



**HAL**  
open science

## Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal studies

Sandra Hamel, Nigel Yoccoz, Jean-Michel Gaillard

► **To cite this version:**

Sandra Hamel, Nigel Yoccoz, Jean-Michel Gaillard. Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal studies. *Methods in Ecology and Evolution*, 2012, 3 (4), pp.731-742. 10.1111/j.2041-210X.2012.00195.x . hal-02296820

**HAL Id: hal-02296820**

<https://univ-lyon1.hal.science/hal-02296820v1>

Submitted on 22 Aug 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal studies

Sandra Hamel<sup>1\*</sup>, Nigel G. Yoccoz<sup>1</sup> and Jean-Michel Gaillard<sup>2</sup>

<sup>1</sup>*Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, University of Tromsø, 9037 Tromsø, Norway; and* <sup>2</sup>*Université de Lyon, Université Lyon 1, CNRS, UMR 5558 “Biométrie et Biologie Evolutive”, F-69622 Villeurbanne, France*

## Summary

1. Autocorrelation and individual heterogeneity are now considered to reflect biological processes rather than simply being a nuisance requiring to be accounted for. Before using parameter estimates that represent autocorrelation and individual heterogeneity to infer biological processes, a statistical evaluation of their precision and accuracy is required to validate their use.
2. Using simulated data, we evaluated accuracy and precision of temporal autocorrelation and individual heterogeneity estimates provided by different statistical models. We compared estimates across different intensity of individual variation and life histories, and sampling effort. We focused on recurrent binary variables because statistical evaluations of models describing binary processes have often been overlooked although several evolutionary and ecological processes are expressed as binary variables (e.g. probability of annual reproduction, plant annual flowering and detection, seasonal migration decision).
3. Our results showed that autocorrelation and individual heterogeneity were generally better estimated using a ‘time series’ modelling approach, but that a ‘state dependence’ modelling approach also provided fair estimates in most cases. The latter method was even more robust when data sets included missing values. Data sets including missing values or consisting of very short times series resulted in important bias in some instances.
4. Models ignoring either individual heterogeneity or autocorrelation performed poorly, illustrating the fundamental association between these two processes, and demonstrating that the complex structure of autocorrelation and individual heterogeneity patterns is difficult to tackle using simple models.
5. Our work’s major finding is the demonstration that autocorrelation and individual heterogeneity need to be both accounted for to provide reliable estimates even in studies focusing on only one of these processes. Our study also offers a set of practical recommendations for helping researchers modelling these two processes depending on their scientific aims and the structure of their data. Finally, our results illustrate that more research is required for estimating individual heterogeneity when positive temporal autocorrelation is expected because none of the models evaluated provided suitable estimates.

**Key-words:** accuracy, first-order autocorrelation, generalized linear mixed models, individual heterogeneity, precision, random intercept model

## Introduction

Autocorrelation and individual heterogeneity are now often included in ecological and evolutionary studies to account for

the potential confounding effects resulting from the non-independence among the repeated measures collected, be it at the individual, temporal or spatial level (van Noordwijk & de Jong 1986; Diniz-Filho, Bini & Hawkins 2003; van de Pol & Verhulst 2006). In life-history studies, evolutionary ecologists usually work with data sets consisting of repeated measures of a set of individuals taken at different ages. This results in

\*Correspondence author. E-mail: sandra.hamel@uit.no

individual heterogeneity, that is, among-individual variation leading to the non-independence of the repeated measures collected on the same individuals, as well as temporal dependency. Dependency can take many forms. In this case, dependency occurs between values of the same variable measured at different moments through time, which is referred to as temporal autocorrelation. For example, repeated measures on individual mass in iteroparous species often result in heterogeneity, because larger individuals will usually remain heavier than smaller individuals (Nussey *et al.* 2011). Temporal autocorrelation in individual mass can also occur, for example, as a result of allocation to reproduction: breeders are often lighter than non-breeders at the next reproductive attempt and are also more likely to skip reproduction to rebuild their condition (Pomeroy *et al.* 1999), leading to temporal oscillation in individual mass. In botany, plant population demography is often inferred from presence/absence data on marked individuals, which are recorded at specific time intervals over several years (Kéry & Gregg 2003; Shefferson *et al.* 2011). The plant detection probability, however, can be affected by the specific characteristics of each plant, so that some individuals are consistently more likely to be detected than others at each census, leading to individual heterogeneity. Previous detection can also affect future detection and lead to temporal autocorrelation (Kéry & Gregg 2003). Similar situations also occur in behavioural ecology, when, for example, animal personality and previous capture history both affect the recapture probability. Bold animals are more likely to be captured/recaptured than shy individuals (Réale *et al.* 2000), but an individual recently captured, be it bold or shy, is less likely than another individual to be recaptured in the next session (Boyer *et al.* 2010).

Although statistically taking into account the confounding effects of autocorrelation and individual heterogeneity is essential, these effects also have an important biological meaning. For instance, works by van de Pol and co-authors (van de Pol & Verhulst 2006; van de Pol & Wright 2009) brought the biological importance of individual heterogeneity a step further. They demonstrated that heterogeneity takes place at two different levels (i.e. between- and within-individuals) with distinct biological meaning. For example, the between-individual variance estimated for a life-history trait recorded at each age usually represents selection processes, resulting from the appearance/disappearance of individuals, whereas the within-individual variance measures growth/senescence processes (van de Pol & Verhulst 2006). These clarifications illustrate the wide potential of the different measures of individual heterogeneity to answer various ecological and evolutionary questions that still remain largely unexplored.

As temporal autocorrelation and individual heterogeneity hold a strong potential to enhance biological knowledge, it is essential to distinguish between the definition of these concepts and the statistical ways of measuring them. Although the validity of the models incorporating autocorrelation and individual heterogeneity has sometimes been assessed, this has mainly been evaluated for fixed parameters, often using specific autocorrelation and individual heterogeneity values (see e.g.

Masaoud & Stryhn 2010). No study, to our knowledge, has yet focused on the estimation of autocorrelation and individual heterogeneity and thereby evaluated whether the methods used to define these two parameters provide accurate measures of their strength and hence confidence in their use. Furthermore, previous statistical evaluations of models including autocorrelation or individual heterogeneity have been performed for linear models (LM; see e.g. van de Pol & Verhulst 2006; Martin *et al.* 2011). Nevertheless, data available in ecology and evolution often fall outside this linear framework, commonly consisting of binary, proportion or counts data, and thus often require to be modelled through generalized LMs (GLM; Bolker *et al.* 2009). Parameters estimated from GLM, however, are not easy to standardize. For instance, the variance of a Bernoulli/binomial process is maximized at a mean probability of 0.5 and is constrained towards 0 as the mean probability approaches 0 or 1 (Gaillard & Yoccoz 2003). Thus, the comparison of parameter estimates is difficult when the mean probability differs among the binary variables to be compared. Consequently, performing comparative analyses of autocorrelation and individual heterogeneity in ecology and evolution is complex. Accurate and standardized measures of the strength of autocorrelation and individual heterogeneity should be highly valuable for determining the relative importance of these two processes and would allow comparing the relative contribution of these processes among studies. Because relying on appropriate metrics and reliable statistical indicators is of prime importance, a statistical assessment of the reliability of these two measures would, therefore, provide an important step forward in ecology and evolution.

Here, our goals are to evaluate the reliability of estimates provided by different statistical models for describing temporal autocorrelation and individual heterogeneity and to develop a standardization method that would allow comparing the strength of these two processes among studies. First, we describe parameters included in different models that are frequently used in evolutionary ecology and explain which parameters could reliably measure the amount of autocorrelation and individual heterogeneity. We then simulate longitudinal data with different levels of autocorrelation and individual heterogeneity to assess the reliability of each model in estimating these parameters. We simulate data on individuals within populations having different mean trait values, number of individuals, average life span, as well as number of missing values, to determine the influence of the data structure on the accuracy and the precision of the autocorrelation and individual heterogeneity measures. We then compare results across all models to provide guidelines for evaluating autocorrelation and individual heterogeneity according to the distribution of the variables and the structure of the data available to the researcher. We mainly focus on recurrent variables – as opposed to non-recurrent variables like mortality – that follow a Bernoulli process (hereafter binary variables) because numerous evolutionary and ecological processes are expressed as binary variables (e.g. annual reproduction, plant annual flowering and detection, seasonal migration decision), and because statistical evaluations of models describing binary processes have been neglected.

In addition, modelling and interpreting parameter estimates are more complex for binary than for LMs. Nevertheless, we compare how results from binary models compared with those of simpler, LMs.

## Materials and methods

### STATISTICAL MODELS AND PARAMETERS EVALUATED

The statistical models we analysed can be divided into two general types that we will refer to as ‘state dependence’ and ‘time series’. From these starting models, we derived the other models that are basically a simplification by removing either the autocorrelation or the individual variability.

#### ‘State dependence’ models

First, we will evaluate the occurrence of autocorrelation using what we will refer to as ‘state dependence’ models, which are commonly used by econometricians to model first-order autocorrelation and heterogeneity (Heckman 2001; Wooldridge 2005; Berridge & Crouchley 2011). These models assess the influence of the state of a variable in the past at time  $t$ , on the expression of the same variable at the current time (i.e. modelling dependence on previous state). In ecology and evolution, the autocorrelation is often assessed with only one time lag ( $t = 1$ ), for example, the survival of offspring produced by a female in a given year in relation with the survival of offspring produced by the same female the previous year. The sample, therefore, consists of observations from  $n_{ID}$  individuals with  $n_{AGE}$  repeated measurements over the lifetime (i.e. at each age), and the response and explanatory variables take the form of  $y_{ij}$  and  $x_{A_j}$ , respectively. When the response follows a Bernoulli process, this model takes the form of a multilevel logistic regression:

$$\begin{aligned} \text{logit}(\Pr(y_{ij} = 1 | \alpha_i)) &= \beta_0 + \beta_T y_{i,j-1} + f(x_{A_j}) + \alpha_i \quad (\text{for } j = 1, \dots, n_{AGE}) \\ \alpha_i &\sim N(0, \sigma_{ID}^2) \quad (\text{for } i = 1, \dots, n_{ID}) \end{aligned} \quad \text{eqn 1}$$

Although this type of model can allow for the inclusion of covariates, we simply included the effect of age that can take various functional forms,  $f(x_{A_j})$ . This model includes a random intercept  $\alpha_i$  that follows a Gaussian distribution with a mean of 0 and a variance  $\sigma_{ID}^2$ , representing heterogeneity among individuals  $i$  (also called between-subject effect, see van de Pol & Wright 2009). This random effect allows fitting a model with a different intercept for each individual  $i$ , therefore, allowing the estimation of the intercept  $\beta_0$  to vary according to individual heterogeneity and hence accounting for bias (either of estimates or of their uncertainty) appearing when  $\sigma_{ID}^2 > 0$  and the sample includes repeated measures. Here, we evaluated individual heterogeneity estimates based on random intercept models rather than random slope and intercept models because we were interested in assessing between-individual variance effect on the dependent variable (but see Appendix S1 and van de Pol & Wright (2009) for more details on random slope models).

The parameter of interest for measuring the dependence in eqn (1) is  $\beta_T$ . This coefficient measures the link between the variables while taking into account other covariates and individual heterogeneity. In a LM with standardized input and response variables,  $\beta_T$  would be equal to the correlation coefficient  $r$  estimated between the two variables (see Schielzeth 2010

p.108), so that if two variables are negatively correlated at  $-0.5$ , then  $\beta_T$  should be estimated at  $-0.5$ . In a binary framework, the interpretation of  $\beta_T$  is not straightforward because of the logit transformation of the data. Nevertheless, the relationship between a probability and its logit can provide some rules of thumb for the interpretation of  $\beta_T$  (see Gelman & Hill 2007 p.82). Indeed, the derivative of the logit function calculated for  $x = 0.5$  is 4, so the slope at the mid-point of the logistic curve equals 4. Hence, we can divide logistic regression parameters by 4 to obtain an approximation of their linear predictions (Gelman & Hill 2007). For probabilities ranging between 0.3 and 0.7, the relationship between the probability and its logit is almost linear. Because it is markedly nonlinear outside of this range, the slope increases as the probability approaches towards 0 and 1 (see Fig. S1), and hence the coefficient required to transform a binary parameter for it to compare with a linear one increases with higher and lower probabilities. To have binary parameters that compare with linear parameters, we could divide binary parameters by the derivative of the mean probability value ( $pv$ ). Thus, we evaluated whether using  $\beta_T * [f'(pv)]^{-1}$  would provide correlations that would compare with simulated correlations and hence with correlations obtained from LMs. In our case, we simulated mean probability ( $pv$ ) of 0.5, 0.7 and 0.9, and so we used values of  $f'(pv)$  equal to 4.00, 4.76 and 11.13, respectively.

The model in eqn (1) includes a random intercept accounting for individual heterogeneity, which requires to be accounted for (Gimeñez & Choquet 2010). This random effect, however, is sometimes removed when not statistically significant. To assess the influence of removing individual heterogeneity while trying to measure the strength of the autocorrelation, we also evaluated variation in parameter estimation for the following model, representing a simple logistic regression:

$$\text{logit}(\Pr(y_{ij} = 1)) = \beta_0 + \beta_T y_{i,j-1} + f(x_{A_j}) \quad (\text{for } j = 1, \dots, n_{AGE}) \quad \text{eqn 2}$$

#### ‘Time series’ models

Because variables used in longitudinal studies consist of repeated measures at successive life-history stages, data correspond to a sequential variation of a variable through time, such that variation of that variable with time forms a time series. For a continuous variable, for example, where the mass of an individual in a given year is put in relation with its mass the previous year, a time series model would evaluate changes in mass,  $y_{ij}$ , in relation to age,  $x_{A_j}$ , while assessing the autocorrelation  $\rho$  among residuals  $\varepsilon_{ij}$ . The linear mixed model with a first-order autoregressive error process with parameter  $\rho$  takes the following form:

$$\begin{aligned} y_{ij} &= \beta_0 + f(x_{A_j}) + \alpha_i + \varepsilon_{ij} \quad (\text{for } j = 1, \dots, n_{AGE}) \\ \alpha_i &\sim N(0, \sigma_{ID}^2) \quad (\text{for } i = 1, \dots, n_{ID}) \\ \text{Corr}[\varepsilon_{ij}, \varepsilon'_{ij}] &\sim \rho^{j-j'} \end{aligned} \quad \text{eqn 3}$$

For a binary variable, such as the example discussed at eqn (1) where survival of offspring produced by a female in a given year is put in relation with survival of offspring produced by the same female the previous year, a time series model would evaluate variation in offspring survival,  $y_{ij}$ , in relation to mother age,  $x_{A_j}$ , while assessing the autocorrelation  $\rho$  between the successive realizations  $y_{ij}$ . The difference with the LM of eqn (3) is that the autocorrelation is evaluated on the realizations  $y_{ij}$  rather than on the residuals  $\varepsilon_{ij}$  because a logistic

regression model is not defined in terms of distribution of residuals. The generalized linear mixed model with a first-order autoregressive error process modelling a binary variable takes the following form:

$$\begin{aligned} \text{logit}(\Pr(y_{ij} = 1 | \alpha_i)) &= \beta_0 + f(x_{Aj}) + \alpha_i \quad (\text{for } j = 1, \dots, n_{\text{AGE}}) \\ \alpha_i &\sim N(0, \sigma_{\text{ID}}^2) \quad (\text{for } i = 1, \dots, n_{\text{ID}}) \\ \text{Corr}[y_{ij}, y'_{ij}] &\sim \rho^{|j-j'|} \end{aligned} \quad \text{eqn 3a}$$

Equation 3a differs from eqn (1) by having the explanatory variable  $y_{i,j-1}$  removed and  $\beta_T$ , which previously linked  $y_{ij}$  and  $y_{i,j-1}$ , replaced with  $\rho$ . This model has the advantage of directly modelling the autocorrelation. To assess the influence of ignoring potential autocorrelation when trying to estimate individual heterogeneity for a variable measured repeatedly through time, we also considered a model without  $\rho$ :

$$\begin{aligned} \text{logit}(\Pr(y_{ij} = 1 | \alpha_i)) &= \beta_0 + f(x_{Aj}) + \alpha_i \quad (\text{for } j = 1, \dots, n_{\text{AGE}}) \\ \alpha_i &\sim N(0, \sigma_{\text{ID}}^2) \quad (\text{for } i = 1, \dots, n_{\text{ID}}) \end{aligned} \quad \text{eqn 4}$$

## SIMULATIONS

### *Variation in the population structure ( $n_{\text{ID}}$ , $n_{\text{AGE}}$ )*

To compare the accuracy and precision of the parameters estimated by the different models, we simulated binary data following a Bernoulli process, that is, simulating one value (0 or 1) for each individual age. We simulated data for representing a variable measured at successive ages. We simulated data sets consisting of 50, 200 or 1000 individuals ( $n_{\text{ID}}$ ) with a longevity ( $n_{\text{AGE}}$ ) of 5, 15 or 40 (see simulation details in Appendix S1), leading to nine combinations of population samples. These combinations were representative of population samples for (i) short-lived species, such as small passerines, rabbits or small mammals, which usually include a large number of individuals monitored over a small number of ages ( $n_{\text{ID}} = 1000$ ,  $n_{\text{AGE}} = 5$ ), (ii) relatively long-lived species, such as ungulates, snakes or lizards, which usually include a lower number of individuals monitored for longer ( $n_{\text{ID}} = 200$ ,  $n_{\text{AGE}} = 15$ ) and (iii) very long-lived species, such as primates, turtles or some seabirds, which usually include a small number of individuals monitored throughout their long life ( $n_{\text{ID}} = 50$ ,  $n_{\text{AGE}} = 40$ ). The combinations we selected also included less typical sample sizes, ranging from few individuals of short-lived species ( $n_{\text{ID}} = 50$ ,  $n_{\text{AGE}} = 5$ ) to large number of individuals of long-lived species ( $n_{\text{ID}} = 1000$ ,  $n_{\text{AGE}} = 40$ ).

### *Variation in the population parameters ( $\beta_0$ , $\beta_A$ , $\sigma_{\text{ID}}^2$ , $\rho$ or $\beta_T$ )*

In addition to varying the structure of the population data, we also simulated data sets with different parameter values: mean population

probability,  $\beta_0$ , age effect,  $\beta_A$ , individual heterogeneity,  $\sigma_{\text{ID}}^2$ , and first-order autocorrelation,  $\rho$  or  $\beta_T$ . As the variance of a Bernoulli process is maximized at a mean probability of 0.5 and is constrained towards 0 as the mean approaches 0 or 1 (Gaillard & Yoccoz 2003), we evaluated the model performance with different mean probabilities, that is, average (0.5), relatively high (0.7) and very high (0.9). We did not simulate very low (0.1) and relatively low (0.3) probabilities, as results would mirror results from 0.9 and 0.7 probabilities, respectively. We simulated weak and strong age effects, standard variation varying from none to very high (0–1.4) and negative first-order autocorrelation varying from almost null to very high (–0.005 to –0.8; Appendix S1 provides further details on simulation procedures). Concretely, the autocorrelation and individual heterogeneity values simulated encompassed a range of situations ranging from individual trajectories that do not vary among individuals and do not fluctuate with time to trajectories that vary among individuals and fluctuate with time (see Fig. S4). We focused on negative first-order autocorrelation because the observed marginal correlation at the individual level is likely to be due to individual heterogeneity, so that trade-offs are often expected to occur once individual heterogeneity has been accounted for (see e.g. Westendorp & Kirkwood 1998). Nevertheless, positive temporal autocorrelation might still occur in addition to heterogeneity, and Appendix S1 provides the details for simulations based on positive autocorrelation.

### *Data sets simulated and estimation of model performance*

All variation in population structure and population parameters resulted in a total of 3888 combinations of population data sets simulated, and each of the 3888 combinations was simulated 100 times. Each time, we ran each of the four models (eqns 1, 2, 3a and 4) to estimate first-order autocorrelation and individual heterogeneity (see Table 1 for a summary of parameters estimated according to each model). We used the R (R Development Core Team 2010) function ‘glmmPQL’ from the package ‘MASS’ (Venables & Ripley 2002), which fits a generalized linear mixed model with multivariate normal random effects using Penalized Quasi-Likelihood, to run models 1, 3 and 4. We used ‘glmmPQL’ because it is one of the rare functions that allow estimating autocorrelated errors in addition to random effects (Zhang *et al.* 2011). For model 2 (which did not include a random effect), we used the R function ‘glm’ that fits a GLM. For each parameter estimated, we computed the mean and the variance over the 100 simulations. We compared the mean estimated value with the simulated value to determine whether bias occurred and used the variance to evaluate the precision of the estimates. We evaluated whether bias and variance varied across the different population structures and population parameters simulated, as well as across the different models tested. Although models 1–4 could also be fitted under a Bayesian approach (e.g. with the R package ‘MCMCglmm’, Hadfield 2010; or

**Table 1.** Summary of each model including first-order autocorrelation and individual heterogeneity parameters for which we estimated accuracy and precision based on simulations

Model description	First-order autocorrelation	Individual heterogeneity
Model 1 <i>State dependence</i> modelling approach including both parameters (eqn 1)	$\beta_T$	$\sigma_{\text{ID}}$
Model 2 <i>State dependence</i> modelling approach excluding individual heterogeneity (eqn 2)	$\beta_T$	–
Model 3 <i>Time series</i> modelling approach including both parameters (eqn 3a)	$\rho$	$\sigma_{\text{ID}}$
Model 4 <i>Time series</i> modelling approach excluding autocorrelation (eqn 4)	–	$\sigma_{\text{ID}}$

‘INLA’, Rue, Martino & Chopin 2009), we chose to assess the performance of non-Bayesian models because it is currently the most common approach used in life-history studies, and because this approach was shown to provide reliable estimates of autocorrelation and individual heterogeneity (see below, Li *et al.* 2011). Furthermore, mixed models fitted with Bayesian methods can be hard to interpret because even models with one-way random effects can end up with bimodal posterior (Liu & Hodges 2003).

### Complexity of simulating binary data

In regression models for binary data, the dependence is often defined with the correlation on the observed values (i.e. 0–1) or in terms of odds ratio (Lipsitz, Laird & Harrington 1991), whereas the random effect has a normal distribution on the logit scale, making the simulation of correlated binary data complex (see Appendix S1 for more details). In addition, because correlation within binary data is constrained by the mean probability (Prentice 1988; Chaganty & Harry 2006), not all combinations of autocorrelation and individual heterogeneity were possible (Fig. S6). For example, strong negative autocorrelation cannot occur for very high or very low probability (Fig. S6). Furthermore, for a given probability, as the standard deviation representing individual heterogeneity increases, the possibility for a strong negative autocorrelation to occur decreases (Fig. S6). Thus, numerous combinations of the 3888 potential ones were impossible to simulate because of this natural constraint. For some other combinations, only a fraction of the 100 simulations were successfully estimated, and so we only used results from the simulations when more than 75% of the simulations of a given combination were successful.

### Comparing binary with normally distributed data

To evaluate how parameter estimates from Bernoulli processes compare with those from Gaussian processes, we simulated a continuous variable with normally distributed residuals (e.g. mass of offspring produced by a female each year). For this example, we set  $\beta_0 = 0$  to simulate a population that is centred on 0. We selected only some combinations of data structure to perform this analysis, that is, short-lived species ( $n_{ID} = 1000$ ,  $n_{AGE} = 5$ ), relatively long-lived species, ( $n_{ID} = 200$ ,  $n_{AGE} = 15$ ) and very long-lived species ( $n_{ID} = 50$ ,  $n_{AGE} = 40$ ). Results were similar for all combinations, so we only report results for a population structure of  $n_{ID} = 200$ ,  $n_{AGE} = 15$ . We simulated the same variation of  $\sigma_{ID}$  and  $\rho$  (or  $\beta_T$ ) as for the binary simulations. We used the R function ‘arima.sim’ to simulate first-order autocorrelation for normally distributed error based on an autoregressive integrated moving average model, that is, ARIMA(p,d,q), where p, d and q are positive integers referring to the order of the autoregressive, integrated and moving average parts of the model, respectively. We used a model of the form ARIMA(1,0,0) to include only a first-order autocorrelation. We simulated the same negative first-order autocorrelation and standard deviation values as for the binary simulations. We ran the four models (eqns 1, 2, 3a and 4 for variables with normally distributed residuals, that is, without the logit transformation) on these simulated data sets to estimate first-order autocorrelation and individual heterogeneity. Because residuals followed a normal distribution, we used the R function ‘lme’ (package ‘nlme’; Pinheiro *et al.* 2010) to fit linear mixed-effects models (models 1, 3, and 4), and the function ‘lm’ to fit LMs (model 2). We fitted the models by maximizing the restricted log-likelihood (i.e. REML method). As for the previous

simulations, we compiled the mean and variance over 100 simulations. Because simulations of continuous data are not constrained like binary data, all simulations were successful. We then compared bias and precision of parameters estimated from linear (mixed) models with those of the binary simulations representing the same population structure.

### Evaluating the influence of incomplete data

As population data consisting of incomplete individual time series are common when working with free-ranging populations, we evaluated the influence of missing values on parameter estimates. We thus randomly removed 10, 25 and 50% of the data from each individual time series and then estimated parameters for models 1–4 for each of the 100 simulations. Again, we compiled the mean and variance over the 100 simulations and only kept values where more than 75% of the simulations were successful. We performed this analysis for both the Gaussian and the binary processes, using only the three combinations of data structure selected previously. As the influence of missing data was similar for both processes and was independent of the mean population probability, we only report results for the binary process with a probability of 0.5.

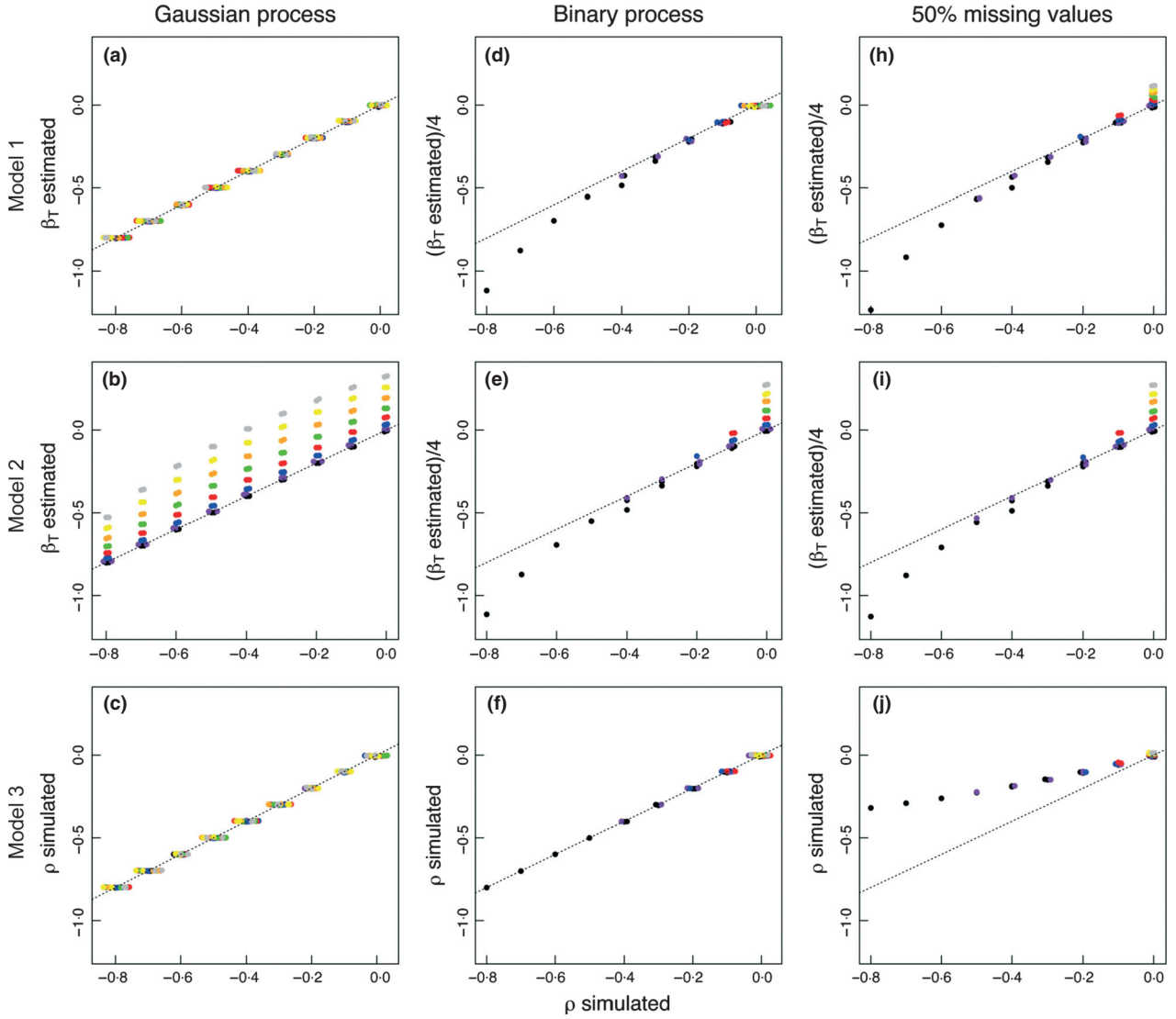
## Results

For all combinations that were successfully simulated ( $n = 686$ ), mean and precision (95th percentile) of the autocorrelation and the individual heterogeneity estimated based on 100 simulations and according to the different models tested are listed in Appendix S2. Below, we provide a summary of the bias patterns we observed for the mean. With respect to precision patterns, the range of precision (i.e. width of the 95th percentile of the estimated values) obtained for the autocorrelation and the individual heterogeneity was overall much smaller for the ‘time series’ models than for the ‘state dependence’ ones (see Appendix S1 and Table S1 for detailed results on precision).

### PARAMETERS SIMULATED VS. ESTIMATED

#### Autocorrelation

Three models included autocorrelation (see Table 1). Estimates of  $\rho$  from model 3 provided the most accurate values for the first-order autocorrelation simulated (Fig. 1f). Estimates of  $\beta_T * [f'(p_v)]^{-1}$  from model 1 also provided fairly accurate values of the autocorrelation simulated, but overestimation began to appear with strong autocorrelation values and increased as stronger autocorrelation was simulated (Fig. 1d). Because small autocorrelation values represent relatively strong autocorrelation for probabilities approaching 0 or 1 (see Fig. S6), the overestimation bias observed with stronger autocorrelation occurred in smaller  $\rho$  simulated for probabilities of 0.7 and 0.9 than for a probability of 0.5 (see Figs 2a–c). Thus, dividing binary parameters by  $f'(p_v)$  provided accurate autocorrelation estimates in comparison with the autocorrelation simulated (and with values estimated from LMs, see Fig. 1a), but only for simulated autocorrelation values that are weak relative to a

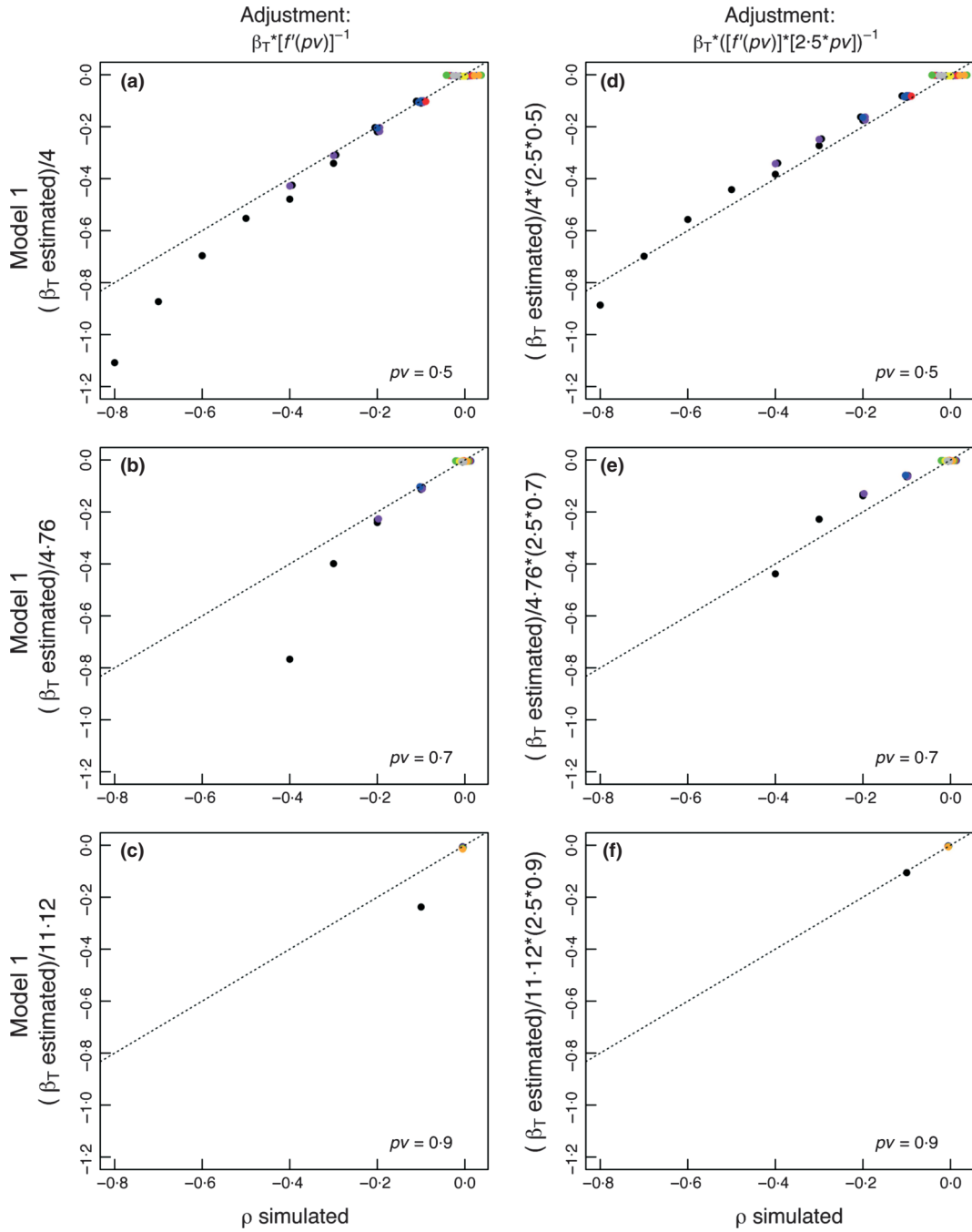


**Fig. 1.** Autocorrelation estimated in relation with the autocorrelation and the standard deviation simulated, for models 1–3 (Table 1), for Gaussian (left panels) and Bernoulli (middle panels) processes. Right panels present the influence of 50% missing values for Bernoulli process (results were similar for Gaussian process; see Fig. S9 for comparison of the results for 10, 25 and 50% missing values). Results are presented for a mean probability of 0.5 for the binary models (and so  $\beta_T * [f'(pv)]^{-1} = \beta_T * 4^{-1}$ ), and for a centered mean value of 0 for the linear models, with  $n_{ID} = 200$  and  $n_{AGE} = 15$ . Colors represent different standard deviation simulated, from low to high (from dark to light colors; black = 0, purple = 0.2, blue = 0.4, red = 0.6, green = 0.8, orange = 1.0, yellow = 1.2, and grey = 1.4). Dashed lines represent points where estimated and simulated values are equal. The same axis ranges were used for all panels to aid comparisons. Autocorrelation values were jittered on the x-axis to better visualize similar estimates.

certain probability. An evaluation *a posteriori* of the relationship between  $\beta_T * [f'(pv)]^{-1}$  and simulated  $\rho$  values suggests that using the formula  $\beta_T * ([f'(pv)] * [2.5 * pv])^{-1}$  would allow correcting for bias associated with strong autocorrelation when estimating autocorrelation for binary processes based on  $\beta_T$  estimates of state dependence models (Figs 2d–f). For model 2, which does not account for individual heterogeneity compared with model 1, estimates of  $\beta_T * [f'(pv)]^{-1}$  provided the worst fit (Fig. 1e). Estimated values showed the same bias as for model 1 with increases in the autocorrelation simulated, but also underestimated the autocorrelation simulated as the standard deviation simulated increased (Fig. 1e).

### Individual heterogeneity

Three models included individual heterogeneity (see Table 1). Model 3 provided again the most accurate estimates, whereas estimates from models 1 and 4 were biased (Fig. 3). For model 3, estimated values were overall not biased, but a very slight underestimation started to appear with very high standard deviation simulated (Fig. 3e). The latter effect became somewhat more important with a small number of individuals and short time series (Appendix S2). When neglecting the autocorrelation using model 4, standard deviation values were underestimated as the strength of the



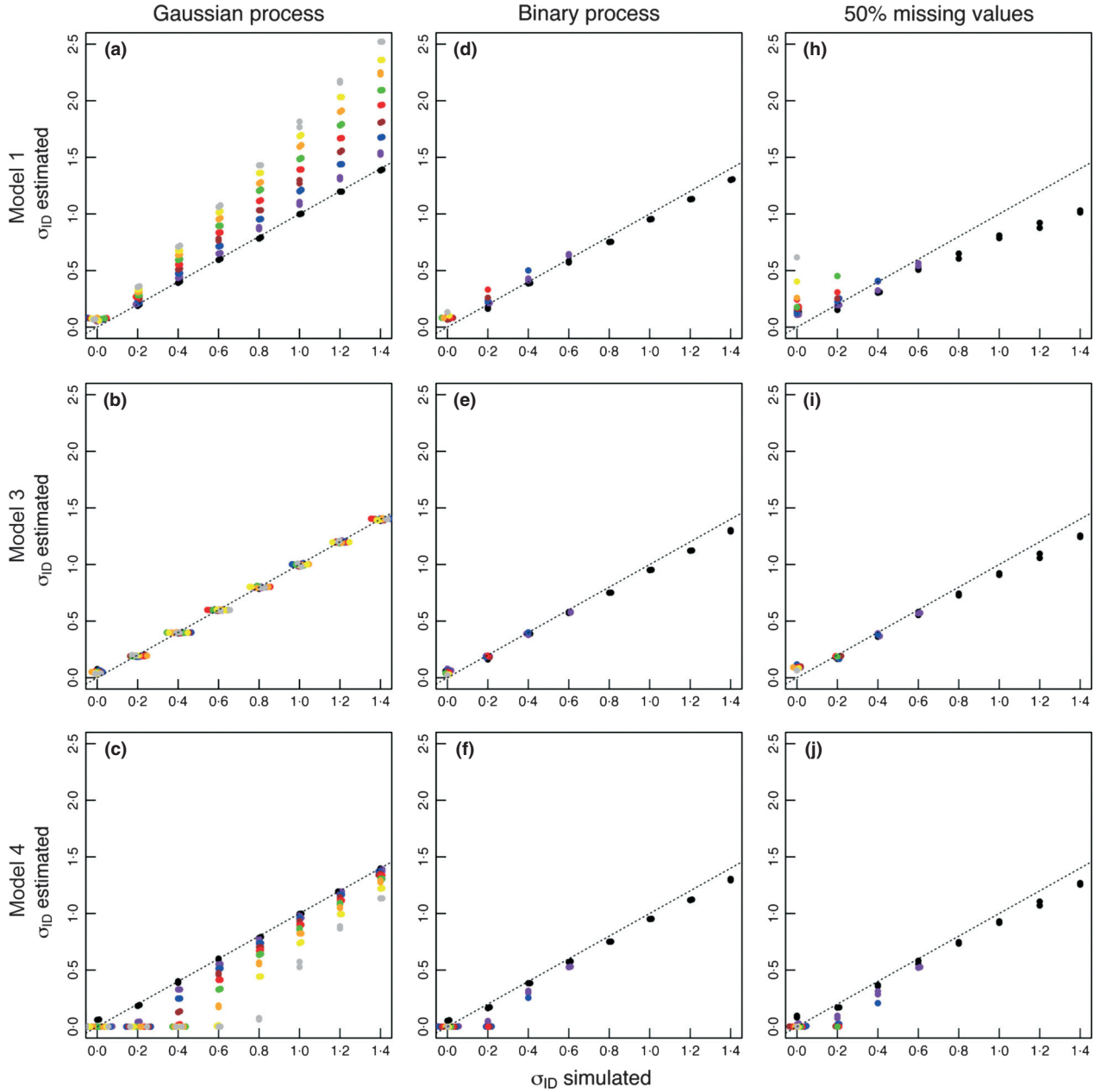
**Fig. 2.** Autocorrelation estimated in relation with the autocorrelation simulated for model 1 following a Bernoulli process, for different probability values ( $pv$ ) and adjustments (left panels: *a priori* adjustment =  $\beta_T * [f'(pv)]^{-1}$ , right panels: *a posteriori* adjustment =  $\beta_T * ([f'(pv)] * [2.5 * pv])^{-1}$ ). As the probability value increases, fewer values of autocorrelation and standard deviation appear because of the limits imposed for simulating binary data (Fig. S6). Results are presented for  $n_{ID} = 1000$  and  $n_{AGE} = 15$ . Colors represent different standard deviation simulated, from low to high (from dark to light colors; black = 0, purple = 0.2, blue = 0.4, red = 0.6, green = 0.8, orange = 1.0, yellow = 1.2, and grey = 1.4). Dashed lines represent points where estimated and simulated values are equal. The same axis ranges were used for all panels to aid comparisons. Autocorrelation values were jittered on the  $x$ -axis to better visualize similar estimates.

autocorrelation simulated increased (lighter colours, Fig. 3f). When incorporating the autocorrelation in the form of  $\beta_T$  using model 1, standard deviation values were overestimated as the autocorrelation simulated increased (lighter colours, Fig. 3d).

#### INFLUENCE OF VARIATION IN POPULATION STRUCTURE

Appendix S1 provides detailed results on the influence of variation in population structure. In general, changes in the





**Fig. 3.** Standard deviation estimated in relation with the autocorrelation and the standard deviation simulated, for models 1, 3 and 4 (Table 1), for Gaussian (left panels) and Bernoulli (middle panels) processes. Right panels present the influence of 50% missing values for Bernoulli process (results were similar for Gaussian process; see Fig. S10 for comparison of the results for 10, 25 and 50% missing values). Results are presented for a mean probability of 0.5 for the binary models, and for a centered mean value of 0 for the linear models, with  $n_{ID} = 200$  and  $n_{AGE} = 15$ . Colors represent different first-order autocorrelation simulated, from weak to strong (from dark to light colors; black =  $-0.005$ , purple =  $-0.1$ , blue =  $-0.2$ , brown =  $-0.3$ , red =  $-0.4$ , green =  $-0.5$ , orange =  $-0.6$ , yellow =  $-0.7$ , and grey =  $-0.8$ ). Dashed lines represent points where estimated and simulated values are equal. The same axis ranges were used for all panels to aid comparisons. Standard deviation values were jittered on the  $x$ -axis to better visualize similar estimates.

number of individuals ( $n_{ID}$ ) and in the length of the time series ( $n_{AGE}$ ) did not bias the parameters estimated, except for short time series,  $n_{AGE} = 5$ . Biases with short time series were almost negligible for the autocorrelation, but were much more pronounced for individual heterogeneity (Appendix S1). The worst bias was a strong overestimation of the standard deviation estimated for a probability of 0.9 for all combinations with  $n_{AGE} = 5$  (Fig. S8). These biases

were slightly smaller with larger  $n_{ID}$  (Appendix S2). Not surprisingly, smaller samples provided less precise estimates than large samples, but, interestingly,  $n_{AGE}$  had a greater influence than  $n_{ID}$  (Appendix S1, Table S1). Modelling short time series also resulted in singularity issues (model 1) and/or failed to converge (models 3 and 4) more often than longer time series ( $n_{AGE} = 5$ : 118,  $n_{AGE} = 15$ : 10, and  $n_{AGE} = 40$ : 3).

## BINARY VS. NORMALLY DISTRIBUTED DATA

Apart from the fact that data describing continuous variables with normally distributed error were not limited by the range of simulated autocorrelation and standard deviation values, results from LMs were overall similar with those obtained from binary ones (Figs 1 and 3). The only exception was that the relationship between estimates of  $\beta_T$  from model 1 and the autocorrelation simulated was linear for continuous variables (Figs 1a,d). The latter was expected because, asymptotically, theory from classical LMs applies to autoregressive models (Mann & Wald 1943; Chatfield 2004). Indeed, one of the methods used to fit an autoregressive model is the ordinary least squares, which corresponds to the LM, and hence, the standardized regression coefficient and the correlation coefficient tally (see Chatfield 2004 p.60).

## INFLUENCE OF INCOMPLETE LIFE-HISTORY DATA

Removing data generally resulted in the underestimation of autocorrelation values (Fig. 1, Fig. S9), but this effect was mainly pronounced for model 3 (Fig. 1j). As the number of missing values increased, the greater was the bias in the autocorrelation estimated (Fig. S9). For model 3, an increase in missing values led to an underestimation of the autocorrelation, which was particularly important for strong values of autocorrelation simulated (Fig. S9 bottom panels). This effect was independent of the standard deviation values simulated (Fig. S9 bottom panels). In model 1, an increase in missing values led to an underestimation of the autocorrelation (i.e. estimated values of  $\beta_T * [\rho(pv)]^{-1}$  decreased) with greater standard deviation values simulated (lighter colours, Fig. S9 top panels). This effect did not occur in model 2 (Fig. S9 middle panels), probably because this effect was already present in the simulations without missing values (Fig. 1e). This last result indicates that missing values had a similar effect on the autocorrelation estimated as neglecting individual heterogeneity in the first place.

The standard deviation estimated was also affected by the presence of missing values (Fig. 3; Fig. S10). With a greater percentage of missing values, model 1 underestimated the standard deviation when standard deviation values simulated were large (Fig. S10 top panels). Still in model 1, the standard deviation was markedly overestimated with increase in the autocorrelation values simulated (lighter colours, Fig. S10 top panels). These effects did not occur in model 3 and 4 (Figs 3i,j; Fig. S10 middle and bottom panels).

## POSITIVE TEMPORAL AUTOCORRELATION

We found that temporal autocorrelation and individual heterogeneity are more difficult to estimate reliably for positive than negative temporal autocorrelation. The two processes seemed to get mixed up, competing for the same information, and autocorrelation generally took over, which resulted in a correct estimation of temporal autocorrelation but an underestimation of individual heterogeneity (see

Appendix S1 and Fig. S5 for detailed results on positive temporal autocorrelation).

## Discussion

Although temporal dependence and individual heterogeneity are fundamentally important processes in evolutionary and ecological studies, a clear and detailed evaluation of how best these processes can be measured and estimated is currently lacking in the literature. Here, we used simulated data sets to assess and compare the reliability of different models in estimating temporal autocorrelation and individual heterogeneity in binary processes, based on realistic situations of individual variation within populations with contrasted life histories and sampling intensity. Our results demonstrate that the complex structure of these patterns is difficult to tackle with simple models. Indeed, although the state dependence model that includes individual heterogeneity (model 1) led to relatively fair autocorrelation and individual heterogeneity estimates, the time series model (3) provided by far the most precise and accurate estimates overall. Furthermore, simpler models ignoring either individual heterogeneity (state dependence model 2) or autocorrelation (time series model 4) performed relatively poorly. By ignoring individual heterogeneity, model 2 underestimated the autocorrelation even when the occurrence of individual heterogeneity was very small. Similarly, model 4 underestimated individual heterogeneity even when only a small autocorrelation level was introduced in the data.

These simulation results illustrate the fundamental association between temporal autocorrelation and individual heterogeneity. Because a complete absence of individual heterogeneity within data containing repeated measures should rarely be observed, we should expect a frequent co-occurrence of autocorrelation and individual heterogeneity within longitudinal data sets used to assess variation in life-history traits. These kinds of expectations coming from our biological understanding of the underlying processes are essential for modelling dependence. It is also important to note, however, that temporal autocorrelation can appear as a result of an inappropriate modelling of the fixed effects (e.g. modelling the fixed effect of age using a linear relationship when a curvilinear effect of age occurs, or neglecting to include an environmental covariate that is correlated with time; Hodges & Reich 2010). Indeed, statistical researches on spatial dependency have demonstrated that the form of the dependence modelled (i.e. residual spatial dependence vs. independence) markedly affects parameter estimates and that statistical analyses alone (e.g. model selection) cannot select the most appropriate form of dependence for modelling the data (Wakefield 2007). Instead, the form of dependence to be modelled should be determined by the biological understanding of the underlying processes, emphasizing that a good biological knowledge is required to model properly both stochastic and fixed effects. Therefore, one must interpret carefully autocorrelation estimates in situations where a non-negligible autocorrelation (i.e.  $> |0.1|$ ) is found but not expected. Furthermore, simple models ignoring

either individual heterogeneity or autocorrelation should be avoided as soon as both processes are suspected to occur in a longitudinal data set (see summary of recommendations in Table 2). Importantly, studies now commonly include individual heterogeneity in models, but the temporal dependence is seldom taken into account. Our results, however, clearly demonstrate that individual heterogeneity and temporal autocorrelation both need to be included in models even when only one of the two processes is investigated (Table 2). Moreover, many researchers fit mixed models to account for possible confounding effects of individual heterogeneity on the focal relationship between traits rather than to estimate individual heterogeneity *per se*. Although our study does not focus on measuring the impact of ignoring either individual heterogeneity or temporal autocorrelation on fixed parameter estimates, the best rule of thumb ecologists should use would be modelling both processes in all case studies. Such a modelling strategy would allow accounting properly for the confounding effects of individual heterogeneity or temporal autocorrelation and thereby reduce the risk of getting biased estimates.

The time series model (3) provided the best estimates for measuring the autocorrelation compared with the state dependence models (1 and 2). State dependence models do not provide autocorrelation values directly, but through the estimation of  $\beta_T$  coefficients. For continuous variables, our

results confirmed that  $\beta_T$  coefficients from LMs are directly comparable with  $\rho$  values, corresponding to autocorrelation values. When variables are binary, however, we need to adjust  $\beta_T$  to provide an estimate of autocorrelation that is equivalent to  $\rho$  (and to  $\beta_T$  estimated from LMs). We found that the precision and the accuracy of adjusted  $\beta_T$  were comparable with  $\rho$ , but an overestimation occurred when data included large autocorrelation values relative to the maximum autocorrelation that can occur for a given level of individual heterogeneity. An evaluation *a posteriori* of our results suggested that standardizing  $\beta_T$  using the formula  $\beta_T * ([f'(pv)] * [2.5 * pv])^{-1}$  allowed correcting for the bias associated with strong autocorrelation when estimating the autocorrelation based on  $\beta_T$  in binary models. Applying this correction did not result in a perfect match between the autocorrelation simulated and the  $\beta_T$  estimated, as was the case for  $\rho$  estimates, but it allowed capturing the true relationship and provided a reasonable fit overall. Therefore, using standardized  $\beta_T$  estimates offers a comparable measure of autocorrelation as using  $\rho$  estimates.

For measuring individual heterogeneity, both the state dependence model (1) and the time series model (3) provided similar estimates, except that the state dependence model (1) was biased when modelling Gaussian processes with strong autocorrelation values. The state dependence model (1), however, was more robust to missing values than the time series model (3). Indeed, our results demonstrated that the larger the

**Table 2.** Summary of recommendations according to the scientific aims

Aim	Model	Recommendations
Quantify temporal autocorrelation only	Time series model 3	Preferred, but only if the data set includes no or a very low proportion of missing values (< 10%)
	State dependence model 1	Recommended if the data set includes > 10% missing values, but $\beta_T$ requires an <i>a posteriori</i> standardization using $\beta_T * ([f'(pv)] * [2.5 * pv])^{-1}$
Quantify individual heterogeneity only	State dependence model 2	Not recommended
	Time series model 3	Preferred
	State dependence model 1	Recommended, but not if the data set includes a large proportion of missing values (> 25%)
Quantify both temporal autocorrelation and individual heterogeneity	Time series model 4	Not recommended
	Time series model 3	Preferred, but only if the data set includes no or a very low proportion of missing values (< 10%)
	State dependence model 1	Recommended, but not if the data set includes a large proportion of missing values (> 25%). $\beta_T$ requires an <i>a posteriori</i> standardization using $\beta_T * ([f'(pv)] * [2.5 * pv])^{-1}$
Quantify temporal autocorrelation and/or individual heterogeneity for short time series ( $n_{AGE} \leq 5$ )	Time series model 3	Recommended, except for estimating heterogeneity if the mean probability approaches 0 or 1
	State dependence model 1	Not recommended, except for estimating autocorrelation if the number of individuals monitored is very large (> 200)
	State dependence model 2	Not recommended
	Time series model 4	Not recommended
Quantify temporal autocorrelation when expecting a positive temporal autocorrelation	Time series model 3	Preferred
	State dependence model 1	Recommended, except when strong autocorrelation (> 0.7) is expected
	State dependence model 2	Not recommended
Quantify individual heterogeneity when expecting a positive temporal autocorrelation	State dependence model 1	Not recommended
	Times series model 3	Not recommended
	Times series model 4	Not recommended

percentage of missing values in the data sets, the greater was the bias in estimating autocorrelation and individual heterogeneity, although this bias was more pronounced for the autocorrelation than for the individual heterogeneity estimates. Results differed importantly between data sets containing < 10% missing values and those containing 25%. Although 25 is a relatively high percentage, it represents a population data set with only 3–4 years missing out of 15 for most individuals, a figure that is far from being uncommon in long-term longitudinal studies. Finally, data sets based on very short time series often had convergence problems and, more importantly, led to seriously biased estimates in some instances. These biases were reduced when a larger number of individuals were included in the analyses, but they were still present. It is essential to keep in mind that measuring temporal autocorrelation and individual heterogeneity based on short time series and/or on data with a high proportion of missing values can be misleading (Table 2). This is noteworthy because most studies usually have data consisting of very short time series, often combined with a small number of individuals not monitored at each census. For instance, about 65% of populations from which survival and reproductive costs of reproduction have been investigated in mammals (as reviewed in Hamel *et al.* 2010) were based on time series shorter or equal to 5 years. Of these, about 35% were also based on a low sample size of about 50 or less individuals. This suggests that several mammalian populations for which life-history data have been collected in the past would not be suitable for reliably measuring temporal autocorrelation and individual heterogeneity processes, a situation that is likely to be similar in other taxa. Our study, therefore, generally advises against estimating the strength of the autocorrelation and the individual heterogeneity present in data sets consisting of very short time series (Clutton-Brock & Sheldon 2010) because these estimates could be misleading (Table 2). Short time series can be the result of either short monitoring periods or short species life span. Researchers can prevent the first problem to occur by monitoring individuals for as long as possible to estimate reliably both autocorrelation and individual heterogeneity, but the second problem cannot be avoided. Nevertheless, autocorrelation and individual heterogeneity can be estimated reliably for very short time series in some specific instances, such as when more than 200 individuals are sampled without missing values and the mean population probability is not close to 0 or 1 (Table 2). Furthermore, although variation in longevity among individuals did not influence results (see Appendix S1), unbalanced data sets will necessarily exclude individuals that die after one time step, which could affect estimates of autocorrelation and individual heterogeneity.

In this study, we evaluated accuracy and precision of temporal autocorrelation and individual heterogeneity estimates provided by different binary models based on simulated data to determine the reliability of different model types. Overall, our work suggests that using a time series approach (model 3) will provide better autocorrelation and individual heterogeneity estimates, except in the presence of missing data (> 10%), where a state dependence approach (model 1) will give better estimates (Table 2, see also Nakagawa

& Freckleton 2008 for suggestions for accounting for missing values in analyses). As measures of autocorrelation and individual heterogeneity are now considered to reflect biological processes rather than simply being a nuisance requiring to be accounted for (van de Pol & Verhulst 2006; van de Pol & Wright 2009; Tuljapurkar, Steiner & Orzack 2009), the insights we provide on the reliability of different models in estimating the intensity of these two processes should be useful in a large range of situations across different research areas. Moreover, our work focused primarily on binary variables, because they are common in life-history studies but are prone to serious modelling difficulties. To generalize our approach, future researches should aim at evaluating the reliability of autocorrelation and individual heterogeneity estimates for other distributions. In addition, our methods are not suitable for time-to-event data such as survival, which are analysed using other types of models (e.g. Cox proportional hazards models). While a rather large literature on frailty models has been published following the seminal paper by Vaupel, Manton & Stallard (1979), no paper to our knowledge has yet addressed some form of temporally autocorrelated hazard functions (only some spatially autocorrelated mixed hazards models have been produced and applied to human survival models; e.g. Hennerfeind, Brezger & Fahrmeir 2006). Such models require to be developed to evaluate whether temporal autocorrelation and individual heterogeneity can be quantified for time-to-event data. Finally, more research is also required for estimating individual heterogeneity when positive temporal autocorrelation is expected because none of the models evaluated in this study provided suitable estimates (Fig. S5, Table 2). Nevertheless, these models are useful if one is only interested in estimating positive temporal autocorrelation (Table 2).

## Acknowledgements

Financial support was provided from Le Fonds québécois de la recherche sur la nature et les technologies (postdoctoral fellowship to SH), from the Norwegian Research Council (Aurora programmet) and from the University of Tromsø. We thank J.M. Ver Hoef for helpful discussions, and D. Nussey and M. van de Pol for very constructive comments.

## References

- Berridge, D.M. & Crouchley, R. (2011) *Multivariate Generalized Linear Mixed Models Using R*. CRC Press, Boca Raton.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**, 127–135.
- Boyer, N., Réale, D., Marmet, J., Pisanu, B. & Chapuis, J.-L. (2010) Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. *Journal of Animal Ecology*, **79**, 538–547.
- Chaganty, N.R. & Harry, J. (2006) Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, **93**, 197–206.
- Chatfield, C. (2004) *The Analysis of Times Series: An Introduction*. CRC Press, New York.
- Clutton-Brock, T. & Sheldon, B.C. (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562–573.
- Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.

- Gaillard, J.-M. & Yoccoz, N.G. (2003) Temporal variation in survival of mammals: a case of environmental canalization? *Ecology*, **84**, 3294–3306.
- Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.
- Gimenez, O. & Choquet, R. (2010) Individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. *Ecology*, **91**, 951–957.
- Hadfield, J. (2010) MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.
- Hamel, S., Gaillard, J.-M., Yoccoz, N.G., Loison, A., Bonenfant, C. & Descamps, S. (2010) Fitness costs of reproduction depend on life speed: empirical evidence from mammalian populations. *Ecology Letters*, **13**, 915–935.
- Heckman, J.J. (2001) Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, **109**, 673–748.
- Hennerfeind, A., Brezger, A. & Fahrmeir, L. (2006) Geoadditive survival models. *Journal of the American Statistical Association*, **101**, 1065–1075.
- Hodges, J.S. & Reich, B.J. (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *American Statistician*, **64**, 325–334.
- Kéry, M. & Gregg, K.B. (2003) Effects of life-state on detectability in a demographic study of the terrestrial orchid *Cleistes bifaria*. *Journal of Ecology*, **91**, 265–273.
- Li, B.Y., Lingsma, H.F., Steyerberg, E.W. & Lesaffre, E. (2011) Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Medical Research Methodology*, **11**, 77.
- Lipsitz, S.R., Laird, N.M. & Harrington, D.P. (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, **78**, 153–160.
- Liu, J.N. & Hodges, J.S. (2003) Posterior bimodality in the balanced one-way random-effects model. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **65**, 247–255.
- Mann, H.B. & Wald, A. (1943) On the statistical treatment of linear stochastic difference equations. *Econometrica*, **11**, 173–220.
- Martin, J.G.A., Nussey, D.N., Wilson, A. & Réale, D. (2011) Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression coefficient models. *Methods in Ecology and Evolution*, **2**, 362–374.
- Masaoud, E. & Stryhn, H. (2010) A simulation study to assess statistical methods for binary repeated measures data. *Preventive Veterinary Medicine*, **93**, 81–97.
- Nakagawa, S. & Freckleton, R.P. (2008) Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, **23**, 592–596.
- van Noordwijk, A.J. & de Jong, G. (1986) Acquisition and allocation of resources: their influence on variation in life history tactics. *American Naturalist*, **128**, 137–142.
- Nussey, D.N., Coulson, T., Delorme, D., Clutton-Brock, T.H., Pemberton, J.M., Festa-Bianchet, M. & Gaillard, J.M. (2011) Patterns of body mass senescence and selective disappearance differ among three species of free-living ungulates. *Ecology*, **92**, 1936–1947.
- Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. & the R Development Core Team. (2010) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-97.
- van de Pol, M. & Verhulst, S. (2006) Age-dependent traits: a new statistical model to separate within- and between-individual effects. *American Naturalist*, **167**, 766–773.
- van de Pol, M. & Wright, J. (2009) A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour*, **77**, 753–758.
- Pomeroy, P.P., Fedak, M.A., Rothery, P. & Anderson, S. (1999) Consequences of maternal size for reproductive expenditure and pupping success of grey seals at North Rona, Scotland. *Journal of Animal Ecology*, **68**, 235–253.
- Prentice, R.L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.
- R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Réale, D., Gallant, B.Y., Leblanc, M. & Festa-Bianchet, M. (2000) Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour*, **60**, 589–597.
- Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **71**, 319–392.
- Schielzeth, H. (2010) Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, **1**, 103–113.
- Shefferson, R.P., McCormick, M.K., Whigham, D.F. & O'Neill, J.P. (2011) Life history strategy in herbaceous perennials: inferring demographic patterns from the aboveground dynamics of a primarily subterranean, myco-heterotrophic orchid. *Oikos*, **120**, 1291–1300.
- Tuljapurkar, S., Steiner, U.K. & Orzack, S.H. (2009) Dynamic heterogeneity in life histories. *Ecology Letters*, **12**, 93–106.
- Vaupel, J.W., Manton, K.G. & Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439–454.
- Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics With S*. Springer, New York.
- Wakefield, J. (2007) Disease mapping and spatial regression with count data. *Biostatistics*, **8**, 158–183.
- Westendorp, R.G.J. & Kirkwood, T.B.L. (1998) Human longevity at the cost of reproductive success. *Nature*, **396**, 743–746.
- Wooldridge, J.M. (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, **20**, 39–54.
- Zhang, H., Lu, N., Feng, C., Thurston, S.W., Xia, Y., Zhu, L. & Tu, X.M. (2011) On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine*, **36**, 2562–2572.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Supplementary methods and results.

**Appendix S2.** Table listing the mean and precision of the autocorrelation and the individual heterogeneity estimated according to the different models tested and to the different population combinations.