Identifying diversity in cattle performance using Bayesian inference and a dynamic, mechanistic model

J.F. Ramirez[1*], L. Puillet[1], and N.C. Friggens[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 75005 Paris, France; [*]johnfredy.ramirezagudelo@inrae.fr

Abstract

Quantitative description of individual animal performance is a key element of breeding. In addition to measurements of traits associated with productivity, dynamic/mechanistic models can be used for identifying diversity of complex traits like feed efficiency. The objective of this work was to use Bayesian inference to quantify differences in a model's "genetic-scaling" parameters (related to resource acquisition and resource allocation) within- and between-breeds. We first describe the experimental data processing before calibration, the function used for the error calculation, and the Bayesian algorithm used for parameters generation. Two databases were used as case studies. The results showed that this approach allowed us to identify variations in the genetic-scaling parameters within- and between-breeds. This approach is a useful tool to quantify complex traits for future selection of efficient and resilient animals.

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. When direct measurements of biological mechanisms are difficult or even impossible, mechanistic models allow estimation of these underlying traits by individual parametrization of a model. Puillet et al. (2016) developed a dynamic/mechanistic model (AQAL) to generate genetically-driven trajectories of energy acquisition and allocation in dairy cows lifespan, with four genetic-scaling parameters (GSP). 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. This model 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 strategies evaluation. In this type of model, Bayesian inference 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). This work presents a protocol to use Bayesian inference for within- and between-breeds characterization of resource acquisition and allocation parameters. The protocol presented here aims to provide a comprehensive description for the individualized calibration of AQAL in Python. First, we describe the observed data processing before calibration. Then, we describe the general structure of the calibration process. We illustrate the application of the protocol for two case studies.

Materials & Methods

Time-series observations, data handling before calibration. For calibrating AQAL, experimental data is necessary to run the model and to compute the error. The model uses a vector of values to represent the increase in intake per lactation through the cow's lifespan. This vector must be defined by using the relationship among the averages of the maximum values (upper quartile) of milk yield observed in each lactation. The energy value of the diet (GE or ME) is another model input. At least three traits must be used to calculate the differences

between simulation and experimental data: Milk yield (MY), Body weight (BW), and Body Condition Score (BCS, 1-5 scale). MY must be expressed in kilocalories of Metabolizable Energy (ME) assuming a metabolic efficiency of 0.64, and Gross Energy content (GE, Kcal/kg) of 5700, 9200, 3950 for protein, fat, and lactose.

Formulation of the objective function. In each iteration of the calibration process, the objective function uses the set of parameters generated by the Bayesian algorithm and returns the mismatch between the simulation and the experimental values of MY, BW, and BCS. We used the Mean Absolute Percentage Error (MAPE) to quantify the mismatch because this metric avoids the effect of the trait scales on the average error calculation. One MAPE is calculated for each variable x lactation, then the overall average MAPE was calculated from these individual variable x lactation MAPEs. To minimize the effect of unbalances among the individual errors, the overall error variance was added to the overall MAPE. The minimum value of this sum was used to identify the best parameter set for a given individual animal.

Parameter optimization. In the Bayesian approach, the goal of the calibration is to infer the parameter posterior distributions given a set of observations (Miles and Smith, 2019). For parameters generation, we adopt an approximate-inference technique based on Monte Carlo methods, which aims at generating a sequence of random parameters from a Markov chain whose distributions approach the posterior distributions of each parameter. In particular, we have used the delayed rejection adaptive Metropolis (DRAM) algorithm (Haario et al., 2006) implemented for Python in the 'pymcmcstat' library (Miles and Smith, 2019). 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 chain. 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). The default program uses a uniform prior, which is a common approach for these types of problems (Miles and Smith, 2019). The parameters optimized were related to acquisition: *acq_bas_pot*, *acq_bas_lac* (basal and an additional lactation acquisition), and allocation of energy resource: *init_G*, *s2pc_pot*, (growth and lactation allocation).

Case studies. Two databases were used to test our parameters inference procedure. The first database (One-Breed) contains data from 26 Holstein cows. This database reported at least four lactations for all cows. The second database (MEMO, see Friggens et al. 2007) contains data from three breeds (18 Holstein, 17 Red Danish, and 19 Jersey cows). This database reported at least three lactations for all cows. Both databases report measures of MY and composition, BW, and BCS. The first and second databases were used to examine within- and between-breed diversity of the four GSP of AQAL, respectively.

Results

Figure 1 shows an example of simulation, using the best set of parameters for a specific cow, and observed data of MY, BW, and BCS. Figure 2 shows an example of the Markov chains for a specific cow. The first half of the chain served to tune the sampler and was discarded (burn-in period). To approximate the posterior distribution for each parameter, parameter sets with low error from the second half of each chain were used. Figure 3 compares the parameter distributions for a specific cow and the breed (One-Breed database). Figure 4 shows the parameter distributions for the three breeds (Holstein, Red Danish, and Jersey) in the MEMO database. Finally, we present a Table with the average parameters values and the associated variance.
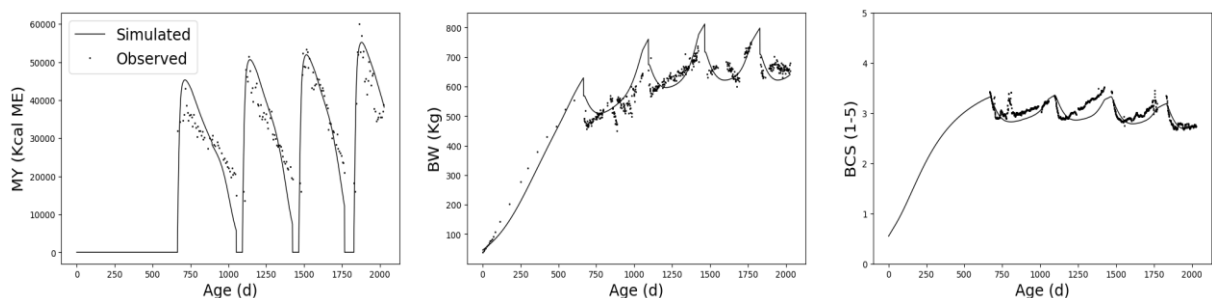
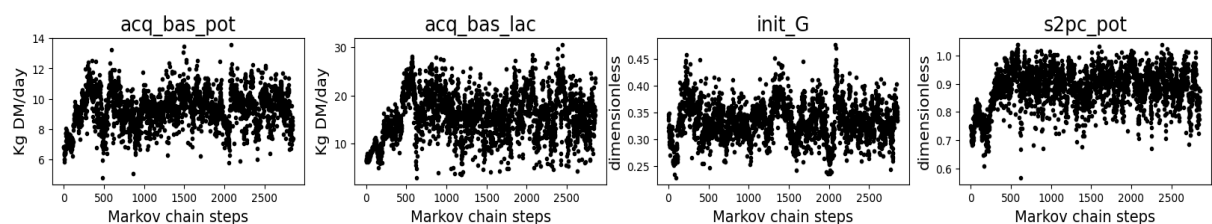**Figure 1. Example of simulation results with the best set of parameters for an individual**

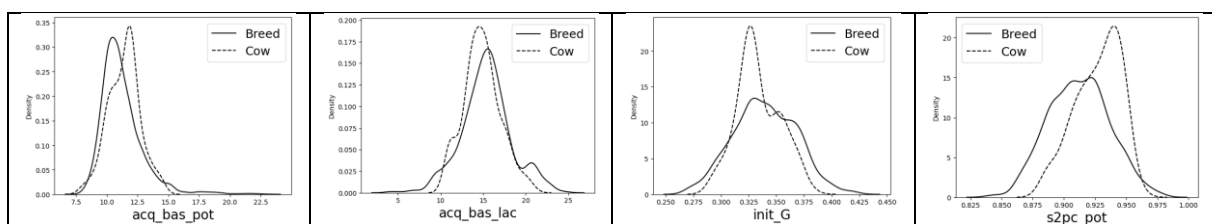**Figure 2. Example of the Markov chain of four genetic-scale parameters**

Figure 3. Posterior parameter distributions for a breed and an example of one individual of that breed
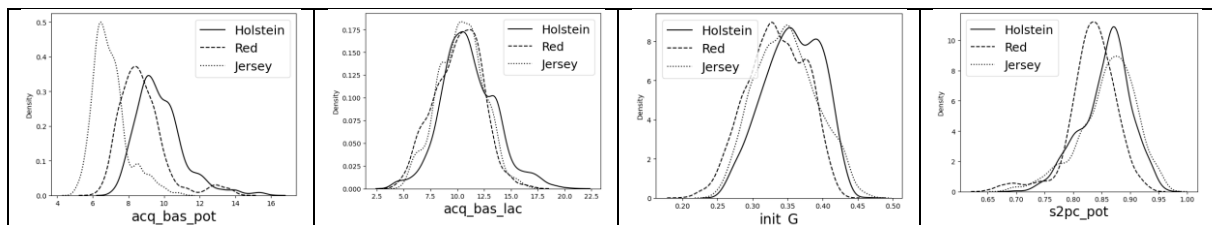

Figure 4. Posterior distributions of the genetic-scale parameters between breeds

Table 1. Parameter values per database and breed.

| | One-Breed | MEMO | | |
| | | Holstein | Danish Red | Jersey |
Parameters	Median (Std)	Median (Std)	Median (Std)	Median (Std)
acq_bas_pot	11.1 (1.3)	9.9 (1.1)	8.8 (1.3)	7.2 (1.3)
acq_bas_lac	16.1 (2.9)	12.0 (2.0)	10.9 (1.4)	10.4 (1.8)
init_G	0.34 (0.025)	0.36 (0.039)	0.33 (0.040)	0.35 (0.037)
s2pc_pot	0.92 (0.022)	0.86 (0.030)	0.84 (0.029)	0.86 (0.042)

Discussion

We chose to use a Bayesian method for parameter estimation because we expected to explore different local minima within the response surface of the objective function (since the model

contains multiple nonlinear functions). In this context, the Bayesian inference used (DRAM) was an efficient method to explore the parameter space, showing the power of the proposed method for this type of task. The results allow us to demonstrate the ability of a dynamic/mechanistic models to identify variations in GSP both within- and between-breeds. This is very interesting for the future identification and selection of efficient and resilient animals because dynamic/mechanistic models (e.g. AQAL) can simulate the consequences of different combinations of acquisition and allocation parameters on animal performance under different nutritional environments (see Puillet et al., 2021).

References
Friggens, N.C., Berg, P., Theilgaard, P., Korsgaard, I.R., Ingvartsen, K.L., Løvendahl, P., and 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*(11), 5291-5305.

Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). DRAM: efficient adaptive MCMC. *Statistics and computing*, *16*(4), 339-354.

Miles, P.R., and Smith, R.C. (2019). Parameter estimation using the python package pymcmcstat. In *Proceedings of the 18th Python in Science Conference (SCIPY 2019)*.

Puillet, L., Réale, D., and 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), 1-16.

Puillet, L., Ducrocq, V., Friggens, N.C., and 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*(5), 5805-5816.