Assessing adaptive phenotypic plasticity by means of conditional strategies from empirical data: the latent environmental threshold model

Mathieu Buoro, Olivier Gimenez, Etienne Prévost

Article

Assessing adaptive phenotypic plasticity by means of conditional strategies from empirical data: the latent environmental threshold model

Mathieu Buoro 1,2,4, Olivier Gimenez 1,6 and Etienne Prévost 2,3,5

1 Centre d'Ecologie Fonctionnelle et Evolutive, campus CNRS, UMR 5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France.

2 INRA, UMR Ecobiop, Quartier Ibarron 64310 Saint Pée s/ Nivelle, France

3 Université de Pau et Pays de l'Adour, UMR Ecobiop, Campus de Montaury, 64600 Anglet, France

4 E-mail: mathieu.buoro@cefe.cnrs.fr

5 E-mail: eprevost@st-pee.inra.fr

6 E-mail: olivier.gimenez@cefe.cnrs.fr

Corresponding author:

Mathieu Buoro

Tel: +33 4 67 61 34 30

Fax: +33 4 67 41 21 38

E-mail: mathieu.buoro@cefe.cnrs.fr

Keywords: Bayesian modeling; Conditional strategies; Environmental Threshold Model; Measurement error; Phenotypic plasticity

ABSTRACT

Conditional strategies are the most common form of discrete phenotypic plasticity. In a conditional strategy, an expressed phenotype is determined by the difference between an environmental cue and a threshold, both of which may vary among individuals. The Environmental Threshold model (ETM) has been proposed for understanding the evolution of conditional strategies. Surprisingly, the ETM has little been applied to empirical studies. Here, we extending the ETM that allows assessment of conditional strategies from observational data. Our model accommodates likely situation where an observable aspect of the environment ('observable cue') is distinct from the unobserved proximate cue. First, we show that ignoring the observable vs. proximate cue distinction can lead to overestimate the variation in the threshold. Second, despite of identifiability issue, we show that the LETM allow to estimate the proximate cue and the threshold at individual level. Third, we propose the use of genetic data to cope with identifiability issues. Finally, we illustrate our approach with empirical data on the size-dependent smolting process for stream-dwelling Atlantic salmon juveniles in the Scorff River (Brittany, France). We argue that coupling our model with quantitative genetic methods could allow disentangling the genetic and environmental components of the threshold phenotypic variance.

INTRODUCTION

Phenotypic plasticity - the ability of a genotype to produce different phenotypes according to environmental conditions – can be an adaptive response of organisms to selection in stochastic environments (Ghalambor et al. 2007). When plasticity refers to a discrete trait (e.g., maturation at given age, polymorphism in defensive structures, and alternative mating tactics), the concept of the conditional strategy (Gross 1996) is a popular framework to study the ability of a genotype to express alternative phenotypes (or tactics). Tomkins and Hazel (2007) define the conditional strategy as a genetically determined decision rule containing a conditional clause. For binary traits, i.e. traits two possible categorical phenotypic states, the phenotype expressed by an individual may be dependent on an environmental cue and the choice between phenotypes may result from a physiological 'comparison' between the cue and a threshold (or switchpoint).

Tomkins and Hazel (2007) critically assess the Status Dependent Selection model (SDSM) initially formulated by Gross (1996) for representing a conditional strategy. They propose to replace the SDSM with the Environmental Threshold Model (ETM; Hazel et al. 1990; Hazel et al. 2004). The ETM is an extension of the SDSM. First, the main difference between the two models lies in that the ETM assumes the threshold can vary among individuals according to their genotype. Second, the ETM facilitates the use of quantitative genetic theory for understanding the evolution of conditional strategies (Hazel et al. 2004). In the ETM the threshold is considered as a trait that varies among individuals, and in this quantitative genetic framework the threshold is considered to be controlled by a trait that is modeled as continuously and normally distributed on some unobserved latent scale. In the ETM model, the threshold can thus be treated as a heritable trait where variation may come from multiple sources, including environmentally-induced variation and heritable differences among individuals (Roff 1997). At the individual level, the phenotype expressed results from the comparison between the cue and the threshold both in the SDSM and the

ETM. At the population level, the proportion of the alternative phenotypes depends only on the cue distribution in the SDSM - the threshold is fixed - whereas it depends both on the cue and the threshold distributions in the ETM. The ETM formally distinguishes phenotypic plasticity from evolution of the phenotype proportions (Tomkins & Hazel 2007). A change in the cue distribution immediately translates into a change in the phenotype proportions, therefore reflecting phenotypic plasticity. Evolutionary changes modify the threshold distribution and thus shift the phenotype proportions as well, and potentially independently of changes in the distribution of the cue.

In the context of environmental change, understanding how organisms can respond to varying environmental conditions is of particular importance. Understanding this potential for phenotypic change requires simultaneous consideration of the consequences of phenotypic plasticity and of evolutionary change, and in the current context, of the evolution of plasticity. The ability of the ETM to accommodate both phenotypic plasticity and its evolution is most appealing for assessing conditional strategies in the wild. Indeed, it allows addressing jointly in a single model the two mechanisms by which adaptation can occur, namely environmental cues that trigger phenotype expression and genetic variation that control phenotype expression. However, confronting the ETM with observation data collected in the wild requires embedding the ETM into a statistical model to deal with uncertainty. Although the thresholds and the parameters of their distribution are clearly not observable, they are conceptual and unknown quantities for which statistical inference can be derived from observed data. In contrast, the phenotypes expressed by individuals are observable. The status of the environmental cue is more ambiguous. Indeed, the proximate mechanism by which an individual assesses its environment to adjust its phenotype remains in most instances unknown (Metcalfe 1998; Tomkins & Hazel 2007). The environmental cue of the ETM should then be split in two distinct but related quantities: the *proximate cue*, for which the comparison with a threshold would trigger the

phenotype expressed by an individual, is hidden; however it may be correlated with an *observable cue* which can be readily measured.

The proposed distinction between observable and (unobservable) proximate cue is supported by Ostrowski et al. (2000) study on the snail *Bultinus truncatus*. They tested the ETM in a set of experiments where both the genotype and the environment were controlled. In contrast with what was expected under the ETM, they observed significant random variation in the phenotype expression for fixed environment X genotype combinations. They hypothesized a micro-environmental and uncontrolled variation in the threshold to explain this residual randomness. Avoiding the rather inconsistent statement that environment varies significantly even when it is experimentally fixed, we contend the proximate vs. observable cue dichotomy is a more sensible alternative hypothesis. In the case Ostrowski et al. (2000)'s study, it would mean that the organisms assess temperature – the observable environmental cue experimentally controlled - through an unknown proximate mechanism with some random "measurement error". The proximate cue would then only be correlated with temperature, which could be used as the observed cue. Note that any observable quantity traits can also be used as environmental cues. In fish, growth rate or size at a given age are known to be strongly influenced by the environment and are thus considered as integrating various environmental factors (Hutchings 2004; Dieckmann et al. 2007). Body size may be particularly useful as an observed cue, because it may integrate many important aspects of the environment an individual has experienced in a single tractable and obtainable measurement. However body size or any other cue should nonetheless still be distinguished from the proximate cue triggering the phenotype expression because there must be an (unknown) inner mechanism by which an organism gets informed about its size or growth rate, because a perfect correlation with the proximate cue is unlikely.

The purpose of this paper is to develop a statistical model extending the ETM that

incorporates the proximate vs. observable cue distinction. It allows the estimation of latent variables of primary interest, i.e. the proximate cue and the threshold trait. We refer to this new model as the Latent Environmental Threshold Model (LETM). We first describe the model and show that ignoring the proximate vs. observable cue distinction leads to overestimation of the variance of the distribution of the threshold. This result is crucial since the potential for evolution of the conditional strategy directly depends on this the existence of this variability in the threshold and its heritability (Hazel et al. 2004). Second, we investigate the properties of our new model using simulated data. In particular, we explore the conditions in which the parameters of interest are, or can be rendered, identifiable. Particularly, we show that the use ofgenetic data render parameters of interest identifiable. Our LETM approach is applicable to any situation where the conditional strategy framework is relevant and when individual data are available for at least the alternative phenotypes involved and a related observable cue. Third, because Atlantic salmon is a conditional strategist (*sensu* Gross 1996) with respect to status-dependent alternative life-history tactics (migrating to sea or not, delaying reproduction or not), we illustrate our approach with a case study on the size-dependent smolting process for stream-dwelling juvenile Atlantic salmon in the Scorff River (Southern Brittany, France).

THE LATENT ENVIRONMENTAL THRESHOLD MODEL (LETM)

We use the notation $A|B \sim Dist[f(B)]$ to denote a set of random variables A distributed conditionally on the set of variables B according to a probability distribution *Dist* which parameters are a function f of B. Observable quantities are denoted with capital roman letters and unknowns with greek letters.

The LETM as an extension of the ETM model

For an individual *i*, the threshold modeling framework stipulates that if the value of a cue η_i is larger (resp. lower) than a critical threshold θ_i, then it triggers the expression of a phenotype, say migrant (resp. resident). If Y_i is the binary variable indicator of the phenotype (e.g., 1 for migrant and 0 for resident), then we have:

$$Y_i = \begin{cases} 1 & if \quad \eta_i > \theta_i \\ 0 & if \quad \eta_i \leq \theta_i \end{cases} \qquad (1)$$

In line with the ETM, the threshold varies among individuals and is a polygenic quantitative trait which is normally distributed with mean μ_θ and standard deviation σ_θ , as typically assumed in quantitative genetics (Hazel et al. 1990; Tomkins & Hazel 2007; Lynch and Walsh 1998):

$$\theta_i | \mu_\theta, \sigma_\theta \sim N(\mu_\theta, \sigma_\theta) \qquad (2)$$

where η_i and θ_i.is the value of the proximate cue and the threshold of individual *i*, respectively. The proximate cue η_i is to be compared with the threshold θ_i. η_i is assumed to vary among individuals as a function of the environment, and θ_i is considered an intrinsic property of the individuals. Although η_i is not observable, an observable cue X_i can be measured which is correlated with η_i. Indeed, the distribution of the unknown proximate cue η_i can be expressed conditionally on the observable cue X_i with some error.

Specifically, η_i is assumed to be normally distributed around X_i with standard deviation σ_η:

$$\eta_i | X_i, \sigma_\eta \sim N(X_i, \sigma_\eta) \quad (3)$$

Note that it corresponds to the Berkson measurement error model in the statistical literature (Congdon 2006). This formulation has the advantage of making no assumption regarding the distribution of η_i, therefore making the statistical analysis convenient since irrespective of the origin of the X_i distributions.

We refer to the model defined by Eqns 1-3 as the Latent Environmental Threshold Model (LETM) since (i) in the statistical modeling terminology, both η_i and θ_i are latent variables (Congdon 2006) and (ii) it is an extension of the ETM which corresponds to the limit case where σ_η is null, i.e., the proximate cue and the observable cue are equal ($X_i = \eta_i$). The conditional structure of the LETM can be summarized by a Directed Acyclic Graph (DAG; Fig. 1).

[Figure 1 about here]

The latent data formulation of the LETM

In the LETM, we assume that Y_i the binary variable indicator of the phenotype (e.g., 1 for migrant and 0 for resident) depends on continuous latent variables, i.e. the proximate

cue and the threshold. To estimate the relevant parameters, we need to derive the distribution of Y_i. First, Y_i is distributed according to a Bernoulli distribution with probability p_i:

$$Y_i \sim Bernoulli\left(p_i\right) \quad (4)$$

Second, using that the random variable $\left(\frac{\theta_i - \mu_\theta}{\sigma_\theta}\right)$ is distributed as a standardized normal distribution N(0,1) (recall that θ_i is normally distributed; Eqns 2), we have:

$$p_i = Pr\left(Y_i = 1\right) = Pr\left(\eta_i > \theta_i\right) = Pr\left(\frac{\eta_i - \mu_\theta}{\sigma_\theta} > \frac{\theta_i - \mu_\theta}{\sigma_\theta}\right) = F\left(\frac{\eta_i - \mu_\theta}{\sigma_\theta}\right) \quad (5)$$

where F is the cumulative function of a standardized normal distribution. The inverse of F is often referred to as the probit function and is often used for parameters that lie in the unit interval i.e., it connects the dichotomous variable Y_i and the corresponding continuous latent variable η_i through the proportion p_i. Third, conditionally on the proximate cue η_i and integrating over the latent variable θ_i, we obtain:

$$Y_i \mid \eta_i, \mu_\theta, \sigma_\theta \sim Bernoulli\left(F\left(\frac{\eta_i - \mu_\theta}{\sigma_\theta}\right)\right) \quad (6)$$

This formulation corresponds to the standard ETM formulation (e.g., Ostrowski et al. 2000).

To make the distinction between the proximate and observable cues , the LETM formulation can be obtained using the same reasoning as above by conditioning on the observable cue X_i and by integrating over the latent variables η_i and θ_i:

$$Y_i | X_i, \mu_\theta, \sigma_\theta, \sigma\eta \sim Bernoulli\left[F\left[\frac{X_i - \mu_\theta}{\sqrt{\sigma_\theta^2 + \sigma_\eta^2}}\right]\right] \quad (7)$$

The comparison of the two formulations of the ETM and LETM (respectively Eqns [6] and [7]) shows that confounding the proximate and the observable cue makes the variance component of the proximate cue falsely attributed to the variation in the threshold and artificially inflates the threshold variability. Given that the variability in the threshold is of primary interest for studying conditional strategy, ignoring the proximate vs. observable cue distinction could result in flawed inference regarding the potential evolution of the conditional strategy (Hazel et al. 2004; Tomkins & Hazel 2007).

Identifiability issues

The LETM is a statistical model which makes the connection between the conditional strategy conceptual framework and observable quantities at the individual

level. The unknowns that need to be estimated are both the individual latent variables (the proximate cue η_i and the threshold θ_i) and the parameters controlling the distribution of these latent variables (the mean μ_θ and the standard deviation σ_θ of the individual threshold distribution and the standard deviation σ_η of the proximate cue distribution around the observable cue X_i). However, when only the observable cues X_i's and the alternative phenotypes Y_i's are available, the LETM is not fully identifiable, that is not all parameters are separately estimable. More specifically, only the total variance $\sigma_T^2=\sigma_\theta^2+\sigma_\eta^2$ is identifiable while the ratio $\kappa=\sigma_\theta^2/\sigma_T^2$ is not. The ratio κ which corresponds to the proportion of total variance explained by the variability in the threshold is of primary interest because if it is not identifiable, then the variance of the threshold is not identifiable either. Indeed, from Eqns [5], the likelihood of the observations (Y_i, X_i) conditionally on (μ_θ, σ_θ, σ_η) (and consequently σ_T) remains unchanged for any combination of (σ_θ, σ_η). To circumvent this issue, additional information is required. When studying wild populations or when performing controlled experiments, genetic data may be available providing additional information about the structure of the θ_i's variability. Assuming the threshold is heritable (Hazel et al. 1990; Tomkins and Hazel, 2007), batches of related individuals have the same threshold value, thus enhancing our ability to estimate the value of the threshold for each individual of each batch and the variability of the threshold in the population. Consequently, knowledge of the individuals kinship or pedigree inform about the relative degree of similarity of the θ_i's between individuals and about the variability of the threshold θ in the population. We explore the introduction of the genetic information in the LETM and its influence on the estimation of parameters below (see “Simulation study” section and “Results” section).

Bayesian statistical inference

To fit our LETM to the data, we adopted a Bayesian approach using Markov chain Monte Carlo (MCMC) algorithms because it provides a flexible framework for analyzing latent variables models and their conditional structure (Clark 2005). The Bayesian approach combines the likelihood (i.e., information derived from the observed data) and the prior distribution of the unknown quantities (i.e., knowledge available before the data were observed) to produce a joint probability distribution of all model unknowns conditionally on the observed data, the so called joint posterior distribution (see Gelman et al. 2004; Ellison 2004; McCarthy 2007 for more details about the Bayesian statistical modeling). If the prior and the posterior distributions of a given parameter look similar, , then there is not enough information in the data to estimate this parameter. The joint posterior distribution of all the model unknowns, i.e., the parameters (μ_θ, σ_θ, σ_η) and the individual thresholds and proximate cues (θ_i, η_i), was obtained by means of MCMC sampling as implemented in the OpenBUGS software (Spiegelhalter et al. 2003). The OpenBUGS code of our model is available at (http://www.cefe.cnrs.fr/biom/zips/LETM_Buoro et al.txt). We ran 3 parallel MCMC chains and retained 50,000 iterations after an initial burn-in of 10,000 iterations. Convergence of MCMC sampling was assessed by means of the Brooks-Gelman-Rubin diagnostic (Brooks & Gelman 1998). A Bayesian analysis requires specifying prior probability distributions for the model parameters, i.e. the unknown quantities which are not conditioned by any other quantity in the model (μ_θ, σ_θ, σ_η; Fig. 1). The prior on the mean of the threshold distribution μ_θ was a normal distribution with mean 0 and a large variance (1000). To make the assessment of identifiability issues easier, priors on the standard deviations σ_θ and σ_η were not defined directly as is usual but rather on the total variance σ_T^2 and the ratio κ. Because there is a one-to-one transformation relating (σ_T^2 ,κ) to (σ_θ, σ_η), assigning a prior to (σ_T^2 ,κ) induces a prior on (σ_θ, σ_η) as well (Gelman et al. 2004). We used a scaled inverse-χ^2 with one degree of freedom for σ_T^2 (Gelman et al. 2004) and a uniform

distribution between 0 and 1 for κ. The empirical variance of the observed cue X_i was used to scale the prior on σ_T^2 to the level of the observed cue.

SIMULATION STUDY

To evaluate the performance of the LETM, we carried out a simulation study. The data were generated according to the LETM using known parameter values and including a simple genetic structure for the θ_i's. Statistical inference was derived from the simulated data to check whether the LETM was able to provide accurate estimates of the known values of the parameters (μ_θ, σ_θ , σ_η) and of the individual latent variables (i.e., the thresholds θ_i and proximate cues η_i). To assess the added value of bringing genetic knowledge about the θ_i's in the fitting process, inferences were successively conducted considering the genetic (batch) structure of the θ_i's was unknown and known. In the first instance, the model assumes the θ_i's are potentially all different. In the second instance, it is explicitly included in the model that the θ_i's are equal within a batch while being potentially different between batches (Eqn [8]).

We considered a set of 200 individuals assuming they were made of 20 batches of 10 individuals, each batch corresponding to the same genotype and a single threshold.First, 20 threshold values were generated for the 20 batches from a normal distribution with mean μ_θ =0 and standard deviation σ_θ =0.25 (Eqns [2]).
We then allocated these latent threshold values to the 200 individuals *i* so that for two distinct individuals *i* and *i'* :

$$B_i = B_{i'} \Leftrightarrow \theta_i = \theta_{i'} \quad (8)$$

where B_i and $B_{i'}$ denotes the batch number of the individuals i and i'. If individuals i and i' are from the same batch B_i , then they have the same threshold value. It means that the variability of the threshold is exclusively genetic and the heritability of this trait is 1. Our aim here is not to identify the genetic component of the threshold trait but the structure of the variability in the threshold through the relatedness between individuals.

Second, we generated an observed cue X_i value for each individual i by drawing from a normal distribution with mean 0 and standard deviation 1. For each individual *i*, given the value of the observed cue, X_i, we generated its proximate cue η_i from a normal distribution with mean X_i and standard deviation σ_η =0.5 (Eqn [3]). Finally, given the values of the proximate cue η_i and of the threshold θ_i, we assigned the phenotype indicator values Y_i (Eqn [1]). Note that simulated data are much more variable in the proximate cue than in their threshold, as should usually be the case (Tomkins and Hazel 2007). A graphical representation of the resulting data is provided in Figure 2.

[Figure 2 about here]

APPLICATION TO ALTERNATIVE LIFE HISTORY TACTICS IN ATLANTIC SALMON

Atlantic salmon is an anadromous species that shares its life cycle between freshwater and the ocean (Guéguen & Prouzet 1994). The juvenile phase takes place in freshwater and lasts one or two years in Brittany (Baglinière et al., 1993). Thereafter, fish migrate to the ocean and return after one or two years to their native stream to breed. Atlantic salmon are conditional strategists with state-dependent choice among alternative life history tactics (Thorpe et al. 1998). During their first year of life in their natal river, young of the year (YOY; i.e., individuals less than one year old, counting from fertilization)

have to 'decide' whether to migrate to the ocean the next spring or to reside in freshwater an additional year (Thorpe et al. 1998). The choice between the migrant vs. the resident alternative tactics (i.e., phenotypes) is related to the size of the individuals at their first autumn (Nicieza et al. 1991; Thorpe & Metcalfe 1998). Although size is an observable cue, it is probably best considered as a proxy for energetic status (Thorpe et al. 1998), i.e. likely a more biologically proximate cue, which is to be compared to a threshold for triggering seaward migration the next spring (Mangel & Satterthwaite 2008; Thorpe et al. 1998; Satterthwaite et al. 2009). The individual energetic status influences this life-history choice (Jonsson & Jonsson 2005) because migration to the ocean is preceded by the smolting process preparing individuals for sea water life which is energetically costly (Mccormick et al. 1998; Thorpe et al. 1998). The energetic status reflects the way energy is acquired, stored and used and is strongly influenced by the environmental conditions experienced by each individual (e.g., food availability, temperature regime or density of fellow beings; (Elliott & Hurley 1997; Forseth et al. 2001; Imre et al. 2005; Jones et al. 2002; Murphy et al. 2006). Under the LETM, we consider migrant vs. resident (at one year of age) as alternative phenotypes, YOY size in autumn as an observable cue indicative of the individual energetic status, the proximate cue triggering phenotype expression.

Data collection

In autumn 2005, YOY juveniles were sampled by electrofishing at 39 stations along the main course of the Scorff. Every fish captured was measured (fork length, to the nearest 1mm) and individually marked with a Passive Integrative Transponder (PIT) tag (11 mm long, 2.2 mm in diameter) inserted into the peritoneal cavity according to the protocol described in Acolas et al. (2007).

One year old seaward migrating juveniles (smolts) previously PIT tagged were identified during their downstream migration in the spring of 2006. They were captured at

two successive traps located at the lower end of the river system below all sites where YOY were marked. At both facilities, their individual PIT tags were identified. Eventually, PIT tagged anadromous salmon were recaptured in 2007 and 2008 when returning into the Scorff river after one or two years at sea. They were sampled at the Princes Mill facility in a trap designed to catch upstream migrating adults. PIT tagged resident juveniles, i.e., future two years old smolts, were identified in autumn 2006 using sampling by electrofishing according to same protocol used for the YOY the previous year. Two year old smolts we also recaptured the following spring (2007) and were also identified by their PIT tags.

Here, we considered the set of YOY juveniles marked in autumn 2005 and recaptured later on (N=124). For each of them, we observed its alternative phenotype (migrant vs. resident) and its observable cue (fork length at first autumn).

Modeling

For each individual *i*, given the observed cue (fork length at first autumn) Fl_i, the proximate cue η_i (energetic status) was assumed to be normally distributed with mean Fl_i and standard deviation σ_η (Eqn [3]). For each individual *i*, the alternative phenotype indicator Y_i (Eqn. [1]) takes the value 1 if the individual migrates to sea at one year of age, and 0 if it stays an additional year in fresh water.

We considered two scenarios as in the simulation study. First, we assumed we have no information available about the structure of the threshold variability. Second, we assumed that YOY captured in the same station in autumn were closely related genetically and had the same threshold value. The latter hypothesis was unrealistic but was used as an (extreme) illustration of the likely greater genetic similarity of YOY salmon within a site than between sites. Indeed, YOY juveniles tend to stay close to their natal spawning nest (Beall 1994; Einum et al. 2008; Foldvik et al. 2010).

RESULTS

Simulation study

The comparison of posterior to prior distributions suggested that the information contained in the data led to considerable updating of the prior distributions. Whether genetic information was included or not, the LETM properly estimated the threshold mean μ_θ and the total variance σ_T^2 ; the true value of these parameters were close to their posterior median (fig. 3). As expected, κ was not identifiable when the genetic data was lacking; its posterior distribution was the same as its prior distribution. When genetic information was included, κ became identifiable as its posterior distribution was much narrower than the prior and the true value was very close to the posterior median (Fig. 3). The standard deviations of the threshold σ_θ and of the proximate cue σ_η were poorly estimated without the genetic information. Although the true values fell within their corresponding 95% PPI (Posterior Probability Interval (PPI) is defined as the posterior probability that the true value of the parameter lies in an interval of two given values with probability 0.95; Congdon, 2006), they were at the edge of the posterior inter-quartile range. The posterior distributions of σ_θ and σ_η were the same, reflecting the essentially uniform posterior distribution of κ. With the inclusion of genetic information, the posterior distributions of σ_θ and σ_η were substantially smaller, and the posterior medians coincided with the true values. Moreover, the posterior distributions of σ_θ and σ_η were well contrasted indicating that the parameters were identifiable, given the inclusion of genetic information.

[Figure 3 about here]

Whether genetic information was included or not, the proximate cue η_i and the threshold θ_i estimates were estimated with little bias, the latter falling within or being close to the 95% PPI (Fig. 4). The uncertainty in the threshold estimates was much reduced when integrating genetic information while the proximate cue estimates remained almost unchanged.

[Figure 4 about here]

Case study

Using the LETM framework, we were able to obtain precise estimates of the mean latent threshold μ_θ and the total variance σ_T^2 (Fig. 5). The posterior distributions of these parameters were little affected by the addition or exclusion (pseudo) genetic information. The ratio κ was identifiable only when the genetic information was included, although it remained imprecisely estimated. In this case, the variance of the threshold corresponds approximately to 30% of the total variance σ_T^2. Recall that σ_T^2 would have been equated with the threshold variance if the observed vs. proximate cue distinction was ignored. The posterior distributions of standard deviations of the proximate cue σ_η and of the threshold σ_θ were narrowed down when incorporating the (pseudo) genetic information. The posterior distributions of σ_θ and σ_η were well contrasted indicating that the identifiability issue was alleviated.

[Figure 5 about here]

Estimates of the proximate cue and of the threshold at the individual level were also obtained (Fig. 6). YOY salmon appeared much more variable in the proximate cue than in their threshold. As for the simulated data, the uncertainty of the threshold estimates was reduced when integrating (pseudo) genetic information while the proximate cue estimates remained almost unchanged.

The proximate cue is a conceptual quantity and as such its scale is arbitrary. Here, given the measurement error structure of the LETM, its scale is the same as that of the observed cue. Then a proximate cue of say 90 can be interpreted as the mean energetic status of a YOY of 90 mm fork length in autumn. For the same reason, the mean threshold μ_θ can be either interpreted as the energetic status (proximate cue; eq. [5]) or the fork length in autumn (observed cue; Eq. [6]) of a YOY salmon having equal odds to become migrant or resident.

[Figure 6 about here]

DISCUSSION

Conditional strategies are the most common form of discrete phenotypic plasticity (phenotype or tactics) within species (Gross 1996). Understanding how these strategies evolve and are maintained by natural selection is a challenge. Tomkins and Hazel (2007) critically reviewed the theoretical models that have been proposed to understand the evolution of phenotypic plasticity in the conditional strategy framework. They argued that

the ETM is “the best model available currently for understanding the evolution and maintenance of conditional strategies” because it accounts for both genetic variation and environmental cues that affect phenotype expression. Surprisingly, the ETM has little been used to study adaptive phenotypic plasticity either in the wild (with theoretical approach; Edeline 2007) or under controlled experimental conditions (Ostrowski et al. 2000). In this paper, we propose a statistical model extension of the ETM, the Latent ETM, which should make the assessment of conditional strategies from observational data easier. The originality of our proposal lies in the proximate vs. observable cue distinction which accounts for the underlying proximate mechanism triggering the phenotype expression (by means of a Berkson type measurement error structure). It allows the estimation of not only the parameters of the threshold distribution but also the proximate cue and the threshold at the individual level.

When data are available only for the alternative phenotypes and the observable cue, the LETM is not fully identifiable. However, this issue does not prevent the LETM from providing reliable estimates of the mean threshold and of the individual thresholds as shown by our simulation study (Fig. 3 and 4). To circumvent the identifiability issue, we recommend integrating supplementary information about the structure of the individual threshold variability. The threshold is considered as a heritable phenotypic trait being under polygenic control and can vary among individuals according to their genotype (Tomkins and Hazel 2007). The genetic component in the threshold variability has been evidenced by Ostrowski et al. (2000) and Piché et al. (2008) show that probabilistic maturation reaction norms vary genetically. Consequently, considering the genetic basis about the threshold trait, knowledge about the relative degree of similarity between individuals’ genotype (by means of kinship or pedigree) could inform about the structure of the threshold variability. We simulated genetic data in this way, by having batches of individuals with the same threshold. We showed that using this additional information

allows separation of the threshold variance from the proximate cue variance (Fig. 3 and 5) and can improve the precision of the individual threshold estimates (Fig. 4 and 6). Alternatively, the Bayesian approach with informative priors could be used to separate the threshold variance from the proximate cue variance (Congdon 2006).

Wrongly ignoring the distinction between proximate vs. observable cue would lead to overestimation of the variability of the latent threshold since the variance component of the proximate cue is falsely attributed to the variation in the threshold and thus artificially inflates the threshold variability. Even when the threshold variance cannot be separated from the proximate cue variance, the LETM protects against overestimating the threshold variability while acknowledging a greater uncertainty in this key parameter. This variability conditions the ability of the conditional strategy to evolve under selective pressure (Hazel et al. 2004). Hence, the LETM can be seen as a cautious approach preventing erroneous inference regarding the potential for an adaptive evolutionary response of the threshold components of reaction norms to environmental change.

Within the conditional strategy framework, the LETM opens up interesting prospects in the study of phenotypic plasticity from observational data. Our new approach is a generic tool and could be applied to a wide range of taxa and to different forms of conditional strategies, including the induction of defences against predators (Hammill et al. 2008), polyphenic traits in insects (Moczek 2010; Tomkins & Moczek 2009), sex-ratio investment, filial cannibalism (Takeyama et al. 2006) and alternative reproductive tactics (Piché et al. 2008; Pitnick et al. 2009; Gross 1996). For the sake of simplicity, only binary traits were considered in this paper. However, the extension to more than two phenotypes can be envisaged by considering multiple thresholds (Gianola 1982; Sorensen et al. 1995).

Further development of the means for including additional information, including genetic data, will be highly profitable. The LETM could be adapted to take advantage of the knowledge about the relatedness of individuals in a population in a more realistic way

than we did in the simulation study. Pedigree data could be incorporated to account for the genetic structure of the threshold and consequently, to ensure model identifiability and to improve estimation precision. Coupling the LETM with quantitative genetic methods (such as the animal model; Kruuk 2004, Wilson et al. 2010) would allow the estimation of the genetic and environmental components of phenotypic variance of the threshold (i.e., its heritability) and a better assessment of the evolutionary potential of the conditional strategy. In a context of rapid and global environmental change, both evolution and plasticity can be critical for species adaptation (Gienapp et al. 2008). The joint appraisal of both phenomena from observational data is required, for which the use and further developments of the LTEM should help.

Acknowledgments

We thank N. Jeannot (INRA, U3E, Pont-Scorff), J. Rives, F. Lange, and F. Guéraud (INRA, Ecobiop, St Pée sur Nivelle), Y. Guilloux (Federation de pêche du Morbihan, Pont-Scorff) and other technical staff members for their help in collecting field data.

Mathieu Buoro and Olivier Gimenez were supported by a grant from the French Research National Agency (ANR), reference ANR-08-JCJC-0028-01.

Literature cited

Acolas, M.L. et al., 2007. Laboratory experiment on survival, growth and tag retention following PIT injection into the body cavity of juvenile brown trout (Salmo trutta). *Fisheries Research*, 86(2-3), 280-284.

Beall, E., 1994. Dispersal patterns and survival of Atlantic salmon (Salmo salar L.) juveniles in a nursery stream. *ICES Journal of Marine Science*, 51(1), 1-9.

Brooks, S. & Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455.

Congdon, P., 2006. *Bayesian Statistical Modelling, 2nd edition.* Wiley seri.,

Dieckmann, U. et al., 2007. Probabilistic maturation reaction norms: their history, strengths, and limitations. *Disentangling the causes of maturation trends in exploited fish populations*, 335, 253–269.

Edeline, E., 2007. Adaptive phenotypic plasticity of eel diadromy. *Marine Ecology Progress Series*, 341, 229-232.

Einum, S. et al., 2008. Nest distribution shaping within-stream variation in Atlantic salmon juvenile abundance and competition over small spatial scales. *Journal of Animal Ecology*, 77, 167-172.

Elliott, J. & Hurley, M., 1997. A functional model for maximum growth of Atlantic salmon parr, Salmo salar, from two populations in northwest England. *Functional Ecology*, 11(5), 592–603.

Ellison, A.M., 2004. Bayesian inference in ecology. *Ecology Letters*, 7(6), 509-520.

Foldvik, A., Finstad, A.G. & Einum, S., 2010. Relating juvenile spatial distribution to breeding patterns in anadromous salmonid populations. *Journal of Animal Ecology*, 79, 501-509.

Forseth, T. et al., 2001. Functional models for growth and food consumption of Atlantic salmon parr, Salmo salar, from a Norwegian river. *Freshwater Biology*, 46(2), 173-186.

Gelman, A. et al., 2004. *Bayesian Data Analysis, Second Edition*, New York: Chapman & Hall/CRC.

Ghalambor, C. et al., 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Ecology*, 21, 394–407.

Gianola, D., 1982. Theory and analysis of threshold characters. *Journal of Animal Science*, 54(5), 1079.

Gienapp, P. et al., 2008. Climate change and evolution: disentangling environmental and genetic responses. *Molecular ecology*, 17(1), 167-78.

Gross, M.R., 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology & Evolution*, 11(2), 92-97.

Guéguen, J. & Prouzet, P., 1994. *Le saumon atlantique* Quae. IFREMER,

Hammill, E., Rogers, a. & Beckerman, a.P., 2008. Costs, benefits and the evolution of inducible defences: a case study with Daphnia pulex. *Journal of evolutionary biology*, 21(3), 705-15.

Hazel, W., Smock, R. & Johnson, M., 1990. A polygenic model for the evolution and maintenance of conditional strategies. *Proceedings of the Royal Society B*, 242(1305), 181–187.

Hazel, W., Smock, R. & Lively, C.M., 2004. The ecological genetics of conditional strategies. *Am. Nat*, 163, 888-900.

Hutchings, J.A., 2004. Norms of Reaction and Phenotypic Plasticity in Salmonid Life Histories. In A. H. Stearns *Evolution Illuminated: Salmon and Their Relatives*. Oxford: Oxford University Press, pp. 154-174.

Imre, I., Grant, J. & Cunjak, R.A., 2005. Density-dependent growth of young-of-the-year Atlantic salmon in Catamaran Brook, New Brunswick. *Journal of Animal Ecology*, 74, 508-516.

Jones, W. et al., 2002. Seasonal patterns of growth, expenditure and assimilation in juvenile Atlantic salmon. *Journal of Animal Ecology*, 71(6), 916-924.

Jonsson, B. & Jonsson, N., 2005. Lipid energy reserves influence life-history decision of Atlantic salmon (Salmo salar) and brown trout (S. trutta) in fresh water. *Ecology of Freshwater Fish*, 14(3), 296-301.

Lynch M., Walsh B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates: Sunderland, MA.

Mangel, M. & Satterthwaite, W.H., 2008. Combining proximate and ultimate approaches to understand lufe history variation in salmonids with application to fisheries, conservation and aquaculture. *Bulletin of marine science*, 83(1), 107-130.

McCarthy, M.A., 2007. *Bayesian Methods for Ecology* Cambridge., Cambridge, UK.

Mccormick, S.D. et al., 1998. Movement , migration , and smolting of Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, 55, 77-92.

Metcalfe, N., 1998. The interaction between behavior and physiology in determining life history patterns in Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, 55, 93-103.

Moczek, A.P., 2010. Phenotypic plasticity and diversity in insects. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1540), 593-603.

Murphy, M.H., Connerton, M.J. & Stewart, D.J., 2006. Evaluation of winter severity on growth of young-of-the-year atlantic salmon. *Transactions of the American Fisheries Society*, 135(2), 420-430.

Nicieza, A., Brana, F. & Toledo, M., 1991. Development of length-bimodality and smolting in wild stocks of Atlantic salmon, Salmo salar L., under different growth conditions. *Journal of Fish Biology*, 38(4), 509–523.

Ostrowski, M.F., Jarne, P. & David, P., 2000. Quantitative genetics of sexual plasticity: the environmental threshold model and genotype-by-environment interaction for phallus development in the snail Bulinus truncatus. *Evolution*, 54(5), 1614-25.

Piché, J., Hutchings, J. & Blanchard, W., 2008. Genetic variation in threshold reaction norms for alternative reproductive tactics in male Atlantic salmon, Salmo salar. *Proceedings of the Royal Society B: Biological Sciences*, 275(1642), 1571.

Pitnick, S. et al., 2009. Size-dependent alternative male mating tactics in the yellow dung fly, Scathophaga stercoraria. *Proceedings of the Royal Society B*, 276(1671), 3229-37.

Roff, D.A., 1997. *Evolutionary Quantitative Genetics*, New York: Chapman & Hall.

Satterthwaite, W. et al., 2009. Steelhead Life History on California's Central Coast: Insights from a State-Dependent Model. *Transactions of the American Fisheries Society*, 138(2007), 532–548.

Sorensen, D. et al., 1995. Bayesian inference in threshold models using Gibbs sampling. *Genetics, Selection, Evolution: GSE*, 27(3), 229.

Spiegelhalter, D.J. et al., 2003. *WinBUGS user manual. Version 1.4. MRC Biostatistics Unit, Cambridge, UK.*, Cambridge, UK.

Takeyama, T., Okuda, N. & Yanagisawa, Y., 2006. Filial cannibalism as a conditional strategy in males of a paternal mouthbrooding fish. *Evolutionary Ecology*, 21(1), 109-119.

Thorpe, J. & Metcalfe, N.B., 1998. Is smolting a positive or a negative developmental decision? *Aquaculture*, 168(1-4), 95-103.

Thorpe, J. et al., 1998. Modelling the proximate basis of salmonid life-history variation, with application to Atlantic salmon, Salmo salar L. *Evolutionary Ecology*, 12(5), 581–599.

Tomkins, J. & Hazel, W., 2007. The status of the conditional evolutionarily stable strategy. *Trends in Ecology & Evolution*, 22, 522-528.

Tomkins, J.L. & Moczek, A.P., 2009. Patterns of threshold evolution in polyphenic insects under different developmental models. *Evolution*, 63(2), 459-68.

Wilson, A.J. et al., 2010. An ecologist’s guide to the animal model. *Journal of Animal Ecology*, 79, 13-26.

Figure captions

Figure 1: Directed Acyclic Graph of the Latent Environmental Threshold Model (LETM). Observable data are presented in squares and unknown quantities to be estimated are in circles. For an individual *i*, the threshold θ_i is normally distributed with mean μ_θ and standard deviation σ_θ. Its proximate cue η_i is assumed to be normally distributed around the observable cue X_i with standard deviation σ_η. Finally, Y_i is a binary indicator variable of the observed phenotype and is modeled as a joint function of the threshold and the proximate cue at the individual level, such that $Y_i = 1$ when $\theta_i < \eta_i]$. Solid and broken arrows stand for stochastic and logical dependence, respectively**.** The model is fit over observations at phenotyped individuals, i.e., the boxes denote a loop over i=1,2, ... N.

Figure 2: Representation of the simulated data for 200 individuals (circles). Upper panel: correlation between the observed cue X_i and the proximate cue η_i. Bottom panel: relationship between the phenotype indicator Y and the proximate cue η_i. The empirical distribution of thresholds θ is also shown (histogram).

Figure 3: Posterior distributions of Latent Environmental Threshold Model (LETM) parameters with and without genetic information for simulated data (based on 50000 MCMC samples). The 2.5, 25, 50 (median), 75, 97.5 percentiles are displayed. The actual values are also displayed (dashed lines).

Figure 4: Posterior distributions of proximate cue η_i and threshold θ_i for one individual picked in each of the 20 batch considering the genetic information (right column) or not (left column) (based on 50000 MCMC samples). The 2.5, 25, 50 (median), 75, 97.5 percentiles are displayed. The actual values are also displayed (stars).

Figure 5: Posterior distributions of latent Environmental Threshold Model (LETM) parameters for Atlantic salmon data with and without (pseudo) genetic information (based on 50000 MCMC samples). The 2.5, 25, 50 (median), 75, 97.5 percentiles are displayed.

Figure 6: Posterior distributions of proximate cue η_i and threshold θ_i for one individual picked in each of the 20 batches considering the (pseudo) genetic information (right column) or not (left column) (based on 50000 MCMC samples). The 2.5, 25, 50 (median), 75, 97.5 percentiles are displayed.

FIGURES

FIGURE 1

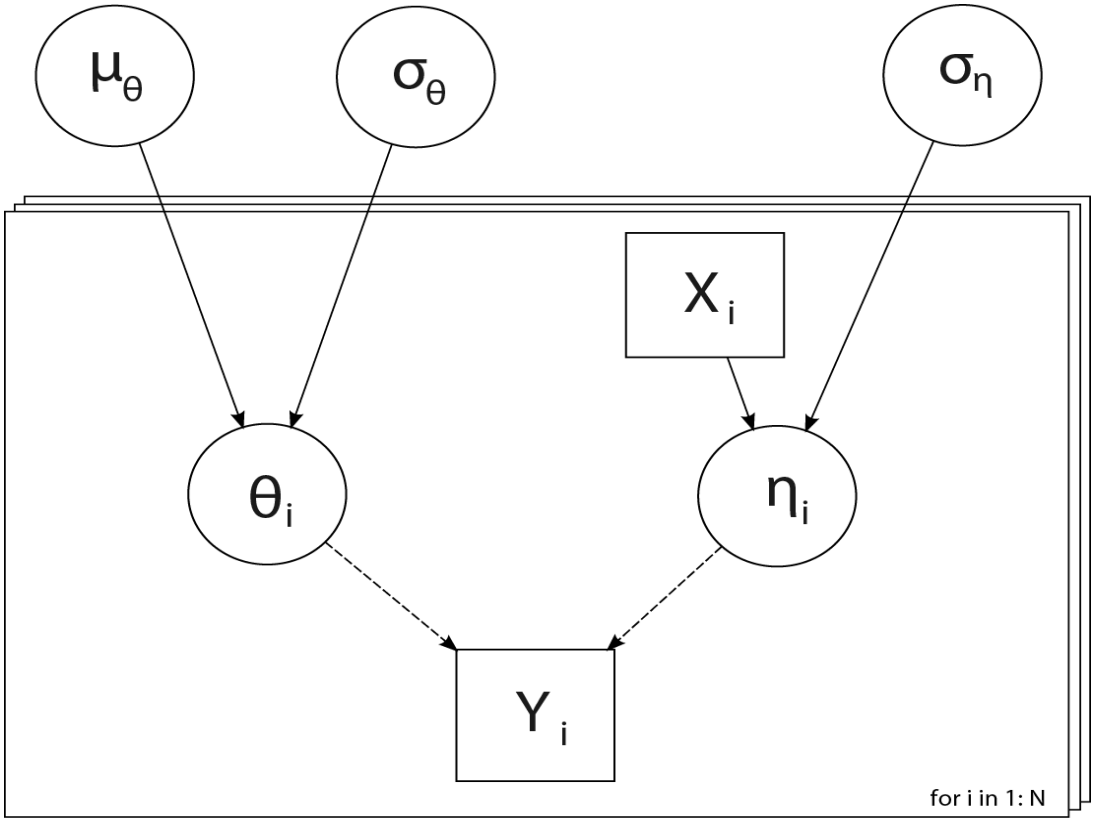


FIGURE 2

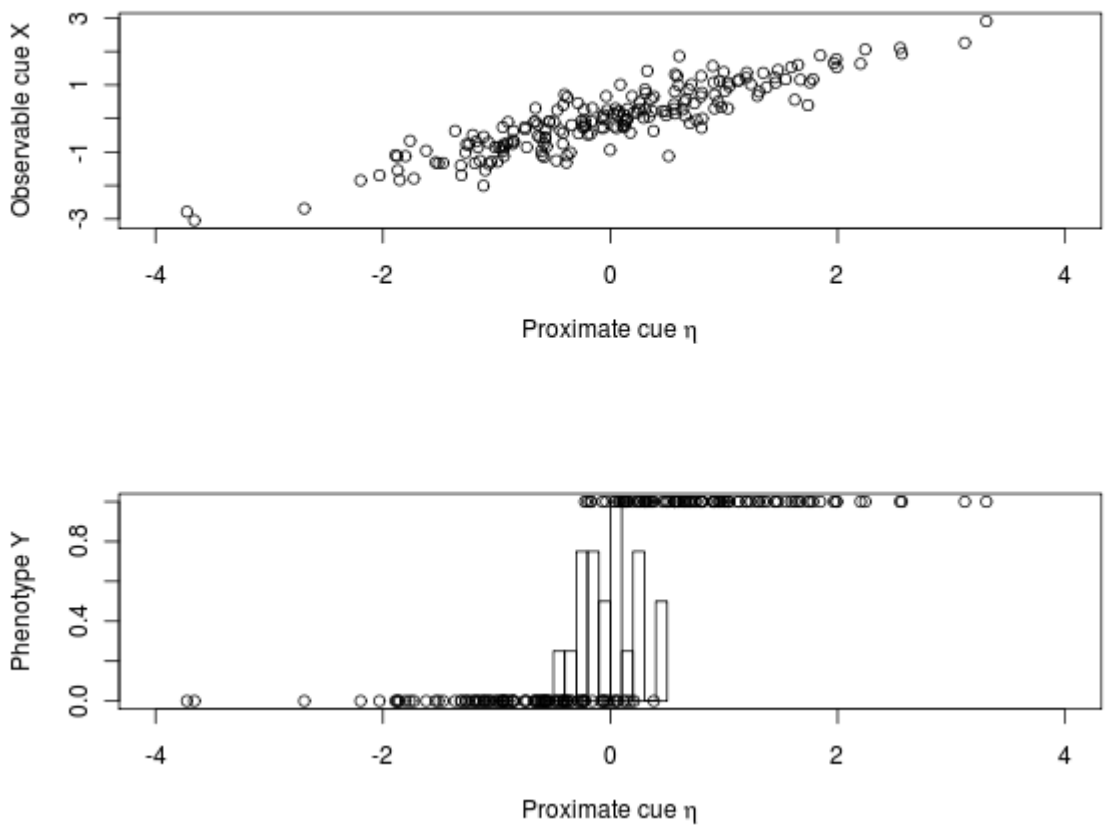

Observable cue X
3
1
-1
-3
-4
-2
0
2
4
Proximate cue η
Phenotype Y
0.8
0.4
0.0
-4
-2
0
2
4
Proximate cue η

FIGURE 3

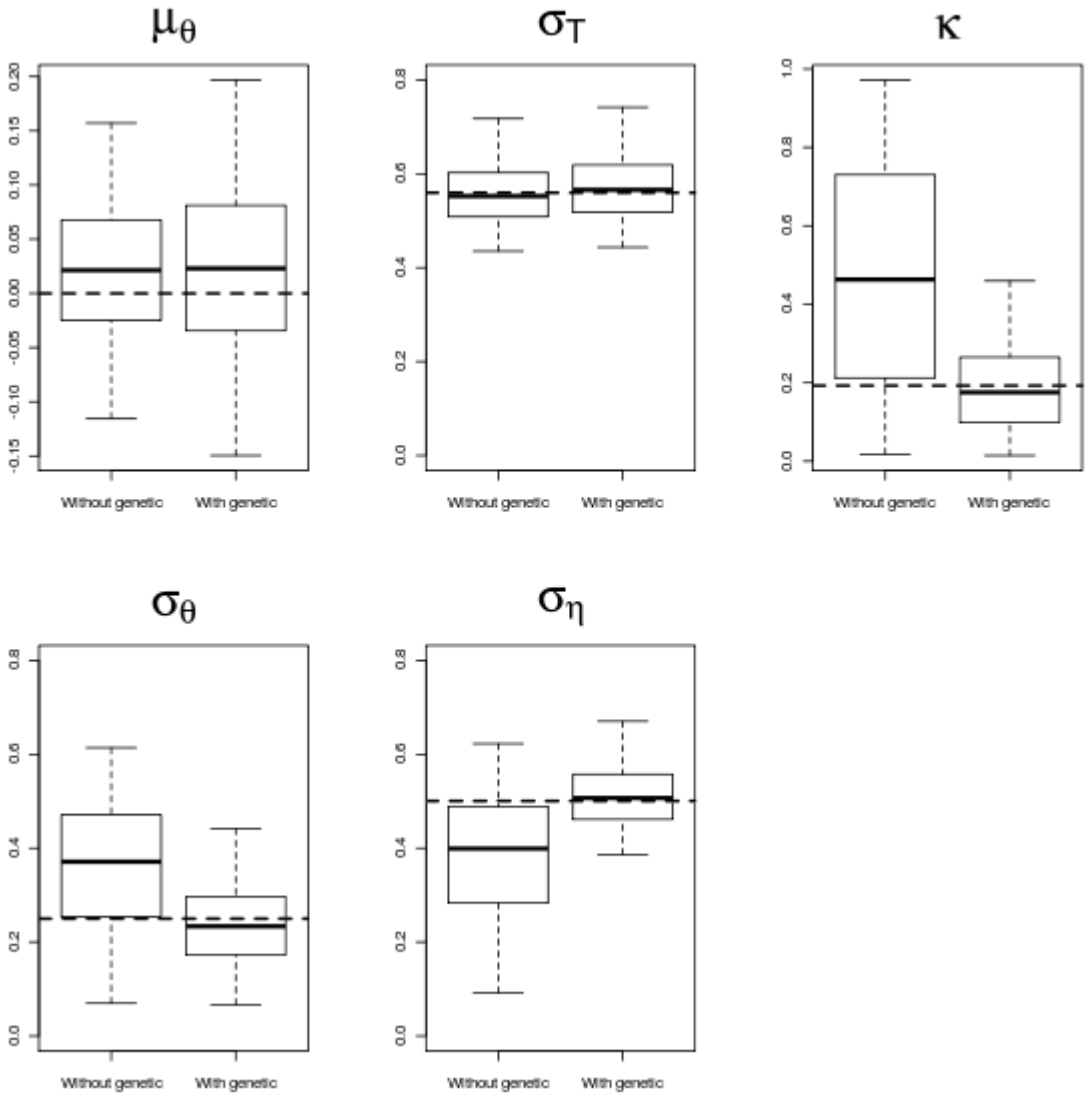

μ_θ
σ_T
κ
σ_θ
σ_η
Without genetic
With genetic

FIGURE 4

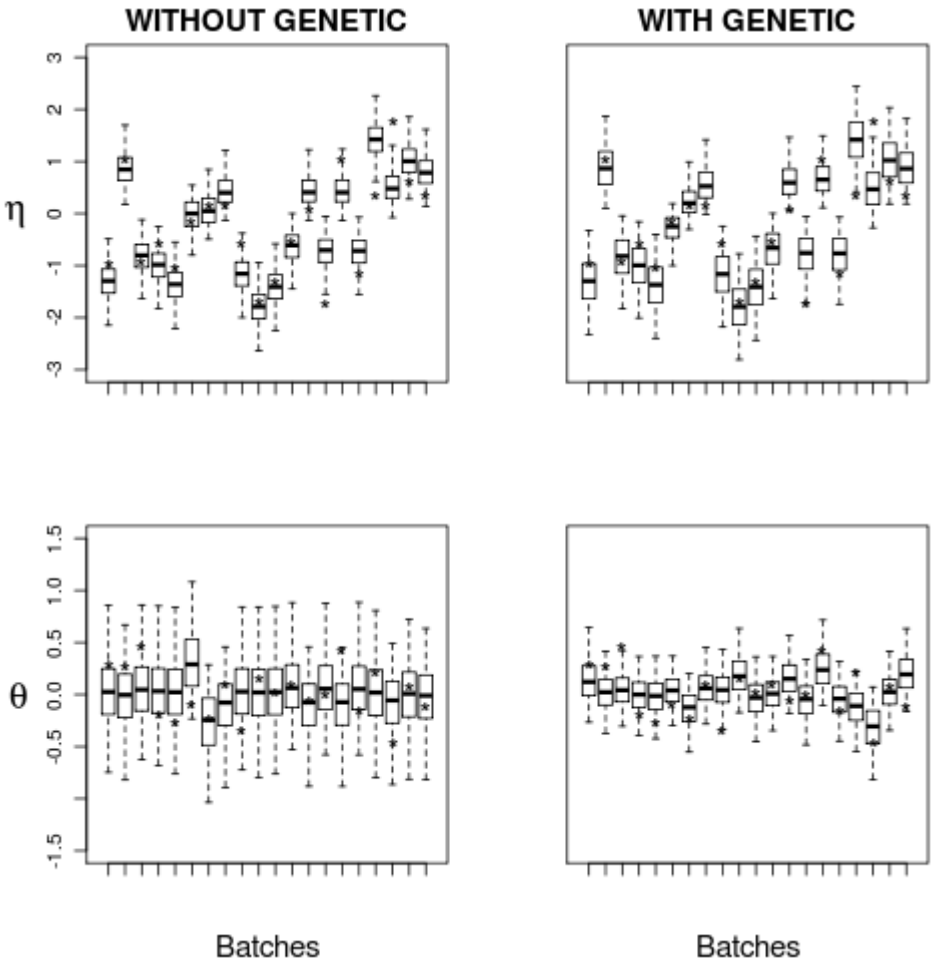

WITHOUT GENETIC
WITH GENETIC
η
θ
3
2
1
0
-1
-2
-3
1.5
1.0
0.5
0.0
-0.5
-1.5
Batches
Batches

FIGURE 5

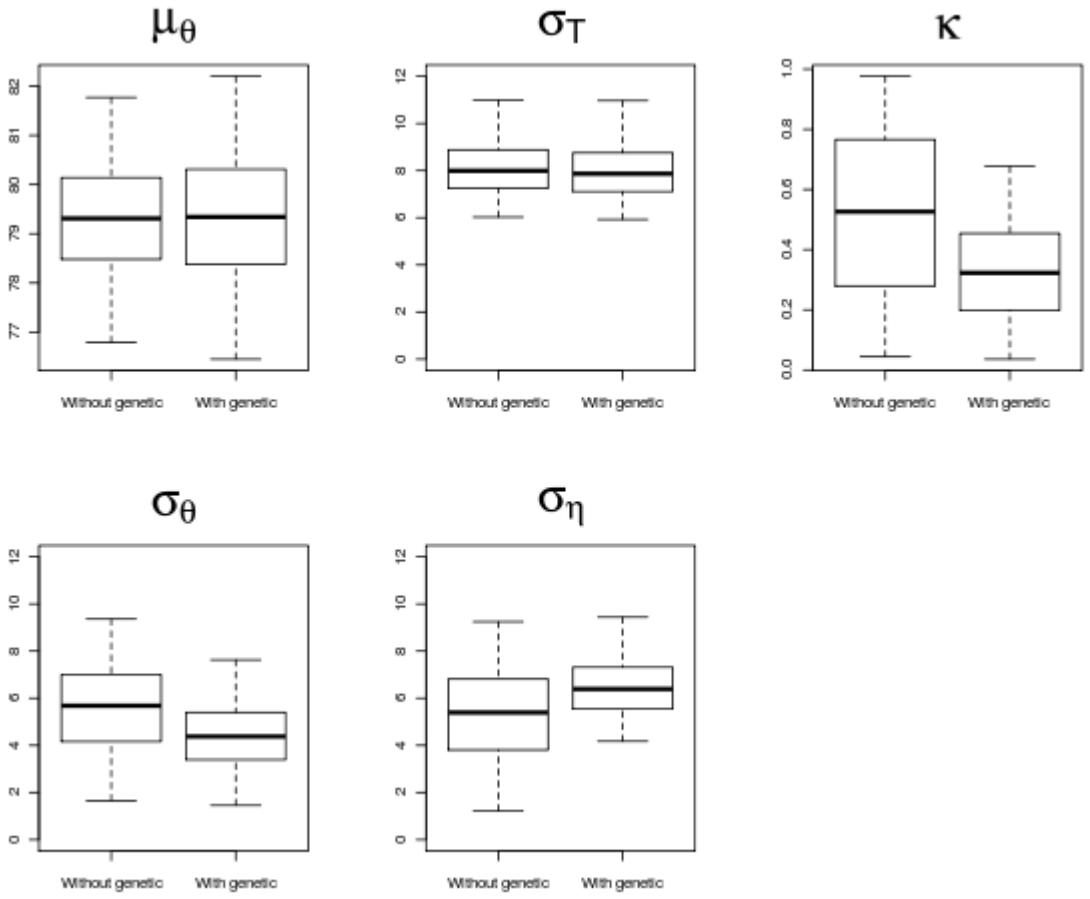
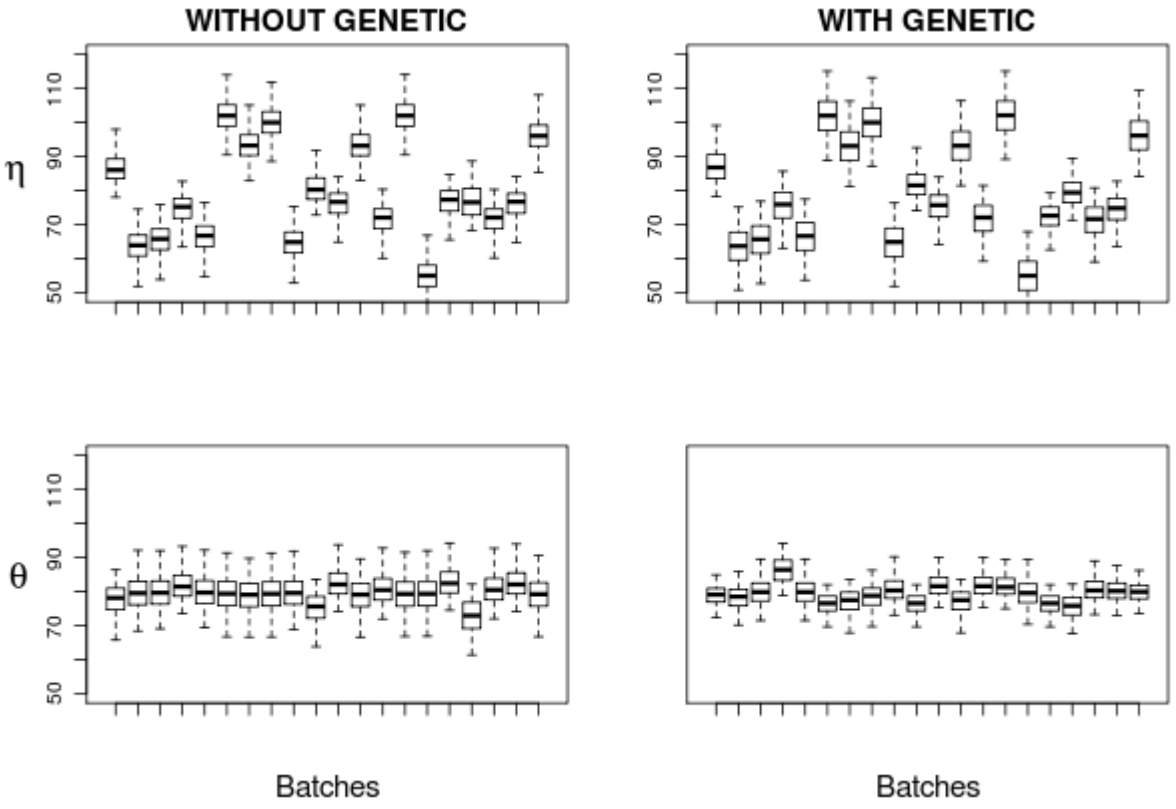

μ_θ
σ_T
κ
σ_θ
σ_η
Without genetic
With genetic

FIGURE 6


WITHOUT GENETIC
WITH GENETIC
η
θ
50
70
90
110
Batches
Batches