



Errors-in-variables and validation problems in reaction norm predictions for wild populations

R. Ergon¹

¹*Department of Electrical engineering, Information Technology and Cybernetics, University of South-Eastern Norway, Porsgrunn, Norway. E-mail: rolf.ergon@usn.no*

Abstract

Studies of phenotypic responses in wild populations are often based on reaction norm models where the environmental drivers in many cases are related to climate change. Such input signals will never be exactly known, and there will always be measurement errors also in the recorded responses. In parameter estimation these errors give rise to errors-in-variables problems, especially in the form of overfitting caused by errors in the input measurements. A second important feature of such phenotypic response problems is that the environmental inputs must be given appropriate but largely unknown reference values. A third problem is that it is difficult to find good validation methods for predicted responses. Essential aspects of these problems are here studied by use of a reaction norm model in its simplest univariate form, characterized by a mean intercept value and a mean plasticity slope value, and the overall conclusion is that validated disentanglement of plasticity and genetic adaptation based on realistically short data for wild populations is a difficult task. In a proposed validation method, the available input-output data is split into one part for modeling and one part for validation, and the feasibility of this approach is studied in simulations with use of a prediction error method, which is essentially a maximum likelihood method. It is also argued that validation of a chosen or estimated reference environment in practice is impossible when the data comes from the (unintended) anthropogenic global warming experiment, where no independent experimental data exists. When the evolution is slow because of small genetic variances, overlapping generations and long lifetimes, or because of near optimal adaptive plasticity, the best quantitative genetics option may be to assume a constant plasticity slope value, equal to the initial value. It turns out to be easy to estimate this value, but that should be done without setting other unknown parameter values to zero. This option is appealing also because it removes the dependence of a guessed or estimated reference environment.

Keywords: Environmental drivers; errors-in-variables problems; prediction error method; reaction norm predictions; reaction norm validation; reference environment.

1. Introduction

Errors-in-variables problems are ubiquitous in ecological and evolutionary modeling where processes driven by environmental variables are modeled for some specific purpose. For example, Kangas (1998) used tree diameter at breast height, tree height and the height to the base of live crown as independent variables in a for-

est growth model and discussed the effects of measurement errors. J. et al. (2015) discussed how measurement errors in spatial climate variables affect species distribution models. A number of articles have studied effects of climate change and discussed whether phenotypic changes are genetically based or the result of phenotypic plasticity. A summary of such studies was given in Merilä and Hendry (2014), with reference to

11 accompanying studies covering a variety of taxa. A main conclusion was that plastic responses on climate change are common, but that there were few examples of adaptive genetic responses. An early example where adaptive responses were found was a study of a northern squirrel population (Réale et al., 2003), but that was for the special case with a constant plasticity slope. In another example Tarka et al. (2015) studied great reed warblers, while Bonamour et al. (2019) used plasticity of blue tit breeding phenology as an example. A common feature in such papers is the use of various ecological drivers of evolutionary processes, and such drivers are obviously not exactly known. In a recent example, Valdés et al. (2023) used 22 years of field observations of the perennial forest herb *Lathyrus vernus* in order to assess phenotypic selection of flowering time as function of spring temperature, and as spring temperature they used the mean April temperature as an approximation of the true but unknown driver.

In all these examples, and numerous others, there are errors-in-variables problems as discussed in Söderström (2019) and depicted in Figure 1. The ecological system has scalar or vector valued and time-varying input and output signals $u_{0,t}$ and $z_{0,t}$, respectively, while the measured signals are $u_t = u_{0,t} + \tilde{u}_t$ and $z_t = z_{0,t} + \tilde{z}_t$, respectively, where \tilde{u}_t and \tilde{z}_t are the measurement errors. Without further assumptions there is a fundamental lack of identifiability in the errors-in-variables problem.

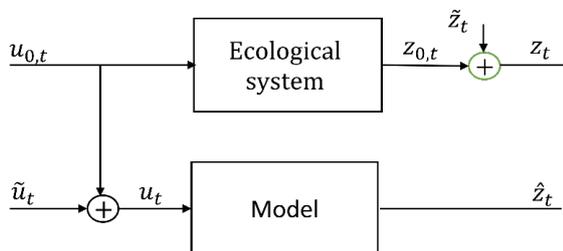


Figure 1: The basic setup for an errors-in-variables model identification problem.

In the ecological and evolutionary setting the true output $z_{0,t}$ is typically multivariate, with vectors $\mathbf{y}_{0,t}$ and $\mathbf{w}_{0,t}$ of individual phenotypic and relative fitness values, respectively. The true input $u_{0,t}$ that drives the evolutionary process is often a climatic variable, which varies from year to year with very little autocorrelation. In such cases the influence of the environmental cue $u_{0,t}$ is found from $u'_{0,t} = u_{0,t} - u_{ref}$, from which follows $u'_t = u_t - u_{ref}$, where u_{ref} is a reference environment as discussed in Ergon (2022). An important climatic variable is spring air temperature, which typically is a Gaussian white noise process around a slow trend in mean value, as shown in Figure 2. Although

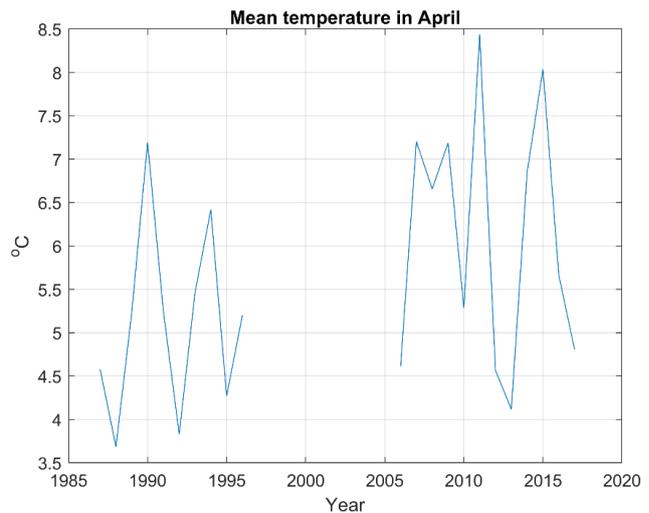


Figure 2: Mean April temperature in Tullgarn south of Stockholm, which is used as environmental driver in Valdés et al. (2023).

the specific choices of such input signals as drivers of evolutionary processes should be based on all available information, they will obviously be more or less error corrupted representations u_t of the true driver $u_{0,t}$. If also the error signal \tilde{u}_t is Gaussian white noise, there is no practical possibility to reconstruct $u_{0,t}$ from known values of u_t . A theoretical possibility could otherwise involve the use of information on non-normal distributions of $u_{0,t}$ and \tilde{u}_t (Söderström, 2019). Without such additional information one must simply accept parameter estimation errors caused by \tilde{u}_t , and as far as possible test the results by validation against data that are not used in the modeling. A primary validation problem is here that data typically come from field observations of the (unintended) anthropogenic global warming experiment, and that no independent experimental data exist. A possible approach for this type of data is to use a part of the recorded time series for modeling and the rest for validation. There is also a second validation problem with such data in that the reference environment is unknown, and this problem cannot be solved by splitting the time series data in two parts (because the two parts will have the same reference environment). The unknown reference environment thus represents a serious problem.

Here, I will focus on the common cases where reaction norms are predicted on the basis of time-varying environmental variables, as discussed above and exemplified in Figure 2. A reaction norm describes the phenotypes that a given set of genes in an individual biological organism can produce across a range of environments. Mean reaction norms in a population of such organisms can evolve by natural selection for example

after a sudden introduction to a new environment or as response on climate change, and this evolution can be modeled by use of several methods. In its simplest univariate form, a mean reaction norm is characterized by a mean intercept value and a mean plasticity slope value, but it may also be necessary to use models with a multiple of reaction norms. In many cases reaction norms are also nonlinear (Arnold et al., 2019).

A simple reaction norm model is the two-trait intercept-slope model as used in Lande (2009),

$$y_{0,i,t} = a_{i,t} + v_{i,t} + b_{i,t}(u_{0,t} - u_{ref}), \quad (1a)$$

where $y_{0,i,t}$ is the individual phenotypic response on an environmental cue $u_{0,t}$ at generation t , and where u_{ref} is the reference environment (which Lande assumed to be zero). In populations with sexual reproduction, an individual will here be a midparent with traits that are mean values of the traits of the two parents. The reference environment is defined as the environment where the phenotypic variance has its minimum, and where thus the expected geometric mean fitness is maximized (Ergon, 2022). In the following I will treat the individual reaction norm parameters $a_{i,t} + v_{i,t}$ and $b_{i,t}$ as quantitative traits in their own right, and where necessary I will thus distinguish between phenotypic traits and reaction norm traits. In Equation (1a), $a_{i,t}$ and $b_{i,t}$ are the additive genetic components of the reaction norm traits, while $v_{i,t}$ is zero mean Gaussian white noise. From Equation (1a) follows the mean reaction norm equation

$$\bar{y}_{0,t} = \bar{a}_t + \bar{b}_t(u_{0,t} - u_{ref}), \quad (1b)$$

where the mean traits \bar{a}_t and \bar{b}_t , and thus also \bar{y}_t , may evolve as results of a changing environment and natural selection. See Lande (2009) for step response simulations of the model in Equations (1a,b), and Figure 1 in Ergon (2023b) for a detailed explanation.

As discussed in Ergon (2023b), it may be difficult to estimate the environmental reference values from data, and I will therefore assume that u_{ref} is known from historical data, for example from the time before the recent increase in global temperature, or before a sudden change in the environment. In simulations, I will assume that u_t is a temperature, and I will then choose a temperature scale such that $u_{ref} = 0$.

A main focus in the presentation will be to show the difficulties involved in disentanglement of the plasticity and microevolutionary contributions to measured changes in mean phenotypic values, given short time series data collected in field studies, where also the reference environment is unknown. For this purpose I will use simulations where assumed true data are generated according to the model in Equations (1a,b), and where parameters in this model are found by use of a

prediction error method (PEM). The PEM approach is well established in the engineering control field (Ljung, 1999, 2010), but although comparisons between measured and predicted responses are discussed in the literature (Morrissey et al., 2010; Merilä and Hendry, 2014), PEM as such appears to be largely unknown in the quantitative genetics community. A main parameter estimation difficulty is that errors in the assumed environmental drivers will result in large prediction errors caused by overfitting, and this difficulty becomes even more pronounced when a part of the already short data is set aside for validation purposes. The situation may be different in controlled laboratory settings, where measurement errors may be small, and where the time series may be longer. It should be noted that overfitting to short data is not a specific PEM problem, errors in the assumed environmental driver will result in overfitting regardless of the choice of parameter estimation method.

Background theory is given in Section 2, together with a more detailed discussion of the input measurement problem and a presentation of the proposed validation methodology under the assumption of a correct reference environment. A special case with constant plasticity slope is described. Different aspects of the identification and validation processes are illustrated and documented by simulations in Section 3. In the simulations I assume the simple model (1a,b), with parameter values as in Lande (2009), and I use two types of environmental changes. The first type is a sudden change in environment, as in Lande (2009), while the other type is a gradual change similar to the global warming since 1970. In both cases I assume non-overlapping generations and realistic temperature variations from generation to generation, but extensions to cases with overlapping generations are also discussed. A final discussion with conclusions is given in Section 4. Additional examples of prediction results are given in Appendices A and B. Typical MATLAB codes are given in Appendix C.

2. Theory and Methods

2.1. The selection gradient model

Under the assumption of random mating in a large population the evolution of correlated phenotypic traits is determined by the multivariate breeder’s equation (Lande, 1979). When the reaction norm parameters in Equation (1a) are treated as quantitative traits in their own right, the multivariate breeder’s equation gives

$$\begin{bmatrix} \Delta \bar{a}_t \\ \Delta \bar{b}_t \end{bmatrix} = \mathbf{GP}^{-1} \begin{bmatrix} cov(w_{0,i,t}, a_{i,t}) \\ cov(w_{0,i,t}, b_{i,t}) \end{bmatrix} = \mathbf{G}\boldsymbol{\beta}_t, \quad (2)$$

where β_t is the selection gradient. Here, $w_{0,i,t} = W_{0,i,t}/\bar{W}_{0,t}$ is the relative individual fitness, where $W_{0,i,t}$ and $\bar{W}_{0,t}$ are the individual and mean fitness, respectively, while \mathbf{G} and \mathbf{P} are given by

$$\mathbf{G} = E \left[\begin{bmatrix} a_{i,t} - \bar{a}_t \\ b_{i,t} - \bar{b}_t \end{bmatrix} \begin{bmatrix} a_{i,t} - \bar{a}_t & b_{i,t} - \bar{b}_t \end{bmatrix} \right] = \begin{bmatrix} G_{aa} & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} G_{aa} + \sigma_v^2 & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix}. \text{ I will assume that the } \mathbf{G} \text{ and } \mathbf{P} \text{ matrices are constant.}$$

The model in Equation (2) cannot be identified by use of available environmental, phenotypic, and fitness data, for the simple reason that the individual latent variables $a_{i,t}$ and $b_{i,t}$ are not available. This problem was solved in Ergon (2022) by use of a linear transformation of the vector $[a_{i,t} \ b_{i,t}]^T$ onto the vector $[a_{i,t} \ b_{i,t} \ y_{i,t}]^T$. This results in the model

$$\begin{bmatrix} \Delta \bar{a}_t \\ \Delta \bar{b}_t \end{bmatrix} = \begin{bmatrix} G_{aa} & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix} \begin{bmatrix} 1 \\ u_{0,t} - u_{ref} \end{bmatrix} \times P_{yy,t}^{-1} cov(w_{0,i,t}, y_{0,i,t}), \quad (3a)$$

where

$$P_{yy,t} = G_{aa} + \sigma_v^2 + 2G_{ab}(u_{0,t} - u_{ref}) + G_{bb}(u_{0,t} - u_{ref})^2. \quad (3b)$$

In Equation (3a), selection with respect to $a_{i,t}$ and $b_{i,t}$, as in Equation (2), is replaced by selection with respect to $y_{0,i,t}$, which is assumed to be known (with a measurement error). The two models (2) and (3a,b) give equal results when the population size $N \rightarrow \infty$, but for practical purposes they are interchangeable also for rather limited population sizes. Since both models in any case are approximations of reality, Equations (3a,b) can be seen as just as correct as Equation (2). Note that Equation (3a) is a dynamical model in the sense that $P_{yy,t}$ varies with time. In order for Equation (3a) to be compatible with the fundamental Price equation (Price et al., 1970; Ergon, 2019) the covariance expression should be computed as

$$cov(w_{0,i,t}, y_{0,i,t}) = \frac{1}{N} \sum_{i=1}^N w_{0,i,t} y_{0,i,t} - \bar{w}_{0,t} \bar{y}_{0,t}. \quad (3c)$$

With given initial mean trait values \bar{a}_1 and \bar{b}_1 , Equation (3a) gives \bar{a}_t and \bar{b}_t , and thus also $\bar{y}_{0,t} = \bar{a}_t + \bar{b}_t(u_{0,t} - u_{ref})$ according to Equation (1b), for use as assumed ‘true’ mean phenotypic data in simulations.

2.2. The prediction error model

Under the assumption of a true model according to Equations (3a,b), and with measurement errors, the

prediction error method (PEM) uses the model

$$\begin{bmatrix} \Delta \hat{a}_t \\ \Delta \hat{b}_t \end{bmatrix} = \begin{bmatrix} G_{aa} & \hat{G}_{ab} \\ \hat{G}_{ab} & \hat{G}_{bb} \end{bmatrix} \begin{bmatrix} 1 \\ u_{0,t} + \tilde{u}_t - \hat{u}_{ref} \end{bmatrix} \times \hat{P}_{yy,t}^{-1} cov(w_{0,i,t} + \tilde{w}_{i,t}, y_{0,i,t} + \tilde{y}_{i,t}), \quad (4a)$$

$$\hat{y}_t = \hat{a}_t + \hat{b}_t(u_{0,t} + \tilde{u}_t - \hat{u}_{ref}), \quad (4b)$$

where

$$\hat{P}_{yy,t} = G_{aa} + \hat{\sigma}_v^2 + 2\hat{G}_{ab}(u_{0,t} + \tilde{u}_t - \hat{u}_{ref}) + \hat{G}_{bb}(u_{0,t} + \tilde{u}_t - \hat{u}_{ref})^2. \quad (4c)$$

In addition, PEM uses the unknown initial mean reaction norm slope value \hat{b}_1 , while the initial intercept value is given by $\hat{a}_1 = \hat{y}_1 - \hat{b}_1(u_{0,1} + \tilde{u}_1 - \hat{u}_{ref})$. Here, \hat{y}_1 can be set to any value, for example $\hat{y}_1 = 0$. Note that with $cov(w_{0,i,t} + \tilde{w}_{i,t}, y_{0,i,t} + \tilde{y}_{i,t})$ given from the data, G_{aa} can be set to any value, such that \hat{G}_{bb} and $\hat{\sigma}_v^2$ are found in relation to this value, i.e., G_{aa} cannot be estimated.

Equation (3) is derived under the assumption of non-overlapping generations. With overlapping generations only a fraction $c_t < 1$ of the population is affected by the changes in fitness, and the expression $cov(w_{0,i,t} + \tilde{w}_{i,t}, y_{0,i,t} + \tilde{y}_{i,t})$ in Equation (4a) should then be replaced by $c_t cov(w_{0,i,t} + \tilde{w}_{i,t}, y_{0,i,t} + \tilde{y}_{i,t})$ (Ergon, 2023a). A constant value of c_t can also be estimated from input-output data.

In PEM, all the unknown parameters \hat{G}_{ab} , \hat{G}_{bb} , $\hat{\sigma}_v^2$, \hat{b}_1 and \hat{u}_{ref} (and possibly \hat{c}_t) are in principle tuned until the criterion function $f = \frac{1}{T} \sum_{t=1}^T (\bar{y}_t - \hat{y}_t)^2$ reaches a minimum, where T is the number of samples. In the simulations in Section 3 the MATLAB function *fmincon* is used for this purpose. This function finds a constrained minimum of a scalar function of several variables. In order to focus on the errors-in-variables and validation problems, the simulations assume that $\hat{G}_{ab} = G_{ab} = 0$ and $u_{ref} = 0$. As will be found in the simulations, the final residuals $\bar{y}_t - \hat{y}_t$ will approximately be white noise, which makes PEM into a maximum likelihood method (Ljung, 2010).

2.3. PEM error analyses

Since model tuning results in $\hat{y}_t \approx \bar{y}_t$, all measurement errors must be compensated by errors in the estimated parameters, including \hat{b}_1 and possibly \hat{u}_{ref} and \hat{c}_t . As an analysis of how combinations of measurement errors affect the results becomes quite unwieldy, I here discuss the effects of such errors one at a time.

2.3.1. Errors in \hat{u}_{ref}

Theory and simulations in Ergon (2022) show that errors in \hat{u}_{ref} give large errors in the predicted mean reaction norms, and with perfect measurements and known parameter values it is theoretically possible to compute these errors. In practice, errors in \hat{u}_{ref} may give large errors in the predicted mean reaction norms, as shown in simulations in Section 3.

2.3.2. Measurement errors in $y_{i,t}$ and/or $w_{i,t}$, with u_{ref} known

For large populations, white noise errors $\tilde{y}_{i,t}$ and $\tilde{w}_{i,t}$ will neither affect \bar{y}_t nor $cov(w_{i,t}, y_{i,t})$, and will thus have no effects on the mean trait predictions. For small population sizes, however, such random errors will not be zero mean, and $cov(\tilde{w}_{i,t}, y_{i,t})$ may not be quite zero, and simulations will show some effects on the predictions.

2.3.3. Measurement errors in u_t , with u_{ref} known

The main problem with measurement errors \tilde{u}_t as shown in Figure 1 is that overfitting may give large errors in the estimated parameter values. The true input $u_{0,t}$ is typically a white noise process added to a slow trend, as shown in Figure 2. With white noise measurement errors \tilde{u}_t it will then in practice be impossible to separate the effects of $u_{0,t}$ and \tilde{u}_t from each other, and there is therefore no option but to live with the overfitting problems caused by \tilde{u}_t . These problems are especially pronounced with short data. The underlying problem is that minimization of $\frac{1}{T} \sum_{t=1}^T (\bar{y}_t - \hat{y}_t)^2$ does not imply minimization of $\frac{1}{T} \sum_{t=1}^T (\bar{a}_t - \hat{a}_t)^2$ and $\frac{1}{T} \sum_{t=1}^T (\bar{b}_t - \hat{b}_t)^2$, unless both the model and the measurements are perfect. This will be demonstrated in simulations in Section 3.

2.4. The model validation problem

In addition to errors-in-variables problems, a second problem is validation of the model found by use of the PEM approach, and thus of the predicted mean reaction norm traits over time. This validation problem is not specific for PEM, i.e., there will be an errors-in-variables problem independent of parameter estimation method.

The typical industrial approach is to test the model against a new and independent set of input-output data (Ljung, 2010), and this is also essential for predictive models in ecology (Tredennick et al., 2021). This is obviously not possible for a model of reaction norm responses on climate change, where time

series of the environmental variable u_t are given by a single natural experiment. Under the assumption of a correct reference environment, model validation can still be performed by splitting the time series into two parts, one for modeling and one for validation, as further discussed below. However, validation of a guessed or estimated reference environment, i.e., of which environment the population is adapted to, will require comparison with data from a population that has evolved from an adapted state in a different environment, and this is not a feasible approach for wild populations. The proposed validation method therefore assumes that the guessed reference environment is correct. The validation process has three steps, using the model in Equations (1a,b) as example. First, identify an initial model by use of for example the first 3/4 of the generations, and find predictions $\hat{a}_{model,t}$ and $\hat{b}_{model,t}$ for all generations, including the remaining 1/4 of the generations. Second, identify a final model by use of all data, and find predictions $\hat{a}_{final,t}$ and $\hat{b}_{final,t}$ for all generations. Third, compare the predicted mean trait changes over all generations by use of the initial model, $\Delta_{T,model} \hat{a}_t = \hat{a}_{model,T} - \hat{a}_{model,1}$ and $\Delta_{T,model} \hat{b}_t = \hat{b}_{model,T} - \hat{b}_{model,1}$, with the same changes computed by use of the final model, i.e., with $\Delta_{T,final} \hat{a}_t = \hat{a}_{final,T} - \hat{a}_{final,1}$ and $\Delta_{T,final} \hat{b}_t = \hat{b}_{final,T} - \hat{b}_{final,1}$. For a good model this comparison will result in validation ratios $\Delta_{T,final} \hat{a}_t / \Delta_{T,model} \hat{a}_t \approx 1$ and $\Delta_{T,final} \hat{b}_t / \Delta_{T,model} \hat{b}_t \approx 1$, but a possible conclusion could be that the predictions \hat{a}_t can be trusted, while \hat{b}_t cannot. This possibility can be seen by a theoretical study of the changes $\Delta_T \hat{a}_t$ and $\Delta_T \hat{b}_t$ over all samples as functions of $\Delta_T u_t = u_T - u_1$. From Equation (1b) follows

$$\begin{aligned} \Delta_T \hat{a}_t &= \hat{a}_T - \hat{a}_1 \\ &= \hat{y}_T - \hat{b}_T u_T - (\hat{y}_1 - \hat{b}_1 u_1) \\ &= \Delta_T \hat{y}_t - \Delta_T \hat{b}_t u_T - \hat{b}_1 \Delta_T u_t, \end{aligned} \quad (5)$$

which shows that we with $\Delta_T \hat{b}_t u_T \approx 0$, and a good estimate \hat{b}_1 , may find a good prediction $\Delta_T \hat{a}_t \approx \Delta_T \hat{y}_t - \hat{b}_1 \Delta_T u_t$ also if the prediction $\Delta_T \hat{b}_t$ is poor.

2.5. Special case with $G_{bb} = 0$

For the special case with $G_{bb} = 0$, Equation (4) as well as Equation (1b) gives

$$\Delta_T \hat{a}_t = \Delta_T \hat{y}_t - \hat{b} \Delta_T u_t. \quad (6)$$

We then have what might be called constant partially adaptive plasticity (Lande, 2009), i.e., the constant plasticity plays a role in the evolution of \bar{a}_t and \bar{y}_t . It follows from Equation (6) that it in this case is not necessary to know u_{ref} in order to find the predicted change in \bar{a}_t and \bar{y}_t over all generations.

3. Simulations

3.1. Description of simulation system

As in Lande (2009), assume the reaction norm model (1a,b) with $u_{ref} = 0$, i.e., with true mean phenotypic and reaction norm traits \bar{y}_t , \bar{a}_t and \bar{b}_t . True responses are generated by the model (3a,b), with $G_{aa} = 0.5$, $G_{bb} = 0.045$, $G_{ab} = 0$, and $\sigma_v^2 = 0.5$. Also assume non-overlapping generations where the individual values $a_{i,t}$, $b_{i,t}$ and $v_{i,t}$ at each generation are drawn from populations with normal distributions around \bar{a}_t , \bar{b}_t and 0, respectively. The individual fitness function is assumed to be

$$W_{i,t} = \exp\left(-\frac{(y_{i,t} - \theta_t)^2}{2\omega^2}\right), \quad (7)$$

where θ_t is the stochastic phenotypic value that maximizes fitness, while $\omega^2 = 50$. We may for example assume that θ_t is the optimal breeding time as function of spring temperature.

Also assume a stochastic environment u_t , with mean $\mu_{U,t}$, and added zero mean Gaussian and white variations $u_{n,t}$ with variance $\sigma_{U_n}^2 = 0.5$, i.e., $u_t = \mu_{U,t} + u_{n,t}$. The population is assumed to be fully adapted to a stationary stochastic environment with mean value $\mu_{U,t} = u_{ref} = 0$. In a corresponding way assume that $\theta_t = \mu_{\Theta,t} + \theta_{n,t}$, where $\theta_{n,t}$ is zero mean Gaussian and white with variance $\sigma_{\Theta_n}^2 = 2$, and where $u_{n,t}$ and $\theta_{n,t}$ are correlated with covariance $\sigma_{\Theta_n U_n} = 0.25$. Following Lande (2009), we may assume that juveniles of generation t are exposed to the environment $u_{t-\tau}$ during a critical period of development a fraction of a generation before the adult phenotype is expressed and subjected to natural selection. Define $\theta_t = 2u_t$, which implies a linear relationship $\mu_{\Theta,t} = 2\mu_{U,t}$, variances $\sigma_{\Theta_n}^2 = 4\sigma_{U_n}^2$, and covariance $\sigma_{\Theta_n U_n} = 2\rho_\tau\sigma_{U_n}^2$, where $\rho_\tau = 0.25$ is the autocorrelation of background environmental fluctuations. The optimal value of the mean plasticity slope in a stationary stochastic environment is then $\bar{b}_{opt} = \sigma_{\Theta_n U_n} / \sigma_{U_n}^2 = 2\rho_\tau = 0.5$ (Lande, 2009), and this value is in the simulations used as initial mean slope value, i.e., $\bar{b}_1 = 0.5$. Since $\mu_{\Theta,t} = 2\mu_{U,t}$, this means that the population in a stationary stochastic environment has 25% adaptive plasticity.

3.2. Two test cases, with step and ramp inputs

Simulation results are given for two different cases, with input signals as shown in Figure 3. In the first case the mean environment $\mu_{U,t}$ is a step function from 0 to 2.5 at $t = 10$, while it in the second case is a ramp function where $\mu_{U,t}$ goes from 0 to 2.5 over 50 generations. The noisy ramp function is similar to the registered yearly mean temperatures in Oslo, Norway, from 1970 to 2020 (Norsk klimaservicesenter), using the temperature around 1960 as zero-point.

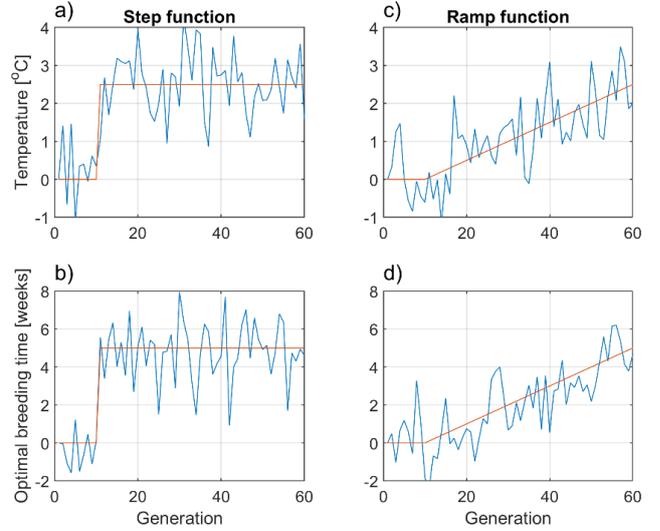


Figure 3: Input signals in the form of noisy step and ramp functions, with variances $\sigma_{U_n}^2 = 0.5$ and $\sigma_{\Theta_n}^2 = 2$, and covariance $\sigma_{\Theta_n U_n} = 0.25$. Panels b) and d) show advancements in breeding time θ_t towards earlier dates as results of increased temperatures.

3.3. Effects of random input measurement errors

As discussed in Subsection 2.3, random errors in the measured input signal u_t implies an error-in-variables problem, resulting in errors in the predicted mean reaction norm traits. The problem is that the estimated parameters \hat{G}_{bb} and $\hat{\sigma}_v^2$ are obtained by overfitting of \hat{y}_t , which results in errors in the predicted mean traits \hat{a}_t and \hat{b}_t , as shown in Figure 4. Here, recall that we in Equation (4a) can set G_{aa} to any fixed value and scale \hat{G}_{bb} and $\hat{\sigma}_v^2$ accordingly. The plots in Figure 4 also show that the very same errors in the input signal that cause large errors in especially \hat{G}_{bb} , and thus in the predicted reaction norm traits, have less influence on the predictions when a model with true parameter

values is used. For clarity it is assumed that the initial mean plasticity slope value \bar{b}_1 is known, such that PEM only has to find \hat{G}_{bb} and $\hat{\sigma}_v^2$. The prediction errors will be larger when also \bar{b}_1 must be estimated.

3.4. PEM identification and validation with correct model structure

3.4.1. Results with u_{ref} known

The true system was simulated with a population of $N = 1,000$ individuals over $T = 60$ generations (increased to $T = 120$ in two cases). Individual measurement errors $\tilde{y}_{i,t}$ and $\tilde{w}_{i,t}$ were introduced as Gaussian white noise with standard deviations that were 20% of the registered mean standard deviations of $y_{0,i,t}$ and $w_{0,i,t}$ over all generations, respectively. Measurement errors \tilde{u}_t were introduced as Gaussian white noise with standard deviation that was 20% of the standard deviation of the known random variations in u_t . Identification and validation were performed in three steps, as described in Subsection 2.4. First, data for the first 3T/4 generations (modeling set) were used for identification of a preliminary model, which was used to find predictions $\hat{a}_{modeling,t}$ and $\hat{b}_{modeling,t}$ for all generations, including the last T/4 generations. Second, a final model was found by use of all data, which gave predictions $\hat{a}_{final,t}$ and $\hat{b}_{final,t}$. Third, validation ratios $\Delta_{T,final}\hat{a}_t/\Delta_{T,model}\hat{a}_t$ and $\Delta_{T,final}\hat{b}_t/\Delta_{T,model}\hat{b}_t$ were computed, where Δ_T stands for change over $T = 60$ generations. Finally, the predicted changes in mean traits were compared with the true changes, i.e., $\Delta_{T,final}\hat{a}_t/\Delta_{T,true}\bar{a}_t$ and $\Delta_{T,final}\hat{b}_t/\Delta_{T,true}\bar{b}_t$ were computed. The lower and upper search limits for $\hat{\sigma}_v^2$ were set to $0.5G_{aa}$ and $2G_{aa}$, respectively, which means that the limits for the heritability $G_{aa}/(G_{aa} + \hat{\sigma}_v^2)$ were 0.33 and 0.67.

Results are given in Tables 1 and 2, presented as *mean* \pm *SE* values from 100 repeated simulations with different realizations of all random variables. The starting values in the numerical searches were set to $\hat{G}_{bb,start} = \hat{\sigma}_{v,start}^2 = \hat{b}_{1,start} = 0$. As pointed out in Subsection 2.4, good and validated predictions are indicated by $\Delta_{T,final}\hat{a}_t/\Delta_{T,model}\hat{a}_t \approx 1$ and $\Delta_{T,final}\hat{b}_t/\Delta_{T,model}\hat{b}_t \approx 1$. Perfect measurements of $u_{0,t}$ results in good predictions (Cases 1 and 4), while measurement errors in $u_{0,t}$ combined with short data gives poor predictions of especially \bar{b}_t (Cases 2 and 5). Note that the changes in the mean trait \bar{a}_t in Cases 2 and 5 were clearly overestimated. Also note that the predictions were improved when T was increased from $T = 60$ to $T = 120$ (Cases 3 and 6), especially in the ramp responses, but that such an increase is unrealistic

for many wild populations.

The system in Equations (3a,b) and (7), with given parameter values and ramp inputs as in Figure 3, gave what is considered to be true mean trait responses as shown in Figure 5, where also typical predicted responses are shown. Figure 5 also shows to which degree the predicted responses \hat{y}_t are similar to the true responses \bar{y}_t , which is a prerequisite for useful predictions of the mean reaction norm traits. Panel d) in Figure 5 shows for example that a model based on samples 1 to 45 gives a poor fit for the validation samples 46 to 60. The autocorrelation plot in Figure 6 shows that typical final residuals $\bar{y}_t - \hat{y}_t$ essentially are white noise sequences, which makes PEM into a maximum likelihood method (Ljung, 2010).

3.4.2. Results of errors in \hat{u}_{ref}

For Case 5 in Table 2, various values of \hat{u}_{ref} give prediction errors as shown in Table 3. Note that use of the mean value of the noisy ramp function in Figure 3, $\hat{u}_{ref} = 1.25$, results in large bias errors in both $\Delta_{T,final}\hat{a}_t$ and $\Delta_{T,final}\hat{b}_t$. Also note that these bias errors cannot be detected by the proposed validation method, for the simple reason that the modeling and validation data use the same erroneous value of \hat{u}_{ref} .

3.4.3. Various results

Since an error in \hat{u}_{ref} may give a large bias error in \hat{a}_t (Table 3), and since \hat{b}_t is difficult to predict, it is tempting to set $\hat{G}_{bb} = 0$, and use the constant estimated slope value, $\hat{b}_t = \hat{b}_1$. From Equation (6) then follows that $\Delta_T\hat{a}_t = \Delta_T\hat{y}_t - \hat{b}_1\Delta_Tu_t$, where Δ_Tu_t is independent of u_{ref} . It is also interesting to find prediction results for reduced values of G_{aa} and G_{bb} . Table 4 summarizes results for various test cases, to be discussed in Section 4. Note that Case 9 gave a very erroneous result for \hat{G}_{bb} .

4. Summary, discussion and conclusions

Studies of phenotypic responses in wild populations are often based on reaction norm models, where the environmental inputs in many cases are related to climate change. Such input drivers will never be exactly known, and there will also be measurement errors in the recorded responses, and problems related to understanding of ecological mechanisms and prediction of responses are thus error-in-variables problems, as depicted in Figure 1. Measurement errors in the input signals are especially problematic because they lead to

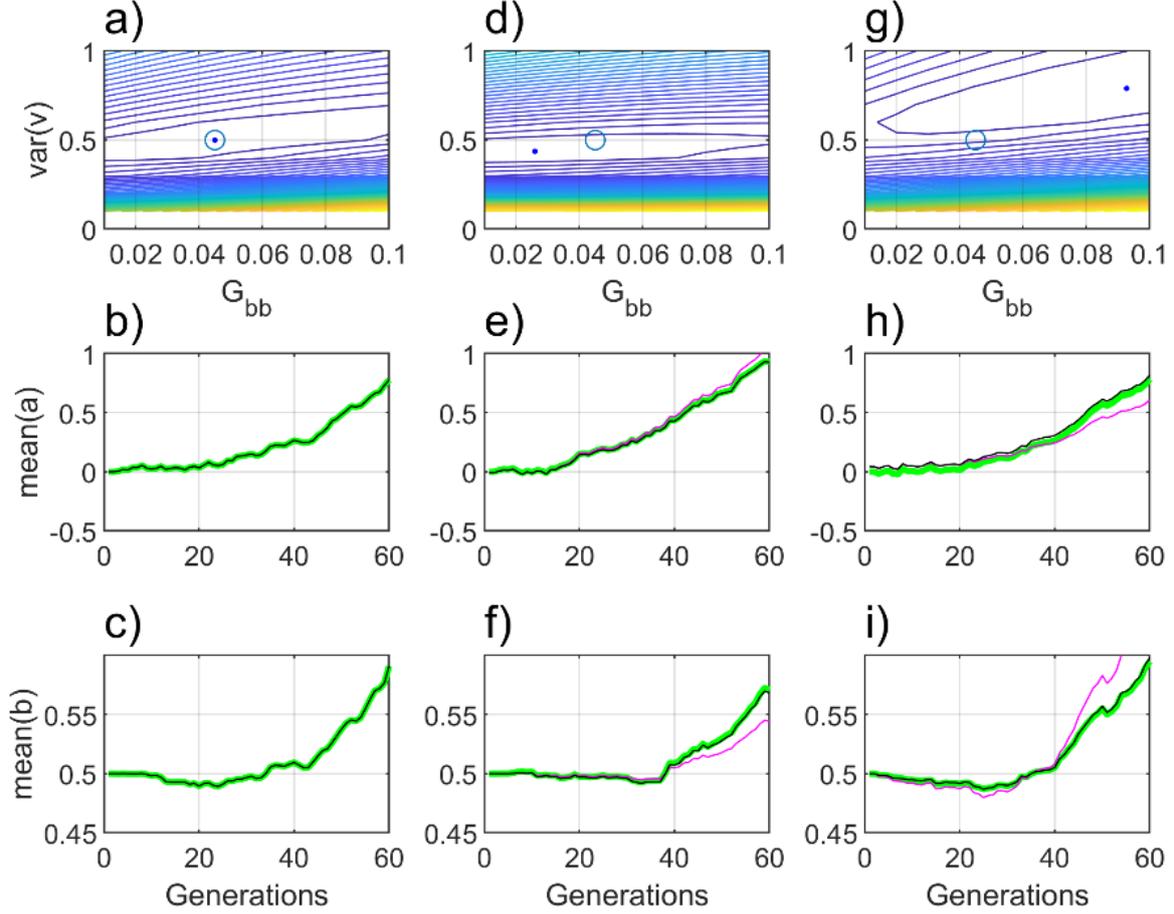


Figure 4: Contour plots with true (circles) and estimated (dots) parameter values (upper panels). Here, the x and y variables are G_{bb} and σ_v^2 , while the dependent variable is $f = \frac{1}{T} \sum_{t=1}^T (\bar{y}_t - \hat{y}_t)^2$ as function of G_{bb} and σ_v^2 . The corresponding true and predicted mean reaction norm traits are shown in the lower panels, as green and red lines, respectively. With perfect input and output measurements, the estimated parameter values and predicted responses are also perfect as shown in panels a), b) and c). Random errors in the measured input u_t result in errors in the estimated parameter values and the predicted mean traits, as shown for two realizations in panels d) to i). The white measurement errors had standard deviations that were 20% of the standard deviation of the random variations in u_t . For the two tuning results marked by dots in panels d) and g), the minimum values of $f = \frac{1}{T} \sum_{t=1}^T (\bar{y}_t - \hat{y}_t)^2$ were 25×10^{-5} and 35×10^{-5} , respectively. Response plots with the very same inputs applied on a model with true parameter values (circles in the contour plots) are shown by black lines, corresponding to $f = 29 \times 10^{-5}$ and $f = 52 \times 10^{-5}$, respectively, i.e., larger than the optimal minimum values, but still resulting in clearly better predictions.

Table 1: Estimation and prediction results for system given by Equations (3a,b) and (7), and use of the step function in Figure 3, panels a) and b). The reference environment $u_{ref} = 0$ is assumed to be known.
 Case 1: Step function, with measurement errors in $y_{0,i,t}$ and $w_{0,i,t}$.
 Case 2: Step function, with measurement errors in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$.
 Case 3: Same as Case 2, but with $T = 120$.

Parameters and prediction ratios	True or optimal values	Case 1	Case 2	Case 3
\widehat{G}_{bb}	0.045	0.045 ± 0.001	0.036 ± 0.011	0.040 ± 0.011
$\widehat{\sigma}_v^2$	0.5	0.50 ± 0.01	0.45 ± 0.06	0.47 ± 0.06
\widehat{b}_1	0.5	0.50 ± 0.00	0.50 ± 0.02	0.49 ± 0.04
f_{final}	0	$(51 \pm 13)10^{-6}$	$(83 \pm 18)10^{-4}$	$(40 \pm 6)10^{-4}$
$\Delta_{T,true}\widehat{a}_t/\Delta_{T,true}\widehat{y}_t$	-	0.40 ± 0.08	0.41 ± 0.08	0.54 ± 0.23
$\Delta_{T,true}\widehat{b}_t/\widehat{b}_1$	-	0.56 ± 0.04	0.57 ± 0.04	0.68 ± 0.04
$\Delta_{T,final}\widehat{a}_t/\Delta_{T,model}\widehat{a}_t$	1	1.00 ± 0.01	0.98 ± 0.08	1.00 ± 0.03
$\Delta_{T,final}\widehat{b}_t/\Delta_{T,model}\widehat{b}_t$	1	1.00 ± 0.02	1.10 ± 0.24	1.00 ± 0.09
$\Delta_{T,final}\widehat{a}_t/\Delta_{T,true}\widehat{a}_t$	1	1.00 ± 0.01	1.10 ± 0.11	1.05 ± 0.10
$\Delta_{T,final}\widehat{b}_t/\Delta_{T,true}\widehat{b}_t$	1	0.99 ± 0.02	0.85 ± 0.17	0.92 ± 0.16

Table 2: Estimation and prediction results for system given by Equations (3a,b) and (7), and use of the ramp function in Figure 3, panels c) and d). The reference environment $u_{ref} = 0$ is assumed to be known.
 Case 4: Ramp function, with measurement errors in $y_{0,i,t}$ and $w_{0,i,t}$.
 Case 5: Ramp function, with measurement errors in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$.
 Case 6: Same as Case 5, but with $T = 12$ and $\mu_{U,t} = 2.5$ for $t > 60$.

Parameters and prediction ratios	True or optimal values	Case 4	Case 5	Case 6
\widehat{G}_{bb}	0.045	0.045 ± 0.004	0.039 ± 0.030	0.041 ± 0.009
$\widehat{\sigma}_v^2$	0.5	0.50 ± 0.02	0.42 ± 0.11	0.44 ± 0.07
\widehat{b}_1	0.5	0.50 ± 0.00	0.48 ± 0.02	0.48 ± 0.03
f_{final}	0	$(52 \pm 12)10^{-6}$	$(73 \pm 37)10^{-4}$	$(42 \pm 10)10^{-3}$
$\Delta_{T,true}\widehat{a}_t/\Delta_{T,true}\widehat{y}_t$	-	0.35 ± 0.09	0.34 ± 0.21	0.30 ± 0.06
$\Delta_{T,true}\widehat{b}_t/\widehat{b}_1$	-	0.21 ± 0.03	0.21 ± 0.03	1.21 ± 0.04
$\Delta_{T,final}\widehat{a}_t/\Delta_{T,model}\widehat{a}_t$	1	0.99 ± 0.03	1.06 ± 0.18	1.01 ± 0.10
$\Delta_{T,final}\widehat{b}_t/\Delta_{T,model}\widehat{b}_t$	1	1.03 ± 0.19	$351 \pm 1357^*$	1.03 ± 0.16
$\Delta_{T,final}\widehat{a}_t/\Delta_{T,true}\widehat{a}_t$	1	1.00 ± 0.02	1.13 ± 0.21	1.09 ± 0.12
$\Delta_{T,final}\widehat{b}_t/\Delta_{T,true}\widehat{b}_t$	1	0.99 ± 0.07	0.85 ± 0.52	0.97 ± 0.10

* Some values of $\Delta_{T,model}\widehat{b}_t$ are close to zero.

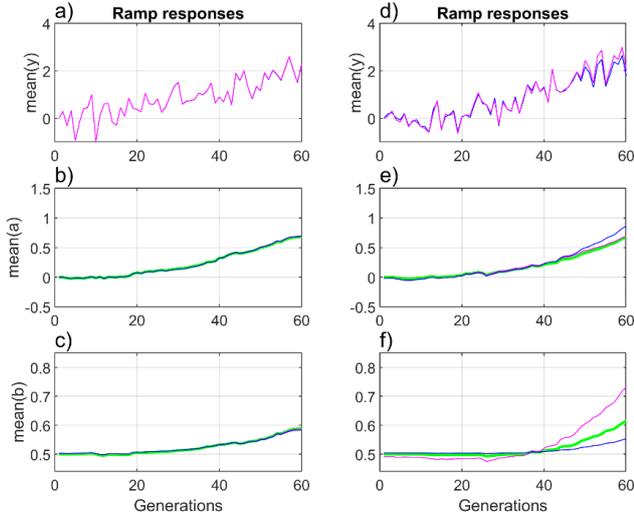


Figure 5: Typical responses on the noisy ramp function as shown in Figure 3. Responses with measurement errors in $y_{0,i,t}$ and $w_{0,i,t}$ are shown in panels a), b) and c), and responses with measurement errors also in $u_{0,t}$ are shown in panels d), e) and f). Panels a) and d) show \bar{y}_t (blue lines) and $\hat{\bar{y}}_t$ based on the modeling data (red lines). Panels b) and e) show the true responses \bar{a}_t (green lines), and the predicted responses $\hat{\bar{a}}_t$ based on the modeling data (red lines) and on all data (blue lines). Panels c) and f) show the true responses \bar{b}_t (green lines), and the predicted responses $\hat{\bar{b}}_t$ based on the modeling data (red lines) and on all data (blue lines).

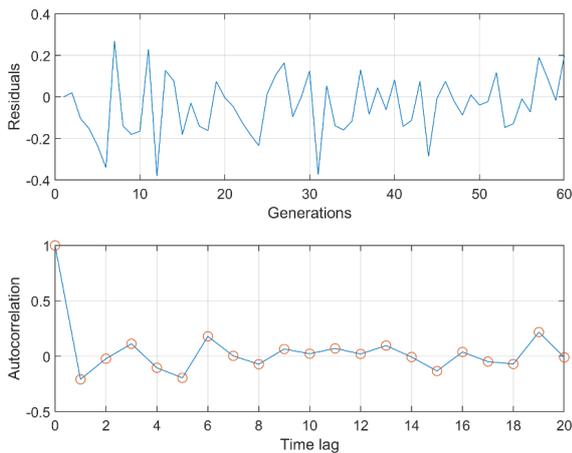


Figure 6: Typical final residuals $\bar{y}_t - \hat{\bar{y}}_t$ corresponding to the ramp responses in Figure 5, panels d) to f), and the resulting autocorrelation plot.

Table 3: Estimation and prediction results for ramp responses, with guessed values of u_{ref} and measurement errors in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$ (as in Case 5 in Table 2).

\hat{u}_{ref}	$\frac{\Delta_{T,final}\hat{a}_t}{\Delta_{T,true}\bar{a}_t}$	$\frac{\Delta_{T,final}\hat{b}_t}{\Delta_{T,true}\bar{b}_t}$
-1.25	0.97 ± 0.29	0.88 ± 0.55
-0.5	1.05 ± 0.26	0.96 ± 0.59
0	1.11 ± 0.21	0.90 ± 0.54
0.5	1.18 ± 0.17	0.82 ± 0.48
1.25	1.27 ± 0.10	0.46 ± 0.31

overfitting in parameter estimation, independent of estimation method.

A second important feature of reaction norm response problems is that the environmental inputs must be given an appropriate reference value. This is obvious with an interval scaled input like temperature, but it is just as necessary for a ratio scaled input like salinity. As discussed in Ergon (2022), the proper reference value is the value where the phenotypic variance has a minimum such that geometric mean fitness is maximized, i.e., the environmental value the population is adapted to.

Prediction of evolutionary responses on various ecological drivers must necessarily be accompanied by validation methods (Ljung, 2010; Tredennick et al., 2021), and in order to focus on the essential errors-in-variables and reference environment problems I have here used a reaction norm model in its simplest univariate form, characterized by a mean intercept value and a mean plasticity slope value. I have proposed a validation method where the available input-output data is split into one part for modeling and one part for validation. As shown in simulations in Section 3 and Appendix A, validation of the mean reaction norm trait predictions is difficult due to the short time-series that are normally available from field studies of wild populations. I have also argued that validation of a chosen or estimated reference environment in practice is impossible, when the data comes from the (unintended) anthropogenic global warming experiment, where no independent experimental data with a different reference environment exists.

In simulations for studies of the overfitting and validation problems, I have used a variant of the multivariate breeder's equation where selection with respect to the unknown individual reaction norm traits is replaced by selection with respect to the individual phenotypic trait, which is assumed to be known from field data (Ergon, 2022). I use a prediction error method (PEM), that is well established in the engineering control field (Ljung, 1999), but it should be noted that

Table 4: Prediction results for ramp responses in various test cases with measurement errors in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$.

 Case 7: $G_{aa} = 0.5$ and $G_{bb} = 0.045$, and $\hat{G}_{bb} = 0$.

 Case 8: $G_{aa} = 0.5$ and $G_{bb} = 0.045$, and $\hat{G}_{bb} = 0$ and $\hat{\sigma}_v^2 = 0$.

 Case 9: $G_{aa} = 0.05$ and $G_{bb} = 0.0045$, and \hat{G}_{bb} and $\hat{\sigma}_v^2$ as free variables.

Test cases	\hat{G}_{bb}	\hat{b}_1	$\frac{\Delta_{T,final}\hat{a}_t}{\Delta_{T,true}\bar{a}_t}$	$\frac{\Delta_{T,final}\hat{b}_t}{\Delta_{T,true}\bar{b}_t}$
Case 7 ($G_{bb} = 0.045$)	0	0.48 ± 0.02	1.25 ± 0.21	0
Case 8 ($G_{bb} = 0.045$)	0	0.36 ± 0.04	2.33 ± 0.08	0
Case 9 ($G_{bb} = 0.0045$)	0.022 ± 0.035	0.49 ± 0.04	1.10 ± 0.46	1.72 ± 2.16

overfitting due to errors-in-variables and short data is not a specific problem for PEM, errors in the assumed environmental driver will result in overfitting regardless of the choice of parameter estimation method.

A typical ecological input signal is shown in Figure 2, where it can be seen that it essentially is white noise on top of a slow trend. Figure 3 shows the true input signals used in simulations, where the noisy ramp function is similar to typical temperature changes caused by global warming. Figure 4 illustrates that random errors in ecological drivers lead to overfitting when the squared prediction error is minimized, and that true parameters give better predictions in spite of the fact that they give a larger squared prediction error. This calls for some sort of regularization, but it is unclear how that should be done, if at all possible. Figure 5 shows typical responses on the noisy ramp input signal in Figure 3, with added input and output measurement errors, and illustrates that it is easier to predict the mean reaction norm incident than the mean reaction norm slope. Figure 6 shows that the PEM residuals in the current setting is a white noise sequence, and that PEM thus becomes a maximum likelihood method (Ljung, 2010).

Table 1 shows that parameter estimates and reaction norm predictions with use of the noisy step function in Figure 3 are quite good also with random output measurement errors (Case 1), but it is then assumed that the reference environment is known. When random input measurement errors are added (Case 2), the results are poorer but still fairly good. Table 2 shows that the noisy ramp input in Figure 3 with random measurement errors results in large SE values for the mean reaction norm predictions, especially for the mean slope prediction (Case 5). When the number of generations is increased from 60 to 120 (Case 6), the prediction results are clearly improved. The practical problem is thus not only the input measurement errors as such, but the combination with short data. A sufficiently large number of generations is, however, not realistic for many of the studies of wild populations. Table 3 shows that errors in the environmental reference value

may give large prediction errors. Mean centering of the ramp function in Figure 3, will for example give around 30% overestimation of the changes in mean intercept value over 60 generations.

Because input measurement noise makes it difficult to find good and validated predictions of the mean plasticity slope, it is tempting to assume a constant plasticity slope by setting $\hat{G}_{bb} = 0$. In Case 7 in Table 4 this resulted in 25% overestimation of the change in \bar{a}_t . The change in \bar{a}_t was overestimated also in Cases 2 and 5, but then by only around 10%. A possible explanation of this overestimation is that the measurement error \tilde{u}_t is filtered through the non-linear filter given by the prediction model in Equation (4), and that the estimation error increases when $\hat{G}_{bb} = 0$. There is thus no point in setting $\hat{G}_{bb} = 0$, and one should definitely not set $\hat{\sigma}_v^2 = 0$, because that results in very large errors in \hat{a}_t and a large bias in the estimated initial mean slope value \hat{b}_1 (Case 8 in Table 4). With small values of G_{aa} and G_{bb} (Case 9 in Table 4) we find a very large error in \hat{G}_{bb} , as further discussed below.

A first main conclusion is that the typically short data from field studies of wild populations make it difficult to predict changes in the mean reaction norm slope value \bar{b}_t , but that the simulation results for changes in the mean intercept value \bar{a}_t are somewhat more promising, at least for the fairly large changes in the main examples. The validated ramp response case with input measurement noise (Case 5) gave $\Delta_{T,final}\hat{a}_t/\Delta_{T,true}\bar{a}_t = 1.13 \pm 0.21$ as mean and SE values based on 100 realizations, and four different realizations in Appendix A gave 1.20, 0.99, 1.14 and 0.85. Provided a correct model, the result for a single realization will thus at least give an indication of the change in mean intercept value over all generations. The size of the white input measurement errors used in simulations may be pessimistic, but the simulations assume a perfectly known model structure, which is not realistic, and the input measurements may also have offset errors. It is also promising that the true initial mean slope value, $\bar{b}_1 = 0.5$, in all normal cases was estimated

well, from 0.48 ± 0.02 (Case 5) to 0.50 ± 0.02 (Case 2). The exception is Case 8 in Table 4, where $\hat{\sigma}_v^2 = 0$ gave $\hat{b}_1 = 0.36 \pm 0.04$.

A second main conclusion is that the proposed validation method is not very reliable for the main cases with $T = 60$ generations. As shown in Appendix A, Examples 1 and 4, there may then occur realizations of ramp responses that give large errors in the mean reaction norm slope predictions, also with seemingly good validation results. This type of problem did not appear when the number of samples was increased to $T = 120$, with $\mu_{U,t} = 2.5$ for $t > 60$.

A third main conclusion is that an assumed or estimated reference environment cannot be validated given data from a specific field study. This is a serious problem since errors in the assumed reference environment may give large prediction errors (Table 3). When studying the effects of the recent global warming it may be tempting to assume that the population was adapted to the mean temperatures in the 1960s, but note that it was around 0.5 °C colder in the mid 1800s, and that evolution often is slow. It should also be noted that estimation of the reference environment by use of the BLUP/PEM optimization method described in Ergon (2023b) gives better results than the gradient method used here, but that this is of limited value without a possibility for validation.

A fourth main conclusion is that small changes in the mean reaction norm traits over all generations result in very unreliable predictions of these changes. Such small changes may be the result of small values of the genetic variances, as in Case 9 in Table 4, where \hat{G}_{bb} was very much larger than the true value. Small changes in the mean reaction norm may also be the result of overlapping generations and long lifetimes, where only a fraction $f_t < 1$ of the population is affected by the changes in fitness such that the evolution will be slowed down accordingly (Ergon, 2023a). As discussed below, however, small changes in \bar{a}_t and \bar{b}_t over all generations may also be a result of near optimal adaptive plasticity.

As a short summary of the four main conclusions above, it seems safe to say that disentanglement of plastic and genetic responses on climate change based on realistically short data is a difficult task, and that especially predicted changes in \bar{b}_t are unreliable, except that the initial mean slope value \bar{b}_1 is estimated well. The best option may thus in many cases be to conclude that the mean plasticity slope is constant, $\bar{b} = \bar{b}_1$, but that does not mean that one should assume that $G_{bb} = 0$ in the tuning of the prediction model in Equation (4). As seen in Table 4, Case 7, $\Delta_{T,final}\hat{a}_t/\Delta_{T,true}\bar{a}_t$ increases to 1.25 ± 0.21 with $\hat{G}_{bb} = 0$, but this increase will be lower with a smaller variance G_{bb} . There is in any case

no point in setting $\hat{G}_{bb} = 0$ in the search for parameter values, but because of the poor predictions of changes in \hat{b}_t it may be a point in setting $\hat{b}_t = \hat{b}_1$.

In the simulations, the plasticity in a stationary stochastic environment is 25% partially adaptive, that is, the plastic phenotypic change is in the adaptive direction, but with only 25% of the optimal magnitude. As shown in Appendix B, an increase to 50% partial adaptivity reduces the changes in \bar{a}_t and \bar{b}_t over 60 generations and makes it more difficult to predict these changes. With 100% adaptivity, the initial mean trait values \bar{b}_1 and \bar{a}_1 will not evolve when the environment changes, and it will then be impossible to estimate the system parameters. Such an optimal plasticity was an initial assumption in Valdés et al. (2023) (where G_{aa} and G_{bb} were far from zero), although the authors eventually found that the plasticity was in fact somewhat larger than optimal (maladaptive plasticity).

The simulations with small values of G_{aa} and G_{bb} (Case 9 in Table 4) revealed an identification problem with near 100% adaptivity as discussed above. It is not enough that the system is excited by variations in the environmental input, as shown in Figure 2, there must also be sufficient changes in \bar{a}_t and \bar{b}_t over time. The apparent reason for this can be seen in Equation (1b), $\bar{y}_t = \bar{a}_t + \bar{b}_t(u_t - u_{ref})$, from which follows that when \bar{a}_t and \bar{b}_t are approximately constant only $\bar{b}_t \approx \bar{b}_1$ can be estimated from input-output data. The difficulties to estimate G_{bb} when there are small changes in \bar{a}_t and \bar{b}_t over time raises an interesting theoretical question about persistent excitation in reaction norm identification, but that must be left for further research.

Finally note that I in this article have assumed non-overlapping generations. Field studies of species with overlapping generations and long lifetimes give opportunities to determine individual plasticity, and possibly also to find the reference environment where the phenotypic variance has a minimum. In an example of this approach, Valdés et al. (2023) used 22 years of field observations of the perennial forest herb *Lathyrus vernus* in order to assess phenotypic selection of flowering time as function of spring temperature, as shown in Figure 2. Although they did not emphasize this result, Figure 1 in their paper shows that the reference environment where the phenotypic variance has a minimum is the mean temperature in the mid 1800s. This is not surprising in light of the estimated lifetime of 43 years, which means very much overlapping generations and a correspondingly slow evolution.

Acknowledgments

I thank University of South-Eastern Norway, Campus Porsgrunn, for support.

References

- Arnold, P. A., Kruuk, L. E., and Nicotra, A. B. How to analyse plant phenotypic plasticity in response to a changing climate. *New Phytologist*, 2019. 222(3):1235–1241. doi:[10.1111/nph.15656](https://doi.org/10.1111/nph.15656).
- Bonamour, S., Chevin, L.-M., Charmantier, A., and Teplitsky, C. Phenotypic plasticity in response to climate change: the importance of cue variation. *Phil. Trans. R. Soc. B*, 2019. 374:20180178. doi:<http://dx.doi.org/10.1098/rstb.2018.0178>.
- Ergon, R. Quantitative genetics state-space modeling of phenotypic plasticity and evolution. *Modeling, Identification and Control*, 2019. 40:51–69. doi:[10.4173/mic.2019.1.5](https://doi.org/10.4173/mic.2019.1.5).
- Ergon, R. The important choice of reference environment in microevolutionary climate response predictions. *Ecology and Evolution*, 2022. 12(4):e8836. doi:[10.1002/ece3.8836](https://doi.org/10.1002/ece3.8836).
- Ergon, R. Dynamical blup modeling of reaction norm evolution, accommodating changing environments, overlapping generations, and multivariate data. *Ecology and Evolution*, 2023a. 13:e10194. doi:[10.1002/ece3.10194](https://doi.org/10.1002/ece3.10194).
- Ergon, R. Microevolutionary system identification and climate response predictions by use of blup prediction error method. *Identification and Control*, 2023b. 44(3):83–110. doi:[10.4173/mic.2023.3.1](https://doi.org/10.4173/mic.2023.3.1).
- J., S., Daly, C., Foster, S., Ashcroft, M., and Warton, D. A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution*, 2015. 6:412–423. doi:[10.1111/2041-210X.12217](https://doi.org/10.1111/2041-210X.12217).
- Kangas, A. Effect of errors-in-variables on coefficients of a growth model and on prediction of growth. *Forest Ecology and Management*, 1998. 102:203–212.
- Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 1979. pages 402–416. doi:[10.2307/2407630](https://doi.org/10.2307/2407630).
- Lande, R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of evolutionary biology*, 2009. 22(7):1435–1446. doi:[10.1111/j.1420-9101.2009.01754.x](https://doi.org/10.1111/j.1420-9101.2009.01754.x).
- Ljung, L. *System Identification: Theory for the User*. 2nd Edition, Prentice-Hall, Upper Saddle River, NJ, 1999.
- Ljung, L. Perspectives on system identification. *Annual Reviews in Control*, 2010. 34(1):1–12. doi:[10.1016/j.arcontrol.2009.12.001](https://doi.org/10.1016/j.arcontrol.2009.12.001).
- Merilä, J. and Hendry, A. P. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary applications*, 2014. 7(1):1–14. doi:[10.1111/eva.12137](https://doi.org/10.1111/eva.12137).
- Morrissey, M., Kruuk, L., and Wilson, A. The danger of applying the breeder’s equation in observational studies of natural populations. *J. Evol. Biol.*, 2010. 23:2277–2288. doi:[10.1111/j.1420-9101.2010.02084.x](https://doi.org/10.1111/j.1420-9101.2010.02084.x).
- Price, G. R. et al. Selection and covariance. *Nature*, 1970. 227:520–521.
- Réale, D., McAdam, A., Boutin, S., and Berteaux, D. Genetic and plastic responses of a northern mammal to climate change. *Proc. Royal Soc. Lond.*, 2003. 270:591–596. doi:[10.1098/rspb.2002.2224](https://doi.org/10.1098/rspb.2002.2224).
- Söderström, T. A user perspective on errors-in-variables methods in system identification. *Control Engineering Practice*, 2019. doi:[10.1016/j.conengprac.2019.05.013](https://doi.org/10.1016/j.conengprac.2019.05.013).
- Tarka, M., B., H., and Hasselquist, D. Selection and evolutionary potential of spring arrival phenology in males and females of a migratory songbird. *J. Evol. Biol.*, 2015. 28:1024–1038. doi:[10.1111/jeb.12638](https://doi.org/10.1111/jeb.12638).
- Tredennick, A., Hooker, G., Ellner, S., and Adler, P. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 2021. 102(6):e03336.
- Valdés, A., Arnold, P., and Ehrlén, J. Spring temperature drives phenotypic selection on plasticity of flowering time. *Proc. R. Soc. B*, 2023. 290:20230670. doi:[10.1098/rspb.2023.0670](https://doi.org/10.1098/rspb.2023.0670).

A. Further examples with 25% adaptive plasticity

Figures [A.1](#) and [A.2](#) show four examples of ramp responses with measurement noise in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$ (Case 5 in [Table 2](#)), with prediction results given in [Table A.1](#). Examples 1 and 4 show uncommon but possible extreme responses, which seems to represent good models, but where especially the predictions of \bar{b}_t are very poor. This type of responses did not appear when the number of samples was increased to $T = 120$, with $\mu_{U,t} = 2.5$ for $t > 60$.

B. Example with 50% adaptive plasticity

In the system in the simulations in Section 3 and in Appendix A, the environment that maximizes fitness is given by $\theta_t = 2u_t$, which implies a linear relationship $\mu_{\Theta,t} = 2\mu_{U,t}$ and the variance $\sigma_{\Theta_n}^2 = 4\sigma_{U_n}^2$. This gives the initial mean plasticity slope $\bar{b}_1 = 0.5$, and thus 25% adaptive plasticity. With $\mu_{\Theta,t} = \mu_{U,t}$ and $\sigma_{\Theta_n}^2 = \sigma_{U_n}^2$, the adaptive plasticity increases to 50%, and as shown in Table B.1 and compared with Case 5, this results in smaller changes of \bar{a}_t and \bar{b}_t over 60 generations, and larger errors in the predicted changes.

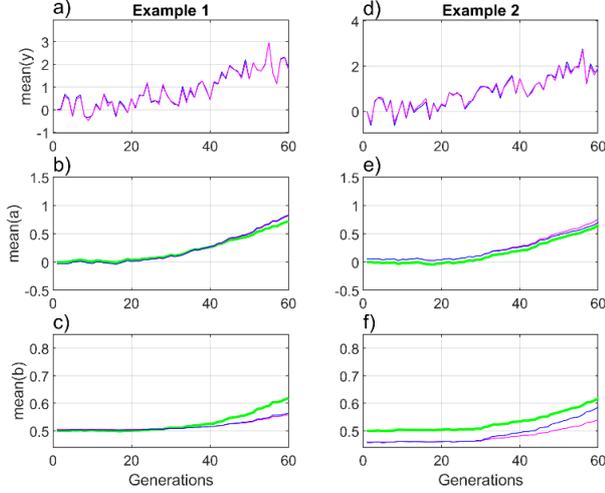


Figure A.1: Two examples of ramp responses with 25% partial adaptivity and measurement noise in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$, where Example 1 clearly underestimates G_{bb} and $\Delta_T \bar{b}_t$. See Figure 5 caption for details.

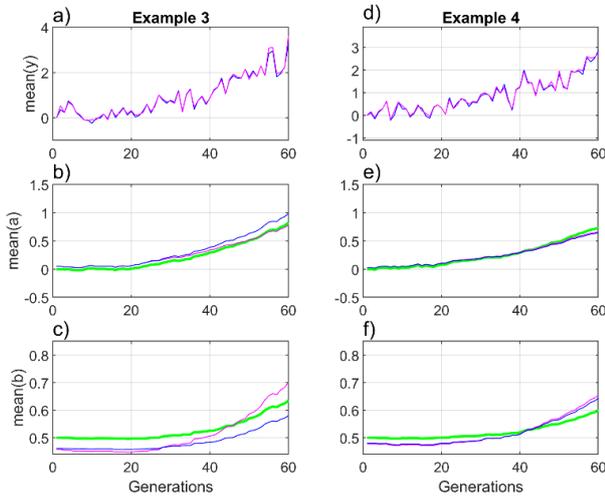


Figure A.2: Two additional examples of ramp responses with 25% partial adaptivity and measurement noise in $y_{0,i,t}$, $w_{0,i,t}$ and $u_{0,t}$, where Example 4 clearly overestimates G_{bb} and $\Delta_T \bar{b}_t$. See Figure 5 captions for details.

Table A.1: Prediction results for the four examples in Figures A.1 and A.2.

Parameters and prediction ratios	True or optimal values	Example 1	Example 2	Example 3	Example 4
\widehat{G}_{bb}	0.045	0.018	0.036	0.035	0.094
$\widehat{\sigma}_v^2$	0.5	0.39	0.49	0.39	0.57
\widehat{b}_{init}	0.5	0.50	0.46	0.46	0.48
$10^4 f_{final}$	0	42	53	95	52
$\Delta_{T,true}\bar{a}_t/\Delta_{T,true}\bar{y}_t$	-	0.39	0.36	0.25	0.26
$\Delta_{T,true}\bar{b}_t/\bar{b}_1$	-	0.24	0.24	0.27	0.20
$\Delta_{T,final}\widehat{\bar{a}}_t/\Delta_{T,model}\bar{a}_t$	1	1.02	0.92	1.28	0.97
$\Delta_{T,final}\widehat{\bar{b}}_t/\Delta_{T,model}\bar{b}_t$	1	1.12	1.59	0.49	0.92
$\Delta_{T,final}\widehat{\bar{a}}_t/\Delta_{T,true}\bar{a}_t$	1	1.20	0.99	1.14	0.85
$\Delta_{T,final}\widehat{\bar{b}}_t/\Delta_{T,true}\bar{b}_t$	1	0.51	1.09	0.88	1.65

Table B.1: Prediction results for example with 50% adaptive plasticity.

Parameters and prediction ratios	True or optimal values	Case 5 with 25% adaptive plasticity	Case 5 with 50% adaptive plasticity
\widehat{G}_{bb}	0.045	0.039 ± 0.030	0.047 ± 0.043
$\widehat{\sigma}_v^2$	0.5	0.42 ± 0.11	0.38 ± 0.18
\widehat{b}_1	0.5	0.48 ± 0.02	0.49 ± 0.03
f_{final}	0	$(73 \pm 37)10^{-4}$	$(74 \pm 36)10^{-3}$
$\Delta_{T,true}\bar{a}_t/\Delta_{T,true}\bar{y}_t$	-	0.34 ± 0.21	0.16 ± 0.05
$\Delta_{T,true}\bar{b}_t/\bar{b}_1$	-	0.21 ± 0.03	0.06 ± 0.01
$\Delta_{T,final}\widehat{\bar{a}}_t/\Delta_{T,model}\bar{a}_t$	1	1.06 ± 0.18	1.10 ± 0.25
$\Delta_{T,final}\widehat{\bar{b}}_t/\Delta_{T,model}\bar{b}_t$	1	$351 \pm 1357^*$	$30 \pm 81^*$
$\Delta_{T,final}\widehat{\bar{a}}_t/\Delta_{T,true}\bar{a}_t$	1	1.13 ± 0.21	1.19 ± 0.27
$\Delta_{T,final}\widehat{\bar{b}}_t/\Delta_{T,true}\bar{b}_t$	1	0.85 ± 0.52	0.96 ± 0.84

* Some values of $\Delta_{T,model}\widehat{\bar{b}}_t$ are close to zero.

C. MATLAB code

```

%-----
% Parameter estimation of intercept-slope model with
% fmincon() in Matlab, GRAD algorithm.

% Rolf Ergon (rolf.ergon@usn.no)
% University of South-Eastern Norway
% March 15, 2024
%-----

clear

for m=1:100
    m
    x=1;
    T=60;
    z=3*T/4;
    N=1000;
    wsquare=50;
    vartheta=0.5;
    varu=2;
    rho=0.25;
    Gaa=0.5;
    varv=Gaa;
    Gbb=0.045;
    % Gbb=0;
    Paa=Gaa+varv;
    Pbb=Gbb;
    G=[Gaa 0 ; 0 Gbb];
    P=[Paa 0 ; 0 Pbb];

    %% Generate u and theta sequences
    uplot=zeros(1,T);
    du=zeros(1,T);
    dtheta=zeros(1,T);
    for i=2:T
        du(i)=sqrt(varu)*randn;
        dtheta(i)=du(i)*rho*sqrt(vartheta/(varu))+sqrt(vartheta*(1-rho^2))*randn
        ;
        if i>10
            uplot(i)=(i-10)/20;           % Ramp function
            if i>60
                uolot(i)=2.5;
            end
        %         uplot(i)=2.5;           % Step function
    end
end
u0=uplot+du;
theta=2*uplot+dtheta;

%% Individual population traits around abar, bbar1 and bbar2
for t=1:T
    a(:,t)=sqrt(Gaa)*randn(N,1);

```

```

    b(:,t)=sqrt(Gbb)*randn(N,1);
    v(:,t)=sqrt(varv)*randn(N,1);
    a(:,t)=a(:,t)-mean(a(:,t));
    b(:,t)=b(:,t)-mean(b(:,t));
    v(:,t)=v(:,t)-mean(v(:,t));
end

%% Simulation of true system
abar0=zeros(1,T);
bbar0=0.5*ones(1,T);
ybar0=0*ones(1,T);
for t=1:T-1
    ybar0(x)=0;
    abar0(x)=ybar0(x)-bbar0(x)*u0(x);
    bterm=(bbar0(t)+b(:,t)).*u0(t);
    Y0(:,t)=abar0(t)+a(:,t)+v(:,t)+bterm;
    W0(:,t)=exp(-(Y0(:,t)-theta(t)).^2/(2*wsquare));
    Wbar0(t)=mean(W0(:,t));
    Pyy=Gaa+varv+u0(t)^2*Gbb;
    covWy=(N-1)*cov(W0(:,t),Y0(:,t))/N;
    covWyt(t)=covWy(1,2);
    betay=inv(Pyy)*covWy(1,2)/(1*Wbar0(t));
    betayplot(t)=betay;
    abar0(t+1)=abar0(t)+Gaa*betay;
    bbar0(t+1)=bbar0(t)+Gbb*u0(t)*betay;
    ybar0(t+1)=abar0(t+1)+bbar0(t+1)*u0(t+1);
end

deltaabar(m)=(abar0(T)-abar0(1))/(ybar0(T)-ybar0(1));
deltabbar(m)=(bbar0(T)-bbar0(1))/0.5;

%% Measurement errors
Y=Y0;
u=u0;
Y=Y0+0.2*randn(N,T-1);
u=u0+0.2*sqrt(varu)*randn(1,T);
ybar=[mean(Y) ybar0(T)];
W=W0;
W=W0+0.08*randn(N,T-1);
Wbar=mean(W);

for i=1:2
    if i==2
        z=T;
    end

%% Constraints
Gbb_min=0; Gbb_max=0.1;
% Gbb_min=0; Gbb_max =0.0000001;
varv_min=Gaa/2; varv_max=2*Gaa;
% varv_min=0; varv_max =0.0000001;

bbarinit_min=0; bbarinit_max=2;

```

```

par_lb=[Gbb_min varv_min bbarinit_min];
par_ub=[Gbb_max varv_max bbarinit_max];

par_guess=zeros(1,3);
par_known=[Gaa];
ybarhat_init=0;

%% fmincon
Aineq=[]; Bineq=[]; Aeq=[]; Beq=[];
fun_objective_SISO_Example_handle=...
    @(par)fun_objective_SISO_Example(par,par_known,u,Y,ybar,W,Wbar,
        ybarhat_init,N,T,x,z);
[par_opt,fval,exitflag,output,lambda,grad,hessian]=...
fmincon(fun_objective_SISO_Example_handle,par_guess,Aineq,Bineq,Aeq,Beq,
    par_lb,par_ub);
output;

if i==1
    opt_results1(m,:)=par_opt;
else
    opt_results2(m,:)=par_opt;
end

%% Model test
Gbbopt=opt_results1(m,1);
varvopt=opt_results1(m,2);
bbaropt11=opt_results1(m,3);

abarhatmodel=0*ones(1,T);
bbarhatmodel=0.5*ones(1,T);
ybarhatmodel=0*ones(1,T);
for t=1:T-1
    ybarhatmodel(x)=0;
    bbarhatmodel(x)=bbaropt11;
    abarhatmodel(x)=ybarhatmodel(x)-bbarhatmodel(x)*u(x);
    Pyy=Gaa+varvopt+u(t)^2*Gbbopt;
    covWy=(N-1)*cov(W(:,t),Y(:,t))/N;
    betay=inv(Pyy)*covWy(1,2)/Wbar(t);
    abarhatmodel(t+1)=abarhatmodel(t)+Gaa*betay;
    bbarhatmodel(t+1)=bbarhatmodel(t)+Gbbopt*u(t)*betay;
    ybarhatmodel(t+1)=abarhatmodel(t+1)+bbarhatmodel(t+1)*u(t+1);
end

%% Model test results
if i==1
fmodel(m)=sum((ybar(x:z)-ybarhatmodel(x:z)).^2)/(z-x);
fvalidation(m)=sum((ybar(z+1:T)-ybarhatmodel(z+1:T)).^2)/(T-z-1);

Totalabar=abar0(T)-abar0(x);
Totalabarmodel=abarhatmodel(T)-abarhatmodel(x);
Rel_abarmodel(m)=Totalabarmodel/Totalabar;

Totalbbar0z=bbar0(T)-bbar0(x);
Totalbbarmodel=bbarhatmodel(T)-bbarhatmodel(x);

```

```

Rel_bbarmodel(m)=Totalbbarmodel/Totalbbar0z;

fmodellast(m)=sum((ybar(45+1:T)-ybarhatmodel(45+1:T)).^2)/(T-45-1);

end
end

%% Final test
Gbbopt=opt_results2(m,1);
varvopt=opt_results2(m,2);
bbaropt=opt_results2(m,3);

abarfina1=0*ones(1,T);
bbarfina1=0.5*ones(1,T);
ybarfina1=0*ones(1,T);
for t=1:T-1
    ybarfina1(x)=0;
    bbarfina1(x)=bbaropt;
    abarfina1(x)=ybarfina1(x)-bbarfina1(x)*u(x);
    Pyy=Gaa+varvopt+u(t)^2*Gbbopt;
    covWy=(N-1)*cov(W(:,t),Y(:,t))/N;
    betay=inv(Pyy)*covWy(1,2)/Wbar(t);
    abarfina1(t+1)=abarfina1(t)+Gaa*betay;
    bbarfina1(t+1)=bbarfina1(t)+Gbbopt*u(t)*betay;
    ybarfina1(t+1)=abarfina1(t+1)+bbarfina1(t+1)*u(t+1);
end

ffina1(m)=sum((ybar(x:T)-ybarfina1(x:T)).^2)/T;
ffina1last(m)=sum((ybar(45+1:T)-ybarfina1(45+1:T)).^2)/(T-45-1);

Totalabar0T=abar0(T)-abar0(x);
Totalabarfina1=abarfina1(T)-abarfina1(x);
Rel_abarfina1(m)=Totalabarfina1/Totalabar0T;

Totalbbar0T=bbar0(T)-bbar0(x);
Totalbbarfina1=bbarfina1(T)-bbarfina1(x);
Rel_bbarfina1(m)=Totalbbarfina1/Totalbbar0T;

%% Final test results
Totalabar0T=abar0(T)-abar0(x);
Totalabarfina1=abarfina1(T)-abarfina1(x);
Rel_abarfina1(m)=Totalabarfina1/Totalabar0T;

Totalbbar0T=bbar0(T)-bbar0(x);
Totalbbarfina1=bbarfina1(T)-bbarfina1(x);
Rel_bbarfina1(m)=Totalbbarfina1/Totalbbar0T;

%% Validation results
Rel_abarvalidation(m)=Totalabarfina1/Totalabarmodel;
Rel_bbarvalidation(m)=Totalbbarfina1/Totalbbarmodel;

end
%% Plots

```

```

figure(1)
q=ybar0-ybarfinal; % Rettet
subplot(2,1,1),plot(q), grid
xlabel('Generations')
ylabel('Residuals')
autocorrq=autocorr(q);
lags=zeros(1,21);
for lag=1:21
    lags(lag)=lag-1;
end
subplot(2,1,2), plot(autocorrq), hold on, grid
plot(autocorrq,'o'), hold off
xlabel('Time lag')
ylabel('Autocorrelation')

figure(2)
subplot(3,2,1)
plot(ybar0,'b'), hold on
plot(ybarhatmodel,'m')
hold off, grid
ylabel('mean(y)')
axis([0 T -1 4])
title('Ramp responses')
text(0,4.6,'a'),'FontSize',12)

subplot(3,2,3)
plot(abar0,'g','LineWidth',1.5),hold on
plot(abarhatmodel,'m')
plot(abarfinal,'b')
ylabel('mean(a)')
hold off, grid
axis([0 T -0.5 1.5])
text(0,1.75,'b'),'FontSize',12)

subplot(3,2,5)
plot(bbar0,'g','LineWidth',1.5), hold on
plot(bbarhatmodel,'m')
plot(bbarfinal,'b')
hold off, grid
xlabel('Generations')
ylabel('mean(b)')
axis([0 T 0.44 0.85])
text(0,0.90,'c'),'FontSize',12)

%% Results
Mean_par=mean(opt_results2)
Std_par=std(opt_results2)
Mean_f=mean(ffinal);
Std_f=std(ffinal);
f_resultsfinal=[Mean_f Std_f]

%% True change
Mean_deltaaabar=mean(deltaaabar);
Std_deltaaabar=std(deltaaabar);

```

```

Change_abar=[Mean_deltaabar Std_deltaabar]
Mean_deltabbar=mean(deltabbar);
Std_deltabbar=std(deltabbar);
Change_bbar=[Mean_deltabbar Std_deltabbar]

%% Validation results
Mean_abarvalidation = mean(Rel_abarvalidation);
Std_abarvalidation=std(Rel_abarvalidation);
Validation_ratio_abar=[Mean_abarvalidation Std_abarvalidation]
Mean_bbarvalidation = mean(Rel_bbarvalidation);
Std_bbarvalidation=std(Rel_bbarvalidation);
Validation_ratio_bbar=[Mean_bbarvalidation Std_bbarvalidation]

%% Final results
Mean_abarfinal = mean(Rel_abarfinal);
Mean_bbarfinal = mean(Rel_bbarfinal);
Std_abarfinal=std(Rel_abarfinal);
Std_bbarfinal=std(Rel_bbarfinal);
Final_abarresults=[Mean_abarfinal Std_abarfinal]
Final_bbarresults=[Mean_bbarfinal Std_bbarfinal]

function f = fun_objective_SISO_Example(par,par_known,u,Y,ybar,W,Wbar,
    ybarhat_init,N,T,x,z)
Gbb=par(1);
varv=par(2);
bbarhat_init=par(3);
Gaa=par_known(1);
bbarhat=bbarhat_init*ones(1,T);
ybarhatobj=ybarhat_init*ones(1,T);
for t=1:T-1
    ybar(x)=0;
    ybarhat(x)=0;
    bbarhat(x)=bbarhat_init;
    abarhat(x)=ybarhat(x)-bbarhat(x)*u(x);
    Pyy=Gaa+varv+u(t)^2*Gbb;
    covWy=(N-1)*cov(W(:,t),Y(:,t))/N;
    betay=inv(Pyy)*covWy(1,2)/Wbar(t);
    abarhat(t+1)=abarhat(t)+Gaa*betay;
    bbarhat(t+1)=bbarhat(t)+Gbb*u(t)*betay;
    ybarhatobj(t+1)=abarhat(t+1)+bbarhat(t+1)*u(t+1);
end
    f=sum((ybar(x:z)-ybarhatobj(x:z)).^2)/(z+1-x);
end

```