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

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

Anne-Sophie Giacobbi[1,*], Leo Meyer[2], Magali Ribot[2], Romain Yvinec[3,4], Hedi Soula[5], Chloe Audebert[1,6,**]

[1] Sorbonne Université, CNRS, Institut de Biologie Paris-Seine (IBPS), Laboratory of Computational and Quantitative Biology UMR 7238, 75005 Paris, France
[2] Institut Denis Poisson, Université d'Orléans, CNRS, Université de Tours, 45067 Orléans, France
[3] PRC, INRAE, CNRS, Université de Tours, 37380 Nouzilly, France
[4] Université Paris-Saclay, Inria, Centre Inria de Saclay, 91120 Palaiseau, France
[5] Nutriomics, La Pitié-Salpétrière, Sorbonne Université, CNRS, 75013 Paris, France
[6] 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

Abstract

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

cell size distribution and explain the differences between the model and the measurements. The proposed framework enables the characterization of adipocyte size distribution with four parameters and can be easily adapted to measurements of cell size distribution in different health conditions.
keywords: parameter estimation, adipocyte size distribution, parameter identifiability, partial differential equation

---

1 Introduction

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

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

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

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

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

2 Mathematical model for adipocyte size distributions

2.1 Model construction

Based on Soula *et al.* [27] work, we introduce a new model for adipocyte size distribution 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 ℓ and variation of radius r to adapt to lipid content are described by,

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

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

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

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

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

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

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

where v is defined by

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

The total amount of lipids λ is assumed to be constant over time and the second term of the right-hand side of (5) describes the intracellular amount of lipids at time t contained within all cells. The transport function v describes the exchange of lipids within the population of cells [27]. The lipid exchanges are based on two biochemical processes: lipogenesis – cell store lipids – and lipolysis – release of lipids in the extracellular environment. Lipogenesis depends on a surface-limited rate α, and it increases with the extracellular amount of lipids L with a saturation effect depending on the value of κ. The parameter ρ is a cell size threshold above which lipogenesis rate slows down. This parameter prevents the cell radius from becoming too large, as lipogenesis rate slows down and lipolysis rate becomes the main mechanism for lipid exchanges. Lipolysis activity includes a basal rate β and a surface-limited rate γ. 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 close to one for larger lipid content through parameter χ.

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

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

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

2.2 Stationary solution

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

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

$$\begin{cases} \partial_r f^\infty(r) = \dfrac{1}{D} v(r, L^\infty) f^\infty(r), & (10) \\ L^\infty = \lambda - \displaystyle\int_{r_{min}}^{r_{max}} (V(r) - V_{em}) \frac{4\pi r^2}{V_\ell^2} f^\infty(r) dr. & (11) \end{cases}$$

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

$$\begin{cases} (f^\infty)'(r) = \dfrac{1}{D} v(r) f^\infty(r), & (12) \\ \displaystyle\int_{r_{min}}^{r_{max}} f^\infty(r) dr = 1, & (13) \\ v(r) = \dfrac{V_\ell}{4\pi} \left(\alpha \dfrac{L}{L + \kappa} \dfrac{\rho^3}{\rho^3 + r^3} - \dfrac{(\beta + \gamma r^2)}{r^2} \dfrac{V(r) - V_{em}}{V(r) - V_{em} + V_\ell \chi} \right), & (14) \end{cases}$$

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	$\in [7.5, 150]$ (μm)	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}	$\frac{4\pi}{3}6^3$ (μm^3)	volume of an empty adipocyte (zero lipid) [1]
V_ℓ	$1.091\,10^6$ $(\mu m^3.nmol^{-1})$	volume taken by 1 $nmol$ of triglyceride [27]
α	- $(nmol.\mu m^{-2}.h^{-1})$	surface-limited rate in lipogenesis
κ	- $(nmol)$	constant of the limiting term in lipogenesis
ρ	- (μm)	cell size threshold of the Hill function in lipogenesis
β	31.25 $(nmol.h^{-1})$	basal lipolysis rate [28]
γ	0.27 $(nmol.\mu m^{-2}.h^{-1})$	surface-limited rate in lipolysis [28]
χ	- $(nmol)$	constant of the limiting term in lipolysis
D	- $(\mu m^2.h^{-1})$	diffusion coefficient for size fluctuations
λ	- $(nmol)$	total amount of lipids

where the unknown parameters to be estimated are α, L, κ, ρ, χ and D.

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

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

This solution can be computed numerically and when possible, the integrals are computed explicitly otherwise a trapezoid rule is used. Typically, in the computation, a radius step of 0.1 μm is considered and an interpolation is applied to compute f at any radius.

2.3 The model can represent a bimodal distribution of cell size

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

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

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

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

3 Model identifiability and parameter estimation

3.1 Parameter identifiability analysis

We perform an identifiability analysis of the unknown parameters of the model: α, L, κ, ρ, χ 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

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

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

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

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

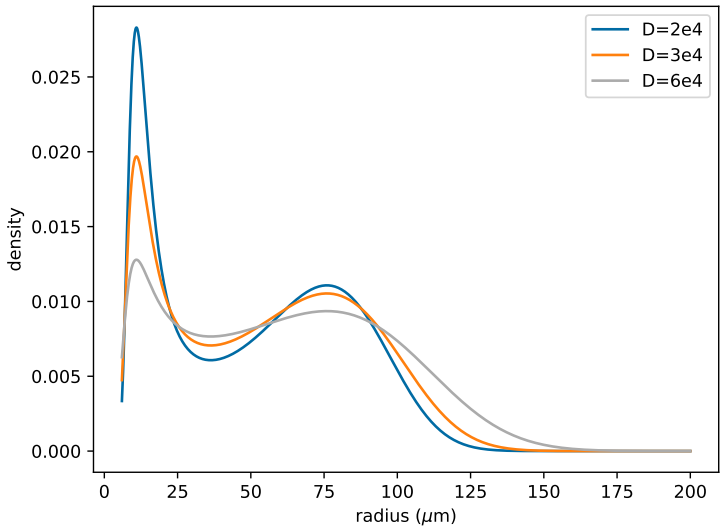


Figure 1: **Computed stationary solutions** from eq. 15 with three values for diffusion parameter. The other parameters are fixed to values reported in Table 1 and $L = 3\,nmol$, $\alpha = 0.29\,nmol.\mu m^{-2}.h^{-1}$, $\kappa = 0.001\,nmol$, $\rho = 200\,\mu m$, $\chi = 0.0035\,nmol$.

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

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

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

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

We recall that the values of V_{em}, β and γ are known (see Table 1).

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

This method ranks parameters in three categories: *globally identifiable*, *locally but not globally identifiable* and *non-identifiable*. A parameter θ_k is said to be locally identifiable if there is a finite set of possible values for θ_k given the observation. When 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_ℓ, β 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

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

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

3.2 Parameter estimation procedure

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

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

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

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

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

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

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

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

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

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

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

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

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

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

One can notice a difference between the two data sets when the estimation is performed with missing observations in the data. The last three columns of Table 2 show that, depending on the considered data set, some information on parameters is lost when cells with radii larger than a threshold are only observed. In both cases, the cost function values only slightly increase compared with the complete data sets. This indicates that the model is still able to correctly represent the data sets with missing observations. In *synthetic data set 1*, the impact on the parameter estimation is relatively small and relative errors remain below 5%. In *synthetic data set 2*, we are able to correctly estimate the values of θ_1, ρ and θ_4 but the information about parameter θ_3 seems lost, and the relative error increases to 65%.

The number of observed cells is reduced in these data sets and not in the same way in each set. On synthetic data we know exactly the percentage of information that is missing. In *synthetic data set 1* when we remove samples larger than $10\,\mu m$, 15% of the observation is missing, whereas in *synthetic data set 2* we remove 28% of the initial distribution. This difference may explain the poor estimation of θ_3 in *synthetic data set 2* with missing observations.

In the case of synthetic data sets, variations along θ_3 mainly affect the first mode of cell size distribution: increasing θ_3 strongly reduces the density of small cells and slightly increases the density of large adipocytes. Therefore, with a data set of samples with radii larger than $10\mu m$, the missing information on the first mode has an impact on the estimation of θ_3. Moreover, this parameter is related to parameter χ that drives the lipolysis mechanism in the model (size reduction). These results show that lipolysis is important for driving small cell distribution.

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

To sample a parameter θ_i, a new parameter $\bar{\theta}_i$ is first generated uniformly in $[0.8\hat{\theta}_i, 1.2\hat{\theta}_i]$ where $\hat{\theta}_i$ is the estimated parameter value obtained with the CMA-ES algorithm. Then, the cost function is computed with parameter $\bar{\theta}_i$ while the other parameters are fixed at their estimated values. The parameter is selected if the cost function is below 0.1% of $\mathcal{L}(\hat{\theta})$. This threshold was set to investigate the parameter space with small changes on cell size distribution. Note that the parameter sampling 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 ($[\theta_1 .10^2, \rho .10^{-3}, \theta_3 .10^{-4}, \theta_4 .10^2]$).

synthetic data set 1			$10,000$ 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.
θ_1	10^{-3}	9.60	9.61	$1\,10^{-8}$	0.2%	9.62	$2\,10^{-8}$	0.3%
ρ	10^{2}	1.50	1.50	$1\,10^{-8}$	0.2%	1.49	$2\,10^{-8}$	0.8%
θ_3	10^{3}	2.18	2.17	$5\,10^{-8}$	0.6%	2.09	$2\,10^{-7}$	4.2%
θ_4	10^{-3}	7.37	7.20	$2\,10^{-7}$	2.3%	7.35	$4\,10^{-7}$	0.3%
synthetic data set 2			$10,000$ 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.
θ_1	10^{-3}	9.92	9.92	$1\,10^{-8}$	0.04%	9.91	$1\,10^{-7}$	0.1%
ρ	10^{2}	2.00	2.00	$1\,10^{-8}$	0.2%	2.01	$5\,10^{-8}$	0.6%
θ_3	10^{2}	3.27	3.12	$2\,10^{-7}$	4.8%	5.39	$4\,10^{-6}$	65%
θ_4	10^{-2}	1.11	1.12	$2\,10^{-8}$	1.7%	1.12	$1\,10^{-7}$	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. ±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

synthetic data set 1			$10,000$ samples			samples $> 10\mu m$		
parameter	order	true	esti. value	esti. ±20%	select. values	esti. value	esti. ±20%	select. values
θ_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
ρ	10^{2}	1.50	1.50	1.20 - 1.80	1.47 - 1.53	1.49	1.19 - 1.79	1.46 - 1.52
θ_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
θ_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
synthetic data set 2			$10,000$ samples			samples $> 10\mu m$		
parameter	order	true	esti. value	esti. ±20%	select. values	esti. value	esti. ±20%	select. values
θ_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
ρ	10^{2}	2.00	2.00	1.60 - 2.40	1.97 - 2.03	2.01	1.61 - 2.41	1.99 - 2.05
θ_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
θ_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

per parameter.

Table 3 shows for each parameter the considered range of values and the selected intervals for each synthetic data set. For synthetic data sets without missing observations, the range of values selected by the procedure is reduced in comparison with the initial one and contains the true parameter. This analysis gives an information on the range of accepted values for each parameter. We note that, in *synthetic data set 1*, the model output seems less sensitive to parameter θ_4 that has the largest range of selected values. In *synthetic data set 2* the largest range of selected values is for parameter θ_3.

In data sets with missing observations, the selected ranges are not impacted for *synthetic data set 1* (small difference for θ_3). In *synthetic data set 2*, the loss of information about small cells leads to the selection of the total initial interval for parameter θ_3 (±20% of the estimated value) and an important increase of the selected range for θ_4 (almost twice the length) compared to the case without missing observations. As observed in section 2.3, parameter D (hence θ_4) controls the relative heights of both modes in the cell size distribution. This can explains that data sets with missing observations on small sizes lead to higher uncertainty on θ_4. These results are in agreement with the computed relative errors of the previous paragraph (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.

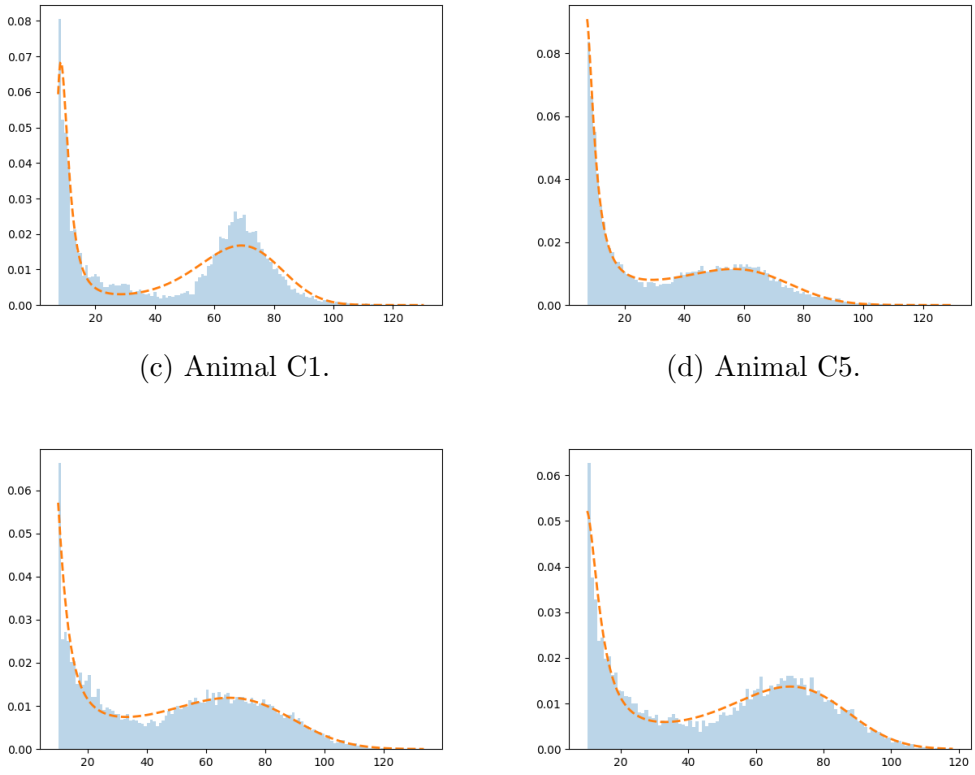
4.2 Parameter estimation with measured data

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`). 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 θ_1 and ρ, showing that the size distribution of adipocytes for rats in the

same experimental conditions can be characterized with parameters in the same value ranges. The variability in the population is larger for parameter θ_3 and θ_4 (larger RSD). However, the previous analysis on synthetic data showed that less confidence in the estimation is expected for these parameters, especially θ_3.

(a) Animal A1. (b) Animal B1.



(c) Animal C1. (d) Animal C5.

Figure 2: **Comparison model-data**. Four examples (over 32) of adipocyte radius distributions (in μm) 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.

For each animal, accepted parameter ranges are also computed following the procedure described in section 3.2 (Table 4). Figure 3 displays for each parameter the estimated value for each animal with the range of selected values (dots and bars). As expected, the parameter ranges are larger for parameters θ_3 and θ_4 compared to parameters θ_1 and ρ. Figure 3 also shows the mean (dash red line) and the standard

deviation (gray area) over the rat population for each estimated parameter. It enables to compare the amplitude of the range of accepted values for each parameter for each animal with the variability within the population. We can see that for each parameter the range of accepted values is always smaller than the standard deviation in the population. It shows that the largest standard deviation within the population obtained for θ_3 and θ_4 (Table 5) should not be attributed to less confidence in the estimations.

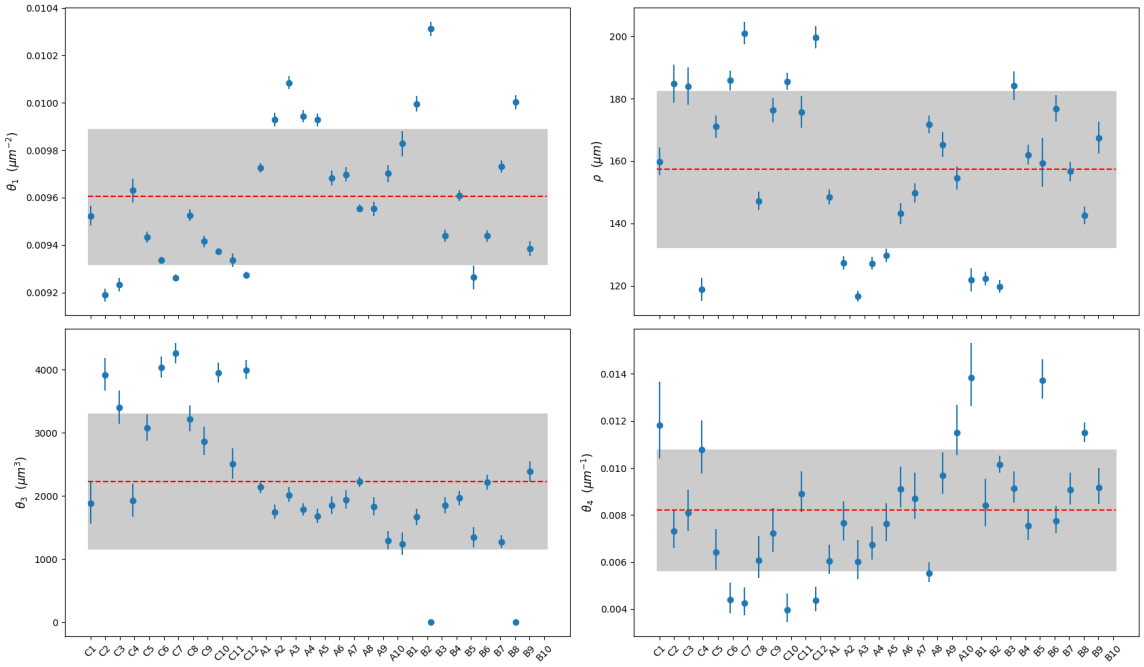


Figure 3: **Group variability and range of selected values.** Upper left and right figures display the results for parameters θ_1 and ρ. Lower left and right figures show the results for parameters θ_3 and θ_4. For each estimated parameter the average over the population is shown with dash red line and 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.

The range of selected values of parameter θ_3 in rats population is between 1070 and 4429 μm^3. From this range of values, we can compute a range of radii for which the lipolysis term becomes mainly a surface based mechanism (i.e. $(V(r) - V_{em})/(V(r) -$

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

Similarly, for lipogenesis, the parameter θ_1 is estimated within the rats population between 0.0092 and $0.010 \mu m^{-2}$. We remind that this quantity is a combination of parameters : $\theta_1 = \frac{\alpha L}{\beta(L+\kappa)}$ and parameter β is known [28]. We then obtain an estimation of $\frac{\alpha L}{(L+\kappa)}$ between 0.29 and $0.31\,nmol.\mu m^{-2}.h^{-1}$. In the case of high available lipids, L is large and we can assume $\frac{L}{L+\kappa} \sim 1$. Under this assumption, the parameter α is estimated between 0.29 and $0.31\,nmol.\mu m^{-2}.h^{-1}$. An alternative case is for low L, then we can assume $\frac{L}{L+\kappa} \sim L$ and the estimated values of θ_1 provide an estimation for αL.

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

4.3 Sensitivity analysis

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

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

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

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

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

Then, Sobol' indices are computed to determine which parameters are most influential on the cell size dynamic. The first-order indices are displayed for several radii and each parameter in Figure 4(b). The results indicate that parameter θ_1 explains the most the variations of cell sizes with a first-order sensitivity index between 0.6 and 1 for all radii. Interestingly, for the cells with radii around 40μm, the index of θ_1 decreases and we notice that ρ index increases (index equals 0.36). It shows that parameter ρ around this point explains the variability of the model output up to 36%. The impacts of θ_3 and θ_4 are almost negligible on cells size distribution. From $r = 90\mu m$, the results show that the influence of θ_1 decreases whereas ρ becomes more influential and explains up to 18% of the output variability. The total-order sensitivity indices are also computed (not shown) and are similar to first-order indices, revealing that parameter interactions have a negligible influence on the adipocyte size distributions.

The sensitivity analysis suggests that the cell size dynamics in rats is mainly driven by the parameters depending on lipogenesis, and especially by θ_1 which represents the combination of the unknown parameters (α, κ, L).

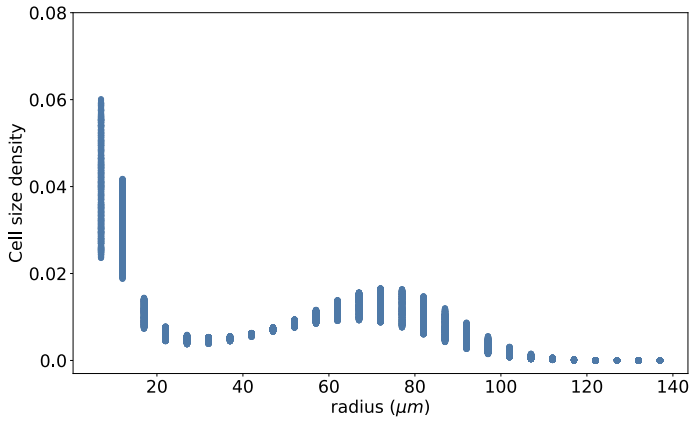
Parameters θ_3 and θ_4, associated with lipolysis (through χ) and diffusion (D) respectively, have a negligible impact on the cell size dynamic along all cell sizes. This result confirms the difficulty to identify these parameters in practice and are in agreement with the largest ranges of selected parameter values. In addition, this study highlights the fact that the nadir is difficult to capture since we observe an opposite change in the parameter sensitivity around this radius. With this study we are able to explain the results of parameter estimation on the measured data.

5 Discussion

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

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

(a) Cell size distributions



(b) Estimation of first-order Sobol' indices

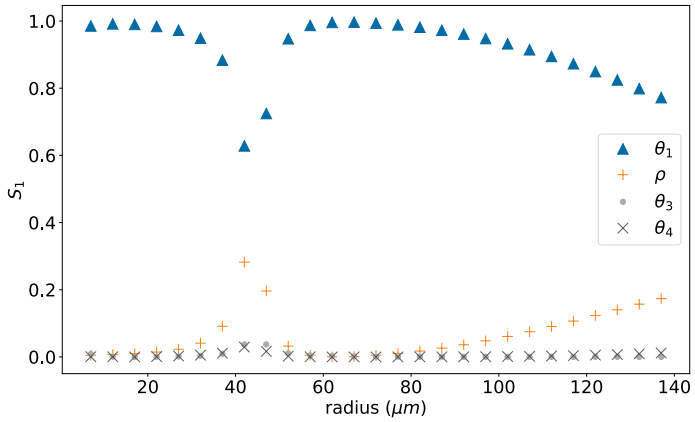


Figure 4: (a) A sample of cell size distributions. The parameter sampling design is constructed using Saltelli algorithm where each parameter is uniformly distributed in a range of values corresponding to $\pm 1\%$ of the mean of its estimated value in rats (Table 5). A number of $20,480$ samples giving bimodal distributions are generated to estimate the Sobol' indices. (b) Estimation of first-order Sobol' indices for θ_1, ρ, θ_3 and θ_4 using a Monte-Carlo based approach [25, 22, 3, 21].

threshold χ as well as the diffusion coefficient D that describes cell size fluctuations.

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

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

captured. However, the nadir part of the distribution as well as the height of the modes were not perfectly reproduced. It is possible that the model is missing some aspect of the adipocyte size dynamics that would help to better capture the nadir. This is supported by the sensitivity analysis, that showed that the nadir part was not sensitive specifically to one of the four considered parameters. Therefore, it makes this part of the distribution difficult to fit. In addition, in the presented model, the diffusion parameter D via θ_4 affects linearly both lipogenesis and lipolysis. It would be interesting to change this modeling assumption with a more complex diffusion process, impacting differently lipogenesis and lipolysis. For instance, considering a size dependent diffusion coefficient could improve the agreement between the model outputs and the observations.

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

Nevertheless, we have estimated parameter values for 32 rats. We found a larger variability between rats in the estimated values of θ_3 and θ_4 (Figure 3). However, the sensitivity analysis showed that the model is less sensitive to these parameters (Figure 4). For θ_1 and ρ, the estimated values were more robust within the population leading us to believe that θ_1, ρ are less individual-specific parameters. However they could change if the estimation is performed with another species. This result suggests lipolysis (driven by χ) is more an individual-dependent process than lipogenesis (driven by θ_1 and ρ) that is more constant within the population.

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

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

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

6 Acknowledgments

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

References

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002) Molecular biology of the cell, fourth edition, 197.

[2] Arner, E., Westermark, P. O., Spalding, K. L., Britton, T., Rydén, M., Frisén, J., Bernard, S., Arner, P. (2010) Adipocyte turnover: relevance to human adipose tissue morphology. Diabetes, 59(1), 105–109. https://doi.org/10.2337/db09-0942

[3] Campolongo, F., Saltelli, A., Cariboni, J. (2011) From screening to quantitative sensitivity analysis. A unified approach. Computer Physics Communications 182, 978–988. https://doi.org/10.1016/j.cpc.2010.12.039.

[4] Catchpole, E. A., MORGAN, B. J. T. (1997) Detecting parameter redundancy. Biometrika, 84(1), 187–196. https://doi.org/10.1093/biomet/84.1.187

[5] Clément, K. (2011) Bariatric surgery, adipose tissue and gut microbiota. Int J Obes 35 (Suppl 3), S7–S15. https://doi.org/10.1038/ijo.2011.141

[6] Cole, D.J., Morgan, B.J., Titterington, D.M. (2010) Determining the parametric structure of models. Math Biosci, 228(1), 16–30. https://doi.org/10.1016/j.mbs.2010.08.004

[7] Drolet, R., Richard, C., Sniderman, A.D., Mailloux, J., Fortier, M., Huot, C., Rhéaume, C., Tchernof, A. (2008) Hypertrophy and hyperplasia of abdominal adipose tissues in women. Int. J. Obes. (London) 32 (2), 283?291. http://dx.doi.org/10.1038/sj.ijo.0803708

[8] Hansen, N. (2016) The CMA Evolution Strategy : A Tutorial. https://doi.org/10.48550/arXiv.1604.00772

[9] Hesse, J., Müller, T., Relógio, A. (2023) An integrative mathematical model for timing treatment toxicity and Zeitgeber impact in colorectal cancer cells. npj Syst Biol Appl 9, 27. https://doi.org/10.1038/s41540-023-00287-4

[10] Hong H., Ovchinnikov A., Pogudin G., Yap, C. (2019) SIAN: software for structural identifiability analysis of ODE models. Bioinformatics, 35(16), 2873–2874. https://doi.org/10.1093/bioinformatics/bty1069

[11] Hong H., Ovchinnikov A., Pogudin G., Yap, C. (2020) Global identifiability of differential models. Communications on pure and applied mathematics, 73(9), 1831–1879. https://doi.org/10.1002/cpa.21921

[12] Jacquier M., Crauste F., Soulage C.O., Soula H.A. (2014) A Predictive Model of the Dynamics of Body Weight and Food Intake in Rats Submitted to Caloric Restrictions. PLOS ONE 9(6): e100073. https://doi.org/10.1371/journal.pone.0100073

[13] Jo, J., Gavrilova, O., Pack, S., Jou, W., Mullen, S., et al. (2009) Hypertrophy and/or Hyperplasia: Dynamics of Adipose Tissue Growth. PLOS Computational Biology 5(3): e1000324. https://doi.org/10.1371/journal.pcbi.1000324

[14] Jo, J., Shreif, Z., Periwal, V. (2012) Quantitative dynamics of adipose cells. Adipocyte, 1(2), 80–88. https://doi.org/10.4161/adip.19705

[15] Jo, J., Shreif, Z., Gaillard, J.R., Arroyo, M., Cushman, S.W., Periwal, V. (2013) Mathematical Models of Adipose Tissue Dynamics. In: Gefen A., Benayahu D. (eds) The Mechanobiology of Obesity and Related Diseases. Studies in Mechanobiology, Tissue Engineering and Biomaterials, 16, Springer, Cham. https://doi.org/10.1007/8415_2013_170

[16] Laforest, S., Michaud, A., Paris, G., Pelletier, M., Vidal, H., Géloën, A., Tchernof, A. (2017) Comparative analysis of three human adipocyte size measurement methods and their relevance for cardiometabolic risk. Obesity, 25, 122–131. https://doi.org/10.1002/oby.21697

[17] Li, Y., Periwal, V., Cushman, S. W., Stenkula, K. G. (2015) Adipose cell hypertrophy precedes the appearance of small adipocytes by 3 days in C57BL/6 mouse upon changing to a high fat diet. Adipocyte, 5(1), 81–87. https://doi.org/10.1080/21623945.2015.1128588

[18] Lönn, M., Mehlig, K., Bengtsson, C., Lissner, L. (2010) Adipocyte size predicts incidence of type 2 diabetes in women. FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 24(1), 326–331. https://doi.org/10.1096/fj.09-133058

[19] MacKellar, J., Cushman, S.W., Periwal, V. (2010) Waves of Adipose Tissue Growth in the Genetically Obese Zucker Fatty Rat. PLoS ONE 5(1): e8197. https://doi.org/10.1371/journal.pone.0008197

[20] McLaughlin, T., Sherman, A., Tsao, P., Gonzalez, O., Yee, G., Lamendola, C., Reaven, G. M., Cushman, S. W. (2007) Enhanced proportion of small adipose cells in insulin-resistant vs insulin-sensitive obese individuals implicates impaired adipogenesis. Diabetologia, 50(8), 1707–1715. https://doi.org/10.1007/s00125-007-0708-y

[21] Owen, A. B. (2020) On dropping the first Sobol' point. arXiv:2008.08051 [cs, math, stat]. http://arxiv.org/abs/2008.08051.

[22] Saltelli, A. (2002) Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications, 145(2), 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1.

[23] Schlichting, A. (2019) The Exchange-Driven Growth Model: Basic Properties and Longtime Behavior. Journal of Nonlinear Science, 30, 793–830. https://doi.org/10.1007/s00332-019-09592-x

[24] Skurk, T., Alberti-Huber, C., Herder, C., Hauner, H. (2007) Relationship between adipocyte size and adipokine expression and secretion. The Journal of clinical endocrinology and metabolism, 92(3), 1023–1033. https://doi.org/10.1210/jc.2006-1055.

[25] Sobol' I.M. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55 (1–3), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.

[26] Song, S., Zhang, B., Chen, X., Xu, Q., Qu, J. (2023) Wart-Treatment Efficacy Prediction Using a CMA-ES-Based Dendritic Neuron Model. Applied Sciences, 13(11), 6542. https://doi.org/10.3390/app13116542

[27] Soula, H.A., Julienne, H., Soulage, C.O., Géloën, A. (2013) Modelling adipocytes size distribution. Journal of Theoretical Biology, 332, 89–95. https://doi.org/10.1016/j.jtbi.2013.04.025

[28] Soula, H. A., Géloën, A., Soulage, C. O. (2015) Model of adipose tissue cellularity dynamics during food restriction. Journal of theoretical biology, 364, 189–196. https://doi.org/10.1016/j.jtbi.2014.08.046

[29] Toni T., Welch D., Strelkowa N., Ipsen A., Stumpf M. P. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society, Interface, 6(31), 187–202. https://doi.org/10.1098/rsif.2008.0172

[30] Van der Meersch, V., Chuine, I. (2023) Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES. Methods in Ecology and Evolution, 14, 1808–1820. https://doi.org/10.1111/2041-210X.14119

[31] Van Harmelen, V., Skurk, T., Röhrig, K., Lee, Y. M., Halbleib, M., Aprath-Husmann, I., Hauner, H. (2003) Effect of BMI and age on adipose tissue cellularity and differentiation capacity in women. International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity, 27(8), 889–895. https://doi.org/10.1038/sj.ijo.0802314

[32] `https://maple.cloud/app/6509768948056064`

[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 θ_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.

	estimated values				selected ranges				normalized cost function values
animal	$\theta_1\,10^{-3}$ (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
θ_1	$9.6\,10^{-3}$	$2.8\,10^{-4}$	0.03
ρ	$1.57\,10^{2}$	$0.25\,10^{2}$	0.16
θ_3	$2.24\,10^{3}$	$1.07\,10^{3}$	0.47
θ_4	$8.21\,10^{-3}$	$2.58\,10^{-3}$	0.31