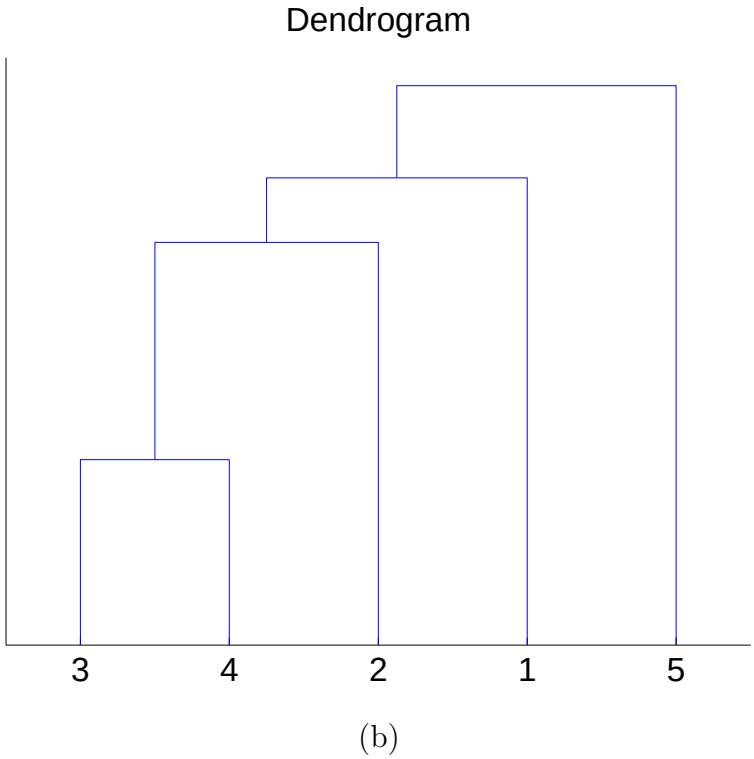


Fig. 9. Landmarks of *Trinomys*.



(a)



(b)

Fig. 10. (a) Feature space of the real data using the 2 features that best separate the species; (b) Dendrogram of the distances of the mean value of each class.