



**HAL**  
open science

## Mathematical modeling of adipocyte size distributions: identifiability and parameter estimation from rat data

Anne-Sophie Giacobbi, Leo Meyer, Magali Ribot, Romain Yvinec, Hedi Soula,  
Chloe Audebert

### ► To cite this version:

Anne-Sophie Giacobbi, Leo Meyer, Magali Ribot, Romain Yvinec, Hedi Soula, et al.. Mathematical modeling of adipocyte size distributions: identifiability and parameter estimation from rat data. 2024. hal-04141173v3

**HAL Id: hal-04141173**

**<https://hal.science/hal-04141173v3>**

Preprint submitted on 25 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Mathematical modeling of adipocyte size distributions: identifiability and parameter estimation from rat data

Anne-Sophie Giacobbi<sup>1,\*</sup>, Leo Meyer<sup>2</sup>, Magali Ribot<sup>2</sup>,  
Romain Yvinec<sup>3,4</sup>, Hedi Soula<sup>5</sup>, Chloe Audebert<sup>1,6,\*\*</sup>

<sup>1</sup> Sorbonne Université, CNRS, Institut de Biologie Paris-Seine (IBPS), Laboratory of Computational and Quantitative Biology UMR 7238, 75005 Paris, France

<sup>2</sup> Institut Denis Poisson, Université d'Orléans, CNRS, Université de Tours, 45067 Orléans, France

<sup>3</sup> PRC, INRAE, CNRS, Université de Tours, 37380 Nouzilly, France

<sup>4</sup> Université Paris-Saclay, Inria, Centre Inria de Saclay, 91120 Palaiseau, France

<sup>5</sup> Nutriomics, La Pitié-Salpêtrière, Sorbonne Université, CNRS, 75013 Paris, France

<sup>6</sup> Sorbonne Université, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions UMR 7598, 75005 Paris, France

\* anne-sophie.giacobbi@sorbonne-universite.fr

\*\* chloe.audebert@sorbonne-universite.fr

---

## 1 Abstract

2 Fat cells, called adipocytes, are designed to regulate energy homeostasis by storing  
3 energy in the form of lipids. Adipocyte size distribution is assumed to play a role in  
4 the development of obesity-related diseases. This population of cells that do not have  
5 a characteristic size, indeed a bimodal size distribution is observed in adipose tissue.  
6 We propose a model based on a partial differential equation to describe adipocyte  
7 size distribution. The model includes a description of the lipid fluxes and the cell  
8 size fluctuations and using a formulation of a stationary solution fast computation of  
9 bimodal distribution is achieved. We investigate the parameter identifiability and es-  
10 timate parameter values with CMA-ES algorithm. We first validate the procedure on  
11 synthetic data, then we estimate parameter values with experimental data of 32 rats.  
12 We discuss the estimated parameter values and their variability within the popula-  
13 tion, as well as the relation between estimated values and their biological significance.  
14 Finally, a sensitivity analysis is performed to specify the influence of parameters on

15 cell size distribution and explain the differences between the model and the mea-  
16 surements. The proposed framework enables the characterization of adipocyte size  
17 distribution with four parameters and can be easily adapted to measurements of cell  
18 size distribution in different health conditions.

19 *keywords: parameter estimation, adipocyte size distribution, parameter identifiability,*  
20 *partial differential equation*

---

## 21 1 Introduction

22 Pathologies related to obesity are characterized by an important accretion of adipose  
23 tissue which is mainly composed of adipose cells, called adipocytes. The adipocytes  
24 are designed to regulate energy homeostasis by storing energy in form of lipids. Dur-  
25 ing an excess of energy, adipocytes compensate with two mechanisms: hypertrophy  
26 (increase in size) and hyperplasia (increase in number)[7]. Adipocyte size variations  
27 are very large with radii ranging from  $10\mu\text{m}$  to more than  $100\mu\text{m}$ , corresponding to  
28 3 orders of magnitude in volume. In addition, cell size distribution among a tissue is  
29 not unimodal but presents two peaks: one for small adipocytes (radius below  $30\mu\text{m}$ )  
30 and one for large adipocytes (above  $80\mu\text{m}$ ) [20]. A bimodal distribution of cell sizes  
31 is striking. Indeed, most cells in the population are small adipocytes, which do not  
32 contribute significantly to the storing capacity. There is no scientific consensus on the  
33 functional importance of this bimodality. However, cell size has been associated with  
34 metabolic properties dysfunction that may be linked to obesity-related pathologies  
35 [31, 24, 20, 18] or to play a role in the development of those diseases [5].

36 Few mathematical models have been proposed for adipocyte size dynamics in var-  
37 ious health conditions. In [13, 14, 15, 17], the authors consider partial differential  
38 equation models that describe adipocyte size distribution dynamics. They have as-  
39 sumed a size-dependent rate described by an imposed function where the associated  
40 parameters are difficult to relate to physiological processes. The adipocyte modeling  
41 in [19] is based on three compartments and has been developed to describe small,  
42 medium and large adipocytes. The cell size evolution depends on lipid fluxes that are  
43 related to protein concentration controlling lipotoxicity – a cellular dysfunction due  
44 to lipid accumulation in non-adipose tissue. All these models provide studies of the  
45 adipose tissue growth dynamic and its bimodality through cell hyperplasia and/or  
46 hypertrophy, but the mechanisms governing lipid fluxes involved in adipocyte hy-  
47 pertrophy have not been considered. Furthermore, model parameters lack biological  
48 meaning.

49 A detailed model of cell hypertrophy based on lipid exchanges has been proposed  
50 in [27]. Adipocyte bimodal distributions have been explained based on mathemati-

51 cal analyses. Individual-based Monte Carlo techniques were performed to solve the  
 52 model. However, this approach is computationally costly so parameter estimation  
 53 using biological measurements is very difficult. A similar simplified model, accounting  
 54 only for lipolysis (deflation), compares well with distributions obtained from fasting  
 55 rats [28].

56 The paper is organized as follows. Based on [27, 28], we formulate the mathe-  
 57 matical model in section 2. It is based on partial differential equations, to describe  
 58 stationary adipocyte size distribution. The contribution of our work is to have a  
 59 diffusion term in the partial differential equation describing the cell size fluctuations  
 60 like in [14]. Through parameter estimation, we aim at comparing the distribution  
 61 obtained with the model to cell size distribution measured in rats before any manip-  
 62 ulation [28, 12]. To perform parameter estimation, we first conduct an identifiability  
 63 analysis in order to select model parameters that can be uniquely estimated with the  
 64 available data. Using these selected parameters, we carry out a study on synthetic  
 65 data (generated with model equations). The model identifiability and the parameter  
 66 estimation on synthetic data are presented in section 3. Once the parameter estima-  
 67 tion problem is verified, in section 4 we perform parameter estimation using adipocyte  
 68 size distributions measured in 32 healthy rats [28, 12]. The estimated parameters are  
 69 presented and then commented through a sensitivity analysis. We conclude this paper  
 70 with some discussions in section 5.

## 71 **2 Mathematical model for adipocyte size distribu-** 72 **tions**

### 73 **2.1 Model construction**

Based on Soula *et al.* [27] work, we introduce a new model for adipocyte size distri-  
 bution that we aim at fitting on experimental measurements. We first briefly recall  
 the main hypotheses of the model in [27]. To represent adipocyte size density, the  
 variation of the content of lipids  $\ell$  and variation of radius  $r$  to adapt to lipid content  
 are described by,

$$\begin{cases} \frac{d\ell}{dt} = \mathcal{T}(r, \ell, L(t)), & (1) \\ \frac{dr}{dt} = \mathcal{R}(r, \ell), & (2) \end{cases}$$

74 where the term  $L(t)$  represents the extracellular amount of lipids at time  $t$ . These  
 75 two equations refer to evolution with different characteristic times: the first equation  
 76 is a rapid evolution of fatty acid content whereas the second is a slower variation of  
 77 radius to adapt to cell lipid content.

78 We first assume a quasi steady state for equation (2) to describe a faster adaptation  
 79 to lipid content. The relation between the lipid content  $\ell$  and the radius  $r$  of a cell is  
 80 then given by  $\mathcal{R}(r, \ell) = 0$ , leading to

$$\ell = \frac{V(r) - V_{em}}{V_\ell}, \quad V(r) = \frac{4}{3}\pi r^3, \quad (3)$$

81 with  $V_{em}$  the volume of the cell with no lipid,  $V_\ell$  the conversion constant: the volume  
 82 taken by 1 *nmol* of triglyceride, and the cell volume  $V(r)$  is assumed to be spherical.  
 83 Second, similarly to [14], we introduce a constant diffusion term  $D$  to represent cell  
 84 size fluctuations.

85 With the above mentioned assumptions, we can re-write the main equation in  
 86 [27], replacing  $\ell$  by (3) and keeping only the radius variable. We then consider the  
 87 cell size density  $f$  expressed as a function of time  $t \in \mathbb{R}_+$  and radius  $r \in [r_{min}, r_{max}]$ ,  
 88 and we introduce the following system:

$$\begin{cases} \partial_t f(t, r) + \partial_r(v(r, L(t))f(t, r)) - D\partial_r^2 f(t, r) = 0, & (4) \\ L(t) = \lambda - \int_{r_{min}}^{r_{max}} (V(r) - V_{em}) \frac{4\pi r^2}{V_\ell} f(t, r) dr, & (5) \\ v(r_{min}, L(t))f(t, r_{min}) - D\partial_r f(t, r_{min}) = 0, & (6) \\ v(r_{max}, L(t))f(t, r_{max}) - D\partial_r f(t, r_{max}) = 0, & (7) \end{cases}$$

89 where  $v$  is defined by

$$v(r, L) = \frac{V_\ell}{4\pi} \left( \alpha \frac{L}{L + \kappa} \frac{\rho^3}{\rho^3 + r^3} - \frac{(\beta + \gamma r^2)}{r^2} \frac{V(r) - V_{em}}{V(r) - V_{em} + V_\ell \chi} \right). \quad (8)$$

90 The total amount of lipids  $\lambda$  is assumed to be constant over time and the second  
 91 term of the right-hand side of (5) describes the intracellular amount of lipids at  
 92 time  $t$  contained within all cells. The transport function  $v$  describes the exchange  
 93 of lipids within the population of cells [27]. The lipid exchanges are based on two  
 94 biochemical processes: lipogenesis – cell store lipids – and lipolysis – release of lipids in  
 95 the extracellular environment. Lipogenesis depends on a surface-limited rate  $\alpha$ , and it  
 96 increases with the extracellular amount of lipids  $L$  with a saturation effect depending  
 97 on the value of  $\kappa$ . The parameter  $\rho$  is a cell size threshold above which lipogenesis  
 98 rate slows down. This parameter prevents the cell radius from becoming too large,  
 99 as lipogenesis rate slows down and lipolysis rate becomes the main mechanism for  
 100 lipid exchanges. Lipolysis activity includes a basal rate  $\beta$  and a surface-limited rate  
 101  $\gamma$ . The term  $\frac{V(r) - V_{em}}{V(r) - V_{em} + V_\ell \chi} = \frac{\ell}{\ell + \chi}$  is small when cells contain few lipids and becomes  
 102 close to one for larger lipid content through parameter  $\chi$ .

103 We assume that in the measurements at the time of the biopsy the adipose tissue  
 104 is at equilibrium, thus we neglect the recruitment of new cells. In addition, it has  
 105 been shown that the life time of a human adipocyte is around 10 years [2], so the cell  
 106 death is not taken into account. It gives the boundary conditions (6)-(7). The total  
 107 number of cells is then constant and we assume the density integral is 1 between  $r_{min}$   
 108 and  $r_{max}$ , which leads to

$$\forall t \geq 0, \int_{r_{min}}^{r_{max}} f(t, r) dr = 1. \quad (9)$$

109 Table 1 reports the details on model variables and parameters. The parameter  
 110 values of  $V_{em}$ ,  $V_l$ ,  $\beta$  and  $\gamma$  are known from literature [27, 28] and will be fixed. We  
 111 choose the values of  $r_{min}$  and  $r_{max}$  as the boundary values of the measured radii in  
 112 the considered adipose tissue.

## 113 2.2 Stationary solution

114 In model (4)-(7), the number of adipocytes is fixed and the total amount of lipids  
 115 is constant, thus we expect the size distribution to reach a steady state [23]. The  
 116 mathematical study of the asymptotic behavior is not the purpose of this work.

117 We denote by  $f^\infty$  and  $L^\infty$  a stationary density of cell size and the extracellular  
 118 amount of lipids respectively. A stationary solution verifies  $\partial_t f^\infty(r) = 0$ . With the  
 119 boundary conditions (6)-(7) and assuming  $D \neq 0$ , we obtain the following system:

$$\left\{ \begin{array}{l} \partial_r f^\infty(r) = \frac{1}{D} v(r, L^\infty) f^\infty(r), \end{array} \right. \quad (10)$$

$$\left\{ \begin{array}{l} L^\infty = \lambda - \int_{r_{min}}^{r_{max}} (V(r) - V_{em}) \frac{4\pi r^2}{V_\ell^2} f^\infty(r) dr. \end{array} \right. \quad (11)$$

120 We note that assuming  $f^\infty(r)$  is known for all  $r \in [r_{min}, r_{max}]$ , then  $L^\infty$  is de-  
 121 termined by the equation (11) and only depends on the unknown parameter  $\lambda$ . In  
 122 parameter identifiability analysis and parameter estimation we assume that the cell  
 123 size distribution is observed. So to simplify the dependency on parameters we con-  
 124 sider  $L$  to be a parameter instead of  $\lambda$ . We thus replace  $L^\infty$  by a parameter  $L$ , and  
 125 it leads to the following simplified model,

$$\left\{ \begin{array}{l} (f^\infty)'(r) = \frac{1}{D} v(r) f^\infty(r), \end{array} \right. \quad (12)$$

$$\left\{ \begin{array}{l} \int_{r_{min}}^{r_{max}} f^\infty(r) dr = 1, \end{array} \right. \quad (13)$$

$$\left\{ \begin{array}{l} v(r) = \frac{V_\ell}{4\pi} \left( \alpha \frac{L}{L + \kappa \rho^3 + r^3} - \frac{(\beta + \gamma r^2)}{r^2} \frac{V(r) - V_{em}}{V(r) - V_{em} + V_\ell \chi} \right), \end{array} \right. \quad (14)$$

Table 1: **Description of model variables and parameters.** Parameter units and known values are summed up in the second column and a description of each variable is given in the third column.

name	value (unit)	description
$t$	- ( $h$ )	time
$r$	$\in [7.5, 150]$ ( $\mu m$ )	adipocyte radius [28, 12]
$L(t)$	- ( $nmol$ )	extracellular amount of lipids at time $t$
$f(t, r)$	-	cell density at time $t$ with respect to radius $r$
$V_{em}$	$\frac{4\pi}{3} 6^3$ ( $\mu m^3$ )	volume of an empty adipocyte (zero lipid) [1]
$V_\ell$	$1.091 \cdot 10^6$ ( $\mu m^3 \cdot nmol^{-1}$ )	volume taken by 1 $nmol$ of triglyceride [27]
$\alpha$	- ( $nmol \cdot \mu m^{-2} \cdot h^{-1}$ )	surface-limited rate in lipogenesis
$\kappa$	- ( $nmol$ )	constant of the limiting term in lipogenesis
$\rho$	- ( $\mu m$ )	cell size threshold of the Hill function in lipogenesis
$\beta$	$31.25$ ( $nmol \cdot h^{-1}$ )	basal lipolysis rate [28]
$\gamma$	$0.27$ ( $nmol \cdot \mu m^{-2} \cdot h^{-1}$ )	surface-limited rate in lipolysis [28]
$\chi$	- ( $nmol$ )	constant of the limiting term in lipolysis
$D$	- ( $\mu m^2 \cdot h^{-1}$ )	diffusion coefficient for size fluctuations
$\lambda$	- ( $nmol$ )	total amount of lipids

126 where the unknown parameters to be estimated are  $\alpha$ ,  $L$ ,  $\kappa$ ,  $\rho$ ,  $\chi$  and  $D$ .

127 Given those parameters, we can compute a stationnary solution of model (12)-(14)  
 128 and we have for  $r \in [r_{min}, r_{max}]$ ,

$$f(r) = \frac{\exp\left(\int_{r_{min}}^r \frac{1}{D} v(s) ds\right)}{\int_{r_{min}}^{r_{max}} \exp\left(\int_{r_{min}}^r \frac{1}{D} v(s) ds\right) dr}. \quad (15)$$

129 This solution can be computed numerically and when possible, the integrals are com-  
 130 puted explicitly otherwise a trapezoid rule is used. Typically, in the computation, a  
 131 radius step of  $0.1 \mu m$  is considered and an interpolation is applied to compute  $f$  at  
 132 any radius.

### 133 **2.3 The model can represent a bimodal distribution of cell** 134 **size**

135 We first study the impact of the diffusion parameter that is the main change with  
 136 respect to model in [27]. Figure 1 shows solutions computed numerically with the

137 equation (15) for a given set of parameters. The model is able to qualitatively re-  
138 produce a bimodal distribution of cell size as measured in rats. Upon investigation  
139 of equations (12)-(14), it is immediate that the number of extremal points of  $f$ , and  
140 their locations, will depend only on the parameters that appear in the velocity  $v$  (14).  
141 We can notice in equation (12) that the introduction of parameter  $D$  does not change  
142 the definition of lipogenesis and lipolysis ( $v$  is only multiplied by  $\frac{1}{D}$ ). In addition, the  
143 diffusion process does not overtake the velocity process in cell size dynamics, other-  
144 wise flat curves would be obtained. However, variations in the value of the diffusion  
145 parameter impact the size distribution: increasing the diffusion reduces the difference  
146 between the height of the two peaks and the density value at the nadir (lowest point  
147 between the two peaks) increases with diffusion.

148 In the model of Soula et al. [27], an individual-based Monte Carlo technique  
149 (20,000 cells) has been performed leading to a large computational time. It was then  
150 very hard to perform quantitative comparison with measurements. The proposed  
151 model enables a fast computation of the cell size distribution by computing directly  
152 a stationary solution with equation (15). It is now possible to perform quantitative  
153 comparison with measured size distribution and estimate parameters.

154 Prior to this parameter estimation, we study which parameters are likely to be  
155 estimated with the available data through model parameter identifiability analysis  
156 and parameter estimation on synthetic data.

## 157 **3 Model identifiability and parameter estimation**

### 158 **3.1 Parameter identifiability analysis**

We perform an identifiability analysis of the unknown parameters of the model:  $\alpha$ ,  
 $L$ ,  $\kappa$ ,  $\rho$ ,  $\chi$  and  $D$ . We define a parameterized model  $\mathcal{M}(\theta)$  derived from equations  
(12)-(14) and study its parameter identifiability which is an intrinsic property: from  
[6], the model  $\mathcal{M}$  is said to be globally identifiable in  $\theta \in \Theta$  if

$$\forall \tilde{\theta} \in \Theta, \mathcal{M}(\theta) = \mathcal{M}(\tilde{\theta}) \Rightarrow \theta = \tilde{\theta}.$$

159 The parametric structure of model (12)-(14) is complex in the sense that it includes  
160 non-linear functions in which some parameters are combined in a product. This might  
161 result in redundancies in the model – only a smaller set of unknown parameters can  
162 be estimated – or in a non-identifiable model [4].

To study the parametric structure of the model, we first set the observed outputs,

$$x_1 = f^\infty, x_2 = r$$



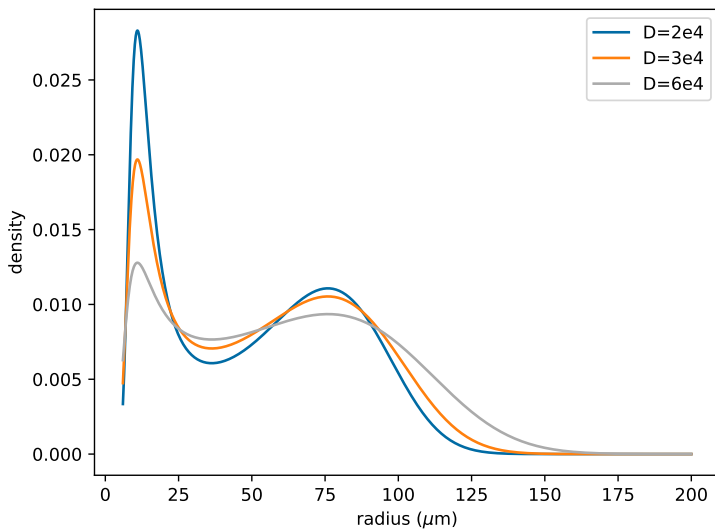


Figure 1: **Computed stationary solutions** from eq. 15 with three values for diffusion parameter. The other parameters are fixed to values reported in Table 1 and  $L = 3 \text{ nmol}$ ,  $\alpha = 0.29 \text{ nmol} \cdot \mu\text{m}^{-2} \cdot \text{h}^{-1}$ ,  $\kappa = 0.001 \text{ nmol}$ ,  $\rho = 200 \mu\text{m}$ ,  $\chi = 0.0035 \text{ nmol}$ .

163 and we introduce the following quantities to re-parameterize the model:

$$\theta_1 = \frac{\alpha L}{\beta(L + \kappa)}, \theta_2 = \rho^3, \theta_3 = V_\ell \chi \text{ and } \theta_4 = \frac{4\pi D}{V_\ell \beta}. \quad (16)$$

164 We obtain the system parameterized by  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$  the vector of unknown  
165 quantities (assumed to be strictly positive),

$$\begin{cases} \frac{dx_1}{dr} = \frac{1}{\theta_4} \left( \theta_1 \frac{1}{1 + \frac{x_2^3}{\theta_2}} - \frac{1 + \frac{\gamma}{\beta} x_2^2}{x_2^2} \frac{\frac{4}{3}\pi x_2^3 - V_{em}}{\frac{4}{3}\pi x_2^3 - V_{em} + \theta_3} \right) x_1, \\ \frac{dx_2}{dr} = 1. \end{cases} \quad (17)$$

166 We recall that the values of  $V_{em}$ ,  $\beta$  and  $\gamma$  are known (see Table 1).

167 We investigate the identifiability of unknown parameters using the **Structural**  
168 **identifiability Toolbox** of Maple [32]. It is based on the Structural Identifiability  
169 ANalyser (SIAN) algorithm which combines differential algebra and Taylor series  
170 approaches [10, 11]. From an input ODE model, a polynomial equations system is  
171 generated and the associated Gröbner basis is computed to assess the identifiability.

172 This method ranks parameters in three categories: *globally identifiable*, *locally but*  
 173 *not globally identifiable* and *non-identifiable*. A parameter  $\theta_k$  is said to be locally  
 174 identifiable if there is a finite set of possible values for  $\theta_k$  given the observation. When  
 175 a parameter is neither locally nor globally identifiable, it is called non-identifiable.

Applied to the system (17), SIAN algorithm returns that all the quantities  $\theta_k, k \in \{1, \dots, 4\}$  are globally identifiable. Going back to the model parameters in equations (12)-(14), the parameters  $V_\ell, \beta$  are known and the function  $\rho \mapsto \rho^3$  is bijective so assuming the cell size distribution is observed, the set of identifiable quantities is

$$\left\{ \frac{\alpha L}{L + \kappa}, \rho, \chi, D \right\}.$$

176 We notice that we need at least the values of  $(L, \kappa)$ ,  $(L, \alpha)$  or  $(\alpha, \kappa)$  to uniquely  
 177 estimate  $\alpha, \kappa$  or  $L$  respectively. Only a combination of these values can be uniquely  
 178 retrieved when a size distribution  $f(r)$  is given for all  $r \in [r_{min}, r_{max}]$ .

### 179 3.2 Parameter estimation procedure

180 Thanks to the parameter identifiability analysis, we know which parameters or pa-  
 181 rameter combinations we can expect to estimate from size distribution. We now need  
 182 a procedure to estimate these parameters and we want to verify this procedure on a  
 183 benchmark case: synthetic data.

184 **Minimization algorithm** To define a procedure to estimate model parameters,  
 185 we first introduce a cost function. We want to minimize this function to compare the  
 186 model output and the measurements. Then, we choose an algorithm to minimize this  
 187 function.

188 Let  $\theta$  be the parameter vector to be estimated. We denote by  $N$  the number of  
 189 measured radii for the considered observation. Given the vector of measured radii,  
 190  $(r_i)_{i=1, \dots, N}$ , we estimate  $\theta$  by minimizing the negative log-likelihood, as a cost function,  
 191 defined as follows,

$$\mathcal{L}(\theta) = - \sum_{i=1}^N \log(f(r_i, \theta)) \quad (18)$$

192 where  $f(r_i, \theta)$  is the value of a density  $f$ , solution of the model, computed at (mea-  
 193 sured) radius  $r_i$  with the parameter vector  $\theta$ . This density provides a likelihood of  
 194 finding a cell of size  $r_i$  in the adipose tissue.

195 To find the optimal parameter values, we use the Covariance Matrix Adaptation  
 196 Estimation Strategy (CMA-ES) algorithm [8]. This optimization method has been  
 197 widely used and has proved its effectiveness for mathematical model parameters esti-  
 198 mation in different fields of application like medicine [9, 26] and ecology [30]. In this

199 algorithm, from initial parameters, new possible solutions are sampled with a mul-  
 200 tivariate normal distribution. The covariance matrix depends on a step-size control  
 201 introduced to enhance the exploration of parameter space. A weighted combination  
 202 of the best candidates is then selected according to the value of the cost function  
 203 (18) and it is updated with the covariance matrix. These steps are repeated until  
 204 termination criteria are reached. At each generation, this method takes into account  
 205 recombination, mutation and selection of the possible candidates as an evolution al-  
 206 gorithm.

207 Estimation of  $(\theta_1, \rho, \theta_3, \theta_4)$  is performed with CMA-ES using cell size distribution  
 208 as observation (we replace  $\theta_2 = \rho^3$  by  $\rho$ ). The vector of parameters is also scaled to  
 209 have components of similar order of magnitude (scaling factors are  $[\theta_1 10^2, \rho 10^{-3}, \theta_3 10^{-4}, \theta_4 10^2]$ ).  
 210 Finally, to test the impact of the initial guess on the algorithm results, we perform 100  
 211 runs of CMA-ES with different initial parameters, we report the mean and standard  
 212 deviation of these runs.

213 In order to run the CMA-ES algorithm, we used `cma` Python package [33]. The  
 214 `fmin2` function of this package is used with default parameters and an initial standard  
 215 deviation of 0.05 (in each coordinate). The files to run parameter estimation are avail-  
 216 able on [https://plmlab.math.cnrs.fr/audebert/adipocyte\\_size\\_modeling](https://plmlab.math.cnrs.fr/audebert/adipocyte_size_modeling).

217 **Parameter estimation on synthetic data** We first estimate parameters with  
 218 data generated with the model (synthetic data). To generate such data, we compute  
 219 the solution of the model for chosen parameters with equation (15). Then, from  
 220 the obtained density, 10,000 samples are drawn leading to a first synthetic data set.  
 221 To mimic the true measurements we also consider a second type of synthetic data  
 222 where on the 10,000 samples only radii greater than  $10\mu m$  are observed. With this  
 223 procedure, we want to assess the impact of missing data on the parameter estimation.  
 224 To quantify the precision of the parameter estimation we compute a relative error  
 225 defined by  $\mathcal{E} = |p - p_e|/p$ , with  $p_e$  the parameter estimated value and  $p$  the true value  
 226 of the parameter (chosen to generate synthetic data).

227 Two different parameter vectors are used to obtain synthetic data sets (*synthetic*  
 228 *data set 1* and *synthetic data set 2*). The second column of Table 2 sums up the  
 229 chosen parameter values (true). The parameter estimation is performed for both  
 230 synthetic data sets without and with missing observations (Table 2 columns 3 to 8).

231 Columns 3 and 4 in Table 2 display the average and the standard deviation of the  
 232 estimated parameter values over the 100 runs. We note that the differences between  
 233 the 100 estimations can be neglected, showing that the initial guess has no impact on  
 234 the estimation.

235 In both synthetic data cases, when the estimation is performed with the complete  
 236 data set, the estimated parameter values are similar to the true values with relative

237 errors smaller than 5% (Table 2 column 5).

238 One can notice a difference between the two data sets when the estimation is  
239 performed with missing observations in the data. The last three columns of Table 2  
240 show that, depending on the considered data set, some information on parameters is  
241 lost when cells with radii larger than a threshold are only observed. In both cases,  
242 the cost function values only slightly increase compared with the complete data sets.  
243 This indicates that the model is still able to correctly represent the data sets with  
244 missing observations. In *synthetic data set 1*, the impact on the parameter estimation  
245 is relatively small and relative errors remain below 5%. In *synthetic data set 2*, we  
246 are able to correctly estimate the values of  $\theta_1$ ,  $\rho$  and  $\theta_4$  but the information about  
247 parameter  $\theta_3$  seems lost, and the relative error increases to 65%.

248 The number of observed cells is reduced in these data sets and not in the same  
249 way in each set. On synthetic data we know exactly the percentage of information  
250 that is missing. In *synthetic data set 1* when we remove samples larger than  $10\ \mu m$ ,  
251 15% of the observation is missing, whereas in *synthetic data set 2* we remove 28%  
252 of the initial distribution. This difference may explain the poor estimation of  $\theta_3$  in  
253 *synthetic data set 2* with missing observations.

254 In the case of synthetic data sets, variations along  $\theta_3$  mainly affect the first mode  
255 of cell size distribution: increasing  $\theta_3$  strongly reduces the density of small cells and  
256 slightly increases the density of large adipocytes. Therefore, with a data set of samples  
257 with radii larger than  $10\ \mu m$ , the missing information on the first mode has an impact  
258 on the estimation of  $\theta_3$ . Moreover, this parameter is related to parameter  $\chi$  that  
259 drives the lipolysis mechanism in the model (size reduction). These results show that  
260 lipolysis is important for driving small cell distribution.

261 **From estimated parameter values to parameter intervals** The identifiability  
262 analysis ensures that the minimization problem should have only one solution and  
263 the estimation procedure computes this solution. Here, we want to compute inter-  
264 vals of parameter values for which the cost function remains close to its minimum.  
265 Our approach follows the strategy of ABC method where parameters are sampled  
266 from a prior distribution and are then selected according to a criterion based on the  
267 evaluation of the model output [29].

268 To sample a parameter  $\theta_i$ , a new parameter  $\bar{\theta}_i$  is first generated uniformly in  
269  $[0.8\hat{\theta}_i, 1.2\hat{\theta}_i]$  where  $\hat{\theta}_i$  is the estimated parameter value obtained with the CMA-ES  
270 algorithm. Then, the cost function is computed with parameter  $\bar{\theta}_i$  while the other  
271 parameters are fixed at their estimated values. The parameter is selected if the cost  
272 function is below 0.1% of  $\mathcal{L}(\hat{\theta})$ . This threshold was set to investigate the parameter  
273 space with small changes on cell size distribution. Note that the parameter sampling  
274 is performed one at a time. This strategy is repeated until 1,000 replicates are selected

Table 2: **Results of parameter estimation procedure performed on synthetic data sets without and with missing data.** The first three columns display the parameter names, orders and true values for both synthetic data sets. Columns 3 and 4 present the estimated parameters for complete data sets (10,000 samples), it shows the average over 100 estimations with different initial guesses and standard deviations. The fifth column sums up the difference between true parameter and its estimation with a relative error in percentage. The three last columns present the same values for the same data sets with missing observations: only radii over  $10\mu m$  are observed (samples  $> 10\mu m$ ). All estimations are performed with CMA-ES algorithm of `fmin2` function from `cma` Python package. For each case, we present a normalized cost function defined by :  $\mathcal{L}_N(\theta) = \frac{1}{N}\mathcal{L}(\theta)$ , with  $N$  the total number of observed radii. We choose the default parameters and an initial standard deviation of 0.05 (in each coordinate). The parameters are scaled to have similar sensitivity ( $[\theta_1 \cdot 10^2, \rho \cdot 10^{-3}, \theta_3 \cdot 10^{-4}, \theta_4 \cdot 10^2]$ ).

<i>synthetic data set 1</i>			10,000 samples - $\mathcal{L}_N(\theta) = 4.20$			samples $> 10\mu m$ - $\mathcal{L}_N(\theta) = 4.26$		
parameter	order	true	esti. value	std	rel. err.	esti. value	std	rel. err.
$\theta_1$	$10^{-3}$	9.60	9.61	$1 \cdot 10^{-8}$	0.2%	9.62	$2 \cdot 10^{-8}$	0.3%
$\rho$	$10^2$	1.50	1.50	$1 \cdot 10^{-8}$	0.2%	1.49	$2 \cdot 10^{-8}$	0.8%
$\theta_3$	$10^3$	2.18	2.17	$5 \cdot 10^{-8}$	0.6%	2.09	$2 \cdot 10^{-7}$	4.2%
$\theta_4$	$10^{-3}$	7.37	7.20	$2 \cdot 10^{-7}$	2.3%	7.35	$4 \cdot 10^{-7}$	0.3%
<i>synthetic data set 2</i>			10,000 samples - $\mathcal{L}_N(\theta) = 4.18$			samples $> 10\mu m$ - $\mathcal{L}_N(\theta) = 4.54$		
parameter	order	true	esti. value	std	rel. err.	esti. value	std	rel. err.
$\theta_1$	$10^{-3}$	9.92	9.92	$1 \cdot 10^{-8}$	0.04%	9.91	$1 \cdot 10^{-7}$	0.1%
$\rho$	$10^2$	2.00	2.00	$1 \cdot 10^{-8}$	0.2%	2.01	$5 \cdot 10^{-8}$	0.6%
$\theta_3$	$10^2$	3.27	3.12	$2 \cdot 10^{-7}$	4.8%	5.39	$4 \cdot 10^{-6}$	65%
$\theta_4$	$10^{-2}$	1.11	1.12	$2 \cdot 10^{-8}$	1.7%	1.12	$1 \cdot 10^{-7}$	1.2%

Table 3: **Range of selected values for the parameters.** The first three columns show the parameter names, orders and true values. For each data set, the estimated parameter value (column "esti. value") with CMA-ES method is subject to a maximum of 20% variation (column "esti.  $\pm 20\%$ "). From this variation, a range of values is selected for each parameter (column "selec. values") allowing a maximum error rate of 0.1% on the value of the estimated cost function  $\mathcal{L}$ . For each parameter 1,000 samples are generated

<i>synthetic data set 1</i>			10,000 samples			samples $> 10\mu m$		
parameter	order	true	esti. value	esti. $\pm 20\%$	selec. values	esti. value	esti. $\pm 20\%$	selec. values
$\theta_1$	$10^{-3}$	9.60	9.61	7.69 - 11.53	9.58 - 9.63	9.62	7.70 - 11.54	9.59 - 9.65
$\rho$	$10^2$	1.50	1.50	1.20 - 1.80	1.47 - 1.53	1.49	1.19 - 1.79	1.46 - 1.52
$\theta_3$	$10^3$	2.18	2.17	1.74 - 2.60	2.05 - 2.29	2.09	1.67 - 2.51	1.91 - 2.29
$\theta_4$	$10^{-3}$	7.37	7.20	5.76 - 8.64	6.54 - 8.02	7.35	5.88 - 8.82	6.58 - 8.32
<i>synthetic data set 2</i>			10,000 samples			samples $> 10\mu m$		
parameter	order	true	esti. value	esti. $\pm 20\%$	selec. values	esti. value	esti. $\pm 20\%$	selec. values
$\theta_1$	$10^{-3}$	9.92	9.92	7.94 - 11.90	9.90 - 9.95	9.91	7.92 - 11.89	9.86 - 9.95
$\rho$	$10^2$	2.00	2.00	1.60 - 2.40	1.97 - 2.03	2.01	1.61 - 2.41	1.99 - 2.05
$\theta_3$	$10^3$	3.27	3.12	2.49 - 3.74	2.69 - 3.58	5.39	4.31 - 6.47	4.32 - 6.47
$\theta_4$	$10^{-2}$	1.11	1.12	0.90 - 1.34	1.05 - 1.21	1.12	0.90 - 1.34	0.98 - 1.28

275 per parameter.

276 Table 3 shows for each parameter the considered range of values and the selected  
277 intervals for each synthetic data set. For synthetic data sets without missing obser-  
278 vations, the range of values selected by the procedure is reduced in comparison with  
279 the initial one and contains the true parameter. This analysis gives an information  
280 on the range of accepted values for each parameter. We note that, in *synthetic data*  
281 *set 1*, the model output seems less sensitive to parameter  $\theta_4$  that has the largest range  
282 of selected values. In *synthetic data set 2* the largest range of selected values is for  
283 parameter  $\theta_3$ .

284 In data sets with missing observations, the selected ranges are not impacted for  
285 *synthetic data set 1* (small difference for  $\theta_3$ ). In *synthetic data set 2*, the loss of  
286 information about small cells leads to the selection of the total initial interval for  
287 parameter  $\theta_3$  ( $\pm 20\%$  of the estimated value) and an important increase of the se-  
288 lected range for  $\theta_4$  (almost twice the length) compared to the case without missing  
289 observations. As observed in section 2.3, parameter  $D$  (hence  $\theta_4$ ) controls the rela-  
290 tive heights of both modes in the cell size distribution. This can explain that data  
291 sets with missing observations on small sizes lead to higher uncertainty on  $\theta_4$ . These  
292 results are in agreement with the computed relative errors of the previous paragraph  
293 (Table 2).

## 4 Application to adipocyte size distribution measured in rats

### 4.1 Measurements of adipocyte size distribution

The measured cell size distributions used to perform parameter estimation come from previous experiments [28] and data from [12], but this part of the experiment has not been published. Here, only adipocyte size distributions of animals in normal physiological conditions are considered. We assume that these distributions represent a stable state for adipose tissue, corresponding to a steady state of the mathematical system.

We use two data sets of size distribution in retroperitoneal adipose tissue for a total of 32 male Wistar rats (20 rats METAJ, aged between 20 and 24 months, Charles River, L'Arbresle, France and, 12 rats EMPA, 12-week-old, Le Genest-Saint-Isle, France). Cell size distributions were measured with Beckman Coulter Multisizer IV (Beckman Coulter, Villepinte, France), which resulted in bimodal distributions [20, 16]. Due to limitations in measurement techniques, only cell radii larger than  $7.5\mu m$  for the first experiment and  $10\mu m$  in the second were measured. The measurement of large diameters is not limited by the measurement techniques. Each animal cell size distribution is composed of a minimum of 6,000 cell radii.

### 4.2 Parameter estimation with measured data

The estimation procedure validated on synthetic data is now applied to measured size distributions. Parameter estimation is performed with CMA-ES algorithm with radius distributions measured for 32 rats in the same experimental conditions. Figure 2 shows four examples of model-data fitting (the model fitting results of the 32 rats are available on [https://plmlab.math.cnrs.fr/audebert/adipocyte\\_size\\_modeling](https://plmlab.math.cnrs.fr/audebert/adipocyte_size_modeling)). These results show the ability of the model to reproduce different types of cell size distribution. The height of each peak is not always correctly captured. This could be related to the loss of information due to missing observation for small cells in experimental data. In addition, the nadir is always underestimated by the model. We hypothesize that we are missing a process in the model to properly capture this point. However, the overall size distribution obtained with the model is in good agreement with the measured one. This result is underlined by the obtained cost function values, which are of the same order of magnitude as those obtained in the synthetic data cases (Table 4).

Table 5 shows the mean, standard deviation and relative standard deviation (RSD) of the estimated parameter values obtained in the 32 rats. The RSD are relatively small for  $\theta_1$  and  $\rho$ , showing that the size distribution of adipocytes for rats in the

330 same experimental conditions can be characterized with parameters in the same value  
 331 ranges. The variability in the population is larger for parameter  $\theta_3$  and  $\theta_4$  (larger  
 332 RSD). However, the previous analysis on synthetic data showed that less confidence  
 333 in the estimation is expected for these parameters, especially  $\theta_3$ .

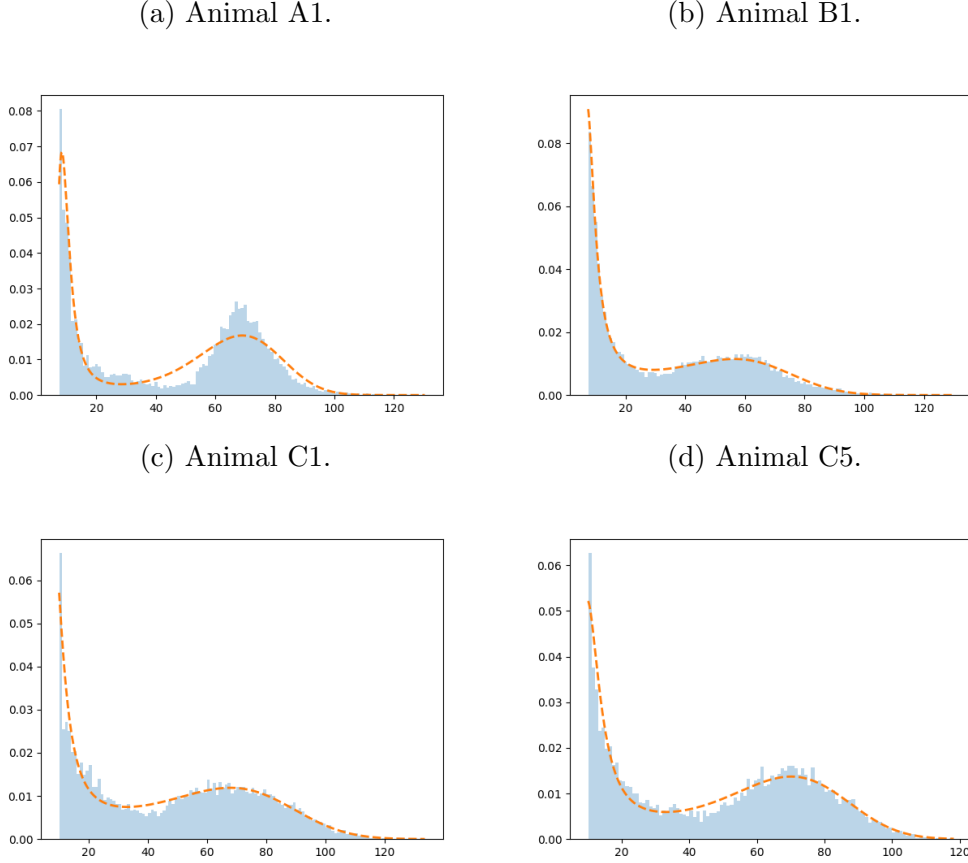


Figure 2: **Comparison model-data.** Four examples (over 32) of adipocyte radius distributions (in  $\mu m$ ) as histograms in rat in normal physiological conditions and model output computed (dash lines) with estimated parameters (see section 4.1). The parameter estimations are performed with CMA-ES algorithm of `cma` Python package by minimizing the function  $\mathcal{L}$  eq. 18.

334 For each animal, accepted parameter ranges are also computed following the pro-  
 335 cedure described in section 3.2 (Table 4). Figure 3 displays for each parameter the  
 336 estimated value for each animal with the range of selected values (dots and bars).  
 337 As expected, the parameter ranges are larger for parameters  $\theta_3$  and  $\theta_4$  compared to  
 338 parameters  $\theta_1$  and  $\rho$ . Figure 3 also shows the mean (dash red line) and the standard



339 deviation (gray area) over the rat population for each estimated parameter. It en-  
 340 ables to compare the amplitude of the range of accepted values for each parameter  
 341 for each animal with the variability within the population. We can see that for each  
 342 parameter the range of accepted values is always smaller than the standard deviation  
 343 in the population. It shows that the largest standard deviation within the population  
 344 obtained for  $\theta_3$  and  $\theta_4$  (Table 5) should not be attributed to less confidence in the  
 345 estimations.

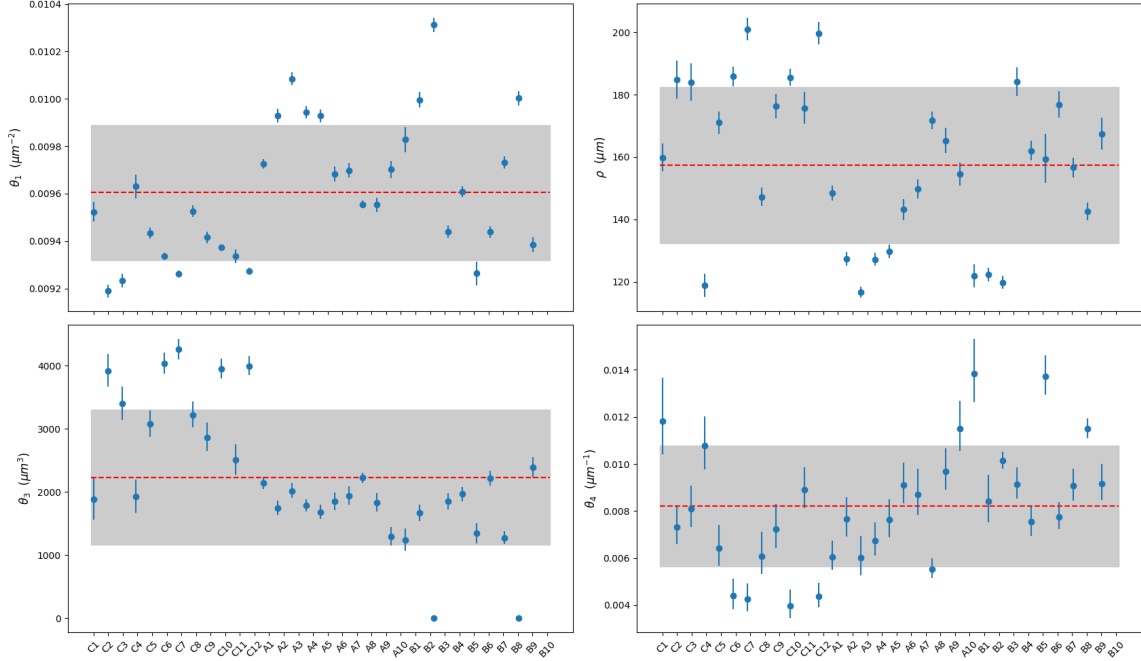


Figure 3: **Group variability and range of selected values.** Upper left and right figures display the results for parameters  $\theta_1$  and  $\rho$ . Lower left and right figures show the results for parameters  $\theta_3$  and  $\theta_4$ . For each estimated parameter the average over the population is shown with dash red line and the gray area is one standard deviation around the average computed over the population (values are reported in Table 5). For each parameter, the estimated value for each animal is displayed with dots and the bar represents the range of selected values. These ranges consist in values of the parameter (assuming the 3 others are fixed) for which the maximal cost function is 0.1% of the obtained cost function with the estimation (see section 3.2). All numerical values are reported in Table 4.

346 The range of selected values of parameter  $\theta_3$  in rats population is between 1070 and  
 347 4429  $\mu\text{m}^3$ . From this range of values, we can compute a range of radii for which the  
 348 lipolysis term becomes mainly a surface based mechanism (i.e.  $(V(r) - V_{em})/(V(r) -$

349  $V_{em} + \theta_3) > 0.95$ ). We find radii in the range  $17.2 - 27.3 \mu m$ .

350 Similarly, for lipogenesis, the parameter  $\theta_1$  is estimated within the rats population  
351 between  $0.0092$  and  $0.010 \mu m^{-2}$ . We remind that this quantity is a combination of  
352 parameters :  $\theta_1 = \frac{\alpha L}{\beta(L+\kappa)}$  and parameter  $\beta$  is known [28]. We then obtain an esti-  
353 mation of  $\frac{\alpha L}{(L+\kappa)}$  between  $0.29$  and  $0.31 nmol.\mu m^{-2}.h^{-1}$ . In the case of high available  
354 lipids,  $L$  is large and we can assume  $\frac{L}{L+\kappa} \sim 1$ . Under this assumption, the parameter  
355  $\alpha$  is estimated between  $0.29$  and  $0.31 nmol.\mu m^{-2}.h^{-1}$ . An alternative case is for low  
356  $L$ , then we can assume  $\frac{L}{L+\kappa} \sim L$  and the estimated values of  $\theta_1$  provide an estimation  
357 for  $\alpha L$ .

358 The cell size threshold  $\rho$  of the Hill function in lipogenesis term is estimated in  
359 the range  $115 - 204 \mu m$ . Above this threshold, the term  $\rho^3/(\rho^3 + r^3)$  is smaller than  
360  $0.5$  and limits the growth of the cell.

### 361 4.3 Sensitivity analysis

362 In order to investigate the differences between model output and measured cell size  
363 distribution, a sensitivity analysis is performed. Sensitivity analysis is a local analysis  
364 and quantifies how sensitive the model output is to parameter changes. We choose to  
365 apply the Sobol' method [25]. The sensitivity indices are based on the decomposition  
366 of the output variance at each cell size point.

367 The first order index measures the singular effect of a parameter on the model  
368 output. It represents the contribution part of the parameter alone in the variability of  
369 model output. A high value of this index indicates a high contribution of the param-  
370 eter, which means that the model output is highly sensitive to this parameter. The  
371 total order index enables to include the effects depending on parameter interactions  
372 (higher order indices).

373 The model output is the cell size distribution  $f$  computed with equation (15) for  
374 radii from  $7.5 \mu m$  to  $140 \mu m$ . To study the influence of the estimated parameters, each  
375 parameter  $\theta_i$  is uniformly distributed in a range of  $\pm 1\%$  of estimated mean over the  
376 population of rats (Table 5). The change of  $\pm 1\%$  in parameters values is chosen such  
377 that the adipocyte size distributions computed with these parameters are bimodal.  
378 Then, Saltelli algorithm is performed to explore the parameter space leading to the  
379 generation of  $n(2d+2)$  parameter samples with a Monte-Carlo approach [22, 25]. We  
380 choose  $n = 2048$  and  $d = 4$  the number of parameters. The sensitivity analysis is  
381 performed using the SALib Python Library [25, 22, 3, 21].

382 Figure 4(a) shows cell size distributions ranges computed with parameters from  
383 the sampling design. With these small perturbations, a large variability is found  
384 between the cell size distributions around the two modes. The first mode of the  
385 adipocyte size density is represented by cells with radii from  $7.5$  to  $10 \mu m$ . Regarding

386 large adipocytes, the higher densities present a high variability and correspond to  
387 adipocyte size values from  $50$  to  $120\mu m$ . These results illustrate the heterogeneity of  
388 cell sizes that can be obtained with the model with small changes in parameters.

389 Then, Sobol' indices are computed to determine which parameters are most influ-  
390 ential on the cell size dynamic. The first-order indices are displayed for several radii  
391 and each parameter in Figure 4(b). The results indicate that parameter  $\theta_1$  explains  
392 the most the variations of cell sizes with a first-order sensitivity index between  $0.6$   
393 and  $1$  for all radii. Interestingly, for the cells with radii around  $40\mu m$ , the index of  
394  $\theta_1$  decreases and we notice that  $\rho$  index increases (index equals  $0.36$ ). It shows that  
395 parameter  $\rho$  around this point explains the variability of the model output up to  
396  $36\%$ . The impacts of  $\theta_3$  and  $\theta_4$  are almost negligible on cells size distribution. From  
397  $r = 90\mu m$ , the results show that the influence of  $\theta_1$  decreases whereas  $\rho$  becomes  
398 more influential and explains up to  $18\%$  of the output variability. The total-order  
399 sensitivity indices are also computed (not shown) and are similar to first-order indices,  
400 revealing that parameter interactions have a negligible influence on the adipocyte size  
401 distributions.

402 The sensitivity analysis suggests that the cell size dynamics in rats is mainly driven  
403 by the parameters depending on lipogenesis, and especially by  $\theta_1$  which represents  
404 the combination of the unknown parameters  $(\alpha, \kappa, L)$ .

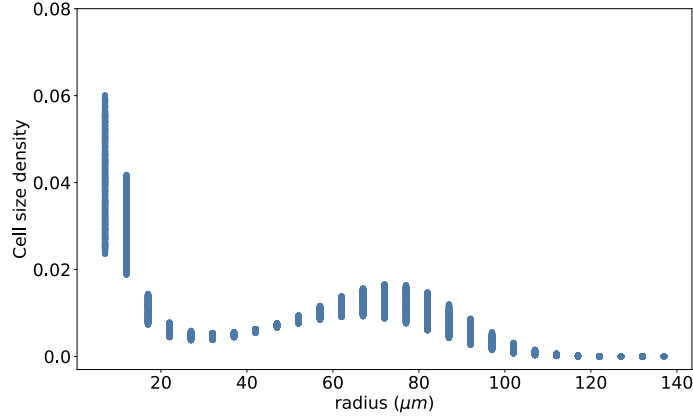
405 Parameters  $\theta_3$  and  $\theta_4$ , associated with lipolysis (through  $\chi$ ) and diffusion ( $D$ )  
406 respectively, have a negligible impact on the cell size dynamic along all cell sizes.  
407 This result confirms the difficulty to identify these parameters in practice and are  
408 in agreement with the largest ranges of selected parameter values. In addition, this  
409 study highlights the fact that the nadir is difficult to capture since we observe an  
410 opposite change in the parameter sensitivity around this radius. With this study we  
411 are able to explain the results of parameter estimation on the measured data.

## 412 5 Discussion

413 We presented a mathematical model to describe adipocytes cell size distribution,  
414 based on a partial differential equation and including lipid exchanges. With the  
415 formulation of a stationary solution, we were able to solve numerically and efficiently  
416 this model. Prior to the estimation of parameter with measurements, we analyzed  
417 which parameter can be identifiable and how reliable are the estimations.

418 The identifiability of unknown parameters was studied with a re-parameterized  
419 form of the model. We showed that only four quantities can be uniquely identified  
420 and that three of our parameters of interest are related. These three parameters  
421 cannot be identified separately with an observation of the cell size distribution only.  
422 However, we can identify the threshold radius  $\rho$  involved in lipogenesis, the lipolysis

(a) Cell size distributions



(b) Estimation of first-order Sobol' indices

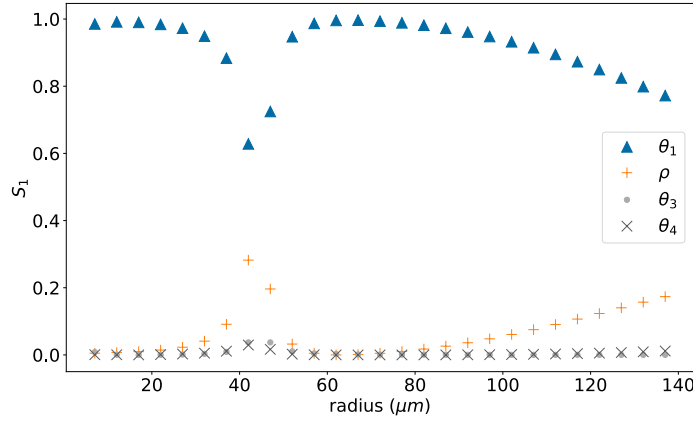


Figure 4: (a) A sample of cell size distributions. The parameter sampling design is constructed using Saltelli algorithm where each parameter is uniformly distributed in a range of values corresponding to  $\pm 1\%$  of the mean of its estimated value in rats (Table 5). A number of 20,480 samples giving bimodal distributions are generated to estimate the Sobol' indices. (b) Estimation of first-order Sobol' indices for  $\theta_1$ ,  $\rho$ ,  $\theta_3$  and  $\theta_4$  using a Monte-Carlo based approach [25, 22, 3, 21].

423 threshold  $\chi$  as well as the diffusion coefficient  $D$  that describes cell size fluctuations.

424 The model calibration on synthetic data sets showed, in practice, an accurate  
 425 estimation of the parameters. When we considered data sets with missing observations  
 426 (similar to the measurements), we found that three over the four quantities can be  
 427 correctly estimated.

428 The model parameters were estimated on 32 adipocyte size distributions measured  
 429 in rats. With these estimated parameters, the overall distribution of cell size was

430 captured. However, the nadir part of the distribution as well as the height of the  
431 modes were not perfectly reproduced. It is possible that the model is missing some  
432 aspect of the adipocyte size dynamics that would help to better capture the nadir.  
433 This is supported by the sensitivity analysis, that showed that the nadir part was not  
434 sensitive specifically to one of the four considered parameters. Therefore, it makes  
435 this part of the distribution difficult to fit. In addition, in the presented model, the  
436 diffusion parameter  $D$  via  $\theta_4$  affects linearly both lipogenesis and lipolysis. It would  
437 be interesting to change this modeling assumption with a more complex diffusion  
438 process, impacting differently lipogenesis and lipolysis. For instance, considering a  
439 size dependent diffusion coefficient could improve the agreement between the model  
440 outputs and the observations.

441 We also think that our assumption regarding the normalization of the cell size  
442 distribution (it integrates to 1 between  $r_{min}$  and  $r_{max}$ ) affects the fits (especially  
443 the height of the 2 modes). However, we have no background knowledge about the  
444 total number of adipocytes in the distribution. In addition, we know that the data  
445 collection does not include cells with a radius below a certain threshold. In [13], a  
446 formulation has been proposed to approximate the total cell number in a fat pad  
447 but to do this estimation, we need to have the fat pad mass which is not the case  
448 in our experimental data. An other way to solve this issue would be to introduce  
449 a parameter that quantifies the total number of cells. However with an additional  
450 parameter, we will lose parameter identifiability. Then, we might need to fix other  
451 unknown quantities, so this solution only shifts the problem.

452 Nevertheless, we have estimated parameter values for 32 rats. We found a larger  
453 variability between rats in the estimated values of  $\theta_3$  and  $\theta_4$  (Figure 3). However,  
454 the sensitivity analysis showed that the model is less sensitive to these parameters  
455 (Figure 4). For  $\theta_1$  and  $\rho$ , the estimated values were more robust within the population  
456 leading us to believe that  $\theta_1$ ,  $\rho$  are less individual-specific parameters. However they  
457 could change if the estimation is performed with another species. This result sug-  
458 gests lipolysis (driven by  $\chi$ ) is more an individual-dependent process than lipogenesis  
459 (driven by  $\theta_1$  and  $\rho$ ) that is more constant within the population.

460 Recruitment of new cells via adipogenesis or cell death were not included in  
461 our model. Since we were looking at the distribution of size at one specific time,  
462 these mechanisms can be neglected. However, if one wants to represent longitudinal  
463 adipocyte size distributions especially in case of diet changes, these processes should  
464 be considered. This will have an impact on the cell size distribution, especially for  
465 small cells, as suggested in [28]. Moreover, it is known that past diets affect the  
466 adipocyte size regulation and may be irreversible [13, 27]. Indeed, past diets could  
467 lead to a larger number of cells in the tissue. However, in the presented model, the  
468 number of cells is not explicitly considered. This assumption should be modified to  
469 take into account longitudinal size distributions and to be able to compare animals

470 with different diets. In past works [13, 14, 15, 17], the authors have considered partial  
471 differential equation models that take into account a recruitment rate of new cells.  
472 Our model could be extended with this extra term for adipogenesis modeling.

473 We believe that the presented framework can be adapted to estimate model pa-  
474 rameters with adipocyte size distribution in other species than rats and in different  
475 health conditions. Our current data set is not rich enough to enable us to study the  
476 relation between model parameters and animal health conditions. With an adequate  
477 data set, the presented framework may enable to establish links between the mathe-  
478 matical model parameters and health conditions based on adipocyte size distribution  
479 observations. The final purpose is to be able to characterize and potentially classify  
480 the different obesity-related pathologies.

## 481 6 Acknowledgments

482 This work was supported the ANR MATIDY grant (ANR-20-CE45-0003).

## 483 References

- 484 [1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002)  
485 Molecular biology of the cell, fourth edition, 197.
- 486 [2] Arner, E., Westermark, P. O., Spalding, K. L., Britton, T., Rydén, M., Frisén, J.,  
487 Bernard, S., Arner, P. (2010) Adipocyte turnover: relevance to human adipose  
488 tissue morphology. *Diabetes*, 59(1), 105–109. <https://doi.org/10.2337/db09-0942>
- 489 [3] Campolongo, F., Saltelli, A., Cariboni, J. (2011) From screening to quantitative  
490 sensitivity analysis. A unified approach. *Computer Physics Communications* 182,  
491 978–988. <https://doi.org/10.1016/j.cpc.2010.12.039>.
- 492 [4] Catchpole, E. A., MORGAN, B. J. T. (1997) Detecting parameter redundancy.  
493 *Biometrika*, 84(1), 187–196. <https://doi.org/10.1093/biomet/84.1.187>
- 494 [5] Clément, K. (2011) Bariatric surgery, adipose tissue and gut microbiota. *Int J*  
495 *Obes* 35 (Suppl 3), S7–S15. <https://doi.org/10.1038/ijo.2011.141>
- 496 [6] Cole, D.J., Morgan, B.J., Titterton, D.M. (2010) Determining  
497 the parametric structure of models. *Math Biosci*, 228(1), 16–30.  
498 <https://doi.org/10.1016/j.mbs.2010.08.004>

- 499 [7] Drolet, R., Richard, C., Sniderman, A.D., Mailloux, J., Fortier, M., Huot,  
500 C., Rhéaume, C., Tchernof, A. (2008) Hypertrophy and hyperplasia of ab-  
501 dominal adipose tissues in women. *Int. J. Obes. (London)* 32 (2), 283–291.  
502 <http://dx.doi.org/10.1038/sj.ijo.0803708>
- 503 [8] Hansen, N. (2016) The CMA Evolution Strategy : A Tutorial.  
504 <https://doi.org/10.48550/arXiv.1604.00772>
- 505 [9] Hesse, J., Müller, T., Relógio, A. (2023) An integrative mathematical model for  
506 timing treatment toxicity and Zeitgeber impact in colorectal cancer cells. *npj*  
507 *Syst Biol Appl* 9, 27. <https://doi.org/10.1038/s41540-023-00287-4>
- 508 [10] Hong H., Ovchinnikov A., Pogudin G., Yap, C. (2019) SIAN: software for struc-  
509 tural identifiability analysis of ODE models. *Bioinformatics*, 35(16), 2873–2874.  
510 <https://doi.org/10.1093/bioinformatics/bty1069>
- 511 [11] Hong H., Ovchinnikov A., Pogudin G., Yap, C. (2020) Global identifiability of  
512 differential models. *Communications on pure and applied mathematics*, 73(9),  
513 1831–1879. <https://doi.org/10.1002/cpa.21921>
- 514 [12] Jacquier M., Crauste F., Soulage C.O., Soula H.A. (2014) A Pre-  
515 dictive Model of the Dynamics of Body Weight and Food Intake in  
516 Rats Submitted to Caloric Restrictions. *PLOS ONE* 9(6): e100073.  
517 <https://doi.org/10.1371/journal.pone.0100073>
- 518 [13] Jo, J., Gavrilova, O., Pack, S., Jou, W., Mullen, S., et al. (2009) Hypertrophy  
519 and/or Hyperplasia: Dynamics of Adipose Tissue Growth. *PLOS Computational*  
520 *Biology* 5(3): e1000324. <https://doi.org/10.1371/journal.pcbi.1000324>
- 521 [14] Jo, J., Shreif, Z., Periwal, V. (2012) Quantitative dynamics of adipose cells.  
522 *Adipocyte*, 1(2), 80–88. <https://doi.org/10.4161/adip.19705>
- 523 [15] Jo, J., Shreif, Z., Gaillard, J.R., Arroyo, M., Cushman, S.W., Periwal, V.  
524 (2013) Mathematical Models of Adipose Tissue Dynamics. In: Gefen A., Be-  
525 nayahu D. (eds) *The Mechanobiology of Obesity and Related Diseases. Studies*  
526 *in Mechanobiology, Tissue Engineering and Biomaterials*, 16, Springer, Cham.  
527 [https://doi.org/10.1007/8415\\_2013\\_170](https://doi.org/10.1007/8415_2013_170)
- 528 [16] Laforest, S., Michaud, A., Paris, G., Pelletier, M., Vidal, H., Géoën, A., Tch-  
529 ernof, A. (2017) Comparative analysis of three human adipocyte size measure-  
530 ment methods and their relevance for cardiometabolic risk. *Obesity*, 25, 122–131.  
531 <https://doi.org/10.1002/oby.21697>

- 532 [17] Li, Y., Periwal, V., Cushman, S. W., Stenkula, K. G. (2015) Adipose  
533 cell hypertrophy precedes the appearance of small adipocytes by 3 days in  
534 C57BL/6 mouse upon changing to a high fat diet. *Adipocyte*, 5(1), 81–87.  
535 <https://doi.org/10.1080/21623945.2015.1128588>
- 536 [18] Lönn, M., Mehlig, K., Bengtsson, C., Lissner, L. (2010) Adipocyte size predicts  
537 incidence of type 2 diabetes in women. *FASEB journal : official publication of*  
538 *the Federation of American Societies for Experimental Biology*, 24(1), 326–331.  
539 <https://doi.org/10.1096/fj.09-133058>
- 540 [19] MacKellar, J., Cushman, S.W., Periwal, V. (2010) Waves of Adipose Tissue  
541 Growth in the Genetically Obese Zucker Fatty Rat. *PLoS ONE* 5(1): e8197.  
542 <https://doi.org/10.1371/journal.pone.0008197>
- 543 [20] McLaughlin, T., Sherman, A., Tsao, P., Gonzalez, O., Yee, G., Lamendola, C.,  
544 Reaven, G. M., Cushman, S. W. (2007) Enhanced proportion of small adipose  
545 cells in insulin-resistant vs insulin-sensitive obese individuals implicates impaired  
546 adipogenesis. *Diabetologia*, 50(8), 1707–1715. [https://doi.org/10.1007/s00125-](https://doi.org/10.1007/s00125-007-0708-y)  
547 [007-0708-y](https://doi.org/10.1007/s00125-007-0708-y)
- 548 [21] Owen, A. B. (2020) On dropping the first Sobol’ point. arXiv:2008.08051 [cs,  
549 math, stat]. <http://arxiv.org/abs/2008.08051>.
- 550 [22] Saltelli, A. (2002) Making best use of model evaluations to compute  
551 sensitivity indices, *Computer Physics Communications*, 145(2), 280–297.  
552 [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1).
- 553 [23] Schlichting, A. (2019) The Exchange-Driven Growth Model: Basic Prop-  
554 erties and Longtime Behavior. *Journal of Nonlinear Science*, 30, 793–830.  
555 <https://doi.org/10.1007/s00332-019-09592-x>
- 556 [24] Skurk, T., Alberti-Huber, C., Herder, C., Hauner, H. (2007) Rela-  
557 tionship between adipocyte size and adipokine expression and secretion.  
558 *The Journal of clinical endocrinology and metabolism*, 92(3), 1023–1033.  
559 <https://doi.org/10.1210/jc.2006-1055>.
- 560 [25] Sobol’ I.M. (2001) Global sensitivity indices for nonlinear mathematical models  
561 and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, 55  
562 (1–3), 271–280. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- 563 [26] Song, S., Zhang, B., Chen, X., Xu, Q., Qu, J. (2023) Wart-Treatment Efficacy  
564 Prediction Using a CMA-ES-Based Dendritic Neuron Model. *Applied Sciences*,  
565 13(11), 6542. <https://doi.org/10.3390/app13116542>



- 566 [27] Soula, H.A., Julienne, H., Soulage, C.O., G elo en, A. (2013) Modelling  
567 adipocytes size distribution. *Journal of Theoretical Biology*, 332, 89–95.  
568 <https://doi.org/10.1016/j.jtbi.2013.04.025>
- 569 [28] Soula, H. A., G elo en, A., Soulage, C. O. (2015) Model of adipose tissue cellularity  
570 dynamics during food restriction. *Journal of theoretical biology*, 364, 189–196.  
571 <https://doi.org/10.1016/j.jtbi.2014.08.046>
- 572 [29] Toni T., Welch D., Strelkova N., Ipsen A., Stumpf M. P. (2009) Approximate  
573 Bayesian computation scheme for parameter inference and model selection in  
574 dynamical systems. *Journal of the Royal Society, Interface*, 6(31), 187–202.  
575 <https://doi.org/10.1098/rsif.2008.0172>
- 576 [30] Van der Meersch, V., Chuine, I. (2023) Estimating process-based model  
577 parameters from species distribution data using the evolutionary algo-  
578 rithm CMA-ES. *Methods in Ecology and Evolution*, 14, 1808–1820.  
579 <https://doi.org/10.1111/2041-210X.14119>
- 580 [31] Van Harmelen, V., Skurk, T., R ohrig, K., Lee, Y. M., Halbleib, M., Aprath-  
581 Husmann, I., Hauner, H. (2003) Effect of BMI and age on adipose tissue cel-  
582 lularity and differentiation capacity in women. *International journal of obesity*  
583 *and related metabolic disorders : journal of the International Association for the*  
584 *Study of Obesity*, 27(8), 889–895. <https://doi.org/10.1038/sj.ijo.0802314>
- 585 [32] <https://maple.cloud/app/6509768948056064>
- 586 [33] <https://github.com/CMA-ES/pycma>

Table 4: **Parameter estimation results on measured adipocyte radius distribution in 32 rats.** First column is the animal identification. Estimation is performed with CMA-ES algorithm of `fmin2` function from `cma` Python package by minimizing the cost function  $\mathcal{L}(\theta)$  (18). The second to fourth columns show each parameter estimated value for each rat averaged over 100 runs with different initial guesses and the standard deviations are in brackets. For each estimated parameter, considering a maximum change of 20% of its estimated value, 1,000 samples are selected with a maximal error rate of 0.1% of the cost function value. The range of selected values of each parameter is given in the next four columns. These ranges consist in values of the parameter (assuming the other are fixed) for which the maximal cost function is 0.1% of the obtained cost function with the estimation. One can note that animals *B3* and *B9* have a value of  $\theta_3$  that is estimated to be zero ( $10^{-12}/10^{-13}$ ). Indeed, these animals show particular cell size distributions with a very large number of small cells which can be due to a measurement artifact. The last column provides the cost function values normalized by the number of observed radii  $N$ ,  $\mathcal{L}_N(\theta) = \frac{1}{N}\mathcal{L}(\theta)$ , that is associated with each parameter estimation.

animal	estimated values				selected ranges				normalized cost function values
	$\theta_1 10^{-3}$ (std $10^{-11}$ )	$\rho 10^2$ (std $10^{-6}$ )	$\theta_3 10^3$ (std $10^{-4}$ )	$\theta_4 10^{-3}$ (std $10^{-10}$ )	$\theta_1 10^{-3}$	$\rho 10^2$	$\theta_3 10^3$	$\theta_4 10^{-3}$	$\mathcal{L}_N(\theta)$
C1	9.52 (3.12)	1.60 (2.94)	1.89 (2.31)	11.8 (7.95)	9.49 - 9.56	1.56 - 1.64	1.59 - 2.21	10.5 - 13.5	4.44
C2	9.19 (1.05)	1.85 (2.16)	3.92 (1.07)	7.31 (3.43)	9.17 - 9.21	1.79 - 1.90	3.70 - 4.16	6.67 - 8.10	4.26
C3	9.23 (1.33)	1.84 (2.54)	3.40 (1.50)	8.10 (4.82)	9.21 - 9.26	1.79 - 1.89	3.17 - 3.65	7.40 - 8.99	4.29
C4	9.63 (2.27)	1.19 (1.63)	1.92 (1.52)	10.8 (5.75)	9.59 - 9.68	1.15 - 1.22	1.70 - 2.16	9.88 - 11.9	4.02
C5	9.43 (1.75)	1.71 (1.99)	3.08 (1.34)	6.43 (3.25)	9.41 - 9.45	1.68 - 1.74	2.90 - 3.27	5.75 - 7.31	4.36
C6	9.34 (1.41)	1.86 (1.92)	4.04 (1.13)	4.39 (2.33)	9.32 - 9.35	1.83 - 1.89	3.89 - 4.19	3.88 - 5.08	4.40
C7	9.26 (1.02)	2.01 (1.89)	4.26 (0.91)	4.26 (2.02)	9.25 - 9.27	1.98 - 2.04	4.12 - 4.42	3.80 - 4.86	4.41
C8	9.53 (2.15)	1.47 (1.86)	3.22 (1.36)	6.09 (3.29)	9.50 - 9.55	1.45 - 1.50	3.04 - 3.41	5.42 - 7.01	4.27
C9	9.42 (1.9)	1.76 (2.25)	2.87 (1.49)	7.24 (3.65)	9.39 - 9.44	1.73 - 1.80	2.68 - 3.08	6.52 - 8.20	4.40
C10	9.37 (1.81)	1.86 (2.19)	3.95 (1.35)	3.97 (2.25)	9.36 - 9.39	1.83 - 1.88	3.81 - 4.10	3.50 - 4.58	4.38
C11	9.34 (1.69)	1.76 (2.46)	2.51 (1.81)	8.90 (5.35)	9.31 - 9.36	1.71 - 1.80	2.30 - 2.73	8.19 - 9.76	4.26
C12	9.27 (0.95)	2.00 (1.80)	4.00 (0.83)	4.37 (1.78)	9.26 - 9.29	1.96 - 2.03	3.87 - 4.14	3.98 - 4.89	4.35
A1	9.73 (1.5)	1.48 (1.04)	2.14 (0.60)	6.06 (1.77)	9.71 - 9.74	1.46 - 1.50	2.05 - 2.23	5.55 - 6.67	4.17
A2	9.93 (2.23)	1.27 (1.12)	1.75 (0.74)	7.65 (2.87)	9.90 - 9.96	1.25 - 1.29	1.65 - 1.85	6.98 - 8.50	4.16
A3	10.1 (2.76)	1.17 (1.01)	2.02 (0.79)	6.01 (2.48)	10.1 - 10.1	1.15 - 1.18	1.92 - 2.13	5.35 - 6.86	4.16
A4	9.94 (2.01)	1.27 (1.03)	1.79 (0.60)	6.74 (2.17)	9.92 - 9.97	1.25 - 1.29	1.70 - 1.88	6.18 - 7.44	4.11
A5	9.93 (1.91)	1.30 (1.05)	1.68 (0.63)	7.62 (2.58)	9.90 - 9.95	1.28 - 1.32	1.59 - 1.79	6.98 - 8.42	4.15
A6	9.68 (1.44)	1.43 (1.15)	1.85 (0.62)	9.11 (2.70)	9.65 - 9.71	1.40 - 1.46	1.73 - 1.98	8.41 - 9.97	4.18
A7	9.70 (1.81)	1.50 (1.38)	1.93 (0.75)	8.72 (3.22)	9.67 - 9.72	1.47 - 1.53	1.81 - 2.07	7.92 - 9.68	4.30
A8	9.55 (0.98)	1.72 (1.07)	2.23 (0.49)	5.53 (1.43)	9.54 - 9.57	1.69 - 1.74	2.16 - 2.30	5.20 - 5.94	4.00
A9	9.55 (1.54)	1.65 (1.66)	1.83 (0.78)	9.70 (3.63)	9.53 - 9.58	1.62 - 1.69	1.71 - 1.97	8.98 - 10.1	4.26
A10	9.70 (1.71)	1.54 (1.35)	1.3 (0.77)	11.5 (4.09)	9.67 - 9.74	1.51 - 1.58	1.17 - 1.44	10.5 - 12.6	4.28
B1	9.83 (2.16)	1.22 (1.36)	1.24 (0.94)	13.8 (6.65)	9.78 - 9.88	1.19 - 1.25	1.09 - 1.40	12.7 - 15.2	4.14
B2	10.0 (2.31)	1.22 (1.06)	1.67 (0.70)	8.42 (2.80)	9.97 - 10.0	1.20 - 1.24	1.56 - 1.79	7.61 - 9.45	4.19
B3	10.3 (1.42)	1.20 (0.74)	9.24 $10^{-16}$ (1.52 $10^{-8}$ )	10.1 (9.13)	10.3 - 10.3	1.18 - 1.22	0.76 - 1.09 $10^{-15}$	9.83 - 10.5	2.79
B4	9.44 (1.04)	1.84 (1.43)	1.85 (0.62)	9.14 (2.53)	9.41 - 9.46	1.80 - 1.88	1.74 - 1.97	8.60 - 9.81	4.14
B5	9.61 (1.16)	1.62 (1.06)	1.97 (0.61)	7.54 (2.35)	9.59 - 9.63	1.59 - 1.65	1.87 - 2.07	7.02 - 8.19	4.18
B6	9.26 (2.11)	1.59 (2.15)	1.34 (0.87)	13.7 (6.14)	9.22 - 9.31	1.52 - 1.67	1.21 - 1.49	13.0 - 14.5	3.70
B7	9.44 (0.93)	1.77 (1.2)	2.22 (0.55)	7.76 (2.44)	9.42 - 9.46	1.73 - 1.81	2.11 - 2.33	7.27 - 8.34	4.09
B8	9.73 (1.69)	1.57 (1.22)	1.28 (0.73)	9.07 (3.23)	9.71 - 9.76	1.54 - 1.59	1.19 - 1.37	8.51 - 9.72	4.08
B9	10 (1.17 $10^5$ )	1.43 (1.25 $10^5$ )	1.74 $10^{-6}$ (17.4)	11.5 (8.22 $10^4$ )	9.97 - 10.0	1.40 - 1.45	1.43 - 2.05 $10^{-6}$	11.1 - 11.9	2.97
B10	9.38 (1.52)	1.67 (1.98)	2.39 (0.81)	9.17 (3.68)	9.36 - 9.41	1.63 - 1.72	2.25 - 2.53	8.55 - 9.93	4.13

Table 5: **Parameter estimation with adipocyte size distributions measured in rats.** The first column is the parameter names. Over 32 estimations with the different animal cell size distributions, the mean is presented in the second column, the standard deviation in the third column and the fourth column is the relative standard deviation *i.e* the ratio of standard deviation over mean. The parameters are estimated with CMA-ES algorithm of `fmin2` function from `cma` Python package (with 100 initial guesses).

parameters	mean	std	RSD
$\theta_1$	$9.6 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	0.03
$\rho$	$1.57 \cdot 10^2$	$0.25 \cdot 10^2$	0.16
$\theta_3$	$2.24 \cdot 10^3$	$1.07 \cdot 10^3$	0.47
$\theta_4$	$8.21 \cdot 10^{-3}$	$2.58 \cdot 10^{-3}$	0.31