E-Growth: a global database on intraspecific body growth variability in earthworms

Jérôme Mathieu

Sorbonne Université, CNRS, UPEC, Paris 7, INRA, IRD, Institut d’Ecologie et des Sciences de l’Environnement de Paris, F-75005, Paris, France

Corresponding author:

Jérôme Mathieu, jerome.mathieu@upmc.fr

Abstract

Earthworms play a key role in soil and ecosystem functioning. Predicting their abundance and spatial distribution is required to understand their ecological role. There is growing evidence that mechanistic models of earthworm population dynamics are promising tools to tackle this issue. However, this approach requires a fair amount of data because it explicitly integrates the three fundamental biological processes: growth, reproduction and mortality. Hitherto, the lack of comprehensive databases on life history parameters related to these three processes hampered the widespread development of mechanistic earthworm population dynamics models. As a consequence, predicting earthworm abundance in a variety of conditions across species is still difficult.
The clear bottleneck for making progress is the lack of databases on the intraspecific variability of earthworm life history traits in response to environmental conditions. Data related to body growth and body size are critical because body size largely determines reproduction and mortality rates. Body growth is therefore the backbone of mechanistic models of earthworm population dynamics.

Here I present EGrowth, the first comprehensive database on intraspecific variability of earthworm body growth in relation to environmental conditions. The EGrowth database contains 1073 growth curves of 51 species of earthworms, representing 16002 measures of body mass. It covers publications on earthworm body size from 1900 to 2016. The environmental conditions in which the growth curves were produced are also reported. The database is open access and can be browsed from a graphical user interface. EGrowth will be updated regularly in the future as new studies are published. I propose a standardized framework for reporting future data on body growth of earthworms.

Key words: Body size; Life history traits; Interspecific variability; Intraspecific trait variability (ITV); Database; Allometry

1. Introduction

Earthworms play an important role in soil functioning (Lavelle, 1988). For instance, they modify soil structure, bulk density and aggregate stability, with direct consequences for water infiltration rates and hydrological conductivity (Blanchart, 1992; Bossuyt et al., 2005; van Schaik et al., 2014). They also affect nutrient and carbon fluxes through their effect on decomposition and microbial activity.
(Pashanasi et al., 1996; Richardson et al., 2016). These modifications impact vegetation (De Deyn et al., 2003; Hattenschwiler and Gasser, 2005; Laossi et al., 2009) and climate dynamics (Lubbers et al., 2013; Zhang et al., 2013). A number of studies have shown that all these effects quantitatively depend on earthworm abundance and traits, which are themselves constrained by environmental conditions (e.g. Jouquet et al., 2008; van Schaik et al., 2014). Understanding how the environment impacts earthworm abundance and traits in a quantitative way is therefore a key requisite for a better grasp of their role in soil functioning. However, predicting earthworm abundance and traits in a specific context is very challenging because soils are extremely heterogeneous even at fine scales.

In addition, ecological preferences vary among species. A major challenge is thus to develop a quantitative understanding of earthworm ecology for predicting earthworm abundance and traits in relation to ecological conditions.

A promising approach for tackling this issue is the development of process based models (Jager et al., 2006; Kooijman, 2010). They have been used for a variety of purposes such as understanding the spatial structure of earthworm populations, the impact of earthworms on infiltration rates, or the effect of pesticides on earthworm population dynamics (Barot et al., 2007; Baveco and De Roos, 1996; Schneider and Schroder, 2012; Vorpahl et al., 2009). These models mimic growth, reproduction and death of individuals or cohorts, in order to predict population dynamics. They require a fair amount of data to parametrize the modelled processes, which has been identified as the primary bottleneck for the development of such approaches (Schneider and Schroder, 2012).

Data on body growth variability, and thus on intraspecific body size variability (IBSV), are critical because body size determines the rate of reproduction and life span (Brown et al., 2004). Indeed cocoon production, cocoon hatchability, food consumption, and longevity all depend on the body size of earthworms (Daniel, 1990; Michon, 1954). In turn, the population dynamics and ecological effects of earthworm are also related to their body size (Brown et al., 2004).
In order to build mechanistic models of body growth, we need to estimate the effects of the environment on growth patterns. For this we need databases that document the respective effect of a variety of environmental factors. This kind of data is typically available from experiments in controlled conditions, where only the factor(s) studied vary. A large number of studies have reported this type of experiments on earthworms. However, as far as I am aware, there is no comprehensive database compiling data on the effect of the environment on intraspecific body size variability in earthworms.

Building a database on IBSV that covers a wide spectrum of earthworm species is challenging in several aspects. The first inherent difficulty is that the shape of body growth curve varies among studies (Grimm et al., 2014). In particular, growth curves can be non-monotonic because body size can regress, present oscillations or can follow a staircase shape (Lakhani and Satchell, 1970; Michon, 1954; Tondoh and Lavelle, 1997). Hence, data on body size without reference to the shape of growth pattern have limited utility for modelling the effect of the environment on body size. The ideal structure of a global database should include this diversity of growth patterns, and should enable generalization to species for which growth form has not been measured so far. A simple way to achieve this is to build IBSV databases on body growth curves – i.e. ontogenetic growth. Body growth curves are measures of body mass at different times on the same individual or on the same population. This approach allows the modelling of body growth and the calculation of a variety of body growth parameters that can be compared among ecological conditions and across species (West et al., 2001).

A second difficulty for building databases on IBSV is the retrieval of the conditions during growth. Having this information is critical for identifying the drivers of IBSV, and to quantify their effect on body size. This information also allows parametrizing reaction norm functions, which are mathematical models that predict body growth in relation to environmental conditions such as temperature (Angilletta et al., 2004; Gillooly et al., 2001; Ray, 1960). Retrieving environmental
conditions in reports on body growth of earthworm is however challenging because there are no standardized guidelines to present this type of information. As a result, these data are often missing or dispersed within the documents.

Lastly, the data by themselves are difficult to access because they are scattered in many articles, in different journals and in different kind of reports. The title and summary of the documents often do not reveal the presence of data on body growth. Growth curve data were often published in early articles that are not recorded in search engines and not available in electronic format. Data are usually difficult to reuse because they are presented only in graphical form, without the corresponding raw data. This implies a manual digitalization of the figures to reuse the data. Furthermore, centralizing existing data on body size and growth of soil animals is necessary, not only for making these data easily reusable, but also to ensure that these data will not be definitively lost in the future.

In this paper, I present the E Growth database. The database compiles existing data about body growth of 51 species of earthworm. It actually contains more than 16000 body mass measures, representing more than 1000 growth curves. The database is open access and can be accessed in a variety of manner. It can be downloaded or accessed through a Graphical Interface (GUI) from R or from internet. This database will be updated in the future with new studies. In order to facilitate this process, I propose a standardized framework for reporting future data on body growth of earthworms.

2. Material and methods

2.1 Database construction

Data were searched in articles published in peer reviewed journals and in PhD theses from 1900 to 2016. Articles were searched in different ways, with the goal to be as exhaustive as possible. An
intensive internet search was carried out through different databases, mainly Web Of Science, Scopus, Google Scholar and Researchgate. References cited within the articles were also searched for online and in various libraries in France (MNHN and IRD Bondy) and USA (UCSB and Stanford). Key words such as earthworm growth rate, body size, and ones related to earthworm ecotoxicology - a field which offers a large amount of data- were used to retrieve publications. In addition, all issues from the most relevant journals – Pedobiologia, Soil Biology & Biochemistry, Biology and Fertility of Soils, Applied Soil Ecology and Megadrilogica were checked manually through table of contents for articles containing data about earthworm growth rate. Articles that were not available in a digital format were scanned, and the text was extracted through an OCR process (Optical character recognition). Then figures and tables with relevant data were extracted, digitalized with the software DataThief, and exported into spreadsheets. Point data from figures were reported with their associated error bars when available. Error bars were converted to Standard Error bars. Metadata such as the number of measures per point, temperature and treatments were searched for manually in the text and included in the database, when available. The authors of the articles with missing data were contacted in order to complete the database. All articles were stored as pdf files in an online folder and can be accessed upon request.

Overall, 414 publications were analysed, from which only the ones with at least four monitoring dates were retained. As a consequence, many articles, in particular the ones that used Instantaneous Growth Rate (IGR) - the difference of (log) body mass between two dates -, were not considered. Studies that used adults at the beginning of the experiments were also discarded. At the end 162 publications were used to build the database. The list of articles is given in supplementary material.

2.2 Database structure
The database is organized in three tables (Fig. 1), which are described in detail in supplementary material.

2.2.1 File "curves.txt"

This table stores the growth curve data points (16002 entries, 4 columns). Each entry is the individual or average biomass of a batch of earthworms at a given time in a given experiment, with the standard error of the mean of the biomass, when available. Each growth curve has a unique identifier called "CURVE_ID". All points with the same CURVE_ID belong to the same growth curve. Be aware that the column "time" is in most cases not the age of animals but the time since the beginning of the experiment.

2.2.2 File "curves_md.csv"

This table (1073 entries, 30 columns) describes the environmental conditions in which each growth curve was produced. This table is linked to the "curves.txt" file through the CURVE_ID field. This field allows the user to retrieve the environmental conditions in which each curve was produced. For each curve the name of the species studied, the types of factors that were tested in the experiment, the level of the factors, the intra and eventually interspecific earthworm density in the container, the room temperature, the soil moisture, the geographic origin of the individuals, the food and the substrate can be obtained. In experiments with fluctuating temperature, the average temperature was reported. There is also a "REF_ID" field that indicates the source of the data. This field is used to join this table to the file "references.csv".

2.2.3 File "references.csv"

This table describes the documents from which the growth curves come from. It is linked to the "curves_md.csv" file through the "REF_ID" field. It contains also the DOI of sources, when available.
This file is essential for tracking the origin of the data and for reproducibility of analyses. Only published data were included in the database. There are 162 entries and 7 columns in this table.

2.2.4 Database access

The database is open and can be accessed in a variety of manners that are presented in details in the user guide provided in supplementary material.

You can access it through a Graphical User interface from internet at http://www.jerome-mathieu.com/open-data/egrowth or from the R software console by typing:

```
if (!require('shiny')) install.packages('shiny');
library(shiny);
runGitHub("EGrowth", "JeromeMathieuEcology")
```

Alternatively you can download the database in your computer from Zenodo


and access it through R.

A minimal example of the database usage is presented now.

This example assumes that the database (the three files "curves.txt", "curves_md.csv" and "references.csv") are stored in the R working directory.

The first step is to read the data.
Then one way to proceed is to merge the tables based on field "CURVE_ID", which is the growth curve unique identifier:

```r
EGrowth <- merge(growth, EGrowth_metadata, by = "CURVE_ID")
```

We can plot all curves of a species of interest.

Example with *Drawida willsi*:

```r
if (!require('lattice ')) install.packages('lattice');
library(lattice)
with(EGrowth[EGrowth$species == "Drawida willsi"], xyplot(bm~time | CURVE_ID, type = c("p", "smooth"), pch=20, cex =1.5,col="grey20", ylab="Biomass (mg)",xlab="Time (days)",layout=c(4, 2),
```


par.settings=list(strip.background=list(col=c("grey80")))

[end of code]

Then you can retrieve the environmental data by typing:

EGrowth_metadata[EGrowth_metadata$species =="Drawida willsi",]

[end of code]

More examples and the code of all the analyses presented in this paper are given in the User Guide.

The code of the GUI is provided in the GitHub repository:

https://github.com/JeromeMathieuEcology/EGrowth

3. Coverage of EGrowth

At the present time, the EGrowth database contains 1073 growth curves, made of 16002 measurements of biomass. These data come from 162 publications.

3.1 Temporal dynamics of publication
No data were found between 1900 and 1953. The production of data on earthworm body growth started in 1954, with the seminal PhD thesis of Jean Michon (Michon, 1954). From this time onwards, the number of published growth curves increased exponentially until the 90’s (Fig. 3a). During this period, five other theses focused on earthworm body growth and life history traits. After the 90’s, the publication rate of growth curves slowed down. The last past ten years showed a dramatic reduction in the production of new data. In particular the number of species covered has not increased since 2011 (Fig. 3b).

3.2 Species covered

Growth curve data on 51 species are included in E Growth (Fig. 4). The most documented species were by far Eisenia foetida (n=244), Lumbricus terrestris (n=131), Eisenia Andrei (n=87), Aporrectodea caliginosa (n=74) and Lumbricus rubellus (n=70) (Fig. 4). The criteria for studying a specific species was rarely reported, except for E. foetida, which is a model species in ecotoxicology (OECD, 1984). In the top five studied species, two are considered as global invasive species (L. terrestris and L. rubellus). Overall, only one endemic species (Hormogaster elisae, Spain) was studied. There are no data for endangered species such as Megascolides australis, the giant gippsland earthworm from Australia (Van Praagh, 1992), or Driloleirus americanus, the giant Palouse earthworm from USA (Sánchez-de León and Johnson-Maynard, 2009).

3.3 Locations covered

Growth curves in E Growth were issued from 86 different sites spread around the world (Fig. 5). Most of the data were produced in Europe and in India. At the moment, no dataset is available from Russia, China and central Africa. Data from Russia and China probably exist, but were not found.
because they are not yet available in English. There are much more body growth curves from temperate areas (n=693) than from tropical (n=283) or from Mediterranean climates (n=97).

[Figure 5]

3.4 The general shape of body growth curves in earthworm

The shape of body growth curves from birth to death has been poorly documented so far. The most complete work regarding this topic is the PhD of Michon (1954). He compared the pattern of body growth of 14 species, from birth to death, with extremely frequent measures of body mass. He identified several stages in the life cycle of earthworm based on the shape of the logged body mass growth curve. In particular, he showed the existence of a senescence phase, which has been poorly documented since then. From this work and from more recent ones, we can define a general pattern of earthworm growth shape (Fig. 6a).

After birth (B), juveniles (J) usually follow a lag phase, then a steep, exponential or linear, growth. In long living species, the J phase can show a staircase shape pattern (Lakhani and Satchell, 1970). When sexual maturity is attained, adults enter the reproduction stage (R) and growth is strongly reduced and sometimes becomes negative. After some time, adults enter a senescence stage (S), during which time body size decreases and sexual organs disappear.

[Figure 6]

3.5 The variety of monitoring schemes of earthworm body growth curve

In most of the publications, growth was monitored only during a fraction of lifespan. Three broad monitoring schemes can be identified (Fig. 6b).

The type 1, called Instantaneous Growth Rate (IGR), is the most often used. It is calculated as the difference of (usually log) body mass between two dates (e.g. Eriksen-Hamel et al., 2009). The monitoring generally occurs during the juvenile stage. The exact age at the beginning and at the end
of the monitoring vary among studies and is often not reported. This kind of data was not included in EGrow database because it is not suitable for fitting growth rate models.

The monitoring scheme type 2 is the most common after type 1. It refers to cases were biomass was measured on more than four occasions, the first measurement being during the juvenile stage, but not at birth. Monitoring ends either during the juvenile stage (J monitoring scheme) or during reproduction stage (JR monitoring scheme). Measures were rarely taken until senescence (JS monitoring scheme) or until death (JD).

In the third broad monitoring scheme, monitoring started at cocoon hatching. This monitoring scheme is the most informative because the age of the individuals is known. Unfortunately, this monitoring scheme is also the least frequent, probably because of the fact that obtaining a batch of synchronized hatchlings with the same age is difficult. Some techniques such as storing cocoons at low temperature has been proposed to solve this difficulty (Bouwman, 1998). However, it still requires to breed the individuals in advance, to wait for cocoon production and for hatching. This can be a severe difficulty in species with slow population dynamics. Within this monitoring scheme, a variety of monitoring length also exists. A number of studies only cover the juvenile stage (BJ). These studies are particularly useful to estimate the age of the individuals in the monitoring scheme type two, in which age at the beginning of the monitoring is not reported. However BJ data are usually not sufficient to estimate the shape of the growth shape without making strong assumption about the shape of the growth curve. In BR monitoring scheme, individuals are monitored until their mature stage, before senescence occurs. In this case, the growth curves are usually monotonic: there is no decrease in body mass. This is the most frequent sampling design in monitoring scheme type three. A very limited number of studies monitored growth rate from birth to senescence (BS) or from birth to death (BD). These monitoring schemes are however the most informative as they give estimates of life spam and the complete shape of the body growth curve.
Overall, the great majority of studies in EGrowth lasted between 100 and 200 days (Fig. 7a), but some lasted over 900 days (e.g. Mulder et al., 2007). Most sampling frequency fell into one of the three following classes: every week, every two weeks, or every month (Fig. 7b).

3.6 Drivers of body growth

Nearly all studies were performed in controlled conditions in the laboratory. Only seven growth curves, all from temperate areas, were obtained in the field (Table 1). In the laboratory, by far the two most frequently documented factors are the type of food provided (368 curves) and the toxic effect of a variety of products (330 curves). These experiments typically come from studies on vermicomposting and from ecotoxicology. The effect of air temperature (140 curves) and density (134 curves) are also well studied. All other factors are much less well studied. Surprisingly, the effect of soil properties on earthworm growth has been poorly studied (48 curves), whereas it would greatly help to understand the distribution of earthworm species in nature. Overall, 172 curves comes from experiments that tested two or more factors.

In ecotoxicology studies, the most frequent treatments are related to the type of waste used as substrate for growth, manure being the most studied (Table 2). The effect of copper and cadmium are the most documented among metals, while Moxidectin, Glyphosate and Dieldrin are the most studied among pesticides.
4. Nine recommendations for reporting body growth curves

The absence of a general framework to report data on growth has resulted in a heterogeneity of available information among studies in the EGrowth database. This has occasionally resulted in the discarding of growth curves because of the absence of critical information. It has also resulted in missing values in the database, which limits the possibilities of disentangling the effect of different factors, like temperature and humidity. This issue could easily be avoided in the future if simple guidelines were followed systematically when reporting results. The general good practices on how to report data for their reuse in meta-analyses have been proposed in various articles and are summarized by the FAIR principle: Findable, Accessible, Interoperable and Reusable (Gerstner et al., 2017; Penev et al., 2017; White et al., 2013; Wilkinson et al., 2016). Here I focus on the specific aspects in reporting growth curves for their easy re-use and integration in databases. I propose a tentative template for reporting such type of results, accompanied by a To Do checklist (Supplementary material).

4.1 Reporting the experimental design

It is critical to clarify as much as possible the experimental design. For this, four components need to be clearly explained: 1° the type of treatments that vary between curves, if any (e.g. temperature or food). 2° the treatment levels for each treatment (e.g. for temperature: 5°C, 15°C, 25°C). If there are more than one factor, the combination of levels covered (full factorial, incomplete design) must be specified. 3° the number of independent measures (replicates) per level of treatment. This is usually the number of containers per level of treatment. 4° the number of individuals per replicate: density within containers (e.g.: 3 ind.replicate-1).
If you will allow it; I would like to remind you here of a few points on experimental design. First, individuals in the same container cannot be considered as independent because of block effects. Using the individuals’ data without accounting for this block effect leads to pseudo replication (Hurlbert, 1984). Indeed, all individuals in the same container are affected similarly by any variation of the container, and hence are not independent. When it is possible to track individuals separately, we can use the raw data and integrate a block effect (i.e. container) in the analysis. However, for earthworms, this is generally not possible without using tags (Mathieu et al., 2017). The correct procedure in general is thus to use the averaged biomass by replicate, but this results in a significant loss of statistical power. In order to circumvent this problem, a solution is to distribute individuals in as many as possible separate replicates, rather than grouping them in a limited number of replicates. This maximizes the power of the study without increasing its cost (the number of individuals does not change). Following these lines, it is more efficient to put only one individual per container rather than several ones. Beyond improving the power of statistical tests, it removes all effects of density dependence such as competition and reproduction, which can interfere with the treatments, and which can vary according to species (Uvarov, 2009). In the end, it gives a better picture of the effect of treatments on growth, all other things being equal. Replicates with only one individual per container thus facilitate the comparison among studies and among species.

4.2 What is biomass and how to report it

Reporting biomass is not as trivial as it can seem at first. Indeed biomass can change significantly with body moisture, gut content, and fat content, without having an ecological or evolutionary significance. Hence, it is important that the type of biomass reported is clearly mentioned. The ideal unit for comparing studies is dry weight of gut voided individuals. This removes the issues of body moisture and gut content, which represent a large proportion body weight and which can vary substantially regardless of dry biomass. However, this unit is not practical because it requires gut voiding then killing a fraction of the individuals at each monitoring event.
A good compromise is to report fresh biomass of non-gut voided individuals, and to measure separately, on a batch of individuals of varying size, gut content as a function of fresh biomass (e.g. Bolton and Phillipson, 1976; Curry and Bolger, 1984), and the relationship between the biomass of fresh gut voided and dry gut voided individuals (Saussey, 1966).

Whatever the type of biomass reported, it is crucial to clearly define it, and to use appropriate terminology and units. For instance, in a number of publications the word "growth" is used instead of biomass. Growth is a variation of biomass over a period of time (units can be mg.day⁻¹), not a biomass at a given time (units: mg).

4.3 Reporting age rather than time

Ideally, body growth curves relate body biomass to age. This permits one to estimate growth parameters that depend on age, such as the age at growth spurt - the maximum growth rate (Parks, 1982). These kind of parameters are sometimes the only ones affected by a treatment. For instance, a particular treatment can produce a shift in phenology such as a decrease of age at maturity, and change population dynamics, without impacting maximal body size.

Reporting age is also needed for comparing growth models across studies. Indeed, parameters in growth models are often age dependent. Hence, they can only be correctly estimated if age is given in data, rather than time since the beginning of the experiment. This is particularly important when the study do not start from the birth of the individuals. In such case, if the age at the beginning of the growth curve is not given, it is difficult to translate the time axis into age. This is particularly true for studies based on IGR. Without age and weight, it is not possible to integrate IGR data in growth curve databases.

In conclusion, reporting age offers a lot more possibilities for modelling and for comparison across studies. If age is not known, then reporting the biomass of juveniles at hatching or at the first date of monitoring can be very helpful to rescale the time axis into age.
4.4 Reporting phenology

Phenology is an important aspect of body growth patterns. Indeed a number of standard indicators, such as biomass at maturity, are defined in relation to the stage of the individuals. Changes in stages can be indicated by arrows on curves (Klok and de Roos, 1996) or directly as a table giving the proportion of individuals at each stage, at each date (e.g. Elvira et al., 1996). When data are presented in an aggregated way, the most useful information is the age at which 50% of the individuals reached a given stage. Reporting the death or quiescence of animals is also interesting because it allows computing the mortality curve, and to take into account density dependent processes. This can be indicated by a sign such as stars on graphs, or directly in the raw data.

4.5 Reporting taxonomy

The correct identification of species is critical for comparing data and building databases. However, taxonomy is still being actively updated, with sometimes major revisions. Most earthworm classifications has been developed on morphological or anatomical traits, and recent molecular studies are revising many taxa (e.g. Csuzdi et al., 2017; Domínguez et al., 2015). The problem is particularly accurate for growth experiments because identifying living individuals is very difficult, and in most cases not possible until maturity. This might not be problematic to compare treatments within a study, but it becomes a matter of importance when comparing results across studies and across species. Most taxonomical errors probably consist in aggregating different species into a single species, because of cryptic species. This leads to artificially increasing intraspecific variability while decreasing interspecific differences. The problem occurs even in the most studied species such as *E. foetida*, *L. terrestris*, and *A. caliginosa*, whose taxonomical status has been revised several times. As it is almost impossible to check the identification of species in previous studies, caution should be taken in any comparative analysis. The only option to circumvent this difficulty in future...
studies is to systematically store a few individuals from each study in pure alcohol, and to check identifications with molecular techniques. This is necessary to reach a homogenous and standardized classification across studies.

4.6 Reporting the conditions of the experiments: the need for metadata

Understanding the drivers of IBSV requires that environmental conditions in which data were acquired are reported. This information should cover at minimum the air temperature, the soil pH, soil moisture, expressed in clear units, the number of individuals per container, and the dimensions of the container. Soil humidity in particular is often poorly reported although it is a critical condition for earthworm growth. The type of food, its amount, and the frequency at which it was provided should also be mentioned. These data should ideally be given for each curve and, if possible, for each measurement date. Indeed the conditions such as temperature or the number of individuals sometimes vary during the experiments.

4.7 Reporting raw data

An obvious limiting factor for building databases on body growth curves is that data are usually only presented graphically. Extracting data from graphs is tedious, time consuming and needs to be done by a careful and experienced worker. In addition, digitalizing growth curves is sometimes not possible when data overlap, or when artwork resolution is too low. A much more efficient way would be to publish the data also in tables. Ideally, data should be stored in long term repositories such as Zenodo or Figshare, which offer four key features to reuse data in the future: long term storage, easy identification and citation of data through unique identifiers (DOI), easy finding of datasets through key words in metadata, and easy download of data. Alternately, raw data can be presented in supplementary material associated with the manuscript. In both cases, it is critical to
use non-proprietary format such as comma separated text (.csv) or tab separated text (.txt), because they will always be readable from all platforms.

4.8 Report individual growth curves

It is much more useful to publish the raw data of each growth curve or each replicate, rather than their average. It clarifies the experimental design. It gives more power to analyses and simplifies them. Indeed, it is much straightforward to include all data in analysis than using aggregated data, which must be weighted by error bars and sample size. In addition, it reduces the digitalization errors, particularly when data overlap in graphs.

4.9 Report error bars and their unit

Despite previous reminders about the necessity of reporting error bars in graphs (e.g. Cumming et al., 2007), they are still absent in a number of publications. When reporting averaged data is the only possibility, it is vital to show the error bars of the mean. It is also critical to clearly mention the type of error bar that was used (e.g. standard error, standard deviation, confidence interval) and the sample size, otherwise errors bars cannot be used.

Conclusion

The E Growth database shows that a large amount of data on earthworm body growth is already published. By centralizing these data, E Growth should help the development of process based models of earthworm ecology. E Growth also reveals knowledge gaps that challenge the applicability of such models to real situations in nature. In particular, there is a clear lack of data from the field. This kind of data is critical for assessing the performances of models. In addition, the effect of critical environmental factors for earthworm growth, such as soil properties and soil humidity, have been little studied so far. Lastly, little is known about rare species that are not dominant, particularly regarding large or endemic species. In order to integrate new data, E Growth will be updated in the future. For this, a framework to report future data on earthworm body growth is proposed.
Acknowledgement

The author would like to thank all the authors who kindly replied to his emails and furnished additional details on their data, and Naoise Nunan for English editing.

References


List of publications used to build the EGrowth database on earthworm body growth variability

Contact: Jerome.mathieu@upmc.fr

a) List of publications from which data were included in the EGrowth database


53. Garcia, J.A. and C. Fragoso, *Influence of different food substrates on growth and reproduction of two tropical earthworm species (Pontoscolex corethrurus and Amynthas corticis): The 7th


114. Perrson, T., L. Lenoir, and A. Taylor, Bioturbation in different ecosystems at Forsmark and Oskarshamn, 2007, Swedish University of Agricultural Sciences. p. 34.


Publications in which data were searched to build the EGrowth database, but from which NO data were included in the EGrowth


60. Ferreira, R.C., S. Papini, and M.M. de Andrea, *Bioavailability and influence of 14C-carbofuran on Eisenia andrei avoidance, growth and reproduction in treated natural tropical soils*. 


70. Gordo, P.H., Estudio de la distribucion horizontal de "Hormogaster elisae" (Oligochaeta, Hormogastridae) en el molar y de los factores que le regulan, in Facultad de ciencias biologicas2006, Universidad complutense de Madrid: Madrid. p. 254.


Checklist for reporting growth curves of Earthworms

Reference : Mathieu J. Soil Biology and Biochemistry 2018

https://doi.org/10.1016/j.soilbio.2018.04.004

1 Experimental design

☐ The treatments are clearly defined.
☐ The treatment levels are clearly defined.
☐ The number of replicates per treatment is clearly indicated.
☐ The number of individuals per replicate is clearly indicated.

2 Biomass

☐ The type of biomass is clearly mentioned (dry vs fresh biomass).
☐ The gut content status is mentioned (void gut vs non voided gut).
☐ The unit of the biomass is given.
☐ "Growth" and "biomass" are used appropriately.
☐ The biomass at hatchling is reported.
☐ The biomass at first date of the survey is given numerically (not only on graphs), with the error bar.

Bonus

☐ Gut content was estimated.
☐ Body moisture was estimated.

3 Age and time

☐ The age of the individuals at the first measurement is clearly given.
☐ The meaning of the X axis (age or time since the beginning of the measures) is defined.
☐ The units of the time/age axis are clearly mentioned (e.g. days, weeks).

If IGR (Instantaneous growth rate) only is reported,

☐ Its formula is mentioned.
☐ The absolute biomass are also reported.
☐ The age at each measurement is reported.

4 Phenology
5 Taxonomy

- Full clitellum development is indicated.
- Death of individuals is indicated.

6 Conditions of the experiments (metadata)

All these information are reported:

- Geographical origin of the specimens.
- Geographical coordinates in decimal degrees.
- The system of coordinates.
- Country were specimens were sampled.
- Species name, with the descriptor.
- Air temperature (°C).
- Soil or substrate pH.
- pH measurement type (e.g. H2O, KCl).
- Container volume.
- Substrate mass or volume.
- Soil and substrate moisture.
- Type of moisture measurement.
- Type of food given.
- Amount of food given.
- Sampling scheme (I,II or III, see figure 4).
- Stages covered by the monitoring (BJ, JR, JRS, JRSD and so on, see figure 4).

7 Data

- Raw data and metadata are available either as supplementary material or in a data repository.
- Data are presented in txt or csv files.
- Data are given at the highest resolution possible (either by individual or by replicate) rather than as average per treatment.
- Error bars are given if averaged data are presented.
- The nature of errors bar is clearly stated.