An explicit definition of earthworm ecological categories – Marcel Bouché's triangle revisited

Nicolas Bottinelli, Mickael Hedde, Pascal Jouquet, Yvan Capowiez

An explicit definition of earthworm ecological categories – Marcel Bouché's triangle revisited

N. Bottinelli [1,2], M. Hedde [3], P. Jouquet [1], Y. Capowiez [4]

[1] Sorbonne Université, UPEC, CNRS, IRD, INRAE, Institut d'écologie et des sciences de l'environnement, IESS, 93143 Bondy, France

[2] Soils and Fertilizers Research Institute (SFRI), Dong Ngac, Tu Liem, Hanoi, Vietnam

[3] INRAE, UMR 1222 Eco & Sols, 2 place Viala, 34070 Montpellier cedex, France

[4] INRAE, UMR 1114 EMMAH, INRAE/Université d'Avignon, Domaine Saint Paul, 84914 Avignon cedex 09, France

Corresponding author: Yvan Capowiez

yvan.capowiez@inra.fr

Abstract

In the early 70s, the soil biologist Marcel Bouché classified French earthworms species into defined ecological categories. This classification system was immensely successful and is still widely used to describe earthworm functional groups even outside of Europe. Bouché used morpho-anatomical traits to differentiate three main categories: epigeic, anecic and endogeic. However, the way species are assigned to a category was not explicitly described in Bouché's work. Thus, some earthworm species can still be assigned to two categories depending on the way researchers interpret Bouché's description. To solve these issues and avoid unnecessary controversies, we applied PCA and random forest models to the seminal data of Marcel Bouché (earthworm morpho-anatomical traits). Their assignment to Bouché's three main categories allowed us to statistically redefine the different categories and determine which traits are the most influential. We found that the three main traits were skin pigmentation (from none to black), body length (mean of the minimal and maximal values) and skin coloration (yes or no), followed by 10 other morphological and anatomical traits. We then used this approach to assign a likely category to all of the species studied by Bouché, resulting in a new triangular graph including other categories such as epi-anecic, endo-anecic, epi-endogeic and intermediate. Finally, we calculated the percentage that each species belongs to each main ecological category. This represents a paradigm shift and may change our vision of earthworm communities enabling the computation of the percentage of anecic, endogeic and epigeic species at the community level and thus overcoming the limits and debate about fixed ecological categories for each species.

Keywords : classification ; ecological traits ; functional group ; endogeic ; epigeic ; anecic ; epi-anecic.

1. Introduction

Marcel Bouché defined a widely used earthworm classification based on three ecological categories (epigeic, anecic and endogeic) at the beginning of the 1970s. The concept was first introduced in a chapter of an edited book (Bouché, 1971), further explained in detail in his book (Bouché, 1972) and finally summarized again in a conference paper a few years later (Bouché, 1977). This last paper was particularly successful since it described, for the first time, a triangle defined by three poles corresponding to three main ecological categories and positioned some typical earthworm species within this triangle. Despite the fact that these reports were all published in French, they had incredible success and the categories were widely adopted, at least for Lumbricidae species. As a consequence, the two main references (Bouché, 1972; Bouché, 1977) have been highly cited and remain consistently cited today: in the last 50 years, about 10% of the papers on earthworms have cited at least one of these two references (Fig. S1).

Marcel Bouché was especially influenced, among others, by the seminal work of Wilcke (1953) and suggested that three main criteria can be used to classify earthworms, their: (i) vertical distribution, (ii) ecology and (iii) morphology and with obvious, although not quantified, relationships between these criteria. However, because vertical distribution can vary depending on the environment, climate and thus the season, Marcel Bouché only developed his classification using the two last criteria and first defined ten morpho-physiological (1972) and later fifteen morpho-anatomical (1977) characteristics (Table S1). The second criteria (i.e. the ecology) remained, however, unquantified and relied on Bouché's personal observations or experience (1972). He also clearly stated that some earthworm species could be classified between the poles of the triangular graph. Thus four additional classes were defined and termed ‘epi-endogeic’, ‘epi-anecic’, ‘endo-anecic’ and ‘intermediate’ (the latter class means that the species is positioned between the three main classes). Examples of such species are *Microscolex dubius* (Fletcher, 1887) for epi-endogeic, *Lumbricus terrestris* (Linnaeus, 1758) for epi-anecic, *Aporrectodea caliginosa meridionalis* (Bouché, 1972) for endo-anecic and *Allolobophora chlorotica* (Savigny, 1826) as intermediate.

So, why revisit this concept? Firstly, because some species studied by Marcel Bouché in his book were not assigned to an ecological category (something that Marcel Bouché did not comment upon). Secondly, because despite the impressive success of this classification and the countless services this classification has rendered to scientists trying to understanding soil ecology, some recurrent controversies have plagued earthworm literature, e.g., the status of *Aporrectodea trapezoides* Duges, 1828 (Baldivieso-Freitas et al. 2018) or *Allolobophora chlorotica* (Le Couteulx et al. 2015). These inconsistencies are difficult to prevent since no general rules were given regarding the use of the 10 to 15 characteristics of Bouché (1972, 1977) and their relative importance. For species that were classified as occupying an intermediate position between two or three poles, it is even more difficult since, in this case, Bouché did not provide a formal definition. Futhermore, although the triangle (1977) is well known to soil ecologists, the method used to generate it is now ignored by most of us (how many morpho-ecological characteristics were used?; were they all useful?) and has even been forgotten by

Marcel Bouché himself (Bouché, personal communication).

Hence, to revisit the ecological categories, we first carried out a literature review looking for the most studied earthworm species worldwide to determine whether their ecological category was unambiguously defined and largely accepted in scientific publications. Then we used both the traits described in Bouché's book (1972) and the ecological category assigned to the species, to study the most influential characteristics and the minimal set of characteristics that could be used to assign an earthworm species to an ecological category. Finally, all the species mentioned by Bouché were projected onto a definitive triangular graph. This also means that, for each species, our approach goes beyond a fixed ecological category (this use was not encouraged by Bouché) and instead computes the percentages by which they belong to the three main ecological categories. Our hypothesis was that this numerical approach, based on morpho-anatomical data, would provide a more accurate classification of earthworm ecological categories.

2. Material and Methods

2.1. Bibliography review

To define the most currently studied earthworm species at the global scale, we benefited from the recent 11th International Symposium of Earthworm Ecology, held in Shanghai in the summer of 2018. Using the book of Abstracts (http://www.isee11.org/Home/Menu/30), we recorded every earthworm species cited in oral and poster presentations. We only retained species that were cited at least twice and obtained an initial list of 36 species. Since our main focus was soil ecology, *Eisenia fetida* and *E. andrei* were removed from the list, since these species are mainly found in composts and mainly studied in ecotoxicology and vermicompost production. For each species in the initial list, we then searched the Web of Science database (http://apps.webofknowledge.com) publications using the earthworm genus and species names as keywords. We applied the following rules to select the manuscripts: 5 years old (2013-2018) and if less than 100 manuscripts were found for that period, we considered older manuscripts until 100 manuscripts were selected (when possible). We then retained species for which at least 10 manuscripts could be found, obtaining a final list of 26 species. For each selected publication, when whole access was possible, we searched 'endogeic', 'epigeic' and 'anecic' as keywords, so that the three intermediate categories were also detected. Thus we determined whether the species under consideration was assigned to an ecological category (only the ecological category 'intermediate' was not detected). When available, we also noted the explanations given by the authors to explain this assignment.

2.2. Definition of ecological categories based on earthworm morpho-anatomical traits

Forty morpho-anatomical traits measured on 125 species were recorded in Bouché's book (1972) even if Bouché used only 10 and then later 15 of them to define the ecological categories. This dataset is available from the Dryad Digital Repository (http://doi.org/10.5061/dryad.g7046). Not all the species defined by Bouché are currently accepted by the scientific community (see for example http://drilobase.org/), but we decided to use all the data available in the book in order to replicate how Bouché defined his ecological categories. As a first important rule, sexual traits were not taken into account in our study because Bouché (1972) never mentioned them in the definition of the ecological categories and we feared that this would have given too much weight to phylogeny. Our dataset (Table S2) included 125 species (18 epigeic, 25 anecic, 20 endogeic, 6 epi-anecic, 10 epi-endogeic, 2 endo-anecic, 5 intermediate and 39 unspecified) characterized by 23 traits (11 qualitative and 12 quantitative) potentially relevant for distinguishing ecological categories. To build a model capable of assigning an earthworm species to an ecological category, we adopted the following procedure: first, we selected the morpho-anatomical traits which were the most important drivers of the differences among the three ecological categories (epigeic, anecic and endogeic) considering that these are the three main poles from which the other four categories were derived. To do this, we carried out a random forest classification to identify the most important predictors characterizing the three groups among the 23 traits. The importance of each predictor was determined by assessing the increase in the mean square error (MSE) between observations and out-of-bag predictions. The significance of the importance of each predictor on the three groups was tested with a permutation method involving 500 replicates of each tree, using the 'rfPermute' package in R. The selected traits were then used to build a PCA model using the 'PCAmixdata' package in R. From this PCA model, we added other species, which were not assigned to a category but described by Bouché (1972) as supplementary individuals (6 epi-anecic, 10 epi-endogeic, 2 endo-anecic, 5 intermediate and 39 unspecified). The triangle was defined by the three barycenters of each main ecological category in the two first axes of the PCA (the barycenters were computed only using species assigned to each category by Bouché 1972). Once the three poles of the triangle were defined, we computed the coordinates of each species as the distance to each pole in the two first axis of the PCA. These distances were first used to determine whether a species was within or outside of the triangle. If within, its coordinates were simply given by the distances to the three poles. If outside, then the species was projected onto the closest pole or side of the triangle and its coordinates determined by recomputing the distances to the three poles after the projection. The triangle was further divided into seven sub-regions to arbitrarily define the seven ecological categories. We set the limits of the main categories to 75% to get true archetypes and then cut the central space into four regions by setting the limits to 60%.

3. Results

3.1. Assignments to an ecological category found in the literature review

Among the 36 species first selected, we found more than 10 available publications for only 26 of them. More than 50% of these species were assigned to only one ecological category (Table 1). Seven species were assigned to two ecological categories. Among these, *A. trapezoides* and *Metaphire guillelmi* were assigned to two of the main categories (anecic-endogeic-epigeic). The remaining five were assigned to an intermediate category and this was the case for the well-known *L. terrestris* which is mainly cited as an anecic but sometimes classified as an epi-anecic (Table 1). Next, two species were classified in three categories: *A. longa* was mainly cited as anecic (84%) but also as endogeic (6%) and endo-anecic (10%) whereas *L. rubellus* was mainly classified as epigeic (72%) but also epi-anecic (2%) and epi-endogeic (25%). Lastly, two species were classified in four categories: *Amynthas corticis* was stated either endogeic (13%), epigeic (37%), epi-anecic (13%) or epi-endogeic (37%), whereas *Amynthas gracilis* was considered as anecic (14%), endogenic (14%), epigeic (43%) or epi-endogeic (29%). In general, in the above-cited studies the authors did not provide data or other evidence to support the assignment of the species to an ecological category. Indeed, the percentage of articles with such supporting information varied from 0 to 43% depending on the species under consideration (Table 1). It is worth noting that authors usually referred to earthworm behavior such as feeding (37.2% of explanations), followed by their vertical distribution in the soil profile (25.5%) and gallery organization (28.3%). In contrast, data related to earthworm morphology and anatomy were rarely used since earthworm pigmentation and size represented only 2.8 and 6.2% of the data, respectively (data not shown).

3.2. Characteristics and minimal dataset for the definition of ecological categories

The random forest classification model performed very well with regards to small out-of-bag errors (4.8%). Among the 23 traits used, only 13 were considered important for differentiating the three main ecological categories (Table S4). The most important traits included skin coloration (qualitative trait), body length (quantitative trait) and skin pigmentation (qualitative trait) (Table S3). This emphasized the importance of skin attributes in the definition of ecological categories. The rest of the traits which discriminated the categories were, in order of importance: body weight, the antero-posterior gradient, the dorso-ventral gradient, diameter, epithelium rigidity, muscle structure, typhlosole type, flattening index, transversal and longitudinal furrows (definitions can be found in Table S4). Other evaluated traits had a negligible effect on the differences found among categories. The PCA model generated with the 13 important traits and 63 species (those assigned to one of the three main categories by Bouché 1972) showed a good separation of the epigeic, anecic and endogeic species (Fig. 1) although intra ecological group variability was rather large. The first axis (22% of total variance explained) clearly separated the three ecological categories whereas the second axis (13% of total variance explained) opposed epigeic species to anecic and endogeic species. The projection of supplementary species (those unassigned or those assigned to ecological categories other than the three main ones) on the two axes of the PCA model showed that the 6 epi-anecic species were located between epigeic and anecic species, whereas the 10 epi-endogeic species overlapped with endogeic species. The five intermediate and the two endo-anecic species had variable positions in the PCA. Several species which Bouché (1972) did not previously assign, preferentially overlapped with the endogeic species.

3.3. Definition of ecological categories

The scores for each species obtained from the PCA were projected on a ternary plot (Fig. 2) and some well-studied species and species mentioned on Bouché's triangle (1977) are highlighted. We observed a relatively good agreement between Bouché's triangle (1977) and our triangle. For each species, the percentage belonging to the three main ecological categories and their assignment to one of the seven ecological categories are presented in Table S4. Our model indicated that among the 125 species, endogeic species dominated (31%), followed by epigeic (17%), epi-anecic (18%) and anecic (14%). The other categories were scantly represented with less than 10% for each (Table 2). The classification accuracy between our model and Bouché's groups (1972) showed high values for epi-anecic (100%), epigeic (82%) and endogeic (70%) medium values for intermediate (60%) and anecic (52%) and very low values for epi-endogeic (25%) and endo-anecic (0%). The 13 traits characterizing the four most abundant ecological categories used in this study (epigeic, anecic, endogeic and epi-anecic) are presented in Table 3. From this analysis we determined a set of characteristics for each category. Epigeic species are small sized and light in weight. They are all pigmented, usually red with a gradient along the body. The caudal part of the body is sometimes flattened and the epithelium is soft. The muscle structure is pinnate. They do not have a longitudinal furrow and few have transversal furrows. The shape of the typhlosole is mostly massive. Anecic species are larger and heavier than epigeics. They are all pigmented, with a brown color and a gradient along the body. The clitellum and the caudal part of the body can be flattened and the epithelium is rigid. The muscle structure is mostly pinnate. They do have longitudinal and transversal furrows. The shape of the typhlosole is only pinnate. Endogeic species are of medium size and weight. They are all non-pigmented, thus their coloration (visible through their skin) is mostly pink with no gradient along the body. The clitellum can be flattened and the epithelium is soft. The muscle structure is mostly elementary and radial. They all have longitudinal furrows but few have transversal furrows. The shape of the typhlosole is mostly bilamelated. Epi-anecics share similar traits with anecics except the absence of longitudinal furrows, which is a trait shared with the epigeic species.

4. **Discussion**

4.1. Accurate and unambiguous definitions of the earthworm ecological categories

Marcel Bouché's ecological categories have been outstandingly successful and there is no doubt that they have been instrumental in unifying studies on earthworm ecology. However, our literature review showed that, even for some very common Lumbricidae such as *A. chlorotica*, *L. terrestris*, and *A. nocturna*, it can be challenging to clearly assign ecological categories It should be noted however that confusion regarding the classifications were generally between a main category (anecic, endogeic, epigeic) and one of the sub-categories. This reflects the fact that for some authors (Hoeffner et al. 2018), these ecological categories are indeed hierarchical: the earthworm species is first anecic and then further described as epi-anecic, for example (as often observed for *L. terrestris;* van Groenigen et al. 2019). This is in contrast with Bouché's opinion that these categories were rather exclusive as showed by his definition of the 'intermediate' ecological category (in between anecic, endogeic and epigeic categories).

There are also difficulties with earthworm species from Asia, with the striking examples of two species (*A. gracilis* and *A. corticis*) assigned to four different ecological categories. We assume that the lack of clear rules impeded the authors from assigning these species to one ecological category. Moreover, the ecological categories of traits defined mainly on Lumbricidae in Europe may not perhaps be fully applicable to earthworms from other families and continents. This makes it clear to us that the time has come to revisit these ecological categories, not to destroy them but to rebuild their foundations in order to apply the approach to either (i) difficult cases or (ii) other continents (if applicable). This is also supported by the availability of modern statistical methods that were out of reach at the beginning of the 1970s (random forest approaches).

Our method successfully reconstructed and accurately defined the three main ecological categories and four additional categories. Every new species for which the 13 traits are measured or estimated can thus be positioned in the triangle and assigned an ecological category (assuming that this classification made mainly on Lumbricidae can be extended to other families). Our approach could also be used to confirm the use of these categories for non-lumbricidae earthworms from other parts of the world such as Asia. The hierarchization of the factors (given by the random forest classification) to define the ecological categories was not unexpected. However, the three main characteristics are very basic traits (skin coloration, body length and skin pigmentation) and can be easily assessed on adult earthworms. The remaining nine characteristics do include some internal anatomy (typhlosole, muscle structure) requiring dissection or the use of categorical characteristics (flattening index and epithelium rigidity) requiring expertise or the establishment of clear descriptions and boundaries. Regarding the last criteria (epithelium rigidity), it is worth mentioning that Briones and Alvarez-Oterao (2018) proposed earthworm body wall thickness as a method for inferring membership to an ecological category.

Despite the general agreement between our classification and Bouché's seminal work there were some discrepancies. The most striking differences concern the anecic vs epi-anecic categories. In brief, in Bouché (1972), anecics combined the brown-headed large earthworms with the giant Mediterranean earthworms (mainly of the *Scherotheca* genus), and the epi-anecic earthworms are mainly red-headed large earthworms of the *Lumbricus* genus. In contrast, in our study anecics only included most of the giant Mediterranean earthworms whereas the red and brown-headed large earthworms from the *Lumbricus* and *Nicodrilus* genera were epi-anecic (with a few exceptions). The main difference between these two categories in our classification was mean size and the existence of transversal furrows (Table 3). This may be seen as non-optimal or illogical for some researchers, especially those that define epi-anecics using behavioral traits (Hoeffner et al. 2018) something Marcel Bouché never did. This suggests that, using multivariate analysis, the more extreme anecic earthworms (giant Mediterranean worms) represent the norm for anecic earthworms. Although this observation may hold true for statistical reasons only, in section 4.2, we explain how this may not be a problem. Besides these discrepancies, we also observed clear agreements with Bouché's work. The most impressive is the position of *A. chlorotica* which we found in the middle of the triangle and was thus assigned to the 'intermediate' category (in the middle of the three main categories), exactly how Bouché predicted in his book (1972, p. 269): « *A. chlorotica* is morphologically assigned to an intermediate position between the three categories, even if the green morph has manners more epigeic than the leucotypic morph ». This is striking since no papers were found that cited or acknowledged this strange position for this current species. In our literature review, *A. chlorotica* was always assigned to the endogeic category with the exception of only one report (Le Couteulx et al. 2015) where it was assigned to the epi-endogeic category. In our opinion, this agreement reinforces our classification and the seminal work and intuition of Marcel Bouché who never explained in his book why this species was given an intermediate position: it was so obvious to him that no further explanations was needed. However, our study makes these statements fully explicit and also fully available to non-French speakers.

4.2. From fixed categories to the use of fuzzy logic

The use of sole categories to classify earthworm species was originally criticized by Bouché (1977) himself who advocated that anecic, epigeic and endogeic are the three main poles of a continuous distribution of earthworm species between these three main ecological strategies. However, despite its success in terms of scientific citations, Bouché's seminal work has been frequently misapplied and most, if not all, the researchers involved in earthworm ecology, including the authors of the present manuscript, have used the ecological categories as categorical data. It is time for a rediscovery of what Bouché had already stated in the 1970s: earthworm species representing a typical earthworm category are rare. In our study, we found that only 16 species out of the 125 belonged 100% to only one ecological category. These were: two *Dendrobaena* and one *Dendrodrilus* as epigeic ; most of the *Prosellodrilus*, *Hemigatsrodrilus monicae*, *Ailoscolex lacteospumosus* and *Allolobophora graffi* and *A. zicsii* as endogeic *;* and *Nicodrilus gognus* as anecic. It is once again striking that only one species was 100% anecic, however most of the *Scherotheca* species were more than 90%.

Slicing of the triangle based on arbitrary thresholds may have induced bias and could be a problem for assigning species close to these borders. However, following the idea of Marcel Bouché, we believe it is time for a paradigm shift regarding these categories and we should now accept that each earthworm species is indeed defined by three percentages, one for each main ecological category (see Table S4). This is not a drastic increase in complexity of the definition of these categories and could help solve recurrent controversies regarding the assignment of an earthworm species to a unique and fixed ecological category (Bastardie et al. 2005, Felten and Emerling 2009, Baldivieso-Freitas et al. 2018), especially the species not assigned by Marcel Bouché. In this study, we have defined, using automatic borders inside the triangle, the four other categories that are used but which were not accurately defined by Bouché (1972). We also provide some information regarding their main characteristics (excluding the two less common categories) to provide support for further discussion on this topic. We do not believe that our proposed separation is definitive. It is worth noting that all the misclassified species using our methodology were part of the mixed categories (epi-anecic, endo-anecic, epi-endogeic and intermediate; Table 2). In contrast, we advocate for the wider use of percentage of belonging to the three main (well accepted and defined) categories.

The perspectives of our work go beyond dispute resolution, it opens the avenue for new kinds of approaches in earthworm ecology. Traditionally, earthworm data are first analyzed using total abundance and biomass and then using ecological categories and their putative reaction to different stresses such as tillage, pesticide application (and thus exposure) or soil compaction. We believe our approach can be used almost unambiguously to compute the percentage of the anecic, endogeic and epigeic categories in whole earthworm communities based on fuzzy logic estimates of their membership to the ecological categories (i.e. the three percentages). This approach is analogous to the Community Weighted Means widely applied in trait-based ecology (Hedde et al., 2012).

We also believe that this is a first important step in the study of relationships between earthworm communities and important soil functions associated with their activities (Frazao et al., 2019; van Groeningen et al., 2019). However, this goal will only be achieved if we develop new traits that can be defined for earthworm species, particularly traits linked to the ecology and behavior of species and not only those based on morpho-anatomical characteristics.

Conclusions and prospects for the future

We re-analyzed the data of Bouché (1972) to redefine the three main ecological categories for earthworms. By doing this, we were also able to compute the percentage by which each species belongs to each category based on data of 13 morpho-anatomical traits. We assume this will change the way soil ecologists will analyze earthworm communities in the future.

One of our aims was also to propose a shared procedure for computing this percentage for species worldwide. It is not possible to provide a universal spreadsheet for doing this since in the PCA we used mixed categorical and continuous data, which does not give access to the computations of the projections on the two main axis. Instead we now propose that if research groups send us their data (13 traits) collected for a species not already described in Bouché's book (1972), we will analyze them and provide the percentages of belonging to the three main categories. We will also share the results with the wider scientist community (upon agreement with the authors) on websites such as 'BETSI' or 'Drilobase'. Since the traits could be measured or estimated in different ways by different researchers, our percentages will be computed on several datasets (for example *Pontoscolex coeruthrus* (Muller, 1856) from different parts of the world) enabling the computation of robust mean values.

Acknowledgments

We thank colleagues from the TEBIS network (http://www.reseau-tebis.fr/) for fruitful discussions about ecological traits and their use.

References

Baldivieso-Freitas, P., Blanco-Moreno, J.M., Gutierrez-Lopez, M., Peigné, J., Perez-Ferrer, A., Trigo-Aza, D., Sans, F.X., 2018. Earthworm abundance response to conservation agriculture practices in organic arable farming under

Mediterranean climate. Pedobiologia 66, 58-64.

Bastardie, F., Capowiez, Y., Renault, P., Cluzeau D., 2005. A radio-labelled study of earthworm behaviour in artificial soil cores in term of ecological types. Biology and Fertility of Soils 41, 320-327.

Bouché, M.B., 1971. Relation entre les structures spatiales et fonctionnelles des écosystèmes illustrés par le rôle pédobiologique des vers de terre. In: Pesson P. (Ed.) La vie des sols. Paris, Gauthier-Villars, 187-209.

Bouché, M.B., 1972. Lombriciens de France. Ecologie et Systématique. Paris, INRA.

Bouché, M.B., 1977. Stratégies lombriciennes. In: Lohm, U., Persson, T., (Eds) Soil organisms as components of ecosystems. Stockholm, Ecology Bulletin 25, 122-132.

Briones, M.J., Alvarez-Otero, R., 2018. Body wall thickness as a potential functional trait for assigning earthworm species to ecological categories. Pedobiologia 67, 26-34.

Felten, D., Emmerling, C., 2009. Earthworm burrowing behaviour in 2D terraria with single- and multi-species assemblages. Biology and Fertility of Soils 45, 789-797.

Frazao, J., de Goede, R.G.M., Capowiez, Y., Pulleman, M.M., 2019. Soil structure formation and organic matter distribution as affected by earthworm species interaction and crop residues placement. Geoderma 338, 453-463.

Hedde, M., van Oort, F., Lamy, I., 2012. Functional traits of soil invertebrates as indicators for exposure to soil disturbance. Environmental Pollution 164, 59-65.

Hoeffner, K., Monard, C., Santonja, M., Cluzeau, D., 2018. Feeding behaviour of epi-anecic earthworm species and their impacts on soil microbial communities. Soil Biology and Biochemistry 125, 1-9.

Le Couteulx, A., Wolf, C., Hallaire, V., Péres, G., 2015. Burrowing and casting activities of three endogeic earthworm species affected by organic matter location. Pedobiologia 58, 97-103.

van Groeningen, J.W., van Groeningen, K.J., Koopmans, G.F., Stokkermans, L., Vos, H.M.J., Lubbers, I.M., 2019. How fertile are earthworm casts? A meta-analysis. Geoderma 338, 525-535.

Wilcke, D.E., 1953. Über die vertikale Verteilung der Lumbriciden im Boden. Einteilung der Lumbricidenfauna nach Michaelsen. Zoologischer Anzeiger 154, 312-318.

Figure and Table captions

Fig.1: Projection of the 125 earthworm species, including non-assigned species and those belonging to four additional ecological categories, described by Bouché (1972) in the PCA defined using the 13 more important morpho-anatomical traits (assessed by a random forest approach using only the three main ecological categories).

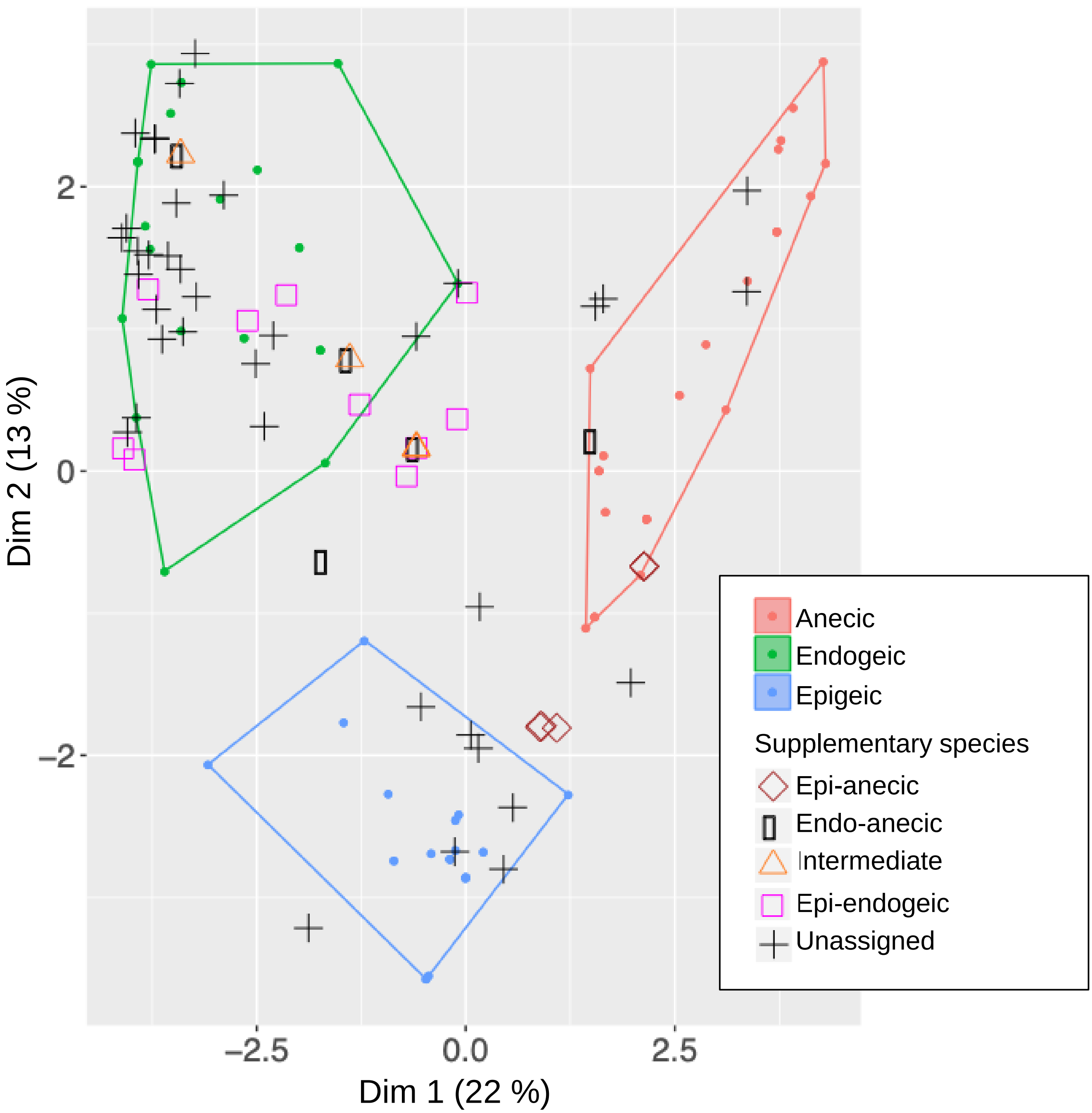
Fig.2: Ternary plot defined by the three main categories (anecic, endogeic, epigeic) where the 125 species were projected. Seven zones were defined to assign each species to one ecological category (but see Table S4), some typical species of each category are listed.

Table 1: Summary of the literature review for the 27 species selected: (i) the number of articles studied for each species; (ii) the assignment of the species in ecological categories; (iii) the percentage of articles where the assignment to an ecological category was justified and (iv) the number of articles where ecological categories were not mentioned.

Table 2: A comparison of the ecological categories assigned by Bouché (1972) with the model-based approach used in the present study.

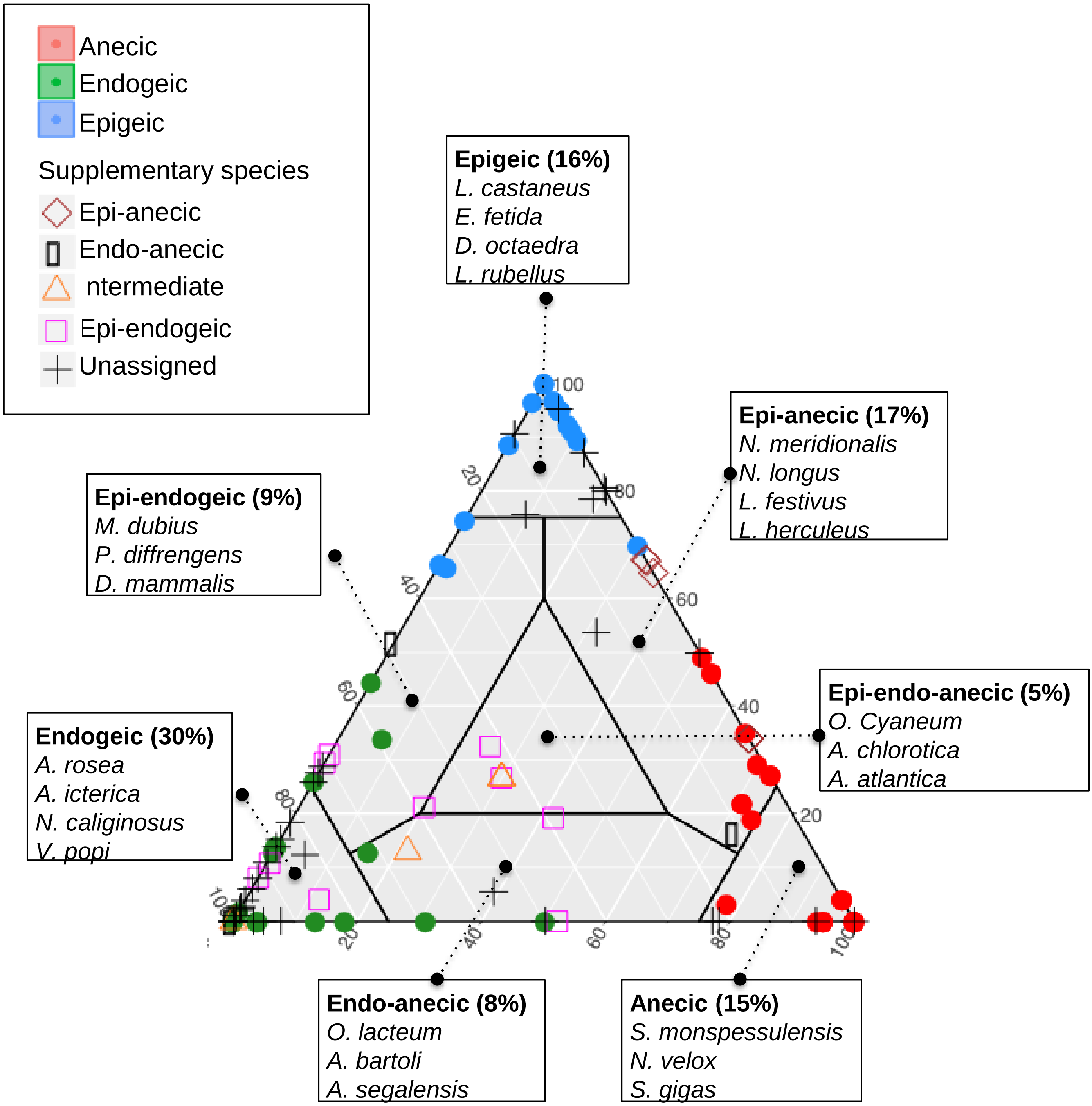
Table 3: Summary of the morpho-anatomical traits characterizing the most frequent ecological categories studied by Bouché (1972).

Fig. 1



2
0
−2
−2.5
0.0
2.5
Dim 1 (22 %)
Dim 2 (13 %)
Anecic
Endogeic
Epigeic
Supplementary species
Epi-anecic
Endo-anecic
Intermediate
Epi-endogeic
Unassigned

Fig. 2



Anecic
Endogeic
Epigeic
Supplementary species
Epi-anecic
Endo-anecic
Intermediate
Epi-endogeic
Unassigned
Epigeic (16%)
L. castaneus
E. fetida
D. octaedra
L. rubellus
Epi-anecic (17%)
N. meridionalis
N. longus
L. festivus
L. herculeus
Epi-endogeic (9%)
M. dubius
P. diffrengens
D. mammalis
Epi-endo-anecic (5%)
O. Cyaneum
A. chlorotica
A. atlantica
Endogeic (30%)
A. rosea
A. icterica
N. caliginosus
V. popi
Endo-anecic (8%)
O. lacteum
A. bartoli
A. segalensis
Anecic (15%)
S. monspessulensis
N. velox
S. gigas
100
80
60
40
20

Table 1: Summary of the literature review for the 26 species selected: (i) the number of articles studied for each species; (ii) the assignment of the species in ecological categories; (iii) the percentage of articles where the assignment to an ecological category was justified and (iv) the number of articles where ecological categories were not mentioned.

	Species	Total	Anecic	Endogeic	Epigeic	Epi-anecic	Epi-endogeic	Endo-anecic	Intermediate	Justification	No mention*
			(%)								(number)
	One category										
	Aporrectodea caliginosa	86		100						28	19
	Aporrectodea icterica	19		100						37	2
	Octolasion lacteum	19		100						5	7
	Octolasion cyaneum	33		100						6	9
	Aporrectodea rosea	64		100						11	29
	Octolasion tyrtaeum	36		100						8	9
	Pontoscolex corethrurus	52		100						8	9
	Amynthas robustus			100							
	Lumbricus castaneus	25			100					16	7
	Eisenia veneta	52			100					6	10
	Dendrobaena octaedra	52			100					2	4

	Dendrodrilus rubidus	52			100					0	16
	Perionyx excavatus	47			100					2	23
	Amynthas huperensis		100								
	Amynthas pingi		100								
	Two categories										
	Lumbricus terrestris	224	**92**		8					9	107
	Allolobophora chlorotica	56		**98**			2			9	8
	Aporrectodea nocturna	29	**95**					5		28	9
	Aporrectodea trapezoides	35	16	**84**						43	16
	Amynthas agrestis				46		56			0	8
	Amynthas hilgendorfi				50		50				
	Metaphire guillelmi	33	**80**	20						18	8
	Three categories										
	Aporrectodea longa	39	**84**	6				10		15	7
	Lumbricus rubellus	84			72	2	26			8	41

	Four categories										
	Amynthas corticis	20		13	38	13	38			0	12
	Amynthas gracilis	18	14	14	43		29			0	11
	* number of studies in which the eartworm species was not assigned to an ecological category by the authors										

Table 2: A comparison of the ecological categories assigned by Bouché (1972) with the model-based approach used in the present study.

Bouché's assignement	Model-based appoach						
	Epigeic	Anecic	Endogeic	Epi-anecic	Epi-endogeic	Endo-anecic	Intermediate
Epigeic	**14**	0	0	1	3	0	0
Anecic	0	**13**	0	12	0	0	0
Endogeic	0	0	**14**	0	3	3	0
Epi-anecic	0	0	0	**6**	0	0	0
Epi-endogeic	0	0	3	0	**2**	2	3
Endo-anecic	0	0	0	1	1	**0**	0
Intermediate	0	0	1	0	0	1	**3**
Unassigned	7	4	21	2	3	2	0
Total	21	17	39	22	12	8	6

Table 3: Summary of the morpho-anatomical traits characterizing the most frequent ecological categories studied by Bouché (1972).

Morpho-anatomical traits	Epigeic	Endogeic	Anecic	Epi-anecic
Length (mm)	62	79	349	143
Diameter (mm)	2	3	9	6
Weight (mg)	518	751	13235	3054
Pigmentation	100%	0%	100%	100%
Skin coloration	Red (86%)	Pink (61%)	Grey (100%)	Grey (55%)
Antero-posterior gradient	100%	0%	100%	100%
Dorso-ventral gradient	95%	0%	84%	100%
Flattening index	No (43%), Tail (29%)	No (45%), Clittelum (29%)	Clitellum-Tail (47%), Tail (47%)	Clitellum-Tail (45%), Tail (23%)
Epithelium rigidity	38%	0%	**89%**	**91%**
Muscle structure	Pinnate (95%)	Elementary (47%), Radial (16%)	Pinnate (63%)	Pinnate (95%)
Longitudinal furrow	0%	34%	89%	5%
Transversal furrow	43%	100%	100%	95%
Typhlosole type	Massive (57%)	Bilamelated (58%)	Pinnate (100%)	Pinnate (68%)