



HAL
open science

Bayesian inference for parameter identification in mechanistic models, exemplified using a cow lifetime performance model

J.F. Ramirez-Agudelo, J.B. Daniel, Laurence Puillet, Nicolas N.C. Friggens

► **To cite this version:**

J.F. Ramirez-Agudelo, J.B. Daniel, Laurence Puillet, Nicolas N.C. Friggens. Bayesian inference for parameter identification in mechanistic models, exemplified using a cow lifetime performance model. *Animal - Open Space*, 2023, 2, pp.100054. 10.1016/j.anopes.2023.100054 . hal-04392783

HAL Id: hal-04392783

<https://hal.science/hal-04392783>

Submitted on 14 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.



Method paper

Bayesian inference for parameter identification in mechanistic models, exemplified using a cow lifetime performance model



J.F. Ramirez-Agudelo ^a, J.B. Daniel ^b, L. Puillet ^a, N.C. Friggens ^{a,*}

^a Université Paris-Saclay, INRAE, AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 91120 Palaiseau, France

^b Trouw Nutrition R&D, PO Box 299, 3800 AG Amersfoort, the Netherlands

ARTICLE INFO

Article history:

Received 9 December 2022

Revised 9 November 2023

Accepted 13 November 2023

Handling editor: J.P. Sanchez

Keywords:

Delayed Rejection Adaptive Metropolis

Differential Evolution

Genetic Algorithm

Particle Swarm Optimisation

Simulated Annealing

ABSTRACT

Mechanistic models are valuable tools for studying the underlying mechanisms of complex biological phenomena. For example, cow lifespan models can be used to identify differences in resource acquisition and allocation strategies between individuals, which is relevant for decision-making in breeding programs. In such models, differences in simulated traits between individuals are consequences of the parameter set that represents the genetic potential of each animal and its interaction with the environment. This indicates that the identification of these differences is essentially a search for individual parameters. In mechanistic models, this search is generally a non-convex problem that has different local minima because the parameters interact within these models. Due to this and to the simulation time length (e.g. years), there is uncertainty associated with the inference of the parameter values for each individual. This uncertainty can be quantified using Bayesian inference since this approach treats the model parameters as random variables with an underlying probability distribution that describes them. The objective of this work was to employ the Delayed Rejection Adaptive Metropolis (**DRAM**) algorithm to identify the parameters of a cows' lifespan model using two datasets of Holstein cows. The datasets contain periodic measurements of Milk Yield (**MY**), BW, and Body Condition Score (**BCS**). Additionally, one of the two datasets has information of BW from birth to first calving. The average Mean Absolute Percentage Error (**MAPE**) minimisation between the simulated and experimental data (MY, BW and BCS) was used as the objective function for parameter search. The Bayesian inference performance was compared with four optimisation metaheuristic approaches: Differential Evolution, Genetic Algorithm, Particle Swarm Optimisation, and Simulated Annealing. Although the results show that all methods are efficient in finding parameter values that reduce the distance between the simulated and experimental data (MAPE < 10%), the DRAM method is more efficient in terms of computational cost, and the parameter distributions obtained with this method offer more information about the statistical properties of each parameter (e.g. median).

© 2023 The Authors. Published by Elsevier B.V. on behalf of The Animal Consortium. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Reader comments

We invite you to comment on the article on the PubPeer platform by clicking on this link [discuss this article](#).

Implications

The growing demand of milk and meat makes it necessary to breed efficient animals that are resistant to harsh conditions and

that recover quickly from adverse events. The identification of this type of animal can be done by using mechanistic models in which productive life is defined by parameters that represent the individual genetic potential and its interaction with the environment. To use models in breeding programs, it is necessary to identify the best set of parameters for each individual. Therefore, in this model-based approach, the methodology with which these parameters are identified is important. This work explores different methods for this task.

* Corresponding author.

E-mail address: nicolas.friggens@agroparitech.fr (N.C. Friggens).

Specification table

Subject	Breeding and Genetics <i>*This paper does not address genetic issues directly.</i>
Type of data	Table, Graph
How data were acquired	Recorded at milking time, weighings at the same time of day using standard commercial equipment, and visual inspection for BCS.
Data format	Raw data
Parameters for data collection	Milk yield and milk composition were recorded twice daily. All animals were weighed at the same time of day. The same person was responsible for the BCS estimation.
Description of data collection	In both datasets, cows were milked twice daily between 0600 and 0800 h and between 1600 and 1800 h. Milk yield and milk composition were recorded at each milking: almost daily for Dataset1 and weekly for Dataset2. In Dataset1, cows were weighed once a week during the dry period and until 3 months after calving, and every 2 weeks from 3 months after calving to the dry period. In Dataset2, cows were weighed at least twice a week. Body condition was scored from 1 to 5 on the same days of weighing in both datasets.
Data source location	Dataset1 Institution: Danish Cattle Breeders Organisation. City/Town/Region: Ammitsbøl, Skovgard Country: Denmark Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: Lat: 55.650014, Long: 9.410401. Dugårdvej 31, 7100 Vejle, Denmark Dataset2 Institution: Trouw Nutrition R&D City/Town/Region: Amersfoort Country: The Netherlands Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: Lat: 51.64377, Long: 5.92376. Sint Anthonisweg 65, 5831 AD Boxmeer, the Netherlands
Data accessibility	Code repository name: Zenodo Code identification number: https://doi.org/10.5281/zenodo.7998116 Data repository name: Data INRAE Data identification number: https://doi.org/10.57745/O4XABU
Related research article	Friggens, N.C., Berg, P., Theilgaard, P., Korsgaard, I.R., Ingvarsen, K.L., Løvendahl, P., Jensen, J., 2007. Breed and parity effects on energy balance profiles through lactation: evidence of genetically driven body energy change. Journal of Dairy Science, 90, 5291–5305.

Introduction

In breeding programs, most of the emphasis has been made on traits directly associated with productivity, such as milk yield or daily weight gain. However, the biological mechanisms underlying productivity are difficult to be taken into account for selection because the direct measurements of these mechanisms are difficult or even impossible. To partially overcome these limitations, it is possible to use mechanistic models which allow estimation of inaccessible traits. It implies such models are able to identify parameters that define biological traits. Puillet et al. (2016) developed a dynamic-mechanistic model (AQAL) to generate genetically driven lifetime trajectories of energy acquisition and allocation in dairy cows, using genetic-scaling parameters. For a given individual, the simulated performance is the result of the combination of a capacity to acquire resources and how these acquired resources are allocated to different life functions. AQAL was used to evaluate the extent of genotype by environment interaction for feed efficiency traits (Puillet et al., 2021), demonstrating the usefulness of mechanistic modelling in breeding strategie evaluation.

To use a model like AQAL in breeding programs, it is necessary to have an accurate methodology for parameter search. In mechanistic models, this search is generally a non-convex problem that has different local optimum because parameters interact within the model (e.g. AQAL), which generates uncertainty in the estimation of parameters. Additionally to the model architecture (Muñoz-Tamayo et al., 2018), this uncertainty is also associated with the noise in the experimental data used during identification (Sherri et al., 2019); making difficult to determine, a priori, the global optimum of the objective function.

Several global methods based on function evaluation have been developed for optimisation problems. The majority are analogies of efficient and simple natural phenomena. Among them, the most efficient techniques are population-based algorithms such as Particle Swarm (Kennedy and Eberhart, 1995), Genetic Algorithms (Fogel, 1994), or Simulated Annealing (Ingber, 1993). However, there is no theoretical guarantee that these algorithms converge to the global optimum. Alternatively, Bayesian methods have good theoretical convergence results, providing information about the parameter distributions.

Bayesian Inference (BI) can be used to quantify the uncertainty in the parameter values, because the parameters are treated as random variables with an underlying probability distribution that describes them (Miles and Smith, 2019). Among them, the DRAM method (Haario et al., 2006), a modified version of the Metropolis-Hastings algorithm, is an efficient BI method for parameter identification. The adaptive Metropolis (AM) component in DRAM means that each parameter is generated following the distribution calculated with all of the previous states in the iterative process. The delayed rejection (DR) mechanism is used to stimulate mixing by sampling from a narrower proposed distribution when a parameter sample is rejected (Haario et al., 2006).

This work presents a protocol to use BI for the identification of resource acquisition and resource allocation parameters using the AQAL model (Puillet et al., 2016) for dairy cows, taking into account the aforementioned uncertainty. The performance of BI was compared with the performance of four other optimisation methods: Differential Evolution (DE), Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), and Simulated Annealing (SA).

Material and methods

The objective of optimisation is to explore the parameter space to identify the global optimum of the objective function, which in

our case is to reduce the distance between the simulated and experimental data. Three steps are required to do this: a) preprocessing of experimental data to fit with the specific model being used (in this case AQAL), b) formulation of the objective function, and c) parameter optimisation. The first two-steps are model specific.

Preprocessing of experimental data

The preprocessing steps may differ depending on the model and the structure of the database being used. For calibrating AQAL, experimental data are necessary to run the model and to compute the error. To run, AQAL uses a sequence of values (e.g. [0.75, 0.83, 0.90, 1, 1, 1]) which represents the proportion of the maximum level of resource acquisition reached in each lactation. In the process of calculating this sequence, milk production should be used.

First, it is important to find the average of the upper MY values in each lactation rather than just using the maximum value. This approach helps mitigate the influence of potential noise or outliers in the data. Once the lactation with maximum production is determined, the next step is to calculate the maximum resource acquisition ratio for each lactation. This is achieved by dividing the MY (upper average) of each lactation by the maximum MY observed in all lactations. The details to carry out this calculation can be reviewed in the calibration script that accompanies this work.

To compute the error, at least three traits must be used: MY (Kcal of Metabolisable Energy per day or Kg/d), BW (Kg), and BCS (1–5 scale). Given that data are commonly recorded with a date, it becomes essential to have the animal’s date of birth in order to determine the age at which each variable was measured. This information is crucial for conducting a meaningful comparison between simulated data and the actual observations recorded on those specific dates. Data about parturition and drying-off, or culling, are also required. The accompanying datasets for this study include columns that present the age of the animals, parturition number and days in milk.

Formulation of the objective function

Parameter inference is an iterative process in which an algorithm evaluates candidates until obtaining a parameter set that most reduces the distance between experimental and simulated data (i.e. objective function). For this, the algorithm needs a metric that represents these distances. We used the Mean Absolute Percentage Error (**MAPE**) to quantify this mismatch because, unlike other metrics like Mean Squared Error, the MAPE metric avoids the effect of traits’ units and scales on the average error calculation. To calculate the MAPE, the following general equation was used:

$$MAPE_{(i)} = \frac{ABS(Experimental_{(i)} - Simulated_{(i)})}{Experimental_{(i)}} * 100.$$

Due to the fact that databases can present imbalances in the data record (e.g. more data in one lactation than in others), it is recommended to calculate an average MAPE for each variable for each lactation. For example, for a cow with four lactations, it will be necessary to calculate 12 MAPEs (three variables × four lactations). These can then be reduced to 1 value by weighting the different MAPEs according to their relative importance. In the present case, we choose equal weights, i.e. the objective function returns the average of those 12 MAPEs.

The algorithm for MAPE estimation is summarised as follows:

1. Generate a set of parameters according to each method.
2. Run the model using the candidate parameters.
3. For each lactation

- 3.1. For each variable (MY, BW, BCS)
 - 3.1.1. Calculate the average MAPE
4. Return the average of the MAPEs.
5. Repeat all the steps until the number of samples required is achieved. It is generally considered that there should be thousands of runs for these kinds of parameter search methods.

Parameter optimisation

The parameter optimisation step involves determining the optimal values for the parameters of the AQAL model or any other predictive model. In this step, recorded traits, which are the observed or measured trait values, and model-predicted trait values are used as the input. The goal is to find the set of parameters that minimises the discrepancy or error between the recorded traits and the model-predicted values. This discrepancy is quantified using a defined function that measures the difference between the observed traits and the predicted values (e.g. MAPE). The optimisation algorithm then iteratively adjusts the parameter values to minimise the discrepancy, ultimately finding the parameter values that best fit the data. The codes used for the parameter optimisation process were deposited in the code repository and properly documented, ensuring transparency and reproducibility for future use.

This section refers to the first step of the MAPE estimation algorithm which involves generating a set of parameters using different methods. Each method employs a unique strategy for parameter generation. Specifically, DRAM generates a sequence of random parameters from a Markov chain whose distributions approach the posterior distributions of each parameter. In DE and GA, the algorithm begins by randomly initialising the population within a certain search space, and then, these initial values are evolved over generations in order to optimise the objective function; this is achieved using genetic operators such as mutation, selection, and crossover. In PSO, the search is based on the movement of individual particles through the search space, which mimics the swarm of birds or insects. SA is based on metallurgical practices whereby a metal is heated to a high temperature and then cooled to obtain a more purified material. According to Piotrowski et al. (2020), we set the population to 100 in DE, GA, PSO and SA methods. The mutation probability was set at 0.01 for GA. In PSO, the parameters w, c1, and c2 were set to 0.8, 0.5, and 0.5, respectively. The maximum and minimum temperature in SA was set to 1 and 1e–9, respectively. Table 1 shows the minimum and maximum values for parameters search in DE, GA, PSO and SA methods, and the values for DRAM initialisation. In Python, we used the ‘pymcmcstat’ library for DRAM and the ‘scikit-opt’ library for DE, GA, PSO, and SA.

Table 1
Minimum, maximum, and DRAM initialisation values for parameter search in the dairy cows AQAL model (Puillet et al., 2016).

Parameters	Min.	Max.	DRAM initialisation
acq_bas_pot	5	20	8.5
acq_bas_lac	0	30	15
init_G	0.2	0.5	0.35
s2pc_pot	0.6	1.1	0.9
k_acq_bas_mat	0.005	0.012	0.008
k_Pc_on_g2s	0.008	0.02	0.015

The parameters “acq_bas_pot” and “acq_bas_lac” are the basal and an additional lactation acquisition (Kg/d), respectively. “k_acq_bas_mat” is the rate of basal acquisition maturation. “init_G” and “s2pc_pot” control energy allocation for growth and lactation, respectively. “k_Pc_on_g2s” relaunches growth after calving. For details, see Puillet et al. (2016). AQAL = dynamic-mechanistic model of lifetime trajectories of energy acquisition and allocation in dairy cows; DRAM = Delayed Rejection Adaptive Metropolis.

Datasets

Two datasets from Holstein cows (18 and 25 animals) were used to test the performance of the five methods (DRAM, DE, GA, PSO, and SA). Both datasets report periodic measures of MY, BW, and BCS at least three lactations per cow. Additionally, one of the two datasets has information of BW from birth to first calving. The data used in this study are freely accessible in the data repository.

Although the size of the datasets in this study may be smaller compared to those in studies targeting direct biological research objectives, it is important to acknowledge that the datasets utilised are sufficient for the purpose of evaluating and comparing optimisation methods. The primary focus of this study is to demonstrate the effectiveness and advantages of these methods, rather than drawing broad biological conclusions. Therefore, considering this objective, the size of the datasets employed is appropriate and enables a meaningful evaluation of the proposed optimisation methods.

Results

The model used in this study (AQAL, Puillet et al., 2016) is designed to prematurely end the simulation (death of the virtual cow). The probability of death depends on extreme values of body reserve proportion. This model characteristic means that, during calibration, some iterations do not return an optimised error value because the chosen parameter set results in an early end of the simulation, i.e. before reaching the last data value in the observed data. In these cases, which are typified by MAPE values more than 10 times greater than for completed simulations, these parameter sets were not included in the registry of accepted parameters per method. This means that the number of parameter sets is always lower than the number of iterations used to obtain these sets. The parameter search was stopped for each animal when at least 2 000 sets of valid iterations were obtained, i.e. excluding parameters estimated in incomplete simulations.

Iterative processes

Sampling strategies differ significantly between methods. Fig. 1 shows, except for DE, that as the iterative process progresses, the generation of samples is focused on a certain region of the parameter space, which in this case was from 5 to 20 (Kg DM/d) for the acquisition parameter “acq_bas_pot”. With the exception of DRAM, in the early iterative process, the methods explore all the parameter space proposed. This behaviour is strongly maintained during the entire search in DE. The DRAM subplot shows that the early search is affected by the initial value of the parameter; however, after a few hundred iterations, this initial effect has been lost and the algorithm explores the parameter space where it is expected to find the global minimum of the response function.

Given the iterative behaviours presented in Fig. 1, it is expected that independent runs of GA, PSO and SA converge to the same parameter value. However, stochasticity results in different parameter values for independent runs of these methods. Fig. 2 shows the smoothed trajectories of three runs of the search for the parameter “acq_bas_pot” with the five methods. The Savitzky Golay filter (Schafer, 2011) was used to smooth all iterative processes, with a third-degree polynomial function and a window size of 100 points.

Methods performance

As presented in Table 2, all methods were able to explore the proposed parameter space and find sets of parameters that reduce the distances between the observed and predicted values of MY, BW, and BCS (MAPE < 10%). However, in terms of computational cost, the DRAM method is the most efficient for this task. Table 2 shows that the minimum error can be found with this method with fewer iterations compared to the other methods.

The quality of the adjusted between the parameters inferred with the DRAM method and the other methods (DE, GA, PSO and SA) was evaluated using the concordance correlation coefficient (CCC, Lawrence and Lin, 1989), and the root mean square prediction error (RMSPE). The results in Table 3 show that there is no

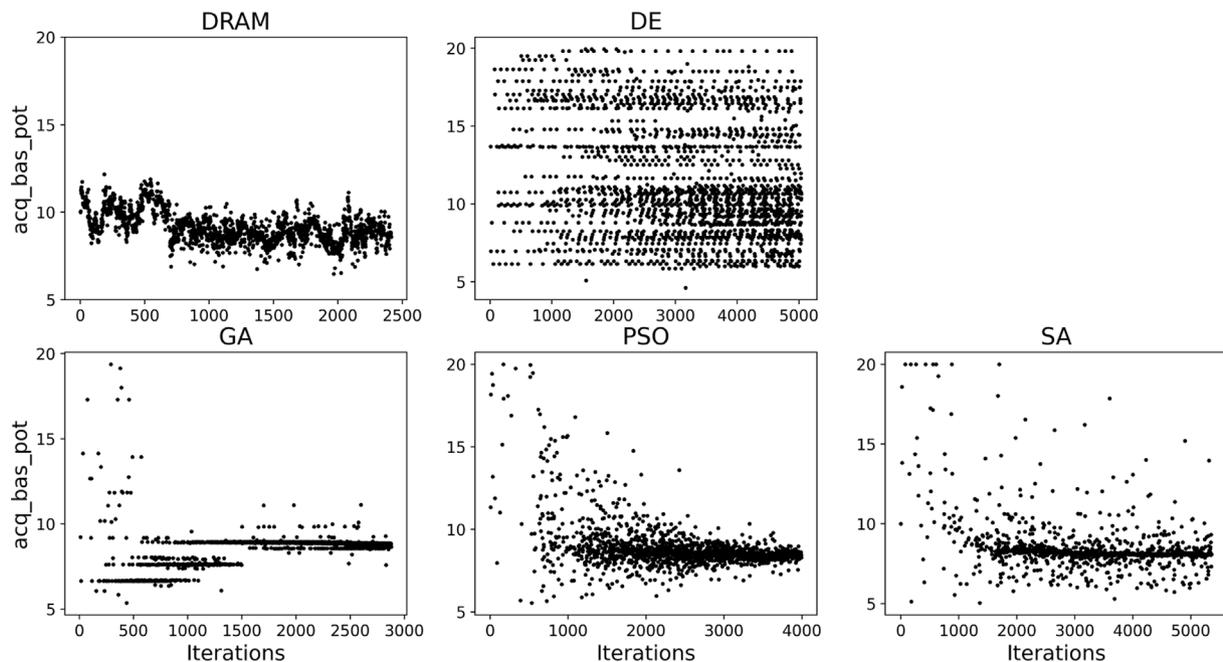


Fig. 1. Iterative processes for identification of the parameter “acq_bas_pot” (basal acquisition, Kg DM/d, in the dairy cows AQAL model, Puillet et al., 2016), using Delayed Rejection Adaptive Metropolis (DRAM), Differential Evolution (DE), Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Simulated Annealing (SA). The points on the graphs are the candidates' values generated for each method during the iterative process.

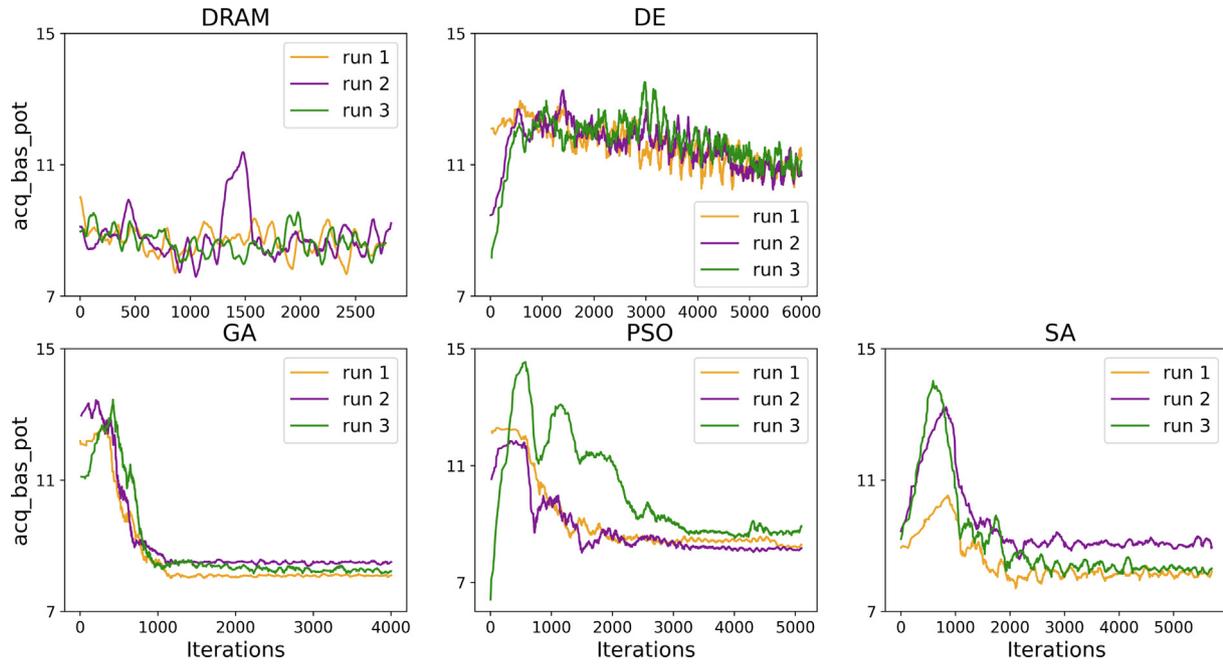


Fig. 2. Smoothed iterative processes of independent runs for the identification of the “acq_bas_pot” parameter (basal acquisition, Kg DM/d, in the dairy cows AQAL model, Puillet et al., 2016), using Delayed Rejection Adaptive Metropolis (DRAM), Differential Evolution (DE), Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Simulated Annealing (SA).

Table 2
Summary of methods performance for parameter search in the dairy cows AQAL model (Puillet et al., 2016).

Method	Minimum MAPE		Iterations to achieve the min MAPE	
	Dataset1 ¹	Dataset2 ¹	Dataset1 ¹	Dataset2 ¹
DRAM	5.9 (1.2)	6.2 (1.5)	1 215 (630)	1 111 (806)
GA	7.0 (1.7)	9.2 (2.8)	2 560 (540)	2 352 (1 097)
DE	8.5 (1.8)	8.0 (1.7)	4 023 (1 107)	4 046 (1 720)
PSO	5.7 (1.0)	6.5 (1.8)	4 176 (269)	4 328 (553)
SA	5.2 (0.9)	6.3 (1.9)	5 261 (281)	4 721 (585)

AQAL = dynamic-mechanistic model of lifetime trajectories of energy acquisition and allocation in dairy cows; DRAM = Delayed Rejection Adaptive Metropolis; GA = Genetic Algorithm; DE = Differential Evolution; PSO = Particle Swarm Optimisation; SA = Simulated Annealing; MAPE = Mean Absolute Percentage Error.

¹ Mean (SD).

good fit between the parameters estimated by the DRAM method and those estimated with the other methods. The 81.3% of the CCC values were below 0.7.

It is important to note that different optimisation methods may not necessarily converge to the exact same values. This variability in parameter estimation is a common occurrence and can be attributed to the inherent differences in how each optimisation algorithm explores the parameter space and finds optimal solutions. We employed various optimisation methods to compare its performance with that of the DRAM method. One of the advantages of using BI, such as the DRAM method, is that it provides a probabilistic framework for parameter estimation. Instead of providing a single point estimate for each parameter, Bayesian methods allow us to obtain additional information about the parameter values, such as their distribution or median value. This information-rich approach provides a more nuanced understanding of the parameter uncertainty and variability.

Parameter distributions obtained with the Bayesian method

Based on the iterative behaviours presented in Fig. 1, it is expected that the DRAM method is not prone to getting trapped

in local minima, and can terminate the procedure still searching the wider parameter domain for better results. The DRAM results are the probability density distribution of each parameter, which allows it to deal with the sources of uncertainty during the parameter identification. Fig. 3 shows the posterior distributions of the acquisition parameter “acq_bas_pot” using DRAM. As shown in this figure, the median value of the parameter distribution and the best value (i.e., the parameter value that produces the least error) are expected to be closer as the number of iterations increases (1 000–2 000).

Although DRAM is more computationally efficient than the other methods, exploring the parameter wider domain with this method is time-consuming. In those cases where it is required to make an individualised identification of lifespan parameters of many animals (see Ramirez et al., 2022), it is necessary to find a balance between the minimum time allocated for the parameter identification and the quality of the identification. Fig. 4 shows that it is possible to obtain good results using a short iterative process (e.g. 2 000 iterations). This figure shows that the median values of the acquisition parameters “acq_bas_pot” are close values after short (2 000) and long (5 000 or 10 000) iterations.

Table 3

Summary of the adjustment between the four Genetic Scale Parameters (GSP) in the dairy cows AQAL model (Puillet et al., 2016) estimated with the DRAM method and the GSP estimated with DE, GA, PSO and SA methods.

Method	GSP	Dataset	Respect to DRAM			
			RMSPE%	CCC	Slope	Intercept
DE	acq_bas_pot	Dataset1	9.2	0.213	0.27	6.48
DE	acq_bas_lac	Dataset1	23.2	0.396	0.40	6.49
DE	init_G	Dataset1	11.4	0.505	0.55	0.17
DE	s2pc_pot	Dataset1	6.8	0.167	0.30	0.56
DE	acq_bas_pot	Dataset2	10.7	0.634	0.75	2.50
DE	acq_bas_lac	Dataset2	19.4	0.450	0.87	1.45
DE	init_G	Dataset2	10.8	0.182	0.26	0.28
DE	s2pc_pot	Dataset2	5	0.036	0.04	0.84
GA	acq_bas_pot	Dataset1	7.2	0.464	0.55	4.10
GA	acq_bas_lac	Dataset1	22.1	0.564	0.79	2.36
GA	init_G	Dataset1	10.2	0.586	0.66	0.15
GA	s2pc_pot	Dataset1	6.5	0.148	0.25	0.59
GA	acq_bas_pot	Dataset2	16.8	0.476	0.99	0.78
GA	acq_bas_lac	Dataset2	36	0.104	0.47	4.47
GA	init_G	Dataset2	14.2	0.202	0.49	0.21
GA	s2pc_pot	Dataset2	8.8	0.190	0.85	0.09
PSO	acq_bas_pot	Dataset1	6.4	0.478	0.51	4.04
PSO	acq_bas_lac	Dataset1	8.7	0.909	0.76	3.20
PSO	init_G	Dataset1	6.2	0.811	0.80	0.07
PSO	s2pc_pot	Dataset1	3	0.712	0.74	0.20
PSO	acq_bas_pot	Dataset2	23.7	0.348	1.17	-0.71
PSO	acq_bas_lac	Dataset2	24.2	0.240	0.57	5.71
PSO	init_G	Dataset2	9.5	0.425	0.67	0.14
PSO	s2pc_pot	Dataset2	5.2	0.244	0.45	0.48
SA	acq_bas_pot	Dataset1	7.8	0.321	0.40	4.71
SA	acq_bas_lac	Dataset1	11.1	0.882	0.75	3.69
SA	init_G	Dataset1	7.8	0.775	0.91	0.01
SA	s2pc_pot	Dataset1	2.4	0.785	0.76	0.19
SA	acq_bas_pot	Dataset2	11.2	0.635	0.78	2.12
SA	acq_bas_lac	Dataset2	17.1	0.459	0.74	3.49
SA	init_G	Dataset2	6.2	0.530	0.41	0.23
SA	s2pc_pot	Dataset2	6.8	0.289	0.87	0.09

AQAL = dynamic-mechanistic model of lifetime trajectories of energy acquisition and allocation in dairy cows; DRAM = Delayed Rejection Adaptive Metropolis; DE = Differential Evolution; GA = Genetic Algorithm; PSO = Particle Swarm Optimisation; SA = Simulated Annealing. The parameters “acq_bas_pot” and “acq_bas_lac” are the basal and an additional lactation acquisition (Kg/d), respectively. The parameters “init_G” and “s2pc_pot” control energy allocation for growth and lactation, respectively. For details, see Puillet et al. (2016). Concordance correlation coefficient (CCC, scale from -1 to 1).

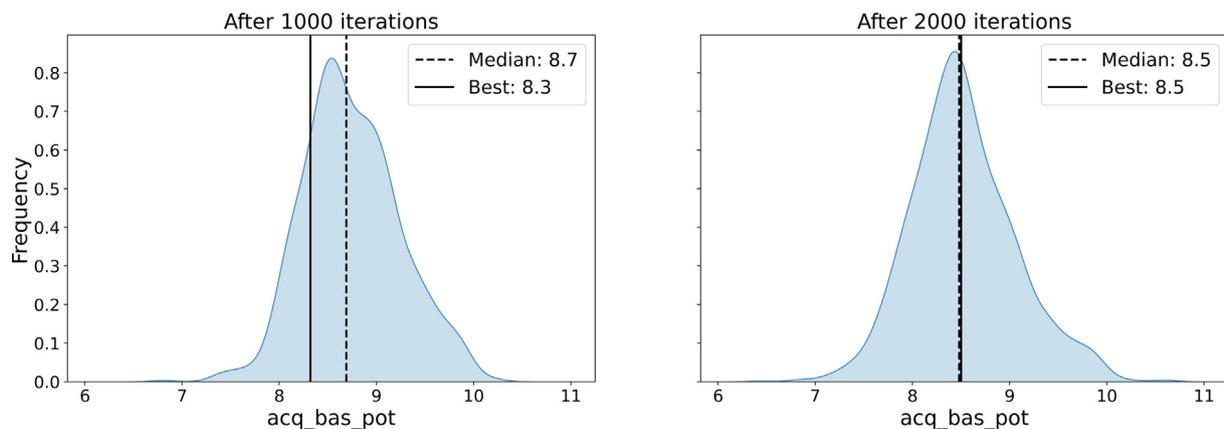


Fig. 3. Examples of the posterior distributions of the acquisition parameter “acq_bas_pot” (basal acquisition, Kg DM/d, in the dairy cows AQAL model, Puillet et al., 2016) using Delayed Rejection Adaptive Metropolis (DRAM). The dashed line is the median value of the parameter distribution, and continuous line is the best parameter (i.e. the parameter that produces the least error) after 1 000 and 2 000 iterations.

Simulation results with the Bayesian method

Fig. 5 shows the simulation results for two cows, from the two datasets used in this work. This figure compares the actual MY, BW, and BCS data and the simulation results using the best set of parameters obtained with the DRAM method.

Author’s point of views

- The main outcome of the study is the comparison of different optimisation methods for parameter identification in a cow’s lifespan model. The DRAM method was chosen as the reference method due to its ability to provide a more informative and

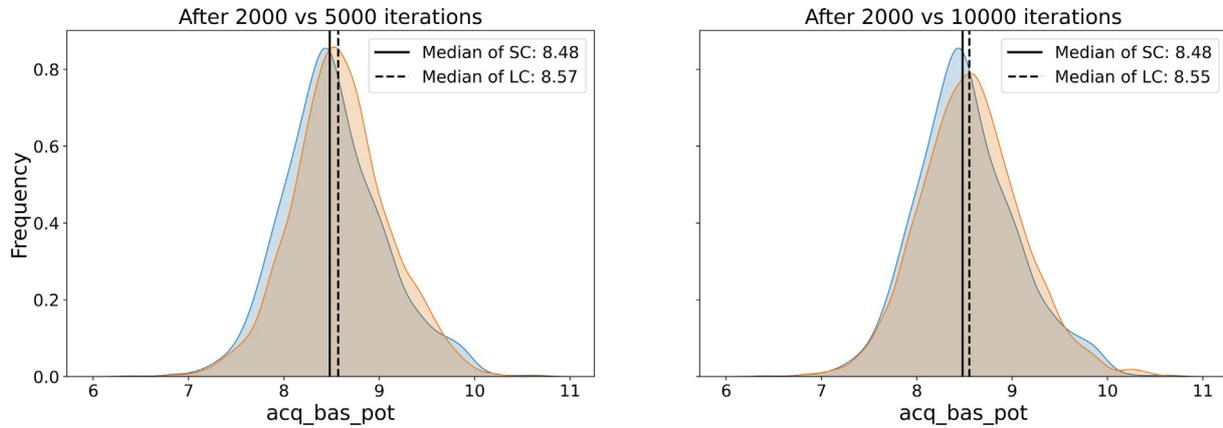


Fig. 4. Examples of posterior distributions of the acquisition parameter “acq_bas_pot” (basal acquisition, Kg DM/d, in the dairy cows AQAL model, PUILLET et al., 2016) using Delayed Rejection Adaptive Metropolis (DRAM). In blue, the distribution of the short chain (SC, n = 2 000); and in orange, the distribution of the long chain (LC, n = 5 000 or n = 10 000). The continuous and dashed lines are the median values of the parameter distribution for SC and LC, respectively.

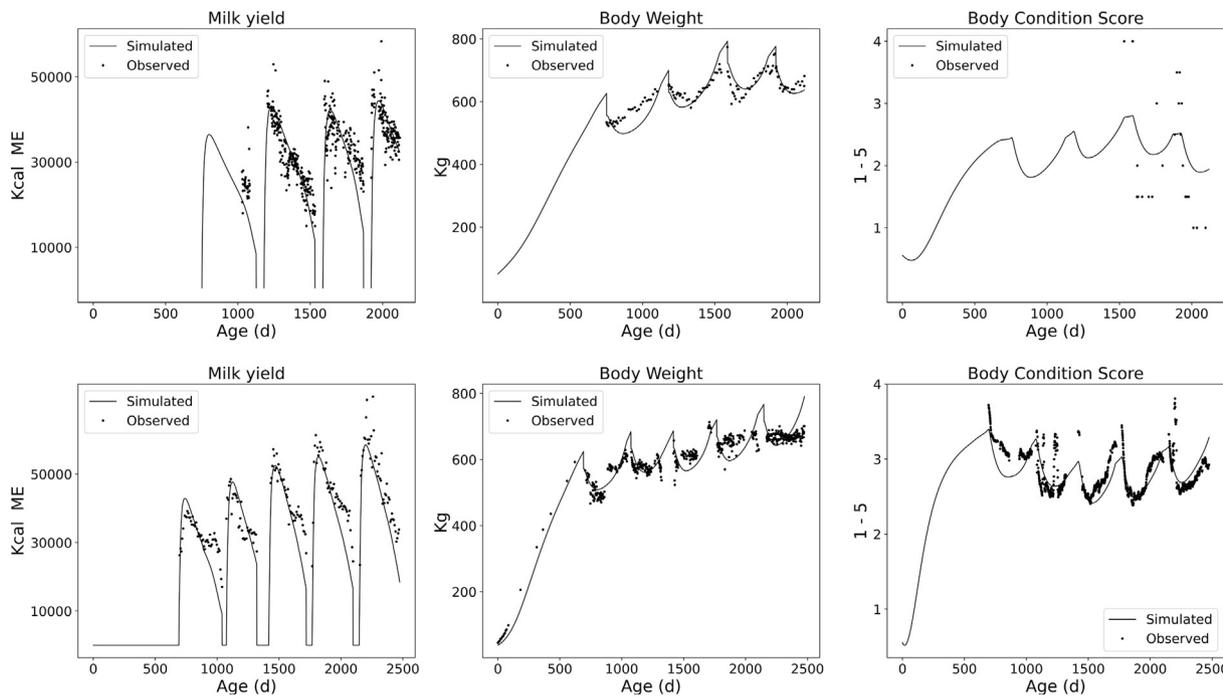


Fig. 5. Simulation results for two cows, from the two datasets used in this work. The figure compares the actual MY, BW, and BCS data and the simulation results of the dairy cows’ AQAL model (PUILLET et al., 2016) using the best set of parameters obtained with the Delayed Rejection Adaptive Metropolis (DRAM) method.

comprehensive representation of the estimated parameters. While it is true that the parameter values obtained using different optimisation methods may differ (Table 3), the Bayesian framework allows us to capture the inherent variability and uncertainty in a more rigorous and systematic way.

- The results show that all methods, including DRAM, are efficient in finding parameter values that reduce the distance between simulated and experimental data.
- There are limitations in the study that might have affected the outcome. One limitation is the uncertainty associated with the inference of parameter values due to the non-convex nature of the parameter search problem and the simulation time length.
- The results from this study can be used and benefit various stakeholders such as researchers in the field of animal breeding programs. Breeders can utilise AQAL and BI to identify efficient animals that are resistant to harsh conditions and recover quickly from adverse events. The study confirms that mechanistic models, such as the AQAL, can be valuable tools for understanding the underlying mechanisms of complex biological phenomena.
- The data from this study can be reused for further analysis by researchers and modellers interested in studying cow lifespan, resource allocation, and breeding strategies. The dataset could be used to compare different modelling approaches, test alternative parameter estimation methods, or explore additional factors influencing cow lifespan, such as environmental or

behavioural factors. Further analysis could provide deeper insights into the dynamics of cow lifespans and the potential for improving breeding programs.

Conclusion

The use of mechanistic models and BI has significant implications for breeding programs, offering a means to identify animals with high efficiency and resilience based on their estimated genetic potential. This particular study focused on parameter identification within a mechanistic model of cow lifetime performance. The objective was to compare different methods for determining individual parameters that accurately represent the genetic potential of each cow and its interaction with the environment. The results demonstrated that various methods, such as DRAM, DE, GA, PSO, and SA, were effective in minimising the disparity between simulated and experimental data. Notably, DRAM proved to be more computationally efficient and provided additional insights into the statistical properties of each parameter. These findings highlight the importance of selecting appropriate parameter optimisation methods in order to obtain reliable and informative results for breeding programs.

Ethics approval

No approval was needed for his study; the original data procurement was in accordance with European ethical regulations.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) did not use any AI and AI-assisted technologies.

Author ORCIDs

J.F. Ramirez: <https://orcid.org/0000-0002-2169-7260>.

J.B. Daniel: <https://orcid.org/0000-0001-7627-3708>.

N.C. Friggens: <https://orcid.org/0009-0008-0410-7223>.

L. Puillet: <https://orcid.org/0000-0003-2358-7302>.

Author contributions

NCF and **JBD** produced the experimental data of the study. **JFR** developed the codes for calibration. **JFR**, **JBD**, **LP** and **NCF** drafted the paper. All authors read and approved the final manuscript.

Declaration of interest

The authors declare that they have no competing interests.

Acknowledgements

We thank Trouw Nutrition R&D and Klaus Lønne Ingvartsen from Aarhus University for sharing the data used for the work presented in this manuscript.

Financial support statement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727213 (GenTORE).

References

- Fogel, D.B., 1994. Asymptotic convergence properties of genetic algorithms and evolutionary programming: analysis and experiments. *Cybernetics and Systems: An International Journal* 25, 389–407.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. *Statistics and Computing* 16, 339–354.
- Ingber, L., 1993. Simulated annealing: Practice versus theory. *Mathematical and Computer Modelling* 18, 29–57.
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: *Proceedings of ICNN'95-international conference on neural networks*. November 27 – December 1, 1995, Perth, Australia, pp. 1942–1948.
- Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 255–268.
- Miles, P.R., Smith, R.C., 2019. Parameter estimation using the python package pymcmcstat. In: *Proceedings of the 18th Python in Science Conference (SCIPY 2019)*. 8–14 July 2019, Austin, TX, pp. 93–100.
- Muñoz-Tamayo, R., Puillet, L., Daniel, J.B., Sauviant, D., Martin, O., Taghipoor, M., Blavy, P., 2018. To be or not to be an identifiable model. Is this a relevant question in animal science modelling? *Animal* 12, 701–712.
- Piotrowski, A.P., Napiorkowski, J.J., Piotrowska, A.E., 2020. Population size in particle swarm optimization. *Swarm and Evolutionary Computation* 58, 100718.
- Puillet, L., Réale, D., Friggens, N.C., 2016. Disentangling the relative roles of resource acquisition and allocation on animal feed efficiency: insights from a dairy cow model. *Genetics Selection Evolution* 48, 1–16.
- Puillet, L., Ducrocq, V., Friggens, N.C., Amer, P.R., 2021. Exploring underlying drivers of genotype by environment interactions in feed efficiency traits for dairy cattle with a mechanistic model involving energy acquisition and allocation. *Journal of Dairy Science* 104, 5805–5816.
- Ramirez, J.F., Puillet, L., Friggens, N.C., 2022. Identifying diversity in cattle performance using Bayesian inference and a model. In: *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, 3–8 July 2022, Rotterdam, The Netherlands, pp. 328–331.
- Schafer, R.W., 2011. What is a Savitzky-Golay filter? [lecture notes]. *IEEE Signal Processing Magazine* 28, 111–117.
- Sherri, M., Boulkaibet, I., Marwala, T., Friswell, M.I., 2019. A differential evolution Markov chain Monte Carlo algorithm for Bayesian model updating. In: *Special Topics in Structural Dynamics, Volume 5: Proceedings of the 36th IMAC, A Conference and Exposition on Structural Dynamics*, 12–15 February 2018, Orlando, FL, pp. 115–125.